UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

***SOBRE AS PROTEÍNAS
TIROSINA-FOSFATASES.***

*Reatividade intrínseca de ésteres de
fosfato e simulação computacional
dos mecanismos de reação
enzimática*

Guilherme Menegon Arantes

Tese apresentada ao Instituto de Química da Universidade de São Paulo, para concorrer ao Título de Doutor, pelo curso de Pós-Graduação em Bioquímica.

São Paulo
2004

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

***SOBRE AS PROTEÍNAS
TIROSINA-FOSFATASES.***

***Reatividade intrínseca de ésteres de
fosfato e simulação computacional
dos mecanismos de reação
enzimática***

Guilherme Menegon Arantes

Tese apresentada ao Instituto de Química da Universidade de São Paulo, para concorrer ao Título de Doutor, pelo curso de Pós-Graduação em Bioquímica.

Orientador: Prof. Hernan Chaimovich

Coorientador: Prof. Michel Loos

São Paulo
2004

*Aos meus pais, Clélia e Eden.
Obrigado pelo amor, apoio, valores e educação.*

*À Verônica.
Obrigado pelo amor e carinho.*

Agradecimentos

Tive a sorte de conviver com diferentes mentores durante estes anos e aprendi algo com todos eles. Mas, sem dúvida, aquele que mais me influenciou e mais me ensinou foi o Luís Gustavo Dias. Foi ele quem me mostrou o lado teórico de como podemos tentar compreender a natureza, discutiu e explicou-me pacientemente muitas idéias. O Luís também sugeriu uma parte do trabalho desenvolvido nesta tese. Tenho que lhe agradecer tudo isso, mas, principalmente e acima de tudo, a sua amizade. Ao Hernan, gostaria de agradecer a liberdade dada para que eu navegasse por mares desconhecidos pelo grupo e fizesse todo meu trabalho com a independência que se espera de um pesquisador. Também lhe agradeço a total confiança depositada. Fui muito bem recebido pelo Peter R. Taylor quando visitei seu novo grupo na Universidade de Warwick, Reino Unido, e agradeço-lhe por fornecer parte dos recursos computacionais usados nesta tese. Ao Martin J. Field de Grenoble, França, agradeço por tornar sua biblioteca DYNAMO (um belo pedaço de software) disponível livremente e por ter respondido algumas das minhas dúvidas sobre sua utilização. Ao Michel, agradeço a sua sinceridade nas discussões e o incentivo inicial ao uso do L^AT_EX (pelo qual esta tese foi concebida).

Também sou grato à Iolanda pelo suporte inicial quando entrei no grupo e ao Pedro, Maurício e Xuds por formarem um diverso ambiente de pesquisa. Aos colegas do IQ, Flávio, Paula e Guilherme, Teresa, Tereza, David, Larissa, Valéria, Mimi, Fábio, Chien, Greg, Marcos, Tiago Révi e Abacaxi, tenho que agradecer a ótima companhia desde os seminários até as cervejadas.

Finalmente, agradeço à FAPESP as bolsas concedidas e, à pró-reitoria de pós-graduação da USP por uma passagem aérea.

Sumário

Lista de Figuras	p. ix
Lista de Tabelas	p. xii
Lista de abreviaturas, siglas e símbolos	p. xiv
Resumo	p. xvii
Abstract	p. xviii
1 Introdução	p. 1
1.1 Catálise enzimática	p. 1
1.1.1 Descrição macroscópica	p. 2
1.1.2 Forças e mecanismos moleculares	p. 3
1.1.2.1 Estabilização ou desestabilização?	p. 5
1.1.2.2 Modelo do sítio dividido	p. 8
1.1.2.3 NACs e hipótese espaço-temporal	p. 9
1.1.2.4 Catálise covalente	p. 11
1.1.2.5 Catálise ácido-base geral	p. 12
1.1.2.6 Interação eletrostática	p. 13
1.1.2.7 Efeitos entrópicos	p. 14
1.1.2.8 Tensão e impedimento estérico	p. 16
1.1.2.9 Efeitos dinâmicos	p. 17
1.2 Reatividade de ésteres de fosfato	p. 18

1.2.1	Reações em solução	p. 19
1.2.2	Reações em fase gasosa	p. 30
1.3	Proteínas tirosina-fosfatases	p. 32
1.3.1	Papel fisiológico, classificação e estrutura	p. 32
1.3.2	Mecanismo de catálise	p. 35
1.3.3	Análise do mecanismo proposto	p. 38
1.4	Cálculo de estrutura eletrônica molecular	p. 44
1.4.1	Estrutura eletrônica <i>ab initio</i>	p. 45
1.4.1.1	Conjunto base atômico	p. 52
1.4.2	Estrutura eletrônica semiempírica	p. 54
1.5	Otimização de parâmetros	p. 61
1.5.1	Otimização calculando derivadas	p. 61
1.5.2	Otimização sem derivadas	p. 63
1.6	Simulação computacional de macromoléculas	p. 64
1.6.1	Campos de força empíricos	p. 65
1.6.2	Dinâmica molecular clássica	p. 68
1.6.3	Cálculo de energia livre	p. 72
1.6.4	Potenciais híbridos	p. 73
1.6.4.1	Interação eletrostática semiempírica	p. 76
1.6.4.2	Fronteira covalente entre as regiões	p. 78
2	Reatividade intrínseca de ésteres de fosfato	p. 80
2.1	Métodos	p. 80
2.2	Resultados	p. 82
2.2.1	Triésteres de fosfato	p. 84
2.2.1.1	Reação A	p. 85
2.2.1.2	Reações B, C e D	p. 88

2.2.2	Monoésteres de fosfato	p. 93
2.2.2.1	Reações E, F, G e H	p. 94
2.2.3	Conjunto base para expansão dos orbitais moleculares	p. 102
2.2.4	Comparação com medidas experimentais	p. 103
2.3	Discussão	p. 108
2.3.1	Calibração da metodologia de cálculo <i>ab initio</i>	p. 108
2.3.2	Tiólise e alcoólise de ésteres de fosfato	p. 116
2.3.3	Modelos para os mecanismos enzimáticos	p. 127
3	Calibração da hamiltoniana híbrida	p. 131
3.1	Métodos	p. 133
3.1.1	Algoritmo de otimização: GA+simplex	p. 133
3.1.2	Reparametrização da hamiltoniana semiempírica	p. 134
3.1.3	Interação não ligante	p. 139
3.1.3.1	Complexos bimoleculares	p. 139
3.1.3.2	Energia de complexação	p. 142
3.1.3.3	Calibração específica dos parâmetros	p. 143
3.1.4	Fronteira e interação covalente	p. 144
3.2	Resultados e discussão	p. 146
3.2.1	Parâmetros semiempíricos específicos	p. 146
3.2.2	Propriedades obtidas pela hamiltoniana QM calibrada	p. 154
3.2.3	Interação não ligante	p. 158
3.2.3.1	Definição da interação eletrostática	p. 158
3.2.3.2	Parâmetros específicos	p. 161
3.2.4	Teste das interações de fronteira	p. 167
4	Simulação das reações catalisadas por PTPs	p. 170

4.1	Métodos	p. 171
4.1.1	Construção das estruturas enzimáticas	p. 171
4.1.2	Potencial híbrido e dinâmica molecular	p. 175
4.1.3	Potencial de força média	p. 178
4.2	Resultados	p. 179
4.2.1	Simulações da fosfatase de proteínas VHR	p. 180
4.2.1.1	Possíveis coordenadas de reação	p. 180
4.2.1.2	Fenilfosfato ou metilfosfato como substrato	p. 182
4.2.1.3	Mutantes ASP→ALA e ARG→ALA	p. 185
4.2.1.4	Protonação do substrato ou do nucleófilo	p. 187
4.2.1.5	Mutante CYS→SER	p. 189
4.2.2	Simulações da fosfatase de proteínas CDC25B	p. 194
4.2.2.1	Estado de protonação do substrato	p. 194
4.2.2.2	Ácido glutâmico do P-loop como ácido geral	p. 198
4.3	Discussão	p. 203
4.3.1	Incertezas nas simulações	p. 203
4.3.2	Escolha da coordenada de reação	p. 209
4.3.3	Mecanismo da reação no sítio ativo das PTPs	p. 213
4.3.3.1	Simulações da catálise por uma LM-PTP	p. 219
4.3.4	Considerações das reações nos mutantes e de metilfosfato	p. 222
4.3.5	Qual é o estado de protonação do substrato?	p. 224
4.3.6	Por que o mutante CYS→SER é inativo?	p. 226
4.3.7	Qual é o ácido geral na CDC25B?	p. 230
5	Conclusões	p. 234
	Apêndice A – Integrais semiempíricas de repulsão eletrônica	p. 242

Apêndice B – Cálculo de propriedades termodinâmicas	p. 245
--	--------

Referências Bibliográficas	p. 247
-----------------------------------	--------

Lista de Figuras

1	Esquema do perfil energético de uma reação enzimática e sua referência em solução	p. 5
2	Esquema do perfil energético de reações enzimáticas para MCs com diferentes estabilidades	p. 7
3	Esquema dos mecanismos propostos para transferência de fosfato . . .	p. 19
4	Estrutura de uma bipirâmide trigonal	p. 21
5	Estrutura do monoéster de fosfato na dissociação assistida por água . .	p. 25
6	Esquema da reação de tiólise intramolecular de um tio nucleotídeo . . .	p. 30
7	Esquema das reações de tiólise (A e B) e alcoólise (C e D) de triésteres de fosfato	p. 33
8	Esquema das reações de tiólise (E e F) e alcoólise (G e H) de monoésteres de fosfato	p. 34
9	Representação da estrutura terciária da VHR	p. 35
10	Estrutura do sítio ativo do MC da enzima VHR	p. 36
11	Esquema do mecanismo de reação proposto para catálise pelas PTPs .	p. 37
12	Representação da partição do sistema macromolecular em regiões QM e MM	p. 75
13	Estruturas otimizadas no nível MP2/6-311++G* para reação A	p. 83
14	Caminho mínimo de energia MP2/6-31G* para a reação A	p. 86
15	Perfil de energia total relativa para as reações A e B	p. 88
16	Estruturas do B:I e D:I otimizadas no nível MP2/6-31+G*	p. 90
17	Perfil de energia total relativa para as reações C e D	p. 92
18	Energia HF/6-31+G* para a dissociação de $(\text{CH}_3\text{O})^-$ do TMP	p. 92

19	Estruturas do G:IMC1a, G:IMC1d e G:Ia otimizadas no nível MP2/6-31+G*	p. 94
20	Variação das distâncias interatômicas ao longo da IRC da reação entre CH ₃ OH e PO ₃ ⁻	p. 96
21	Perfil de energia total relativa para a reação E	p. 98
22	Perfil de energia total relativa para a reação F	p. 99
23	Perfil de energia total relativa para a reação G	p. 100
24	Perfil de energia total relativa para a reação H	p. 101
25	Energias absolutas para segunda etapa da reação F (D _n +A _n) em função do conjunto base	p. 104
26	Complexos bimoleculares usados na calibração de $E_{QM/MM}$	p. 141
27	Energia de complexação entre fenol e metanol obtida por diferentes hamiltonianas	p. 159
28	Energia de complexação entre formamida e metafosfato obtida por diferentes hamiltonianas	p. 160
29	Energia de complexação entre formamida e metafosfato obtida com diferentes parâmetros de LJ para parte QM	p. 163
30	Energia de complexação entre guanidina e metafosfato obtida com diferentes parâmetros de LJ para parte QM	p. 164
31	Energia de complexação entre água e fenol, e entre água e ácido acético	p. 166
32	Energias de dissociação para CYS-PO ₃ H ⁻ e SER-PO ₃ H ⁻	p. 169
33	Representação das regiões usadas nas simulações enzimáticas	p. 175
34	Esquema das distâncias e grupos reativos usados nas simulações enzimáticas.	p. 176
35	Perfil de energia livre para diferentes coordenadas da reação de PhOP ²⁻ catalisada pela VHR	p. 181
36	Perfil de energia livre para reação de PhOP ²⁻ e MeOP ²⁻ catalisada pela VHR	p. 184

37	Estrutura do sítio ativo no TS da reação de fenilfosfato catalisada pela VHR	p. 184
38	Perfil de energia livre para reação de PhOP^{2-} catalisada pelos mutantes ASP→ALA e ARG→ALA da VHR	p. 186
39	Perfil de energia livre para reação com os grupos reativos protonados catalisada pela VHR	p. 188
40	Perfil de energia livre para reação de PhOP^{2-} catalisada pela VHR mutante CYS→SER ionizada	p. 191
41	Perfil de energia livre para reação de PhOP^{2-} catalisada pela VHR mutante CYS→SER protonada	p. 192
42	Perfil de energia livre para reação de PhOP^{2-} e PhOP^- catalisada pela CDC25B	p. 195
43	Perfil de energia livre para aproximação entre a cadeia lateral do GLU no P-loop e o substrato na CDC25B	p. 199
44	Perfil de energia livre para reação de PhOP^{2-} catalisada pela CDC25B, com GLU no P-loop como ácido geral	p. 201
45	Distribuição dos valores de coordenada de reação	p. 210
46	Exemplo de configuração de carga para cálculo de integrais semiempíricas de repulsão eletrônica	p. 243

Lista de Tabelas

1	Efeito isotópico cinético em reações de hidrólise de ésteres de fosfato em solução	p. 23
2	Constantes de velocidade medidas para a reação de pNPP catalisada por PTPs	p. 38
3	Efeito isotópico cinético em reações de hidrólise de pNPP catalisada por PTPs	p. 40
4	Energias e entropias relativas para a reação A	p. 87
5	Energias e entropias relativas para a reação B	p. 90
6	Energias e entropias relativas para a reação C	p. 91
7	Energias e entropias relativas para a reação D	p. 91
8	Caracterização dos TS e da participação da transferência de H^+ na coordenada de reação	p. 95
9	Energias e entropias relativas para a reação E	p. 98
10	Energias e entropias relativas para a reação F	p. 99
11	Energias e entropias relativas para a reação G	p. 100
12	Energias e entropias relativas para a reação H	p. 101
13	Energias relativas para segunda etapa da reação F ($D_n + A_n$) em função do conjunto base	p. 104
14	Energias relativas para reação C em função do conjunto base	p. 105
15	Energias relativas para segunda etapa da reação E $A_n + D_n$ em função do conjunto base	p. 105
16	Energias das reações de transferência de H^+ em fase gasosa calculadas e experimentais	p. 106

17	Comparação entre medidas experimentais e valores calculados em fase gasosa	p. 107
18	RMSD máximo e médio das coordenadas internas para as espécies reotimizadas no nível B3LYP/6-31+G(d)	p. 109
19	RMSD das coordenadas internas e <i>ZPE</i> para as espécies reotimizadas no nível B3LYP/6-31+G(d)	p. 110
20	Parâmetros semiempíricos otimizados: fase I	p. 134
21	Parâmetros semiempíricos otimizados: fase II	p. 135
22	Parâmetros semiempíricos originais para os elementos H, C, O, P e S	p. 136
23	Parâmetros semiempíricos MNDO+G/CHOPS para os elementos H, C, O, P e S	p. 152
24	Desvios das propriedades calculadas pelos métodos semiempíricos	p. 155
25	Parâmetros de LJ e cargas parciais do campo de força OPLS-AA e calibrados	p. 162
26	Energias relativas calculadas para as reações 3.18 a 3.21	p. 167
27	Grupos reativos ou catalíticos e coordenadas de reação usadas nas simulações enzimáticas	p. 176
28	Energia livre de ativação calculada para as diferentes escolhas de coordenada de reação	p. 181
29	Distâncias calculadas para reação de PhOP^{2-} catalisada pela VHR	p. 183
30	Energia livre de reação e de ativação, e coordenada de reação calculadas para catálise pela VHR	p. 183
31	Energia livre de ativação e coordenada de reação calculadas para catálise pela VHR mutante CYS124→SER	p. 190
32	Distâncias calculadas para reação catalisada pela VHR mutante CYS124→SER	p. 193
33	Energia livre de reação e de ativação, e coordenada de reação calculadas para a catálise pela CDC25B	p. 195
34	Distâncias calculadas para reação de PhOP^{2-} catalisada pela CDC25B	p. 197

Lista de abreviaturas, siglas e símbolos

‡	Complexo ativado ou estado de transição,	p. 3
§	Seção,	p. 2
ξ	Coordenada de reação,	p. 72
ALA	Alanina,	p. 38
ARG	Arginina,	p. 36
ASN	Asparagina,	p. 41
ASP	Ácido aspártico,	p. 39
ATP	Trifosfato de adenosina,	p. 32
au	Unidades atômicas,	p. 45
bohr	Unidade atômica de distância,	p. 77
BSSE	Erro de superposição de conjunto base,	p. 54
cap.	Capítulo,	p. 1
CDK	Quinase dependente de ciclina,	p. 35
CPU	Unidade central de processamento,	p. 81
CYS	Cisteína,	p. 36
d(XY)	Distância entre os centros X e Y,	p. 82
DFT	Teoria do funcional da densidade,	p. 51
DMPP	Dimetil fenilfosfato,	p. 33
DMSO	Dimetil sulfóxido,	p. 25
DRC	Coordenada de reação distinta,	p. 82
DSP	Fosfatases de dupla especificidade,	p. 33
e	Carga do elétron,	p. 144
eV	Elétron-volt,	p. 107
$E^{(n)}$	Energia da perturbação de n -ésima ordem,	p. 50
EVB	Ligação de valência empírica,	p. 43

GA	Algoritmo genético,	p. 61
GLN	Glutamina,	p. 233
GLU	Ácido glutâmico,	p. 42
GLY	Glicina,	p. 171
h	Constante de Planck,	p. 18
HF	Hartree-Fock,	p. 48
HIS	Histidina,	p. 39
HOMO	Orbital molecular ocupado de mais alta energia,	p. 137
IMC	Complexo íon-molécula,	p. 84
IRC	Coordenadas intrínsecas de reação,	p. 81
IUPAC	União internacional de química pura e aplicada,	p. 19
k_B	Constante de Boltzmann,	p. 18
k_{obs}	Constante de velocidade observada,	p. 30
KIE	Efeito isotópico cinético,	p. 22
LFER	Relação linear de energia livre,	p. 20
lg	Grupo de saída,	p. 23
LJ	Lennard-Jones,	p. 66
LM-PTP	Fosfatase de tirosina de proteínas de baixo peso molecular,	p. 33
LYS	Lisina,	p. 172
MC	Complexo de Michaelis,	p. 2
MeOP ⁻	Metilfosfato monoaniônico,	p. 34
MeOP ²⁻	Metilfosfato dianiônico,	p. 128
MeSP ⁻	Tiometilfosfato monoaniônico,	p. 94
MET	Metionina,	p. 198
MM	Molecular mecânico,	p. 75
MP	Møller-Plesset,	p. 49
NAC	Conformação próxima ao ataque,	p. 9
NDDO	Negligência do recobrimento diferencial diatômico,	p. 54
NOX	Número de oxidação,	p. 151
O1	Oxigênio do grupo de saída,	p. 182
O2, O3 e O4	Oxigênios não ligantes do fosfato,	p. 182
PBC	Condições de contorno periódicas,	p. 71

PDB	Banco de dados de proteínas,	p. 171
PES	Superfície de energia potencial,	p. 45
PHE	Fenilalanina,	p. 171
PhOH	Fenol,	p. 30
PhOP ⁻	Fenilfosfato monoaniônico,	p. 34
PhOP ²⁻	Fenilfosfato dianiônico,	p. 128
PI	Potencial de ionização eletrônica,	p. 107
PMF	Potencial de força média,	p. 72
pNPP	<i>p</i> -nitrofenil fosfato,	p. 23
PTP	Proteína tirosina-fosfatase,	p. 1
QM	Quanto-mecânico,	p. 75
QM/MM	Híbrido de mecânica quântica e mecânica molecular,	p. 74
R_e	Distância de equilíbrio,	p. 143
$\mathbf{R}_e$	Posições atômicas de um ponto estacionário,	p. 62
RMN	Ressonância magnética nuclear,	p. 30
RMSD	Desvio médio quadrático,	p. 108
SER	Serina,	p. 36
T	Temperatura,	p. 18
TBP	Bipirâmide trigonal,	p. 20
THR	Treonina,	p. 36
TMDMP	Tiometil dimetil fosfato,	p. 85
TMP	Trimetil fosfato,	p. 33
TS	Estado de transição,	p. 3
TYR	Tirosina,	p. 171
VAL	Valina,	p. 39
ZPE	Energia do ponto-zero,	p. 81

Resumo

Proteínas tirosina-fosfatases (PTPs) catalisam a hidrólise de fosfotirosina de outras proteínas e, assim, regulam importantes processos bioquímicos. Dois representantes desta família são as fosfatases de dupla especificidade VHR e CDC25B. A primeira etapa de reação catalisada é um ataque nucleofílico da cadeia lateral de uma cisteína sobre o fósforo do substrato, com uma possível transferência de H^+ de um ácido geral para o grupo de saída, formando uma PTP intermediária tiofosforilada e desfosforilando o substrato. Dúvidas ainda persistem sobre esta etapa, envolvendo os estados de protonação do substrato e do nucleófilo enzimático, a inatividade de certos mutantes e a identificação do ácido geral nas CDC25s. Procuramos solucionar estas questões por simulações computacionais das reações catalisadas pela VHR e pela CDC25B. Inicialmente, caminhos das reações de tiólise e alcoólise de ésteres de fosfato na fase gasosa foram determinados por cálculos de estrutura eletrônica *ab initio* e analisados como referências da reatividade intrínseca de ésteres de fosfato e como modelos da atividade enzimática. Um potencial híbrido de mecânica quântica e mecânica molecular foi amplamente testado e calibrado, empregando estes caminhos de reação e outros dados *ab initio*. A calibração permitiu que conclusões semiquantitativas pudessem ser obtidas a partir das simulações enzimáticas. Potenciais de força média foram determinados com o potencial híbrido para desfosforilação de diferentes substratos catalisada pelas PTPs selvagens e por suas mutantes. Os resultados mostram que o mecanismo da reação catalisada segue uma adição e eliminação simultânea, com um estado de transição dissociativo com caráter de metafosfato. As barreiras calculadas são bastante próximas às energias de ativação experimentais. O substrato enzimático é um diânion desprotonado e o nucleófilo está ionizado. As reações do substrato ou do nucleófilo protonados apresentam barreiras, no mínimo, 15 kcal/mol mais altas que os valores experimentais. A VHR mutante cisteína→serina no sítio ativo é inativa, porque a serina está protonada. As CDC25s não utilizam um ácido geral para catálise, ao contrário das outras PTPs. Propostas de que o ácido geral poderia ser o próprio substrato ou um dos ácidos glutâmicos presentes no sítio ativo são energeticamente inacessíveis.

Abstract

Protein tyrosine phosphatases (PTPs) catalyse the hydrolysis of phosphotyrosine from other proteins and, hence, regulate important biochemical processes. Two members from this family are the dual specificity phosphatases VHR and CDC25B. The first step of the catalysed reaction is the nucleophilic attack from the side chain of a cysteine towards the substrate, with a possible H^+ transfer from a general acid to the leaving group, forming a PTP thiophosphorylated intermediate and dephosphorylating the substrate. There are still some doubts about this step, involving the protonation states of the substrate and the enzymatic nucleophile, the inactivity of certain mutants and the identification of the general acid in CDC25s. We tried to solve this questions by computer simulations of the reactions catalysed by VHR and by CDC25B. Initially, reaction pathways of phosphate esters thiolysis and alcoholysis in the gas-phase were determined by *ab initio* electronic structure calculations and analysed as benchmarks for the intrinsic reactivity of phosphate esters and as models of the enzymatic activity. A hybrid potential of quantum mechanics and molecular mechanics was fully tested and calibrated, employing these reaction pathways and other *ab initio* data. The calibration allowed that semiquantitative conclusions could be obtained from the enzymatic simulations. Potentials of mean force were determined with the hybrid potential for the dephosphorylation of different substrates catalysed by the wild-type PTPs and their mutants. The results show that the catalysed reaction mechanism follows a concerted addition and elimination, with a dissociative transition state with metaphosphate-like character. The calculated barriers are very close to the experimental activation energies. The enzymatic substrate is a deprotonated dianion and the nucleophile is ionised. The reactions of the protonated substrate or nucleophile have barriers, at least, 15 kcal/mol higher than the experimental results. The active site cysteine→serine VHR mutant is inactive because the serine is protonated. The CDC25s do not employ a general acid for catalysis, differently from the other PTPs. Proposals that the general acid is the substrate or one of the glutamic acids present in the active site are not energetically accessible.

Raízes - Sem elas você seria como uma árvore em um vento forte, facilmente derrubada. Ritmos e música são a única ligação sobrevivente ao que somos e a única voz de nossos antepassados, e alma é a continuação de suas experiências. Sem raízes, música e alma nós seríamos cegos, perdidos, sem futuro e incapazes de superar o que enfrentamos diariamente.

O Escritor Desconhecido. UR-042. Detroit, EUA, 1997.

1 *Introdução*

Este capítulo está dividido em seis seções. Nas três primeiras, dados obtidos experimental e computacionalmente foram incluídos e apresentados conjuntamente, pois a interação entre ambos tem gerado interessantes discussões em relação aos diferentes mecanismos propostos sobre: os paradigmas atuais dos mecanismos de catálise enzimática; a reatividade de ésteres de fosfato em solução e em fase gasosa e a catálise na superfamília das proteínas tirosina-fosfatases (PTP). A quarta, quinta e sexta parte contêm uma breve introdução das idéias que sustentam os cálculos de estrutura eletrônica usados para a obtenção de caminhos de reações de ésteres de fosfato, e as simulações de reações químicas em macromoléculas usadas para estudar os mecanismos de catálise nas PTPs.

Os três capítulos (cap.) posteriores são divididos em partes metodológicas, de apresentação dos resultados e de discussões dos três tópicos tratados nesta tese:

- Reatividade intrínseca de ésteres de fosfato;
- Calibração específica do potencial híbrido;
- Simulações híbridas dos mecanismos de reação nas PTPs.

O último capítulo apresenta as conclusões.

Uma nota sobre a forma das referências citadas: referências na linha de texto e, normalmente, entre parênteses correspondem às partes desta tese; referências sobrescritas e entre colchetes correspondem às citações bibliográficas.

1.1 *Catálise enzimática*

Enzimas são catalisadores de extraordinária eficiência. As acelerações de velocidade de reações realizadas por enzimas podem chegar a vinte ordens de magnitude^[1]! Por outro lado, enzimas possuem uma grande especificidade pelos substratos e sua atividade é controlável e totalmente seletiva.

O modelo e as equações básicas postuladas para a atividade enzimática macroscópica serão revisados na seção 1.1.1 (§1.1.1) e uma parte dos mecanismos propostos para explicar a catálise num nível microscópico será apresentada na §1.1.2. Há muita controvérsia sobre a importância de cada um dos diferentes fatores e mecanismos propostos para explicar a catálise. Cerca de 21 teorias sobre catálise enzimática foram contabilizadas há 10 anos^[2]. Este número elevado é consequência de diferentes interpretações semânticas e imprecisas definições de quantidades termodinâmicas nas teorias propostas, mas, principalmente, porque a maior parte destas teorias não foi (ou não pode ser) testada quantitativamente. Poucas (ou quase nenhuma) enzimas tiveram seu mecanismo molecular analisado quantitativamente e aceito por todos pesquisadores da área^[3]. A definição precisa de quantidades e, principalmente, de ciclos termodinâmicos^[4] que possam ser testados experimental ou computacionalmente é necessária para explicar os mecanismos microscópicos da catálise.

Apenas algumas das teorias propostas serão apresentadas aqui. A escolha é incompleta e parcial, pois reflete as propostas que o autor considera mais relevantes para catálise enzimática em geral e para as PTPs, bem como aquelas que foram testadas e quantificadas por simulações computacionais.

1.1.1 Descrição macroscópica

O modelo básico usado para descrever cinética enzimática para um único substrato, S , é o mecanismo de Michaelis-Menten^{[5, cap. 3][6, §III.13]}:

$$K_S = \frac{[E][S]}{[E \cdot S]} = \frac{k_{-1}}{k_1} \quad (1.2)$$

$$K_M = \frac{k_{-1} + k_{cat}}{k_1} \simeq K_S \Leftrightarrow k_{-1} \gg k_{cat} \quad (1.3)$$

onde E denota enzima, P é o produto, K_S é a constante de equilíbrio para dissociação do complexo de Michaelis (MC), $E \cdot S$, gerando a enzima e o substrato livres, k_{cat} é a constante de velocidade de reação observada e K_M é a constante de Michaelis. Na maioria dos casos, $k_{-1} \gg k_{cat}$ e, portanto, $K_M \simeq K_S$ (equação 1.3). Neste modelo, a velocidade inicial de formação do produto, $v = k_{cat} [E \cdot S]$, também pode ser escrita como:

$$v = \frac{k_{cat}[E]_T[S]}{K_M + [S]} \quad (1.4)$$

onde definimos a concentração total de enzima, $[E]_T = [E] + [E \cdot S]$. Uma equação com a mesma forma também é válida para mecanismos mais complexos, mas com diferentes definições da constante de equilíbrio K_S .

O seguinte ciclo termodinâmico pode ser usado para comparar a velocidade de reação catalisada com a reação em solução aquosa sem catálise, k_{non} , caso as duas reações procedam-se segundo o mesmo mecanismo [logo, pelo mesmo estado de transição (TS)] [7, cap. 6].

A velocidade de reação na enzima será maior do que a reação não catalisada em solução se $k_{cat}/K_M > k_{non}$. Neste caso, a afinidade da enzima pelo TS ou complexo ativado $S^\ddagger$, $K_{TS} = K_M/(k_{non}/K_{cat}) > 0$ ou o complexo ativado é mais estável quando ligado à enzima do que quando em solução.

Considerando este modelo macroscópico, existem duas maneiras da evolução maximizar a velocidade de catálise aparente ($\equiv k_{cat}/K_M$) e a seletividade enzimática em relação a reação não catalisada (equação 1.5). A alteração da K_M para um determinado substrato pode acelerar a velocidade de catálise aparente pelo aumento da $[E \cdot S]$, além de modificar a seletividade por este substrato. Um dos possíveis mecanismos empregados é o “encaminhamento”, em que o volume livre para difusão do substrato é reduzido ou por mudanças estruturais na enzima (como em complexos enzimáticos multiunidades) ou por interações eletrostáticas que guiam o substrato para o sítio ativo^[3]. A outra maneira de acelerar a velocidade de catálise, sem alterar a seletividade pelo substrato, é o aumento da k_{cat} , ou seja, acelerar a etapa de transformação química na enzima (equação 1.5).

Não discutiremos processos que alterem apenas K_M ou a associação entre enzima e substrato. Nosso interesse, ao longo dos próximos parágrafos e capítulos, está focalizado nos processos que aumentam a velocidade de catálise, principalmente pelo aumento de k_{cat} .

1.1.2 Forças e mecanismos moleculares

Em 1969, Jencks^[8, p. 6] escreveu que “o estudo dos mecanismos moleculares de catálise enzimática é necessariamente empírico e qualitativo”. No entanto, simulações computa-

cionais que permitem determinações quantitativas de propriedades termodinâmicas têm alterado este panorama e mecanismos especiais ou “misteriosos” têm sido descartados^[9], já que as simulações reproduzem as observáveis experimentais sem recorrer a tais mecanismos.

Ao contrário de medidas experimentais, simulações computacionais de reações enzimáticas^[3, 10, 11] podem ser usadas para desmontar as observáveis termodinâmicas obtidas em interações e forças definidas e quantificáveis.

A análise do efeito da enzima e seus mecanismos moleculares de catálise por um ciclo termodinâmico, como a equação 1.5, requer que o mecanismo no sítio ativo seja comparado com o mesmo mecanismo num estado de referência, *i.e.*, um microagregado ou complexo de encontro, contendo os reagentes solvatados^[4] (RS, figura 1). O mecanismo operante na reação catalisada pode ser energeticamente inacessível ou, no mínimo, diferente do mecanismo dominante da reação em solução na ausência do catalisador. Nestes casos, a comparação direta entre o mecanismo enzimático e aquele observado em solução não pode ser usada para o estudo dos fatores que influenciam a catálise. Simulações computacionais também podem calcular o perfil energético das reações de referência que não são acessíveis experimentalmente em solução^{[10, cap. 3][11, 12]}.

O mecanismo proposto e aceito por grande parte da comunidade como principal responsável pela catálise enzimática é a estabilização eletrostática do TS^[4, 13]. Contudo, efeitos eletrostáticos não são sempre dominantes (por exemplo, reações de espécies radiculares estão menos sujeitas à catálise por este mecanismo)^[9]. Sem dúvida, cada enzima possui sua própria “receita”^[2], em que uma combinação dos fatores propostos nas §1.1.2.1 a §1.1.2.9 pode ser responsável pela catálise. Portanto, o interesse desta seção é analisar e, se possível, quantificar a contribuição de cada possível mecanismo para o abaixamento da energia de ativação $\Delta G_{enz}^\ddagger$ (figura 1) em relação a $\Delta G_{sol}^\ddagger$ (que pode chegar a 30 kcal/mol!)^[1].

Duas perguntas básicas podem auxiliar a compreensão dos mecanismos moleculares empregados por enzimas:

1. O que sofre catálise?
2. Como é a catálise?

A primeira pergunta refere-se a espécie, ao longo da coordenada de reação, que sofre influência determinante do ambiente enzimático. Possíveis respostas serão apresentadas

Figura 1: Esquema do perfil energético de uma reação enzimática e sua referência em solução. ΔG_B , ΔG_{comp} , $\Delta G_{enz}^\ddagger$ e $\Delta G_{sol}^\ddagger$ são respectivamente a energia livre de formação do MC (ES) entre a enzima E e o substrato S, de complexação dos reagentes R e S em solução, de ativação da reação no ambiente enzimático e de ativação da reação no complexo RS dos reagentes em solução. A região com linhas diagonais indica o solvente.

nas §1.1.2.1, §1.1.2.2 e §1.1.2.3. Entretanto, estudar qual a espécie que sofre a catálise não responde a segunda pergunta. As possíveis forças e processos moleculares responsáveis pela catálise serão apresentados nas seções entre §1.1.2.4 a §1.1.2.9.

1.1.2.1 Estabilização ou desestabilização?

A sugestão inicial de Fischer^[14] para a atividade enzimática, conhecida como a hipótese da “chave e fechadura”, foi extendida pelas propostas de Haldane^[14], em que o substrato não “encaixa” exatamente na enzima, mas sofre uma distorção após a formação do complexo $E \cdot S$, e de Pauling^[15], em que a enzima se liga mais fortemente ao TS do que ao substrato.

A proposta de Haldane, que é equivalente a desestabilização do substrato em relação a espécie livre em solução, é defendida por Jencks^[8, cap. 5 e apêndice] e Menger^[2], entre outros^[16]. Porém, nenhum experimento quantificou o efeito da possível desestabilização.

Portanto, esta proposta é fundamentada em especulações^[8, p. 794] e na interpretação de experimentos de reatividade de compostos orgânicos em solução^{[8, cap. 5][17]}. A hipótese de Pauling apóia-se no conceito de estado de transição^[7, cap. 6] e diz que este é mais estabilizado no sítio ativo da enzima em relação a mesma espécie em solução aquosa (ou numa reação de referência, vide supra). Esta hipótese é amplamente aceita^[3, 10, 11, 14, 18]. Centenas de estruturas tridimensionais determinadas de enzimas complexadas a análogos de TS apóiam a noção de complementaridade entre as estruturas do sítio ativo enzimático e do TS. Estes complexos são de grande utilidade prática no desenho de drogas que funcionam como inibidores enzimáticos competitivos^[6, p. 819]. No entanto, comprovar a hipótese de Pauling por experimentos também não é diretamente possível, já que TS são construções mentais (da teoria de estado de transição) e, logo, não podem ser isolados ou observados.

Claramente, as propostas de Haldane e Pauling implicam que a interação da enzima com o substrato no E·S é diferente da interação no E·S[‡]. Simulações computacionais de reações enzimáticas mostram repetidamente^[3, 9, 16, 19] que a interação eletrostática (§1.1.2.6) entre o ambiente enzimático e o TS das reações é a maior responsável pela diminuição de $\Delta G_{enz}^{\ddagger}$ em relação a $\Delta G_{sol}^{\ddagger}$ (figura 1), sugerindo que o TS é a espécie que sofre catálise ao ser estabilizado eletrostaticamente, em acordo com a hipótese de Pauling. Desta maneira, a enzima apenas estabilizaria o TS equivalente àquele atingido na reação não catalisada. Porém, além da estabilização, a enzima pode *alterar* o TS e até selecionar um mecanismo de reação diferente daquele observado em outros microambientes. Neste caso, a determinação dos fatores ou das forças moleculares responsáveis pela catálise deve ser feita pela comparação do mecanismo de reação na enzima e do *mesmo* mecanismo numa reação de referência (vide supra). Já o desenho de drogas análogas ao TS deve ser orientado pelo mecanismo e o TS atingidos na enzima^[20].

Em relação ao perfil de energia livre de ligação e reação representado pela curva A na figura 2, a desestabilização do substrato pode ser representada pelos perfis B, C e D. Uma alteração estrutural da proteína anterior à ligação do substrato, por exemplo, em consequência da ligação de um regulador alostérico ou de uma mutação no sítio ativo, pode resultar em aumento da K_M (perfil B, figura 2). Uma alteração estrutural expressiva resultaria em tamanha desestabilização do MC que $K_M > 1$ (perfil C, figura 2) e corresponderia ao perfil de energia livre em que $[S] \ll K_M$ ^[5, cap. 12]. A desestabilização também pode ser consequência de uma mudança estrutural na enzima^[8, cap. 5] após a formação do complexo de Michaelis ES, gerando o complexo ES' (perfil D, figura 2). Jencks propôs que a energia de ligação do substrato observada nos processos representados pelos perfis B e D é $\Delta G_B - \Delta\Delta G_B$, ou seja, a diferença entre a energia de ligação intrínseca e a energia

Figura 2: Esquema do perfil de energia livre de reações enzimáticas para MCs com diferentes estabilidades. ΔG_B , $\Delta\Delta G_B$, $\Delta G_{enz}^\ddagger$ e ΔG_{cat} são respectivamente a energia livre de formação do MC, de desestabilização do MC, de ativação na enzima ($\Delta G_{enz}^\ddagger = -RT\ln[k_{cat}/A]$, onde A é o fator pré-exponencial) e de ativação aparente ($\Delta G_{cat} = -RT\ln[\frac{k_{cat}/A}{K_M}]$).

usada para desestabilizar o substrato. A quantidade $\Delta\Delta G_B$ seria usada para reduzir a barreira de ativação química na enzima^[8, p. 755].

Pela proposta de Jencks, a desestabilização ocorre no MC (ES) e não (ou pouco) altera a energia do TS^[8, p. 720]. Portanto, a energia de ativação aparente (ΔG_{cat}) é mantida e a energia de ativação na enzima é modificada para cada perfil, ou seja, $\Delta G_{enz}^\ddagger(A) > \Delta G_{enz}^\ddagger(D) > \Delta G_{enz}^\ddagger(B) > \Delta G_{enz}^\ddagger(C)$ (figura 2). Esta proposta implica que enzimas têm alta afinidade intrínseca pelo substrato.

A análise do efeito da evolução separadamente sobre k_{cat} e K_M de diversas enzimas e de processos hipotéticos dentro do modelo de Michaelis-Mentem (equação 1.1) com a relação k_{cat}/K_M fixada, indica que enzimas evoluíram para ligar os substratos fracamente (alto K_M) e o TS fortemente (baixo $\Delta G_{enz}^\ddagger$ ou alto k_{cat})^[5, §12.C]. A desestabilização do MC (perfis B e D da figura 2) não muda $\Delta G_{cat} = RT\ln[\frac{k_{cat}/A}{K_M}]$, portanto não é claro como esta proposta pode ser usada pela evolução para *variar ou maximizar a relação k_{cat}/K_M* ^[4].

A importância da desestabilização do substrato para catálise deve ser limitada, pois o aumento da energia do complexo ES acima da energia de E+S (perfil C, figura 2) não é catalítico, porque a velocidade máxima de uma reação reflete a diferença entre as energias livres do ponto mais alto ($ES^\ddagger$) e do mais baixo (E+S, neste caso) ao longo do perfil de

reação. A quantidade de energia disponível para catálise pela desestabilização ($\Delta\Delta G_B$) parece ser pequena em comparação com a diferença entre as barreiras da reação catalisada e em solução ($\Delta G_{sol}^\ddagger - \Delta G_{enz}^\ddagger$, vide supra)^[4].

Enzimas mutantes cujo ΔG_{cat} é significativamente diferente da enzima selvagem podem ser divididas em três classes: mutantes que sofrem alteração da energia do TS, resultando em diminuição apenas de k_{cat} e manutenção de K_M ^[21] (*i.e.*, $\Delta G_{enz}^\ddagger$ é aumentado, pois o TS é desestabilizado); mutantes que sofrem modificação de um sítio de ligação distante dos grupos reativos do substrato, resultando em diminuição do K_M , mantendo k_{cat} (ou seja, o TS e o MC têm suas energias aumentadas pela mesma quantidade, logo, $\Delta G_{enz}^\ddagger$ não muda)^[4]. Não foram encontrados^[13] mutantes da terceira classe, em que ambos K_M e k_{cat} seriam modificados (ou seja, $\Delta G_{enz}^\ddagger$ e ΔG_B apresentariam aumentos da mesma magnitude). Este seria o resultado esperado para uma mutação que anule ou diminua a capacidade de desestabilização do substrato pela enzima selvagem.

Em resumo, as evidências computacionais existentes e a análise das implicações da proposta de desestabilização não suportam esta hipótese como o principal mecanismo de catálise enzimática.

A afirmação de “maior ligação ou afinidade da enzima pelo TS” é por vezes interpretada literalmente, como se o TS fosse atingido em solução e, depois, ligado e estabilizado pela enzima. Esta interpretação é equivocada porque o TS tem um tempo de vida muito pequeno em solução, da ordem de 10^{-15} s e, portanto, não existe por um tempo suficiente para que difusão e ligação à enzima ocorram. Na interpretação correta, a enzima complexa o substrato no estado fundamental e a ativação ocorre a partir do MC. O TS formado possui melhores contatos com a enzima do que o substrato, maximizando a estabilização apenas quando o complexo é ativado. Assim, o TS tem maior estabilidade no (ou afinidade pelo) sítio ativo do que em solução (ou numa reação de referência não catalisada, figura 1).

1.1.2.2 Modelo do sítio dividido

Como a velocidade de uma reação reflete a maior diferença ao longo do perfil de energia livre, a formação do complexo ES é irrelevante (perfil C, figura 2) ou anticatalítica (perfis A, B e D) para velocidade de reação enzimática. Então, por que um MC é formado em todas as reações enzimáticas? A regulação e a seletividade enzimática são necessárias e podem ser determinadas pela formação de ES ^[5, p. 364]. Portanto, a evolução não maximizou apenas o fluxo da reação $E + S \rightleftharpoons ES^\ddagger$ para direita.

Menger^[2] propôs uma função catalítica para o MC. No seu modelo de sítio dividido, o substrato pode ser dividido em sítios ou regiões de ligação (lig) e sítios ou regiões de reação (reac). As energias livres do MC e TS são descritas pela soma das energias de cada região, $\Delta G = \Delta G(\text{lig}) + \Delta G(\text{reac})$. As seguintes hipóteses estão no cerne do modelo:

1. A energia de formação do MC proveniente da região ligante, $\Delta G_B(\text{lig})$, é conservada quando o TS é alcançado, já o $\Delta G_B(\text{reac})$ não é conservado no TS, porque a estrutura molecular nesta região é modificada;
2. As forças operantes na região de ligação são sempre atrativas (estabilizantes) e as forças na região reativa são desestabilizantes.

Assim, o modelo de Menger defende que a desestabilização da parte reativa do substrato é um importante mecanismo de catálise, pois esta pode ser compensada por uma estabilização na região ligante, resultando numa energia de ligação total (ΔG_B) maior do que num MC com apenas uma região (reativa) desestabilizada (ou seja, como os exemplos da figura 2), além de diminuir a barreira de ativação para reação.

No entanto, o ganho em energia livre neste modelo é limitado pela diferença entre a estabilização da região ligante e a desestabilização da região reativa. Estimativas^[4] e medidas cinéticas para catálise de substratos com diferentes regiões ligantes mas com a mesma região reativa^[5, tabela 12.1] sugerem que acelerações de apenas três ordens de magnitude podem ser obtidas pelo mecanismo de sítio dividido.

Todavia, este modelo pode explicar as acelerações (cerca de 10^3 vezes) observadas na catálise da desfosforilação do substrato natural (a proteína CDK2, que deve conter diversas regiões de ligação com a enzima), em relação a desfosforilação de substratos artificiais (como fenilfosfato, que não deve possuir regiões somente ligantes), catalisadas por PTPs da subfamília CDC25^[22].

1.1.2.3 NACs e hipótese espaço-temporal

Reações intramoleculares diversas vezes se processam mais rapidamente do que as respectivas reações intermoleculares. O mecanismo desta aceleração, em decorrência da proximidade dos reagentes, tem sido comparado aos mecanismos empregados por enzimas para catálise^[17, 23].

Bruice e Lightstone^[23] introduziram o conceito de conformação próxima ao ataque (NAC), definido como a conformação que os reagentes precisam assumir para chegar ao

TS. A análise conformacional, utilizando campos de força clássicos (§1.6.1)^[24, cap. 4], de uma série de diésteres indicou que a fração molar dos confôrmeros em NACs (definidas de acordo com critérios geométricos como distância e ângulo da ligação em formação) era diretamente proporcional à aceleração da reação intramolecular de formação do anidrido^[23]. Posteriormente, a distribuição das conformações dos substratos nos MC em algumas reações enzimáticas que não envolvem a formação de intermediários covalentes enzima-substrato (por exemplo, o rearranjo de corismato para perfenato, catalisado pela enzima corismato mutase) foi determinada por campos de força clássicos, que indicaram que NACs são formados no sítio ativo enzimático em maior proporção do que na reação de referência em solução^[18].

A energia livre para formação de um NAC em solução aquosa foi estimada em 5 a 10 kcal/mol por Kollman, Kuhn e Perakyla^[9]. Este cálculo, contudo, apresenta algumas inconsistências que foram severamente criticadas (§1.1.2.7)^[12, 25]. A energia livre para formação da NAC no sítio ativo de uma enzima pode ser estimada em cerca de 2 kcal/mol, a partir da fração molar do substrato em NACs^[18]. Portanto, Bruice^[18] acredita que o aumento da fração de NACs é um mecanismo de catálise enzimática operante, ao lado da estabilização do TS (§1.1.2.1).

No entanto, uma NAC não é uma espécie termodinamicamente estável (um ponto estacionário na superfície de energia potencial) e, portanto, sua definição não é única (depende dos critérios escolhidos para distância e ângulo da ligação em formação ou quebra). A escolha de reações enzimáticas sem intermediários covalentes ressalta que o uso de campos de força clássicos^[18] para simulação de conformações de espécies reativas não é adequado, já que estes campos de força não descrevem formação ou quebra de ligações covalentes e foram parametrizados para simulação de conformações no estado fundamental, próximas à geometria de equilíbrio. Potenciais híbridos (§1.6.4) calibrados seriam mais adequados para o estudo no sítio enzimático.

Após a definição de um ciclo termodinâmico que relaciona o efeito de NACs à energia do TS e a construção de um potencial híbrido que força os reagentes e TS a assumirem as conformações restringidas observadas no sítio ativo enzimático, Warshel et al.^[26] estimaram a contribuição da formação de NACs para a catálise em menos que 2.5 kcal/mol.

A idéia de controle estereopopulacional em reações intramoleculares, como a sugestão de aumento da fração molar de NACs, também inspirou Menger^[17] a propor sua teoria espaço-temporal, que postula que a “velocidade da reação entre os grupos X e Y é proporcional ao tempo que X e Y residem a uma distância crítica”. A magnitude dos parâmetros

de espaço e tempo pode ser modificada pelos possíveis arranjos moleculares que podem ser obtidos, sintetizando moléculas em que os grupos X e Y estão imobilizados. A teoria pode ser aplicada para enzimas, que, ao invés da imobilização covalente, utilizam interações não covalentes, provenientes da energia de ligação do substrato, para fixar os grupos reativos e acelerar sua reatividade^[17]. A quantificação desta proposta também não foi efetuada e, portanto, não podemos afirmar qual a importância deste mecanismo para catálise apenas pela interpretação de experimentos em compostos modelos.

1.1.2.4 Catálise covalente

A descoberta de intermediários enzima-substrato, ligados covalentemente, desacreditou a existência de mecanismos “misteriosos” para catálise e sugeriu que o mecanismo de ação enzimática não é diferente de qualquer outra reação química^[8, cap. 2].

A catálise covalente pode ser descrita como uma modificação química transiente na enzima que ativa o substrato ou transfere um grupo reativo do substrato para outro aceptor. Para catálise covalente ser efetiva, as seguintes condições devem ser atingidas^[8, p. 68]:

- O catalisador deve ter uma maior reatividade frente ao substrato do que o aceptor final;
- O intermediário formado entre o catalisador e o substrato deve ser mais reativo do que o substrato;
- O intermediário deve ser termodinamicamente instável em relação ao produto final, de maneira que o intermediário não acumule.

Catalisadores como proteases, estearases, descarboxilases, fosfatases (incluindo as PTPs), entre outras enzimas, formam intermediários covalentes e parecem seguir as condições acima. Por exemplo, fosfatases que utilizam um resíduo de cisteína para ataque nucleofílico sobre fósforo formam um intermediário tiofosforilado (§1.3). A energia de ativação para alcoólise^[27] ou hidrólise^[28] de tioésteres de fosfato (quebra da ligação P–S) em solução aquosa é de 5 a 10 kcal/mol mais baixa do que a respectiva reação de oxiésteres de fosfato (quebra da ligação P–O), sugerindo que a vantagem energética obtida pela formação do intermediário ou pela catálise covalente é desta magnitude.

Jencks^[8, p. 704 e 705] também propôs que, evolucionariamente, uma enzima pode selecionar um grupo catalítico que será usado em duas ou mais etapas diferentes com maior facilidade do que posicionar dois ou mais grupos distintos para serem usados na mesma

etapa catalítica. Por exemplo, o ácido geral presente na primeira etapa da reação das PTPs pode funcionar como base geral na segunda etapa (§1.3).

A formação de intermediários resulta na divisão do processo total em diversas etapas ou reações^[29] e, portanto, o mecanismo enzimático pode ser diferente do mecanismo da reação total observado em solução. A comparação da atividade catalítica deve ser feita entre cada etapa e a respectiva reação de referência não catalisada em solução. Por exemplo, a hidrólise de ésteres de fosfatos catalisada por PTPs é dividida em duas etapas: formação do intermediário fosforilado via tiólise da fosfotirosina; e hidrólise do intermediário tiofosfoéster (§1.3). Embora a hidrólise de fosfotirosina seja diretamente observada em solução na ausência de catalisador (§1.2), as reações de referência em solução para comparação com a atividade enzimática, devem ser a tiólise de fosfotirosina e a hidrólise de tiofosfoéster.

1.1.2.5 Catálise ácido-base geral

A catálise ácido-base geral, ou seja, doação (ou recepção) de H^+ por um grupo enzimático que funciona como um ácido (ou uma base), é uma das etapas envolvidas em diversas reações biológicas^[8, cap. 3]. A transferência pode ocorrer em um pré-equilíbrio rápido e anterior à etapa limitante da velocidade de reação, ocorrer simultaneamente ou mesmo constituir a etapa limitante (por exemplo, na transferência de H^+ de carbonos alifáticos). No primeiro caso, as barreiras intrínsecas de transferência de H^+ são pequenas. Porém, a transferência pode formar intermediários instáveis (como ácidos ou bases conjugadas dos reagentes) e, conseqüentemente, elevar a barreira total da reação. No laboratório, esta elevação não é um problema, pois a adição de hidróxido ou hidrônio à mistura de reação pode forçar a transferência de H^+ . No entanto, enzimas em pH neutro não podem simplesmente aumentar a atividade de hidróxido ou hidrônio no sítio ativo e, portanto, devem evitar a formação de intermediários instáveis, possivelmente formados em solução^[8, p. 167].

A transferência também pode neutralizar ou alterar a carga formal dos grupos reagentes, facilitando outras etapas do mecanismo de reação, por exemplo, o ataque nucleofílico sobre o substrato (como observado para PTPs, §1.3). Na ausência da transferência de H^+ , as outras etapas do mecanismo serão energeticamente desestabilizadas ou até impedidas.

A quantificação do abaixamento da altura de barreira da reação total catalisada não é trivial, porque a transferência H^+ pode influenciar diferentes etapas ou ocorrer em etapas não limitantes. Em solução aquosa, a catálise ácido-base geral propicia aumentos na

constante de velocidade de cerca de 10^2 a 10^3 vezes^[8, p. 617] e, portanto, podemos estimar que contribuições da mesma magnitude serão obtidas pela catálise enzimática.

1.1.2.6 Interação eletrostática

A importância das interações eletrostáticas foi rapidamente identificada para a ligação de substratos e para estabilidade e enovelamento de proteínas, através da formação de pares iônicos ou pontes salinas^[8, cap. 8]. Porém, sua importância como mecanismo molecular para catálise não foi comprovada até o advento de experimentos de mutação sítio-dirigida e de simulações computacionais^[13]. Experimentos de mudança da força iônica do meio mostraram alterações muito pequenas na catálise, já que o interior de uma enzima (onde, normalmente, se localizam os sítios ativos) é pouco perturbado por modificações externas; experimentos em compostos modelos em solução também não apontaram grande efeitos eletrostáticos em razão da capacidade dielétrica do solvente aquoso^[13].

No entanto, alterações do microambiente dielétrico de uma reação podem diminuir a altura da barreira de ativação em mais de 20 kcal/mol (por exemplo, p. 30). Aparentemente, enzimas fornecem microambientes cujo potencial eletrostático complementa as mudanças de cargas observadas ao longo de uma reação^[10, §9.4.1]. Mutações em resíduos que estabilizam cargas formais ou parciais presentes no TS resultam em diminuição de k_{cat} ^[25]. Simulações computacionais mostram repetidamente^[3, 9, 16, 19] que a interação eletrostática entre o ambiente enzimático e o TS das reações catalisadas é a maior responsável pela diminuição de $\Delta G_{enz}^\ddagger$ em relação a $\Delta G_{sol}^\ddagger$ (figura 1). Mesmos nos casos onde o TS é menos polar do que os reagentes (MC na enzima ou complexo de encontro RS em solução aquosa), a estabilização do TS na enzima é maior do que em solução^[12]. A partir de simulações computacionais, a aceleração *média* calculada e atribuída à interação eletrostática na reação enzimática é de 10^{11} vezes, em comparação à reação em solução^[16].

A coordenação (não covalente) com íons metálicos e ligações de hidrogênio são normalmente consideradas interações eletrostáticas, pois são modeladas com razoável sucesso por este tipo de interação^[10, 19].

Solventes polares, como a água, podem reorientar seus dipolos para estabilizar a distribuição de carga no TS. Em uma enzima, no entanto, os dipolos que estabilizam o TS já estão *pré-orientados* no MC para estabilizar as cargas do TS. Em solução, a energia para *reorientar* os dipolos do solvente em torno das cargas do TS é elevada. Pode-se dizer que a energia de orientação dos dipolos no sítio ativo já foi gasta durante o enovelamento da enzima para formação de sua estrutura nativa. Portanto, enzimas

são “supersolventes”, que funcionam por substituição de solvatação do TS, com pequena energia de reorganização^[10, §9.4.2].

Uma proposta semelhante, em que enzimas propiciariam microambientes apolares que desestabilizariam os reagentes por dessolvatação^[8, apêndice], é baseada na observação que algumas reações em fase gasosa ou em solventes apolares ocorrem mais rapidamente do que as mesmas reações em soluções polares (por exemplo, p. 30). Contudo, esta hipótese é descartada em decorrência do efeito de mutações no sítio ativo^[4] e se um ciclo termodinâmico (como a equação 1.5) inclui a transferência dos fragmentos reativos polares do solvente aquoso (referência) para o possível microambiente apolar. Esta transferência envolve uma alta energia, que não está disponível no processo de catálise^[10, §9.2.2]. Sítios ativos são ambientes polares e heterogêneos, bastante diferentes de solventes apolares ou da fase gasosa.

1.1.2.7 Efeitos entrópicos

A aceleração de reações por efeitos entrópicos é freqüentemente considerada como um dos principais mecanismos moleculares de catálise^{[8, apêndice][9, 30]}. Esta proposta, que já foi enunciada de diversas maneiras, implica que o espaço configuracional disponível para os reagentes em solução é restringido no sítio ativo enzimático, resultando em significativas diferenças entrópicas entre as barreiras de ativação na enzima e na reação de referência em solução.

Na estimativa dos efeitos entrópicos, a definição do estado de referência em solução é crucial. O efeito de concentração, equivalente a aproximação dos reagentes, é dado por $\approx -RT \ln 55$ para cada par de reagentes formados. Assim para uma reação bimolecular R+S (figura 1) esta contribuição (ΔG_{comp}) é 2.4 kcal/mol a 300 K. Descontado o efeito de concentração, a diferença entre as barreiras entrópicas (ΔS) para ativação da reação em solução e na enzima pode ser escrita por:

$$\begin{aligned}
 (\Delta \Delta S^\ddagger)_{sol \rightarrow enz} &= [(\Delta S_{enz}^\ddagger) - (\Delta S_{sol}^\ddagger)] \\
 &= [(S_{enz}^{TS} - S_{enz}^{ES}) - (S_{sol}^{TS} - S_{sol}^{RS})] \\
 &= [(S_{enz}^{TS} - S_{sol}^{TS}) - \Delta S_B]
 \end{aligned} \tag{1.6}$$

onde $\Delta S_B = (S_{enz}^{ES} - S_{sol}^{RS})$, é a entropia de ligação do substrato, considerando os estados *ES* e *RS* (figura 1).

As propostas de contribuição entrópica para catálise assumem *implicitamente* que os

graus de liberdade do TS em solução ou na enzima estão congelados, ou seja $S_{enz}^{TS} = S_{sol}^{TS} \simeq 0$ e, portanto, superestimam $(\Delta\Delta S^\ddagger)_{sol \rightarrow enz} = -\Delta S_B$ ^[9, 11, 25]. Esta suposição é baseada no raciocínio de Jencks para reações bimoleculares utilizando reagentes *esféricos*^[8, p. 671] cuja conclusão é que até -35 unidades entrópicas (cerca de 10 kcal/mol a 300 K) seriam obtidas pela restrição entrópica no sítio ativo. Porém, este raciocínio não é válido para estruturas *moleculares*, porque parte da movimentação livre no complexo RS (ou ES na figura 1) também está livre no TS (parte dos graus de liberdade translacionais e rotacionais tidos como “congelados” são na verdade transformados em vibrações de baixa frequência nos TS cuja contribuição entrópica é bastante apreciável)^{[10, §9.3.2][25, 31]}.

Simulações computacionais das contribuições entrópicas ainda sofrem de problemas de convergência e, portanto, as estimativas não são totalmente seguras^[32]. Simulações semiquantitativas^{[10, §9.3.1][32]} indicam que $(\Delta\Delta S^\ddagger)_{sol \rightarrow enz} < 3$ kcal/mol, incluindo os efeitos de concentração (vide supra). Entretanto, a definição utilizada por Warshel et al. para estas estimativas não é idêntica daquela tradicionalmente proposta^{[8, apêndice][9]}, pois esta última não permite a construção de um ciclo termodinâmico.

Kollman et al.^[11] propuseram uma metodologia para estimar as contribuições entrópicas baseadas na superfície potencial da reação em fase gasosa. Este método foi criticado porque estimativas na fase gasosa superestimam as contribuições entrópicas e, principalmente, porque o cálculo da entropia deve ser feito a partir de potenciais bem-definidos, mas o uso de superfícies obtidas na fase gasosa para estimar contribuições em solução ou no sítio enzimático não é totalmente justificado^[10, §9.3.1]. Curiosamente, uma das justificativas de Kollman, Kuhn e Perakyla^[9] para inclusão das contribuições entrópicas estimadas na fase gasosa é que o valor destas contribuições é semelhante à diferença entre os valores computados (sem tais contribuições) e os experimentais. Tunón^[12] repetiu parte dos cálculos de Kollman et al.^[11] e obteve contribuições entrópicas até 7 kcal/mol diferentes, evidenciando que a superfície potencial deve ser bem definida para que as estimativas da entropia sejam reproduzíveis. Estimativas obtidas por uma metodologia diferente^[12] resultaram em valores pequenos para $(\Delta\Delta S^\ddagger)_{sol \rightarrow enz}$, em acordo com os resultados de Warshel et al. e até 20 kcal/mol diferentes daquelas obtidas por Kollman et al..

Nenhum experimento demonstrou diretamente que um dado efeito catalítico em enzimas estivesse associado a alterações de $(\Delta\Delta S^\ddagger)_{sol \rightarrow enz}$, por exemplo, ao observar a restrição das configurações disponíveis ao MC. A análise das contribuições para energia livre de reações enzimáticas e sua referência em solução (determinadas experimentalmente a partir de gráficos de Arrhenius)^[18, 32] indica que a contribuição entrópica é equivalente em

solução e na enzima e que a catálise é atribuída a efeitos entálpicos. Contudo, não pode-se afirmar qual parte destas diferenças entrópicas está associada ao grau de liberdade dos reagentes e qual está associada à entropia de solvatação. Dados experimentais obtidos em solução^[33] sugerem que a água já impõe restrições à movimentação translacional e rotacional de solutos polares e, portanto, o ganho em termos destas contribuições à entropia pela transferência dos reagentes para o sítio ativo enzimático seria pequeno.

As acelerações observadas em reações intramoleculares (§1.1.2.3), em que os reagentes estão covalentemente ligados e parcialmente imobilizados são citadas como forte evidência para hipótese de catálise por fatores entrópicos. Porém, a informação obtida destas reações modelo não é diretamente transferível para reações enzimáticas, porque as acelerações observadas nos modelos refletem diferentes fatores estéricos, de distorção (§1.1.2.8), entrópicos, etc., que não são facilmente separáveis^[10, §9.3.2]. Simulações computacionais indicaram que acelerações de reações intramoleculares não estão necessariamente associadas à contribuições entrópicas^[23]. De qualquer forma, a restrição das configurações disponíveis na reação intramolecular não está necessariamente ligada às possíveis restrições observadas no sítio ativo, em comparação com a reação intermolecular correspondente de referência em solução^[10, §9.3.2].

Portanto, embora efeitos entrópicos para catálise não possam ser totalmente ignorados, a magnitude destas contribuições é menor do que tradicionalmente proposto^[8, apêndice]. A avaliação destas contribuições requer a correta definição das quantidades termodinâmicas e o uso de simulações computacionais com potenciais calibrados.

1.1.2.8 Tensão e impedimento estérico

A tensão ou impedimento estérico resultantes da aproximação entre moléculas ou partes da mesma molécula provocam distorções estruturais e influenciam quais conformações são assumidas. Uma parte das teorias sobre catálise enzimática, principalmente aquelas ligadas à desestabilização do substrato (§1.1.2.1), sugere que o impedimento estérico contribui para as acelerações observadas^[8, cap. 5].

Esta hipótese assume que, caso o sítio ativo seja relativamente rígido, sua estrutura será complementar somente a uma das espécies ao longo da coordenada de reação: reagente, TS ou produto; portanto, as outras espécies sofrerão alguma distorção geométrica para se acomodar no sítio ativo^[8, p. 286]. A distorção do reagente facilitaria a catálise, porque a energia livre do MC seria aumentada e, portanto, a altura da barreira de reação, diminuída (perfil B, figura 2). Caso o sítio ativo seja flexível e o substrato rígido, a

formação do MC seria acompanhada por uma mudança conformacional na enzima que tencionaria e desestabilizaria a geometria do substrato, aumentando a energia livre (perfil D, figura 2)^[8, p. 294]. Em ambos os casos, a energia de ligação observada (ΔG_B) seria menor (por $\Delta\Delta G_B$) do que aquela disponível se o ajuste das estruturas no MC fosse ideal (perfil A, figura 2).

A proposta de “ajuste induzido”^[8, p. 292] é semelhante mas não está relacionada à catálise, pois a energia de ligação é utilizada na modificação da estrutura enzimática para regular a especificidade pelo substrato (ao invés de aumentar a energia livre do MC).

Nenhuma observação experimental prova diretamente estas hipóteses^[8, p. 283] e as evidências existentes são consideradas ambíguas^[5, §12.D.5]. O mecanismo de catálise na lisozima é tradicionalmente citado como evidência da tensão que pode ser exercida sobre o substrato, pois a estrutura do sítio ativo parece complementar à conformação “sofá” do substrato (um sacarídeo, cuja conformação “cadeira” é cerca de 5 kcal/mol mais estável em solução aquosa). No entanto, a diferença entre a energia livre associada à mudança da conformação cadeira para conformação sofá em solução aquosa e no ambiente enzimático é menor que 1 kcal/mol^[10, §6.2] e, portanto, esta diferença não é suficiente para explicar o abaixamento de $\Delta G_{enz}^\ddagger$ em relação a $\Delta G_{sol}^\ddagger$ (figura 1).

Medidas de RMN e simulações computacionais mostram que enzimas são moléculas bastante flexíveis, que podem acomodar mudanças na estrutura do substrato sem aumento significativo da energia livre^[16, 29]. Mudanças geométricas do reagente, normalmente associadas à formação do TS, são menores do que 1 Å e deformações desta magnitude são facilmente acomodadas pelos modos vibracionais de baixa energia da proteína (que, pictoricamente, pode ser comparada a um conjunto de molas com baixa constante de força)^[10, §9.2.1].

Numa simulação computacional, o impedimento estérico pode ser definido como o potencial repulsivo das interações de van der Waals e as contribuições da distorção de ligações, ângulos e diedrais (§1.6.1). A contribuição destas interações para diminuição de $\Delta G_{enz}^\ddagger$ foi calculada em menos que 2 kcal/mol para diferentes enzimas^{[26][10, §6.2]}.

1.1.2.9 Efeitos dinâmicos

Movimentações da estrutura enzimática também são propostas como um possível fator de catálise^[14, 16]. Estes movimentos incluem os modos vibracionais e as mudanças da estrutura terciária da proteína que facilitariam a reação em comparação com a reação

em solução. Os movimentos podem influenciar a altura de barreira $\Delta G_{enz}^\ddagger$ da reação enzimática pelas chamadas “vibrações promotoras”^[34], cujo efeito é normalmente incluído nas simulações computacionais^[3, 9, 10] (chamado de reorganização da estrutura enzimática)^[19] e não deve representar um mecanismo particular de catálise^[16, 25].

Pela teoria generalizada do estado de transição^{[35][7, cap. 7]}.

$$k(T) = \kappa(T) \left(\frac{k_B T}{h} \right) \exp \left[\frac{-\Delta G^\ddagger}{RT} \right] \quad (1.7)$$

as movimentações também podem alterar a constante de velocidade da reação, $k(T)$, pela modificação do coeficiente de transmissão, $\kappa(T)$. Na equação acima, k_B é a constante de Boltzmann, T é a temperatura e h é a constante de Planck. Na teoria generalizada, $\kappa(T) \leq 1$, ou seja, uma trajetória pode ultrapassar o TS, mas retornar aos reagentes. Na teoria de estado de transição canônica, o recruzamento do TS não é permitido, pois toda trajetória que ultrapassa o TS é reativa e resulta nos produtos da reação, logo, $\kappa(T) \equiv 1$.

Além de um recruzamento não reativo clássico, $\kappa(T)$ pode ser alterado por efeitos de solvatação de não-equilíbrio (por exemplo, proveniente de uma distribuição configuracional dos reagentes fora do equilíbrio, ocasionada pela formação do MC) e por efeitos quânticos como energia do ponto-zero e tunelamento^[14, 16, 25, 30]. Os efeitos de recruzamento clássico e de não-equilíbrio podem apenas tornar $\kappa(T)$ menor do que um e, usualmente, já são pequenos em solução aquosa, ou seja, $\kappa(T)$ é quase unitário^[14]. Portanto, parece não existir possibilidade de enzimas utilizarem estes efeitos para maximizar $\kappa(T)$ e amplificar a catálise de ordens de magnitude, de acordo com os resultados observados em simulações computacionais^[14, 16]. O tunelamento pode resultar em $\kappa(T) > 1$ e pode contribuir em até 3 ordens de magnitude para aceleração de reações enzimáticas, em relação a reação em solução, se uma transferência de hidrogênio (tanto nas formas H^+ , H neutro e H^-) for a etapa determinante da velocidade. Apenas transferências de hidrogênio e de elétrons estão sujeitas a efeitos de tunelamento e somente uma pequena fração das reações enzimáticas contém estas transferências na etapa determinante da velocidade. De qualquer forma, a magnitude do efeito não é suficiente para explicar toda aceleração observada na catálise^[16, 30] (§1.1.2.5).

1.2 Reatividade de ésteres de fosfato

Uma parte significativa da química de sistemas biológicos depende de reações de ésteres de fosfato. Além de atuarem como agentes sinalizadores pela fosforilação de

Figura 3: Esquema dos mecanismos propostos para as reações de transferência de fosfato.

proteínas (§1.3), ésteres de fosfato estão envolvidos no material genético e na transferência de energia química^[36]. Pesticidas e agentes neuro-tóxicos usados em armas químicas também são derivados de ésteres do ácido fosfórico.

Nas últimas décadas, numerosos estudos dedicaram-se a compreender os mecanismos das reações de ésteres de fosfato em solução^[37–39], em fase gasosa^[40, 41] e catalisadas por enzimas^[42, 43].

Três tipos de mecanismos, que variam no grau de associação do nucleófilo e dissociação do grupo de saída, foram propostos para descrever as reações de transferência de fosfato (figura 3). A nomenclatura sugerida pela IUPAC foi adotada para designar os mecanismos de reação^[44]. O mecanismo A_n+D_n é uma adição-eliminação com ataque nucleofílico na primeira etapa, formação do intermediário pentacoordenado estável e dissociação do grupo de saída na segunda etapa. O mecanismo D_n+A_n equivale a uma reação S_N1 (na antiga nomenclatura) e corresponde à eliminação do grupo de saída na primeira etapa e formação do metafosfato, PO_3^- , uma espécie altamente reativa, que reage com o nucleófilo na segunda etapa de reação. No mecanismo A_nD_n , a adição do nucleófilo e a eliminação do grupo de saída ocorrem simultaneamente, em uma única etapa, sem formação de um intermediário estável.

1.2.1 Reações em solução

O fator que determina o mecanismo observado é o estado de alquilação do éster de fosfato^[37–39, 45]. Dados experimentais para reações em solução são normalmente interpretados^[37–39, 45] como evidências de que monoésteres de fosfato reagem segundo um

mecanismo D_n+A_n ou sob um mecanismo A_nD_n cujo TS apresenta significativa quebra da ligação P—O entre o fosfato e o grupo de saída. Os dados experimentais também sugerem que diésteres e triésteres reagem segundo mecanismo A_nD_n se o grupo de saída for ativado (como, por exemplo, *p*-nitrofenol), mas com maior participação do nucleófilo no TS do que as reações de monoésteres; ou formam intermediários estáveis, reagindo segundo um mecanismo A_n+D_n .

Condições do meio, natureza do nucleófilo e do grupo de saída, e o estado de protonação para os di e monoésteres também influenciam qual mecanismo é observado.

O ataque nucleofílico sobre triésteres de fosfato apresenta uma cinética de segunda ordem^[38] e pode ocorrer sobre o fósforo ou sobre um dos átomos de carbono. A natureza e basicidade do nucleófilo afetam tanto a regioselectividade como a velocidade das reações. Em solução aquosa, nucleófilos duros (segundo a classificação de Parr e Pearson^[46, §5.3]), como OH^- , atacam preferencialmente o fósforo. Já nucleófilos moles, como H_2O , Br^- e I^- , atacam preferencialmente um átomo de carbono do triéster^[38]. Considerando o ataque sobre o fósforo, nucleófilos de pequena basicidade reagem mais lentamente^[37] do que nucleófilos mais básicos ou aqueles com “efeito α ”^[8, §2.D.9]^[47, 48]. Por exemplo, a constante de velocidade de segunda ordem de ataque de OH^- é 10^4 vezes maior do que a constante de ataque de H_2O sobre o fósforo de um triaril fosfato^[38]. A hidrólise de triésteres não é catalisada por H^+ , ao contrário da hidrólise de monoésteres (vide infra).

A variação da constante de velocidade de reação com o $\text{p}K_a$ do nucleófilo ou do grupo de saída pode ser representada por uma relação linear de energia livre (LFER)^[49, cap. 6]. Nas reações de triésteres de fosfato, o β_{nuc} (inclinação da LFER usando o $\text{p}K_a$ do nucleófilo) varia de 0.30 a 0.48^[50]. O β_{nuc} significativamente diferente de zero indica efetiva participação do nucleófilo na etapa determinante da velocidade de reação e, portanto, é considerado uma evidência de mecanismo associativo (A_nD_n ou A_n+D_n)^[37, 39, 50].

A entropia de ativação ($\Delta S^\ddagger$) medida para reações de triésteres em solução, cerca de $-35 \text{ cal mol}^{-1} \text{ K}^{-1}$, é interpretada como perda de entropia translacional dos reagentes, geralmente associada a reações bimoleculares^[49, §5.4.4] e, portanto, também é considerada como evidência de um mecanismo A_nD_n ou A_n+D_n ^[51]. A distinção entre mecanismos unimoleculares e bimoleculares em função do $\Delta S^\ddagger$ é possível, caso a contribuição dominante para $\Delta S^\ddagger$ seja a entropia translacional. Esta condição deve ser verdadeira nos experimentos com triésteres^[51], cujos valores de $\Delta S^\ddagger$ são bastante negativos, mas deve ser vista com ressalvas no caso dos experimentos com monoésteres (vide infra)^[52].

Ésteres de fosfatos pentacoordenados têm geometria de bipirâmide trigonal (TBP),

Figura 4: Exemplo de uma bipirâmide trigonal. O plano equatorial está indicado por um triângulo pontilhado. Legenda: átomo central em preto, átomos equatoriais em linhas diagonais e átomos axiais em linhas cruzadas.

com três grupos na posição equatorial e dois grupos na posição axial (figura 4). Estas espécies podem sofrer pseudorrotações^[53], ou seja, mudanças dos ângulos de ligação que resultam em troca de posição entre um grupo equatorial e outro axial na TBP. Uma espécie pentacoordenada que forma um intermediário estável de reação de éster de fosfato é denominada *fosforana*. A formação de fosforanas foi proposta para reações de di e triésteres nos casos em que uma pseudorrotação seja necessária para posicionar axialmente o grupo de saída^[40, 45, 53, 54]. Fosforanas foram observadas experimentalmente para alguns sistemas especiais^[55, 56], mas evidências computacionais^[57, 58] sugerem que fosforanas não são necessariamente formadas em todas as reações de triésteres de fosfato.

As seguintes regras empíricas foram propostas para o arranjo dos grupos ligantes numa fosforana^[39]:

1. Nucleófilos e nucleófugos entram e saem da TBP por uma posição axial;
2. Ligantes mais eletronegativos preferem posições axiais;
3. Ligantes doadores de elétrons π preferem posições equatoriais;
4. Efeitos estéricos são minimizados, posicionando-se os ligantes impedidos equatorialmente;
5. Caso o fósforo e dois de seus átomos ligantes sejam membros de um anel de 4 ou 5 membros, estes dois átomos estarão posicionados axial e equatorialmente.

Estas regras, em grande parte baseadas na reatividade de fosfatos cíclicos, têm sido criticadas, principalmente em relação à necessidade de entrada e saída dos ligantes pela posição axial^[39].

Efeito isotópico cinético (KIE) foi medido para reações de ésteres de fosfato em diferentes condições^[20]. O KIE é a relação entre a constante de velocidade de reação obtida com o reagente marcado com um isótopo leve sobre aquela obtida com um isótopo pesado^[59, cap. 7] e pode ser usado para investigar a estrutura do TS da etapa determinante da velocidade de reação^[60, cap. 12, 13 e 14]. Um KIE é chamado de primário se a ligação que se quebra ou se forma durante a reação envolve o átomo marcado e de secundário se o isótopo se encontra em outras posições. O KIE é consequência das diferenças nos modos vibracionais dos quais o átomo marcado participa, entre os reagentes e o TS do passo limitante da reação^[59, cap. 7]. O KIE também pode ser interpretado em termos de mudanças da ordem de ligação observada no TS limitante^[20]. O grupo de saída *p*-nitrofenol é usado em medidas de KIE em reações de ésteres de fosfato. Assim, o KIE primário, $^{18}k_{bridge}$, refere-se à ligação P–O entre o fosfato e o grupo de saída, e mede o grau de dissociação desta ligação no TS. Um $^{18}k_{bridge} > 1$ indica que a quebra da ligação P–O é substancial no TS da etapa determinante da velocidade de reação. Os KIE secundários medidos são aqueles do nitrogênio do grupo de saída, ^{15}k , e dos oxigênios não esterificados do grupo fosfato, $^{18}k_{nonbridge}$. O ^{15}k indica a densidade de carga negativa deslocalizada no grupo nitro^[20]. Caso o grupo de saída se dissocie como um ânion, o ^{15}k será maior que 1. Se alguma protonação neutraliza parte da densidade de carga formada da quebra da ligação P–O, $^{15}k \approx 1$. A diminuição ou anulação do KIE em razão da protonação também é observada no $^{18}k_{bridge}$. O $^{18}k_{nonbridge}$ indica se o grupo fosfato no TS tem característica de um metafosfato dissociado, em que $^{18}k_{nonbridge} < 1$, ou se o grupo fosfato tem características mais associativas, como uma fosforana, em que $^{18}k_{nonbridge} > 1$. O $^{18}k_{nonbridge}$ também sofre influência da protonação dos oxigênios não esterificados do grupo fosfato.

Para as reações de diésteres e, principalmente, triésteres, o $^{18}k_{nonbridge}$ é bem maior que 1, indicando um TS na etapa determinante com significativo caráter associativo (tabela 1). A reduzida magnitude de ^{15}k e $^{18}k_{bridge}$ para as reações de triésteres e diésteres indica que, no TS determinante da velocidade de reação, a densidade de carga negativa no grupo de saída e a quebra da ligação P–O são reduzidas e sugere que a etapa determinante da velocidade é o ataque nucleofílico.

Para di e triésteres com um grupo de saída ativado, como o *p*-nitrofenol, o TS tem caráter associativo, mas não há evidências que comprovem a existência de uma fosforana e, assim, o mecanismo proposto é A_nD_n ^[39, 45]. Para grupos de saída mais básicos como fenóxido e metóxido, o $^{18}k_{nonbridge}$ é mais elevado e, portanto, a formação de uma fosforana é mais provável^[39].

Tabela 1: Efeito isotópico cinético em reações de hidrólise de ésteres de fosfato em solução^a

reagente	¹⁵ k	¹⁸ k _{bridge}	¹⁸ k _{nonbridge}
monoéster diânion em H ₂ O	1.0028	1.0189	0.9994
monoéster diânion em <i>t</i> -butanol	1.0039	1.0202	0.9997
monoéster monoânion	1.0005	1.0094	1.0199
diéster ^b	1.0007(16)	1.0042(63)	1.0028(56)
triéster ^b	1.0007	1.0063	1.0063(250)

^a o erro associado às medidas^[20] é na última casa decimal. ^b KIEs para diésteres e triésteres correspondem à hidrólise alcalina. Os números em parênteses mostram os valores máximos nas últimas casas decimais indicadas obtidos para diferentes reagentes.

A reatividade de monoésteres de fosfato é mais complexa, porque diferentes estados de protonação são acessíveis a estes compostos. O primeiro e o segundo pK_a dos oxigênios não esterificados em monoésteres de fosfato são 1.5 e 6.5, respectivamente^[61]. Os pK_a s podem variar em torno destes valores, dependendo do grupo ligado ao fosfato^[61]. O perfil de variação da constante de velocidade de hidrólise de monoésteres versus o pH do meio tem forma de um sino, com máximo em torno de $pH = 4$, e torna-se independente do pH para valores abaixo de 1 e acima de 9 unidades^[62, 63]. A cinética de hidrólise é de primeira ordem e nenhuma troca de ¹⁸O entre H₂¹⁸O e CH₃OH é observada na hidrólise de metilfosfato, indicando que o ataque nucleofílico ocorre sobre o fósforo^[63].

Monoésteres neutros (duplamente protonados) estão presentes em baixíssimas concentrações em solução neutra *in vitro* ou *in vivo* e, portanto, não são importantes mecanisticamente.

O mecanismo proposto para reações de monoésteres dianiônicos (não protonados) em solução aquosa é A_nD_n , cujo TS apresenta significativa quebra da ligação P–O com o grupo de saída e mínima formação de ligação com o nucleófilo. Esta interpretação é baseada nas seguintes observações experimentais:

- Alta dependência da velocidade de reação com o pK_a do grupo de saída ($\lg$) ($\beta_{lg} = -1.2$)^[62];
- Baixa dependência da velocidade de reação com o pK_a do nucleófilo ($\beta_{nuc} = 0.1$)^[64, 65];
- $\Delta S^\ddagger$ positiva e pequena, cerca de 3 cal mol⁻¹ K⁻¹ para *p*-nitrofenil fosfato (pNPP)^[51];
- Inversão de configuração quando o grupo fosfato é quiral^[66];
- Medidas de KIE primário e secundário (tabela 1)^[20];

- Intermediário metafosfato não foi isolado e sua existência não foi diretamente comprovada^[39].

Ao contrário de triésteres, o β_{nuc} é quase nulo na reação de monoésteres, o que indica pouca ou nenhuma participação do nucleófilo no TS determinante da velocidade de reação. O alto β_{lg} indica que a etapa determinante é efetivamente composta pela quebra da ligação P–O.

A pequena entropia de ativação, sugerida como indicação do mecanismo A_nD_n , deve ser vista com ressalvas, porque outros fatores além da entropia translacional, como, por exemplo, efeitos de solvatação causados pela redistribuição e deslocalização de carga no TS, podem contribuir significativamente para $\Delta S^\ddagger$.

A inversão de configuração na reação de um fosfato quiral (contendo $[^{16}\text{O}, ^{17}\text{O}, ^{18}\text{O}]$ -fosfato) pode ser vista tanto como evidência de um mecanismo com caráter associativo ou como um mecanismo dissociativo em que o metafosfato não é uma espécie difusionalmente equilibrada. Neste caso, o intermediário reativo metafosfato não tem tempo de vida suficiente para difundir na solução e é capturado por uma molécula de água antes de escapar da esfera de solvatação onde a reação ocorre.

Observando os KIEs do monoéster dianiônico em H_2O (tabela 1), temos que os valores de ^{15}k e $^{18}k_{bridge}$ são 4 vezes superiores àqueles observados para reações de di e triésteres, indicando significativa quebra da ligação P–O e formação de carga negativa no grupo de saída na etapa determinante da velocidade de reação. O valor de $^{18}k_{nonbridge} < 1$ indica que o TS é dissociativo, com característica de metafosfato.

Em solução aquosa, o metafosfato nunca foi diretamente observado^[39]. Porém a reação do monoéster dianiônico em *t*-butanol tem os valores de ^{15}k e $^{18}k_{bridge}$ maiores do que em H_2O (tabela 1), indicando que a quebra da ligação P–O neste solvente é mais avançada e o TS é mais dissociativo do que em água. Neste solvente alcoólico, a reação de um fosfato quiral produz uma mistura racêmica dos produtos^[66]. Estas duas observações sugerem que em *t*-butanol, um composto pouco nucleofílico e impedido estericamente, o grupo metafosfato é formado e tem tempo de vida suficientemente grande para difundir pelo solvente e, portanto, permitir que o ataque nucleofílico ocorra com a mesma chance pelas duas faces do PO_3^- planar.

Grande parte dos resultados experimentais comentados acima foram obtidos com *p*-nitrofenil fosfato. Parte das conclusões discutidas acima não deve ser válida para reações com grupos de saída pouco ou desativados, pois as reações de monoésteres de fosfato

Figura 5: Estrutura do possível TS para dissociação assistida por água de monoéster de fosfato monoprotonado^[68].

dependem significativamente da basicidade do grupo de saída (vide supra)^[62]. Em solução aquosa, a hidrólise de metilfosfato e fenilfosfato dianiônicos são retardadas com a adição de DMSO à solução, mas a reação de pNPP é acelerada cerca de 10^4 vezes^[67]. Estes resultados foram interpretados como evidência de que a reação de hidrólise de monoésteres dianiônicos com grupos de saída de basicidade igual ou superior a de fenol ocorre segundo um mecanismo diferente da reação de pNPP^[67].

Em geral, a hidrólise de monoésteres de fosfato monoaniônicos (monoprotonados) se processa mais rapidamente do que a reação do diânion e a reação de diésteres no mesmo pH^[62]. Esta observação foi atribuída à protonação, via uma transferência intramolecular, do oxigênio do grupo de saída. A transferência de H^+ pode ocorrer num passo rápido de pré-equilíbrio, formando um zwitterion, $RO^+(H)PO_3^{2-}$, ou simultaneamente à dissociação do grupo de saída. Kirby et al. propuseram que, para grupos de saída com basicidade menor do que fenóxido, a transferência de H^+ ocorre conjuntamente à quebra da ligação P—O e para grupos de saída mais básicos (como alquilfosfatos), a transferência é rápida e anterior à dissociação^[62]. Também foi proposto^[62] o envolvimento de uma molécula de água na transferência protônica. O anel de seis membros formado na reação de transferência com participação da água (figura 5) é menos tencionado do que o anel de quatro membros formado na ausência da molécula de água.

Na hidrólise do monoéster monoaniônico, a entropia de ativação é pequena, ocorre inversão de configuração de um fosfato quiral e o $\beta_{lg} = -0.27$ ^[39, 62]. Estes dados são consistentes com um mecanismo dissociativo alternativo, com formação de um zwitterion. O KIE para hidrólise de pNPP monoânion (tabela 1) também difere significativamente do diânion, pois os baixos valores de ^{15}k e $^{18}k_{bridge}$ indicam que a densidade de carga formada pela dissociação do grupo de saída foi neutralizada e o valor de $^{18}k_{nonbridge}$ significativamente maior que 1 sugere que o oxigênio não esterificado foi desprotonado.

A hipótese da influência da basicidade do grupo de saída na ordem de transferência de H^+ é suportada pela diferença no efeito isotópico do solvente, k_{H_2O}/k_{D_2O} , que passa de 0.87 na hidrólise de metilfosfato para 1.45 na hidrólise de 2,4-dinitrofenil fosfato^[62]. Espera-se um efeito significativo caso a transferência de H^+ participe da coordenada de reação da etapa limitante, e um efeito pequeno (próximo de 1) caso a transferência seja anterior a etapa limitante da reação^[59, cap. 7].

Um interessante estudo computacional de estrutura eletrônica molecular acoplada a uma representação do solvente aquoso estabeleceu recentemente a existência de um zwitterion aniônico, $RO^+(H)PO_3^{2-}$, como um intermediário de reação estável^[68]. O intermediário zwitteriônico foi encontrado quando $R = H$, metil e fenil, mas não foi encontrado quando $R = 2,4$ -dinitrofenil^[69]. Estes cálculos concordam com a proposta de Kirby et al. para a reação de hidrólise. A barreira de ativação calculada ($\Delta G_{calc}^\ddagger$) para reação de hidrólise de metilfosfato corresponde à experimental ($\Delta G_{calc}^\ddagger = 30.3$ e $\Delta G_{exp}^\ddagger = 30.7$ kcal/mol a $100^\circ C$ ^[38]) e o efeito isotópico do solvente calculado é idêntico ao experimental (vide supra). Os mecanismos D_n+A_n e A_nD_n com TS dissociativo, calculados^[68] para reação de hidrólise de $CH_3OPO_3^-\cdot H_2O$, ou seja, metilfosfato coordenado com uma molécula de água (que forma o anel de 6 membros na transferência de H^+ , figura 5), apresentam barreiras de ativação de 30 kcal/mol e, portanto, não podem ser distinguidos com base apenas nos valores computados para altura de barreira.

Ao analisar as constantes de velocidade de hidrólise de ésteres de fosfato, Åqvist, Warshel e colaboradores propuseram um mecanismo alternativo para reação de monoésteres, que inclui uma etapa de pré-equilíbrio, com a transferência de H^+ do nucleófilo para o substrato^[52, 70]:

O ataque de OH^- ocorreria segundo um mecanismo associativo, em que a dissociação do grupo de saída seria limitante da velocidade de reação. Observações cinéticas são consistentes com esta proposta, baseada na equação 1.8, ou com um mecanismo D_n+A_n ou A_nD_n , como tradicionalmente aceito (figura 3, vide supra).

Åqvist, Warshel e colaboradores também analisaram as LFER experimentais^[39] em termos de parábolas de Marcus^[19], em que as energias dos TSs estão correlacionadas com o cruzamento de parábolas que representam a superfície de energia diabática dos reagentes e produtos. Este tipo de representação é válida para sistemas desacoplados em que a coordenada de reação na região do TS não inclui as distorções nucleares dos reagentes,

como, por exemplo, transferências de elétrons^[71]. As reações de fosfatos não representam sistemas desacoplados e, portanto, a análise das LFER por parábolas de Marcus deve ser feita cautelosamente. A conclusão de Åqvist, Warshel e colaboradores referente à análise das LFER é que estes dados são consistentes com as propostas tradicionais^[37, 39, 45] mas também com um mecanismo A_n+D_n cuja a segunda etapa, *dissociativa* (D_n), é determinante da velocidade de reação. Os KIE medidos para hidrólise de monoésteres (tabela 1) também podem ser interpretados de acordo com um mecanismo A_n+D_n com a segunda etapa determinante.

A análise de Åqvist, Warshel e colaboradores é sustentada por computações de uma LFER por cálculos de estrutura eletrônica molecular acoplados a uma representação da solução aquosa para a reação $\text{CH}_3\text{O}^- + \text{ROPO}_3^-$, onde RO são derivados de fenol com $\text{p}K_a$ entre 5 e 10^[72]. Estes resultados mostram que um mecanismo A_n+D_n com a segunda etapa determinante reproduz o β_{lg} experimental para reações de monoésteres dianiônicos (vide supra)^[72]. Porém, limitações (otimizações parciais de geometria e um conjunto base atômico reduzido utilizado para expansão da função de onda de um diânion na fase gasosa, §1.4) da metodologia usada neste estudo computacional sugerem que os resultados devem ser vistos cautelosamente.

Åqvist, Warshel e colaboradores também apontaram que os TSs importantes para um mecanismo A_n+D_n tradicional (sem pré-transferência de H^+ , figura 3) devem ser bastante similares aos TSs de um mecanismo associativo baseado na equação 1.8. Estes autores comparam os mecanismos associativos e dissociativos para hidrólise de metilfosfato neutro, mono e dianiônico, utilizando cálculos de estrutura eletrônica acoplados a uma representação de solvente aquoso simplificada (que não contém moléculas de água explícitas, ou seja, que possam catalisar ou participar da reação)^[73]. As barreiras de ativação computadas, cerca de 40 kcal/mol, são semelhantes para ambos mecanismos, porém 10 kcal/mol maiores do que as energias de ativação medidas.

Outro recente estudo computacional da hidrólise de metilfosfato^[74] empregou uma metodologia mais rigorosa, incluindo explicitamente moléculas de água além da representação do solvente aquoso. A barreira de ativação calculada para o mecanismo D_n+A_n é de 29 kcal/mol, um valor bastante próximo ao experimental (vide supra). Contudo, o efeito isotópico cinético do solvente calculado neste mecanismo não corresponde ao efeito medido, ao contrário dos resultados de Bianciotto, Barthelat e Vigroux^[68], em que um intermediário zwitteriônico é obtido (vide supra). A barreira calculada para o mecanismo A_n+D_n é cerca de 10 kcal/mol mais alta e, portanto, Hu e Brinck^[74] concluem que um

mecanismo como a equação 1.8 não é pertinente para hidrólise do metilfosfato aniônico. A vantagem de 10 kcal/mol a favor do mecanismo dissociativo parece ser proveniente da formação de um TS com 6 membros para a transferência de H^+ com a assistência de uma molécula de água (figura 5). Para o mecanismo A_n+D_n , não existe vantagem energética significativa na formação de um TS de 6 membros assistido pela água.

Os resultados para hidrólise de metilfosfato catalisada por uma molécula de água^[68, 74] foram criticados^[75] pela exclusão da contribuição da entropia de “congelamento”, decorrente da presença da água explícita, nos valores calculados para as espécies reativas. No entanto, a molécula de água tem menor liberdade conformacional quando o TS é atingido do que nos reagentes, logo, a contribuição da entropia de “congelamento” resultará numa barreira de ativação calculada ligeiramente maior e, portanto, mais próxima aos valores experimentais. De qualquer forma, esta contribuição é pequena e não deve alterar nenhuma das conclusões^[68, 74].

Quais são, então, as conclusões bem fundamentadas sobre a reatividade de ésteres de fosfato em solução?

- A reatividade de di e triésteres depende da natureza do nucleófilo^[38, 76] e o mecanismo é associativo. A formação de fosforanas estáveis não é obrigatória e depende da basicidade do grupo de saída^[39];
- A reatividade de monoésteres, como os substratos de PTPs, é mais complexa e o mecanismo de reação pode ser D_n+A_n ou A_nD_n com TS dissociativo^[39, 73]. A formação de metafosfato, seu tempo de vida e, portanto, sua relevância mecanística, dependem das condições do meio^[51];
- Monoésteres monoprotonados reagem pela formação de um intermediário zwitteriônico, $RO^+(H)PO_3^{2-}$, ou por uma transferência de H^+ acoplada à coordenada de reação^[62, 68].

A maioria destas conclusões foram obtidas do estudo de reações de hidrólise de ésteres de fosfato. Embora os produtos das reações de ésteres de fosfato observadas em sistemas biológicos sejam os mesmos da reação de hidrólise, diferentes intermediários são formados e, portanto, parece-nos importante conhecer qual a influência do nucleófilo na reatividade. Para reações de monoânions, a influência do nucleófilo parece ser muito pequena ($\beta_{nuc} = 0.1$, p. 23). A formação de um metafosfato implica que este será provavelmente capturado por uma molécula de solvente (Vigroux et al. sugeriram que a seletividade do grupo PO_3^-

é induzida pela capacidade de solvatação e não pela nucleofilicidade do aceptor^[68]) e, neste caso, apenas reações de solvólise poderiam ser observadas.

Existem poucos estudos conclusivos sobre a alcoólise de ésteres de fosfato^[37]. A solvólise de 2,4-dinitrofenilfosfato foi observada em metanol e etanol anidros^[77]. A alcoólise ocorre sobre o fósforo e sobre o carbono (cuja substituição nucleofílica aromática segue cinética de segunda ordem). Em baixas concentrações de alcoóxido, a formação de 90% de fenol como produto de reação indica que o ataque ocorre majoritariamente sobre o fósforo. Pode-se estimar que a energia de ativação para ataque de CH_3O^- sobre fósforo do 2,4-dinitrofenilfosfato dianiônico é cerca de 23 kcal/mol a 25°C. Em *t*-butanol anidro, o diânion do pNPP sofre alcoólise 1500 vezes mais rapidamente do que o monoânion, ao contrário do que é observado para hidrólise em solução aquosa^[51]. A entalpia de ativação ($\Delta H^\ddagger$) e a entropia de ativação ($\Delta S^\ddagger$) são $\Delta H^\ddagger = 31.6$ kcal/mol e $\Delta S^\ddagger = 24.5$ cal mol⁻¹ K⁻¹ para a alcoólise do diânion por *t*-butanóxido, e $\Delta H^\ddagger = 37.1$ kcal/mol e $\Delta S^\ddagger = 27.7$ cal mol⁻¹ K⁻¹ para a alcoólise do monoânion^[51]. Embora o pK_a do grupo de saída seja alterado pelo solvente, a diferença entre a energia de ativação para alcoólise do monoéster dianiônico entre 2,4-dinitrofenilfosfato e pNPP está em acordo com o $\beta_{lg} = -1.2$, estimado da hidrólise de monoésteres de arilfosfatos (vide supra). Ao lado das observações de inversão de configuração para reação de um fosfato quiral em metanol e racemização para a reação em *t*-butanol^[66], estes resultados para a alcoólise de monoésteres de fosfato estão de acordo com as interpretações mecanísticas dadas para as reações de hidrólise, em que o mecanismo de reação de monoésteres dianiônicos é A_nD_n com um TS dissociativo ou D_n+A_n em que o metafosfato é difusionalmente equilibrado somente quando a reação é transferida para um solvente menos nucleofílico e impedido, como o *t*-butanol (vide supra). A pequena ou nula participação do nucleófilo na formação do TS limitante da velocidade de reação implica que o mecanismo é relativamente independente da nucleofilicidade do solvente e, portanto, os mecanismos são semelhantes para a hidrólise em solvente aquoso e para solvólise em solvente alcoólico. A reação do monoânion em *t*-butanol é mais lenta, porque o mecanismo especial com transferência de H^+ intramolecular (vide supra)^[62] é dificultado, já que o pK_a do oxigênio não esterificado do grupo fosfato aumenta neste solvente, mas, principalmente, porque o *t*-butanol é impedido estericamente e não deve catalisar a transferência de H^+ via anel de 6 membros, como a água (figura 5).

Ainda menos estudos são conhecidos para reações de tiólise de ésteres de fosfato. A reatividade do composto O,O-difenil fosforocloridotioato, $(\text{PhO})_2\text{P}(=\text{S})\text{Cl}$, foi testada frente a diversos nucleófilos, incluindo tiofenóxido^[78]. Contudo, este composto e sua reatividade são bastante diferentes de um éster de ácido fosfórico.

Figura 6: Esquema da reação de tiólise intramolecular de um tio nucleotídeo^[76]. Abreviações: Ura é uracila e PhOH é fenol.

O único relato conhecido de tiólise de ésteres de fosfato em solução aquosa é a reação intramolecular de um tio-nucleotídeo^[76] (figura 6). Este análogo é um diéster de fosfato aniônico em $\text{pH} > 2$. Ataques no carbono aromático (35%) e no fósforo do grupo fosfato (65%) foram observados em meio fortemente básico ($\text{pH} = 13$) por ^{31}P RMN. Entre $\text{pH} = 7$ e 10, o fosforotioato (figura 6) é formado majoritariamente. A variação da constante de velocidade de reação com o pH do meio é sigmoidal, com níveis basais para $\text{pH} < 5$ ($k_{\text{obs}} = 10^{-6} \text{ s}^{-1}$), aumentando com o pH e estabilizando novamente em $\text{pH} > 10$ ($k_{\text{obs}} = 10^{-4} \text{ s}^{-1}$). Para o oxi-nucleotídeo normal, a reação de alcoólise ocorre 30 vezes mais rapidamente do que a tiólise em $\text{pH} = 7$. Supondo que o nucleófilo deva estar ionizado para realizar o ataque e corrigindo as concentrações pelos $\text{p}K_a$ s medidos para o álcool e o tiol derivado, Dantzman e Kiessling^[76] concluíram que a reação de tiólise do diéster de fosfato é 10^7 vezes mais lenta do que a alcoólise! Portanto, tiolatos parecem nucleófilos extremamente reticentes frente a ésteres de fosfato em solução aquosa.

1.2.2 Reações em fase gasosa

A solvatação altera significativamente os perfis de energia de reações químicas^[79, §5.5] [49, cap. 8]. Por exemplo, a reação de $\text{Cl}^- + \text{CH}_3\text{Cl}$ tem sua barreira de ativação aumentada em cerca de 24 kcal/mol quando transferida da fase gasosa para a solução aquosa^[49, p. 196]. A observação de reações em fase gasosa permite que a reatividade intrínseca seja estudada independentemente da influência da solvatação na energia relativa das espécies e de moléculas de solvente que podem realizar alguma catálise (por exemplo, pela transferência de H^+).

As reações de nucleófilos com di e monoésteres de fosfato não foram estudadas em fase gasosa, ao conhecimento do autor. A reação entre duas espécies carregadas negativamente, como, por exemplo, o ataque de um nucleófilo aniônico sobre um di ou monoéster

ionizado, é energeticamente proibida na fase gasosa. Os di e monoésteres neutros são protonados e têm elevada acidez^[40, 80], portanto, a principal reação destas espécies com ânions nucleofílicos seria a transferência de H^+ .

A reatividade de triésteres de fosfato em fase gasosa recebeu mais atenção^[40, 41, 81]. Dois estudos^[41, 81] foram conduzidos a baixas pressões e fora do equilíbrio térmico, pela técnica de ressonância ciclotrônica de íons. Outro estudo^[40] empregou a técnica “flowing afterglow”, em que os experimentos são conduzidos em um fluxo de gás hélio e, logo, são termicamente equilibrados (em 298 ± 3 K). Parte das diferenças dos resultados observados pode ser atribuída a estas diferenças metodológicas.

Asubiojo, Brauman e Levin^[81] compararam o ataque de OH^- em TMP e metil etileno-fosfato (um triéster cíclico). Em contraste com a reatividade em solução, a abertura do anel cíclico não foi observada e concluiu-se que existe uma barreira para ambas reações, já que as constantes de velocidade medidas são inferiores às taxas de colisão calculadas para os reagentes. Asubiojo, Brauman e Levin acompanharam o aparecimento de CH_3O^- e, portanto, não puderam distinguir qual o sítio de ataque nucleofílico, pois o ataque de OH^- sobre carbono ou fósforo do TMP gera os mesmos produtos.

Os estudos de Hodges, Sullivan e Beauchamp^[41] e Lum e Grabowski^[40] testaram a reatividade de diversos nucleófilos (mais de 15 diferentes espécies, centradas em átomos de C, N, O, F, S e Cl) frente a TMP em fase gasosa. Todos nucleófilos estudados realizam substituição nucleofílica predominantemente no carbono do TMP. Este resultado está em contraste com as observações em solução, em que nucleófilos como F^- e OH^- atacam o átomo de fósforo do TMP^[38], e demonstra que o microambiente também modula a regioselectividade de uma reação (esta modulação pode ser racionalizada por alterações da dureza molecular e local dos reagentes^{[82][46, §5.3]}). Abstração de H^+ seguida por eliminação de formaldeído também é observada para alguns nucleófilos (NH_2^- , OH^- e $C_6H_5^-$)^[40, 41]. Hodges, Sullivan e Beauchamp observaram o aparecimento de CH_3O^- na reação de NH_2^- e OH^- com TMP e atribuíram este resultado ao ataque sobre o átomo de fósforo. Esta interpretação foi criticada por Lum e Grabowski, que defenderam que a observação de CH_3O^- não pode ser usada para provar o ataque nucleofílico no fósforo, pois o ataque de OH^- no carbono ou no fósforo gera os mesmos produtos e CH_3O^- também pode ser formado por contaminação ou por pirólise dos reagentes.

Ataque nucleofílico sobre o fósforo é observado apenas para nucleófilos como alco-óxidos, CD_3O^- (metóxido- d_3) e $CH_3CH_2O^-$, pelo aparecimento dos triésteres (transes-terificados) produzidos como traços de reação (menos do que 2% do total de reação

observada)^[40]. O ataque sobre fósforo parece depender da basicidade do nucleófilo, já que nenhuma reação no fósforo foi detectada para *t*-butanóxido. Em desacordo com estes resultados, Hodges, Sullivan e Beauchamp observaram 100% de seletividade por carbono para reação de CD_3O^- com TMP. Esta diferença pode ser atribuída às diferentes condições experimentais empregadas (vide supra). A reação de HS^- com TMP resulta em ataque no carbono com 100% de seletividade^[40]. Lum e Grabowski sugerem que um mecanismo D_n+A_n para reação de TMP é muito improvável, já que a formação de um “dimetil metafosfato” é energeticamente inacessível na fase gasosa e que a preferência pelo ataque no carbono deve-se, em parte, ao melhor grupo de saída, dimetil fosfato, formado nesta reação.

Nenhum dos estudos de TMP em fase gasosa obteve valores de constantes de velocidade ou de equilíbrio para o ataque sobre fósforo. A única exceção é a reação de CD_3O^- com TMP que resulta em 98.5% de seletividade sobre carbono e 1.5% sobre fósforo. Todavia, energias de reação ou barreiras de ativação não foram descritas. Apenas uma relação de 43:1 entre as constantes de velocidade para o ataque de CD_3O^- sobre carbono e sobre fósforo do TMP é reportada^[40]. Pela teoria de estado de transição, esta relação equivale a uma diferença de 2.2 kcal/mol entre as duas barreiras de ativação.

Estudos computacionais podem auxiliar na compreensão de processos não observados experimentalmente. Os mecanismos de alcoólise e tiólise de ésteres de fosfato na fase gasosa serão descritos no capítulo 2 por cálculos de estrutura eletrônica de reações modelo de tri (figura 7) e monoésteres de fosfato (figura 8).

Estas reações serão usadas para o estudo da reatividade intrínseca de ésteres de fosfato (cap. 2) e para a calibração de metodologias aproximadas (cap. 3) utilizadas na simulação de reações enzimáticas.

1.3 Proteínas tirosina-fosfatases

1.3.1 Papel fisiológico, classificação e estrutura

Processos bioquímicos dependentes de sinalização celular, como crescimento, diferenciação e proliferação são regulados pelo nível de fosforilação de proteínas na célula^[43, 83, 84]. O nível de fosforilação é, por sua vez, controlado por proteínas quinases e fosfatases. Proteína quinases catalisam a fosforilação dependente de ATP de resíduos de tirosina, serina e treonina de outras proteínas. Esta atividade é contrabalançada pelas proteínas fos-

Figura 7: Esquema das reações de tiólise (reações A e B) de trimetil fosfato (TMP) e dimetil fenilfosfato (DMPP) e alcoólise (reações C e D) de TMP e DMPP.

fatases, que regeneram os resíduos desfosforilados e produzem fosfato inorgânico^[43, 83, 84].

As proteínas fosfatases são divididas em duas grandes superfamílias, que utilizam mecanismos de catálise diferentes^[42, 43]. As enzimas da superfamília das proteínas fosfatases (PP) hidrolisam somente resíduos serina e treonina, são dependentes de íons metálicos divalentes e o mecanismo proposto para catálise apresenta apenas uma etapa^[42]. Já as proteínas tirosina-fosfatases (PTP) não contêm íons metálicos^[42] e desfosforilam resíduos de tirosina em duas etapas, com formação de um intermediário tiofosforilado^[85].

As PTPs são divididas em pelo menos quatro subfamílias:

1. Fosfatases específicas para tirosina clássicas;
2. Fosfatases específicas para tirosina de baixo peso molecular (LM-PTP);
3. Fosfatases de dupla especificidade (DSP);
4. Fosfatases de dupla especificidade CDC25.

Estas quatro subfamílias têm pouca similaridade na sequência primária, mas são consideradas parte da mesma superfamília em razão da conservação do sítio ativo e do mecanismo catalítico^[83, 86] (§1.3.2).

A subfamília das PTPs clássicas é a mais numerosa e contém cerca de 100 diferentes enzimas, incluindo enzimas citossólicas e transmembranares associadas a unidades

Figura 8: Esquema das reações de tiólise (reações E e F) de metilfosfato (MeOP^-) e fenilfosfato (PhOP^-) e alcoólise de MeOP^- e PhOP^- (reações G e H).

receptoras extracelulares e a unidades regulatórias (como, por exemplo, motivos SH2)^[84]. A estrutura terciária das PTPs clássicas apresenta um único domínio $\alpha + \beta$, composto centralmente por uma folha- β altamente dobrada, com 4 fitas- β paralelas e 5 fitas- β antiparalelas, rodeada por 6 α -hélices^[43]. As enzimas mais estudadas deste grupo são a PTP1B humana, que é um regulador negativo da sinalização de insulina e leptina^[87], e a YOP51 de *Yersinia*. Este micro-organismo é causador da peste bubônica e a YOP51 é um fator de virulência essencial, assim como a fosfatase de maior atividade conhecida^[42, 86].

As DSPs desfosforilam resíduos de serina e treonina, além de tirosina, e são encontradas no citossol ou associadas a membranas celulares. A principal representante é a VHR (*Vaccinia* vírus H1 Relacionada), isolada de fibroblastos humanos, porém, sua função celular é desconhecida^[88]. A estrutura terciária das DSPs é semelhante às PTPs clássicas, mas a folha- β é menor, contendo 4 fitas- β paralelas e apenas 1 fita- β antiparalela (figura 9). A principal diferença estrutural entre as duas subfamílias é a profundidade do sítio ativo: cerca de 9 Å nas PTPs clássicas e 6 Å nas DSPs. Esta diferença foi sugerida como causa da seletividade das PTPs clássicas, pois o comprimento da cadeia lateral da tirosina é cerca de 3 Å maior do que a cadeia lateral da serina e treonina^[89, 90].

As enzimas da subfamília das LM-PTPs são específicas para desfosforilação de tirosina. Estas enzimas têm topologia diversa das outras PTPs, mas compartilham a mesma estrutura do sítio ativo^[91]. As LM-PTP são chamadas de baixo peso molecular, porque sua unidade catalítica tem cerca de 140 resíduos, ao contrário dos cerca de 400 encontra-

Figura 9: Representação da estrutura terciária da VHR em dois diferentes ângulos de visão.

dos nas PTPs clássicas. LM-PTPs já foram isoladas de humanos, bovinos e leveduras. Não há consenso sobre sua função *in vivo*^[42], mas sua superexpressão em células de fígado bovino inibe a proliferação celular^[92].

As CDC25s, com três formas homólogas isoladas de humanos, desfosforilam fosfotreonina e fosfotirosina de quinases dependentes de ciclina (CDK2), o que resulta na ativação destas quinases e na progressão do ciclo celular^[93]. Estudos mostraram que o aumento dos níveis de expressão de CDC25A e CDC25B está associado ao aparecimento de câncer e inibidores específicos destas enzimas têm sido investigados como possíveis fármacos^[22]. Embora as CDC25s também sejam DSPs, elas são classificadas como uma outra subfamília, porque têm estrutura primária e terciária distantes das outras subfamílias^[42, 86]. A estrutura terciária das CDC25s é composta por um domínio α/β de 5 fitas- β paralelas centrais, rodeadas por 3 α -hélices na parte inferior e 2 α -hélices na parte superior. Esta estrutura é equivalente à rodanase bacteriana, que catalisa a transferência de enxofre de tiosulfato para cianeto^[94] e cuja função biológica não é conhecida.

1.3.2 Mecanismo de catálise

A identificação dos aminoácidos importantes e de sua função para catálise foi estabelecida por uma série de experimentos cinéticos, de mutação sítio-dirigida e de determinação estrutural (cristalografia e raio-X) nos últimos 15 anos^[42, 43, 86].

A sequência de aminoácidos (HIS ou VAL)–**CYS**–XXX₅–**ARG**–(SER ou THR) é o motivo comum que caracteriza todas as PTPs conhecidas. XXX indica qualquer amino-

Figura 10: Estrutura do P-loop (sítio ativo) do MC entre fenilfosfato e a enzima VHR. As ligações de hidrogênio estão indicadas por linhas tracejadas entre a cadeia protéica principal e lateral da ARG conservada e os oxigênios não ligantes do fenilfosfato, e entre a SER conservada e o enxofre da CYS. Parte dos átomos das cadeias laterais, da cadeia principal e do substrato foram apagados para facilitar a visualização. Legenda para os tipos atômicos: H(branco), C(cinza), C_α (linhas diagonais cruzadas), N(preto), O(linhas diagonais), P(ondulado) e S(linhas cruzadas).

ácido, e os resíduos de cisteína (CYS) e arginina (ARG) são absolutamente conservados e essenciais para catálise. Nas CDC25s, o motivo não contém o resíduo de serina (SER) ou treonina (THR). O motivo é denominado “P-loop”, porque reúne os resíduos do sítio ativo onde o grupo fosfato do substrato se liga. A estrutura forma um laço em que os hidrogênios ligados aos nitrogênios da cadeia principal e do grupo guanidina da ARG coordenam os oxigênios equatoriais do grupo fosfato (figura 10). A geometria do P-loop é complementar aos dois TSs das reações catalisadas (equações 1.9 e 1.10) e exatamente 5 aminoácidos devem estar presentes entre os resíduos CYS e ARG^[42].

Partindo do complexo de Michaelis (MC), as PTPs realizam a hidrólise de monoéster de fosfato ao dividir o processo total em duas reações distintas:

Figura 11: Esquema do mecanismo de reação proposto para catálise pelas PTPs.

$$k_{cat} = \frac{k_2 \cdot k_3}{k_2 + k_3} \quad (1.11)$$

O mecanismo tradicionalmente aceito (figura 11) consiste de uma primeira reação de ataque nucleofílico do grupo tiol ionizado da cadeia lateral da cisteína conservada sobre o fósforo do grupo fosfato do substrato dianiônico. A ligação P–O esterificada é quebrada e um H^+ é doado ao oxigênio do grupo de saída por um ácido geral devidamente posicionado. Os produtos são o grupo de saída protonado, o substrato desfosforilado e a PTP tiofosforilada, via uma ligação covalente P–S tioéster^[85, 95, 96]. Uma PTP1B mutante e tiofosforilada foi isolada, cristalizada e teve sua estrutura determinada por raios-X^[85]. A VHR tiofosforilada também foi diretamente observada por ^{31}P -RMN^[95] e utilizando substrato ^{32}P radioativamente marcado^[96]. Portanto, a reação 1.9 é uma tiólise de monoéster de fosfato e está desacoplada da segunda reação.

A enzima tiofosforilada sofre hidrólise, atacada por uma molécula de água na segunda reação, regenerando a enzima PTP e produzindo fosfato inorgânico. A molécula de água é ativada por uma base geral, que pode ser o mesmo resíduo que funciona como ácido geral na primeira reação. A reação 1.10 é uma hidrólise de monotioéster de fosfato.

Tabela 2: Constantes de velocidade medidas para as duas etapas (reações 1.9 e 1.10) da reação total de hidrólise de pNPP catalisada por PTPs ^a

Enzima	k_2 (s ⁻¹)	k_3 (s ⁻¹)
PTP1B ^[97]	241	12
VHR ^[98]	34	7.6
LM-PTP ^{b [92]}	790	31
CDC25B ^[99]	0.5	1.5

^a condições experimentais: pH = 6 para PTP1B, VHR e LM-PTP, e pH = 8.2 para CDC25B. Temperatura de 30°C para VHR e CDC25B, 3.5°C para PTP1B e 37°C para LM-PTP. As enzimas PTP1B, VHR e LM-PTP têm sua atividade máxima em pH = 6. Já a CDC25B tem atividade máxima e independente do pH para pH > 6. ^b LM-PTP isolada de células de fígado bovino.

A existência de duas reações distintas também é caracterizada por experimentos de cinética transiente, pré-estado estacionário^[60, cap. 1], que permitem a determinação das constantes k_2 e k_3 (tabela 2).

A etapa determinante da velocidade de reação de hidrólise do fosfoéster é a segunda (reação 1.10) para as enzimas PTP1B, VHR e LM-PTP e a primeira (reação 1.9) para a CDC25B. A diferença relativa entre as constantes das duas etapas é menor que 30 vezes para as enzimas PTP1B e LM-PTP e menor que 5 vezes para a VHR e CDC25B. Pela teoria do estado de transição, as diferenças nas constantes correspondem a uma diferença nas barreiras de ativação de cerca de 2 kcal/mol para PTP1B e LM-PTP e 1 kcal/mol para VHR e CDC25B. Portanto, as duas reações têm alturas de barreira similares em todas PTPs. As energias livres de ativação ($\Delta G^\ddagger$) variam de 14 kcal/mol para a reação 1.9 da LM-PTP até 18 kcal/mol para a mesma reação na enzima CDC25B.

1.3.3 Análise do mecanismo proposto

A interação do substrato com o grupo guanidina da arginina no P-loop é essencial para catálise. Mutações ARG→ALA resultam em redução de 10⁴ vezes em k_{cat} e de 10⁶ vezes na velocidade do primeiro evento irreversível, sugerido como a reação 1.9^[91, 100, 101].

O grupo tiol da cisteína no P-loop está em posição ideal para o ataque nucleofílico, alinhado com a ligação P–O que é quebrada^[88, 102]. A mutação CYS→SER, uma simples substituição de enxofre por oxigênio na cadeia lateral, inativa a enzima^[91, 96]. Esta observação é totalmente diferente da reatividade de ésteres de fosfato em solução aquosa, em que nucleófilos centrados em oxigênio são até 10⁷ vezes mais eficientes do que aqueles centrados em enxofre (§1.2.1, p. 30)^[76]. A enzima mutante CYS→SER retém a capacidade

de formar complexos de Michaelis e, portanto, liga-se ao substrato^[88, 102]. Este mutante é de grande utilidade, pois permite que os substratos naturais fosforilados das PTPs sejam identificados^[43, 84].

Nas PTPs clássicas, DSPs e LM-PTPs, o resíduo que funciona como ácido geral na primeira reação e base geral na segunda é um ácido aspártico (ASP), localizado de 30 a 40 resíduos do sítio ativo na direção N-terminal^[42]. O ASP está localizado num laço flexível, chamado “WPD-loop” (do código de uma letra para aminoácidos), que também contém um resíduo triptofano (ou ALA para as DSPs) e uma prolina. Na enzima livre, o ASP está cerca de 10 Å distante do sítio ativo, mas, após a ligação do substrato, o WPD-loop sofre uma mudança conformacional^[86] e a cadeia lateral do ASP fica a cerca de 4 Å do oxigênio esterificado do substrato, posicionada para uma transferência de H^+ ^[88, 102]. A mutação ASP→ALA resulta em redução da atividade catalítica em até três ordens de magnitude^[92, 103, 104]. Embora a existência do ácido geral seja postulada para as CDC25s, a identificação deste resíduo não foi efetuada^[22, 105]. A observação da estrutura cristalográfica indica que nenhuma alça ou resíduo está corretamente posicionado para funcionar como um ácido geral na CDC25A^[106] ou CDC25B^[107]. No entanto, dois resíduos GLU são conservados no P-loop das CDC25s e foram propostos como possíveis ácidos gerais nessas enzimas^[42, 93, 105].

A função postulada para os resíduos de histidina (HIS) conservado no P-loop é de diminuir o pK_a e estabilizar a forma ionizada da CYS nucleofílica via uma rede de ligações de hidrogênio^[42, 102]. A mutação HIS→ALA resulta em diminuição de 10 vezes na k_{cat} e em aumento de 1.5 unidades no pK_a atribuído à CYS^[108]. Na LM-PTP, em que um resíduo de valina (VAL) substitui a HIS, o pK_a atribuído à CYS também é cerca de 1.5 unidades superior àquele atribuído à CYS de outras fosfatases^[108] (vide infra).

Postula-se que o resíduo SER ou THR conservado no P-loop é necessário para catálise, principalmente da reação 1.10, pois estabiliza a regeneração do tiolato na CYS por uma ligação de hidrogênio da cadeia lateral (figura 10)^[42]. A mutação SER→ALA resulta em redução de 100 vezes em k_{cat} , mas em pouca ou nenhuma alteração de k_{cat}/K_M ou de k_2 para as enzimas PTP1B e VHR^[95, 109].

O efeito isotópico cinético (p. 22) pode ser usado para investigar a estrutura do TS de reações enzimáticas^[60, cap. 12, 13 e 14]. Já que o método competitivo é usado para as determinações do KIE em enzimas^[20], o efeito é relativo ao k_{cat}/K_M (normalmente chamado de V/K), o que inclui a porção do mecanismo até a primeira etapa irreversível. A invariância dos KIEs enzimático com o pH e a magnitude dos valores (substancialmente

Tabela 3: Efeito isotópico cinético em reações de hidrólise de pNPP catalisada por PTPs^a

Enzima ^b	¹⁵ (V/K)	¹⁸ (V/K) _{bridge}	¹⁸ (V/K) _{nonbridge}
PTP1B, VHR, YOP51 ^[20]	0.9999–1.0001	1.0118–1.0152	0.9998–1.0003
CDC25A ^[105]	1.0030	1.0357	0.9988
YOP51 e VHR (ASP→ALA) ^[20]	1.0024–1.0030	1.0275–1.0297	1.0019–1.0024
YOP51 (ARG→ALA) ^[20]	1.0012	1.0200	0.9990
VHR (SER→ALA) ^[42]	1.0002	1.0119	1.0001

^a o erro associado às medidas é na última casa decimal. ^b se mais de uma enzima está indicada, os valores dados são as faixas de variação. Mutantes são indicados entre parênteses pela substituição do aminoácido no P-loop.

diferentes de 1 e semelhantes aos valores em solução, tabela 1) indicam que *uma etapa química (em que ocorre reação) é limitante da velocidade* para cada uma das enzimas estudadas. *Portanto, os KIEs caracterizam o TS da formação da enzima tiofosforilada.* A ligação do substrato e a mudança conformacional do WPD-loop parecem ser rápidas e reversíveis, sem impor barreiras cinéticas para catálise^[86].

Os KIE medidos para enzimas PTPs clássicas e VHR selvagens são próximos, sugerindo que os TSs de cada enzima são semelhantes. O valor de ¹⁵(V/K) para as enzimas PTPs clássicas e VHR é próximo de 1, indicando que pouca densidade de carga é deslocada no anel aromático no TS da reação 1.9. Este dado é interpretado como evidência da neutralização do grupo de saída pela transferência do H⁺ do ácido geral para o oxigênio da ligação P–O que se quebra (p. 22). Os mutantes ASP→ALA tem um ¹⁵(V/K) significativamente maior que um e próximo ao de monoésteres em solução (tabela 1), indicando que, neste caso, o grupo de saída é o ânion p-nitrofenóxido. A CDC25A selvagem apresenta um ¹⁵(V/K) típico das enzimas mutantes sem o ácido geral, indicando que, para pNPP como substrato, a catálise acontece sem protonação do grupo de saída. O ¹⁸(V/K)_{nonbridge} menor ou próximo de 1 observado para as enzimas selvagens e para a mutante ARG→ALA indica diminuição da ordem de ligação P–O dos oxigênios não ligantes. Estes valores de ¹⁸(V/K)_{nonbridge} são interpretados como evidência da formação de um TS com caráter mais dissociativo, semelhante ao metafosfato^[20]. Já os valores de ¹⁸(V/K)_{nonbridge} > 1 observados para os mutantes ASP→ALA são semelhantes aos observados na hidrólise de triésteres de fosfato (tabela 1) e indicam um TS mais associativo, com participação do nucleófilo. O valor de ¹⁸(V/K)_{nonbridge} para CDC25A é ainda mais baixo do que nas outras enzimas e foi interpretado como evidência de que as CDC25s utilizam um mecanismo de estabilização do TS diferente das outras fosfatases^[105]. O ¹⁸(V/K)_{bridge} é elevado para todas as enzimas e seus mutantes, e é semelhante ao efeito medido para hidrólise do monoéster de fosfato dianiônico (§1.2.1 e tabela 1). Estes valores sugerem uma quebra substancial da

ligação P–O no TS e caracterizam um mecanismo cuja etapa determinante da velocidade é dissociativa (figura 3). Os KIEs para VHR mutante SER→ALA são próximos dos KIEs para as enzimas selvagens, indicando que esta mutação não altera o TS da reação 1.9.

A velocidade de hidrólise de uma série de derivados de monoésteres de arilfosfatos foi obtida para enzimas selvagens e mutantes (ASP→ASN) que não possuem um ácido geral posicionado para catálise. Para as enzimas PTP1B e VHR selvagens, nenhuma dependência de k_{cat} com o pK_a do grupo de saída foi observada, sugerindo que o grupo de saída é protonado^[95] ou que a quebra da ligação P–O para dissociação do grupo de saída não é uma etapa limitante. Todavia, a CDC25^[93, 105] e as enzimas mutantes^[95] apresentam uma dependência considerável da k_{cat} com o pK_a do grupo de saída (por exemplo, para VHR mutante o $\beta_{lg} = -0.5$)^[95], sugerindo que, nestas enzimas, a quebra da ligação P–O é limitante da velocidade total e que não ocorre protonação do grupo de saída.

As observações estruturais (figura 10) e os experimentos cinéticos e de mutagênese comentados acima sugerem que a singular eficiência catalítica^[1] das PTPs é obtida por estabilização eletrostática do TS (§1.1.2.6)^[20, 42, 86]. A estabilização pode ser atribuída às ligações de hidrogênio entre os oxigênios não ligantes do grupo fosfato e os hidrogênios ligados aos nitrogênios da cadeia principal e da cadeia lateral da ARG^[42, 110–114]. A transferência de H^+ para o grupo de saída também deve contribuir para a catálise (§1.1.2.5).

Assim, as bases moleculares para catálise estão razoavelmente bem estabelecidas, porém, dúvidas ainda persistem. Em particular, os estados de protonação do substrato e dos grupos catalíticos designados experimentalmente têm sido questionados, porque as medidas podem ter diferentes interpretações^[115].

A influência do pH do meio sobre a atividade enzimática pode ser usada para estimar as ionizações de resíduos importantes para catálise. A variação de k_{cat}/K_M com pH descreve a ionização da enzima livre e do substrato nas etapas que envolvem desde a ligação do substrato até a catálise de reação. A dependência de k_{cat} com o pH refere-se à catálise a partir do complexo de Michaelis, especialmente no caso das PTPs cuja etapa química deve ser limitante da velocidade total. Um perfil de pH versus velocidade de reação de hidrólise de monoéster de fosfato catalisada por PTPs é tipicamente em forma de sino, com pH ótimo = 6^[42, 83]. No perfil de k_{cat}/K_M vs. pH a inclinação ascendente (do lado ácido) é igual a +2 e a inclinação descendente (do lado básico) é –1. Para curva de k_{cat} vs. pH, a inclinação ascendente é igual a 1 e a descendente é –1^[103, 104, 116].

A identificação dos grupos tituláveis na enzima e no substrato não é direta uma vez que o substrato, o nucleófilo (grupo tiol da CYS) e o ácido geral (cadeia lateral

da ASP) parecem ter seus pK_a s deslocados para valores próximos de 7^[42]. Tradicionalmente, a inclinação ascendente é atribuída à ionização da cadeia lateral da CYS, porque os pK_a s (pK_a^{CYS}) obtidos dos perfis de atividade contra pH para diferentes enzimas concordam com aqueles determinados por inativação da enzima por iodoacetato: $pK_a^{CYS} = 4.7$ para YOP51, 5.6 para PTP1B, 5.5 para VHR, 7.5 para LM-PTP^[42, 115] e 5.9 para CDC25B^[22, 93]. Embora os pK_a s sejam obtidos de ajustes dos dados a uma equação cinética (portanto, possuem incertezas de cerca de 0.5 unidade), a concordância dos valores de pK_a^{CYS} obtidos por dois métodos experimentais distintos e as evidências cinéticas que suportam o mecanismo de ataque nucleofílico (figura 11) sugerem que a cadeia lateral da CYS está ionizada.

Como a inclinação ascendente no perfil de k_{cat}/K_M vs. pH é +2, outro grupo, além do tiol da CYS, deve estar ionizado para que a ligação do substrato e a catálise ocorram eficientemente. A interpretação mais comum atribui este pK_a no lado ácido à ionização do substrato e, portanto, o diânion do monoéster de fosfato deve se ligar a PTP^[42, 83, 117]. Os resultados de KIE [$^{18}(V/K)_{nonbridge}$, tabela 3] também indicam que os oxigênios não ligantes devam estar desprotonados^[20, 42]. No entanto, a interpretação dos perfis de pH é complicada, porque resíduos estruturalmente importantes, mas não envolvidos diretamente com a catálise, podem afetar as inclinações dos gráficos. Por exemplo, a mutação do ácido glutâmico (GLU) conservado em todas PTPs e que realiza uma ponte salina com a cadeia lateral da ARG do P-loop resultou em anulação da inclinação ascendente do lado ácido do perfil de k_{cat} vs. pH^[116].

A inclinação descendente dos perfis de atividade enzimática contra pH foi atribuída à ionização tanto do substrato como do ácido geral^[42, 115]. PTPs mutantes ASP→ALA apresentam o valor de k_{cat}/K_M constante em pH > 6 e, portanto, a inclinação descendente foi atribuída ao ácido geral que deve estar protonado para catálise nas enzimas selvagens (figura 11)^[103, 104, 116]. Contudo, perfis de pH de enzimas selvagens e de mutantes não podem ser diretamente comparados em alguns casos. A mutação ASP→ALA em LM-PTP^[92], por exemplo, resulta na acumulação da enzima tiofosforilada, pois a energia livre desta espécie é menor do que a energia do MC e, portanto, o k_3 (que compõem k_{cat} , equação 1.11) será medido em relação a espécie tiofosforilada cujos grupos catalíticos podem ter um estado de protonação diferente do MC.

Estudos computacionais utilizando a equação de Poisson-Boltzmann também estimaram o pK_a dos grupos catalíticos nas PTPs^[118, 119]. Embora os resultados dependam da constante dielétrica empregada, que é de difícil estimativa para o interior de uma

macromolécula, entre outras aproximações (por exemplo, a estrutura protéica não é relaxada), os pK_a s obtidos para CYS no P-loop de diferentes PTPs, utilizando uma constante dielétrica igual a 20, concordam^[118] com as medidas experimentais obtidas por inativação da enzima. A orientação dos microdipolos gerados pela cadeia principal do P-loop nas PTPs parece ter um maior impacto na estabilização da forma ionizada do que o campo eletrostático gerado pela α -hélice localizada abaixo do P-loop^[118]. O uso de uma constante dielétrica igual a 4^[119] resultou em pK_a s negativos calculados para o grupo tiol da CYS, em óbvio desacordo com as medidas experimentais. Porém, os pK_a s calculados para o ácido geral da cadeia lateral do ASP no WPD-loop estão próximos daqueles estimados experimentalmente e sugerem que o ASP está protonado no complexo de Michaelis. As principais contribuições para estabilização da forma ionizada no caso da CYS e da forma protonada no caso do ASP também foram atribuídas aos microdipolos da cadeia principal, em especial ao P-loop, e ao grupo guanidina da ARG conservada^[119].

Métodos computacionais e experimentais mostram que a cadeia lateral da CYS nas PTPs apresenta um pK_a menor do que a CYS livre ($pK_a = 8.3$) e que o grupo tiol nas PTPs deve estar ionizado no pH fisiológico (figura 3). No entanto, grande parte dos estudos cinéticos e interpretações mecanísticas é realizada no pH ótimo de atividade das PTPs, em torno de 6^[42], e o grupo tiol pode estar protonado nestas condições.

O estado de protonação tradicionalmente atribuído ao substrato, um diânion, tem sido mais contestado^[115], já que os perfis de atividade enzimática versus pH não têm uma única interpretação (vide supra). A análise das estruturas tridimensionais, determinadas por raios-X dos cristais de PTPs selvagens e mutantes ligados a diferentes substratos e inibidores, sugere que, preferencialmente, duas cargas negativas são acomodadas no sítio ativo^[115]. Portanto, se a cadeia lateral da CYS está ionizada, o substrato deveria se ligar como um monoânion monoprotonado^[115, 120]. Esta hipótese é baseada num argumento eletrostático intuitivo: um substrato dianiônico não “gosta” de interagir com o nucleófilo ionizado (tiolato da cadeia lateral da CYS). Contudo, o P-loop contém uma ARG com uma carga positiva na cadeia lateral e os hidrogênios ligados aos nitrogênios da cadeia principal estão direcionados ao substrato aniônico e, portanto, a repulsão eletrostática entre o substrato e o nucleófilo deve estar minimizada^[111]. Simulações computacionais empregando o método de ligação de valência empírica (EVB, “Empirical Valence Bond”)^{[10, 19][49, cap. 3 e 4]} testaram a reatividade dos substratos mono e dianiônicos para a LM-PTP bovina^[112, 113, 120]. As barreiras de ativação para a reação 1.9 calculadas para a LM-PTP selvagem ligada ao substrato mono ou dianiônico e para o mutante ASP→ALA ligado ao substrato monoaniônico, reproduzem os valores experimentais^[112]. A simulação

do mutante ligado ao substrato dianiônico não foi descrita. No entanto, a energia livre para ligação do substrato dianiônico à LM-PTP foi estimada em 14 kcal/mol acima da energia de ligação do monoânion^[112].

Assim, levantamos as seguintes perguntas sobre as PTPs:

- Qual o estado de protonação do substrato e do grupo nucleofílico no qual a catálise se processa?
- Por que enzimas mutantes CYS→SER são inativas?
- Qual resíduo funciona como ácido geral na catálise das CDC25s?

O objetivo das simulações enzimáticas desenvolvidas (cap. 4) foi responder estas questões. As proteínas tirosina-fosfatases de dupla especificidade, VHR^[88] e CDC25B^[107], foram os modelos protéicos escolhidos. A análise das evidências experimentais e computacionais apresentadas nesta seção como, por exemplo, enfatizado na p. 40, indica que apenas a primeira etapa de catálise, reação 1.9, está relacionada com as questões acima e, portanto, somente esta reação foi estudada.

1.4 Cálculo de estrutura eletrônica molecular

Química quântica ou a determinação da estrutura eletrônica de moléculas é uma das ciências computacionais mais bem sucedidas dos últimos 30 anos. Nesta seção apresentaremos um parte das idéias e do formalismo básico para o cálculo de estrutura eletrônica *ab initio* (§1.4.1) ou por métodos semiempíricos (§1.4.2). Esta informação pode ser usada para otimizar geometrias moleculares (§1.5.1), obter propriedades termodinâmicas (apêndice B) e, assim, calcular caminhos de reação de ésteres de fosfato (cap. 2) e simular a catálise enzimática (cap. 4).

Embora o autor prefira a linguagem da segunda quantização^[121, cap. 1 e 2] devido a sua clareza e elegância, as equações da química quântica serão apresentadas na primeira quantização, porque a descrição dos métodos semiempíricos e suas aproximações nesta linguagem são mais usuais.

Uma nota sobre esta seção e a §1.6. Apenas uma pequena parte das idéias sobre cálculo de estrutura eletrônica e simulação molecular serão apresentadas. A escolha das equações e dos métodos mostrados foi baseada nas técnicas usadas nos capítulos seguintes

desta tese. Não nos aprofundaremos na discussão dos métodos, porque o foco desta tese foi a utilização, ao invés do desenvolvimento ou aprimoramento das metodologias.

1.4.1 Estrutura eletrônica *ab initio*

Ao estudar a estrutura eletrônica de uma molécula, partimos da aproximação de Born-Oppenheimer, que permite-nos tratar o problema da movimentação eletrônica e nuclear separadamente. Esta aproximação considera que os elétrons numa molécula estão se movendo sob o campo dos núcleos fixos, o que é bastante razoável, pois os núcleos movem-se mais lentamente que os elétrons^[122]. O operador hamiltoniano deste sistema, $\hat{\mathcal{H}}$, contém os operadores de energia cinética e potencial dos elétrons, e a repulsão dos núcleos fixos:

$$\hat{\mathcal{H}} = - \sum_i^N \frac{1}{2} \nabla_i^2 - \sum_A^M \sum_i^N \frac{Z_A}{r_{Ai}} + \sum_{i>j}^N \frac{1}{r_{ij}} + \sum_{A>B}^M \frac{Z_A Z_B}{R_{AB}} \quad (1.12)$$

onde ∇^2 é o operador laplaciano, Z_A é o número atômico do átomo no centro A , r_{Ai} é a distância entre o centro A e o elétron i , r_{ij} é a distância entre os elétrons i e j , e R_{AB} é a distâncias entre os centros atômicos A e B . As somatórias são efetuadas sobre os M núcleos atômicos e os N elétrons da molécula e a equação é dada em unidades atômicas (au, 1 au = 627.0595 kcal/mol).

O sistema é descrito por uma *função de onda* Ψ , que é determinada pela solução da equação de Schrödinger eletrônica, não-relativística e independente do tempo:

$$\hat{\mathcal{H}}\Psi = E\Psi \quad (1.13)$$

onde E é a energia do sistema referente ao estado eletrônico descrito por Ψ que é função de $3N$ coordenadas eletrônicas.

A função de onda Ψ depende parametricamente das coordenadas nucleares ($\{\mathbf{R}_A\}$), ou seja, para cada arranjo dos núcleos, teremos uma Ψ e E diferentes. A solução da equação 1.13 sob diferentes arranjos nucleares fornece um potencial efetivo, $E(\{\mathbf{R}_A\})$, pelo qual os núcleos podem se mover, permitindo a descrição dinâmica do sistema molecular (§1.6.2). Este potencial é geralmente uma hipersuperfície, denominada superfície de energia potencial (PES). A exploração de uma porção local e limitada da PES permite que pontos estacionários sejam encontrados (caso eles existam, §1.5.1) e que propriedades termodinâmicas em torno deste pontos sejam calculadas pela separação da movimentação nuclear em contribuições vibracionais, rotacionais e translacionais (apêndice B). Os pontos estacionários na PES conectam os conceitos de estrutura eletrônica com os de reatividade

molecular^[71]. Um ponto de mínimo na PES caracteriza, por exemplo, um reagente ou um intermediário estável de uma reação. Os pontos de sela de primeira ordem (com apenas um único autovalor negativo da matriz hessiana $\mathbf{H}$, §1.5.1) que conectam os mínimos na PES caracterizam TSs. A altura de barreira de uma reação é a diferença entre $E(\mathbf{R}_A)$ quando $\mathbf{R}_A$ assume o arranjo de TS e de reagente.

Podemos definir um *orbital molecular espacial* $\psi(\mathbf{r})$, que será usado na construção da função de onda, como um elemento de um conjunto ortonormal:

$$\int d\mathbf{r} \psi_i^*(\mathbf{r})\psi_j(\mathbf{r}) = \langle \psi_i | \psi_j \rangle = \delta_{ij} \quad (1.14)$$

onde $\delta_{ij} = 0$ se $i \neq j$ ou $\delta_{ij} = 1$ se $i = j$ e ψ^* é o complexo conjugado de ψ . Se este conjunto for completo, qualquer função arbitrária, $\Xi(\mathbf{r})$, poderá ser expandida exatamente em termos de $\psi_i(\mathbf{r})$:

$$\Xi(\mathbf{r}) = \sum_i C_i \psi_i(\mathbf{r}) \quad (1.15)$$

Aqui, usaremos uma mistura indiscriminada da notação introduzida por Dirac onde os “kets” ($|J\rangle$) representam vetores de estado num espaço vetorial complexo, da notação usual de funções $[j(\mathbf{r})]$ da mecânica ondulatória e da notação matricial $(\mathbf{J})$ ^[123, cap. 1].

Na prática, os orbitais moleculares também são expandidos numa *combinação linear de orbitais atômicos*, $\chi_\mu(\mathbf{r})$ ^{[124, cap. 3][125]}:

$$\psi_i = \sum_\mu C_{\mu i} \chi_\mu(\mathbf{r}) \quad (1.16)$$

onde $C_{\mu i}$ são coeficientes constantes a serem determinados. Os $\chi_\mu(\mathbf{r})$ são funções das coordenadas espaciais de um único elétron e formam um conjunto não ortogonal.

O elétron não é descrito apenas por uma função das posições, mas também pelo spin, w , que pode assumir somente dois valores, geralmente denotados $\{\alpha, \beta\}$ (ou spin para cima $\uparrow$ e para baixo $\downarrow$). Podemos construir um *orbital de spin* $\phi(\mathbf{x})$ pela união das funções espaciais $\psi(\mathbf{r})$ e de spin w . Se o conjunto $\{\psi(\mathbf{r})\}$ for ortonormal e completo no espaço de posições, o conjunto $\{\phi(\mathbf{x})\}$ também será ortonormal e completo no espaço de $\mathbf{x} = \{\mathbf{r}, w\}$, que denota a união das coordenadas de spin α, β e de posição $\mathbf{r}$. Assim, definimos a base do espaço de 1-elétron (monoeletrônico) que será usada na construção da função de onda eletrônica.

O produto antissimetrizado dos orbitais de spin forma uma função N -eletrônica,

$\Phi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$, ou um determinante de Slater, $|J\rangle$:

$$|J\rangle = \Phi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \dots & \phi_N(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \dots & \phi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_N) & \phi_2(\mathbf{x}_N) & \dots & \phi_N(\mathbf{x}_N) \end{vmatrix} \quad (1.17)$$

que satisfaz o princípio de exclusão de Pauli^[124, §2.1.3]:

$$\hat{P}_{ij}|J\rangle = -|J\rangle \quad (1.18)$$

onde $\hat{P}_{ij}$ é o operador que permuta a posição dos elétrons i e j .

Já que os determinantes $|J\rangle$ são autofunções do operador hermitiano de 1-elétron, $\hat{h}$ (parte de 1-elétron da hamiltoniana 1.12), e, portanto, constituem um conjunto completo de funções N -eletrônicas^[121, §5.1], podemos representar a função de onda exata do sistema como uma combinação:

$$\Psi = \sum_J C_J |J\rangle \quad (1.19)$$

Assim, determinamos duas hierarquias, que levam a solução exata da equação de Schrödinger (1.13), se^[121, p. 145]:

1. Os orbitais moleculares (equação 1.16), usados para construir os determinantes de Slater, formarem um conjunto completo dentro do espaço de 1-elétron.
2. Todos determinantes que podem ser gerados destes orbitais foram incluídos na expansão da função de onda N -eletrônica (equação 1.19).

Erros serão introduzidos na função de onda caso um destes pontos não seja totalmente satisfeito. O uso de um conjunto incompleto de orbitais atômicos na expansão monoelétrica (equação 1.16) introduz erros que são qualitativamente diferentes daqueles introduzidos por truncagem (ou aproximações) na descrição N -eletrônica (equação 1.19) e vice-versa. Uma descrição pobre no espaço monoelétrônico não pode ser compensada por uma melhor descrição no espaço N -eletrônico (e, novamente, vice-versa), no entanto, algumas vezes um fortuito cancelamento de erro pode ocorrer. Na prática, ambas expansões são incompletas ou truncadas porém a inclusão de mais termos dentro das duas hierarquias sempre levará a uma melhor descrição da função de onda.

A solução da equação de Schrödinger eletrônica é obtida variacionalmente^[121, §4.2].

Podemos reescrever a equação 1.13 em termos de um valor esperado:

$$E[\tilde{\Psi}] = \frac{\langle \tilde{\Psi} | \hat{\mathcal{H}} | \tilde{\Psi} \rangle}{\langle \tilde{\Psi} | \tilde{\Psi} \rangle} \quad (1.20)$$

onde $\tilde{\Psi}$ é uma função tentativa, uma aproximação à função de onda exata Ψ . Assim, para um modelo proposto de espaços mono e N -eletrônicos, a função tentativa é otimizada, ou seja, os coeficientes $\mathbf{C}$ das expansões 1.16 e 1.19 são variados de maneira que a energia calculada $E[\tilde{\Psi}]$ seja estacionária. Esta energia é sempre um limite superior para a energia exata E do sistema. A função de onda exata Ψ corresponde a um ponto estacionário do valor esperado da hamiltoniana $\hat{\mathcal{H}}$. Este ponto é um mínimo para a função de onda no estado fundamental.

O modelo mais simples para se determinar a estrutura eletrônica *ab initio* de uma molécula é o método de Hartree-Fock (HF)^{[121, §5.4 e cap. 10][124, cap. 3][126, cap. 17]}, em que a função de onda eletrônica é aproximada por um único determinante de Slater e a energia é otimizada com relação apenas às variações dos orbitais moleculares (equação 1.16). O operador hamiltoniano é substituído pelo operador:

$$\hat{f} = \hat{h} + \hat{V} \quad (1.21)$$

A parte de 1-elétron ($\hat{h}$) da hamiltoniana 1.12 é preservada, mas a parte de 2-elétrons é substituída pelo potencial de Fock, $\hat{V}$:

$$\hat{V}_i \phi_i(\mathbf{x}_1) = \sum_j \int d\mathbf{x}_2 \phi_j^*(\mathbf{x}_2) \frac{(1 - \hat{P}_{12})}{r_{12}} \phi_j(\mathbf{x}_2) \quad (1.22)$$

onde temos que o efeito do potencial de Fock sobre um único orbital molecular $\phi_i(\mathbf{x}_1)$, depende do efeito *médio* de todos outros orbitais j . Ou seja, a aproximação de Hartree-Fock é uma aproximação de campo médio, isto é, os elétrons não interagem detalhadamente entre si mas com um potencial médio, $\hat{V}$.

A minimização do valor esperado da energia (equação 1.20) utilizando a aproximação de HF resulta nas equações de autovalor de Hartree-Fock canônicas^{[121, §10.3][124, §3.2]}:

$$\hat{f}_i \phi_i(\mathbf{x}_1) = \varepsilon_i \phi_i(\mathbf{x}_1) \quad (1.23)$$

onde o autovalor ε_i é a energia do orbital de spin $\phi_i(\mathbf{x}_1)$. Entretanto, como o $\hat{f}$ é definido em termos de seus próprios autovetores (equação 1.22), as energias e os orbitais canônicos só podem ser obtidos através de um procedimento iterativo, conhecido como campo autoconsistente.

Para uma molécula de camada fechada com orbitais de spin restritos, *i.e.*, que tem a mesma parte espacial para o orbital com spin α ou spin β (podemos lidar com casos mais gerais, mas estas suposições simplificam bastante o desenvolvimento aqui), a introdução de um conjunto base de orbitais atômicos (equação 1.16) e a integração das coordenadas de spin resultam nas equações de Roothaan-Hall^{[125][121, §10.6][124, §3.4]}:

$$\mathbf{FC} = \mathbf{SC}\varepsilon \quad (1.24)$$

onde usamos a notação matricial. $\mathbf{F}$ é a matriz de Fock, $\mathbf{S}$ é a matriz de recobrimento dos orbitais atômicos (não ortonormais), $\mathbf{C}$ é a matriz dos coeficientes $C_{\mu i}$ otimizados, que formam os orbitais moleculares, e ε é a matriz diagonal com as energias dos orbitais. Os elementos da matriz do operador de Fock (equação 1.21) são:

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma) \right] \quad (1.25)$$

$$P_{\mu\nu} = 2 \sum_i^{N/2} C_{\mu i} C_{\nu i}^* \quad (1.26)$$

onde $\mathbf{H}^{core}$ é a representação matricial de $\hat{h}$ (equação 1.21), $\mathbf{P}$ é a matriz de densidade, μ, ν, λ e σ são orbitais atômicos e $(\mu\nu|\lambda\sigma)$ são as integrais moleculares de repulsão eletrônica^[121, cap. 9]. O valor esperado da energia eletrônica (E_{elec}^{HF}) e a energia total (E_{tot}) são:

$$E_{elec}^{HF} = \frac{1}{2} \sum_{\mu\nu} P_{\nu\mu} (H_{\mu\nu}^{core} + F_{\mu\nu}) \quad (1.27)$$

$$E_{tot}^{HF} = E_{elec}^{HF} + \sum_{A>B}^M \frac{Z_A Z_B}{R_{AB}} \quad (1.28)$$

Para uma base atômica completa, a diferença entre a energia exata (obtida por exemplo, por um procedimento de interação total de configurações)^[121, cap. 11] e a energia HF (E_{elec}^{HF}) é chamada de energia de correlação, E_{corr} . Além do procedimento de interação de configurações, o método de clusters acoplados^[121, cap. 13] pode resgatar grande parte da energia de correlação, e, portanto, aproximar a energia eletrônica calculada na aproximação de HF da energia exata. Porém, ambos métodos são computacionalmente custosos e sua aplicação é limitada a átomos e moléculas pequenas. Um procedimento menos dispendioso para resgatar parte de E_{corr} é a teoria de perturbação, em particular a forma desenvolvida por Møller e Plesset^[127] (MP), que pode ser aplicada em moléculas relativamente maiores e nos casos onde Ψ é dominada por uma única configuração N -eletrônica. A solução exata da equação de Schrödinger eletrônica pode ser atingida sistematicamente

por uma expansão ordem-a-ordem da função de onda e da energia. A expansão é extensível por tamanho ordem-a-ordem^[121, §14.2.6] mas não é variacional, ou seja, não representa um limite superior para energia exata. Na teoria de perturbação MP, a hamiltoniana é particionada em^{[121, cap. 14][124, cap. 6]}:

$$\hat{\mathcal{H}} = \hat{f} + \hat{\Theta} + h_{nuc} \quad (1.29)$$

onde h_{nuc} é o termo de repulsão nuclear (equação 1.12) e $\hat{\Theta}$ é o *potencial de flutuação*:

$$\hat{\Theta} = \sum_{i>j} \frac{1}{r_{ij}} - \hat{V} \quad (1.30)$$

ou seja, $\hat{\Theta}$ representa a diferença entre o potencial coulômbico de repulsão de 2-elétrons e o potencial de Fock.

Usando o maquinário padrão de teoria de perturbação^{[121, §14.1 e §14.2][124, §6.1 e §6.5]}, temos que:

$$E_{elec}^{HF} = E_{MP}^{(0)} + E_{MP}^{(1)} \quad (1.31)$$

onde $E^{(n)}$ representa a energia da perturbação de n -ésima ordem. Portanto, a primeira correção à energia HF ocorre somente na segunda ordem:

$$E^{(2)} = - \sum_{n \neq 0} \frac{|\langle 0 | \hat{\Theta} | n \rangle|^2}{E_n^{(0)} - E_{MP}^{(0)}} \quad (1.32)$$

onde a somatória corre sobre todas configurações eletrônicas do sistema, $|n\rangle$, exceto a de estado fundamental, $|0\rangle$, e $E_n^{(0)}$ é a energia de ordem zero da configuração $|n\rangle$. Estas configurações são excitações dos elétrons dos orbitais moleculares $\phi_i(\mathbf{x}_1)$ ocupados (denotados pelos subscritos i e j) para os orbitais moleculares desocupados (denotados pelos subscritos a e b) determinados no procedimento do campo autoconsistente de solução da equação 1.24. De acordo com o teorema de Brillouin^[121, p. 441] e com o fato de que excitações triplas ou de maior ordem não se acoplam com $|0\rangle$ ^[121, p. 741], apenas duplas excitações interagem com $|0\rangle$ e compõem as configurações $|n\rangle$. Assim, a equação 1.32 assume a forma:

$$E_{MP}^{(2)} = - \sum_{\substack{a>b \\ i>j}} \frac{|g_{aibj}^a|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \quad (1.33)$$

onde g_{aibj}^a são as integrais de repulsão eletrônica antissimetrizadas na base de orbitais moleculares que são obtidas da transformação (via a matriz de coeficientes $\mathbf{C}$) das integrais de repulsão eletrônica na base de orbitais atômicos, $(\mu\nu|\lambda\sigma)$ ^[128]. A energia de segunda ordem MP é $E_{MP2} = E_{tot}^{HF} + E_{MP}^{(2)}$. Na prática, é comum que apenas os elétrons

de valência sejam excitados (ou seja i e j designariam orbitais moleculares ocupados de valência) nos cálculos de energia de correlação^[121, §8.3.1]. Esta aproximação é justificada, porque a estrutura eletrônica dos elétrons de caroço^[129] muda muito pouco durante diversos processos moleculares (como reações químicas) e, portanto, a correlação entre estes elétrons se cancelaria ao longo das etapas do processo.

Em média, cerca de 99.5% da energia eletrônica fundamental exata (equação 1.13) de uma molécula é obtida pelo método de HF. A energia de correlação proveniente de excitações duplas (resgatadas no método MP2) é responsável por mais 0.5% da energia exata^[121, p. 853]. A energia de correlação de excitações triplas, correção para energia do ponto-zero (apêndice B) e efeitos relativísticos contribuem, em média, respectivamente 0.02%, -0.03% e 0.05%^[121, p. 853]. Portanto, a E_{MP2} contém quase que a totalidade da energia eletrônica molecular e, assim, deve ser um nível de cálculo apropriado para os estudos de reatividade desenvolvidos nesta tese.

Outra metodologia para cálculo da estrutura eletrônica molecular, que pode resgatar (pelo menos parte da) E_{corr} , é a teoria do funcional da densidade (DFT)^[46, 130]. Nesta teoria, a função de onda $\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ é substituída como variável básica pela densidade eletrônica $\rho(\mathbf{r})$ que depende de apenas três coordenadas espaciais (ao contrário de Ψ que depende de $3N$ coordenadas espaciais). Todas propriedades moleculares do estado fundamental podem ser obtidas da $\rho(\mathbf{r})$, que também é uma quantidade mensurável em experimentos de difração. As equações de Kohn-Sham são um esquema computacional prático e geralmente aplicável, baseado na DFT^[46, cap. 7]:

$$\left(-\frac{1}{2}\nabla^2 + v(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \right) \varphi_j(\mathbf{r}) = \epsilon_j \varphi_j(\mathbf{r}) \quad (1.34)$$

onde $v(\mathbf{r})$ é um potencial externo (os núcleos atômicos fixos, por exemplo), $\varphi_j(\mathbf{r})$ é um orbital de Kohn-Sham e ϵ é a energia deste orbital. Esta equação é similar à 1.23, mas é exata por princípio, caso a expressão exata para energia E_{xc} , que inclui os efeitos de “troca” (proveniente do princípio de exclusão de Pauli)^[124, §2.1.3] e correlação eletrônica, seja usada. No entanto, a forma exata de E_{xc} é desconhecida. Uma forma bastante usada atualmente é a expressão semiempírica:

$$E_{xc} = E_{xc}^{DFT} + a_0(E_x^{exata} - E_x^{DFT}) \quad (1.35)$$

que mistura a energia de “troca” exata com as energias de “troca” e “troca”-correlação calculadas por uma aproximação em DFT (como a densidade local ou o gradiente generalizado)^[46, §6.1 e §6.7]. A E_x^{exata} é a energia de “troca” do determinante de Slater dos

orbitais $\varphi_j(\mathbf{r})$ de Kohn-Sham, que embora conceitualmente diferente, é numericamente muito próxima da energia de “troca” obtida do determinante de Slater dos orbitais $\phi_i(\mathbf{x})$ de HF^[131]. O parâmetro de mistura a_0 é ajustado para que a energia total calculada reproduza dados experimentais de termoquímica^[131, 132]. O método B3LYP é um esquema híbrido entre o método HF e o de Kohn-Sham, que utiliza uma equação semiempírica similar à 1.35 e calcula observáveis moleculares de boa qualidade^[132]. Não desenvolveremos este assunto aqui, porque a DFT é pouco usada nesta tese. O leitor interessado pode consultar outras monografias^[46, 133].

1.4.1.1 Conjunto base atômico

As bases atômicas χ normalmente usadas em cálculos moleculares são composições [121, cap. 6 e 8].

$$\chi_{\alpha lm} = R_{l\alpha}(r)Y_{lm}(\theta, \varphi) \quad (1.36)$$

$$R_{l\alpha}(r) = K(\alpha, l, r)\exp[-\alpha r^2] \quad (1.37)$$

entre uma parte angular que possui uma representação por harmônica esférica, $Y(\theta, \varphi)$ (equação 1.36), ou cartesiana e uma parte radial, $R_{l\alpha}(r)$, composta por uma função gaussiana e por um termo $K(\alpha, l, r)$ que engloba a constante de normalização e os termos monomiais em r . Os parâmetros l e m são remanescentes dos números atômicos de momento angular e magnético, respectivamente, obtidos da solução da equação de Schrödinger para um elétron num campo central. A parte radial também pode ser composta por uma combinação linear de gaussianas $R_{l\alpha}(r)$ (chamadas primitivas), formando um conjunto contraído. Dizemos que uma contração é generalizada se todas as primitivas de um determinado momento angular (l) em um átomo misturam-se livremente e a contração é segmentada se cada primitiva contribui apenas para um orbital atômico.

Para designar um conjunto base e sua contração utilizamos uma notação geral. Por exemplo, um conjunto com 9 primitivas de momento angular s e 5 primitivas p contraídas em 3 orbitais atômicos s e 2 orbitais p é definido pela notação $(9s, 5p)/[3s, 2p]$.

As bases atômicas desenvolvidas por Pople et al.^[134], denominadas 6-31G são do tipo “valência-dividida”, pois possuem uma representação de único- ζ , em referência ao expoente ζ usado na parte radial de orbitais atômicos de Slater^[121, §6.5]:

$$R_{l\zeta_{nl}}(r) = K(l, \zeta_{nl}, r)\exp[-\zeta_{nl}r] \quad (1.38)$$

para a camada eletrônica de caroço (interna) e duplo- ζ para a camada de valência. O termo $K(l, \zeta_{nl}, r)$ tem o mesmo significado descrito acima (equação 1.37). Para os átomos do primeiro período (carbono, oxigênio, etc.), a contração da base 6-31G é $(10s, 4p)/[3s, 2p]$, seguindo o esquema segmentado, onde 6 primitivas s são contraídas para 1 orbital de caroço, e as 4 primitivas de valência (tanto s como p) são contraídas em 2 orbitais (um com 3 primitivas contraídas e outro com 1 primitiva). Para átomos do segundo período (como enxofre e fósforo) a contração é $(16s, 10p)/[4s, 3p]$ e para átomos leves (hidrogênio e hélio) apenas a camada de valência (31) está presente e sua contração é $(4s)/[2s]$. O conjunto 6-311^[135] segue o mesmo esquema de “valência-dividida”, porém, a camada de valência é de triplo- ζ (311). A contração para átomos leves, do primeiro e do segundo período é respectivamente: $(5s)/[3s]$, $(11s, 5p)/[4s, 3p]$ e $(12s, 9p)/[6s, 5p]$. O conjunto base triplo- ζ desenvolvido por Ahlrichs et al.^[136], TZV, também é segmentado e apresenta as seguintes contrações para átomos leves, do primeiro e do segundo período: $(5s)/[3s]$, $(10s, 6p)/[6s, 3p]$ e $(12s, 9p)/[7s, 5p]$, respectivamente.

Os conjuntos base acima contém apenas funções da mesma simetria que os orbitais atômicos ocupados nos átomos isolados. Em virtude da menor simetria das moléculas em relação aos átomos isolados e das interações com o campo elétrico dos núcleos e elétrons vizinhos, funções de simetria diferente daquela dos orbitais ocupados podem contribuir para função de onda^[137, cap. 6]. Portanto, primitivas que permitem a polarização da nuvem eletrônica são incluídas no conjunto base. Funções com maior momento angular, por exemplo, com simetria d , são ortogonais às funções s e p e acomodam as rápidas respostas da função de onda em regiões próximas ao núcleo (ou entre dois elétrons) que não são acomodadas pelas funções s e p , e aumentam a flexibilidade da descrição angular da função de onda eletrônica. Por outro lado, a inclusão de funções com o mesmo momento angular, mas com maior número de nós (maior número quântico principal n), resulta em aumento da flexibilidade radial da função de onda. Ambos tipos de flexibilização também são importantes para resgatar contribuições para energia de correlação dinâmica^[138]. A adição de primitivas com expoente α bem menor (equação 1.36), chamadas primitivas difusas, auxilia a descrição da camada externa de valência, necessária para moléculas com a nuvem eletrônica expandida, como ânions ou estados de Rydberg^[138].

A notação “*” empregada, por exemplo, nas bases 6-31G* e 6-311G*, representa a adição de uma função de polarização para cada átomo pesado, ou seja, a adição de uma primitiva não contraída com momento angular d . O uso de “**”, como em 6-31G**, significa a adição de uma função de polarização (p) aos átomos leves (H e He), além daquela dos pesados. A notação também pode indicar explicitamente quais funções de

polarização são adicionadas à base. Por exemplo, os conjuntos 6-31G* e 6-31G(d) são equivalentes e o conjunto 6-311+G(2df,2p) indica a adição de 2 funções tipo d e 1 função f aos átomos pesados e 2 funções p aos átomos leves. A notação “+” empregada, por exemplo, nas bases 6-31+G e 6-311+G^[139], representa a adição aos átomos pesados de uma primitiva difusa não contraída para cada momento angular (s e p) da camada de valência. O uso de “++”, como em G-31++G, significa a adição de uma função difusa (s) aos átomos leves, além daquelas dos átomos pesados.

Os conjuntos base apresentados acima são incompletos. Uma fonte de erro em computações de estrutura eletrônica que utilizam tais conjuntos é o erro de superposição do conjunto base (BSSE)^{[121, §8.5][140, 141]}. O BSSE manifesta-se principalmente na estabilização artificial de energias de interação e complexação calculadas, e pode ser corrigido utilizando-se o procedimento contra-ponto^[140]:

$$\Delta E_{CP}(R) = E^{AB}(R) - E^{A(AB)}(R) - E^{B(AB)}(R) \quad (1.39)$$

onde E_{CP} representa a energia corrigida ou livre do BSSE, E^{AB} representa a energia calculada para o complexo AB e $E^{X(AB)}(R)$ representa a energia calculada para a espécie $X = A$ ou B , usando o conjunto base do complexo AB na distância R .

1.4.2 Estrutura eletrônica semiempírica

Os métodos *ab initio* e os recursos computacionais existentes antes dos anos 1970 eram ineficientes e incapazes de tratar quantitativamente moléculas com mais do que um par de átomos pesados. Conseqüentemente, aproximações foram sugeridas^[142] e métodos simplificados desenvolvidos com o objetivo de determinar a estrutura eletrônica e calcular propriedades de moléculas de interesse químico com maior rapidez. Estes métodos simplificados são chamados de semiempíricos em virtude do uso de parâmetros retirados diretamente ou derivados de medidas experimentais nas hamiltonianas aproximadas

Os métodos semiempíricos mais utilizados atualmente (MNDO, AM1, etc.)^[143–145] são derivados do formalismo da combinação linear de orbitais atômicos e do método de Hartree-Fock (§1.4.1). A principal aproximação é a negligência do recobrimento orbital diferencial diatômico (NDDO), que pode ser formalmente justificada se o conjunto base atômico for *ortogonal*^[146]. Todos os termos provenientes do recobrimento de orbitais atômicos localizados em átomos diferentes são zerados, a matriz de recobrimento, $\mathbf{S}$,

reduz-se à matriz identidade e a equação 1.24 do método de HF reduz-se a^[142, 143, 147]:

$$\mathbf{FC} = \mathbf{C}\varepsilon \quad (1.40)$$

Uma consequência da aproximação NDDO, responsável em grande parte pela eficácia computacional destes métodos, é a anulação de todas integrais de repulsão eletrônica $(\mu\nu|\lambda\sigma)$ de 3 ou 4-centros atômicos^[148]. A segunda aproximação importante é que somente os elétrons de valência entram no cálculo da estrutura eletrônica e os elétrons de caroço^[129] são “condensados” juntos às cargas nucleares. Um conjunto base atômico mínimo (único- ζ , contendo apenas funções de momento angular s para hidrogênio e s e p para os átomos pesados do primeiro e segundo períodos, ou seja, μ, ν, λ e $\sigma = \{s, p\}$), é utilizado para descrever os elétrons de valência. Se assumirmos que os orbitais atômicos χ_μ e χ_ν estão no centro atômico A , e χ_λ e χ_σ estão no centro atômico B , os elementos da matriz de Fock (equação 1.25), após a aproximação NDDO, são^[143]:

$$F_{\mu\mu} = U_{\mu\mu} + \sum_B V_{\mu\mu,B} + \sum_\nu P_{\nu\nu} \left[(\mu\mu|\nu\nu) - \frac{1}{2}(\mu\nu|\mu\nu) \right] + \sum_B \sum_{\lambda,\sigma} P_{\lambda\sigma} (\mu\mu|\lambda\sigma) \quad (1.41)$$

$$F_{\mu\nu} = \sum_B V_{\mu\nu,B} + \frac{1}{2}P_{\mu\nu}[3(\mu\nu|\mu\nu) - (\mu\mu|\nu\nu)] + \sum_B \sum_{\lambda,\sigma} P_{\lambda\sigma} (\mu\nu|\lambda\sigma) \quad (1.42)$$

$$F_{\mu\lambda} = \beta_{\mu\lambda} - \frac{1}{2} \sum_\nu \sum_\sigma P_{\nu\sigma} (\mu\nu|\lambda\sigma) \quad (1.43)$$

onde os termos têm os seguintes significados:

1. $U_{\mu\mu}$: energias de 1-centro e de 1-elétron (provenientes do operador $\hat{h}$, equação 1.21);
2. $V_{\mu\nu,B}$: atração de 2-centros e 1-elétron, da distribuição eletrônica $\chi_\mu\chi_\nu$ em A pelo caroço em B ;
3. $\beta_{\mu\lambda}$: integral de ressonância de 1-elétron e 2-centros;
4. $(\mu\mu|\nu\nu) = g_{\mu\nu}$ e $(\mu\nu|\mu\nu) = h_{\mu\nu}$: integrais coulômbica ($g_{\mu\nu}$) e de “troca” ($h_{\mu\nu}$) de 1-centro e 2-elétrons;
5. $(\mu\nu|\lambda\sigma)$: integrais de repulsão de 2-elétrons em 2-centros.

Os orbitais moleculares (ou a matriz de coeficientes $\mathbf{C}$) são obtidos pela solução auto-consistente da equação 1.40, assim como no método HF, e a energia eletrônica (E_{elet}^{NDDO}) é dada pela expressão 1.27, utilizando os elementos da matriz de Fock das equações 1.41

a 1.43 e os elementos da matriz $H_{\mu\nu}^{core}$ obtidos dos termos de 1-elétron $U_{\mu\mu}$, $V_{\mu\nu}$ e $\beta_{\mu\nu}$. Os termos nos itens 1 a 5 acima, necessários para construção da matriz de Fock e avaliação da repulsão entre os caroços nucleares (E_{core}^{NDDO} , equação 1.47), não são obtidos como no procedimento *ab initio*, mas são substituídos por dados experimentais e expressões empíricas que contêm parâmetros ajustáveis otimizados para reproduzir observáveis experimentais ou cálculos de alta acurácia. Os parâmetros são obtidos para cada elemento químico e seus respectivos orbitais no conjunto base mínimo. Espera-se que este ajuste corrija pelo menos parte dos erros introduzidos pelas aproximações (NDDO e elétrons de valência com base atômica mínima) e compense a ausência da correlação eletrônica natural do método HF. A energia total calculada no método semiempírico é:

$$E_{tot}^{NDDO} = E_{elet}^{NDDO} + E_{core}^{NDDO} \quad (1.44)$$

Os parâmetros semiempíricos de 1-centro $U_{\mu\mu}$, $g_{\mu\nu}$ e $h_{\mu\nu}$ [149, cap. 15] são obtidos usando um procedimento [150] baseado no método de Oleari [151], em que as energias calculadas para diferentes estados de valência (ou configurações eletrônicas) de átomos e íons atômicos são ajustados aos valores espectroscópicos correspondentes. Nas parametrizações MNDO [143] e AM1 [144] originais, g_{ss} , g_{sp} , $g_{pp'}$, g_{p^2} e h_{sp} são aqueles obtidos diretamente das medidas espectroscópicas pelo procedimento de Oleari para cada átomo [150, 151], mas U_{ss} e U_{pp} são otimizados no procedimento de ajuste (vide infra).

Todas integrais de repulsão de 2-elétrons em 2-centros, $(\mu\nu|\lambda\sigma)$, são avaliadas por uma expansão multipolar truncada, com os limites $R_{AB} \rightarrow \infty$ e $R_{AB} \rightarrow 0$ empiricamente corrigidos (apêndice A).

Os termos de 2-centros são calculados por:

$$\beta_{\mu\lambda} = \left(\frac{\beta_{\mu}^A + \beta_{\lambda}^B}{2} \right) S_{\mu\lambda} \quad (1.45)$$

$$V_{\mu\nu,B} = -z_B(\mu^A \nu^A | s^B s^B) \quad (1.46)$$

onde β_s^A e β_p^A são parâmetros ajustáveis para cada átomo A e $z_B = (Z_B - n_{core})$, onde n_{core} é o número de elétrons de caroço do átomo B . Um orbital atômico tipo s é idealizado no átomo B (s^B) para representar o caroço atômico. A matriz de recobrimento dos orbitais atômicos, $S_{\mu\lambda}$, é avaliada analiticamente pelo recobrimento entre orbitais atômicos de Slater (equação 1.38) [121, §6.5] cujos expoentes ζ_s e ζ_p são otimizados no procedimento de ajuste (vide infra). A expressão 1.45 é a maior responsável pela descrição da ligação química entre os átomos nos métodos semiempíricos [143].

A repulsão entre os caroços nucleares é dada por:

$$E_{core}^{NDDO} = \sum_{A < B} z_A z_B [(s^A s^A | s^B s^B) + f_{core}] \quad (1.47)$$

onde $f_{core}(A, B)$ assume diferentes formas empíricas para reduzir a excessiva repulsão eletrostática gerada pelo caroços em $z_A z_B / R_{AB}$ ^[143, 144]. Na parametrização MNDO, temos^[143]:

$$f_{core}^{MNDO} = (s^A s^A | s^B s^B) \{ \exp[-\alpha_A R_{AB}] + \exp[-\alpha_B R_{AB}] \} \quad (1.48)$$

e nas parametrizações AM1 e PM3, f_{core} é acrescentada de funções gaussianas que novamente buscam corrigir a excessiva repulsão entre os caroços^[144]:

$$f_{core}^{AM1} = f_{core}^{MNDO} + (R_{AB})^{-1} \left\{ \sum_{i=1}^4 K_i^A \exp[L_i^A (R_{AB} - M_i^A)^2] + \sum_{j=1}^4 K_j^B \exp[L_j^B (R_{AB} - M_j^B)^2] \right\} \quad (1.49)$$

Portanto, para cada átomo pesado (para hidrogênio os parâmetros que contêm o orbital p são nulos, já que utilizamos um conjunto base mínimo) no método MNDO original, temos:

- sete parâmetros ($U_{ss}, U_{pp}, \zeta_s, \zeta_p, \beta_s, \beta_p$ e α) otimizados;
- cinco parâmetros ($g_{ss}, g_{sp}, g_{pp'}, g_{p^2}$ e h_{sp}) retirados de dados experimentais^[150, 151];
- cinco parâmetros (D_1, D_2, ρ_0, ρ_1 e ρ_2) derivados dos parâmetros otimizados e experimentais (apêndice A).

Nas parametrizações originais dos métodos AM1 e PM3, temos:

- mais doze parâmetros (K_i, L_i e M_i , onde i varia de 1 a 4) otimizados, além dos sete parâmetros otimizados no MNDO

Adicionalmente, no método PM3 e em algumas parametrizações do AM1 (por exemplo, para os átomos de enxofre e fósforo)^[152, 153], os termos de 1-centro ($g_{\mu\nu}$ e h_{sp}) também são otimizados no procedimento de ajuste.

A qualidade de um método semiempírico depende fortemente da parametrização^[148]. O procedimento de otimização dos parâmetros consiste em encontrar o valor ótimo do conjunto de K parâmetros $\{X_k\} (k = 0, 1, \dots, K)$ pelo ajuste dos valores $P_m (m = 0, 1, \dots, M)$ calculados aos observados de M propriedades de um grupo de moléculas ou reações de

referência que formam o chamado conjunto de treinamento. Assim, procura-se minimizar a função erro:

$$P = \sum_m^M W_m [P_m^{calc} - P_m^{obs}]^2 \quad (1.50)$$

onde P_m^{calc} e P_m^{obs} são respectivamente as propriedades moleculares calculadas pelo método NDDO e observadas (experimentalmente ou obtidas de cálculos de alta acurácia) e W_m é o peso dado a cada propriedade molecular. As propriedades geralmente utilizadas^[143–145] são energia total (ou calor de formação), geometrias (distâncias de ligação, ângulos entre ligações e diedrais), momentos de dipolo, potenciais de ionização, etc.

O número de parâmetros ajustáveis por elemento pode chegar a $K = 24!$ Na otimização simultânea dos parâmetros de alguns átomos (digamos H, C, O, P e S), a procura de um mínimo global no espaço de parâmetros é um problema bastante complexo (§1.5.1). Diferentes procedimentos de otimização não-linear já foram empregados:

- otimização por mínimos quadrados, sem o uso de derivadas da função P (equação 1.50)^[143];
- otimização de primeira-ordem (§1.5.1)^[124, §C.5] utilizando derivadas de P com respeito aos parâmetros^[145, 153];
- otimização por algoritmo simplex (§1.5.2)^{[154, §10.4][150]}
- otimização global com algoritmos genéticos (§1.5.2)^[155–157].

Duas complicações são notórias no procedimento de otimização: determinadas propriedades moleculares são mais sensíveis do que outras à variação dos parâmetros; e existem interrelações (dependências) entre os parâmetros otimizados. Ao conhecimento do autor, não existe um estudo das interrelações dos parâmetros e da sensibilidade das propriedades que facilitariam o procedimento de otimização. Portanto, a otimização de parâmetros semiempíricos é um problema difícil e tedioso, e transformou-se numa espécie de “arte”^[144, 158].

O principal problema no uso de métodos semiempíricos é que *a priori* não sabemos se o método escolhido será adequado ao estudo do sistema e das propriedade que estamos interessados e, portanto, o método semiempírico deve ser testado ou calibrado em comparações com valores de referência. Caso contrário, em razão da natureza das aproximações e do caráter empírico da parametrização, não existe garantia que o estudo de uma propriedade ou de uma molécula diferente daquelas usadas no conjunto de treinamento em que o método foi ajustado resultará em propriedades calculadas com a mesma

qualidade obtida no conjunto treino. Ou seja, o tipo e a magnitude do erro, associados ao cálculo semiempírico, não são sistemáticos^[148]. Da mesma forma, também é difícil afirmar se o acordo entre os valores calculados e observados é causado por uma correta descrição da estrutura eletrônica pela hamiltoniana semiempírica ou por um fortuito cancelamento de erro. Portanto, é difícil afirmar que obtivemos “*a resposta certa pela razão correta*”^[138].

Uma comparação constatou que as parametrizações semiempíricas (MNDO e AM1) calculam propriedades mais próximas às observadas experimentalmente do que o método HF utilizando um conjunto base mínimo^[159]. Outra comparação^[158] para um conjunto de 56 moléculas neutras contendo apenas átomos de H, C, N e O, de camada fechada e no seu estado fundamental, demonstrou que o erro absoluto médio para os calores de formação calculados em comparação com os calores experimentais foi de 6.8, 5.1, 4.1, 2.3 e 1.3 kcal/mol para os métodos MNDO, AM1, PM3, B3LYP^[132] e G2^[160], respectivamente. Os métodos B3LYP e G2 (um esquema extrapolativo de cálculo) utilizaram conjuntos base de qualidade igual ou superior a duplo- ζ para camada de valência. Portanto, os métodos NDDO calculam energias com cerca do dobro ao triplo do erro médio observado para o método B3LYP. O erro observado nos métodos semiempíricos aumenta para moléculas contendo átomos em estados de oxidação ou funções orgânicas diferentes daquelas usadas na parametrização. Os erros absolutos médios para calores de formação de moléculas neutras e de camada fechada contendo átomos do primeiro e do segundo período ultrapassam 9 kcal/mol para o método MNDO e 7 kcal/mol para os métodos AM1 e PM3^[148].

As seguintes propriedades ou espécies são relevantes para os estudos conduzidos no capítulo 4 e podem apresentar incertezas significativas se obtidas pelos métodos semiempíricos originais^[147, 161]:

- Energias de ligação, se um dos átomos possui pares eletrônicos não ligantes;
- Energias relativas para compostos estericamente volumosos, hipervalentes^[162], ou com elétrons deslocalizados;
- Alturas de barreira de reação (superestimadas);
- Afinidades protônicas.
- Interações intermoleculares, principalmente, ligações de hidrogênio. AM1 corrige a energética para ligações de hidrogênio, mas introduz mínimos inexistentes na PES;

Diferentes parametrizações e modificações das aproximações NDDO foram sugeridas para contornar estas limitações e ampliar o uso dos métodos semiempíricos^[148, 158, 163–165].

A equação 1.40 assume que o recobrimento entre os orbitais atômicos é nulo^[142] porém o conjunto base usado não é ortogonal e, assim, os efeitos de “troca” de Pauli não são incorporados à hamiltoniana semiempírica corretamente. Thiel^[148] introduziu correções para ortogonalização dos orbitais atômicos nos termos $V_{\mu\nu}$ e $\beta_{\mu\lambda}$, e, após a reparametrização do método, o erro absoluto médio dos calores de formação calculados para moléculas neutras e de camada fechada, contendo átomos do primeiro e do segundo período, diminuiu para cerca de 3.5 kcal/mol (vide supra).

A inclusão de orbitais atômicos de simetria d no conjunto base para os átomos do segundo período^[164, 165] trouxe melhorias em relação ao conjunto sp usado nos métodos MNDO e AM1 originais, principalmente, para compostos hipervalentes^[148]. Não parece-nos trivial avaliar se esta melhoria foi fruto apenas da extensão e flexibilização do conjunto base ou se é consequência da necessidade intrínseca dos compostos hipervalentes por orbitais de momento angular d para descrição de sua estrutura eletrônica e ligações químicas (§3.2.1).

Modificações de f_{core} na equação 1.47 são freqüentemente propostas^[158, 166], mas, a menos que uma extensiva reparametrização seja efetuada^[166], os resultados não são significativamente superiores aos obtidos com os métodos originais.

A reparametrização dentro do esquema NDDO usual apresentado nesta seção e específica para um sistema ou reação (ou família de reações) de interesse é outra maneira de corrigir os erros listados acima^[157, 167, 168]. Na ausência de dados experimentais que possam ser usados no conjunto de treinamento da parametrização específica, propriedades obtidas por cálculos *ab initio* de alta acurácia podem ser usadas. Por exemplo, Rossi e Truhlar^[157] obtiveram uma parametrização específica para a reação de $\text{Cl} + \text{CH}_4$ que reproduz em ± 2 kcal/mol as energias relativas calculadas no nível MP2 para 23 espécies ao longo da PES da reação.

O procedimento de reparametrização específica foi adotado para corrigir as incertezas observadas nos métodos originais e calibrar a hamiltoniana semiempírica usada nas simulações enzimáticas (cap. 4). Os valores de referência usados no conjunto de treinamento foram as propriedades calculadas *ab initio* (cap. 2) para as espécies e as reações de tiólise e alcoólise de ésteres de fosfato (figuras 7 e 8) e um algoritmo de otimização global (§1.5.2) foi empregado para ajuste dos parâmetros semiempíricos (cap. 3).

1.5 Otimização de parâmetros

A localização de pontos estacionários em espaços de parâmetros de uma função é um problema matemático bastante estudado^{[24, §6.3][154, cap. 10]}. Diversos métodos de otimização são conhecidos e podem ser divididos em quatro tipos gerais:

1. *Sem derivadas*: Utilizam apenas o valor da função na otimização. Por exemplo, o método simplex^[154, §10.4] e os algoritmos genéticos (GA)^[155, 156];
2. *Primeira derivada*: Utilizam o valor da função e sua primeira derivada. Por exemplo, o método de gradiente conjugado^[154, §10.6];
3. *Primeira e segunda derivadas exatas*: Também chamados de métodos de segunda-ordem ou métodos de Newton^[169]. Utilizam a função, e sua primeira e segunda derivada, calculadas exatamente.
4. *Primeira derivada exata e segunda aproximada*: Utilizam a função e sua primeira derivada exata, mas a segunda derivada é aproximada de diferentes maneiras. Por exemplo, os métodos quasi-Newton^[170].

A escolha do melhor método depende da natureza da forma funcional avaliada, do ponto estacionário procurado e se as derivadas desta função em relação aos parâmetros são conhecidas e facilmente calculadas.

Para otimização de geometrias moleculares, as derivadas da energia em função das coordenadas atômicas são geralmente conhecidas e os métodos de Newton ou de quasi-Newton são mais usados (§1.5.1). Para sistemas macromoleculares, o cálculo das segundas derivadas, mesmo aproximadas, pode ser computacionalmente custoso e os métodos de primeira-ordem são preferidos.

Para otimizações em superfícies rugosas e de alta dimensionalidade, como por exemplo, a superfície de parâmetros semiempíricos (equação 1.50), o mínimo global é procurado e milhares ou milhões de passos de otimização são necessários. Portanto, os métodos que não necessitam do cálculo de derivadas (geralmente custoso) são indicados (§1.5.2).

1.5.1 Otimização calculando derivadas

Em princípio, qualquer função cujas derivadas sejam conhecidas pode ser expandida em torno de seus parâmetros e, portanto, a discussão a seguir é válida para problemas de otimização em geral.

Por exemplo, a energia E de um sistema molecular sob a aproximação de Born-Oppenheimer é uma função paramétrica das coordenadas nucleares (§1.4.1), $\{\mathbf{R}_A\} = \mathbf{R}^\dagger = (R_1, R_2, \dots, R_{3N})$, assim, podemos escrever a série de Taylor^{[124, apêndice C.2][171, cap. 2]}:

$$E(\mathbf{R}_1) = E(\mathbf{R}) + \mathbf{q}^\dagger \mathbf{g}(\mathbf{R}) + \frac{1}{2} \mathbf{q}^\dagger \mathbf{H}(\mathbf{R}) \mathbf{q} + \dots \quad (1.51)$$

onde $\mathbf{q} = (\mathbf{R}_1 - \mathbf{R})$ e $\mathbf{R}_i$ designa diferentes matrizes em que as coordenadas $(R_1, R_2, \dots, R_{3N})$ são modificadas. O gradiente $\mathbf{g}(\mathbf{R})$ tem elementos:

$$g_i = \frac{\partial E(\mathbf{R})}{\partial R_i} \quad (1.52)$$

e a hessiana $\mathbf{H}(\mathbf{R})$ tem elementos:

$$H_{ij} = \frac{\partial^2 E(\mathbf{R})}{\partial R_i \partial R_j} \quad (1.53)$$

Nosso interesse é encontrar pontos estacionários $\mathbf{R}_e$ da função E , que, por definição, têm $\mathbf{g}(\mathbf{R}_e) = \mathbf{0}$ e são caracterizados pela análise dos autovalores da hessiana. Para otimização de geometria de espécies reativas, os pontos estacionários que procuramos são mínimos, que possuem todos autovalores da hessiana positivos, e TSs, que possuem um único autovalor negativo e todos outros positivos. Embora a série de Taylor (equação 1.51) seja infinita, podemos truncá-la após o termo quadrático, se o deslocamento $\mathbf{q}$ for pequeno, constituindo a aproximação harmônica:

$$E(\mathbf{R}_1) = E(\mathbf{R}_e) + \frac{1}{2} \mathbf{q}^\dagger \mathbf{H}(\mathbf{R}_e) \mathbf{q} \quad (1.54)$$

Também podemos escrever:

$$\mathbf{g}(\mathbf{R}_1) = \mathbf{g}(\mathbf{R}) + \mathbf{H}(\mathbf{R}) \mathbf{q} \quad (1.55)$$

Novamente, para um ponto estacionário ($\mathbf{R}_1 = \mathbf{R}_e$) e para $\mathbf{H}$ não-singular, temos:

$$\mathbf{g}(\mathbf{R}) = -\mathbf{H}(\mathbf{R}) \mathbf{q} \quad (1.56)$$

$$\mathbf{R}_e = \mathbf{R} - \mathbf{H}^{-1}(\mathbf{R}) \mathbf{g}(\mathbf{R}) \quad (1.57)$$

No problema de otimização de geometria, as singularidades da hessiana, ou seja, os seis autovetores (para estruturas não lineares) que representam os graus de liberdade rotacionais e translacionais da estrutura, devem ser projetados^{[171, §2.5][124, p. 439]}. As equações 1.54 e 1.57 podem ser solucionadas numérica ou analiticamente para sistemas sob potenciais clássicos ou quânticos. A solução da equação 1.57 é o ponto de partida para

os métodos de otimização de segunda-ordem. Estes são os mais eficientes métodos de otimização não-linear local e pode-se mostrar que a convergência para um mínimo (local) ocorre quadraticamente^[169].

No entanto, a avaliação da hessiana em cada passo de otimização é bastante custosa, principalmente em cálculos de estrutura eletrônica. Assim, a hessiana inicial pode ser uma estimativa aproximada, que é melhorada ao longo dos passos de otimização, usando a informação do gradiente calculado^[170]. Estes métodos são menos custosos, mas a convergência ao mínimo local não é garantida e é mais lenta (ocorre em mais passos de otimização) do que os métodos verdadeiramente de segunda-ordem.

1.5.2 Otimização sem derivadas

Algoritmos mais robustos são necessários para encontrar o mínimo global de hipersuperfícies de parâmetros com uma quantidade de mínimos locais muito grande. Métodos que não utilizam derivadas são computacionalmente mais rápidos e, portanto, são mais indicados para explorações extensas destas hipersuperfícies.

A otimização por GAs é uma técnica estocástica que não necessita de derivadas e é bastante eficiente na procura do mínimo (ou máximo) global de uma superfície multidimensional^[155, 156]. GAs simulam o processo de evolução natural, como proposto por Darwin, onde *“os mais adaptados sobrevivem”*. Cada conjunto de parâmetros gerado é chamado de indivíduo e sua coleção forma uma população. A “adaptação” de cada conjunto de parâmetros é avaliada pela função que busca-se otimizar, denominada função erro (por exemplo, equação 1.50), ou seja, o indivíduo é mais adaptado quanto menor o erro. Os indivíduos são compostos por cromossomos, que representam cada um dos parâmetros otimizados. Os cromossomos têm uma representação binária, que facilita as operações de cruzamento e mutação (vide infra), e cada dígito é chamado de gene. As populações evoluem para próxima geração (passo seguinte da otimização) segundo operadores de cruzamento, mutação e elitismo^[156]. A evolução ocorre até que a população convirja, ou seja, até que as diferenças dos erros calculados para todos indivíduos sejam menores que um valor estipulado. Pares de indivíduos são selecionados de acordo com suas adaptações e suas seqüências de genes são divididas e cruzadas, gerando dois novos indivíduos. Uma mutação ocorre quando um gene é trocado de acordo com uma probabilidade pré-definida. As mutações preservam a diversidade dos cromossomos e impedem que uma população convirja prematuramente a um mínimo local. O operador de elitismo obriga que o indivíduo melhor adaptado da população seja inteiramente reproduzido na

próxima geração.

Entretanto, GAs não convergem rapidamente a um ponto estacionário local^[156]. Uma vez que a superfície de parâmetros tenha sido razoavelmente explorada, uma região de mínimo (possivelmente o mínimo global) somente é “refinada” pelo GA após muitas gerações. O algoritmo de otimização simplex, sugerido por Nelder e Mead^[154, §10.4], também não necessita de derivadas da função otimizada e é bem mais eficiente para encontrar um mínimo (“fundo do poço”) de uma região *local* da superfície de parâmetros. O algoritmo simplex equivale à otimização geométrica de uma figura com $n + 1$ vértices num espaço n dimensional, onde n é o número de parâmetros otimizados. Os vértices da figura são: refletidos, quando o ponto com maior erro é refletido no plano definido pelos demais pontos; expandidos, para que uma maior parte da superfície seja explorada; ou contraídos, para encolher o volume da figura e permitir a exploração de vales estreitos na superfície de parâmetros. Por outro lado, o simplex não é um algoritmo indicado para explorações globais de superfícies.

Tanto algoritmos genéticos^[157, 167, 168] como simplex^[150] já foram usados para parametrização de hamiltonianas semiempíricas. No capítulo 3, utilizamos uma combinação dos dois procedimentos de otimização, em que uma extensa exploração inicial da superfície de parâmetros foi executada pelo GA e os mínimos encontrados foram refinados pelo método simplex.

1.6 Simulação computacional de macromoléculas

Na §1.4, apresentamos duas maneiras diferentes para computação da estrutura eletrônica e, logo, da PES cuja informação local pode ser usada para calcular geometrias (§1.5.1) e propriedades termodinâmicas moleculares (apêndice B) em torno de pontos estacionários.

Porém, o cálculo da PES para milhares de configurações nucleares por métodos *ab initio* (§1.4.1) ou por métodos aproximados, como os semiempíricos (§1.4.2), torna-se bastante custoso computacionalmente para moléculas com alguns átomos pesados e impraticável para macromoléculas. Assim, apresentaremos na §1.6.1 os potenciais empíricos que são uma maneira ainda mais simplificada e menos custosa de computação da PES, usados pela técnica de dinâmica molecular (§1.6.2) para descrição de sistemas complexos. Na §1.6.3, discutiremos como obter diferenças de energia livre, a partir da simulação dos sistemas moleculares. Finalmente, apresentaremos na §1.6.4 uma mistura dos métodos

quanto-mecânicos (§1.4) com os potenciais empíricos que permite a simulação de mudanças da estrutura eletrônica, como reações químicas ou excitações eletrônicas, em sistemas complexos ou macromoleculares como enzimas.

1.6.1 Campos de força empíricos

Os potenciais efetivos empíricos ou *campos de força* tiveram sua origem na interpretação de espectroscopia vibracional, mas as funções empíricas apresentadas aqui foram desenvolvidas entre o final dos anos 1960 e início dos 1970^[24, §4.4]. Os campos de força clássicos são apropriados para o estudo de moléculas próximas de suas geometrias de equilíbrio e sua aplicação para o estudo de conformações moleculares é conhecida como *mecânica molecular*.

Diferentes formas funcionais são propostas para os campos de força já que estes são empíricos. No entanto, formas razoáveis podem ser deduzidas de observações experimentais ou da teoria de forças intermoleculares^[122, 172]. Trataremos aqui somente das formas funcionais simples e equivalentes àquelas utilizadas nas simulações do capítulo 4^[24, §4.4]. O potencial empírico total, $\mathcal{V}_{tot}$, é tipicamente dividido em interações ligantes (ou covalentes), $\mathcal{V}_{cov}$, e não ligantes, $\mathcal{V}_{non}$:

$$\mathcal{V}_{tot} = \mathcal{V}_{cov} + \mathcal{V}_{non} \quad (1.58)$$

$$\mathcal{V}_{cov} = \mathcal{V}_{lig} + \mathcal{V}_{ang} + \mathcal{V}_{died} + \mathcal{V}_{imp} \quad (1.59)$$

$$\mathcal{V}_{non} = \mathcal{V}_{elet} + \mathcal{V}_{LJ} \quad (1.60)$$

onde cada um dos termos são descritos por funções específicas.

A energia de ligação e a energia entre ângulos de ligação assumem uma forma harmônica:

$$\mathcal{V}_{lig} = \sum_{lig} \frac{1}{2} k_m (b - b_0)^2 \quad (1.61)$$

$$\mathcal{V}_{ang} = \sum_{ang} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \quad (1.62)$$

onde as somatórias são feitas sobre todas as ligações (*lig*) e ângulos entre ligações (*ang*) definidos no sistema, k_m é a constante de força da ligação, k_θ é a constante de força do ângulo entre ligações, b é a distância entre os dois átomos que definem a ligação, b_0 é a distância de *equilíbrio* da ligação, θ é o ângulo definido ente duas ligações e θ_0 é o valor de equilíbrio deste ângulo. Como o potencial tem uma forma harmônica, a energia

aumentará rapidamente a medida que a distância ou o ângulo de ligação se distorcerem dos valores de equilíbrio.

A energia de rotação em torno de uma ligação, ou seja, a energia de torção de um ângulo diedral ϱ , é descrita por uma função periódica contínua pois os ângulos de 0° a 360° são energeticamente acessíveis na maioria dos casos:

$$\mathcal{V}_{died} = \sum_{diedrais} \sum_n \frac{1}{2} K_n [1 + \cos(n\varrho - \delta)] \quad (1.63)$$

onde n é a periodicidade do ângulo, que varia normalmente de 1 a 4, δ é a fase, usualmente tomada como 0° , e K_n é a constante de força. O potencial $\mathcal{V}_{imp}$ refere-se a um ângulo diedral impróprio, necessário em sistemas planares, pois descreve movimentações fora do plano (como, por exemplo, a inversão da molécula de amônia). Sua forma funcional é no entanto, equivalente a equação 1.63 onde a primeira somatória corre sobre os diedrais impróprios definidos no sistema.

Os termos não ligantes descrevem a interação entre os átomos de diferentes moléculas ou entre partes distantes da mesma molécula e, portanto, ajudam a determinar a conformação total do sistema molecular. A densidade de carga da molécula pode ser descrita por diferentes formas ou distribuições, mas, usualmente, cargas parciais (q_i), puntiformes, constantes e fracionárias ($|q_i| < 1$, resultante da soma das cargas do núcleo e da população eletrônica em torno do átomo i) são colocadas nos respectivos centros atômicos. Assim, a energia eletrostática, $\mathcal{V}_{elet}$, considera a interação coulômbica clássica entre estas cargas parciais de cada átomo na molécula:

$$\mathcal{V}_{elet} = \sum_{i < j} \frac{q_i q_j}{r_{ij}} \quad (1.64)$$

onde a soma corre sobre todos os pares de átomos i e j , e r_{ij} é a distância entre estes centros atômicos.

O potencial de Lennard-Jones (LJ), $\mathcal{V}_{LJ}$, é usado normalmente para mimetizar o efeito de repulsão (“troca” quanto-mecânica), proveniente do recobrimento das nuvens eletrônicas, e o efeito atrativo de dispersão, produzido pela correlação das flutuações da densidade eletrônica de dois grupos:

$$\mathcal{V}_{LJ} = \sum_{i < j} \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (1.65)$$

onde a soma corre sobre todos os pares de átomos ij , ϵ_{ij} é a profundidade (mínimo de energia) do potencial de LJ e σ_{ij} é a posição do fundo do poço de potencial. Embora os

parâmetros ϵ_{ij} e σ_{ij} dependam formalmente de pares de átomos, eles são especificados por regras de combinação dos parâmetros de cada átomo. Normalmente, a combinação de ϵ_{ij} é geométrica, $\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$, e, dependendo da parametrização do potencial, a combinação de σ_{ij} é aritmética^[173], $\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj})$, ou geométrica^[174], $\sigma_{ij} = \sqrt{\sigma_{ii}\sigma_{jj}}$.

A parte repulsiva da equação 1.65 é produzida pelo termo $1/r_{ij}^{12}$ e a parte atrativa por $1/r_{ij}^6$. Este último termo, assim como a interação eletrostática, provém naturalmente da teoria de forças intermoleculares^[172], mas o termo em r_{ij}^{-12} é menos fundamentado e diferentes formas funcionais podem ser empregadas^[175]. O termo eletrostático (equação 1.64) é aquele de maior alcance já que o potencial decai como a distância recíproca. A interação de dispersão também tem alcance considerável, mas decai mais rapidamente. Todas as outras interações (ligantes e não ligantes) são de curto alcance. As energias eletrostáticas e de LJ são aditivas e determinadas par-a-par, ou seja, contém apenas contribuições de dois corpos. Contribuições não ligantes de muitos-corpos, como energia de polarização, podem ser incluídas e são importantes para descrição correta do potencial efetivo de determinados sistemas^[172].

Um detalhe adicional nos campos de força é a exclusão das interações não ligantes entre átomos ligados (chamadas de interações 1–2) ou separados por duas ligações (interações 1–3) e o escalonamento (geralmente por um fator de 0.5)^[174] das interações separadas por três ligações (interações 1–4)^[24, p. 58]. A exclusão e o escalonamento são em parte justificados, porque energias muito elevadas seriam obtidas se estas interações não fossem excluídas, já que a distância r_{ij} entre os centros ligados é pequena e, porque parte da interação já deve estar incluída nos termos ligantes.

Assim como num método semiempírico, a qualidade de um campo de força empírico depende da sua parametrização^[24, §12.8]. As formas funcionais empregadas nos diversos potenciais empíricos existentes^[173, 174] são muito similares, assim, a filosofia de parametrização adotada determina a diferença entre os campos. O problema matemático de otimização discutido na §1.5 aplica-se novamente à parametrização de campos de força, em que procura-se minimizar uma função P (equação 1.50), comparando os valores calculadas pelo campo de força e os valores observados (experimentalmente ou em cálculos *ab initio*). O procedimento de otimização para os campos de força também é tedioso e sem uma receita pronta, mas, neste caso, uma maior parte dos parâmetros pode ser obtida diretamente de medidas experimentais (como distâncias de ligação de equilíbrio, ou constantes de força, etc.). O número de parâmetros variados simultaneamente para o ajuste é relativamente menor do que na otimização de hamiltonianas semiempíricas. Ou-

tra diferença é que, num campo de força, diferentes tipos atômicos são indicados para o mesmo elemento (por exemplo, no campo de força OPLS^[174] temos carbono alifático CT, carbono aromático CA, carbono de carbonila C, etc.) a fim de reproduzir os diferentes ambientes químicos que cada um destes tipos atômicos está submetido. O resultado final da parametrização será um compromisso entre as dificuldades do processo de otimização e a forma funcional escolhida para o potencial, pois ambos limitam a precisão e o tipo de observáveis que podem ser simuladas.

1.6.2 Dinâmica molecular clássica

A evolução dinâmica de um sistema molecular é descrita por uma equação de Schrödinger dependente do tempo. Entretanto, esta descrição é computacionalmente custosa e desnecessária em muitos casos. Para simulação da dinâmica de macromoléculas, as equações da mecânica clássica são suficientes para tratar a maioria dos problemas sendo que correções para efeitos quânticos (como tunelamento ou energia do ponto-zero) podem ser adicionadas *a posteriori*. Assim, a hamiltoniana clássica, $\mathcal{H}$, pode ser descrita pela energia cinética e potencial do sistema^[24, §8.2]:

$$\mathcal{H}(\mathbf{p}_i, \mathbf{r}_i) = \sum_{i=1}^N \frac{1}{2m_i} \mathbf{p}_i^2 + \mathcal{V}(\mathbf{r}_i) \quad (1.66)$$

onde $\mathbf{p}_i$ é o momento da partícula i , m_i sua massa e $\mathcal{V}$ é o potencial efetivo, como descrito na §1.6.1 para o caso de simulações moleculares. Portanto, a hamiltoniana para N partículas é uma função de $3N$ posições de coordenadas e $3N$ momentos. Podemos derivar as equações de movimento:

$$\begin{aligned} \frac{\partial \mathbf{p}_i}{\partial t} &= - \frac{\partial \mathcal{H}}{\partial \mathbf{r}_i} \\ &= - \frac{\partial \mathcal{V}}{\partial \mathbf{r}_i} \\ &= \mathbf{F}_i \end{aligned} \quad (1.67)$$

$$\begin{aligned} \frac{\partial \mathbf{r}_i}{\partial t} &= \frac{\partial \mathcal{H}}{\partial \mathbf{p}_i} \\ &= \frac{\mathbf{p}_i}{m_i} \end{aligned} \quad (1.68)$$

que são equações diferenciais de primeira-ordem com relação ao tempo t ou as chamadas equações de Hamilton. Da equação 1.68 notamos que $\mathbf{p}_i = m_i \partial \mathbf{r}_i / \partial t$, substituindo esta relação na equação 1.67, temos a equação diferencial de segunda-ordem ou a equação de

movimento de Newton:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i \quad (1.69)$$

onde $\mathbf{F}_i$ é força sobre a partícula i do sistema, derivada da energia potencial $\mathcal{V}$.

A simulação da dinâmica molecular pode ser feita pela integração, a partir de valores iniciais ($t = 0$), das equações de primeira-ordem (1.67 e 1.68) ou de segunda-ordem (1.69). Esta integração é um outro exemplo de um problema matemático bastante estudado com uma grande variedade de algoritmos e métodos de solução numérica existentes. A integração das equações de primeira-ordem resulta em métodos mais acurados para evolução dinâmica, mas a integração da equação de Newton é computacionalmente mais eficiente. Um dos esquemas de integração da equação 1.69 é o de *velocidade de Verlet*^[176, §3.2.1], que pode ser derivado a partir de uma expansão de Taylor truncada das posições atômicas do sistema, $\mathbf{R}$, no tempo $t + \tau$, resultando em^[24, p. 137]:

$$\mathbf{R}(t + \tau) = \mathbf{R}(t) + \tau \mathbf{V}(t) + \frac{\tau^2}{2} \mathbf{M}^{-1} \mathbf{F}(t) \quad (1.70)$$

$$\mathbf{V}(t + \tau) = \mathbf{V}(t) + \frac{\tau}{2} \mathbf{M}^{-1} [\mathbf{F}(t) + \mathbf{F}(t + \tau)] \quad (1.71)$$

onde τ é o tamanho do passo temporal segundo o qual as equações de movimento são integradas, $\mathbf{V}(t)$ é a matriz de velocidades das partículas do sistema, $\mathbf{M}$ é uma matriz com a massa de cada partícula e $\mathbf{F}(t)$ é a matriz de forças sobre as partículas do sistema.

Termos além de $\mathbf{F}_i$ podem ser incluídos na equação 1.69^[177]:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i + m_i \gamma_i \frac{\partial \mathbf{r}_i}{\partial t} + \mathcal{R}_i(t) \quad (1.72)$$

O segundo termo do lado direito da equação indica uma força de fricção contrária ao movimento, proporcional à velocidade e determinada para cada partícula i pelo coeficiente de fricção γ_i . O termo $\mathcal{R}_i$ indica uma força randômica e rapidamente flutuante, cuja média é zero e que relaciona-se à temperatura e ao coeficiente de fricção pelo teorema de flutuação-dissipação^[178, §VIII.8]:

$$\langle \mathcal{R}_i(t) \cdot \mathcal{R}_i(t') \rangle = \frac{2k_B T \gamma_i}{m_i} \delta(t - t') \quad (1.73)$$

onde δ é a função δ de Dirac, e $\langle \cdots \rangle$ significa uma média no espaço de configurações (§1.6.3). A força $\mathcal{R}_i$ é usualmente definida como um *barulho branco gaussiano*^[178, §VIII.8]. Os dois termos adicionais em 1.72 podem representar o efeito de moléculas não incluídas explicitamente no sistema simulado ou o acoplamento das partículas a um banho térmico.

A equação diferencial estocástica 1.72, conhecida como equação de movimento de Lan-

gevin, pode ser integrada da mesma forma que 1.69. O esquema resultante, chamado de *velocidade de Verlet-Langevin*, é constituído por equações semelhantes a 1.70 e 1.71, onde os termos do lado direito são multiplicados por constantes que dependem de γ_i ^[176, §9.3]. Este esquema de integração possibilita um melhor controle de temperatura e uma melhor amostragem do espaço de configurações do que a integração da equação de Newton (1.69). No entanto, as propriedades dinâmicas obtidas em simulações pelo esquema de velocidade de Verlet-Langevin não representam corretamente as propriedades dinâmicas do sistema.

A acurácia numérica da integração das equações de movimento pode ser verificada pela conservação de certas propriedades do sistema, como a energia e o momento total. O tamanho do passo temporal, τ , é limitado pelo menor período de vibração do sistema, normalmente, uma ligação C–H, para assegurar que as equações de movimento sejam integradas corretamente. Assim, o *método de simulação de dinâmica molecular* pode ser definido pelo seguinte algoritmo:

1. Definir a composição do sistema (tipos e números de átomos, massas e campo de força), tamanho do τ e duração total da simulação;
2. Atribuir as posições $\mathbf{R}(t = 0)$ e velocidades $\mathbf{V}(t = 0)$ iniciais;
3. Calcular as forças iniciais $\mathbf{F}(t = 0)$ e seguir nos passos:
 - a) Calcular as posições no passo $t + \tau$ usando a equação 1.70;
 - b) Calcular as forças no passo $t + \tau$;
 - c) Calcular as velocidades no passo $t + \tau$ usando a equação 1.71;
 - d) Realizar qualquer tipo de análise ou impressão do presente passo;
 - e) Se o tempo t é menor que a duração total da simulação, incrementar t de τ e retornar ao item (a), senão ir para o item (4)
4. Finalizar a simulação

Duas aproximações adicionais devem ser abordadas quando simulamos sistemas estendidos em fase condensada, como líquidos, cristais ou moléculas solvatadas. As contribuições não ligantes para o campo de força são avaliadas para todos os pares de partículas presentes na simulação (§1.6.1). Isto implica que o número destas interações aumenta como $\mathcal{O}(N^2)$, o que inviabiliza a simulação de milhares de átomos (como tipicamente observado em simulações enzimáticas). Uma das maneiras de lidar com o escalonamento quadrático é a soma de Ewald^[24, §9.8] que é exata dentro de uma determinada precisão

adotada. Outra maneira é diminuir o número de interações calculadas, modificando a expressão de potencial, principalmente a eletrostática (equação 1.64), que tem o decaimento mais lento. A modificação mais comum é a truncagem da interação eletrostática após um raio de corte, r_c , realizada por uma função de corte, $S(r > r_c)$, multiplicada ao potencial original (equações 1.64 e 1.65) para cada par de centros atômicos ij . Esta função pode ser abrupta, como numa função degrau^[24, p. 165], o que resulta em mudanças abruptas na energia e nas forças calculadas durante a dinâmica, levando a artifícios indesejados (como, por exemplo, aquecimento do sistema). Uma maneira mais apropriada é o esquema de truncagem com mudança da força (“force-switching”), em que a função de corte^{[24, p. 166][179]}:

$$S(r) = \begin{cases} 1 & r \leq r_{on} \\ \frac{(r_{off}^2 - r^2)^2 (r_{off}^2 + 2r^2 - 3r_{on}^2)}{(r_{off}^2 - r_{on}^2)^3} & r_{on} < r \leq r_{off} \\ 0 & r > r_{off} \end{cases} \quad (1.74)$$

tem dois raios de truncagem definidos, $r_{on} < r_{off}$, e multiplica a força, $S(r)F(r)$, garantindo que a energia e a força calculadas serão funções contínuas e diferenciáveis. Portanto, a forma funcional do potencial (eletrostático e LJ) é modificada^[24, p. 170]:

$$\mathcal{V}' = \begin{cases} \mathcal{V}(r) + \delta\mathcal{V} & r \leq r_{on} \\ -\int_{r_{off}}^r dw S(w)F(w) & r_{on} < r \leq r_{off} \\ 0 & r > r_{off} \end{cases} \quad (1.75)$$

A energia absoluta do potencial em $r < r_{on}$ é alterada por $\delta\mathcal{V}$ (que depende apenas dos valores de r_{on} e r_{off} , fixos no início da simulação). Naturalmente, as energias relativas (que são as quantidades de interesse numa simulação com campos de força) são preservadas.

Outra aproximação necessária para simulação de sistemas estendidos em fase condensada é a suposição de condições de contorno periódicas (PBC), em que uma pequena parte (caixa central) do sistema total é escolhida para simulação e replicada infinitamente. A periodicidade significa que uma partícula que sai da caixa central num determinado passo de simulação, entra imediatamente no lado oposto da caixa. A caixa replicada é tipicamente cúbica ou ortorrômbica e suas dimensões devem ser escolhidas (ao lado dos raios de corte do método de truncagem) de modo que cada partícula interaja diretamente apenas com a cópia mais próxima das outras partícula do sistema periódico, empregando a chamada *convenção de imagem mínima*. De qualquer forma, a suposição de PBC impõe uma ordem artificial ao sistema completo que pode influir nas propriedades dinâmicas e estruturais obtidas da simulação.

1.6.3 Cálculo de energia livre

A energia livre é um conceito importante para caracterizar equilíbrio (equação B.12) e cinética (equação 1.7) de processos moleculares termalizados. O potencial de força média (PMF), $\mathcal{W}(\xi)$, é a função trabalho ou a energia livre de um sistema em função de uma coordenada ξ ^[180]. Processos dinâmicos ativados (como reações químicas) ou de equilíbrio conformacional podem ser convenientemente expressados em termos de $\mathcal{W}(\xi)$ ^[181]. O PMF é definido a partir da função de distribuição $\langle \rho(\xi) \rangle$ ^{[182][24, §10.6]}:

$$\mathcal{W}(\xi) = w - k_B T \ln[\langle \rho(\xi) \rangle] \quad (1.76)$$

onde w é uma constante arbitrária e $\langle \cdots \rangle$ significa uma média com peso de Boltzmann:

$$\langle \rho(\xi) \rangle = \frac{\int d\mathbf{R} \delta[\xi'(\mathbf{R}) - \xi] \exp[-\mathcal{V}(\mathbf{R})/k_B T]}{\int d\mathbf{R} \exp[-\mathcal{V}(\mathbf{R})/k_B T]} \quad (1.77)$$

onde $\mathcal{V}(\mathbf{R})$ é a energia potencial do sistema em função das coordenadas $\mathbf{R}$ e $\xi'(\mathbf{R})$ é uma função dependente de alguns graus de liberdade do sistema, por exemplo, uma distância, um ângulo ou uma combinação mais complicada destes, caracterizando uma coordenada de reação. A remoção dos graus de liberdade de $\xi'(\mathbf{R})$ da média $\langle \cdots \rangle$ é assegurada pela função δ de Dirac que “filtra” as coordenadas $\mathbf{R}$, resultando no valor de referência do PMF, ξ .

Em princípio, a computação de um PMF ou uma função de distribuição $\langle \rho(\xi) \rangle$ poderia ser feita por uma simulação no ensemble canônico (número de partículas, volume e temperatura constantes), contando-se a ocorrência dos diferentes valores da variável ξ em uma trajetória de dinâmica molecular. Entretanto, na prática, a presença de barreiras (com altura considerável em relação a $k_B T$) ao longo dos valores de ξ prejudica a amostragem do espaço de configurações. Assim, técnicas especiais de amostragem devem ser empregadas para o cálculo do PMF. A técnica de amostragem “guarda-chuva” (umbrella)^[183] introduz um potencial artificial na simulação, $\mathcal{V}_{umb}(\xi)$, que melhora a amostragem do espaço de configurações na vizinhança do valor escolhido de ξ . Como o potencial guarda-chuva não tem uma forma definida (que obviamente varia com a escolha de tipo de coordenada ξ), tipicamente um potencial harmônico:

$$\mathcal{V}_{umb,i}(\xi) = \frac{1}{2} k_{umb} (\xi - \xi_i)^2 \quad (1.78)$$

é usado e centrado em diferentes valores de referência ξ_i , que confina a amostragem a uma região pequena em torno de ξ_i , chamada de janela. Então, uma série de simulações são efetuadas em sucessivas janelas (com diferentes ξ_i) ao longo dos valores da coordenada ξ

em que o PMF deva ser avaliado. Ao final das simulações, os resultados das diferentes janelas devem ser agrupados e a influência do potencial guarda-chuva retirada para que o PMF, $\mathcal{W}(\xi)$, seja obtido. Em cada i -ésima janela, temos^[182]:

$$\mathcal{W}_i(\xi) = w' - k_B T \ln[\langle \rho(\xi) \rangle_i] - \mathcal{V}_{umb,i}(\xi) + \mathcal{K}_i \quad (1.79)$$

onde $\langle \rho(\xi) \rangle_i$ é a função de distribuição alterada pelo i -ésimo potencial guarda-chuva, $\mathcal{V}_{umb,i}$. Diferentes procedimentos são usados para determinação da constante (desconhecida) $\mathcal{K}_i$ e a conseqüente recombinação das janelas. Aparentemente, um dos mais estáveis é o método de análise dos histogramas ponderados (WHAM)^[182]. A única exigência para que o procedimento WHAM funcione corretamente é que as janelas vizinhas se recobram no espaço de configurações (o que pode ser verificado pelo recobrimento das janelas ao longo da coordenada ξ)^[182].

Existem outras maneiras para estimar variações de energia livre a partir de simulações de dinâmica molecular, como o método de perturbação de energia livre, em que a hamiltoniana do sistema é interpolada entre dois estados $\mathcal{H}_0$ e $\mathcal{H}_1$ ^{[180, 181][24, §11.5]}, ou o método de integração termodinâmica, em que uma derivada da energia livre em relação a um parâmetro η é integrada entre $\eta = 0$ e $\eta = 1$, que correspondem a dois estados termodinâmicos diferentes^[180]. Estes métodos são mais empregados para o estudo de processos que não dependem de uma coordenada de reação, como constantes de equilíbrio ou a energia livre de ligação de um substrato a uma enzima.

Perfis de energia livre também podem ser estimados por outros métodos que amostram o espaço de configurações, como, por exemplo, o método Monte Carlo^[24, cap. 11]. Este é um método estocástico, baseado no conceito de amostragem por importância e que apresenta vantagens sobre dinâmica molecular em algumas circunstâncias, mas não pode ser usado para o estudo de propriedades dependentes do tempo.

1.6.4 Potenciais híbridos

A apresentação desta seção é focada na utilização de potenciais híbridos para simulações macromoleculares, em especial, reações enzimáticas. Contudo, potenciais híbridos e o formalismo abaixo também podem ser aplicados em estudos de propriedades eletrônicas de substratos em microambientes heterogêneos ou em soluções, representados explicitamente^[184, 185].

O cálculo da energia livre ao longo do caminho de uma reação química catalisada por

uma enzima é inviável se a PES do sistema *total*, composto por milhares de átomos, for descrita por primeiros princípios (§1.4.1) ou mesmo por métodos semiempíricos (§1.4.2). Por outro lado, métodos mais rápidos e eficientes como os campos de força clássicos (§1.6.1) não foram construídos para descrever a quebra ou a formação de ligações químicas. Uma alternativa óbvia para estudar reações em sistemas biológicos é misturar ambas metodologias num só potencial, formando os chamados métodos híbridos de mecânica quântica e mecânica molecular (QM/MM)^[184, 185] Nos últimos 15 anos, estes métodos mostraram-se uma ferramenta bastante poderosa para a investigação de transformações de estrutura eletrônica em fase condensada, principalmente em sistemas macromoleculares^[3, 11, 19, 186].

Os métodos QM/MM e os potenciais híbridos são baseados em um particionamento estrutural e energético do sistema. Uma pequena região, que sofre as principais alterações eletrônicas durante a reação química (por exemplo, o substrato e os aminoácidos reativos do sítio ativo), é representada por núcleos e elétrons, descritos quanto-mecanicamente na aproximação de Born-Oppenheimer (p. 45). O restante do sistema (a porção não-reativa da proteína e o solvente) é representado por átomos, descritos pelo campo de força clássico (figura 12). Desta maneira, definimos uma fronteira artificial entre as duas regiões que deve ser tratada cuidadosamente (vide infra). O particionamento em duas regiões é intuitivo porque a grande maioria das reações químicas em macromoléculas são localizadas em um pequeno fragmento do sistema total.

Uma hamiltoniana efetiva, $\hat{\mathcal{H}}_{eff}$, que contém as interações de ambas as regiões, é definida por uma soma sobre as partes em que o sistema foi dividido:

$$\hat{\mathcal{H}}_{eff} = \hat{\mathcal{H}}_{QM} + \mathcal{H}_{MM} + \hat{\mathcal{H}}_{QM/MM} \quad (1.80)$$

onde $\hat{\mathcal{H}}_{QM}$ é equivalente à hamiltoniana da equação 1.12 e pode ser descrita tanto pelos métodos *ab initio* da §1.4.1, como pelos semiempíricos da §1.4.2. $\mathcal{H}_{MM}$ é um número que corresponde ao potencial do campo de força clássico da equação 1.58, $\mathcal{H}_{MM} = \mathcal{V}_{tot}$. A hamiltoniana de interação entre as partes clássicas e quânticas, $\hat{\mathcal{H}}_{QM/MM}$, é o principal ingrediente do método^[179, 187]:

$$\hat{\mathcal{H}}_{QM/MM} = - \sum_i^{N_e} \sum_m^M q_m \hat{\mathcal{E}}(\mathbf{r}_i, \mathbf{r}_m) + \sum_q^{N_q} \sum_m^M z_q q_m \hat{\mathcal{N}}(\mathbf{r}_q, \mathbf{r}_m) + \sum_q^{N_q} \sum_m^M \mathcal{H}_{QM/MM}^{vdW} \quad (1.81)$$

onde i e q referem-se respectivamente aos N_e elétrons e N_q núcleos da região QM, z_q é a carga do núcleo QM (ou caroço QM, se o método semiempírico for usado na região QM), m refere-se aos M átomos da região MM, q_m é a carga parcial do átomo MM definida pelo campo de força clássico (equação 1.64) e $\mathbf{r}$ é um vetor de coordenadas. Os

Figura 12: Representação da partição do sistema macromolecular em regiões quanto-mecânicas (QM) e molecular-mecânicas (MM).

operadores $\hat{\mathcal{E}}$ e $\hat{\mathcal{N}}$ determinam respectivamente como a interação eletrostática entre as cargas parciais MM e os elétrons, e os núcleos QM é calculada. Para uma descrição *ab initio* da função de onda QM, estes operadores são r_{im}^{-1} e r_{qm}^{-1} , respectivamente, onde r_{qm} é a distância entre os átomos q e m . Se os métodos semiempíricos são usados para região QM, os operadores eletrostáticos assumem formas alternativas (§1.6.4.1). As interações de van der Waals, $\mathcal{H}_{QM/MM}^{vdW}$, são calculadas por um potencial de Lennard-Jones (equação 1.65) onde ϵ e σ são obtidos pelas mesmas regras de combinação usadas nos campos de força (p. 67), a partir dos parâmetros atômicos atribuídos aos centros QM e MM. Para os átomos MM, os parâmetros de LJ são idênticos àqueles do campo de força. Porém, estes parâmetros não são definidos para os núcleos (ou caroços) QM. Uma alternativa é utilizar os parâmetros do campo de força diretamente nos centros QM. Outra possibilidade mais trabalhosa, que, a princípio, produz resultados de melhor qualidade, é otimizar os parâmetros de LJ da parte QM para que propriedades calculadas pelos potenciais híbridos sejam ajustadas a observáveis experimentais (como energia de solvatação do soluto^[188] ou energia de ligação de dímeros^[189]) ou a cálculos de estrutura eletrônica de boa acurácia, em complexos moleculares modelo^[167, 190]. Caso ligações covalentes sejam separadas entre a fronteira QM e MM (por exemplo, a ligação C—C na figura 12), o

tratamento da hamiltoniana de interação e dos termos ligantes ($\mathcal{V}_{cov}$, equação 1.58) é diferenciado (§1.6.4.2).

Portanto, a região quântica tem sua estrutura eletrônica polarizada pelo campo eletrostático das cargas do microambiente clássico. No entanto, outras interações, como a transferência de carga entre as regiões ou a polarização das cargas MM, não são incluídas na equação 1.81, que é a forma mais comum de hamiltoniana de interação QM/MM encontrada^[184, 185, 187]. Variações podem ser empregadas para descrição das forças de van der Waals, principalmente a repulsão de curto alcance^[175, 185], e para descrição das cargas MM em que a polarização destas em resposta a densidade eletrônica da parte QM pode ser implementada auto-consistentemente^[191, 192].

A expressão 1.80 é inserida na equação de Schrödinger (1.13) cuja solução resulta no valor esperado da energia total E :

$$E = E_{hib} = E_{QM} + E_{MM} + E_{QM/MM} \quad (1.82)$$

Para cada geometria dos núcleos QM e dos átomos MM do sistema, todas contribuições energéticas da equação 1.82 são recomputadas, gerando uma PES Born-Oppenheimer (p. 45). Outras quantidades, como a força sobre os átomos QM e MM, necessárias para otimizações de geometria (§1.5.1) ou para a evolução temporal em uma simulação de dinâmica molecular (§1.6.2), são calculadas analítica ou numericamente uma vez que E é conhecida.

1.6.4.1 Interação eletrostática semiempírica

A computação de PMF a partir de simulações de reações em macromoléculas com potenciais híbridos ainda é bastante demorada, se a região QM for descrita por métodos *ab initio* (§1.4.1). Todavia, o PMF pode ser obtido mais rapidamente, se a estrutura eletrônica da parte quântica for determinada por métodos semiempíricos (§1.4.2). Neste último caso, as interações eletrostáticas entre as partes QM e MM não são unicamente definidas. Diferentes propostas foram sugeridas^[185, 187, 188, 190], com os seguintes objetivos práticos:

- Os parâmetros previamente existentes para os termos QM e MM isolados não seriam afetados;
- Um número mínimo de novos parâmetros QM/MM seriam introduzidos e derivados dos parâmetros QM e MM existentes;

- O balanço entre as diferentes forças presentes nas contribuições QM e MM isoladas seria mantido na interação QM/MM^[172, 187].

Em geral, os operadores $\hat{\mathcal{E}}$ e $\hat{\mathcal{N}}$ assumem as formas:

$$\hat{\mathcal{E}}(\mathbf{r}_i, \mathbf{r}_m) = \sum_{\mu\nu} P_{\mu\nu}(\mu\nu | s^m s^m) \quad (1.83)$$

$$\hat{\mathcal{N}}(\mathbf{r}_q, \mathbf{r}_m) = (s^q s^q | s^m s^m) + f_{core} \quad (1.84)$$

onde cada elétron i é descrito pelos orbitais atômicos μ e ν presentes em cada átomo da região QM (submetidos à aproximação NDDO, p. 55) e $P_{\mu\nu}$ são os elementos da matriz de densidade (equação 1.26). A interação nuclear $\hat{\mathcal{N}}$ é similar à equação 1.47 e f_{core} assume as formas das equações 1.48 e 1.49 se o método semiempírico adotado for respectivamente MNDO ou AM1 (PM3). As integrais de repulsão eletrônica em 2-centros são calculadas segundo a mesma expansão multipolar truncada usada nos métodos semiempíricos (apêndice A)^[193], por exemplo:

$$(s^q s^q | s^m s^m) = [R_{qm}^2 + (\rho_0^q + \rho_0^m)^2]^{-1/2} \quad (1.85)$$

onde os s^m são orbitais idealizados nos centros MM (p. 56) e introduzidos para assegurar a consistência do potencial híbrido com a forma funcional da energia eletrostática QM semiempírica^[187, 188].

Destas expressões e dependendo da escolha do método semiempírico (MNDO ou AM1), temos que os parâmetros semiempíricos ρ_0^m , α^m (equação 1.48) e $\{K_j^m, L_j^m, M_j^m\}$, com j correndo de 1 a 4 (equação 1.49), são designados para os m centros MM. As diferenças entre as formas propostas para interação QM/MM eletrostática semiempírica estão principalmente relacionadas a quais destes parâmetros são usados nas equações 1.83 e 1.84, e como eles são obtidos^[190]. Field et al.^[187, 190] manteve os parâmetros nos centros QM inalterados e otimizou os parâmetros semiempíricos designados aos centros MM para reproduzir dados *ab initio* em complexos moleculares modelo. Nas expressões finais^[187], ρ_0^m e $\{K_j^m, L_j^m, M_j^m\}$ foram zerados e apenas um parâmetro, $\alpha^m = 5.0 \text{ bohr}^{-1}$, é usado para todos centros MM, independente de seu tipo atômico. Outros autores^[188, 190, 191] anularam também os termos ρ_0^q e $\{K_j^q, L_j^q, M_j^q\}$ nos centros QM, aproximando as expressões 1.83 e 1.84 de uma interação puramente coulômbica.

Um procedimento ligeiramente diferente foi adotado pelo grupo de Nancy^[190, 194], em que o termo para interação nuclear na equação 1.81 é substituído por:

$$q_m \hat{\mathcal{N}}(\mathbf{r}_q, \mathbf{r}_m) = z_m [(s^q s^q | s^m s^m) + f_{core}] - L_m (s^q s^q | s^m s^m) \quad (1.86)$$

onde z_m é carga do caroço do centro m (resultante da pseudocondensação dos “elétrons de caroço à carga nuclear do elemento m ”) e $L_m = z_m - q_m$ é uma “população eletrônica implícita” no centro m . Usamos o prefixo pseudo e as aspas, porque elétrons são definidos para os centros MM, embora sua carga de caroço possa ser determinada pelo elemento representado pelo tipo atômico. Neste procedimento, todos parâmetros originais do método semiempírico são usados para os centros QM e MM nas expressões 1.86 e 1.83^[190], sem simplificações.

As interações MM puras não ligantes ($\mathcal{V}_{non}$) são truncadas como apresentado na §1.6.2, e todas as interações QM são incluídas, porque espera-se que a região QM seja pequena. Para as interações QM/MM não ligantes, um esquema de truncagem similar ao das interações MM (p. 71) é aplicado às integrais de repulsão eletrônica (apêndice A).

1.6.4.2 Fronteira covalente entre as regiões

Em algumas aplicações, principalmente simulações enzimáticas, a fronteira entre as regiões clássica e quântica é definida sobre uma ou mais ligações covalentes dentro da mesma molécula (por exemplo, a ligação C—C na figura 12). Já que os elétrons da região MM não são tratados explicitamente, a densidade eletrônica QM ao longo da ligação localizada na fronteira precisa ser terminada de maneira satisfatória. Esta situação não ocorre se as regiões MM e QM estão em moléculas diferentes.

Novamente, não existe um modo exato ou único de tratar este problema. Com o objetivo de assegurar que as interações entre os átomos MM e QM na fronteira sejam similares ao resultado que seria obtido se todo sistema fosse quanto-mecânico, dois esquemas gerais foram introduzidos^[185, 195]:

1. *Orbitais híbridos ou localizados* são colocados no átomo QM de fronteira e dirigidos para região MM^[196–198]. A população eletrônica destes orbitais interage com os outros elétrons i , mas é congelada e, logo, não participa do procedimento auto-consistente (equações 1.24 ou 1.40).
2. *Átomos fictícios ou átomos-de-ligação* (“*link-atoms*”) são adicionados ao sistema para saturar a valência da parte QM em cada ligação covalente de fronteira^[187].

As incertezas introduzidas nas simulação são similares para ambos esquemas e podem ser minimizadas, se a região QM for ampliada, ou seja, se a distância entre a fronteira e os átomos QM reativos for aumentada^[179, 195]. Pequenas variações existem dentro destes

esquemas, principalmente, quanto ao tipo atômico ou o elemento usado como átomo fictício e quanto à definição do orbital híbrido. Portanto, diferentes implementações são usadas^[179, 185, 195].

Diferentes propostas também existem para quais termos ligantes do campo de força ($\mathcal{V}_{cov}$) são retidos na fronteira^[185, 195]. A maioria das implementações inclui todos termos ligantes se pelo menos um átomo MM participa da interação. Por exemplo, o potencial de torção de um diedral formado por um centro MM e três QM ou o termo de ligação entre um centro MM e outro QM são incluídos.

As interações não ligantes ($\mathcal{V}_{non}$) entre os átomos QM e MM de fronteira também variam entre cada implementação^[185]. No caso das interações de LJ, as regras de escalonamento de campos de força (p. 67) são normalmente respeitadas para as interações QM/MM. Já para as interações eletrostáticas, diferentes esquemas foram implementados^[185]. A carga total da região clássica deve permanecer inteira e consistente com o campo eletrostático original, assim, os esquemas mais comuns são aqueles em que todas interações eletrostáticas entre os centros QM e os centros MM vizinhos à fronteira são incluídas (sem nenhum tipo de escalonamento)^[179] ou aqueles em que a carga do primeiro átomo MM após a fronteira é anulada e seu valor distribuído pelas cargas MM vizinhas, de maneira que a carga total no resíduo MM seja preservada^[185]. As interações de LJ e eletrostáticas do átomo fictício com os átomos MM nunca são consideradas^[187].

2 *Reatividade intrínseca de ésteres de fosfato*

2.1 Métodos

Todas computações neste capítulo foram realizadas para moléculas isoladas na fase gasosa.

Os caminhos das reações indicadas nas figuras 7 e 8 foram otimizados, utilizando o conjunto base 6-31+G*^[134] (definições da nomenclatura de bases atômicas na p. 52) para a expansão da função de onda eletrônica, exceto para reação A, em que o conjunto 6-31G* foi inicialmente usado e os pontos estacionários foram reotimizados com o conjunto 6-311++G*^[135]. O nível MP2 restrito (p. 48), com a aproximação do congelamento dos elétrons de caroço (p. 51)^[121, §8.3.1], foi usado para calcular as estruturas eletrônicas, as energias e os gradientes necessários para determinação do caminho de reação.

As geometrias foram otimizadas pelo método do gradiente analítico, utilizando o algoritmo de otimização desenvolvido por Berny Schlegel et al.^[170]. Todas as coordenadas internas das geometrias de mínimos e de estados de transição foram otimizadas. As geometrias não estacionárias na PES foram parcialmente otimizadas, pois a distância da ligação que se formava (ou se quebrava) foi congelada. Assim, os caminhos de reação que levavam os reagentes aos produtos foram construídos pela variação desta distância congelada. Este caminho é definido como um caminho de reação distinto^[157, 199], em que uma coordenada é congelada e todas outras são otimizadas.

A energia dos pontos estacionários foi recalculada no nível MP2, utilizando o conjunto base 6-311+G(2df,2p)^[135]. A energia eletrônica e de repulsão nuclear resultante, E_{MP2} , é apresentada para cada espécie em §2.2.

Frequências vibracionais harmônicas^[171, cap. 2 e 4] foram calculadas analiticamente para todos os pontos estacionários. O nível de cálculo MP2/6-31+G* foi usado para determinar as frequências das espécies das reações C, E(A_n+D_n e D_n+A_n), F(D_n+A_n), G(A_n+D_n e D_n+A_n) e H(D_n+A_n); para a reação A, o nível usado foi MP2/6-31G* e para as reações

B, D, $F(A_n+D_n)$ e $H(A_n+D_n)$, o nível usado foi o B3LYP/6-31+G*. As reações são indicadas pela primeira letra (figuras 7 e 8) e os mecanismos (figura 3) são indicados entre parênteses. Nenhuma frequência imaginária foi encontrada para os mínimos e apenas uma foi encontrada para os TSs^[169].

As espécies nas reações B, D, $F(A_n+D_n)$ e $H(A_n+D_n)$ contém mais que 250 funções base, se suas funções de onda são expandidas pelo conjunto 6-31+G*. O sistema de programas utilizado neste estudo (vide infra) não calcula analiticamente, no nível MP2, frequências vibracionais para espécies com mais de 225 funções base. As frequências podem ser obtidas numericamente, no nível MP2/6-31+G*. Porém, esta computação é demorada e dura cerca de 10 dias de processamento (CPU) nos computadores utilizados neste estudo (p. 82) para cada ponto estacionário. Portanto, para as espécies das reações B, D, $F(A_n+D_n)$ e $H(A_n+D_n)$, as frequências analíticas foram calculadas no nível B3LYP/6-31+G*, que é menos demorado, após a reotimização das geometrias (para assegurar a anulação do gradiente, §1.5.1)^[124, apêndice C] no mesmo nível de cálculo.

Durante a análise das frequências vibracionais dos pontos estacionários, os 6 menores autovalores reais da Hessiana pesada por massa atômica foram conferidos, verificando-se que estes valores eram inferiores a 3 cm^{-1} ^[200]. Frequências vibracionais também foram usadas para obtenção da energia do ponto-zero (*ZPE*) e contribuições termodinâmicas vibracionais, na aproximação de oscilador harmônico^[201, seções 8.1], para energia (E_{vib}) e entropia (S_{vib}) a $T = 298\text{ K}$ (apêndice B). Nenhum escalonamento empírico foi aplicado às frequências^[202]. Frequências vibracionais calculadas para TMP no nível MP2/6-31G* são superestimadas de 1.2% a 4%, em comparação com valores experimentais^[203] (tabela 17). Contribuições termodinâmicas translacionais (E_{trans} e S_{trans}) e rotacionais (E_{rot} e S_{rot}), na aproximação do rotor rígido^[201, seções 8.2], também foram obtidas (apêndice B). Adição de *ZPE* e E_{MP2} resultou na energia total, E_T , que é equivalente a uma energia livre a 0 K.

Coordenadas intrínsecas de reação (IRC)^[204] foram calculadas^[205] a partir de todos TSs, utilizando o mesmo nível de cálculo das frequências vibracionais. As IRCs obtidas corresponderam ao caminho de reação esperado na direção dos reagentes e produtos.

Momentos de dipolo elétrico foram calculados pelo operador de dipolo quanto-mecânico^[124, p. 150 e 151], aplicado às funções de onda obtidas no nível MP2/6-311+G(2df, 2p). Potenciais de ionização eletrônica foram obtidos pela diferença de energia MP2/6-311+G(2df,2p) entre a espécie de camada fechada e a espécie ionizada (camada aberta) na geometria da molécula com camada fechada, ou seja, a ionização calculada é vertical

(sem relaxação da geometria nuclear), com contribuições de correlação eletrônica e de relaxação dos orbitais^[124, p. 389]. A energia da espécie de camada aberta foi obtida pelo método irrestrito^[121, p. 497], com aniquilação dos contaminantes de spin^[124, p. 107].

Todas computações mencionadas acima foram efetuadas com o sistema de programas Gaussian 98^[206].

O estudo da convergência da energia MP2 em relação ao conjunto base utilizou o conjunto base 6-311+G^[135], acrescido de funções de polarização. Os conjuntos usados foram 6-311+G(d), 6-311+G(d,p), 6-311+G(2d,2p), 6-311+G(2df,2p), 6-311+G(2df,2pd) e 6-311+G(3df,3pd) (p. 52). Também foram empregados os conjuntos TZVP e TZVPP, que são compostos pela união do conjunto TZV, desenvolvido por R. Ahlrichs et al.^[136], e das funções de polarização das bases cc-pVDZ e cc-pVTZ^[207], respectivamente. O conjunto TZVPP foi o maior entre aqueles usados neste estudo e, portanto, as energias obtidas com este conjunto representam os valores mais próximos daqueles que seriam obtidos utilizando um conjunto base completo. As computações de energia no estudo de base foram realizadas com o sistema Gaussian 98 e com o sistema Turbomole^[208].

Os computadores usados nesta tese foram PCs Pentium IV, usando o sistema operacional GNU/Linux Debian^[209, 210].

2.2 Resultados

Caminhos de reação e energias relativas para as reações de tiólise e alcoólise sobre fósforo de triésteres de alquil e de arilfosfatos (A, B, C e D, figura 7), e para monoésteres de alquil e de arilfosfatos (E, F, G e H, figura 8) foram obtidos. Com exceção da reação A, os mecanismos estudados correspondem somente a conformação em linha (axial), ou seja, o centro nucleofílico (átomos O ou S do nucleófilo), eletrofílico (P) e nucleófugo (O) estão alinhados. Esta conformação corresponde ao mecanismo proposto para reação no sítio ativo das PTPs. Para a reação A, a saída de CH_3O^- com o grupo tiolato equatorial também foi computada (figura 13).

Para os mecanismos associativos, a coordenada de reação distinta (DRC, p. 80) é definida como a distância de ligação envolvendo o fósforo e o oxigênio do grupo de saída $[\text{d}(\text{PO})]$ menos a distância de ligação do fósforo e o enxofre (ou oxigênio) do nucleófilo $[\text{d}(\text{PS}) \text{ ou } \text{d}(\text{PO})]$. Por exemplo, para reação A, a $\text{DRC} = \text{d}(\text{PO}) - \text{d}(\text{PS})$. Para os mecanismos dissociativos, a DRC não pode ser definida da mesma maneira, porque o nucleófilo encontra-se infinitamente separado durante a etapa de quebra de ligação (D_n), e o grupo

Figura 13: Estruturas otimizadas no nível MP2/6-311++G* para reação A. Distâncias atômicas calculadas estão mostradas em Å. Legenda para os tipos atômicos: H(branco), C(cinza), P(preto), O(linhas diagonais) e S(linhas cruzadas). A numeração dos centros atômicos esta descrita na estrutura A:TS1.

de saída encontra-se infinitamente separado durante a etapa de formação de ligação (A_n). Uma DRC $D_n + A_n$ pode ser definida se a distância que corresponde ao grupo infinitamente separado for substituída por uma constante arbitrária. Porém, esta DRC $D_n + A_n$ é descontínua e não seria comparável à DRC $A_n + D_n$. Já que os mecanismos $D_n + A_n$ são compostos por apenas três reações diferentes (§2.2.2), parece-nos desnecessário definir uma DRC $D_n + A_n$ para apresentação e análise dos resultados das reações dissociativas. Para estes mecanismos, as distâncias de ligações pertinentes são mencionadas e nenhuma DRC é definida. Assim, as reações A a H ocorrem no sentido da esquerda para direita nas figuras 7 e 8 e ao longo dos gráficos de caminho de reação e perfil de energia apresentados.

O zero de energia para cada reação corresponde à energia absoluta dos reagentes infinitamente separados. Contribuições termodinâmicas, como energias térmicas translacionais, rotacionais e vibracionais (E_{TRV}), assim como as equivalentes contribuições entrópicas (S_{TRV}), foram calculadas e estão apresentadas nas tabelas, mas *não* foram

adicionadas às energias totais, E_T (p. 81).

A barreira de ativação total para a reação é a maior diferença de energia entre um TS e qualquer um dos mínimos anteriores no caminho de reação^[7, cap. 6]. As alturas de barreira mostradas nos perfis de energia correspondem a cada etapa de reação separadamente e são relativas ao mínimo exatamente anterior.

Os pontos estacionários (complexos íon-molécula, TSs e intermediários) são indicados no texto por uma sigla contendo a letra correspondente à reação, seguida da abreviação da espécie. Todos os complexos íon-molécula (IMC), TSs, intermediários, e reagentes e produtos isolados, obtidos ao longo do caminho de reação, são denominados de espécie. Por exemplo, o intermediário da reação B é indicado pela sigla B:I ou o TS da primeira etapa da reação D é indicado por D:TS1. Para as reações E a H, em que dois mecanismos são possíveis, indicamos as espécies do mecanismo dissociativo por uma letra “d” e do mecanismo associativo por uma letra “a”. Por exemplo, o TS da segunda etapa da reação F ($A_n + D_n$) é indicado por F:TS2a e o TS da segunda etapa para o mecanismo $D_n + A_n$ é indicado por F:TS2d.

Caminhos de reação associativa são apresentados na seguinte sequência: reagentes infinitamente separados, IMC1 dos reagentes, TS1 correspondente ao ataque nucleofílico, intermediário pentacoordenado, TS2 correspondente a dissociação do grupo de saída, IMC2 dos produtos e produtos infinitamente separados. A reação dissociativa ocorre em duas etapas. Na primeira, o nucleófilo está infinitamente separado e na segunda o grupo de saída está infinitamente separado. A sequência de reação dissociativa é: reagentes, TS1 correspondente a dissociação do grupo de saída do éster de fosfato, IMC1 do metafosfato e grupo de saída, intermediário com todas espécies (metafosfato, grupo de saída e nucleófilo) infinitamente separadas, IMC2 do metafosfato e nucleófilo, TS2 correspondente ao ataque nucleofílico ao metafosfato e produtos. O intermediário com as espécies separadas aparece como um máximo no caminho de reação, porque as energias de complexação foram fornecidas ao sistema para descomplexar as espécies nos IMCs.

As geometrias otimizadas de todas espécies estacionárias, em coordenadas internas, podem ser obtidas com o autor.

2.2.1 Triésteres de fosfato

Esta seção apresenta os resultados para as reações A, B, C e D (figura 7). O caminho de mínima energia e as geometrias e perfis de energia dos pontos estacionários são mostrados

detalhadamente apenas para a reação A. Para as reações B, C e D apenas os perfis de energia e algumas geometrias selecionadas são apresentados. Os caminhos mínimos e as geometrias para estas reações são qualitativamente semelhantes àqueles da reação A.

Os triésteres de fosfato podem sofrer ataque nucleofílico no fósforo ou no carbono (alquilico ou arílico). Apenas resultados do ataque sobre fósforo são apresentados, já que estamos interessados em modelar a reatividade das PTPs. Um estudo *ab initio* da tiólise sobre carbono foi conduzido pelo autor para a reação de $\text{TMP} + \text{CH}_3\text{S}^-$ e a reatividade foi comparada ao ataque sobre fósforo ^[211]. A alcoólise sobre o carbono do TMP também foi estudada^[40, 212].

2.2.1.1 Reação A

A figura 14 mostra o caminho de energia mínima para o ataque nucleofílico de CH_3S^- sobre fósforo do TMP. Um IMC dos reagentes é inicialmente formado e a reação prossegue em duas etapas. O ataque nucleofílico ocorre via o estado de transição A:TS1, formando um intermediário pentacoordenado, A:I1, com geometria de bipirâmide trigonal. A eliminação de metóxido de A:I1 pode ocorrer por dois caminhos diferentes, que passam pelos A:TS2 e A:TS3, e resultam nos complexos A:IMC2 e A:IMC3, respectivamente. Os dois complexos se dissociam para formar os mesmos produtos, CH_3O^- e tiometil dimetil fosfato (TMDMP, figura 7).

A ligação P–O axial à ligação P–S (figura 13) é usada para compor a DRC da primeira etapa. O A:IMC1, formado na DRC = $-3.50 \text{ \AA}$, tem simetria C_s com os grupos metila apontando em direção ao CH_3S^- . Para formação do A:TS1, o ângulo diedral C3–O3–P–O1 (veja numeração dos átomos na figura 13) torciona-se de 180° (C_s) para 60° (C_1) por volta da DRC = $-2.5 \text{ \AA}$. A região entre as coordenadas $-1.6 \text{ \AA}$ e $-0.6 \text{ \AA}$ contém o A:TS1 e torções nos ângulos diedrais C1–S–P–O1 e C2–O2–P–O1. Os dois diedrais foram congelados e variados de -25° a 180° e de -65° a 60° , respectivamente, em conjunto à restrição da ligação P–S (p. 80). Várias otimizações de geometria resultaram em uma extensa avaliação da região da PES correspondente às distâncias de ligação P–S e aos ângulos diedrais fixados. O caminho de reação de menor energia consiste de uma torção de C1–S–P–O1 de -25° a 60° na DRC = $-1.50 \text{ \AA}$ e uma torção de C2–O2–P–O1 de 0° a 57° na DRC = $-1.26 \text{ \AA}$. Nenhum outro caminho com perfil energético similar foi encontrado. Por exemplo, restringindo-se o ângulo diedral C1–S–P–O1 a 60° e variando-se a DRC acima de $-1.7 \text{ \AA}$, a energia calculada aumenta cerca de 1 kcal/mol para cada $0.05 \text{ \AA}$ de aumento na coordenada de reação. O A:TS1

Figura 14: Caminho mínimo de energia MP2/6-31G* relativa em kcal/mol para a reação A. Os reagentes e produtos separados infinitamente estão indicados por um sinal de mais (+).

totalmente otimizado (figura 13) é formado na coordenada $-1.12 \text{ \AA}$ com ângulos diedrais C3–O3–P–O1, C1–S–P–O1 e C2–O2–P–O1 de 48° , 64° e 57° . O ângulo de ligação O3–P–S tem 166° , num arranjo axial típico, enquanto que O1, O2 e O4 estão localizados equatorialmente. A geometria do A:TS1 corresponde a uma TBP distorcida (figura 13).

Na coordenada $-0.62 \text{ \AA}$, uma pseudorrotação move os centros O3 e S axiais para posição equatorial. Esta mudança apresenta barreira energética desprezível. O intermediário pentacoordenado A:I1, com geometria de TBP, é formado na coordenada $-0.51 \text{ \AA}$. Os átomos O1, S e O3 estão em posição equatorial e os oxigênios O2 e O4 em posição axial (figura 13).

Três intermediários com diferentes arranjos de TBP poderiam ser formados e foram analisados. No nível MP2/6-31G*, a estrutura otimizada com os centros O1 e O2 axiais, e S, O3 e O4 equatoriais resultou num mínimo de energia estável. Porém, no nível MP2/6-311++G* este arranjo é instável e gerou um TS. Os outros possíveis intermediários são instáveis em ambos níveis de cálculo.

Dois caminhos com energias semelhantes são possíveis para a segunda etapa de reação. O grupo de saída CH_3O^- pode se posicionar axialmente a outro metóxido, como no ar-

Tabela 4: Energias e entropias relativas em kcal/mol para a reação A

Espécie	ΔE_{MP2}^a	ΔZPE^b	ΔE_{TRV}^c	$T\Delta S_{TRV}^d$	ΔE_T^e
TMP+(CH ₃ S) ⁻	0.0	0.0	0.0	0.0	0.0
A:IMC1	-16.3	0.9	1.8	-3.1	-15.4
A:TS1	-0.2	0.9	0.8	-17.3	0.7
A:I1	-4.9	1.2	1.1	-19.9	-3.5
A:TS2	17.1	-0.6	-0.5	-16.7	16.5
A:TS3	16.5	-0.6	-0.5	-16.7	15.9
A:IMC2	12.8	-0.7	-0.1	-8.7	12.1
A:IMC3	13.5	-0.7	-0.1	-10.6	12.8
TMDMP+(CH ₃ O) ⁻	37.2	-2.7	-2.4	0.1	34.5

^a energia eletrônica e de repulsão nuclear no nível MP2/6-311+G(2df,2p). ^b energia do ponto-zero no nível MP2/6-31G*. ^c $E_{trans} + E_{rot} + E_{vib}$, energias térmicas obtidas no nível MP2/6-31G* a 298 K. ^d $S_{trans} + S_{rot} + S_{vib}$, entropias obtidas no nível MP2/6-31G* a 298 K. ^e $E_T = E_{MP2} + ZPE$.

ranjo em A:I1. Este caminho passa pela TBP distorcida A:TS2 na DRC = 0.70 Å com os ângulos diedrais C3–O3–P–O1 e C1–S–P–O1 torcidos para -26° e 34° , respectivamente. Deslocamento do grupo de saída forma o A:IMC2 na coordenada 2.11 Å. Outro possível caminho envolve uma pseudorrotação, sem barreira energética, partindo da estrutura A:I1 e posicionando o grupo de saída axialmente ao tiolato. O A:TS3 é formado na coordenada 0.62 Å, com geometria de TBP distorcida (figura 13). Os ângulos diedrais C3–O3–P–O1 e C4–O4–P–O1 mudam para 0° e -25° , respectivamente. O A:IMC3 é formado na coordenada 2.01 Å numa conformação similar ao A:TS3. O caminho correspondente ao TS mais energético (A:TS3) tem um IMC de menor energia (A:IMC3).

As frequências imaginárias obtidas são de $120i\text{ cm}^{-1}$ para o A:TS1 e $114i\text{ cm}^{-1}$ para ambos A:TS2 e A:TS3. A pequena magnitude destas frequências sugere uma PES plana na região do estado de transição. A visualização destes modos imaginários indica que a movimentação predominante corresponde à vibração da ligação que se quebra (A:TS2 e A:TS3) ou que se forma (A:TS1).

A reação de tiólise é endoérgica (34.5 kcal/mol, tabela 4) e tem um mecanismo A_n+D_n , ou seja, um mecanismo associativo em duas etapas com formação de um intermediário estável. A altura de barreira para a primeira etapa é 16.0 kcal/mol (figura 15), correspondente à diferença de energia entre o A:TS1 e o A:IMC1. A reação inversa (A:I1 $\rightarrow$ A:TS1) tem uma barreira de 4.2 kcal/mol. A energia de complexação dos reagentes e dos produtos é de cerca de 15 kcal/mol (liberados) para a formação do A:IMC1 e de 22 kcal/mol (consumidos) para a dissociação do A:IMC2 e do A:IMC3. O A:I1 pentacoorde-

Figura 15: Perfil de energia total relativa em kcal/mol dos pontos estacionários para as reações A e B. Os números indicados correspondem às alturas de barreira dos TS (por exemplo, a altura de 10.7 kcal/mol indicada pela linha pontilhada) e energias de formação (ou de reação) dos intermediários (por exemplo, a energia de 3.9 kcal/mol indicada pela linha pontilhada).

nado tem energia de formação em relação ao A:IMC1 de 11.8 kcal/mol. Para a etapa de dissociação do grupo de saída, a altura de barreira é 19.6 kcal/mol para o A:TS2 e de 20.1 kcal/mol para o A:TS3 (figura 15). As barreiras para as reações inversas são 3.1 kcal/mol para A:IMC2 $\rightarrow$ A:I1 e 4.4 kcal/mol para A:IMC3 $\rightarrow$ A:I1. A energia de complexação dos produtos é cerca de 22 kcal/mol. A barreira total para reação A é 31.3 kcal/mol e corresponde à diferença de E_T entre o A:TS2 e o A:IMC1. O mecanismo com o grupo CH_3S^- posicionado equatorialmente ao grupo de saída tem uma barreira total de 31.9 kcal/mol.

2.2.1.2 Reações B, C e D

Os pontos estacionários nas reações B, C e D têm em geral arranjos conformacionais e coordenadas de reação semelhantes àqueles da reação A. As exceções e estruturas com conformação diferenciada serão comentadas abaixo.

Todos IMCs dos reagentes e produtos das reações B, C e D têm geometrias semelhantes a TBP distorcidas, com o grupo de ataque (ou de saída) posicionado axialmente, e com ambos grupos metila equatoriais torcidos na direção da ligação que se forma (ou quebra). As DRCs para os IMCs da primeira e segunda etapa variam respectivamente de -3.5 a -2.4 Å e de 2.0 a 2.9 Å.

Todos intermediários obtidos para reações B, C e D têm configuração de uma TBP. O B:I é formado na DRC = -0.43 Å e contém os grupos tiolato e fenóxido em posição equatorial e grupos metóxido em posição axial (figura 16). A E_T do pseudorrotâmero com metóxidos equatoriais e tiolato e fenóxido axiais é 3 kcal/mol maior que a energia de B:I. O D:I é formado na coordenada 0.12 Å e contém o grupo fenóxido em posição axial e os grupos metóxido equatoriais (figura 16). A energia do pseudorrotâmero também é 3 kcal/mol maior que a energia de D:I. Na reação C, todos grupos ligantes são metóxidos e, logo, não distinguimos confôrmeros. Como a reação C é simétrica, o intermediário é formado na coordenada 0.0 Å.

O B:TS1 é formado na DRC = -1.37 Å e tem os grupos metilas equatoriais torcidos da mesma maneira que nos A:TS1 e B:I (figura 16), ou seja, um grupo torcido na direção do tiolato e outro na direção oposta. Os TSs da primeira etapa das reações de alcoólise C e D são formados nas coordenadas -1.08 Å e -1.25 Å, respectivamente. Os grupos metilas equatoriais estão torcidos da mesma maneira que os IMCs, ou seja, apontando na direção do grupo metóxido que realiza o ataque nucleofílico.

Os TSs da segunda etapa agrupam-se em função de seus grupos de saída. O C:TS2 é formado na coordenada 1.08 Å (simétrico ao C:TS1) e, assim como A:TS2 e A:TS3, têm os grupos equatoriais torcidos como nos IMCs dos produtos da reação A (figura 13). Já os B:TS2 e D:TS2 têm DRC = 0.07 e 0.65 Å, respectivamente, e os grupos equatoriais torcidos como nos intermediários (figura 16). As distâncias da ligação que se forma ou que se rompe e as frequências imaginárias obtidas para os TSs das reações B, C e D estão apresentadas na tabela 8.

A reação de tiólise B é endoérgica (1.8 kcal/mol). A altura de barreira total é 10.7 kcal/mol e corresponde à primeira etapa (tabela 5 e figura 15). A reação de alcoólise C é uma reação de identidade (tabela 6) e a altura de barreira total é 14.0 kcal/mol e corresponde à segunda etapa (tabela 6 e figura 17). A reação de alcoólise D é exoérgica por 36.6 kcal/mol e a altura de barreira total, 3.0 kcal/mol (tabela 7 e figura 17), corresponde à segunda etapa.

A diferença de E_T entre as duas reações de tiólise (A e B, figura 15) é 5 kcal/mol no

Figura 16: Estruturas dos intermediários B:I e D:I otimizadas no nível MP2/6-31+G*. Distâncias de ligação calculadas estão mostradas em Å. A legenda para os tipos atômicos é idêntica àquela da figura 13.

Tabela 5: Energias e entropias relativas em kcal/mol para a reação B

Espécie	ΔE_{MP2}^a	ΔZPE^b	ΔE_{TRV}^c	$T\Delta S_{TRV}^d$	ΔE_T^e
DMPP+(CH ₃ S) ⁻	0.0	0.0	0.0	0.0	0.0
B:IMC1	-17.4	-0.2	1.0	-7.2	-17.6
B:TS1	-6.5	-0.4	0.1	-11.2	-6.9
B:I	-13.5	-0.2	0.1	-12.6	-13.7
B:TS2	-8.6	-0.6	-0.8	-13.7	-9.2
B:IMC2	-17.4	-0.8	-0.2	-8.1	-18.2
TMDMP+(PhO) ⁻	-2.0	0.2	0.2	1.1	-1.8

^a energia eletrônica e de repulsão nuclear no nível MP2/6-311+G(2df,2p). ^b energia do ponto-zero no nível B3LYP/6-31+G*. ^c $E_{trans} + E_{rot} + E_{vib}$, energias térmicas obtidas no nível B3LYP/6-31+G* a 298 K. ^d $S_{trans} + S_{rot} + S_{vib}$, entropias obtidas no nível B3LYP/6-31+G* a 298 K. ^e $E_T = E_{MP2} + ZPE$.

Tabela 6: Energias e entropias relativas em kcal/mol para a reação C

Espécie	ΔE_{MP2}^a	ΔZPE^b	ΔE_{TRV}^c	$T\Delta S_{TRV}^d$	ΔE_T^e
TMP+(CH ₃ O) ⁻	0.0	0.0	0.0	0.0	0.0
C:IMC1	-22.5	1.5	2.1	-10.3	-21.0
C:TS1	-16.9	1.4	1.3	-12.7	-15.5
C:I	-32.7	3.2	2.8	-13.9	-29.5
C:TS2	-16.9	1.4	1.3	-12.7	-15.5
C:IMC2	-22.5	1.5	2.1	-10.3	-21.0
TMP+(CH ₃ O) ⁻	0.0	0.0	0.0	0.0	0.0

^a energia eletrônica e de repulsão nuclear no nível MP2/6-311+G(2df,2p). ^b energia do ponto-zero no nível MP2/6-31+G*. ^c $E_{trans} + E_{rot} + E_{vib}$, energias térmicas obtidas no nível MP2/6-31+G* a 298 K. ^d $S_{trans} + S_{rot} + S_{vib}$, entropias obtidas no nível MP2/6-31+G* a 298 K. ^e $E_T = E_{MP2} + ZPE$.

Tabela 7: Energias e entropias relativas em kcal/mol para a reação D

Espécie	ΔE_{MP2}	ΔZPE	ΔE_{TRV}	$T\Delta S_{TRV}$	ΔE_T
DMPP+(CH ₃ O) ⁻	0.0	0.0	0.0	0.0	0.0
D:IMC1	-25.2	0.4	1.1	-9.9	-24.8
D:TS1	-23.5	0.6	0.6	-12.8	-22.9
D:I	-46.2	2.4	2.3	-13.5	-43.8
D:TS2	-42.6	1.8	1.7	-13.1	-40.8
D:IMC2	-53.8	1.8	2.6	-7.7	-52.0
TMP+(PhO) ⁻	-39.2	2.6	2.3	0.5	-36.6

As quantidades estão definidas na legenda da tabela 5.

Figura 17: Perfil de energia total relativa em kcal/mol dos pontos estacionários para as reações C e D. Os números indicados correspondem às alturas de barreira dos TS e energias de formação (reação) dos intermediários (dos produtos).

Figura 18: Energias HF/6-31+G* em kcal/mol para o dissociação de CH_3O^- do TMP. Distâncias de ligação P-O em Å.

TS de ataque de CH_3S^- ao éster de fosfato e aumenta progressivamente para 8 kcal/mol na formação do intermediário, 15 kcal/mol na ativação do TS do grupo de saída (metóxido na reação A e fenóxido na B) e chega a 30 kcal/mol na formação dos IMCs dos produtos. A mesma observação pode ser feita quanto à alcoólise dos ésteres de fosfato (C e D, figura 17), pois a diferença de E_T entre as reações C e D aumenta de 4 kcal/mol no TS do ataque de CH_3O^- sobre o éster, para 10 kcal/mol no intermediário, 11 kcal/mol na ativação da segunda etapa e chega a 31 kcal/mol na formação dos IMCs dos produtos.

Os mecanismos de todas reações de triésteres de fosfato são associativos (A_n+D_n), com formação de um intermediário pentacoordenado estável (figuras 13 e 16).

O mecanismo D_n+A_n foi testado para as reações A, B, C e D. As barreiras de ativação são, no mínimo, 50 kcal/mol mais altas do que as barreiras A_n+D_n . Por exemplo, na figura 18, observa-se a energia relativa computada no nível HF/6-31+G* para quebra da ligação P–O do TMP, que corresponde à primeira etapa do possível mecanismo D_n+A_n para a reação A ou C. Nenhum TS ou intermediário D_n+A_n foi obtido para os triésteres de fosfato. Na fase gasosa, este mecanismo de reação é considerado energeticamente inacessível (§1.2.2)^[40].

2.2.2 Monoésteres de fosfato

Esta seção apresenta os resultados para os mecanismos associativos e dissociativos (figura 3) obtidos para as reações de monoésteres de fosfato.

Na figura 8, os nucleófilos e grupos de saída (CH_3SH , CH_3OH e fenol) estão protonados e os ésteres de fosfato (MeSP^- , MeOP^- e PhOP^-) estão ionizados como monoânions. Estes são os estados de protonação resultantes para reagentes e produtos em fase gasosa, pois as basicidades dos nucleófilos e grupos de saída são maiores que as basicidades dos ésteres de fosfato.

As duas etapas nas reações dissociativas são independentes e o caminho da reação total é formado pela união destas etapas. Se uma etapa corresponde ao ataque nucleofílico sobre o metafosfato para formação do monoéster, a mesma reação, *no sentido inverso*, corresponde à quebra de ligação esterificada, resultando no grupo metafosfato e de saída dissociados. Por exemplo, a primeira etapa da reação E (D_n+A_n) é a dissociação do MeOP^- gerando metafosfato e CH_3OH . Esta etapa corresponde, no sentido inverso, à segunda etapa das reações D_n+A_n de alcoólise G e H. Assim, a união de apenas três reações distintas forma as reação dissociativas E, F, G e H. Estas três reações são: quebra

Figura 19: Estruturas do G:IMC1a, G:IMC1d e G:Ia otimizadas no nível MP2/6-31+G*. Distâncias calculadas estão mostradas em Å. A legenda para os tipos atômicos é idêntica àquela da figura 13.

da ligação P–S de tiometilfosfato (MeSP^-), e quebra da ligação P–O de metilfosfato (MeOP^-) e de fenilfosfato (PhOP^-).

2.2.2.1 Reações E, F, G e H

Os IMCs associativos têm o hidrogênio ligado ao nucleófilo (ou grupo de saída) coordenado a um dos oxigênios ionizados do grupo fosfato (figura 19). As DRCs para os IMCs da primeira e segunda etapa variam de -2.5 Å a -2.0 Å e de 1.2 Å a 2.2 Å, respectivamente.

Os IMCs dissociativos têm o hidrogênio ligado ao nucleófilo (ou grupo de saída) coordenado a um dos oxigênios do metafosfato (figura 19). No IMC da dissociação de MeSP^- , $d(\text{P}–\text{S}) = 4.23$ Å e $d(\text{O}–\text{H}) = 2.00$ Å. No IMC da dissociação de MeOP^- , $d(\text{P}–\text{OMe}) = 3.86$ Å e $d(\text{PO}–\text{H}) = 1.82$ Å, e no IMC da dissociação de PhOP^- , $d(\text{P}–\text{OPh}) = 3.99$ Å e $d(\text{PO}–\text{H}) = 1.72$ Å.

Os intermediários associativos são TBPs com nucleófilo e grupo de saída sempre em posição axial (em linha) e com os oxigênios do grupo fosfato (os protonados e o não ligante) em posição equatorial (figura 19). O E:Ia é formado na $\text{DRC} = -0.68$ Å, o F:Ia na $\text{DRC} = 0.45$ Å, o G:Ia na $\text{DRC} = 0.0$ Å (reação simétrica) e o H:Ia na $\text{DRC} = 0.14$ Å. O metafosfato tem uma geometria plana com simetria D_{3h} .

A distância de ligação que se forma ou se rompe nos TSs está mostrada na tabela 8.

Tabela 8: Caracterização dos TS e da participação da transferência de H^+ na coordenada de reação

Reação	TS1			TS2		
	d(P-X) ^a	Im(freq) ^b	% ^c	d(P-X)	Im(freq)	%
A	2.80	120i	0	2.76	114i	0
B	3.07	100i	0	2.25	140i	0
C	2.70	100i	0	2.70	100i	0
D	2.93	108i	0	2.34	116i	0
E(A _n +D _n)	2.63	112i	10	2.21	758i	70
E(D _n +A _n)	2.04	1380i	100	2.98	1072i	100
F(A _n +D _n)	2.79	140i	0	2.28	126i	10
F(D _n +A _n)	2.38	1047i	100	2.98	1072i	100
G(A _n +D _n)	2.27	575i	50	2.27	575i	50
G(D _n +A _n)	2.04	1380i	100	2.04	1380i	100
H(A _n +D _n)	2.19	650i	50	2.30	97i	0
H(D _n +A _n)	2.38	1047i	100	2.04	1380i	100

^a distância de ligação em Å entre o P e o átomo do nucleófilo ou grupo de saída, onde X=S ou O.

^b frequência imaginária em cm^{-1} . ^c porcentagem aproximada da contribuição das distâncias de ligação envolvendo a transferência de H^+ para o autovetor correspondente ao autovalor negativo da hessiana (*i.e.*, a frequência imaginária). As frequências imaginárias foram calculadas no nível MP2/6-31+G* ou B3LYP/6-31+G* (§2.1).

Os TS associativos são TBPs distorcidas com o hidrogênio de um dos grupos OH torcido na direção do nucleófilo (ou grupo de saída) e o outro hidrogênio torcido na direção oposta, como nos intermediários (figura 19). As DRCs para os TSs associativos são -0.91 Å para E:TS1a e 0.01 Å para o E:TS2a, -1.06 Å para F:TS1a e 0.11 Å para F:TS2a, ± 0.56 Å para G:TS1a(−) e o G:TS2a(+) (simétricos já que a reação G é uma identidade), e -0.44 Å para o H:TS1a e 0.61 Å para o H:TS2a.

Os TS dissociativos são complexos com ligações de hidrogênio distorcidas e uma menor distância entre P e S (ou O). No TS da dissociação de $MeSP^-$, as distâncias da ligação S–H e O–H são 1.71 Å e 1.18 Å, respectivamente. No TS da dissociação de $MeOP^-$, $d(MeO-H) = 1.20$ Å e $d(PO-H) = 1.29$ Å, e no TS da dissociação de $PhOP^-$, $d(PhO-H) = 1.29$ Å e $d(PO-H) = 1.17$ Å.

A $d(PS)$ no $MeSP^-$ é 2.14 Å e a $d(PO)$ é 1.72 Å no $PhOP^-$ e 1.68 Å no $MeOP^-$.

Em todos caminhos de reação de monoésteres, duas transferências intramoleculares de H^+ são observadas; uma do grupo fosfato para o grupo de saída e outra do nucleófilo para o grupo fosfato. Nas reações de triésteres, em que não ocorre transferência de H^+ , observamos TSs com frequências imaginárias por volta de $100i$ cm^{-1} (tabela 8). Nas reações de monoésteres, os TSs possuem frequências imaginárias que variam de $100i$

Figura 20: Variação das distâncias interatômicas em Å ao longo da IRC da reação entre CH_3OH e PO_3^- . A estrutura do TS corresponde à $\text{IRC} = 0$. O2 denota o oxigênio do grupo PO_3^- e O5, o oxigênio do grupo CH_3OH . A reação, no sentido da esquerda para direita, equivale à segunda etapa das reações G ($\text{D}_n + \text{A}_n$) e H ($\text{D}_n + \text{A}_n$), e, no sentido da direita para esquerda, equivale à primeira etapa das reações E ($\text{D}_n + \text{A}_n$) e G ($\text{D}_n + \text{A}_n$).

a 1400 i cm^{-1} . Freqüências por volta de 100 i cm^{-1} indicam IRCs^[204] contendo apenas a distância da ligação P–S ou P–O. Freqüências acima de 1000 i cm^{-1} indicam IRCs contendo apenas uma transferência de H^+ , sem contribuição da formação (ou quebra) da ligação P–S (ou P–O) e freqüências intermediárias, por volta de 500 i cm^{-1} , indicam IRCs contendo contribuições tanto da ligação P–O como de uma transferência de H^+ .

As IRCs na região dos TSs $\text{D}_n + \text{A}_n$ são dominadas por transferências de H^+ , que ocorrem antes da formação (ou quebra) da ligação P–S (ou P–O), como já observado para reações de hidrólise de monoésteres de fosfato^[73]. Na figura 20, observa-se que a distância da ligação P–O varia menos de 0.2 Å ao longo dos $2\text{ bohr.amu}^{1/2}$ da IRC na região do TS (unidades atômicas pesadas por massa das partículas são usadas para IRCs). No intervalo $-0.5 < \text{IRC} < 0.5\text{ bohr.amu}^{1/2}$, a variação de $d(\text{OH})$ para O2–H e O5–H (legenda na figura 20) é de 0.6 e 0.8 Å . A $d(\text{O5–H})$ é praticamente constante antes de $-0.5\text{ bohr.amu}^{1/2}$, assim como $d(\text{O2–H})$, após $0.5\text{ bohr.amu}^{1/2}$. Portanto, a transferência de H^+ na reação entre CH_3OH e PO_3^- ocorre num intervalo estreito de IRC. O comportamento para as dissociações de MeSP^- e PhOP^- é semelhante e não será mostrado.

Nos TSs associativos E:TS2a, G:TS1a, G:TS2a e H:TS1a, a transferência de H^+ con-

tribui significativamente para IRC. Estes TSs envolvem a quebra ou a formação da ligação P–O de um grupo metóxido. No caso dos TSs associativos de formação de ligação P–S ou de quebra de ligação P–O de um grupo fenóxido, a IRC é dominada pela formação (ou ruptura) destas ligações, sem contribuição de transferência de H^+ .

Na determinação do caminho da reação E, encontramos o E:TS1a e o E:Ia como pontos estacionários na PES Born-Oppenheimer (p. 45). A energia E_{MP2} do E:TS1a é 0.1 kcal/mol mais alta que a energia do E:Ia, mas a adição da *ZPE* resulta num TS com energia total inferior ao intermediário (tabela 9). Portanto, o E:Ia não é uma espécie estável e o mecanismo da reação é, na verdade, simultâneo (A_nD_n , figura 3), com um único TS (E:TS2a). A altura de barreira para reação E associativa é 37.7 kcal/mol e a barreira para reação E dissociativa é aquela da primeira etapa, 30.6 kcal/mol (figura 21). A reação E é endoérgica (3.1 kcal/mol).

A reação de tiólise F é endoérgica (7.1 kcal/mol). A barreira para reação F(A_n+D_n) é 28.7 kcal/mol e corresponde à diferença de energia total entre o F:TS2a e o F:IMC1a. Para reação F(D_n+A_n), a barreira é 33.2 kcal/mol e corresponde à diferença de energia total entre o F:TS2d e os reagentes (tabela 10 e figura 22).

A reação de alcoólise G é uma reação de identidade. A barreira para reação G(A_n+D_n) é 34.8 kcal/mol e para reação G(D_n+A_n) é 30.6 kcal/mol (tabela 11). Ambas barreiras correspondem ao TS da primeira etapa: ataque nucleofílico do metanol sobre o grupo fosfato, com a transferência de H^+ para o substrato, na reação associativa; e transferência de H^+ do $MeOP^-$ para o metóxido, na reação dissociativa.

A reação de alcoólise H é endoérgica (4.0 kcal/mol, tabela 12). A barreira para reação H(A_n+D_n) é 34.1 kcal/mol e corresponde ao ataque nucleofílico de CH_3OH ao $PhOP^-$, com transferência de H^+ e formação da ligação P–O. A barreira para reação H(D_n+A_n) é 34.6 kcal/mol e também corresponde ao ataque nucleofílico de CH_3OH , desta vez sobre o metafosfato.

Todas as energias de formação dos intermediários nos monoésteres são altas, variando de 24 a 28 kcal/mol para as reações A_n+D_n e de 27 a 31 kcal/mol para as reações D_n+A_n . Para os triésteres de fosfato, as energias de formação dos intermediários variam de –19 a 12 kcal/mol.

As energias de formação dos intermediários nas reações dissociativas são calculadas em relação aos reagentes isolados, ao invés dos IMCs anteriores, como nas reações associativas. Utilizamos duas definições diferentes para esta energia de formação a fim de eliminar a

Tabela 9: Energias e entropias relativas em kcal/mol para a reação E

Espécie	ΔE_{MP2}	ΔZPE	ΔE_{TRV}	$T\Delta S_{TRV}$	ΔE_T
(MeOP) ⁻ + CH ₃ SH	0.0	0.0	0.0	0.0	0.0
Mecanismo A_n+D_n					
E:IMC1a	-13.9	0.9	1.8	-8.7	-13.0
E:TS1a	11.0	2.4	2.3	-13.0	13.4
E:Ia	10.9	2.7	3.0	-12.3	13.6
E:TS2a	25.0	-0.3	-0.1	-12.4	24.7
E:IMC2a	-15.1	3.1	3.7	-10.7	-12.0
Mecanismo D_n+A_n					
E:TS1d	33.3	-2.7	-3.0	-0.8	30.6
E:IMC1d	14.1	-0.4	0.3	2.9	13.7
E:Id	29.0	-1.7	-1.8	11.3	27.3
E:IMC2d	17.4	-0.8	0.0	2.9	16.6
E:TS2d	32.0	-2.8	-2.5	0.5	29.2
(MeSP) ⁻ + CH ₃ OH	1.9	1.2	1.4	-0.6	3.1

As quantidades estão definidas na legenda da tabela 6.

Figura 21: Perfil de energia total relativa em kcal/mol dos pontos estacionários para a reação E. Os números indicados correspondem às alturas de barreira nos TS e energias de formação (reação) dos intermediários (dos produtos).

Tabela 10: Energias e entropias relativas em kcal/mol para a reação F

Espécie	ΔE_{MP2}^a	ΔZPE	ΔE_{TRV}	$T\Delta S_{TRV}$	ΔE_T^b
(PhOP) ⁻ + CH ₃ SH	0.0	0.0	0.0	0.0	0.0
Mecanismo A_n+D_n^c					
F:IMC1a	-13.1	0.6	1.0	-10.5	-12.5
F:TS1a	12.7	1.1	1.3	-12.0	13.8
F:Ia	10.0	1.9	2.4	-11.7	11.9
F:IMC2a	15.2	1.0	1.3	-11.6	16.2
F:TS2a	-17.4	1.8	2.7	-8.4	-15.6
Mecanismo D_n+A_n^d					
F:TS1a	32.3	-2.9	-2.9	0.8	29.4
F:IMC1a	10.6	-0.1	0.3	2.1	10.5
F:Ia	33.3	-2.1	-2.2	12.7	31.2
F:IMC2a	21.7	-1.2	-0.3	4.3	20.5
F:TS2a	36.4	-3.2	-2.9	1.9	33.2
(MeSP) ⁻ + PhOH	6.3	0.8	1.0	0.7	7.1

^a MP2/6-311+G(2df,2p). ^b $E_T = E_{MP2} + ZPE$. ^c Para o mecanismo A_n+D_n, ZPE , E_{TRV} e S_{TRV} calculados no nível B3LYP/6-31+G*. ^e Para o mecanismo D_n+A_n, ZPE , E_{TRV} e S_{TRV} calculados no nível MP2/6-31+G*.

Figura 22: Perfil de energia total relativa em kcal/mol dos pontos estacionários para a reação F. Os números indicados correspondem às alturas de barreira nos TS e energias de formação (reação) dos intermediários (dos produtos).

Tabela 11: Energias e entropias relativas kcal/mol para a reação G

Espécie	ΔE_{MP2}	ΔZPE	ΔE_{TRV}	$T\Delta S_{TRV}$	ΔE_T
(MeOP) ⁻ + CH ₃ OH	0.0	0.0	0.0	0.0	0.0
Mecanismo A_n+D_n					
G:IMC1a	-17.3	1.4	2.2	-8.8	-15.9
G:TS1a	19.9	-1.0	-1.1	-12.6	18.9
G:Ia	9.6	2.2	1.9	-13.5	11.8
G:TS2a	19.9	-1.0	-1.1	-12.6	18.9
G:IMC2a	-17.3	1.4	2.2	-8.8	-15.9
Mecanismo D_n+A_n					
G:TS1d	33.3	-2.7	-3.0	-0.8	30.6
G:IMC1d	14.1	-0.4	0.3	2.9	13.7
G:Id	29.0	-1.7	-1.8	11.3	27.3
G:IMC2d	14.1	-0.4	0.3	2.9	13.7
G:TS2d	33.3	-2.7	-3.0	-0.8	30.6
(MeOP) ⁻ + CH ₃ OH	0.0	0.0	0.0	0.0	0.0

As quantidades estão definidas na legenda da tabela 6.

Figura 23: Perfil de energia total relativa em kcal/mol dos pontos estacionários para a reação G. Os números indicados correspondem às alturas de barreira nos TS e energias de formação (reação) dos intermediários (dos produtos).

Tabela 12: Energias e entropias relativas em kcal/mol para a reação H

Espécie	ΔE_{MP2}	ΔZPE	ΔE_{TRV}	$T\Delta S_{TRV}$	ΔE_T
(PhOP) ⁻ + CH ₃ OH	0.0	0.0	0.0	0.0	0.0
Mecanismo A_n+D_n					
H:IMC1a	-16.9	1.3	1.7	-10.3	-15.6
H:TS1a	20.4	-1.9	-1.9	-12.0	18.5
H:Ia	5.5	1.2	1.2	-12.3	6.7
H:TS2a	7.7	0.6	0.4	-12.6	8.3
H:IMC2a	-22.5	0.8	1.4	-9.2	-21.7
Mecanismo D_n+A_n					
H:TS1d	32.3	-2.9	-2.9	0.8	29.4
H:IMC1d	10.6	-0.1	0.3	2.1	10.5
H:Id	33.3	-2.1	-2.2	12.7	31.2
H:IMC2d	18.5	-0.8	0.0	4.2	17.7
H:TS2d	37.7	-3.1	-3.3	0.6	34.6
(MeOP) ⁻ + PhOH	4.4	-0.4	-0.3	1.3	4.0

As quantidades estão definidas na legenda da tabela 10.

Figura 24: Perfil de energia total relativa em kcal/mol dos pontos estacionários para a reação H. Os números indicados correspondem às alturas de barreira nos TS e energias de formação (reação) dos intermediários (dos produtos).

influência das energias de complexação das espécies e minimizar o BSSE (p. 54)^[140, 141].

Para as reações de tri e monoésteres, as barreiras para etapa de ataque nucleofílico dependem do mecanismo estudado mas não dependem da natureza dos grupos de saída (alquila ou arila). As barreiras de saída do grupo alquila ou do arila também não dependem do grupo nucleofílico.

As reações $D_n + A_n$ dos monoésteres, cujas IRCs são dominadas pela transferência de H^+ (tabela 8), apresentam barreiras que, aparentemente, são pouco influenciadas pela natureza do grupo reativo. Para o ataque de CH_3SH e de CH_3OH ao metafosfato, ou seja, na etapa em que o H^+ é transferido para o metafosfato, as barreiras são 13 e 17 kcal/mol, respectivamente. Já na dissociação de CH_3OH e $PhOH$, ou seja, na etapa em que o H^+ é transferido para o grupo de saída, as barreiras são 31 e 29 kcal/mol, respectivamente.

Os ataques nucleofílicos associativos de CH_3OH sobre monoésteres ($G:TS1a$ e $H:TS1a$), cujas IRCs contêm significativa contribuição da transferência de H^+ , têm barreiras de ativação de 34 e 35 kcal/mol. Estas barreiras são 7 a 18 vezes maiores do que as barreiras das reação de alcoólise de triésteres, em que não ocorre transferência de H^+ . As barreiras de saída de CH_3OH são 7 e 11 kcal/mol para reações $A_n + D_n$ dos monoésteres. Nestas reações, a transferência de H^+ do grupo fosfato para o metóxido contribui para diminuição da energia de ativação em cerca de 7 a 9 kcal/mol, em relação às reações nos triésteres.

As barreiras de ativação para saída do grupo fenóxido ou fenol variam de 2 a 5 kcal/mol nas reações $A_n + D_n$ dos triésteres e os monoésteres. Em ambas reações, a IRC é composta principalmente pela distância da ligação P–O (tabela 8).

2.2.3 Conjunto base para expansão dos orbitais moleculares

As reações C, E ($A_n + D_n$) e F ($D_n + A_n$) foram escolhidas como representativas daquelas estudadas (figuras 7 e 8) e a energia E_{MP2} foi calculada em função de diferentes conjuntos base. O objetivo foi escolher o menor conjunto que resultasse em energias *relativas* aparentemente convergidas, isto é, energias relativas calculadas em torno de ± 1 kcal/mol das energias relativas obtidas com o maior conjunto base usado, TZVPP (§1.4.1.1 e §2.1)^[136].

A figura 25 apresenta as energias absolutas HF e de correlação eletrônica em perturbação de segunda ordem ($E^{(2)}$)^[127] para as 3 estruturas da segunda etapa da reação F ($D_n + A_n$). As energias HF e $E^{(2)}$ são apresentadas separadamente, porque a convergência

de cada uma é diferente. Observa-se que a convergência da energia HF é mais rápida do que da energia de correlação eletrônica. O comportamento das energias absolutas nas outras reações é equivalente e não será mostrado.

O tamanho dos conjuntos base chega a quadruplicar durante este estudo (por exemplo, para a reação F, temos 179 funções no conjunto 6-311+G(d) e 752 para o conjunto TZVPP). O menor conjunto, 6-311+G(d), é semelhante ao 6-31+G*, usado nas otimizações de geometria.

Para as três reações estudadas (tabelas 14, 15 e 13), o menor conjunto base cuja energia relativa HF difere menos que 1 kcal/mol da energia relativa calculada pelo conjunto TZVPP, é o 6-311+G(2d,2p). Esta convergência relativa ocorre para $E^{(2)}$ no conjunto 6-311+G(2df,2p). As E_T s obtidas pelos conjuntos 6-311+G(2df,2p), 6-311+G(2df,2pd), 6-311+G(3df,3pd) diferem da energia TZVPP em cerca de $2 \cdot 10^{-3}$ au. Esta variação é maior que $4 \cdot 10^{-3}$ au para os conjuntos menores.

2.2.4 Comparação com medidas experimentais

A calibração de cálculos de estrutura eletrônica de espécies isoladas deve ser feita pela comparação com valores experimentais obtidos em fase gasosa. Porém, dados experimentais, cinéticos ou termodinâmicos, para reações de ataque nucleofílico sobre fósforo^[40, 41], comparáveis aos resultados obtidos para as reações A a H, não estão disponíveis (ao conhecimento do autor).

As energias das seguintes reações de transferência de H^+ podem ser obtidas a partir de dados de entalpia de dissociação de H^+ em fase gasosa^[213–215] e comparadas diretamente com computações *ab initio*:

Estas reações são relevantes para a catálise enzimática nas PTPs e foram usadas na parametrização específica da hamiltoniana semiempírica (§3.2.1).

A entalpia de dissociação de H^+ de $PhOPO_3H_2$ foi estimada em 320 kcal/mol, pois medidas experimentais para esta espécie não foram encontradas. Esta estimativa foi baseada nos valores de dissociação de H^+ de H_3PO_4 e $(CH_3O)_2PO_2H$ em fase gasosa, respectiva-

Figura 25: Energias absolutas (unidades atômicas) Hartree-Fock (A) e de correlação eletrônica (B) para segunda etapa da reação F ($D_n + A_n$) em função do tamanho do conjunto base.

Tabela 13: Energias relativas Hartree-Fock e de correlação eletrônica para segunda etapa da reação F ($D_n + A_n$) em função do conjunto base^a

Base ^b	funções ^c	E_{HF}			$E^{(2)}$		
		I^d	II^e	III^f	I	II	III
d	130	-0.021	-0.065	0.044	0.003	0.015	-0.013
d,p	172	-0.024	-0.064	0.041	0.005	0.020	-0.015
2d,2p	214	-0.022	-0.062	0.040	0.003	0.019	-0.016
2d,2pd	234	-0.022	-0.062	0.040	0.004	0.019	-0.015
2df,2p	256	-0.022	-0.063	0.041	-0.001	0.017	-0.017
2df,2pd	276	-0.022	-0.063	0.040	-0.001	0.017	-0.017
3df,3pd	318	-0.022	-0.062	0.041	0.001	0.017	-0.016
TZVPP	386	-0.023	-0.062	0.039	0.000	0.017	-0.017

^a valores de energia em unidades atômicas. ^b tipo e quantidade de funções de polarização adicionadas ao conjunto 6-311+G, excetuando-se a base TZVPP^[136] (§1.4.1.1). ^c número total de funções base para cada uma das espécies (F:IMC2d, F:TS2d e produtos). ^d $E(\text{produtos}) - E(\text{F:IMC2d})$, ^e $E(\text{produtos}) - E(\text{F:TS2d})$ e ^f $E(\text{F:TS2d}) - E(\text{F:IMC2d})$, onde E é energia.

Tabela 14: Energias relativas Hartree-Fock e de correlação eletrônica para reação C em função do conjunto base^a

Base ^b	funções ^c	E_{HF}			$E^{(2)}$		
		I^d	II^e	III^f	I	II	III
d	179	-0.018	-0.028	-0.007	-0.010	-0.003	-0.007
d,p	218	-0.002	-0.018	-0.008	-0.016	-0.008	-0.008
2d,2p	300	0.000	-0.016	-0.009	-0.016	-0.007	-0.009
2d,2pd	386	0.003	-0.014	-0.008	-0.017	-0.008	-0.008
TZVP	403	-0.001	-0.017	-0.010	-0.018	-0.009	-0.010
2df,2p	456	0.003	-0.015	-0.009	-0.019	-0.010	-0.009
2df,2pd	516	0.003	-0.015	-0.009	-0.019	-0.010	-0.009
3df,3pd	602	0.002	-0.015	-0.009	-0.018	-0.009	-0.009
TZVPP	752	0.002	-0.015	-0.010	-0.020	-0.010	-0.010

^a valores de energia em unidades atômicas. ^b tipo e quantidade de funções de polarização adicionadas ao conjunto 6-311+G, excetuando-se as bases TZVP e TZVPP^[136] (§1.4.1.1). ^c número total de funções base para cada uma das espécies (C:IMC1, C:TS1 e C:I). ^d $E(\text{C:I}) - E(\text{C:IMC1})$, ^e $E(\text{C:I}) - E(\text{C:TS1})$ e ^f $E(\text{C:TS1}) - E(\text{C:IMC1})$, onde E é energia.

Tabela 15: Energias relativas Hartree-Fock e de correlação eletrônica para segunda etapa da reação E $A_n + D_n$ em função do conjunto base^a

Base ^b	funções ^c	E_{HF}			$E^{(2)}$		
		I^d	II^e	III^f	I	II	III
d	176	0.063	0.004	0.059	-0.018	-0.004	-0.014
d,p	240	0.061	0.003	0.058	-0.017	-0.003	-0.013
2d,2p	304	0.061	0.004	0.057	-0.020	-0.004	-0.016
2d,2pd	344	0.061	0.004	0.057	-0.020	-0.004	-0.016
2df,2p	360	0.063	0.004	0.059	-0.024	-0.004	-0.020
2df,2pd	400	0.063	0.004	0.059	-0.024	-0.004	-0.020
3df,3pd	464	0.065	0.005	0.060	-0.023	-0.004	-0.019
TZVPP	572	0.062	0.005	0.058	-0.024	-0.005	-0.020

^a valores de energia em unidades atômicas. ^b tipo e quantidade de funções de polarização adicionadas ao conjunto 6-311+G, excetuando-se a base TZVPP^[136], (§1.4.1.1). ^c número total de funções base para cada uma das espécies (E:IMC1a, E:TS1a e E:Ia). ^d $E(\text{E:Ia}) - E(\text{E:IMC1a})$, ^e $E(\text{E:Ia}) - E(\text{E:TS1a})$ e ^f $E(\text{E:TS1a}) - E(\text{E:IMC1a})$, onde E é energia.

mente, 330.5 e 331.6 kcal/mol^[40, 80], e na diferença de 8.5 kcal/mol entre a entalpia de dissociação de H⁺ do ácido benzóico (PhCOOH) e do ácido acético (MeCOOH), 340.0 e 348.5 kcal/mol, respectivamente^[213].

A tabela 16 mostra os valores calculados (E_T) e experimentais para as energias das reações 2.1 a 2.4.

Tabela 16: Energias das reações de transferência de H⁺ em fase gasosa calculadas e experimentais^a

Reação	E_T	Experimental ^{b[213-215]}
2.1	-2.3	-1.8 ± 2.8
2.2	-35.5	-33.4 ± 3.2
2.3	26.8	28.6 ± 9
2.4	35.9	37.6 ± 8

^a Energias em kcal/mol. ^b O erro das medidas experimentais é a soma dos erros das entalpias de dissociação de H⁺ medidas para cada uma das espécies. O erro para o fenilfosfato foi estimado em 7 kcal/mol.

Outra comparação com valores experimentais pode ser feita entre as barreiras torcionais para ligação C—O no metanol (H₃C—OH)^[216] e C—C no etano (H₃C—CH₃)^[217] que, em fase gasosa, são respectivamente 0.7 e 2.9 kcal/mol, ambas com erro de ± 0.4 kcal/mol. Os valores calculados no nível MP2/6-311+G(2df,2p)//MP2/6-31+G* são 0.79 kcal/mol para torção no metanol e 2.97 kcal/mol para torção no etano.

A tabela 17 mostra os dados experimentais disponíveis e os respectivos valores calculados de potenciais de ionização eletrônica, momento de dipolo, geometrias e frequências vibracionais para alguns dos reagentes estudados. As propriedades estruturais calculadas para CH₃OH, CH₃SH e CH₃COOH são muito próximas das experimentais. As frequências vibracionais calculadas para o TMP diferem menos que 2% dos valores experimentais e os momentos de dipolo calculados diferem menos que 15%. Os potenciais de ionização calculados são próximos aos experimentais para TMDMP, PhOH e CH₃SH, mas o valor calculado para o TMP é 1 eV maior do que o experimental. Em geral, os valores calculados para moléculas contendo enxofre estão mais próximos (e em alguns casos idênticos) aos experimentais do que os valores calculados para outras moléculas.

Estruturas obtidas por espectroscopia de raios-X de cristais sólidos de dois derivados de ésteres de fosfato estão disponíveis. A comparação com valores calculados na fase gasosa para moléculas com composição química e estrutura diferentes deve ser analisada cuidadosamente. O S-4-nitrofenil O,O-difenil tiofosfato^[218] tem distâncias de ligação P—O de 1.592 e 1.579 Å, P=O de 1.464 Å, C—O de 1.372 e 1.385 Å, ângulos de ligação O—P=O

Tabela 17: Comparação entre os potenciais de ionização, momento de dipolo, frequências vibracionais e geometrias calculados e medidos em fase gasosa

Espécie	Propriedade ^a	Calculado ^b	Experimental
TMP	PI	11.8	10.8 ^[220]
	$\nu(\text{P}=\text{O})$	1270	1291 ^[221, 222]
	$\nu(\text{P}-\text{OC})$	1079	1060
	$\nu(\text{PO}-\text{C})$	853	866
TMDMP	PI	9.54	9.55 ^[220]
PhOH	PI	8.89	8.75 ^[223]
	dipolo	1.45	1.41 ^[224]
CH ₃ OH	PI	11.36	10.96 ^[225]
	dipolo	1.87	1.70 ^[224]
	O-C	1.431	1.425 ^[226]
	O-H	0.972	0.945
	C-H	1.096	1.094
	C-O-H	108.7	108.5
	H-C-O	105.9	107.0
	H-C-H	109.0	108.6
CH ₃ SH	PI	9.42	9.44 ^[227]
	dipolo	1.67	1.52 ^[224]
	S-C	1.816	1.819 ^[228]
	S-H	1.340	1.335
	C-H	1.091	1.091
	C-S-H	96.8	96.5
	H-C-H	109.8	109.8
CH ₃ COOH	dipolo	2.07	1.74 ^[224]
	C-C	1.500	1.520 ^[229]
	C=O	1.221	1.214
	C-O	1.365	1.364
	O-H	0.980	0.970
	C-C=O	126.4	126.6
	C-C-O	111.2	110.6
	C-O-H	106.4	107.0

^a PI é o potencial de ionização eletrônica em eV, dipolo é o momento de dipolo em debyes, $\nu(\text{X}-\text{Y})$ indica a frequência vibracional em cm^{-1} e $\text{X}-\text{Y}$ indica distâncias de ligação entre os átomos X e Y em Å e $\text{X}-\text{Y}-\text{Z}$ indica ângulos entre as ligações $\text{X}-\text{Y}$ e $\text{Y}-\text{Z}$ em °. O erro das medidas de PI é ± 0.3 eV e os erros das geometrias experimentais são variáveis, porém, nunca ultrapassam a última casa decimal mostrada. Não temos estimativas dos erros experimentais para momento de dipolo. Todos potenciais de ionização, com exceção do TMDMP, são verticais.

^b no nível MP2/6-311+G(2df,2p)//MP2/6-31+G*.

de 116.2° , O–P–O de 102.3° e P–O–C de 128.2° . Estes valores podem ser comparados à estrutura do DMPP, com distâncias de ligação P–OPh de $1.623 \text{ \AA}$, P=O de $1.490 \text{ \AA}$, C–OPh de $1.407 \text{ \AA}$, ângulos de ligação PhO–P=O de 116.4° , PhO–P–OMe de 101.6° e P–OPh–C de 118.5° . O ácido S-2-(3-aminopropilamino)etilfosforotióico^[219] tem distâncias de ligação P–S de $2.116 \text{ \AA}$, P–O de 1.521 e $1.511 \text{ \AA}$, S–C de $1.824 \text{ \AA}$, e ângulos de ligação P–S–C de 103.7° , S–P–O de 107.1 , 103.0 e 106.6° , O–P–O de 115.5° . Estes valores podem ser comparados à estrutura do TMDMP, com distâncias de ligação P–S de $2.077 \text{ \AA}$, P–O de 1.609 e $1.600 \text{ \AA}$, S–C de $1.822 \text{ \AA}$, e ângulos de ligação P–S–C de 97.7° , S–P–O de 114.3 , 103.8 e 104.5° , O–P–O de 116.7° . O erro nas medidas experimentais é de $0.004 \text{ \AA}$ para as distâncias de ligação e 0.1° para os ângulos. As diferenças entre os valores calculados e experimentais são menores que 5% dos valores experimentais.

2.3 Discussão

2.3.1 Calibração da metodologia de cálculo *ab initio*

As frequências vibracionais dos reagentes, produtos, mínimos e TSs foram obtidas utilizando a aproximação harmônica, que assume que as estruturas são um ponto estacionário na PES Born-Oppenheimer, *i.e.*, o gradiente (negativo da força) em relação às coordenadas atômicas é nulo (§1.5.1). Estruturas cujo gradiente é nulo num nível de cálculo e conjunto base não têm necessariamente o gradiente nulo se a estrutura eletrônica é computada noutro nível de cálculo ou conjunto base e, portanto, frequências harmônicas devem ser calculadas pelo mesmo nível e base empregados na otimização de geometria.

Otimizamos as geometrias no nível MP2/6-31+G* para todas reações estudadas, exceto a reação A, em que empregamos o nível MP2/6-31G*. No entanto, em face de uma limitação computacional (§2.1), não obtivemos frequências vibracionais para as reações B, D, F(A_n+D_n) e H(A_n+D_n) no nível MP2/6-31+G*. As geometrias das espécies destas reações foram reotimizadas e suas frequências harmônicas foram computadas no nível B3LYP/6-31+G*.

O desvio quadrático médio (RMSD) foi calculado para os três tipos de coordenadas internas definidas para cada espécie reotimizada no nível B3LYP/6-31+G*, em relação às coordenadas obtidas no nível MP2. Ao contrário do RMSD calculado para coordenadas cartesianas, os RMSDs obtidos para as distâncias e ângulos de ligação, e ângulos diedrais permitem-nos avaliar o desvio de cada tipo de coordenada interna separadamente. Não

Tabela 18: RMSD máximo e médio das coordenadas internas para as espécies reotimizadas no nível B3LYP/6-31+G(d) ^a

Reação ^b	RMSD			RMSD _{max}		
	ligações	ângulos	diedrais	ligações	ângulos	diedrais
B	0.026	3.25	6.45	0.050	5.66	9.30
D	0.013	1.20	4.39	0.036	1.86	6.63
F(A _n +D _n)	0.024	1.49	6.78	0.082	3.41	10.44
H(A _n +D _n)	0.023	3.04	10.4	0.036	6.80	30.89

^a RMSD de distâncias de ligação em Å e para ângulos entre ligações e ângulos diedrais em °.

^b Reação cujas espécies foram reotimizadas. Veja o texto para detalhes de como os RMSDs são calculados.

mostramos aqui os RMSDs para cada estrutura, mas somente a média dos RMSDs das 5 estruturas (IMCs, TSs e intermediário) para cada reação, $\overline{\text{RMSD}}$, e o máximo RMSD para cada reação, RMSD_{max} (tabela 18). Os reagentes e produtos não foram incluídos na tabela, porque o desvio nestes compostos é significativamente inferior ao desvio nas outras espécies.

Os desvios máximos para as distâncias e os ângulos de ligação são pequenos e nunca ultrapassam 0.1 Å e 7°, respectivamente. O desvio nos ângulos diedrais é maior, chegando a 30.9° no H:IMC1. Praticamente todos RMSD_{max} correspondem a IMCs, porque estes são complexos entre duas espécies que não estão ligadas covalentemente. Assim, as coordenadas internas que envolvem a ligação que foi quebrada ou ainda não foi formada são aquelas que sofrem maior variação com a reotimização das geometrias.

Ângulos diedrais possuem modos vibracionais de baixa frequência, que contribuem pouco para a *ZPE*. As maiores contribuições provêm dos modos de alta frequência (estiramento e angulação de ligações). Portanto, a diferença das *ZPE* calculadas nos níveis B3LYP e MP2 deve ser pequena, já que as principais diferenças entre as geometrias calculadas nos dois níveis são observadas nos ângulos diedrais. A *ZPE* também é pouco sensível ao nível *ab initio* usado no cálculo de frequências vibracionais^[202]. As diferenças entre *ZPE*s calculadas nos níveis B3LYP e MP2 para pontos estacionários das reações E, F e G (tabela 19) não ultrapassam 2.10^{-3} au (cerca de 1.25 kcal/mol) para os valores absolutos e são menores do que 5.10^{-4} au (cerca de 0.3 kcal/mol) para valores relativos, mesmo para espécies com RMSD em torno de 0.05 Å para distâncias de ligação e 10° para os ângulos diedrais. Concluímos que os valores de E_T podem ser comparados entre as diferentes reações estudadas, independentemente do nível de cálculo (B3LYP ou MP2) empregado para obtenção das *ZPE*.

Tabela 19: RMSD das coordenadas internas e ZPE para as espécies reotimizadas no nível B3LYP/6-31+G(d)

Espécie	RMSD_{lig}^a	RMSD_{ang}^b	RMSD_{die}	ZPE_{B3LYP}^c	ZPE_{MP2}
E:IMC2d	0.011	1.77	9.66	0.05877	0.05986
E:TS2d	0.046	1.91	2.36	0.05573	0.05670
E:produto	0.014	0.65	3.69	0.06174	0.06316
F:reagente	0.004	1.12	5.78	0.11732	0.11689
G:IMC1a	0.033	3.65	4.45	0.11818	0.12046
G:TS1a	0.014	1.02	3.37	0.11531	0.11659
G:1a	0.007	0.74	3.25	0.11949	0.12172

^a RMSD de distâncias de ligação (lig) em Å. ^b RMSD de ângulos entre ligação (ang) e die-drais (die) em °. ^c ZPE em unidades atômicas obtidas no nível B3LYP (ZPE_{B3LYP}) e MP2 (ZPE_{MP2}).

Todos pontos estacionários obtidos para espécies diferentes dos reagentes e produtos isolados apresentam cerca de $\frac{1}{3}$ de seus modos vibracionais com frequências inferiores a 500 cm^{-1} . A maioria destes modos são torções internas e angulações coletivas e, logo, não devem ser tratados como vibrações harmônicas, caso contrário, erros significativos são observados nas propriedades termodinâmicas calculadas, especialmente em $S_{vib}^{[31]}$ (apêndice B). O cálculo apropriado das contribuições termodinâmicas dos modos vibracionais de baixa frequência é demorado, não é automatizado e apenas o tratamento de torções internas (como um rotor obstruído)^[31] está implementado no sistema de programas usado nas computações (§2.1). Angulações coletivas e misturas com as torções necessitam de um tratamento diferenciado. Um exemplo da magnitude dos erros introduzidos pelo tratamento inapropriado de modos de baixa frequência é observado na comparação entre as computações para as mesmas reações de hidrólise de ésteres de fosfato realizadas por Florián e Warshel^[73] (reações 13 e 17 do trabalho citado) e por Pedersen et al. (reações 4 e 3, respectivamente)^[230]. Os níveis de cálculo empregados nos dois estudos foram semelhantes (MP2/6-31+G**//HF/6-31G*^[73] e MP2/6-31+G*^[230]), mas, embora a diferença observada entre a energia eletrônica e de repulsão nuclear para as duas reações seja apenas 0.2 e 0.3 kcal/mol, as diferenças na entropia de reação, calculada na aproximação do oscilador harmônico (apêndice B), são de 4.8 e 4.6 kcal/mol a 298 K! Estas diferenças resultam em considerável desacordo entre as energias totais (livres) calculadas nestes dois estudos^[73, 230].

Portanto, as energias totais, E_T , apresentadas na §2.2 são compostas apenas por contribuições eletrônicas e de repulsão nuclear, E_{MP2} , e por ZPE . A obtenção de valores acurados de energia total a 0 K parece-nos mais útil do que obscurecer os resultados com contribuições térmicas e entrópicas pouco precisas.

Nenhuma correção para o BSSE (p. 54) foi adotada para as reações estudadas, pois energias de complexação dos reagentes em fase gasosa não são propriedades relevantes para a reatividade de ésteres de fosfato e para calibração específica do método semiempírico (§3.2.1). Usando o conjunto base 6-311++G**, o BSSE nas energias de complexação dos reagentes e produtos da reação A foi estimado em $\frac{1}{8}$ do valor das energias. Espera-se que energias de complexação obtidas pelo conjunto 6-311+G(2df,2p) apresentem BSSE igual ou inferior a $\frac{1}{8}$ dos valores já que este conjunto é mais completo do que o 6-311++G**.

O uso de conjuntos base atômico e expansões N -eletrônicas (§1.4.1) maiores e mais completos é certamente necessário para o cálculo de energias com alta precisão, principalmente, energias de correlação eletrônica. Todavia, a quantidade de recursos computacionais necessária para o cálculo *ab initio* cresce exponencialmente com o tamanho das moléculas^[121, §9.12]. Por exemplo, a computação da energia MP2 para cada espécie da reação C usando o conjunto base TZVPP^[136] (§1.4.1.1), o maior conjunto empregado nesta tese, dura cerca de 20 horas de CPU. O tempo estimado para computação da energia MP2 para cada espécie da reação B ou D, com os mesmos computadores e conjunto base, é cerca de 400 horas! Portanto, temos que encontrar um balanço entre a completeza dos conjuntos mono e N -eletrônico e o custo computacional.

Uma vez que testes (citados na p. 118, mas não mostrados explicitamente nesta tese) sugeriram que efeitos de correlação eletrônica não-dinâmica^[121, §5.2.10] não eram importantes para as reações de ésteres de fosfato, o tratamento adotado para resgatar a energia de correlação eletrônica foi a teoria de perturbação em segunda ordem (§1.4.1). Este é o método que parece-nos mais indicado para computações da estrutura eletrônica de moléculas do tamanho daquelas estudadas aqui, para a precisão que estamos interessados e para os recursos computacionais disponíveis (p. 82).

Discutiremos abaixo os testes da convergência da energia eletrônica calculada em função de diferentes conjuntos base (§2.2.3) e as comparações com valores experimentais (§2.2.4) usadas como calibração do nível de cálculo *ab initio* empregado nesta tese.

Na §2.2.3, procuramos pelo menor conjunto base que resultasse em energias *relativas* convergidas em relação a uma referência, mas que também apresentasse um custo computacional acessível para cálculos no nível MP2 em todas reações estudadas. Na ausência de dados experimentais para as reações de ésteres de fosfato (§2.2.4), os valores de energia obtidos pelo conjunto TZVPP foram escolhidos como referência.

Embora um estudo da convergência da energia eletrônica em função do conjunto base deva utilizar conjuntos que são ampliados consistentemente^[121, §8.3], como os conjuntos

cc-pVXZ propostos por Dunning^[207], onde X é conhecido como o número cardinal, ou seja, $X = 2$ para bases de qualidade duplo- ζ , $X = 3$ para bases de qualidade triplo- ζ e assim por diante, o custo computacional para cálculos de espécies das reações estudadas é muito elevado se $X > 3$. Assim, fixamos a qualidade dos conjuntos estudados em triplo- ζ e testamos o efeito da adição de funções de maior momento angular ao conjunto 6-311+G (cuja contração é dada na p. 52). Ao contrário do aumento cardinal dos conjuntos cc-pVXZ, esta seqüência, usando a base 6-311+G, não é bem definida, mas é o maior conjunto que pode ser empregado dentro dos recursos computacionais disponíveis.

As energias absolutas mostradas na figura 25 sugerem que a convergência absoluta ou o limite assintótico da energia HF é próximo da energia calculada pelo conjunto TZVPP (pontos com maior número de funções base na figura 25). O conjunto 6-311+G(2df,2p) (pontos com 256 funções base) resulta em energias HF que são 0.013 au (cerca de 8 kcal/mol) menos negativas do que a energia calculada pelo conjunto TZVPP.

Energias absolutas HF, calculadas para moléculas simples contendo átomos do primeiro período, empregando conjuntos base de qualidade quádruplo- ζ , diferem menos que 2 kcal/mol dos valores HF obtidos com um conjunto base completo^[121, p. 316]. Já um conjunto base triplo- ζ , comparável ao TZVPP, apresenta diferenças de 6.1 e 6.4 kcal/mol nas energias calculadas respectivamente para N_2 e H_2O , em comparação aos valores HF obtidos com um conjunto base completo^[121, p. 316]. Assim, supomos que o conjunto 6-311+G(2df,2p) resulta em energias absolutas HF que diferem cerca de 15 kcal/mol daquelas que seriam obtidas com um conjunto base completo.

Porém, a convergência da energia absoluta de correlação eletrônica em segunda ordem, $E^{(2)}$, é mais lenta (figura 25). As $E^{(2)}$ calculadas por conjuntos de qualidade triplo- ζ , acrescidos de funções de polarização (§1.4.1.1), como o conjunto TZVPP, ainda estão longe do limite assintótico (p. 53)^{[121, cap. 15][138]}.

Contudo, as energias de reação e alturas de barreira que determinam a reatividade química dependem apenas de valores de energia *relativa*. Assim, esperamos que os erros introduzidos nas energias absolutas pela truncagem das hierarquias mono e N -eletrônica, usadas na solução da equação de Schrödinger (§1.4.1), sejam constantes ao longo das reações. Os erros resultantes destas duas aproximações muitas vezes apresentam sinais contrários e podem se cancelar fortuitamente^{[121, cap. 15][138]}.

As energias relativas que diferem menos que 1 kcal/mol da energia relativa obtida pelo conjunto TZVPP, são denominadas energias *relativas* convergidas (§2.2.3). Para as reações C, E ($A_n + D_n$) e F ($D_n + A_n$) (tabelas 14, 15 e 13), a menor base cujas energias relativas

HF e $E^{(2)}$ estão convergidas é o conjunto 6-311+G(2df,2p). O conjunto 6-311+G(2df,2p) também é empregado para o cálculo da energia MP2 no esquema de extrapolação G2^[160]. Energias de reação e ativação obtidas por este esquema apresentam RMSD menores que 2 a 3 kcal/mol em comparação com valores experimentais para outras reações simples^[160].

Duas funções bases tipo p para o hidrogênio, e duas tipo d e uma tipo f para os átomos pesados precisam ser adicionadas ao conjunto 6-311+G para convergência relativa de $E^{(2)}$. Funções base tipo d no hidrogênio e f nos átomos pesados parecem desnecessárias para convergência da energia relativa HF nas reações calculadas. Funções p e d além de (2df,2p), como no conjunto 6-311+G(3df,3pd), parecem ser dispensáveis tanto para E_{HF} e $E^{(2)}$.

A aproximação de congelamento dos elétrons de caroço (p. 51) foi usada e, logo, as funções base que descrevem estes elétrons contribuem pouco para energia de correlação eletrônica. Não avaliamos a contribuição da correlação eletrônica de caroço para as bases TZVPP ou a seqüência de bases 6-311+G. Para outros sistemas, esta contribuição é pequena e se cancela ao longo do caminho de reação^{[121, cap. 15][138]}.

A partir da entalpia de dissociação de H^+ medida em fase gasosa, construímos 4 reações de transferência de H^+ que podem ser diretamente comparadas aos resultados *ab initio*. Nas reações 2.1, 2.2 e 2.3 (p. 103), o ácido acético é um modelo para a cadeia lateral do ASP envolvido na catálise realizadas pelas PTPs e os outros grupos são modelos para os possíveis grupos de saída e substrato enzimáticos (§1.3). A reação 2.4 (p. 103) é um passo relevante para a reação F e um modelo para transferência de H^+ entre o substrato e a cadeia lateral da CYS no sítio ativo de PTPs, como proposto recentemente^[52]. A construção de reações de transferências de H^+ entre dois pares ácido-base, permite a comparação entre valores relativos de entalpia de reação, ao invés da comparação entre valores absolutos de entalpia de dissociação de H^+ . As entalpias experimentais^[213–215], obtidas em fase gasosa a baixa pressão, são diretamente comparáveis com os valores de E_T , já que este contém as principais contribuições ($E_{MP2} + ZPE$) presentes nas entalpias experimentais. Os resultados da tabela 16 mostram que a diferença entre os valores calculados e experimentais é menor que 2.4 kcal/mol. Esta diferença é menor que erro experimental para as energias de reação e, portanto, os valores calculados e experimentais são coincidentes.

Não encontramos medidas experimentais para a entalpia de dissociação de H^+ de fenilfosfato. Este valor foi estimado com base nas entalpias de dissociação de H_3PO_4 e $(CH_3)_2PO_2H$, corrigidas para a presença de um grupo fenil. A correção corresponde

à diferença da entalpia de dissociação entre ácido benzóico (PhCOOH) e ácido acético (MeCOOH), ou seja, ao efeito da substituição de um grupo metil por um fenil. A maior acidez do fenilfosfato deve-se a deslocalização da carga negativa pelo anel aromático, que estabiliza o fenilfosfato monoaniônico em fase gasosa. Este efeito também é observado na entalpia de dissociação de H^+ de fenol, que é cerca de 30 kcal/mol inferior a do metanol.

As barreiras conformacionais de metanol e etano calculadas também são coincidentes com aos valores experimentais (p. 106). Os valores calculados para energias conformacionais e de reação estão muito próximos dos valores experimentais. Estes resultados sugerem que energias totais calculadas como descrito na §2.1 são suficientemente acuradas para análise da reatividade intrínseca de ésteres de fosfato (§2.3.2) e para a parametrização específica da hamiltoniana semiempírica (§3.2.1).

Uma comparação entre entalpias de reação experimentais e calculadas para um conjunto de 17 reações isogíricas (entre moléculas de camada fechada), contendo apenas átomos do primeiro período, sugeriu que melhorias significativas sobre os resultados obtidos no nível MP2, empregando um conjunto base triplo- ζ , somente são obtidas caso um nível de cálculo com correções para excitações triplas, como CCSD(T), e um conjunto base quádruplo- ζ fossem empregados no cálculo da estrutura eletrônica^[121, §15.8]. Para espécies como as da reação B, o custo de cálculos no nível CCSD(T), com um conjunto base quádruplo- ζ , é impraticável, mesmo para as maiores capacidades computacionais existentes hoje. Esta comparação também mostra que correções na estrutura eletrônica e na energia *ab initio* referentes a efeitos relativísticos, anamorticidade dos modos vibracionais e correlação dos elétrons de caroço resultam em mudanças de menos de 0.5 kcal/mol nas entalpias de reação calculadas^[121, cap. 15]. Portanto, estas correções podem ser negligenciadas.

Os potenciais de ionização *eletrônica* foram calculados verticalmente, ou seja, a geometria usada para o cálculo das espécies ionizada e neutra é a mesma. Ionizações adiabáticas são obtidos relaxando-se a geometria da espécie ionizada, ou seja, as espécies, antes e após a ionização, têm suas respectivas geometrias de equilíbrio. Todos valores experimentais comparados são PI verticais, com exceção do PI adiabático medido para o TMDMP. Os PIs calculados (tabela 17) diferem dos experimentais em 1 eV para o TMP, 0.4 eV para CH_3OH , 0.14 eV para PhOH e cerca de 0.02 eV para CH_3SH e TMDMP. O ótimo desempenho do cálculo de PI calculados para PhOH e, principalmente, para as moléculas contendo enxofre (também observado para $\text{CH}_3\text{CH}_2\text{SH}$, não incluído na tabela 17: $\text{PI}_{exp} = 9.29 \text{ eV}$ ^[227] e $\text{PI}_{calc} = 9.28 \text{ eV}$), sugere que PIs podem ser calculados quantitativamente

pela metodologia empregada aqui. Para CH_3OH e, principalmente, TMP, os erros nos PI calculados são significativos, porém, não ultrapassam 10% dos valores experimentais.

O desempenho inferior no cálculo de PI para os compostos contendo oxigênio pode ser consequência da maior deformação estrutural que a espécie ionizada sofreria durante o processo de ionização. Neste caso, o princípio de Franck-Condon não é respeitado e o PI vertical reflete a medida da espécie ionizada em algum estado vibracional excitado^[231]. Espécies de oxigênio têm uma nuvem eletrônica menos polarizável do que espécies de enxofre e aromáticas e, portanto, devem sofrer uma deformação nuclear maior em resposta a alteração da estrutura eletrônica. Esta explicação está de acordo com a maior diferença observada entre o PI vertical e o PI adiabático do TMP, que é 0.8 eV^[220, 232], e a menor diferença observada para as espécies cujo cálculo de PI teve bom desempenho (por exemplo, a diferença entre o PI adiabático e o PI vertical para o CH_3SH é 0.02 eV^[227]).

Momentos de dipolo elétrico medidos na fase gasosa^[224] correspondem a dipolos no estado vibracional fundamental, μ_0 . Os dipolos calculados correspondem a dipolos de equilíbrio (mínimo da PES Born-Oppenheimer), μ_e . O μ_0 pode ser computado por uma média vibracional de dipolos^[121, §15.5]. Este procedimento é custoso e possível apenas para moléculas diatômicas. As diferenças entre μ_0 e μ_e , medidos experimentalmente para moléculas simples, variam até 0.09 D, portanto, podemos estimar uma incerteza de cerca de 0.1 D para comparação entre os μ_e calculados aqui e os μ_0 experimentais. A diferença entre os valores experimentais e calculados é pequena: 0.04 D para o PhOH, 0.17 D para CH_3OH , 0.15 D para CH_3SH e 0.27 D para o ácido acético (tabela 17). Não conhecemos o erro das medidas experimentais, mas, em virtude da incerteza inerente a comparação de μ_e e μ_0 , consideramos que os valores de momento de dipolo calculado estão próximos aos experimentais. A flexibilização da descrição eletrônica da região exterior da camada de valência, acrescentando funções difusas ao conjunto base^[121, p. 840], resultaria em valores calculados ainda mais próximos dos experimentais. Esta flexibilização aumenta o custo computacional e não resulta em melhorias significativas em outras propriedades moleculares^[121, p. 840].

O espectro vibracional observado para TMP em fase gasosa contém contribuições de mais de uma população de confôrmeros^[233]. Em experimentos de isolamento em matriz, três bandas vibracionais foram atribuídas ao confôrmero com simetria C_3 , o mais estável na fase gasosa e aquele adotado em todas computações realizadas nesta tese. Estas três frequências diferem em até 21 cm^{-1} dos valores calculados no nível MP2/6-31+G* (tabela 17). A pequena diferença (menos de 2% dos valores) sugere que as geometrias e as

freqüências calculadas para ésteres de fosfato são bastante próximas aos valores experimentais.

Para as geometrias medidas em fase gasosa, o erro máximo experimental estimado é 0.5 pm para distâncias de ligação e 0.5° para os ângulos entre as ligações. Observamos que os desvios nas distâncias de ligação calculadas em comparação com os valores experimentais são inferiores a 0.5 pm, com exceção das distâncias O–H no metanol, e C–C e O–H no ácido acético, cujos desvios não superam 3 pm (tabela 17). Testes para moléculas contendo átomos do primeiro período mostram que as distâncias de ligação calculadas diminuem com o aumento do conjunto base e aumentam com o maior resgate da correlação eletrônica^[121, p. 822]. Estes erros se cancelam em grande parte no nível MP2 e um conjunto base triplo- ζ . As distâncias de ligação obtidas neste nível de cálculo apresentam erros menores do que 0.5 pm^[121, p. 831]. Nas geometrias calculadas nesta tese, as distâncias de ligação são em geral superiores aos valores experimentais. Este desvio ocorre porque o conjunto base (6-31+G*) usado nas otimizações de geometria é menor que triplo- ζ .

Os erros esperados no cálculo de ângulos de ligação são similares^[121, §15.3 e §15.4]. A maior diferença entre os ângulos calculados e experimentais é 1.1° , correspondente ao ângulo H–C–O do metanol. Para os demais ângulos, a diferença entre os valores experimentais e calculados é igual ou inferior a 0.4° (tabela 17).

Em geral, os resultados estruturais calculados estão bastante próximos aos experimentais, considerando a faixa de erro estimada para as medidas. Portanto, podemos considerar que as geometrias obtidas neste estudo são representativas das reações de ésteres de fosfato. Os resultados para propriedades como momento de dipolo e PI também são suficientemente acurados para nossos propósitos.

Concluimos que a metodologia adotada para cálculo de estrutura eletrônica (§2.1) nesta tese, composta por um tratamento perturbativo em segunda ordem para a correlação eletrônica (MP2) e um conjunto base razoavelmente estendido [6-311+G(2df,2p)], reproduz observações experimentais e tem a capacidade preditiva e a acurácia desejadas.

2.3.2 Tiólise e alcoólise de ésteres de fosfato

A alcoólise e tiólise de ésteres de fosfato não é facilmente observada experimentalmente, mesmo em solução (§1.2). Portanto, as reações A a H são uma referência para a análise da reatividade intrínseca de ésteres de fosfato e dos possíveis mecanismos de catálise em PTPs. Nesta seção, os caminhos de reação (§2.2) serão discutidos, compa-

rando as estruturas das espécies e os perfis de energia de tiólise e alcoólise dos dos aril e alquil ésteres de fosfato.

A geometria dos IMCs é facilmente deformável, pois os dois reagentes (ou os produtos) não estão covalentemente ligados. Em solução aquosa, estes complexos são ainda menos estáveis, pois a solvatação blindas as cargas formais das espécies, reduzindo a interação eletrostática^[73]. Os complexos de encontro (RS, figura 1) gerados têm geometrias diferentes das geometrias em fase gasosa. O sítio ativo enzimático possui estrutura complementar ao substrato, que estabiliza a formação do complexo de Michaelis (§1.1) e, assim, a geometria deste complexo é diferente daquela observada em fase gasosa. Portanto, não discutiremos a estrutura dos IMCs calculados na fase gasosa.

As DRCs para as espécies estacionárias nas reações associativas de tiólise de mono e triésteres são em geral cerca de 0.5 Å menores do que as DRCs das espécies formadas na alcoólise dos mesmos reagentes. Esta diferença é consequência da maior distância da ligação P–S. Por exemplo, $d(\text{PS}) = 2.08 \text{ Å}$ para o TMDMP e $d(\text{PO}) = 1.60 \text{ Å}$ para o TMP. No entanto, os TSs da etapa de ataque nucleofílico nas quatro reações de triésteres possuem DRCs (p. 89) e distâncias de ligação (tabela 8) próximas.

Nas reações de triésteres (figura 13), os TSs de quebra ou formação de ligação P–O entre CH_3O^- e o grupo fosfato apresentam os grupos equatoriais torcidos como nos IMCs. Os hidrogênios dos grupos metila equatoriais estão coordenados ao oxigênio do CH_3O^- . Nos TSs de formação da ligação P–S e de quebra da ligação P–O de um grupo fenil, os grupos equatoriais estão torcidos como nos intermediários (figura 13). Todos TSs das reações associativas de monoésteres têm os grupos OH equatoriais torcidos como nos intermediários (figura 19). Portanto, a maior basicidade do metóxido não é suficiente para retirar um hidrogênio das metilas equatoriais nos triésteres, mas resulta na torção dos grupos equatoriais nos TSs. A coordenação aos hidrogênios dos grupos metila estabiliza os TSs de reações de metóxido. Nos monoésteres, os TSs das reações de metóxido são estabilizados pela transferência intramolecular de H^+ e apenas um grupo equatorial é coordenado ao CH_3O^- .

As IRCs para todos TSs das reações $\text{D}_n + \text{A}_n$ são dominadas pela transferência intramolecular de H^+ . Para a reação de dissociação de MeOP^- , a distância P–O quebrada permanece constante na região do TS e as distâncias $d(\text{HX})$, onde X é o doador ou o aceptor de hidrogênio, sofrem mudanças de mais de 0.6 Å (figura 20). As reações de dissociação de MeSP^- e PhOP^- têm um comportamento semelhante. A transferência de H^+ no mecanismo $\text{D}_n + \text{A}_n$ pode ser comparada a transferência de elétrons, em que a

reação ocorre somente quando o “ambiente” (a ligação que se rompe, no caso das reações de fosfatos) assume uma configuração ideal, que permanece praticamente imóvel durante a transferência^[71].

A distância de ligação P–S em todos TSs para tiólise varia de 2.6 a 3.1 Å. A distância P–O varia de 2.2 a 2.4 Å (tabela 8) nos TSs de saída do grupo aril e, varia de 2.0 a 2.9 Å (tabela 8) nos TSs envolvendo reação de CH_3O^- , sugerindo que a transferência de H^+ influencia as distâncias P–OMe mais que nos outros grupos. A IRC das reações $\text{A}_n + \text{D}_n$ de monoésteres envolvendo CH_3SH e PhOH é exclusivamente composta pela formação ou pela quebra da ligação P–S ou P–O. Quando a reação envolve CH_3OH , a IRC é uma combinação da ligação P–O e da transferência de H^+ .

As frequências imaginárias de baixa magnitude observadas em todos TSs das reações de triésteres e em alguns TSs das reações de monoésteres sugerem que a PES na região destes TSs é plana (tabela 8). Frequências de maior magnitude são características das reações de monoésteres envolvendo metóxido e indicam PES com maior curvatura na região dos TSs. As frequências imaginárias acima de $1000i\text{ cm}^{-1}$ indicam barreiras com acentuada curvatura, associadas apenas à transferência de H^+ e características das reações $\text{D}_n + \text{A}_n$.

A PES é relativamente rugosa nas regiões de quebra e formação das ligações P–O e P–S e, portanto, o arranjo dos grupos equatoriais influenciou a otimização da geometria dos TSs (p. 85). Por exemplo, no cálculo do caminho da reação A, os ângulos diedrais responsáveis pelas torções dos grupos equatoriais combinavam-se à coordenada de reação. Pequenas variações nestes ângulos diedrais e na distância da ligação (P–S) em formação resultavam em grandes variações de energia, com diversos mínimos locais na PES (p. 85). Sugerimos que esta rugosidade era causada pela falta de flexibilidade *N*-eletrônica na expansão da função de onda e que uma descrição multiconfiguracional seria necessária no cálculo da PES em torno dos TSs de formação da ligação P–S^[211]. No entanto, investigações posteriores, incluindo cálculos multiconfiguracionais^{[234][121, cap. 12]}, população de orbitais naturais^[235], quebra de simetria da função de onda eletrônica irrestrita^[138], reotimizações de geometria e o cálculo de IRC com conjunto base ampliado indicaram que a rugosidade na PES era causada por falta de flexibilidade no conjunto base de 1-elétron, especialmente pela ausência de funções difusas (p. 53). Estas investigações também indicaram que uma configuração eletrônica (ou um determinante de Slater, §1.4.1) era suficiente para descrição da função de onda das espécies de reação estudadas. Portanto, o conjunto 6-31+G* foi usado nas otimizações de geometria para todas outras reações

estudadas. Também concluímos que a inclusão da correlação eletrônica, pelo menos em segunda ordem (§1.4.1), é essencial para descrição energética, pois barreiras de ativação calculadas no nível HF são até 18 kcal/mol superiores às barreiras calculadas no nível MP2^[74, 211].

As estruturas formadas ao longo do caminho das reações A a H apresentam os átomos ligantes coordenados ao fósforo nas posições de uma TBP (figuras 4, 13 e 16). O fósforo e os ligantes equatoriais são coplanares apenas nos intermediários (são exatamente coplanares nos C:I e G:Ia, pois essas reações são simétricas). As outras espécies formadas são TBPs distorcidas. Os mecanismos dissociativos e os IMCs das reações associativas de monoésteres (figura 19) são as únicas espécies cuja geometria otimizada não equivale a TBPs. O arranjo de TBP é característico dos intermediários propostos para reações de ésteres de fosfato^[37, 39, 45] e é observado experimentalmente em compostos de fósforo (V) hipervalente^[162] e pentacoordenado, como fosforanas (p. 21)^[39, 55].

Para alguns intermediários otimizados, o posicionamento dos grupos na TBP não segue as regras estabelecidas empiricamente (§1.2.1, p. 21)^[39, 53]. Segundo os valores sugeridos por Pearson^[46, tabela F.3], a eletronegatividade (χ) dos ligantes segue a ordem: $\chi_{\text{HO}} > \chi_{\text{CH}_3\text{O}} > \chi_{\text{PhO}} > \chi_{\text{CH}_3\text{S}}$. Os intermediários D:I, E:Ia, F:Ia, G:Ia e H:Ia possuem ligantes de menor eletronegatividade, como PhO^- e CH_3S^- , em posições axiais (contrariando o item 2, p. 21). O D:I contém o grupo fenóxido em posição axial (contrariando os itens 1, 2 e 3, p. 21). Na reação A, barreiras similares são observadas para dissociação de metóxido com o grupo CH_3S^- posicionado axial (A:TS3) ou equatorialmente (A:TS2, figura 13 e tabela 4), sugerindo que o posicionamento dos grupos não-reativos na TBP não influencia as alturas de barreiras de reação. Portanto, concluímos que as regras empíricas para posicionamento dos ligantes não são respeitadas para os intermediários de tiólise e alcoólise formados na fase gasosa e que o posicionamento dos grupos em relação a TBP não influencia as barreiras de ativação.

Barreiras energéticas envolvidas na pseudorrotação (§1.2) foram propostas como causa da dificuldade de formação (ou reação) de intermediários de ésteres de fosfato^[40, 53, 54]. Pseudorrotações são observadas nas reações associativas estudadas (por exemplo, na reação A, entre as espécies A:TS1 e A:I1, p. 86), mas não apresentam barreiras de ativação. Não procuramos por caminhos de reação nem TSs para interconversões de pseudorrotâmeros a partir dos intermediários. Todavia, se as alturas destas barreiras são iguais às diferenças de energia entre os pseudorrotâmeros, 3 kcal/mol para as reações de triésteres (p. 89), as barreiras serão facilmente transpostas. A barreira para a pseudorrotação do intermediário

pentacoordenado da reação de $\text{OH}^- + \text{TMP}$, obtida por cálculos de estrutura eletrônica, é menor que 1 kcal/mol^[212]. Concluimos que a interconversão de pseudorrotâmeros não é um processo mecanisticamente importante e não deve dificultar a formação ou a reação das fosforanas na tiólise e alcoólise.

A única fosforana que pode ser observada experimentalmente é C:I, pois a barreira para sua decomposição é alta e suas energias de formação são exotérmicas (figura 17). Aparentemente, a metodologia empregada por Lum e Grabowski^[40], no estudo da reação de TMP com diversos nucleófilos em fase gasosa, permite a detecção de intermediários aniônicos, mas a observação de uma fosforana não foi reportada. Estes autores sugerem que fosforanas não são formadas e que a etapa determinante da velocidade de reação é a adição do nucleófilo ao TMP^[40] já que o composto deuterado, $\text{C}_2\text{D}_3\text{H}_3\text{O}_4\text{P}^-$, que indica ataque de CD_3O^- sobre fósforo do TMP, foi detectado em pequenas quantidades (o ataque sobre fósforo é observado apenas como traços de reação, cerca de 1.5% do total). Lum e Grabowski argumentam que a formação de uma fosforana resultaria em maior incorporação de CD_3O^- nos produtos, porque o intermediário poderia se decompor para TMP-d₃ e ser novamente atacado por CD_3O^- , gerando $\text{C}_2\text{D}_6\text{O}_4\text{P}^-$, que também não foi detectado. O perfil energético calculado para a reação C sugere outra interpretação destes resultados experimentais. A barreira para a reação total corresponde à decomposição da fosforana. A adição de CD_3O^- a TMP é cerca de 10^7 vezes mais rápida do que a dissociação do grupo de saída (figura 15). A baixa quantidade de composto deuterado detectada experimentalmente deve-se à maior barreira total para reação sobre fósforo em relação à reação sobre o carbono do TMP e ao ataque sobre carbono ser termodinamicamente favorável, ao contrário do ataque sobre fósforo^[212]. O produto deuterado duplamente substituído não é observado, porque a barreira para reação no fósforo tem que ser transposta *duas* vezes, sem que nenhuma outra reação concorrente (ataque sobre o carbono, ou dissociação do CD_3O^-) ocorra. Esta hipótese é energética e estatisticamente desfavorável e a quantidade de $\text{C}_2\text{D}_6\text{O}_4\text{P}^-$ formada deve ser inferior ao limite de detecção experimental (0.2% do total do produto de reação).

Os produtos formados na tiólise de triésteres são mais instáveis do que os produtos de alcoólise e, logo, as energias de reação de tiólise são maiores do que as energias de alcoólise (figuras 15 e 17). A diferença da energia de reação entre tiólise e alcoólise para o mesmo grupo de saída (reações A vs. C e B vs. D) é cerca de 35 kcal/mol. A diferença entre as reações com o mesmo nucleófilo, mas grupos de saída distintos (reações A vs. B e C vs. D), também é cerca de 35 kcal/mol. As reações e formações de intermediários nos monoésteres são endotérmicas e a faixa de variação das energias é restrita. Os inter-

mediários formados no ataque dos triésteres arílicos são cerca de 10 kcal/mol mais estáveis do que os intermediários alquílicos (A:I1 vs. B:I e C:I vs. D:I).

A energia de dissociação de H^+ de fenol em fase gasosa é cerca de 30 kcal/mol mais baixa do que a dissociação do metanol. Portanto, o fenóxido deve ser um melhor nucleófilo que metóxido e as energias de reação cujo grupo de saída é fenóxido são mais favoráveis. A basicidade do fenóxido é menor por causa da deslocalização da carga negativa pelo anel aromático. Para as reações de monoésteres, a transferência de H^+ do grupo fosfato neutraliza a carga formada no nucleófilo e, portanto, as energias das reações cujos grupo de saída são metanol ou fenol são semelhantes. A protonação do grupo de saída por um ácido geral é um dos mecanismos utilizados pelas PTPs para catalisar a desfosforilação do substrato^[42].

Portanto, a estabilidade termodinâmica dos triésteres depende da natureza dos grupos nucleofílicos e de saída. A desesterificação e protonação resultam em considerável independência da estabilidade termodinâmica dos monoésteres de fosfato em relação à natureza dos grupos reativos. As energias de reação e de formação dos intermediários das reações E e H são menores do que 7 kcal/mol. Nas reações $D_n + A_n$, a etapa dissociativa é sempre endoérgica e a etapa associativa é sempre exoérgica. Ao contrário dos triésteres, as altas energias de formação de intermediários sugerem que mecanismos divididos em duas etapas não são favorecidos para as reações de monoésteres. A barreira para decomposição de G:Ia é 7.1 kcal/mol. Assim como C:I, este intermediário também pode ser detectado experimentalmente, caso o mecanismo operacional na fase gasosa seja associativo (vide infra). Estes intermediários correspondem a reações identidade de alcoólise. Os outros intermediários formados nas reação associativas de monoésteres têm barreiras de decomposição menores do que 2 kcal/mol e, portanto, são instáveis na fase gasosa.

As alturas de barreira calculadas para as reações de monoésteres diferem consideravelmente das barreiras de triésteres. A reação com menor barreira é a alcoólise de DMPP (reação D, figura 17). Esta reação deve ser observada experimentalmente já que a barreira para alcoólise sobre carbono alquílico pode ser estimada em cerca de 8.5 kcal/mol, pela diferença entre a energia relativa do TS de alcoólise sobre o carbono na reação de $CH_3O^- + TMP$, calculada por Chang e Lim^[212] (neste trabalho o C:TS1 é denominado TS-6, com $\Delta E_T = \Delta E_{MP2} + \Delta ZPE = -12.7$ kcal/mol, mas a ΔE_T do IMC não é reportada), e a energia do C:IMC1 (o IMC formado para ataque sobre fósforo é o mesmo daquele formado quando o ataque ocorre sobre o carbono). A reação de tiólise de DMPP também apresenta uma barreira transponível (figura 17) e deve ser observada experimen-

talmente. A barreira para ataque de CH_3S^- sobre carbono alquílico foi calculada em 12.6 kcal/mol^[211] e, portanto, deve ser competitiva com a tiólise sobre fósforo do DMPP (reação B, figura 15). O ataque nucleofílico sobre o carbono aromático é improvável em ambas reações. Portanto, sugerimos que quantidades significativas de ataque sobre fósforo devem ser observadas na fase gasosa, para reação de triésteres de fosfato com um grupo de saída aril-derivado, como DMPP, com nucleófilos como CH_3O^- e CH_3S^- , ao contrário da reatividade de triésteres de alquila (§1.2.2).

Nenhum dos estudos experimentais de ataque nucleofílico sobre TMP em fase gasosa apresentou constantes de velocidade para ataque sobre fósforo (§1.2.2)^[40, 41, 81]. O único dado cinético experimental que pode ser comparado com as alturas de barreira calculadas nesta tese é uma relação entre as constantes de velocidade para ataque sobre carbono e sobre fósforo da reação $\text{CD}_3\text{O}^- + \text{TMP}$. Esta relação corresponde a uma diferença de 2.2 kcal/mol entre as barreiras, onde a reação sobre carbono é mais rápida^[40]. A barreira estimada para ataque de CH_3O^- sobre carbono do TMP é 8.5 kcal/mol (vide supra) e a barreira calculada para ataque sobre fósforo é 14.0 kcal/mol (figura 15). Assim, a diferença entre as barreiras calculadas é 5.5 kcal/mol, isto é, 3.3 kcal/mol acima da diferença entre as barreiras experimentais. A proximidade das diferenças de barreira calculada e experimental demonstra que as computações efetuadas tem capacidade preditiva, pelo menos, semiquantitativa. As diferenças de altura de barreira não são mais próximas em virtude das incertezas possivelmente associadas às seguintes fontes:

- Medida experimental, pois a observação do produto de ataque sobre fósforo foi pequena e próxima do limite de detecção^[40];
- Suposição que o mecanismo de reação é A_nD_n , para calcular a relação entre as constantes a partir das distribuições de produtos observadas^[40];
- Nível de cálculo de estrutura eletrônica (MP2/6-31+G*//HF/6-31+G*) usado na determinação da energia relativa do TS de ataque sobre carbono^[212].

A comparação entre os valores calculados e experimentais é resultado da diferença entre dois valores, computados ou medidos, e, portanto, o erro final é a soma dos erros de cada valor. O erro associado ao nível de cálculo usado na computação da altura de barreira de ataque sobre fósforo (§2.2.4 e §2.3.1) deve ser menor do que as três possíveis fontes de erro mencionadas acima.

A constante de velocidade de alcoólise intramolecular de um aril diéster de fosfato em solução aquosa é cerca de 10^7 vezes maior do que a tiólise intramolecular de um substrato

equivalente. Esta diferença corresponde a um $\Delta\Delta G^\ddagger = 9.5$ kcal/mol a 25°C^[76]. O efeito da solvatação aquosa na diferença de energia relativa ($\Delta\Delta G$) entre estas reações deve ser cancelado em grande parte, já que a reação é intramolecular e a estrutura do solvente em torno do álcool e do tio derivado deve ser similar. Observamos um bom acordo entre a diferença das alturas de barreira das reações B e D, 7.7 kcal/mol, e a diferença entre as energias de ativação experimentais. Os mecanismos de reação de diésteres de fosfato são associativos e similares aos mecanismos de triésteres (§1.2).

A reação de $\text{HS}^- + \text{TMP}$ medida em fase gasosa é totalmente regioseletiva para carbono e o ataque sobre fósforo não é observado^[40]. O perfil energético calculado para a reação A (figura 15) sugere que a reação de tiólise sobre fósforo não é observada, porque a reação é termodinamicamente desfavorável ($\Delta G_{\text{reac}} > 30$ kcal/mol), apresenta um alta barreira para eliminação do grupo de saída e uma pequena barreira para decomposição do intermediário na direção dos reagentes (4.2 kcal/mol, figura 15). Lum e Grabowski^[40], e Chang e Lim^[212] propõem que o passo limitante para o ataque ao centro de fósforo em triésteres é a adição do nucleófilo. Os resultados apresentados na §2.2.1 mostram que a etapa limitante é o ataque sobre fósforo apenas para reação B e a etapa de dissociação do grupo de saída é limitante nas outras reações de triésteres. Portanto, parece-nos que, para as reações de triésteres em geral, a etapa de dissociação do grupo de saída é determinante da velocidade de reação.

As energias de descomplexação dos IMCs dos produtos nas reações B, C e D são maiores que as barreiras associadas aos TS1 e TS2 de cada reação. Embora as energias de complexação estejam superestimadas em até $\frac{1}{8}$ dos valores, em consequência do BSSE^[140] (p. 54), a acumulação dos IMCs dos produtos ou a reação inversa (ataque por metóxido nas reações A e C, e por fenóxido nas reações B e D) podem ser observadas em fase gasosa.

Dois tipos de mecanismos foram calculados para as reações de monoésteres (figura 3). As diferenças entre as alturas de barreira observadas no mecanismo $A_n + D_n$ e no $D_n + A_n$ não ultrapassam 7 kcal/mol, assim, os dois mecanismos podem ser competitivos. Para a reação E ($D_n + A_n$), a barreira de ativação total é 7 kcal/mol mais baixa que a barreira para o mecanismo associativo. A altura de barreira total da reação F ($A_n + D_n$) é 28.7 kcal/mol. A dissociação de PhOP^- apresenta uma barreira de 29.4 kcal/mol e, assim, a formação do F:IMC1d pode ser competitiva com a formação do F:IMC2a (figura 21 e tabela 9). Todavia, o mecanismo $A_n + D_n$ deve ser preferido, porque a energia de formação do F:Id é 31.2 kcal/mol (6.8 kcal/mol superior a energia de formação do F:Ia) e a altura de barreira total associativa é 4.5 kcal/mol mais baixa que a barreira $D_n + A_n$. Para a

reação G, a barreira para formação do G:Ia é 5 kcal/mol mais alta do que a barreira de dissociação do MeOP^- (que corresponde a etapa determinante do mecanismo D_n+A_n). Portanto, o mecanismo A_n+D_n e o G:Ia não devem ser observados experimentalmente em fase gasosa (figura 23 e tabela 11). Os mecanismos A_n+D_n e D_n+A_n são competitivos para reação H em fase gasosa já que as barreiras totais são próximas (figura 24 e tabela 12).

A etapa determinante da velocidade das reações de monoésteres é a adição do nucleófilo para a reação H, e a dissociação do grupo de saída nas reações E, F e G. Ao contrário do que é observado em solução aquosa (§1.2.1)^[37, 76] e calculado para as reações de triésteres (§2.2.1), a barreira para etapa de tiólise associativa nos monoésteres é cerca de 8 kcal/mol mais baixa do que a barreira para alcoólise e o ataque de CH_3SH sobre metafosfato tem barreira 4 kcal/mol mais baixa do que o ataque de CH_3OH . Ou seja, na fase gasosa, nucleófilos centrados em enxofre são mais reativos do que os nucleófilos centrados em oxigênio no ataque sobre monoésteres de fosfato. Esta diferença deve-se a maior basicidade do metóxido, desta vez em relação ao tiolato. Na alcoólise associativa de monoésteres, a transferência de H^+ e a formação da ligação P–O contribuem para IRC na região dos TSs, modificando o perfil energético em relação à alcoólise de triésteres, cuja IRC é composta majoritariamente pela ligação P–O e nenhuma transferência de H^+ é observada. Na tiólise associativa de monoésteres, a transferência de H^+ do CH_3SH para o grupo fosfato ocorre antes da formação do TS, assim, o nucleófilo está ionizado e o fosfato neutralizado e, portanto, o ataque nucleofílico tem uma barreira de ativação menor que a alcoólise. Nas reações D_n+A_n , a transferência de H^+ domina a IRC para alcoólise e para a tiólise, porém, a maior basicidade do metóxido implica que a transferência do H^+ desta espécie requer mais energia que a transferência do H^+ do tiol

Para reação E, os E:TS1a e o E:Ia constituem pontos estacionários na PES Born-Oppenheimer, mas a adição da *ZPE* resulta num TS com energia total inferior ao intermediário (tabela 9). Portanto, o E:Ia não é estável e o mecanismo associativo é simultâneo (A_nD_n , figura 3). O TS é associativo, com a ligação P–S praticamente formada [$d(\text{PS}) = 2.20 \text{ \AA}$ no E:TS2a e $2.14 \text{ \AA}$ no produto MeSP^-]. As pequenas barreiras para decomposição e, conseqüentemente, o reduzido tempo de vida dos intermediários formados nas reações D, E, F e H pelo mecanismo A_n+D_n sugerem que a cinética destas reações é equivalente a um mecanismo simultâneo A_nD_n . Observáveis experimentais, como LFER e KIE, obtidas para reações em solução de di e triésteres de fosfato com grupos de saída ativados (por exemplo, fenol-derivados como nas reações D, F e H) são tradicionalmente interpretadas como evidência para mecanismos A_nD_n , com formação “parcial” da ligação

entre nucleófilo e o fosfato no TS (§1.2.1).

Quais são, então, os fatores que determinam as alturas de barreiras para a alcoólise e tiólise intrínsecas de ésteres de fosfato? Triésteres reagem por mecanismos A_n+D_n e a formação de um intermediário dissociativo tipo “dimetil metafosfato” é energeticamente proibida. Portanto, o principal fator que determina a reatividade é o estado de alquilação. Monoésteres, no entanto, não reagem obrigatoriamente por mecanismos dissociativos como tradicionalmente aceito^[37–39, 45]. Florián e Warshel propuseram, a partir de cálculos de estrutura eletrônica^[72, 73], que monoésteres também podem reagir por um mecanismo associativo. A basicidade do nucleófilo e do grupo de saída é o segundo fator que influencia o mecanismo. Logo, a interpretação de LFER de reações de monoésteres em solução deve ser cautelosa (§1.2.1). As altas barreiras observadas para as etapas envolvendo ataque ou saída de CH_3OH estão sempre relacionadas ao custo energético para transferência de H^+ , em razão da elevada basicidade do metóxido. Assim, o mecanismo observado nas reações de monoésteres com ataque ou dissociação de CH_3OH é aquele que minimiza o “gasto” energético com a transferência de H^+ . Não obstante, a reação F, que não envolve a reação de CH_3OH , apresenta a menor barreira de ativação entre as reações de monoésteres. Resultados de experimentos cinéticos para hidrólise de monoésteres di-aniônicos em diferentes solventes^[62, 67] também indicam que o mecanismo de reação pode depender da basicidade do grupo de saída (p. 25) e da participação de uma possível transferência de H^+ intramolecular^[62, 68].

Todas reações de monoésteres calculadas (§2.2.2) contêm duas transferências de H^+ ao longo do caminho de reação. Estas transferências alteram significativamente a termodinâmica e a cinética das reações de monoésteres, em comparação às reações de triésteres (vide supra). Em solução, a presença de um catalisador (por exemplo, uma molécula de água) para transferência de H^+ resulta na diminuição de cerca de 10 kcal/mol da barreira calculada para dissociação de MeOP^- em PO_3^- e CH_3OH (§1.2.1)^[68, 74]. A molécula de água forma um anel de 6 membros (figura 5), que facilita a transferência de H^+ . Esta estabilização não é observada para os mecanismos associativos na hidrólise de MeOP^- (§1.2.1). Portanto, o mecanismo D_n+A_n , que já é favorecido para alcoólise e tiólise de MeOP^- na fase gasosa, deve ser observado também em solução aquosa^[68, 74].

A barreira energética para o ataque de água sobre metafosfato (formando fosfato inorgânico, H_2PO_4^-) em solução aquosa, foi calculada entre 4.5 e 5.5 kcal/mol. Este valor deve ser adicionado a -2.4 kcal/mol para corrigir a diferença (entrópica) entre o estado padrão de 1M, em que o cálculo é efetuado^[68, 74], e a concentração de água líquida,

55M, resultando numa barreira estimada, em energia livre, entre 2.1 e 3.1 kcal/mol. A barreira de ataque de CH_3OH sobre metafosfato (formando MeOP^-) calculada é 6 a 8 kcal/mol^[68, 74] (compare com a barreira de 16.9 kcal/mol observada na fase gasosa, figuras 23 e 24) e a barreira para ataque de CH_3SH (formando MeSP^-) pode ser estimada em 7 kcal/mol (baseada na barreira em fase gasosa, 13 kcal/mol, figuras 21 e 22, e na menor energia de solvatação das espécies contendo CH_3SH , em relação àsquelas contendo CH_3OH , vide infra). Portanto, a adição de água ao metafosfato é mais rápida que a adição de CH_3OH ou de CH_3SH e que a difusão do PO_3^- em solução. (A barreira para difusão desta espécie pode ser estimada em 3.2 kcal/mol com base na constante de difusão $k_d = 10^{10} \text{ s}^{-1}$). Assim, alcoólise e tiólise de monoésteres de fosfato não devem ocorrer em fase aquosa. O metafosfato formado é rápida e seletivamente capturado por uma molécula de água antes de escapar da esfera de solvatação^[68].

No entanto, especulamos que a tiólise de monoésteres de fosfato será observada se a reação for conduzida num meio não-aquoso, com solvente pouco nucleofílico e impedido estericamente, como por exemplo, *t*-butanol. Neste solvente, a reação de um monoéster de fosfato quiral resulta em racemização dos produtos formados^[66], sugerindo que o metafosfato é formado e tem um tempo de vida suficientemente grande para difundir e se equilibrar pelo solvente. A presença de um melhor nucleófilo na solução, como CH_3SH , resultaria em ataque deste sobre o PO_3^- .

A metodologia empregada por Bianciotto, Barthelat e Vigroux^[68], e Hu e Brinck^[74] para o cálculo dos caminhos de reação e barreiras para hidrólise de metilfosfato em fase aquosa é uma combinação de determinação de estrutura eletrônica molecular e de uma representação do solvente aquoso por um dielétrico contínuo^[236, 237]. As moléculas explicitamente usadas por estes autores como reagentes não consistem apenas das moléculas reativas, mas também de moléculas de solvente (água) adicionais, que respondem pelos efeitos específicos negligenciados pelos modelos dielétricos^[236, 237]. Esta metodologia é chamada de solvatação de “supermolécula”. Os efeitos específicos se manifestam na catálise da transferência de H^+ efetuada pela molécula de água explícita. A negligência destes efeitos, como nos cálculos de Warshel et al.^[70, 72, 73], resulta em barreiras calculadas cerca de 10 kcal/mol mais altas que as observações experimentais.

A comparação dos perfis de energia calculados na §2.2 com medidas de reatividade de ésteres de fosfato em solução requer a inclusão de efeitos de solvatação. Para reação de triésteres, em que não ocorre transferência de H^+ , supomos que efeitos específicos podem ser negligenciados e a solvatação pode ser modelada apenas por um dielétrico^[182, 236]. Mas,

para as reações de monoésteres, a necessidade da metodologia supermolecular é clara^[68, 74]. Não efetuamos computações do efeito de solvatação, porque os perfis de reação de ésteres de fosfato levantados nesta tese serão utilizados para calibração da metodologia usada nas simulações de reações enzimáticas (cap. 3). Neste caso os resultados da reatividade intrínseca, livre de efeitos de solvatação, são necessários.

No entanto, podemos estimar que espécies com densidade de carga localizada, como reagentes, produtos, IMCs e intermediários, terão energias de solvatação maiores (em módulo) do que espécies com densidade de carga deslocalizada, como TSs. Portanto, as diferenças nas energias de solvatação podem contribuir para elevar as alturas de barreira das reações e estabilizar a formação de fosforanas. Grupos ou compostos contendo metóxido são menos polarizáveis do que compostos contendo tiolato e fenóxido e, logo, também terão energias de solvatação maiores (em módulo). Assim, em solução aquosa, as reações cujo grupo de saída é o metóxido (ou metanol) devem ser mais exoérgicas e ter alturas de barreira abaixadas em relação às reações em que o grupo de saída é fenóxido (ou fenol). Da mesma forma, as reações de tiólise em solução aquosa devem ter alturas de barreira aumentadas em comparação às reações de alcoólise.

2.3.3 Modelos para os mecanismos enzimáticos

Mono e triésteres de fosfato foram usados como modelos das reações catalisadas por PTPs, porque estes eram os compostos mais próximos dos substratos selvagens das enzimas e, porque poderiam ter suas estruturas eletrônicas e energias determinadas com boa acurácia. Embora os substratos enzimáticos sejam monoésteres, a reação destas espécies é influenciada pela transferência intramolecular de H^+ (§2.2 e §2.3), assim, as reações de triésteres são modelos para descrever a quebra e a formação intrínsecas das ligações P–O e P–S, livre de efeitos de protonação. A alquilação dos oxigênios equatoriais em triésteres também pode mimetizar o efeito da coordenação dos oxigênios equatoriais do substrato com a cadeia lateral da ARG e com os hidrogênios (ligados ao nitrogênio) da cadeia protéica principal no MC das PTPs (figura 10)^[51]. Ésteres com grupos de saída aril e alquil derivados foram incluídos para modelar as reações das PTPs com fosfotirosina e das DSPs com fosfoserina e fosfotreonina. Além do estudo das PTPs selvagens, que utilizam um resíduo CYS para ataque nucleofílico ao substrato (formação da ligação P–S, §1.3), também modelamos a reação catalisada pelos mutantes CYS→SER, em que o ataque nucleofílico envolveria uma alcoólise do éster de fosfato.

Embora a reação entre um nucleófilo ionizado (CH_3S^- ou CH_3O^-) e um monoéster

de fosfato aniônico (mono ou duplamente carregado) seja um modelo mais fiel ao estado de protonação e ao mecanismo de catálise tradicionalmente propostos para atividade das PTPs (figura 11 e §1.3), a determinação da estrutura eletrônica de derivados de fosfato di ou trianiônicos na fase gasosa sugere que estas espécies são instáveis e podem sofrer autoionização eletrônica (p. 149)^[238]. Portanto, escolhemos as reações de tiólise e alcoólise de tri e monoésteres com carga total igual a -1 , como modelos cuja estrutura eletrônica pode ser determinada com acurácia na fase gasosa.

As energias de ativação ($\Delta G_{enz}^\ddagger$) para reação catalisada pelas PTPs variam entre 14 e 18 kcal/mol para os diversos membros da superfamília (§1.3). O mecanismo de catálise ocorre em duas etapas (reações 1.9 e 1.10). Já a reação em solução ocorre geralmente em uma única etapa de hidrólise A_nD_n (§1.2.1). Portanto, não podemos analisar a enzima em termos de sua afinidade pelo TS da reação em solução (§1.1)^[239]. A primeira etapa na catálise (reação 1.9) pode ser comparada à tiólise de monoéster em solução, mas esta reação nunca foi observada (p. 30). A hidrólise não catalisada de MeOP^{2-} e PhOP^{2-} (metilfosfato e fenilfosfato dianiônicos) em solução aquosa, fortemente básica, tem $\Delta G_{sol}^\ddagger = 43$ e 36 kcal/mol a 39°C, respectivamente^[1].

Nos próximos parágrafos, as contribuições e os processos envolvidos na primeira reação catalítica (1.9) serão dissecados e comparados com os perfis de energia calculados.

A repulsão eletrostática presente na barreira de ataque do nucleófilo ionizado ao substrato carregado negativamente (figura 11) pode ser estimada aproximadamente, em solução aquosa, em 5 kcal/mol [$\Delta\Delta G^\ddagger(Q_P = 0 \rightarrow Q_P = -1) \approx 5$ kcal/mol, onde Q_P é a carga formal do grupo fosfato]^[52], a partir da diferença entre as energias de ativação medidas para ataque de OH^- sobre um triéster (neutro e cíclico) e sobre um diéster monoaniônico. Assumindo que a contribuição eletrostática é linearmente aditiva, a repulsão entre um nucleófilo ionizado e um substrato dianiônico (figura 11) contribuiria com $\Delta\Delta G^\ddagger(Q_P = 0 \rightarrow Q_P = -2) \approx 10$ kcal/mol para altura da barreira de reação em solução. As PTPs podem estabilizar esta repulsão eletrostática graças às cargas formais positivas (por exemplo, a cadeia lateral da ARG conservada) e ao potencial eletrostático positivo^[240] observados no sítio ativo (este potencial foi calculado por métodos aproximados^[145] e, portanto, deve ser analisado qualitativamente).

A barreira intrínseca para ataque de CH_3S^- sobre fosfatos neutros é representada pelas barreiras da primeira etapa das reações A e B (com alturas respectivamente de 16.0 e 10.7 kcal/mol), porque as reações de monoésteres são influenciadas pela transferência intramolecular de H^+ (§2.3.2). Portanto, a formação da ligação P–S no processo de catálise

das PTPs não necessita de estabilização, já que as barreiras intrínsecas são menores ou similares a energia de ativação observada experimentalmente.

A formação de um intermediário fosforana ou metafosfato durante a catálise necessita de estabilização, independentemente do mecanismo adotado na enzima ($A_n + D_n$ ou $D_n + A_n$), já que as energias de formação destas espécies na tiólise de monoésteres varia entre 24.4 a 31.2 kcal/mol (reações E e F, figuras 21 e 22).

Se o mecanismo $D_n + A_n$ é adotado pela PTP, a dissociação do grupo de saída, cujas barreiras são cerca de 30 kcal/mol (reações E e F), necessita de estabilização, possivelmente por uma transferência de H^+ proveniente de um ácido geral da enzima. Uma estimativa da contribuição da transferência de H^+ para estabilização do grupo de saída não pode ser facilmente obtida a partir das barreiras das etapas dissociativas das reações E e F $D_n + A_n$, já que a transferência de H^+ nestas reações é intramolecular e a protonação estabiliza o grupo de saída (metanol ou fenol), mas deve desestabilizar o metafosfato.

Se o mecanismo enzimático é associativo, a barreira para dissociação do grupo de saída do intermediário pentacoordenado não será limitante caso o grupo esteja parcialmente protonado (segunda etapa das reações E e F $A_n + D_n$) ou caso o grupo seja fenóxido (segunda etapa das reações A e B). Se o grupo de saída for metóxido, alguma estabilização será necessária, provavelmente via protonação, já que a barreira para saída de metóxido de uma fosforana é cerca de 20 kcal/mol (reação A).

Em resumo, as reações modelo de tiólise de mono e triésteres de fosfato calculadas permitem-nos especular que os processos que necessitam de estabilização na primeira reação (1.9) da atividade das PTPs são:

- Repulsão eletrostática entre o nucleófilo ionizado e o substrato aniônico;
- Formação do intermediário de tiólise, seja este um metafosfato ou uma fosforana;
- Dissociação do grupo de saída caso este seja um alquil derivado e não esteja protonado.

Ao contrário do que é sugerido pelas regras empíricas de posicionamento dos ligantes coordenados a uma TBP (p. 21), não existe gasto energético para posicionar um grupo menos eletrofílico, como o tiolato, em posição axial (p. 119). Portanto, as PTPs não precisam estabilizar o arranjo em linha no ataque nucleofílico ou na possível formação de uma fosforana.

Também podemos especular quais são as possíveis causas da inatividade das PTPs mutantes CYS→SER (§1.3.3). O oxigênio da cadeia lateral da SER mutante pode se posicionar desfavoravelmente para o ataque nucleofílico sobre o substrato fosforilado, localizando-se a uma distância inapropriada do fósforo do substrato [$d(\text{PS})=2.14 \text{ \AA}$ no MeSP^- e $d(\text{PO})=1.68 \text{ \AA}$ no MeOP^-] ou coordenando-se a outros centros, impedindo o ataque nucleofílico^[111]. As distâncias da ligação em formação no TS associativo das reações de triésteres variam em torno de $2.8 \text{ \AA}$ e são quase independentes da natureza do nucleófilo (tabela 8), assim, a distância da ligação P–O num possível TS na PTP mutante seria similar a distância P–S do TS na PTP selvagem. No entanto, a $d(\text{PO})$ num possível TS na PTP mutante pode estar aumentada, pois a ligação C–O na cadeia lateral da serina é $0.4 \text{ \AA}$ mais curta do que ligação C–S da cisteína.

A maior basicidade do grupo metóxido resulta na torção dos grupos equatoriais nas reações de alcoólise dos triésteres (vide supra) e, portanto, os mutantes CYS→SER podem ter o sítio ativo, em especial os grupos coordenados ao nucleófilo, como a cadeia lateral da SER conservada no P-loop e os hidrogênios da cadeia principal do P-loop, torcidos ou mal posicionados para catálise^[111]. A PTP foi otimizada pela evolução para catalisar a tiólise de fosfato e, portanto, os grupos catalíticos coordenam-se para estabilização do TS desta reação e não para reação de alcoólise.

Mutantes CYS→SER são usados para aprisionar o substrato natural de PTPs, já que a mutação não impede a formação do MC, mas inativa a desfosforilação do substrato (p. 39). Uma vez que a ativação da ligação P–O na etapa de dissociação do grupo de saída é pouco influenciada pela natureza do grupo nucleofílico (p. 102), sugerimos que a inatividade dos mutantes CYS→SER está relacionada à etapa de ataque nucleofílico, qualquer que seja o mecanismo empregado nas PTPs.

Em razão da transferência de H^+ nas reações de monoésteres, a reatividade dos alquil e arilfosfatos nas reações E a H é semelhante. No entanto, os perfis de energia das reações de triésteres de alquil e arilfosfatos apresentam diferenças de até 30 kcal/mol , além de diferenças de até $0.5 \text{ \AA}$ na distância da ligação P–O que é quebrada (tabela 8). Portanto, supomos que as DSPs são capazes de catalisar reações com características estruturais e perfis energéticos razoavelmente distintos. Os mecanismos utilizados por estas enzimas para catálise serão apresentados no capítulo 4.

3 *Calibração da hamiltoniana híbrida*

Aproximações são necessárias na modelagem de sistemas complexos tanto para facilitar a compreensão dos fenômenos quanto para reduzir o custo computacional e viabilizar o cálculo ou a simulação destes sistemas. Naturalmente, as aproximações introduzem incertezas e, portanto, a metodologia empregada deve ser devidamente testada e, se necessário, calibrada para assegurar que os resultados dos cálculos e das simulações tenham sentido físico e possam ser interpretados, comparados ou discutidos à luz de observações experimentais.

Neste capítulo, apresentamos testes e calibrações do potencial híbrido que foi usado nas simulações de reações catalisadas pelas PTPs.

Rossi e Truhlar^[157] sugeriram que uma parametrização mais acurada de métodos semiempíricos (§1.4.2) pode ser obtida se o conjunto de treinamento (p. 58) for composto por espécies ao longo do caminho da reação (ou uma família de reações) de interesse e se as propriedades destas espécies, usadas como valores de referência, forem computadas por métodos de estrutura eletrônica *ab initio* de alto nível. Utilizando GAs (§1.5)^[155, 156] para otimização dos parâmetros, Rossi e Truhlar obtiveram uma parametrização calibrada especificamente para a reação de $\text{Cl} + \text{CH}_4$ que reproduz os resultados do cálculo *ab initio* com apenas uma fração do custo computacional.

Esta metodologia foi usada (§3.1.2) para calibração da hamiltoniana semiempírica especificamente para reações de ésteres de fosfato (§3.2.1), porque os elevados desvios observados nas energias relativas para espécies destas reações calculadas com as hamiltonianas originais^[143–145] (§1.4.2) impedem que conclusões, pelo menos semiquantitativas, sejam tiradas dos resultados de cálculos de estrutura eletrônica ou de simulações com potenciais híbridos que empregam tais parametrizações originais.

Já que a interação QM/MM eletrostática usada no potencial híbrido não é definida unicamente se a região quântica for descrita por uma hamiltoniana semiempírica, as propostas de Field et al.^[187] e do grupo de Nancy^[190, 194] (§1.6.4.1) foram testadas (§3.2.3).

Uma vez que a forma de interação eletrostática é fixada, a interação QM/MM total pode ser calibrada, otimizando-se os parâmetros semiempíricos designados aos centros MM (p. 77). Esta alternativa é trabalhosa, porque muitos parâmetros semiempíricos seriam atribuídos aos tipos atômicos definidos para cada campo de força, e, portanto, é contrária aos objetivos práticos adotados na construção dos potenciais híbridos (p. 77). A interação QM/MM também pode ser calibrada se as cargas q_m (equações 1.64 ou 1.81) ou os parâmetros de LJ dos centros MM (equação 1.65), definidos pelo campo de força clássico, forem alterados. Esta alternativa também é trabalhosa e as interações MM puras seriam alteradas, resultando no desbalanceamento das forças nesta região. Um dos principais objetivos na construção e calibração da interação QM/MM em potenciais híbridos é a conservação da consistência entre as forças, observado usualmente para potenciais QM ou MM puros (p. 77).

A alternativa normalmente adotada para calibração da interação QM/MM é a otimização dos parâmetros de LJ atribuídos aos centros QM (p. 75). Bash et al.^[167] estenderam a metodologia de Rossi e Truhlar para otimização destes parâmetros. Neste caso, as estruturas e as propriedades de complexos intermoleculares entre modelos do “solvente” e os reagentes e produtos são determinadas *ab initio* e usadas como valores de referência. Assim, parâmetros específicos para uma reação em um determinado meio (por exemplo, solução aquosa^[167] ou um sítio enzimático^[241]) podem ser obtidos, viabilizando simulações mais precisas de sistemas complexos por potenciais híbridos^[167].

Esta metodologia foi usada (§3.1.3) para calibração dos parâmetros de LJ atribuídos aos centros QM e espera-se que parte dos erros introduzidos pelas aproximações semiempíricas das equações 1.83 e 1.84 possam ser compensados por esta calibração.

Nos complexos intermoleculares, um molécula foi descrita pela parte QM do potencial híbrido e a outra, pela parte MM. Assim, não existe ligação covalente entre as duas regiões e apenas interações não ligantes (equações 1.58 e 1.81) são consideradas. A presença de uma ligação covalente na fronteira entre as regiões QM e MM (por exemplo, a ligação C–C na figura 12) implica que a definição completa do potencial híbrido ainda dependerá de como a densidade eletrônica QM será terminada e de como as interações entre as duas regiões serão tratadas na fronteira (§1.6.4.2). A definição destas interações segundo Field et al.^[179] e o esquema de *átomo fictício* para terminação da região QM^[187] foram adotados nas simulações enzimáticas (cap. 4). Nenhuma modificação ou calibração específica destas aproximações foi usada nesta tese, porém, a influência nos resultados obtidos foi testada (§3.1.4 e §3.2.4).

O algoritmo GA+simplex (§3.1.1) substituiu o GA, utilizado anteriormente na otimização dos parâmetros semiempíricos e da interação QM/MM^[157, 167].

3.1 Métodos

3.1.1 Algoritmo de otimização: GA+simplex

Diferentes métodos para otimização de parâmetros foram apresentados na §1.5. Aqui, utilizamos uma combinação seqüencial de GA e simplex (GA+simplex)^[242]. O código fonte do GA escrito por Carroll^[243] e o simplex de Nelder e Mead^[154, §10.4], ambos em FORTRAN 77^[244], foram usados.

O procedimento de otimização iniciou-se com uma população gerada randomicamente pelo GA, dentro de uma faixa de variação fixada para cada um dos K parâmetros otimizados. Tipicamente, as seguintes opções do GA foram usadas^[155, 156, 243]:

- População de $2K$ indivíduos;
- Cada cromossomo representado por 14 a 20 dígitos binários;
- Probabilidades de mutação normal e “creep”^[155, 156] fixadas em 0.08;
- Elitismo e nicho (que preserva a diversidade da população)^[155, 156] ativados;
- Cruzamento por simples ponto, com probabilidade fixada em 0.8;
- Semente randômica igual a -1000 (para gerar a população inicial);

A população evoluiu pelo GA durante um número fixo de gerações e os $K + 1$ indivíduos mais adaptados da última geração foram usados como pontos (vértices) iniciais no simplex. O indivíduo mais adaptado gerado durante toda evolução do GA esteve necessariamente presente na última geração, já que o elitismo foi ativado.

Os indivíduos fornecidos pelo GA foram otimizados pelo simplex por um número fixo de avaliações da função erro ou até que a diferença entre valores consecutivos da função erro fosse inferior a uma tolerância fixa (10^{-5}). As seguintes opções do algoritmo simplex foram usadas: $\alpha = 1.0$, $\beta = 0.5$ e $\gamma = 2.0$ ^[154, §10.4]. A faixa de variação dos parâmetros otimizados no simplex foi igual àquela do GA. Os indivíduos cujos parâmetros ultrapassassem estas faixas de variação, durante a otimização no simplex, tiveram sua função erro penalizada (ou seja, a função erro assumiu um valor arbitrariamente grande).

Tabela 20: Parâmetros semiempíricos otimizados: fase I

Parâmetros ^a	Elementos ^b					Valores iniciais ^c
	H	C	O	P	S	
$U_{\mu\mu}, \zeta_{\mu}, \beta_{\mu}, \alpha$	o	o	o	o	o	MNDO ^[143, 245]
$g_{\mu\nu}, h_{\mu\nu}, \rho_a$ ($a = 0, 1, 2$)	f	f	f	f	f	MNDO ^[143, 245]
K_i, L_i e M_i ($i = 1$ a 4)	z	z	z	z	o	AM1 ^[153]

^a indica os parâmetros semiempíricos (§1.4.2). ^b indica pela letra “o” quais elementos químicos que tiveram seus parâmetros otimizados, pela letra “f” os parâmetros que foram fixados e pela letra “z” os parâmetros zerados. ^c indica qual parametrização foi inicialmente fixada ou usada para delimitar as faixas de variação durante a otimização. Os valores das parametrizações originais (MNDO e AM1) estão mostrados na tabela 22. Os parâmetros derivados D_1 e D_2 foram recalculados pelas equações A.4 e A.5 para cada conjunto de parâmetros otimizado.

Embora refinados, ou seja, com menores erros, os indivíduos otimizados pelo simplex eram geralmente similares ao indivíduo mais adaptado obtido no GA. Os indivíduos refinados foram introduzidos no GA para checar sua sobrevivência durante novas evoluções. A otimização dos parâmetros prosseguia a um novo ciclo GA+simplex se os parâmetros otimizados, as faixas de variação ou a função erro adotada eram alterados.

3.1.2 Reparametrização da hamiltoniana semiempírica

As propriedades de referência usadas no conjunto de treinamento da calibração específica contém apenas contribuições eletrônicas e de repulsão nuclear. Nenhuma contribuição de ZPE , de energia térmica ou de entropia foi adicionada às energias de referência ou calculada pelo método semiempírico. Os valores de referência para as espécies e reações de ésteres de fosfato foram apresentadas no capítulo 2. O restante das propriedades incluídas na função erro e das espécies incluídas no conjunto de treinamento estão descritas abaixo.

Além de alterações no número de parâmetros e nas faixas de variação otimizadas, diferentes funções erro e conjuntos de treinamento foram usados. Assim, três fases de otimização podem ser definidas.

Na fase I (tabela 20), um total de 41 parâmetros foram otimizados simultaneamente, com faixas de variação de $\pm 10\%$ em relação aos parâmetros iniciais (tabela 22). Funções gaussianas foram adicionadas à f_{core} (equação 1.49) apenas para enxofre, ou seja, os K_i dos outros elementos foram zerados. A função erro (equação 1.50) foi definida como:

$$P = W_{energia} \sum_i^{N-3} (\Delta E_i^{NDDO} - \Delta E_i^{ref})^2 + W_{norma} \sum_i^N \|\mathbf{g}_i^{NDDO}\|^2 \quad (3.1)$$

Tabela 21: Parâmetros semiempíricos otimizados: fase II

Parâmetros ^a	Elementos ^b					Valores iniciais ^c
	C	H	O	P	S	
$U_{\mu\mu}$	o	o				MNDO ^[143]
$\zeta_{\mu}, \beta_{\mu}, \alpha$	f	o				MNDO ^[143]
$U_{\mu\mu}, \zeta_{\mu}, \beta_{\mu}, \alpha$			o	o	o	fase I
$g_{\mu\nu}, h_{\mu\nu}$	o	o	o	o	o	MNDO ^[143, 245]
K_i, L_i e M_i ($i=1,2,3$)	z	o	o	o		AM1 ^{[144, 152]^d}
K_i, L_i e M_i ($i=1,2,3$)					o	fase I ^d

^{a,b,c} veja legenda na tabela 20. ^d parâmetros iniciais e faixas de variação para as funções gaussianas em H, O, P e S descritas no texto. Os campos em branco na coluna ^b estão especificados nas linhas adjacentes. Os parâmetros derivados D_1 e D_2 foram recalculados pelas equações A.4 e A.5, e ρ_a ($a = 0, 1, 2$) foram recalculados pelo procedimento iterativo A.6 para todos elementos (exceto os parâmetros D_1 e D_2 para C que foram fixados nos valores MNDO) para cada conjunto otimizado.

O conjunto de treinamento foi formado por $N = 11$ estruturas ao longo da reação A ($\text{CH}_3\text{S}^- + \text{TMP}$, figura 7): reagentes, produtos, A:IMC1, A:TS1, A:I1, A:TS2, A:TS3, A:IMC2, A:IMC3 (figura 13). O zero de energia corresponde à energia absoluta dos reagentes infinitamente separados (§2.2) e $N - 3$ corresponde ao número de espécies da reação A com energia relativa diferente de zero. As energias relativas calculadas pelo método semiempírico ΔE_i^{NDDO} e de referência MP2/6-311+G(2df,2p), $\Delta E_i^{ref} = \Delta E_{MP2}$ (tabela 4), foram obtidas na mesma geometria (MP2/6-311++G*, figura 13). A norma do gradiente *ab initio* é nula, pois os pontos estacionários foram totalmente otimizados. Portanto, o valor absoluto da norma semiempírica, $\|\mathbf{g}_i^{NDDO}\|$, compôs diretamente a função erro. Os pesos adotados foram $W_{energia} = 1$ (mol/kcal)² e $W_{norma} = 3$ (mol Å/kcal)². Os parâmetros evoluíram por cerca de 2.10^3 gerações no GA e 4.10^3 avaliações da função erro no simplex.

Na fase II (tabela 21), 84 parâmetros foram otimizados simultaneamente. As faixas de variação permitidas em relação aos parâmetros iniciais (tabela 21) foram de até 30% para os elementos H e C, 40% para O e P e 50% para S. Funções gaussianas foram adicionadas em f_{core} (equação 1.49) para todos os elementos, exceto C. A magnitude das gaussianas (parâmetro K_i) foi restringida entre ± 0.02 eV.Å, a fim de permitir a correção da repulsão excessiva entre os caroços (§1.4.2, p. 57), mas evitar a introdução de artefatos (mínimos artificiais)^[147, 157, 158] na PES semiempírica. A faixa de variação dos parâmetros L_i foi de 0.0 Å^{-2} até os respectivos parâmetros no AM1 (tabela 22) e dos M_i foi de 0 até 6.0 Å .

O conjunto de treinamento foi composto por todos os reagentes e produtos isolados, e pontos estacionários das reações A a H (figuras 7 e 8), que foram agrupados de acordo com

Tabela 22: Parâmetros semiempíricos originais (MNDO e AM1) para os elementos H, C, O, P e S

Parâmetros e unidades		MNDO ^[143, 245, 246] +(AM1) ^{[144, 152, 153]_a}				
		H	C	O	P	S
U_{ss}	eV	-11.906276	-52.279745	-99.644309	-56.143360	-72.242281
U_{pp}	eV		-39.205558	-77.797472	-42.851080	-56.973207
ζ_s	bohr	1.331967	1.787537	2.699905	2.108720	2.312962
ζ_p	bohr		1.787537	2.699905	1.785810	2.009146
β_s	eV	-6.989064	-18.985044	-32.688082	-6.791600	-10.761670
β_p	eV		-7.934122	-32.688082	-6.791600	-10.108433
α	$\text{\AA}^{-1}$	2.544134	2.546388	3.160604	2.415280	2.478026
g_{ss}	eV	12.848000	12.230000	15.420000	11.560000	12.880000
g_{sp}	eV		11.470000	14.480000	10.080000	11.260000
g_{pp}	eV		11.080000	14.520000	8.640000	9.900000
g_{p^2}	eV		9.840000	12.980000	7.680000	8.830000
h_{sp}	eV		2.430000	3.940000	1.920000	2.260000
D_1	bohr		0.807466	0.534602	1.012970	0.918993
D_2	bohr		0.685158	0.453625	0.937009	0.832851
ρ_0	bohr	1.058920	1.112428	0.882296	1.176903	1.056289
ρ_1	bohr	1.058920	0.813078	0.521237	1.024082	0.901794
ρ_2	bohr	1.058920	0.747843	0.526541	1.004136	0.895216
K_1	eV. $\text{\AA}$	-0.018336	-0.020061	0.081430	-0.031827	-0.509195
K_2	eV. $\text{\AA}$	0.005090	-0.001260	0.280962	0.018470	-0.011863
K_3	eV. $\text{\AA}$	0.122796	0.011355		0.033290	0.012334
K_4	eV. $\text{\AA}$		0.045924			
L_1	$\text{\AA}^{-2}$	2.000000	5.000000	7.000000	6.000000	4.593691
L_2	$\text{\AA}^{-2}$	5.000000	5.000000	5.000000	7.000000	5.865731
L_3	$\text{\AA}^{-2}$	5.000000	5.000000		9.000000	13.557336
L_4	$\text{\AA}^{-2}$		5.000000			
M_1	$\text{\AA}$	2.100000	2.050000	1.445071	1.474323	0.770665
M_2	$\text{\AA}$	1.800000	2.650000	0.847918	1.779354	1.503313
M_3	$\text{\AA}$	1.200000	1.600000		3.006576	2.009173
M_4	$\text{\AA}$		1.850000			

^a parâmetros do método MNDO^[143, 245, 246], exceto para as funções gaussianas de caroço (equação 1.49) cujos parâmetros (K_i , L_i e M_i) foram retirados do método AM1^[144, 152, 153].

os mecanismos $A_n + D_n$ e $D_n + A_n$ para cada reação (§2.2). No total, $N_e = 70$ valores de energia relativa para espécies ao longo das reações de tiólise e alcoólise de ésteres de fosfato foram introduzidos na função erro 3.2. As moléculas de ácido acético, acetato (modelos da cadeia lateral de ASP, p. 39) e fenilfosfato neutro (PhOPO_3H_2), cujas geometrias foram otimizadas no nível MP2/6-31+G*, também foram incluídas no conjunto de treino, totalizando $N = 60$ estruturas. A função erro foi ampliada e redefinida como:

$$\begin{aligned}
 P = & W_{energia} \sum_i^{N_e} (\Delta E_i^{NDDO} - \Delta E_i^{ref})^2 + W_{norma} \sum_i^N \|\mathbf{g}_i^{NDDO}\|^2 + \\
 & W_{PI} \sum_i^N (PI_i^{NDDO} - PI_i^{ref})^2 + W_{pKa} \sum_i^{N_p} (\Delta E_{pKa,i}^{NDDO} - \Delta E_{pKa,i}^{ref})^2 + \\
 & W_{corr} \sum_i^{N_c} (\Delta E_{corr,i}^{NDDO} - \Delta E_{corr,i}^{ref})^2 + W_{dian} \sum_i^{N_d} (\Delta E_{dian,i}^{NDDO} - \Delta E_{dian,i}^{ref})^2 \quad (3.2)
 \end{aligned}$$

As energias relativas calculadas pelo método semiempírico, ΔE_i^{NDDO} , e de referência MP2/6-311+G(2df,2p), $\Delta E_i^{ref} = \Delta E_{MP2}$ (tabelas 4 a 7 e 9 a 12), foram obtidas na mesma geometria para cada espécie (cap. 2) e, novamente, a norma do gradiente *ab initio* é nula (vide supra). PI é a energia do orbital molecular ocupado de mais alta energia (HOMO) e equivale ao negativo do potencial de ionização eletrônica vertical, segundo o teorema de Koopmans^{[124, §3.3.1][121, §10.5.1]}. As energias $\Delta E_{pKa,i}$ das reações 2.1, 2.2 e 2.4 (§2.2.4), $N_p = 3$, foram incluídas na função erro 3.2 para calibrar as reações de transferência de H^+ relevantes para a química das PTPs selvagens e mutantes (§1.3). Testes e otimizações de geometria utilizando os parâmetros semiempíricos calibrados na fase I mostraram os seguintes erros ou artifícios:

1. Ângulos diedrais C–O–P=O reduzidos para PhOP^- e MeOP^- ;
2. Ângulos diedrais H–O–P=O invertidos para E:TS2a e G:TS1a (§2.2);
3. Pseudorrotâmero de B:I com os grupos fenóxido e tiolato axiais mais estável que B:I (p. 89 e figura 16);
4. Metóxido arrancava H do grupo metila equatorial (CH_3) em C:TS1 e D:TS1;
5. Carbono aromático na posição *orto* no fenóxido era hidrogenado, formando $\text{C}_6\text{H}_6\text{O}$.

Energias MP2/6-311+G(2df,2p) foram calculadas para estas estruturas otimizadas pelos parâmetros semiempíricos e a diferença entre E_{MP2} para esta geometria e para a geometria *ab initio* correspondente (por exemplo, a espécie obtida no item 5 acima e fenol ou a

espécie obtida no item 3 acima e B:I), ΔE_{corr}^{ref} , foi introduzida na função erro 3.2, para corrigir os artifícios nas $N_c = 8$ espécies listadas acima. As energias relativas ΔE_{dian} de um total de $N_d = 18$ pontos com geometrias parcialmente otimizadas (a ligação P–X foi fixada, onde X = O ou S do grupo de saída) no nível MP2/6-31+G* ao longo das reações de dissociação dos diânions PhOPO_3^{2-} , $\text{CH}_3\text{OPO}_3^{2-}$ e $\text{CH}_3\text{SPO}_3^{2-}$ em PO_3^- e fenóxido, metóxido e tiolato, respectivamente, também foram incluídas na função erro 3.2. Para todas espécies i , as energias PI_i , $E_{pK_a,i}$, $E_{corr,i}$ e $E_{dian,i}$ de referência foram obtidas no nível MP2/6-311+G(2df,2p) para as espécies isoladas (fase gasosa). Os pesos adotados foram: $W_{energia} = 1$ (mol/kcal)², $W_{norma} = 3$ (mol Å/kcal)², $W_{PI} = 4$ (mol/kcal)², $W_{pK_a} = 1$ (mol/kcal)², $W_{corr} = 6$ (mol/kcal)², $W_{dian} = 3$ (mol/kcal)² e os parâmetros evoluíram por cerca de 10⁴ gerações no GA e 3.10⁴ passos de simplex.

Na fase III, os parâmetros reotimizados foram os mesmos da tabela 21, mas os valores iniciais foram os parâmetros obtidos na fase II. As faixas de variação permitidas não foram alteradas, ou seja, os limites absolutos para variação dos parâmetros foram exatamente os mesmos da fase II. A função erro foi modificada para:

$$P = W_{opt} \sum_i^{N_o} (\Delta E_{opt,i}^{NDDO} - \Delta E_i^{ref})^2 + W_{corr} \sum_i^{N_c} (\Delta E_{corr,i}^{NDDO} - \Delta E_{corr,i}^{ref})^2 + W_{pK_a} \sum_i^{N_p} (\Delta E_{pK_a,i}^{NDDO} - \Delta E_{pK_a,i}^{ref})^2 \quad (3.3)$$

As energias $\Delta E_{pK_a,i}$ das reações 2.1, 2.2 e 2.4 (§2.2.4) foram complementadas pela reação 2.3 e pelas seguintes reações:

onde os aminoácidos CYS e SER foram tratados na forma neutra, ou seja, o grupo carboxila esteve protonado (COOH) e o grupo amina, desprotonado (NH₂). No total, $N_p = 13$

energias de reação de transferência de H^+ foram incluídas na função erro 3.3. As energias $E_{pK_a,i}^{ref}$ foram obtidas no nível MP2/6-311+G(2df,2p) para as espécies isoladas (fase gasosa) e as geometrias das moléculas que não estavam incluídas no conjunto treino $N = 60$ (vide supra) foram totalmente otimizadas no nível MP2/6-31+G*. Os parâmetros semiempíricos usados para o nitrogênio na CYS e SER foram retirados do método MNDO^[143]. $\Delta E_{opt,i}^{NDDO}$ correspondeu aos $N_o = 43$ valores de energia relativa diferente de zero para espécies ao longo das reações A a H cujas geometrias foram *reotimizadas* pelo método semiempírico, partindo-se das geometrias *ab initio*. Apenas as estruturas de reagentes e produtos isolados, TSs e intermediários foram otimizadas. Os IMCs não foram incluídos. Assim, as energias $\Delta E_{opt,i}^{NDDO}$ descreveram parte dos caminhos de reação de tiólise e alcoólise de ésteres de fosfato obtidos na fase gasosa com os parâmetros semiempíricos calibrados. A $\Delta E_{corr,i}$ usada na equação 3.3 foi a mesma daquela descrita para a função erro 3.2. Os pesos adotados foram: $W_{opt} = 10$ (mol/kcal)², $W_{pK_a} = 0$ nos primeiros $2 \cdot 10^3$ gerações do GA e $W_{pK_a} = 4$ (mol/kcal)² nas últimas 10^3 gerações do GA, $W_{corr} = 2$ (mol/kcal)². Os parâmetros evoluíram por cerca de $3 \cdot 10^3$ gerações no GA e 10^4 passos de simplex.

Todas computações *ab initio* foram efetuadas com o sistema de programas Gaussian 98^[206] (§2.1). As computações semiempíricas foram efetuadas com o programa GEOMOP^[194], no formalismo restrito para camada fechada (p. 48) do método AM1 (§1.4.2). Um algoritmo quasi-Newton (§1.5.1), com o gradiente calculado analiticamente, foi usado nas otimizações de geometria. Até 20 passos de otimização foram permitidos para as estruturas de mínimo e até 10 passos para os TSs. O algoritmo de otimização combinada GA+simplex (§3.1) foi interfaciado ao programa GEOMOP por meio de rotinas em BASH (GNU Bourne-Again SHell)^[210] e as funções erro foram avaliadas por rotinas escritas em FORTRAN 77^[244].

3.1.3 Interação não ligante

3.1.3.1 Complexos bimoleculares

A energia de interação entre as regiões QM e MM (§1.6.4) foi testada e calibrada a partir de comparações com as energias de interação entre complexos bimoleculares modelo na fase gasosa, obtidas por cálculos de estrutura eletrônica *ab initio* (§1.4.1).

Devemos enfatizar que as ligações de hidrogênio entre os oxigênios não ligantes do grupo fosfato e os hidrogênios da cadeia lateral da ARG conservada e da cadeia protéica

principal no P-loop (figura 10) são interações essenciais para atividade catalítica nas PTPs (§1.3 e §4.3.3). Nas simulações apresentadas no cap. 4, o grupo fosfato pertence à região QM e as cadeias protéica principal e lateral da ARG estão na região MM, assim, as ligações de hidrogênio são descritas pela parte híbrida do potencial ($E_{QM/MM}$) e, portanto, precisam ser testadas e calibradas.

Os complexos moleculares usados na calibração são modelos das ligações de hidrogênio descritas. A cadeia lateral da ARG foi representada pelo cátion guanidina, $C^+(NH_2)_3$, e pela espécie metilada, $H_3C-NH-C^+(NH_2)_2$. A cadeia protéica foi representada pela formamida, $HCONH_2$, e pela N-metilacetamida, $H_3C-CONH-CH_3$. O grupo fosfato foi representado pelo metafosfato, PO_3^- , e pelo éster metilfosfato monoprotonado. Complexos de van der Waals entre fenol e metanol também foram usados, pois representam a interação do substrato com as cadeias carbônicas laterais (dos aminoácidos denotados por XXX no P-loop) presentes na região do sítio ativo enzimático (§1.3). As ligações de hidrogênio entre água e fenol, e água e ácido acético também foram testadas, pois são modelos para solvatação do grupo de saída e do ácido geral, respectivamente (§1.3).

Os complexos (figura 26) foram construídos a partir das geometrias das moléculas isoladas, otimizadas no nível MP2/6-31+G*. Para os complexos entre metafosfato e formamida, e entre metafosfato e guanidina, orientações semelhantes àsquelas observadas no MC das PTPs (figura 10) foram escolhidas. A geometria dos complexos de formamida e guanidina foi parcialmente reotimizada no nível MP2/6-31+G* e a orientação foi preservada, fixando-se os ângulos $N-H\cdots O$ e $H\cdots O-P$, e o ângulo diedral $C-N-H\cdots O$ para o complexo de formamida ou o diedral $N-H\cdots O-P$ para o complexo de guanidina. O ângulo diedral foi fixado de modo que as duas moléculas estivessem coplanares. Nos complexos zwitteriônicos (ou seja, entre um cátion e um ânion, por exemplo, complexo I), as distâncias das ligações $N-H$ dos hidrogênios que coordenavam-se ao grupo fosfato também foram fixadas.

A partir das geometrias reotimizadas dos dois complexos, diferentes orientações foram geradas, variando-se os ângulos de ligações e diedral fixados. Os valores escolhidos para os ângulos de ligação estão indicados na figura 26. Por exemplo, o complexo V tem as mesmas coordenadas que IV, exceto os ângulos $H\cdots O-P$ e $C-N-H\cdots O$, cujos valores no complexo V são 90° e 60° (neste caso, o metafosfato não está no mesmo plano que a formamida, figura 26), respectivamente. Da mesma maneira, o complexo II foi obtido de I, alterando-se apenas o ângulo $H\cdots O-P$ de 120° para 180° (figura 26). A partir das geometrias dos complexos I e IV, os complexos III, VI e VIII (figura 26) foram obtidos

Figura 26: Estrutura dos complexos bimoleculares usados na calibração de $E_{QM/MM}$. As linhas tracejadas indicam as distâncias que foram variadas para obtenção das energias de interação.

por otimizações de geometria parciais, com restrições similares.

No complexo entre fenol e metanol, o ângulo C1···C2–C3 (numeração no complexo VII, figura 26) e o ângulo diedral C2···H5–C1–O4 foram fixados em 90° e 180°, respectivamente. A geometria foi reotimizada no nível MP2/6-31+G*. Nos complexos com água, nenhuma restrição foi imposta às geometrias otimizadas no nível MP2/6-31+G* e apenas uma orientação de cada complexo bimolecular foi testada (§3.2.3). No total, foram definidas 8 diferentes orientações para o complexo I, 12 orientações para o complexo IV, 3 orientações para o complexo III, 2 para o complexo VI, 1 para o complexo VIII e 2 para o complexo VII (apenas algumas orientações são mostradas na figura 26).

3.1.3.2 Energia de complexação

Para cada complexo e orientação, energias foram calculadas no nível MP2/6-311+G (2df,2p), em diferentes distâncias R da ligação de hidrogênio ou dos centros carbônicos (indicados pela linha pontilhada na figura 26). O BSSE foi corrigido para cada ponto pelo procedimento de contra-ponto (p. 54)^[140, 141], ou seja, as energias de complexação *ab initio*, ΔE_{MP2}^{comp} , foram obtidas para o complexo formado entre as moléculas A e B na distância R pela expressão 1.39. O sistema de programas Gaussian 98^[206] foi usado para as computações *ab initio*.

A energia de complexação ΔE_{hib}^{comp} foi computada pelo potencial híbrido (equação 1.82, §1.6.4) para exatamente as mesmas geometrias usadas nos complexos *ab initio*, porém, uma molécula formou a região QM e a outra, a região MM. Neste caso, não existe BSSE e a energia de interação entre a molécula A (QM) e B (MM) pode ser escrita por:

$$\Delta E_{hib}^{comp}(R) = E_{hib}^{AB}(R) - E_{QM}^A - E_{MM}^B \quad (3.13)$$

onde E_{hib}^{AB} representa a energia calculada pelo potencial híbrido para o complexo AB , E_{QM}^A representa a energia para a molécula A isolada, calculada pela parte QM do potencial e E_{MM}^B representa a energia para a molécula B isolada, calculada pela parte MM do potencial (vide infra).

Nos complexos I a V e VIII (figura 26), o metafosfato pertence à região QM e a outra molécula, à região MM. No complexo VI, o metilfosfato está na região QM e no complexo VII, ambos os casos, ou seja, fenol na região QM com o metanol MM e o fenol na região MM com o metanol QM foram usados. Nos complexos entre água e ácido acético, e entre água e fenol, água foi representada pelo potencial clássico TIP3P^[247] e ácido acético e fenol pertenceram à região QM.

A região QM foi descrita pela hamiltoniana semiempírica calibrada MNDO+G/CHOPS (§3.1.2 e §3.2.1). A contribuição MM do potencial híbrido foi descrita pelo campo de força OPLS-AA^[174].

As propostas de Field et al. (correspondente às equações 1.83 e 1.84) e do grupo de Nancy (correspondente às equações 1.83 e 1.86) foram usadas para interação QM/MM eletrostática semiempírica (§1.6.4.1). A proposta de Field et al. foi implementada no programa GEOMOP para viabilizar as comparações entre as duas definições de interação e para facilitar a calibração dos parâmetros de LJ. As computações com potenciais híbridos cujos resultados são apresentados em §3.2.3 foram realizadas pelo programa

GEOMOP^[194].

Se as contribuições de van der Waals, $\mathcal{H}_{QM/MM}^{vdW}$, são excluídas da hamiltoniana de interação QM/MM (equação 1.81), as energias de complexação calculadas pelo potencial híbrido na equação 3.13 podem ser denominadas de energias eletrostáticas, ΔE_{elet}^{comp} .

A faixa de variação de R para cada complexo foi delimitada por R_{min} e R_{max} ($R_{min} < R_{max}$), com variações em R de 0.1 Å a 1 Å. R_{min} correspondeu à distância em que $|\Delta E_{MP2}^{comp}(R_e) - \Delta E_{MP2}^{comp}(R_{min})| > 0.2$ au, onde R_e é a distância de equilíbrio dos complexos. R_{max} correspondeu à distância em que $R_{max} = R_e + 4$ Å ou à distância em que $|\Delta E_{MP2}^{comp}(R_{max}) - \Delta E_{MP2}^{comp}(R_{max-1})| < 1.10^4$ au, onde R_{max-1} é o ponto anterior à R_{max} . A condição que fosse satisfeita no menor R_{max} foi adotada.

3.1.3.3 Calibração específica dos parâmetros

O procedimento de calibração dos parâmetros de LJ para os centros QM foi dividido em duas fases. Ambas otimizaram a seguinte função erro:

$$P = \sum_R W_R \left| \frac{\Delta E_{hib}^{comp}(R) - \Delta E_{MP2}^{comp}(R)}{\Delta E_{MP2}^{comp}(R)} \right| \quad (3.14)$$

onde a somatória corre sobre cada uma das distâncias R definidas entre os complexos bimoleculares. O peso W_R dependeu das energias de interação: se $|\Delta E_{MP2}^{comp}(R)| < 5.10^{-4}$ au, $W_R = 0$; se 5.10^{-4} au $\leq |\Delta E_{MP2}^{comp}(R)| \leq |\Delta E_{MP2}^{comp}(R_e)| + 2.10^{-2}$ au, $W_R = 1$; e se $|\Delta E_{MP2}^{comp}(R)| > |\Delta E_{MP2}^{comp}(R_e)| + 2.10^{-2}$ au, $W_R = 0.25$. Esta escolha de pesos assegura que as maiores contribuições para função erro correspondem aos complexos mais próximos do mínimo de energia de interação.

Na fase I, apenas o complexo VII em duas orientações (ângulo C1...C2-C3 = 90° em 9 distâncias R e C1...C2-C3 = 180° em 5 distâncias R) formou o conjunto de treinamento. Os parâmetros calibrados foram somente os parâmetros de LJ atribuídos ao carbono aromático e hidrogênio ligado à carbono aromático quando o fenol era QM, e carbono alifático e hidrogênio ligado a carbono alifático quando o metanol era QM. Os parâmetros de LJ originais do campo de força OPLS-AA foram atribuídos ao oxigênio e hidrogênio do grupo hidroxila QM.

Na fase II, o complexo IV foi usado em 4 orientações diferentes (por exemplo, IV e V na figura 26), cada uma com 9 pontos em diferentes R ; o complexo I foi usado em 4 orientações (por exemplo, I e II na figura 26), cada uma com 10 diferentes R ; e o complexo V em 3 orientações, cada uma com 6 diferentes R . A escolha das orientações

de cada complexo também foi baseada nas diferenças entre os perfis de $\Delta E_{MP2}^{comp}(R)$.

Os parâmetros de LJ atribuídos ao fósforo e aos oxigênios não ligantes do grupo metafosfato foram calibrados. Os parâmetros de LJ originais para fósforo de grupo fosfato foram tirados do campo de força AMBER^[173]. Os parâmetros de LJ originais para os oxigênios não ligantes do grupo fosfato e para o oxigênio esterificado do metilfosfato (ausentes do campo de força OPLS-AA) foram tirados dos parâmetros OPLS-AA para oxigênio ionizado do grupo carboxilato e oxigênio de metanol, respectivamente. Os parâmetros para oxigênio ionizado do grupo carboxilato no campo de força AMBER são iguais aos parâmetros do campo OPLS-AA para este mesmo tipo atômico. No AMBER, estes parâmetros de LJ também são usados para os oxigênios não ligantes do grupo fosfato^[173].

As cargas parciais q_m (equação 1.64) atribuídas ao nitrogênio e hidrogênio envolvidos nas ligações de hidrogênio com o grupo fosfato também foram otimizadas. As cargas atribuídas aos mesmos tipos atômicos que não participavam diretamente da ligação de hidrogênio não foram alteradas. A soma dos valores das cargas parciais otimizadas foi fixada. Por exemplo, as cargas OPLS-AA de N e H na guanidina são -0.80 e 0.46 e , respectivamente. A soma destes valores foi mantida durante a calibração, isto é, se uma carga -0.85 e fosse atribuída ao N, a carga 0.51 e seria atribuída ao H. Desta forma, o valor da carga formal da molécula (0 para formamida ou +1 para guanidina e metilguanidina) foi preservado, sem que as cargas atribuídas aos outros tipos atômicos MM fossem alteradas.

Variações em até 0.2 e foram permitidas para as cargas parciais. Os ϵ e σ foram variados em até 100% dos parâmetros LJ originais do campo de força OPLS-AA (ou AMBER para o fósforo). Os parâmetros otimizados evoluíram por cerca de $5 \cdot 10^2$ gerações no GA e 10^3 avaliações da função erro no simplex, em cada uma das duas fases de otimização.

3.1.4 Fronteira e interação covalente

Uma das maneiras de tratar a ligação covalente entre as regiões QM e MM é o uso de átomos fictícios (§1.6.4.2). A implementação adotada nesta tese^[179] utiliza hidrogênio como átomo fictício (L), posicionado a uma distância $b_{LQ} \equiv 1$ Å entre a ligação dos átomos QM e MM na fronteira covalente, denominados Q e M , respectivamente. As coordenadas do átomo fictício ($\mathbf{R}_L$) são geradas a partir das coordenadas de Q e M por:

$$\mathbf{R}_L = \mathbf{R}_Q + b_{LQ} \frac{\mathbf{R}_M - \mathbf{R}_Q}{|\mathbf{R}_M - \mathbf{R}_Q|} \quad (3.15)$$

Assim, L estará sempre posicionado entre os átomos Q e M . O átomo fictício interage normalmente com os centros QM, participando da determinação de E_{QM} , mas nenhuma interação deste átomo com a região MM é computada^[187].

Embora o uso do átomo fictício seja suficiente para saturar a valência da região QM (em razão do uso do hidrogênio, as ligações de fronteira devem ser ligações simples), ele não assegura que a ligação entre as regiões QM e MM seja mantida. Portanto, o termo de estiramento de ligação $\mathcal{V}_{lig}$ (p. 65) entre os átomos Q e M é mantido no potencial híbrido. Todas interações covalentes ($\mathcal{V}_{cov}$) que incluem pelo menos um centro MM além de M também são preservadas. Por exemplo, a energia correspondente à torção de um diedral contendo dois centros MM e dois centros QM é computada. Os parâmetros covalentes do campo de força (§1.6.1) são usados para os centros QM

As interações de LJ ($\mathcal{V}_{LJ}$) entre os átomos QM e MM na fronteira seguem o esquema de exclusão das interações 1–2 e 1–3, e de escalonamento das interações 1–4 (p. 67). As interações eletrostáticas (§1.6.4.1) são calculadas sem exclusão ou escalonamento, entre todos átomos QM (excluindo o átomo fictício, vide supra) e centros MM.

Ao lado da definição da forma da interação eletrostática semiempírica (§3.2.3) e das expressões apresentadas na §1.6.4, os detalhes apresentados acima completam a descrição do potencial híbrido implementado na biblioteca de rotinas DYNAMO^[179], usada nas simulações de reações enzimáticas (cap. 4).

Os possíveis desvios introduzidos pelo potencial híbrido em virtude da fronteira covalente entre as regiões QM e MM foi testada para os seguintes sistemas modelo, isolados na fase gasosa:

onde todos aminoácidos foram tratados na forma neutra, ou seja, o grupo carboxila esteve protonado (COOH) e o grupo amina, desprotonado (NH₂). As reações 3.16 a 3.19 são dissociações que produzem metafosfato, a partir do aminoácido com a cadeia lateral fosforilada. As reações 3.20 e 3.21 são transferências de H⁺ (3.20 é idêntica à 3.6).

Os grupos CH_3OH e PO_3^- , e a cadeia lateral dos aminoácidos foram tratados pela parte QM do potencial híbrido. Por exemplo, os átomos $\text{C}_\beta\text{H}_2-\text{SH}$ da CYS estavam na região QM. Portanto, a fronteira entre as duas regiões foi a ligação $\text{C}_\alpha-\text{C}_\beta$. Na parte QM, os parâmetros semiempíricos calibrados MNDO+G/CHOPS (§3.2.1) foram usados para os átomos de H, C, O, P e S. O campo de força OPLS-AA foi usado para a parte MM do potencial híbrido. Computações em que toda molécula foi descrita pela hamiltoniana semiempírica pura também foram realizadas. Neste caso, os parâmetros AM1 originais foram usados para N. As geometrias usadas foram determinadas *ab initio* (vide infra). Os cálculos com potencial híbrido ou semiempírico puro foram executados com a biblioteca de rotinas DYNAMO^[179].

Valores *ab initio* foram computados para as reações segundo a mesma metodologia usada no cap. 2. Geometrias foram otimizadas no nível MP2/6-31+G* e energias ΔE_{MP2} foram obtidas no nível MP2/6-311+G(2df,2p). Para as reações 3.16 e 3.17, otimizações de geometria parciais foram realizadas em diferentes distâncias da ligação P–X (X = S ou O) quebrada. O ângulo diedral $\text{C}_\alpha-\text{C}_\beta-\text{X}-\text{P}$ foi fixado para evitar que o grupo fosfato se dissociasse em direção a grupo carboxila (protonado). As computações *ab initio* utilizaram o sistema de programas Gaussian 98.

3.2 Resultados e discussão

A calibração específica da hamiltoniana semiempírica e as comparações feitas abaixo para as espécies e reações de ésteres de fosfato sempre usaram os valores *ab initio* como referência. Os valores *ab initio*, também obtidos a partir de simplificações para resolução da equação de Schrödinger (1.13, §2.3), são bastante próximos aos valores experimentais conhecidos. Portanto, os valores adicionais computados neste capítulo pela mesma metodologia *ab initio* (§3.1.2) devem ser próximos aos valores experimentais.

3.2.1 Parâmetros semiempíricos específicos

O procedimento de otimização dos parâmetros semiempíricos foi dividido em três fases *a posteriori* para explicitar e racionalizar as diferentes estratégias, conjuntos treino e funções erro usadas. Esta divisão existiu formalmente apenas entre a fase I e II. As fases II e III não têm uma fronteira definida pois a mudança da função erro 3.2 para 3.3 ocorreu gradativamente durante o procedimento de otimização, ou seja, cada termo foi retirado de 3.2 ou adicionado à 3.3 em diferentes gerações evoluídas pelo algoritmo

GA+simplex. Independentemente, a repetição do procedimento descrito em §3.1.2 resultaria em parâmetros semiempíricos equivalentes àqueles mostrados nesta seção.

Acreditamos que a precisão desejável para hamiltoniana semiempírica, em comparação com os resultados *ab initio*, seja de ± 3 kcal/mol, pois este valor permite que estudos semiquantitativos sejam realizados e, por exemplo, possam ser usados para responder as dúvidas sobre as reações enzimáticas levantadas na p. 44. Os parâmetros obtidos na fase I reproduziram as energias relativas das espécies ao longo da reação A (tabela 4) em ± 2 kcal/mol. No entanto, testes das energias relativas para reações de arilésteres de fosfato (reação B, por exemplo) apresentaram erros de mais de 10 kcal/mol. Portanto, decidimos ampliar o conjunto de treinamento na fase II para reações contendo arilderivados de fosfato. As reações de monoésteres e de alcoólise também foram incluídas a partir da fase II de otimização.

Na fase I, a função erro empregada foi idêntica àquela proposta por Rossi e Truhlar^[157], que estudaram um sistema simples, a reação de $\text{Cl} + \text{CH}_4$ com simetria C_{3v} . Para sistemas com mais graus de liberdade, parece-nos que somente a inclusão da norma do gradiente na função erro não garante que a PES semiempírica, calculada pelos parâmetros calibrados, seja semelhante a PES de referência, nem ao menos nas regiões de mínimo. Ou seja, a inclusão da norma é útil para aproximar a PES semiempírica da PES de referência, mas não garante que os mínimos e TSs calculados pela hamiltoniana calibrada serão energética e geometricamente similares aos mínimos e TSs de referência. Assim, a energia relativa das estruturas *reotimizadas* pelo método semiempírico, $\Delta E_{\text{opt},i}^{\text{NDDO}}$, foi incluída na função erro, pois, embora as estruturas não sejam idênticas às referências, espera-se que as estruturas sejam semelhantes e, principalmente, que a energia relativa $\Delta E_{\text{opt},i}^{\text{NDDO}}$ das espécies ao longo das reações sejam próximas às energias de referência ΔE_i^{ref} . Os conjuntos de parâmetros cujas geometrias reotimizadas diferiam significativamente das geometrias *ab initio* foram penalizados, porque $\Delta E_{\text{opt},i}^{\text{NDDO}}$ geralmente diferia da ΔE_i^{ref} calculada na geometria *ab initio*. A inclusão do RMSD ou da diferença entre as coordenadas internas da geometria reotimizada e da geometria de referência não garante que os desvios das energias relativas $\Delta E_{\text{opt},i}^{\text{NDDO}}$ serão pequenos (vide infra). Portanto, propriedades geométricas não foram incluídas na função erro da fase III. Ao contrário de ΔE_i^{NDDO} que amostra apenas um ponto da PES semiempírica, $\Delta E_{\text{opt},i}^{\text{NDDO}}$ é obtida após a avaliação de diversos pontos durante os ciclos de otimização de geometria, em especial os pontos de menor energia próximos a mínimos ou pontos de sela da PES. Estas são as regiões mais importantes para o estudo de reações químicas e, no caso dos mínimos, são também as regiões mais populadas durante a amostragem do espaço de configurações em uma simulação de

dinâmica molecular (§1.6.2). As otimizações de geometria necessárias para obtenção de $\Delta E_{opt,i}^{NDDO}$, especialmente dos TSs, tornaram-se custosas, pois foram repetidas milhões de vezes (nas milhares de gerações no GA). Portanto, as otimizações de geometria foram introduzidas na função erro apenas na fase III, quando pareceu-nos que uma boa parte da superfície de parâmetros semiempíricos já havia sido explorada e que as outras propriedades de interesse (equação 3.2) estavam bem calibradas. O número de ciclos de otimização de geometria também foi reduzido, assim, as otimizações convergiram (dentro dos limites padrão estipulados pelo programa GEOMOP) somente para as moléculas com menor número de graus de liberdade. O número de ciclos foi suficiente, no entanto, para permitir que a norma do gradiente diminuísse pelo menos 10^3 vezes e que mudanças estruturais significativas ocorressem nos cálculos com os parâmetros semiempíricos menos adaptados. Os IMCs não foram incluídos no conjunto treino para otimização de geometria, porque não são espécies relevantes para a reatividade em fase condensada, já que os complexos de encontro (RS e ES, figura 1) têm geometrias e propriedades bastante distintas dos IMCs de fase gasosa.

A construção da função erro e dos pesos W atribuídos para cada propriedade foram baseadas nas otimizações originais^[143–145] e específicas^[157, 167], e nas tentativas e erros feitos ao longo do procedimento de calibração. As expressões 3.2 e 3.3, e os W adotados são aqueles que mostraram-se mais eficientes dentro de diversas tentativas não apresentadas aqui. As faixas de variação dos parâmetros permitidas nas fase II e III do procedimento de calibração foram escolhidas em consequência do resultado da fase I e das gerações iniciais na fase II cujos parâmetros para S, O e P foram os que mais variaram, nesta ordem.

Os artifícios introduzidos pelos parâmetros obtidos na fase I e listados na p. 137, foram corrigidos pelo termo $\Delta E_{corr,i}$ na fase II. Os itens 4 e 5 (p. 137) eram os mais preocupantes, pois as estruturas obtidas com os parâmetros semiempíricos da fase I têm E_{MP2} cerca de 20 a 30 kcal/mol maiores do que a E_{MP2} das estruturas *ab initio* correspondentes (p. 137). O termo $\Delta E_{corr,i}$ foi mantido na fase III, caso contrário, os parâmetros obtidos voltavam a apresentar os mesmos artifícios. Portanto, calibrações específicas de parâmetros semiempíricos devem ser testadas ao longo do procedimento de calibração, de modo que a mudança dos parâmetros não introduza artifícios ou erros no cálculo das propriedades e dos processos que não são diretamente relacionados às reações de interesse. O uso de termos como $\Delta E_{corr,i}$ foi suficiente para eliminar os artifícios e para corrigir os parâmetros apresentadas neste capítulo.

A energia do HOMO, PI_i , também foi empregada nas funções erro das parame-

trizações originais^[143]. Além de ser comparável ao potencial de ionização eletrônica, esta propriedade é relevante, porque indica a qualidade da estrutura eletrônica determinada pelos parâmetros semiempíricos calibrados em comparação com a estrutura eletrônica *ab initio* e, porque é diretamente relacionada à reatividade das espécies^[46, §5.4].

As reações de dissociação dos diânions foram incluídas no conjunto de treinamento como modelos das reações do substrato ligado e do intermediário fosforilado formado no sítio ativo das PTPs (§1.3). A faixa de variação do potencial de ionização eletrônica vertical, estimado pelo teorema de Koopmans, para as espécies ao longo das reações de dissociação de PhOPO_3^{2-} , $\text{CH}_3\text{OPO}_3^{2-}$ e $\text{CH}_3\text{SPO}_3^{2-}$ foi respectivamente -0.06 a 0.92 eV, 0.13 a 0.37 eV, e -0.57 a 0.26 eV. A energia para ionização e, logo, a estabilidade eletrônica destas espécies, é pequena e as espécies podem sofrer autoionização em fase gasosa^[238]. No entanto, em fase condensada, as espécies duplamente carregadas podem ser observadas, pois o ambiente influencia a estrutura eletrônica (principalmente das espécies pouco estáveis em fase gasosa) e estabiliza os diânions^[248]. A expansão da função de onda eletrônica por um conjunto base incompleto equivale a um confinamento espacial dos elétrons^[249] e pode ser comparada ao efeito do ambiente sobre a estrutura eletrônica do diânion. Contudo, os resultados desta expansão incompleta para espécies multianiônicas na fase gasosa devem ser vistos cautelosamente e usados semiquantitativamente, pois os diânions podem ser instáveis a medida que o conjunto base é ampliado por funções difusas (p. 53).

O conjunto de treinamento para as reações de transferência de H^+ foi composto por modelos das possíveis reações no sítio ativo das PTPs, como a transferência de H^+ do ácido geral para o substrato ou para o grupo de saída (reações 2.1 a 2.3 e 3.4) e do nucleófilo para o substrato (reações 2.4 e 3.9 a 3.12)^[22, 120]. As diferenças entre a acidez de CH_3SH e CH_3OH (reação 3.5), e entre SER e CYS (reação 3.6) foram incluídas para calibrar a acidez do nucleófilo nas PTPs selvagens em relação aos mutantes $\text{CYS} \rightarrow \text{SER}$. A diferença de energia entre os possíveis estados de protonação dos substratos metil (reação 3.7) e fenilfosfato (reação 3.8) também foram incluídas para calibrar a estabilidade relativa de cada estado (§1.3.3). As observações sobre o uso de diânions no conjunto de treinamento feitas no parágrafo anterior são igualmente válidas para as espécies duplamente carregadas usadas para calcular ΔE_{pK_a} .

Os erros encontrados em cálculos semiempíricos de compostos contendo elementos hipervalentes^[162] (p. 59) do segundo período (§1.4.2) podem ser atribuídos às seguintes causas:

- Parâmetros semiempíricos inadequados, ou seja, um mínimo local e não global no espaço de parâmetros^[152, 153];
- Conjunto base atômico incompleto, pela ausência de funções de simetria d ^[148, 152, 164];
- Elementos com diferentes estados de oxidação nas moléculas do conjunto de treinamento.

A parametrização inadequada pode ser contornada por uma calibração específica e por uma maior exploração da superfície de parâmetros^[145].

A inclusão de orbitais atômicos de simetria d no conjunto base semiempírico trouxe melhorias para compostos hipervalentes após a reparametrização do método^[27, 148, 168]. Porém, não é claro quanto desta melhora deve-se à necessidade intrínseca por funções de momento angular d para descrição das ligações químicas e dos orbitais moleculares de valência ou à simples extensão e flexibilização do conjunto base e a consecutiva reparametrização.

A comparação dos resultados de cálculos no nível HF empregando funções base de simetria sp (extendidos ou não por funções d)^[250] para o estado fundamental de uma série de moléculas indicou que o papel das funções d nos compostos com elementos hipervalentes do segundo período não é diferente do papel destas funções na descrição de moléculas sem elementos hipervalentes ou com apenas átomos do primeiro período. Ou seja, funções d seriam responsáveis apenas pela *polarização* da função de onda, flexibilizando a limitação imposta pelo uso apenas de funções de simetria s e p (§1.4.1.1, p. 53). Otimizações de geometrias, utilizando conjuntos base somente com funções sp , resultam em geometrias qualitativamente corretas (por exemplo, as distâncias de ligação P–F axiais são maiores que as d(PF) equatoriais para a molécula PF₅). No entanto, a adição de funções d ao conjunto base é necessária para obtenção de geometrias quantitativamente corretas^[250].

A análise da população natural^[251] é um método de análise da população de orbitais moleculares que também foi usado para estimar a contribuição de funções d para os orbitais de valência de compostos de fósforo^[162] e enxofre^[252]. Em ambos os casos, a contribuição destas funções para a população eletrônica dos orbitais ocupados não ultrapassa 10%. Esta contribuição apenas aumenta o caráter iônico ou polariza as ligações. Avaliamos a contribuição dos orbitais atômicos d pela análise de população natural^[251] das espécies ao longo da reação A. A população correspondente às funções d não superou 0.1 e para o fósforo e 0.02 e para o enxofre. Portanto, as funções d são responsáveis pela polarização função de onda eletrônica nos derivados de fosfato e não são intrinsicamente

necessárias para descrição da hibridização ou da formação das ligações químicas e dos orbitais moleculares de valência.

As populações de orbitais ou cargas parciais de átomos em moléculas não são observáveis físicas. Logo, não existe uma maneira única de se definir e analisar estas populações e os diferentes esquemas existentes podem sugerir interpretações conflitantes que devem ser analisadas cautelosamente.

O uso dos mesmos parâmetros semiempíricos para diferentes estados de oxidação de um elemento preserva a generalidade do método, mas deve ser a principal razão dentre aquelas listadas acima, para os erros encontrados em cálculos semiempíricos com os parâmetros tradicionais para compostos com elementos (hipervalentes) do segundo período. Por exemplo, o conjunto de treinamento da parametrização do método AM1 para enxofre^[153] foi composto por moléculas orgânicas e inorgânicas, como CH₃SH, SO₂ e H₂SO₄. O número de oxidação (NOX) do enxofre nestas moléculas varia em até 7 unidades! Parece-nos muito improvável que um único conjunto de parâmetros semiempíricos para o enxofre seja capaz de descrever a estrutura eletrônica de um átomo em ambientes químicos tão distintos. O NOX dos elementos de todas as moléculas utilizadas no conjunto de treinamento (§3.1.2) é preservado (a situação seria diferente para reações de oxirredução) e, portanto, o problema descrito neste parágrafo não é pertinente no procedimento de calibração específica na §3.1.2.

Os parâmetros iniciais adotados no procedimento de calibração (tabela 20) foram os valores originais do MNDO^[143, 245, 246], porque apenas compostos de enxofre com NOX = -2, como o nucleófilo CH₃S⁻, foram incluídos no conjunto de treinamento deste método. Nenhuma parametrização tradicional utilizou apenas compostos de P(V), como fosfatos, portanto, escolhemos o mesmo conjunto de parâmetros iniciais para todos os elementos.

Apresentaremos aqui apenas os resultados correspondentes ao conjunto final de parâmetros (tabela 23), obtido na fase III, pois este é o mais consistente e cujas propriedades calculadas apresentam os menores erros, principalmente $\Delta E_{opt,i}^{NDDO}$, com RMSD inferior a 3 kcal/mol (tabela 24).

Os valores dos parâmetros semiempíricos estão correlacionados a observáveis físicas. Por exemplo, $g_{\mu\nu}$ e $h_{\mu\nu}$ são obtidos dos valores espectroscópicos observados para átomos e seus íons (p. 56)^[149, cap. 15]. Porém, esta correlação foi perdida durante o procedimento de calibração, porque todos os parâmetros são variados simultaneamente e compensações podem ocorrer já que os parâmetros possuem interdependências desconhecidas (§1.4.2). Uma racionalização da dependência dos parâmetros e da sensibilidade das propriedades

Tabela 23: Parâmetros semiempíricos MNDO+G/CHOPS, especificamente calibrados para H, C, O, P e S

Parâmetros e unidades		MNDO+G/CHOPS				
		H	C	O	P	S
U_{ss}	eV	-13.320770	-52.772930	-81.496830	-29.926330	-107.17470
U_{pp}	eV		-44.135150	-70.963590	-46.743540	-56.396210
ζ_s	bohr	1.126840	1.787537	3.334590	1.792970	3.570580
ζ_p	bohr		1.787537	2.467500	1.715640	1.859890
β_s	eV	-6.758660	-18.985044	-43.021230	-6.393650	-16.671710
β_p	eV		-7.934122	-25.674230	-4.632620	-9.115900
α	$\text{\AA}^{-1}$	2.958707	2.546388	3.237480	2.385950	2.492397
g_{ss}	eV	16.020432	11.495681	13.464052	11.188208	12.390890
g_{sp}	eV		11.949816	9.546084	6.267279	11.013247
g_{pp}	eV		12.851185	16.368546	10.974892	9.760673
g_{p^2}	eV		11.683615	13.129728	9.718182	8.751465
h_{sp}	eV		2.334164	3.557707	2.602274	2.379333
D_1	bohr		0.807466	0.470220	1.149910	0.516220
D_2	bohr		0.685158	0.496350	0.975330	0.899690
ρ_0	bohr	0.849220	1.183490	1.010470	1.216010	1.097980
ρ_1	bohr	0.849220	0.827950	0.508370	0.949580	0.637340
ρ_2	bohr	0.849220	0.762300	0.436100	0.942970	0.961230
K_1	eV. $\text{\AA}$	-0.015119	0.000000	-0.003600	-0.000288	-0.017855
K_2	eV. $\text{\AA}$	0.000432	0.000000	-0.001584	0.000720	0.003168
K_3	eV. $\text{\AA}$	0.003456	0.000000	0.011375	0.001008	0.003600
K_4	eV. $\text{\AA}$	0.000000	0.000000	0.000000	0.000000	0.000000
L_1	$\text{\AA}^{-2}$	1.915663		4.772553	0.367927	2.545350
L_2	$\text{\AA}^{-2}$	0.157448		0.286400	0.185017	4.576466
L_3	$\text{\AA}^{-2}$	3.049013		2.407364	0.170983	2.453788
M_1	$\text{\AA}$	2.175055		5.076263	0.132665	1.083146
M_2	$\text{\AA}$	0.054415		5.277657	0.310188	2.670814
M_3	$\text{\AA}$	3.059781		1.533687	0.145587	1.762779

Os parâmetros ρ_a podem ser transformados nos parâmetros AX usados nos programas de cálculo baseados no código MOPAC^[147], como a biblioteca DYNAMO^[179] (cap. 4), pela relação: $AX = 0.5/\rho_a$ ($a = 0, 1$ ou 2 e $X = M, D$ ou Q , respectivamente). Todos outros parâmetros apresentados acima estão nas unidades diretamente usadas pelo código MOPAC.

calculadas, especialmente dos parâmetros de 1-centro ($U_{\mu\mu}$, $g_{\mu\nu}$ e $h_{\mu\nu}$) e dos expoentes de Slater (ζ_μ), não foi bem sucedida.

O parâmetro U_{ss} calibrado (tabela 23) é 50% menor para S e 50% maior para P do que os parâmetros originais (tabela 22). Como $U_{\mu\mu}$ pode ser correlacionado à eletronegatividade atômica^[253], podemos afirmar que a calibração atribuiu um caráter mais iônico para S e mais covalente para P (os valores de $U_{\mu\mu}$ e β_μ são sempre negativos). Esta observação corresponde àquela feita para otimização original dos elementos P^[152] e S^[153] no método AM1: os $U_{\mu\mu}$ otimizados serão menos negativos, caso compostos com elementos de alta valência (ou alto NOX) e com maior número de ligações (maior caráter covalente), como P(V) nos fosfatos, sejam incluídos no conjunto de treinamento. A observação acima também é uma evidência de que a ausência das funções de simetria d no conjunto base semiempírico pode ser absorvida pela parametrização específica. O caráter iônico das ligações contendo átomos hipervalentes foi atribuído, em parte, à polarização descrita por funções d (vide supra)^[162, 250]. Na ausência destas funções, a polarização é absorvida pela calibração dos parâmetros $U_{\mu\mu}$.

Os expoentes ζ_s de O e S também foram significativamente aumentados pela calibração, o que significa que os orbitais atômicos do tipo s foram contraídos nestes elementos. Os parâmetros β_s calibrados para O e S são respectivamente 30 e 55% menores que os parâmetros originais. Já β_p calibrado para P é 30% maior. Assim, a calibração atribui maior caráter s para as ligações contendo O, P e S, principalmente pra O e S cuja relação $\beta_s/\beta_p > 1.65$. O parâmetro α é aumentado significativamente apenas para H, resultando numa menor repulsão entre este elemento e os outros centros na molécula. Os parâmetros g_{ss} para H e g_{sp} para O também sofrem alterações significativas. Todas integrais $g_{\mu\nu}$ e $h_{\mu\nu}$ calibradas para P, com exceção de g_{ss} , são significativamente alteradas dos valores originais. Ao contrário dos parâmetros de 1-centro e 1-elétron, $U_{\mu\mu}$, as mudanças nas integrais de 1-centro e 2-elétrons parecem influenciar pouco as propriedades calculadas^[150].

As três gaussianas da função f_{core} calibradas para P e uma gaussiana calibrada para H ($i = 2$) apresentam baixas magnitudes, $K_i \leq 1.10^{-3}$ eV Å, e estão muito próximas ao centro atômico, $M_i \leq 0.3$ Å. Suas contribuições para f_{core} são muito pequenas e podem ser descartadas (ou seus K_i zerados), sem prejuízo para o cálculo de propriedades. As duas gaussianas atrativas ($K_i < 0$) para O também têm magnitude reduzida e estão centradas a uma grande distância do núcleo atômico, $M_i \approx 5$ Å ($i = 1$ ou 2), mas não podem ser retiradas sem prejudicar os resultados calculados. Estas gaussianas atrativas são responsáveis pela diminuição da repulsão entre caroços da ordem de 1.10^{-4} au, a partir

de 3 Å de separação entre os centros atômicos, até 1.10^{-3} au, a 5 Å de separação, onde as gaussianas estão centradas. Os parâmetros apresentados na tabela 23, denominados MNDO+G/CHOPS, foram aqueles usados nos testes comentados abaixo e nas simulações enzimáticas do cap. 4.

As geometrias otimizadas com os parâmetros calibrados na fase I, em que todos parâmetros para carbono foram variados, apresentaram artifícios, como perda de ressonância (o anel benzênico assumia uma conformação cadeira, como hexano). A partir da fase II, os parâmetros para carbono foram fixados nos valores MNDO, evitando que sua alteração introduzisse erros na estrutura dos esqueletos carbônicos (anéis benzênicos e grupos metila). Contudo, a relação entre as eletronegatividades atômicas é desbalanceada ou perturbada se $U_{\mu\mu}$ para carbono não responde às mudanças observadas para este parâmetro nos outros elementos calibrados. Esta perturbação deve ser responsável pelos artifícios listados nos itens 4 e 5 na p. 137. Assim, os parâmetros de 1-centro $U_{\mu\nu}$, $g_{\mu\nu}$ e $h_{\mu\nu}$ para carbono também foram calibrados, sem prejudicar as geometrias otimizadas. A adição de funções gaussianas em f_{core} para carbono não foi necessária. A calibração específica de apenas parte dos parâmetros para carbono não impediu ou influenciou significativamente a precisão das propriedades calculadas pelo conjunto MNDO+G/CHOPS, porque este elemento não participa das ligações quebradas ou formadas durante as reações incluídas no conjunto de treinamento.

3.2.2 Propriedades obtidas pela hamiltoniana QM calibrada

As comparações feitas abaixo para espécies e reações de ésteres de fosfato tomam sempre os valores *ab initio* como referência. Os desvios observados nas propriedades calculadas pelo conjunto calibrado MNDO+G/CHOPS são significativamente menores do que os desvios nas propriedades obtidas a partir das parametrizações semiempíricas tradicionais MNDO, AM1 e PM3, e pelo conjunto usado por Gao et al. (parâmetros AM1 para os elementos C, H e O, e MNDO para P e S)^[110] para simular a primeira etapa de reação catalisada por uma LM-PTP (tabela 24).

As propriedades ΔE , PI , ΔE_{dian} e ΔE_{pKa} (tabela 24) foram calculadas nas geometrias *ab initio* como descrito em §3.1.2. Para 70 estruturas, o RMSD da energia relativa, ΔE , calculada pelas parametrizações tradicionais é de 3 a 7 vezes maior do que o RMSD de ΔE obtido pelo conjunto MNDO+G/CHOPS. O RMSD para esta propriedade foi inferior a 2.5 kcal/mol para os parâmetros obtidos ao final da fase II de calibração, porém, esta propriedade não foi incluída na função erro da fase III (3.3) e, portanto, o desvio observado

Tabela 24: Comparação do módulo do desvio máximo (max) e quadrático médio (RMSD) para propriedades calculadas por diferentes parametrizações semiempíricas NDDO em relação a valores de referência *ab initio*

Propriedade		Parametrização semiempírica					
	desvio	N^a	MNDO	AM1	PM3	Gao ^b	MNDO+G/CHOPS
ΔE	RMSD	70	27.06	17.06	12.11	14.94	3.97
PI^c		60	0.70	0.62	0.90	0.72	0.42
ΔE_{dian}		18	8.17	4.15	10.41	3.34	2.17
ΔE_{pK_a}		13					
	RMSD		11.09	12.54	16.76	9.63	3.75
	max		20.97	20.04	29.12	16.54	9.88
	num ^d		12	10	10	10	4
ΔE_{opt}		43					
	RMSD ^e		15.80	16.22	10.38	14.73	2.79
	tri	16	12.82	16.11	8.88	17.61	4.09
	mono	27	17.85	16.87	11.53	13.34	1.75
	max		41.48	28.71	27.15	27.72	7.31
	num ^d		34	42	36	40	7
Ligações ^f		668					
	$\overline{\text{RMSD}}$		0.16	0.05	0.11	0.05	0.03
	RMSD_{max}		0.63	0.24	0.53	0.21	0.08
Ângulos ^f		629					
	$\overline{\text{RMSD}}$		9.33	5.56	5.46	4.04	3.41
	RMSD_{max}		21.30	28.40	18.98	25.98	8.84
Diedrais ^f		590					
	$\overline{\text{RMSD}}$		27.30	20.47	21.79	23.06	17.30
	RMSD_{max}		80.58	53.00	58.49	83.14	35.93

Definições das propriedades dadas em §3.1.2. ^a indica o número de espécies ou reações consideradas. ^b indica o uso dos parâmetros AM1 para os elementos C, H e O, e MNDO para P e S^[110]. Valores dos desvios em kcal/mol, exceto em ^c onde os valores estão em eV, em ^d onde os valores indicam número de espécies com módulo do desvio superior a 3 kcal/mol e em ^f onde valores indicam desvios em Å para as ligações e em ° para os ângulos. ^e indica RMSD total e RMSD entre as reações de triésteres (tri) e monoésteres (mono) de fosfato. ^f indica a média do RMSD ($\overline{\text{RMSD}}$) e o RMSD máximo (RMSD_{max}) das distâncias de ligações, ângulos entre ligações e ângulos diedrais das 39 geometrias otimizadas pelos métodos semiempíricos (veja o texto para detalhes).

nos cálculos com os parâmetros finais foi maior.

A PI foi a propriedade que se ajustou mais facilmente aos valores de referência. O ajuste não foi maior, porque o peso W_{PI} atribuído a esta propriedade foi reduzido em relação às energias relativas totais, já que esta não é a propriedade de maior interesse na calibração. Parece-nos que uma parametrização específica para PI pode resultar em incertezas médias em módulo para esta propriedade menores do que 0.1 eV. O RMSD (0.4 eV) das energias do HOMO calculadas pelo conjunto MNDO+G/CHOPS é cerca de 5% dos valores absolutos dos potenciais de ionização eletrônica experimental das moléculas neutras do conjunto de treinamento (tabela 17).

O RMSD da energia calculada para a dissociação dos reagentes dianiônicos, ΔE_{dian} , pelas parametrizações originais é 1.5 a 4 vezes maior do que obtido pelo conjunto calibrado. Os maiores desvios (em módulo) observados em ΔE_{dian} calculado pelo conjunto MNDO+G/CHOPS correspondem à reação de dissociação de $\text{CH}_3\text{OPO}_3^{2-}$, principalmente nas três estruturas cuja $d(\text{PO})$ varia de 2.2 a 2.8 Å. Os módulos dos desvios para as reações de PhOPO_3^{2-} e $\text{CH}_3\text{SPO}_3^{2-}$ são inferiores a 2.3 kcal/mol.

A calibração das reações de transferência de H^+ é relevante, porque esta é umas das reações que ocorrem no sítio ativo das PTPs e, porque este processo está diretamente ligado às perguntas sobre o mecanismo de catálise nas PTPs que estamos interessados em responder (p. 44). O RMSD para ΔE_{pK_a} calculada pelos parâmetros calibrados é menor do que 4 kcal/mol, cerca de $\frac{1}{4}$ do RMSD observado para as energias de reação calculadas pelas parametrizações tradicionais. Nestas parametrizações, quase todas reações têm os desvios em módulo superiores a 3 kcal/mol. A redução do desvio máximo observado para menos do que 10 kcal/mol e, principalmente, a redução do número de reações que têm desvio em módulo superior a 3 kcal/mol para apenas 4 reações (tabela 24) demonstram o sucesso da calibração específica. O módulo do desvio máximo computado pelo conjunto MNDO+G/CHOPS corresponde à reação 3.8 e representa menos de 10% do valor da energia de reação. O segundo maior desvio em módulo é 5.00 kcal/mol para a reação 2.1, a favor da formação dos produtos. Se as geometrias das espécies da reação 2.1 são otimizadas pelo conjunto MNDO+G/CHOPS, o módulo do desvio na energia de reação cai para 3.25 kcal/mol. Portanto, a basicidade calculada para fenóxido pelo conjunto MNDO+G/CHOPS é artificialmente aumentada. Os resultados das simulações enzimáticas, que empregaram este conjunto semiempírico, serão analisados (§4.3) à luz deste artifício.

A propriedade que melhor representa a diferença entre a PES de referência e a PES

calculada pelos métodos semiempíricos é ΔE_{opt}^{NDDO} . O RMSD para ΔE_{opt}^{NDDO} obtido para as parametrizações tradicionais é 4 a 10 vezes superior ao desvio obtido para o conjunto MNDO+G/CHOPS. O RMSD observado para as reações de monoésteres de fosfato é 1.75 kcal/mol (tabela 24), indicando que o desvio nas energias e nas alturas de barreira de reação de monoésteres calculadas pelos parâmetros MNDO+G/CHOPS é bastante reduzido. O RMSD observado para as reações de triésteres de fosfato é o dobro do desvio observado para os monoésteres. A calibração foi mais eficiente para as reações de monoésteres, porque o conjunto de treinamento e as propriedades empregadas nas funções erro otimizadas eram compostas majoritariamente por monoésteres. Por exemplo, nenhum triéster pertenceu aos conjuntos usados na avaliação de ΔE_{pK_a} e ΔE_{dian} . Temos dois mecanismos ($A_n + D_n$ e $D_n + A_n$) para reações de monoésteres e somente um para triésteres. Cada espécie contribuiu com o mesmo peso W para as diversas propriedades incluídas nas funções erro, independentemente do mecanismo ou do estado de alquilação do éster de fosfato. O módulo do desvio para ΔE_{opt} obtido, por exemplo, pelos conjuntos AM1 e Gao (tabela 24) é maior que 3 kcal/mol para quase todas as $N = 43$ espécies com geometria otimizada. O número de espécies com desvio em módulo maior do que 3 kcal/mol obtido com os parâmetros MNDO+G/CHOPS foi reduzido para apenas sete, listadas a seguir e com o desvio em kcal/mol indicado entre parênteses: A:TS1 (-5.5), $\text{CH}_3\text{O}^- + \text{TMDMP}$ (-7.3), B:TS1 (-3.8), D:TS1(6.9), D:I(3.5), D:TS2(4.4) e $\text{PhO}^- + \text{TMDMP}$ (6.7). Os desvios negativos significam que ΔE_{opt}^{NDDO} é menor do que ΔE^{ref} . O desvio máximo, ou seja, o erro no pior caso, ocorre no produto da reação A e corresponde a 20% da energia desta reação ($\Delta E_{MP2} = 37.2$, tabela 4). Os desvios observados para os TSs do ataque nucleofílico de CH_3S^- sobre fósforo nas reações A e B são negativos assim como o desvio para o produto da reação A. Portanto, as reações de tiólise de triésteres de fosfato são artificialmente favorecidas pela calibração MNDO+G/CHOPS. As outras quatro espécies com desvio superior a 3 kcal/mol ocorrem na mesma reação D ($\text{CH}_3\text{O}^- + \text{DMPP}$). Já que todos estes desvios são positivos e ΔE_{opt}^{NDDO} corresponde à energia relativa em que o zero é a energia dos reagentes, parece-nos mais provável que a soma da energia absoluta (em módulo) dos reagentes CH_3O^- e DMPP é superestimada pelo conjunto calibrado. Nenhuma espécie ao longo das reações de monoésteres tem desvio em módulo superior a 3 kcal/mol. Esta é mais uma evidência de que os parâmetros MNDO+G/CHOPS calculam propriedades com ótima precisão para monoésteres de fosfato.

O RMSD dos três tipos de coordenada interna foi calculado para cada uma das 39 geometrias reotimizadas, necessárias para obtenção de ΔE_{opt}^{NDDO} das reações de ésteres de fosfato. Não mostramos aqui cada um destes valores, mas somente a média dos 39 RMSD,

$\overline{\text{RMSD}}$, e o máximo RMSD entre as 39 geometrias reotimizadas, RMSD_{max} (tabela 24). Embora nenhuma informação geométrica seja explicitamente incluída nas funções erro otimizadas, o uso de propriedades energéticas e da norma do gradiente no procedimento de calibração é suficiente para diminuir $\overline{\text{RMSD}}$ para os três tipos de coordenadas internas, em relação aos desvios observados nas geometrias reotimizadas com os parâmetros tradicionais. O RMSD máximo também é reduzido de 3 a 8 vezes para as distâncias de ligação e 1.5 a 3 vezes para os ângulos entre as ligações e diedrais. Os erros esperados para as geometrias calculadas pelo conjunto MNDO+G/CHOPS não devem ultrapassar 0.1 Å para as distâncias de ligação e 10 ° para os ângulos entre as ligações. Os ângulos diedrais possuem desvios maiores, porque a torção de um diedral envolve energias muito menores do que a deformação de distâncias ou ângulos de ligação.

Concluimos que o procedimento de calibração dos parâmetros semiempíricos foi bem sucedido, já que todas propriedades importantes para a descrição dos caminhos de reações de ésteres de fosfato e dos modelos para a catálise realizada por PTPs tiveram seu RMSD e desvio máximo consideravelmente reduzidos. Em especial, $\Delta E_{\text{opt}}^{\text{NDDO}}$ apresenta RMSD (e também o desvio máximo, no caso das reações de monoésteres) inferior ao desvio máximo desejável para hamiltoniana semiempírica, estipulada em 3 kcal/mol. Além da redução do erro das propriedades calculadas pelos parâmetros MNDO+G/CHOPS, os testes finais indicam quais reações ou espécies apresentam desvios e, portanto, o erro associado ao método semiempírico *calibrado* é conhecido e previsível, ao contrário das parametrizações originais.

3.2.3 Interação não ligante

3.2.3.1 Definição da interação eletrostática

As propostas de Field et al.^[187] e do grupo de Nancy^[190, 194] para a interação QM/MM eletrostática foram testadas (§1.6.4.1). As energias de complexação, $\Delta E_{\text{hib}}^{\text{comp}}$, e sua contribuição eletrostática, $\Delta E_{\text{elet}}^{\text{comp}}$, foram calculadas para os complexos entre fenol (QM) e metanol (MM, complexo VII, figura 26) e entre metafosfato (QM) e formamida (MM, complexo IV, figura 26).

No complexo de van der Waals (VII), o mínimo de energia é observado em $R_e \approx 3.5$ Å (figura 27), onde R é a distância entre o carbono do metanol e o carbono em posição *para* no anel aromático (figura 26). O modulo de $\Delta E_{\text{MP2}}^{\text{comp}}$ nesta distância corresponde a menos que 2 kcal/mol e para $R < 3$ Å, a interação entre as moléculas é repulsiva, como é esperado

Figura 27: Energia de complexação obtida por diferentes hamiltonianas para distâncias C...C (linha pontilhada no complexo VII, figura 26) do complexo entre fenol e metanol na fase gasosa. O quadro inserido corresponde a um aumento de escala na região do mínimo de energia. Legenda: ΔE_{MP2}^{comp} (círculos); ΔE_{hib}^{comp} (linha tracejada) e ΔE_{elet}^{comp} (linha cheia), calculadas segundo a proposta de Field et al.; ΔE_{hib}^{comp} (linha traço-pontilhada) e ΔE_{elet}^{comp} (linha pontilhada), calculadas segundo a proposta do grupo de Nancy.

para complexos desta natureza^[172]. As ΔE_{hib}^{comp} calculadas segundo ambas propostas para interação eletrostática são maiores que ΔE_{MP2}^{comp} para $R < 3.5$ Å. A interação eletrostática, ΔE_{elet}^{comp} (§3.1.3.2), calculada segundo a proposta de Field é praticamente nula ao longo de R , como também é esperado para interação de complexos de van der Waals^[172]. A ΔE_{hib}^{comp} calculada segundo Field tem a mesma forma do perfil de ΔE_{MP2}^{comp} e na região do mínimo de energia (quadro inserido na figura 27), a diferença $|\Delta E_{hib}^{comp} - \Delta E_{MP2}^{comp}|$ é menor que 1.5 kcal/mol. Porém, para $R < 3$ Å, a energia de complexação obtida pelo potencial híbrido é muito maior que ΔE_{MP2}^{comp} . Este comportamento pode ser corrigido pela redução da interação repulsiva (termo em $1/r^{12}$ de $\mathcal{H}_{QM/MM}^{vdW}$, equação 1.81)

No entanto, a ΔE_{elet}^{comp} calculada segundo a proposta do grupo de Nancy é altamente repulsiva e maior que ΔE_{MP2}^{comp} , para qualquer R , indicando que esta proposta não contém apenas interações eletrostáticas. A forma do perfil de ΔE_{hib}^{comp} calculada segundo Nancy é diferente do perfil *ab initio*, principalmente na região do mínimo de energia e, portanto,

Figura 28: Energia de complexação obtida por diferentes hamiltonianas para distâncias $\text{H}\cdots\text{O}$ (linha pontilhada no complexo IV, figura 26) do complexo entre formamida e metafosfato na fase gasosa. A legenda é idêntica àquela da figura 27.

parece-nos que esta forma é inadequada para descrever a interação de complexos de van der Waals.

O complexo entre metafosfato e formamida representa um IMC com uma ligação de hidrogênio (complexo IV). O mínimo de energia de complexação é observado em torno de $R_e \approx 1.8 \text{ \AA}$ de distância $\text{H}\cdots\text{O}$ da ligação de hidrogênio (figura 28). A ΔE_{MP2}^{comp} nesta região é cerca de -20 kcal/mol , semelhante às energias de complexação dos IMCs apresentados na §2.2. Embora ΔE_{hib}^{comp} calculada segundo ambas propostas para parte eletrostática sejam próximas à ΔE_{MP2}^{comp} para $R > 2.2 \text{ \AA}$, novamente, ΔE_{hib}^{comp} e ΔE_{hib}^{elet} calculadas segundo Nancy são altamente repulsivas em distâncias menores que R_e e a ΔE_{hib}^{elet} é claramente desestabilizante, mesmo para $R < 2 \text{ \AA}$. Este comportamento é qualitativamente incorreto, pois espera-se que a interação eletrostática seja responsável por grande parte da estabilização de um IMC^[122]. Portanto, a contribuição dos dois termos no lado direito da equação 1.86 deve estar desbalanceada^[190]. A ΔE_{hib}^{comp} calculada segundo Field é coincidente à ΔE_{MP2}^{comp} . Embora ΔE_{hib}^{elet} calculada segundo Field apresente o comportamento estabilizante esperado para um IMC, esta contribuição não é puramente

eletrostática, porque uma repulsão de curto alcance ($R \approx 1.25$ Å na figura 28 e $R < 2.5$ Å na figura 27) é observada. Esta pequena repulsão é resultante do termo exponencial (α^m) designado para os centros MM na equação 1.84.

Concluimos que a definição da contribuição eletrostática semiempírica segundo a proposta de Field et al. (equações 1.83 e 1.84)^[187] apresenta a melhor forma e, portanto, é mais adaptada para construção do potencial híbrido^[190]. A proposta do grupo de Nancy não é indicada em razão da elevada repulsão observada e erros significativos devem ser obtidos em computações com esta definição do potencial híbrido^[254].

3.2.3.2 Parâmetros específicos

Uma vez determinada a forma da interação eletrostática semiempírica do potencial híbrido, a interação não ligante entre a região quântica e a clássica foi calibrada para reproduzir energias de complexação *ab initio* na fase gasosa (figura 26).

A forma da função erro usada nesta seção (equação 3.14) foi alterada em relação às formas usadas anteriormente (por exemplo, 1.50 ou 3.3), porque diferenças entre as energias de referência (ΔE_{MP2}^{comp}) e a energia calculada pelo potencial híbrido (ΔE_{hib}^{comp}) na região repulsiva do perfil de energia de complexação resultavam em contribuições absolutas para função erro muito maiores do que as contribuições nas regiões de mínimo ou em $R > R_e$. No entanto, as regiões de mínimo de energia são aquelas mais visitadas durante uma simulação molecular e, logo, o potencial deve ser preferencialmente calibrado para que os valores próximos ao mínimo sejam similares à referência. Nas demais regiões, a forma correta (e não, necessariamente, os valores) do potencial são suficientes. Assim, a função erro 3.14 é uma soma de valores relativos de energia e W_R foi escolhido de modo que os complexos com ΔE^{comp} mais próximas ao mínimo (dentro de uma faixa de $\pm 2.10^{-2}$ au) contribuíssem em maior intensidade.

Testes demonstraram que a otimização apenas dos parâmetros de LJ atribuídos aos átomos QM era suficiente para reduzir o RMSD de ΔE_{hib}^{comp} calculada para os complexos de van der Waals para menos que 3 kcal/mol (para $2.4 < R < 4.5$ Å, complexo VII). Todavia, o RMSD para os complexos contendo ligações de hidrogênio não era significativamente reduzido, a menos que as cargas q_m dos centros MM também fossem otimizadas, principalmente do hidrogênio que participava da ligação.

Nos parâmetros calibrados (tabela 25), as cargas atribuídas ao H e N da formamida (modelo da cadeia principal protéica) são idênticas às cargas do campo de força OPLS-AA

Tabela 25: Parâmetros de LJ e cargas parciais do campo de força OPLS-AA e calibrados (OTM) para complexos bimoleculares modelo do MC nas PTPs^a

OPLS-AA		OTM		OPLS-AA		OTM	
Tipo atômico	σ	ϵ	σ	ϵ	Tipo atômico	q_m	q_m
C aromático	3.550	0.070	3.200	0.110	N formamida	-0.76	-0.76
H aromático	2.420	0.030	1.500	0.010	H formamida	0.38	0.38
C alifático	3.500	0.066	3.320	0.090	N guanidina	-0.80	-0.86
H alifático	2.500	0.030	1.500	0.017	H guanidina	0.46	0.52
P fosfato ^b	3.740	0.200	3.840	0.300	N metilguanidina	-0.70	-0.72
O fosfato	2.960	0.210	2.830	0.100	H metilguanidina	0.44	0.46

^a unidades de σ em Å, ϵ em kcal/mol e q_m em e . ^b parâmetros do campo de força AMBER.

original. As cargas atribuídas ao N e H da guanidina e metilguanidina (modelos da cadeia lateral da ARG) são respectivamente 0.06 e 0.02 e maiores em módulo do que as cargas do OPLS-AA. O aumento das cargas indica que uma maior polarização da região MM é necessária para descrição da ligação de hidrogênio por um potencial híbrido.

Nos parâmetros de LJ calibrados, σ é sempre menor que o valor original OPLS-AA, com exceção do parâmetro para fósforo, indicando que a alta repulsão observada em $R < R_e$ nos perfis obtidos com o potencial híbrido e os parâmetros originais (por exemplo, em $R < 3$, linha tracejada na figura 27) pode ser corrigida pela redução da distância de fundo de poço do potencial de LJ (equação 1.65). Esta diminuição já foi observada em outras otimizações dos parâmetros de LJ atribuídos aos centros QM de potenciais híbridos^[184, 190]. Os valores de energia de complexação mostraram-se pouco dependentes dos parâmetros de LJ atribuídos ao fósforo, já que este átomo encontra-se no interior do grupo fosfato. Os valores de ϵ otimizados são menores que os parâmetros originais para os átomos de hidrogênio e oxigênio, e maiores que os originais para carbono (tabela 25).

O RMSD da energia de complexação de 53 estruturas usadas na calibração dos parâmetros de LJ na fase II e com $|\Delta E_{MP2}^{comp}(R)| \leq |\Delta E_{MP2}^{comp}(R_e)| + 2 \cdot 10^{-2}$ au (p. 143) foi reduzido de 6.1 kcal/mol, usando os parâmetros OPLS-AA, para 4.4 kcal/mol, usando os parâmetros otimizados. A redução no desvio após a calibração da interação não ligante é bem menos significativa do que a redução observada na calibração dos parâmetros da hamiltoniana semiempírica (§3.1.2).

Em comparação aos valores obtidos com os parâmetros OPLS-AA, as curvas de ΔE_{hib}^{comp} utilizando os parâmetros calibrados estão deslocadas para maiores energias em módulo e menores R_e (figuras 29 e 30); as energias de complexação calculadas em $R > R_e$ também são mais próximas à ΔE_{MP2}^{comp} , mas o valor de R_e e a forma da energia de com-

Figura 29: Energia de complexação entre formamida e metafosfato na fase gasosa obtida *ab initio* ou pelo potencial híbrido usando diferentes parâmetros de LJ para parte QM. Na legenda, MP2 indica ΔE_{MP2}^{comp} , OPLS indica ΔE_{hib}^{comp} obtida com os parâmetros de LJ originais OPLS-AA para os átomos QM e OTM indica ΔE_{hib}^{comp} obtida com os parâmetros de LJ otimizados para os átomos QM. Energias para duas orientações estão apresentadas: complexo IV e V, figura 26.

plexação na região de mínimo são mais distantes dos valores *ab initio*. Assim, a principal consequência da otimização foi reduzir a repulsão entre as moléculas, principalmente a curta distância^[184]. Porém, a otimização não foi eficiente na calibração de ΔE^{comp} na região de mínimo. A contribuição dos pontos na região repulsiva ($R < R_e$) dominou o valor da função erro 3.14 e, embora um menor peso W_R para esta região ou diferentes formas da função erro tenham sido testados, os parâmetros obtidos (tabela 25) apresentaram o melhor desempenho. Em diversas tentativas, observamos que a diminuição do desvio de ΔE^{comp} calculada pelo potencial híbrido na região próxima a R_e era possível somente se a região repulsiva fosse comprometida e vice-versa. Portanto, concluímos que um potencial de LJ não é a forma ideal para calcular as interações de van der Waals ($\mathcal{H}_{QM/MM}^{vdW}$) do potencial híbrido.

Nos estudos em que o potencial de LJ foi adotado para $\mathcal{H}_{QM/MM}^{vdW}$ e os parâmetros foram calibrados para sistemas específicos^[120, 167, 190], uma redução significativa do desvio

Figura 30: Energia de complexação entre guanidina e metafosfato na fase gasosa obtida *ab initio* ou pelo potencial híbrido usando diferentes parâmetros de LJ para parte QM. A legenda é a mesma da figura 29. Energias para duas orientações dos complexos estão apresentadas: complexo I e II, figura 26.

na energia de interação QM/MM foi observada apenas nos casos em que σ e ϵ assumiram valores sem relevância física. Por exemplo, os parâmetros otimizados $\epsilon_O = 40.0$ kcal/mol^[190] ou $\epsilon_S = 1.7$ kcal/mol^[120] são respectivamente 100 e 10 vezes maiores que os parâmetros OPLS-AA.

O uso de potenciais como a função de Buckingham, em que um termo exponencial, $1/\exp[\alpha r]$ (onde α é mais um parâmetro ajustável), substitui o termo $1/r^{12}$ da equação 1.65, foi proposto para o cálculo de $\mathcal{H}_{QM/MM}^{vdW}$ ^[175]. Para o dímero de água e o complexo entre água e amônia, as energias de complexação calculadas são bastante satisfatórias e mais próximas de uma referência *ab initio* do que as energias calculadas com uma função de LJ^[175]. Ao contrário do termo $1/r^{12}$, a forma repulsiva do potencial de Buckingham é justificável pela teoria de forças intermoleculares^[122, 172].

Todavia, a implementação de potenciais mais apropriados para cálculo das interações de van der Waals não foi realizada nesta tese, assim, esta interação foi descrita por um potencial de LJ.

A diminuição do desvio nos cálculos com os parâmetros calibrados (tabela 25) deve-se em grande parte à alteração das cargas parciais atribuídas aos átomos de hidrogênio e nitrogênio. Para os complexos entre formamida e metafosfato, estas cargas foram otimizadas, mas os valores finais são idênticos às cargas OPLS-AA. Para guanidina e metilguanidina, as cargas parciais otimizadas são diferentes das OPLS-AA e, portanto, a interação eletrostática MM pura (equação 1.64), por exemplo, entre a cadeia lateral da ARG no P-loop e a cadeia protéica principal das PTPs, representadas na região MM, seria também modificada, com resultados imprevisíveis. O uso apenas dos parâmetros de LJ otimizados não apresenta redução significativa dos desvios calculados em relação ao desvio observado com os parâmetros de LJ do OPLS-AA para os complexos contendo ligações de hidrogênio (vide supra). Portanto, nas simulações enzimáticas (cap. 4), os parâmetros de LJ e as cargas parciais originais OPLS-AA foram usados para interação não ligante. Em alguns casos (por exemplo, complexo IV na figura 29), $\Delta E_{hib}^{comp}(R)$ obtido pelos parâmetros OPLS-AA é mais próximo da energia *ab initio* do que $\Delta E_{hib}^{comp}(R)$ calculado com os parâmetros calibrados. O R_e , a $\Delta E_{hib}^{comp}(R_e)$ e a forma do perfil da energia de complexação na região de mínimo são calculados corretamente quando os parâmetros OPLS-AA são usados. Por exemplo, para o complexo IV, $R_e = 1.80 \text{ \AA}$ e $\Delta E_{MP2}^{comp}(R_e) = 2.81 \cdot 10^{-2}$ au calculados *ab initio*, e $R_e = 1.70 \text{ \AA}$ e $\Delta E_{hib}^{comp}(R_e) = 2.93 \cdot 10^{-2}$ au calculados pelo potencial híbrido (figura 29), ou seja, os desvios em R_e são menores que $0.1 \text{ \AA}$ e os desvios em $\Delta E_{hib}^{comp}(R_e)$ são menores que 1 kcal/mol . Desvios da mesma magnitude em R_e e $\Delta E_{hib}^{comp}(R_e)$ são observados para interação não ligante no potencial híbrido para os complexos contendo ligações de hidrogênio, mesmo quando os parâmetros de LJ são especificamente calibrados^[167, 190]. No entanto, $|\Delta E_{hib}^{comp}|$ obtida com os parâmetros OPLS-AA para os complexos entre o grupo fosfato e formamida ou guanidina é sistematicamente menor que $|\Delta E_{MP2}^{comp}|$ (figuras 29 e 30). Assim, as energias das ligações de hidrogênio entre as partes QM (substrato) e MM (cadeia protéica principal e lateral da ARG) serão sistematicamente subestimadas.

Os perfis de ΔE_{hib}^{comp} calculados para os complexos VI e VIII com os parâmetros OPLS-AA (não mostrados aqui) são semelhantes ao complexo de formamida e metafosfato. Os parâmetros de LJ atribuídos ao oxigênio esterificado do grupo fosfato no complexo VI podem ser os mesmos daqueles atribuídos ao oxigênio de fenol ou oxigênio ionizado de carboxilato (§3.1.3), sem que alterações significativas sejam observadas nas energias de complexação.

Modelos da interação do solvente com o ácido geral e o grupo de saída da reação catalisada por PTPs também foram testados. A diferença entre a ΔE_{hib}^{comp} calculada com os parâmetros de LJ OPLS-AA para parte QM e os valores *ab initio* é menor que 1

Figura 31: Energia de complexação entre água e fenol, e entre água e ácido acético na fase gasosa obtidas *ab initio* ou pelo potencial híbrido. Legenda: ΔE_{MP2}^{comp} (quadrados) e ΔE_{hib}^{comp} obtida com parâmetros OPLS-AA para parte QM (linha contínua) para o complexo entre água e ácido acético; e ΔE_{MP2}^{comp} (círculos) e ΔE_{hib}^{comp} obtida com parâmetros OPLS-AA para parte QM (linha tracejada) para o complexo entre água e fenol. A orientação de ambos complexos está mostrada na figura.

kcal/mol na região do mínimo de potencial (figura 31). Para $R > R_e$, o comportamento é inverso aos complexos entre os modelos do grupo fosfato e da cadeia protéica, pois ΔE_{hib}^{comp} dos complexos com água é superestimado em comparação aos valores *ab initio*. A interação do solvente com o grupo fosfato ou com a cadeia lateral da CYS nucleofílica não foram estudadas, porque ambos os grupos estão localizados no interior do sítio ativo, hidrofóbico e dessolvatado. No caso das CDC25s, o sítio ativo é relativamente mais exposto ao solvente e o grupo de saída deve estar em contato com o solvente na primeira etapa de reação catalisada. No entanto, pela análise das estruturas determinadas por cristalografia e raios-X (§1.3), nenhuma molécula de água penetra no interior do sítio ativo no MC.

Concluimos que, dentro da limitação da descrição das interações de van der Waals entre as regiões QM e MM por um potencial de LJ, os parâmetros OPLS-AA podem ser usados para os centros QM com resultados satisfatórios e da mesma qualidade dos

Tabela 26: Energias relativas calculadas para as reações 3.18 a 3.21

Reagentes	equação ^b	Hamiltoniana ^a		
		MP2	QM	QM/MM
CYS-PO ₃ ²⁻	3.18	-59.2	-51.5	-60.0
SER-PO ₃ ²⁻	3.19	-43.0	-35.0	-45.8
CYS + SER ⁻	3.20	-23.9	-24.5	-30.5
ASP + CH ₃ O ⁻	3.21	-39.1	-35.8	-40.5

^a MP2 é a energia *ab initio*, QM é a energia semiempírica pura e QM/MM é a energia em que a cadeia lateral fosforilada é QM e o restante do aminoácido é MM. ^b equação de reação, p. 145.

resultados obtidos se os parâmetros de LJ atribuídos à parte QM fossem calibrados.

O uso da energia de complexação e da estrutura de dímeros ou complexos intermoleculares isolados para calibração de potenciais que serão utilizados em simulações moleculares tem sido criticado^[189], porque as propriedades (como funções de distribuição radial, observáveis termodinâmicas, etc.) obtidas nas simulações apresentam grandes desvios em comparação com os valores medidos experimentalmente. Neste esquema de calibração, a interação entre muitos-corpos é negligenciada e este efeito é importante na descrição das propriedades das soluções ou solventes macroscópicos. Um esquema de calibração mais eficaz é o ajuste dos parâmetros do potencial (por exemplo, os parâmetros de LJ para parte QM num potencial híbrido) por comparações de valores calculados nas simulações com as observáveis experimentais^[174]. Este último esquema foi adotado em algumas calibrações de potenciais híbridos e pelo menos a energia de solvatação calculada é próxima aos valores experimentais^[120, 188].

3.2.4 Teste das interações de fronteira

As aproximações introduzidas no potencial híbrido para permitir a divisão espacial do sistema molecular em uma região QM e outra MM, ou seja, o esquema de átomo fictício para completar a valência da região QM e as interações covalentes e não ligantes dos átomos na fronteira QM/MM, não foram calibradas. Todavia, apresentamos alguns testes que indicam a natureza e a magnitude das incertezas introduzidas por estas aproximações.

As energias das reações 3.18 e 3.19 equivalem, no sentido inverso, à fosforilação da cadeia lateral de CYS e SER por metafosfato, respectivamente. Estas reações podem ser comparadas à segunda etapa (A_n) da reação catalisada pelas PTPs, se o mecanismo $D_n + A_n$ for utilizado pelas PTPs. A energia de reação calculada pelo potencial híbrido é próxima aos valores *ab initio*. O desvio não ultrapassa 3 kcal/mol, indicando que a

incerteza associada à presença da fronteira QM/MM é desta magnitude (tabela 26). A reação de transferência de H^+ entre ASP e metóxido, 3.21, equivale à doação de H^+ para o grupo de saída se o substrato enzimático for metilfosfato. A diferença entre a energia desta reação calculada pelo potencial híbrido e *ab initio* é menor que 1.5 kcal/mol (tabela 26). Usamos metóxido como grupo de saída ao invés de fenóxido, porque a basicidade calculada para fenóxido pelo conjunto MNDO+G/CHOPS está artificialmente aumentada (p. 156).

Por outro lado, os resultados obtidos com o potencial híbrido para as reações envolvendo dissociação de $CYS-PO_3H^-$ (equação 3.16) e $SER-PO_3H^-$ (3.17) ou para transferência de H^+ entre CYS e SER (3.20) apresentam desvios mais elevados em comparação com a referência *ab initio* (figura 32). As reações 3.16 e 3.17 podem ser comparadas às reações de dissociação de $MeSP^-$ e $MeOP^-$, respectivamente e, portanto, são dominadas pela transferência de H^+ (§2.2.2). As maiores diferenças entre as energias *ab initio* e QM/MM ocorrem nas espécies em que H^+ é transferido do grupo fosfato para a cadeia lateral do aminoácido (figura 32). Para a reação de $CYS-PO_3H^-$, o desvio em relação aos valores *ab initio* é menor que 2 kcal/mol para todas espécies, exceto a última [$d(SP) = 3.38 \text{ \AA}$], em que H^+ é transferido para CYS e o desvio chega a 6 kcal/mol; para reação de $SER-PO_3H^-$, o maior desvio corresponde à $d(PO) = 2.45 \text{ \AA}$, a espécie em que H^+ é transferido para SER. Portanto, sugerimos que os desvios observados nas computações com o potencial híbrido para as reações 3.16, 3.17 e 3.20 são consequência de um desbalanceamento nas basicidades das espécies em razão da introdução da fronteira QM/MM e do uso do método semiempírico MNDO+G/CHOPS (§3.2.1).

Os resultados obtidos com todo sistema tratado pela hamiltoniana QM semiempírica apresentam altos desvios (até 10 kcal/mol, figura 32) para as reações 3.16 a 3.19 e 3.21, porque:

- os parâmetros semiempíricos para nitrogênio não foram especificamente calibrados;
- os parâmetros MNDO+G/CHOPS são usados em ambientes químicos ou funções orgânicas diferentes daquelas usadas na calibração específica;
- a mistura dos parâmetros AM1 para N e MNDO+G/CHOPS para o restante dos átomos não foi calibrada.

A energia da reação 3.20 calculada pela hamiltoniana QM semiempírica é próxima à *ab initio*, mas este acordo deve ser resultado de um cancelamento de erro, já que todas outras reações contendo SER e CYS apresentam desvios elevados.

Figura 32: Energias relativas de dissociação calculadas para CYS- PO_3H^- (A) e SER- PO_3H^- (B) na fase gasosa. Na legenda, MP2 é a energia calculada *ab initio*, QM é a energia calculada pela hamiltoniana semiempírica pura (MNDO+G/CHOPS para C, H, O, P e S, e AM1 para N), e QM/MM é a energia calculada pela hamiltoniana híbrida em que a cadeia lateral fosforilada é QM e o restante do aminoácido é MM. O zero de energia corresponde à energia de CYS- PO_3H^- ou SER- PO_3H^- na geometria totalmente otimizada.

Concluimos que a terminação de fronteira pelo método do átomo fictício e o esquema adotado para computação das interações na fronteira QM/MM não deve introduzir erros significativos nas reações de ataque nucleofílico da cadeia lateral de CYS ou SER ao grupo fosfato, ou na reação de transferência de H^+ do ASP para o grupo de saída. Estes dois processos são os mais relevantes para a simulação das reações catalisadas pelas PTPs. No entanto, as reações que envolvem transferência de H^+ da cadeia lateral de CYS ou SER devem ser analisadas com mais cuidado, pois o potencial QM/MM calibrado introduz incertezas de cerca de 6 a 8 kcal/mol.

4 *Simulação das reações catalisadas por PTPs*

Definimos *mecanismo de reação enzimática* como a seqüência de transformações (quebra e formação de ligações químicas) e mudanças de estrutura do complexo enzimático (que impliquem em mudanças significativas das interações intermoleculares) observadas ao longo do progresso da reação no sítio ativo. Esta definição é diferente do *mecanismo de catálise enzimática*, apresentado na §1.1 para descrever as forças ou interações microscópicas utilizadas pelas enzimas para acelerar a velocidade de reações.

Simulações computacionais de reações enzimáticas podem contribuir para a compreensão dos mecanismos de catálise (§1.1). Outra aplicação, menos ambiciosa, mas igualmente útil, é simular os possíveis mecanismos de reação e, a partir de comparações das quantidades calculadas com observáveis experimentais, decidir qual destes possíveis mecanismos é energeticamente viável^[9]. Algumas vezes nem ao menos a natureza dos grupos que participam da reação catalisada é conhecida exatamente (por exemplo, o estado de protonação do substrato pode ser desconhecido ou alterado pelo sítio enzimático) e simulações também podem ser usadas para discernir entre as possibilidades.

Neste capítulo, apresentamos as simulações computacionais, utilizando a hamiltoniana híbrida calibrada e testada no cap. 3, da primeira etapa da reação catalisada pelas proteínas tirosina-fosfatases de dupla especificidade, VHR e CDC25B (equação 1.9). Nosso objetivo foi responder as perguntas formuladas ao final da §1.3 (p. 44). Portanto, não investigamos os mecanismos moleculares usados pelas PTPs para acelerar a velocidade de desfosforilação do substrato nem simulamos a reação de referência em solução aquosa, na ausência da enzima (§1.1.2). As simulações foram conduzidas para enzimas selvagens e mutantes, com diferentes propostas de ácido geral, nucleófilo enzimático, grupo de saída no substrato e estados de protonação dos grupos reativos (substrato e nucleófilo enzimático). A partir da comparação dos perfis de energia livre calculados com observáveis experimentais, esperamos responder as questões de interesse, mas também indicar qual o mecanismo da reação de desfosforilação no sítio ativo das PTPs.

4.1 Métodos

4.1.1 Construção das estruturas enzimáticas

Os modelos estruturais usados nas simulações enzimáticas foram construídos a partir das estruturas tridimensionais cristalográficas das proteínas obtidas no Banco de dados de proteínas (PDB)^[255].

A VHR humana foi cristalizada (código PDB 1vhr)^[88] na presença de HEPES [4-(2-hidroxietil)-1-piperazina ácido etanosulfônico], um éster de sulfato, que se liga ao P-loop analogamente ao substrato natural. A proteína foi cristalizada sem os 7 primeiros aminoácidos (apenas os resíduos numerados de 8 a 185 estão presentes na estrutura do PDB), mas engloba toda unidade catalítica. O resíduo que funciona como ácido geral é o ASP92 e o P-loop (§1.3) é: HIS123–**CYS124**–ARG125–GLU126–GLY127–TYR128–SER129–**ARG130**–SER131 onde os aminoácidos estão representados pelo código de três letras e os resíduos em negrito indicam a cisteína e a arginina conservadas na família das PTPs. Duas moléculas protéicas estão presentes na estrutura código 1vhr. Apenas a cadeia A (complexada a HEPES) foi usada. A estrutura foi determinada numa resolução de 2.1 Å.

A CDC25B humana foi cristalizada (código PDB 1qb0)^[107] na presença de íons sulfato (SO_4^{2-}), que também se ligam ao sítio ativo como o substrato natural. Apenas a unidade catalítica da enzima foi cristalizada (resíduos numerados de 374 a 550 estão presentes na estrutura do PDB). O P-loop é composto por: HIS472–**CYS473**–GLU474–PHE475–SER476–SER477–GLU478–**ARG479**–GLY480. A estrutura foi determinada em 1.9 Å de resolução.

As duas estruturas não apresentaram mutações e ambos fragmentos têm atividade catalítica comprovada na hidrólise de ésteres de fosfatos^[88, 93, 107]. Todas as águas cristalográficas e outras moléculas diferentes da cadeia protéica, e do análogo do substrato complexado no P-loop foram removidas da estrutura obtida no PDB.

A construção das posições dos hidrogênios ausentes e do substrato para a VHR e para CDC25B foram ligeiramente diferentes. Fenilfosfato foi usado como modelo da fosfotirosina de proteínas fosforiladas, que são os substratos naturais das PTPs (§1.3.1). Para a VHR, Dillet, Van Etten e Bashford^[119] construíram as coordenadas dos hidrogênios, substituíram o HEPES por fenilfosfato dianiônico e otimizaram somente as posições dos hidrogênios, do substrato, e do enxofre (ionizado) da cadeia lateral da CYS124, usando um

algoritmo de otimização por gradiente conjugado (§1.5) e o campo de força CHARMM^[256] para descrever a energia de todo sistema. O restante da proteína foi fixada nas posições originais dadas no PDB. As coordenadas otimizadas foram disponibilizadas pelos autores^[119] e usadas nesta tese. Para a CDC25B, as coordenadas dos hidrogênios ausentes foram construídas e fenilfosfato dianiônico foi sobreposto às coordenadas do sulfato ligado ao sítio ativo (o fósforo foi sobreposto ao enxofre e os oxigênios do fosfato foram sobrepostos aos oxigênios do sulfato, de modo que o grupo fenil se localizasse no exterior do sítio ativo). Somente as posições dos hidrogênios e do fenilfosfato foram otimizadas, usando um algoritmo de otimização por gradiente conjugado até que a média quadrática do gradiente fosse inferior a $0.001 \text{ kJ mol}^{-1} \text{ \AA}^{-1}$. O restante da proteína foi fixada nas posições originais dadas na estrutura do PDB. O campo de força OPLS-AA foi usado para descrever a energia de todo sistema. Os parâmetros de LJ do campo de força AMBER foram usados para o fósforo (tabela 25) e os parâmetros de oxigênio ionizado do grupo carboxilato no OPLS-AA foram atribuídos aos oxigênios não ligantes do grupo fosfato (p. 144, §3.1.3.3). Os parâmetros necessários para cálculo de $\mathcal{V}_{cov}$ (§1.6.1) dos termos que envolviam fósforo ou oxigênio não ligante também foram tirados do AMBER. Os parâmetros covalentes e de LJ do OPLS-AA para enxofre da CYS foram usados para o enxofre ionizado da CYS no P-loop e os parâmetros de oxigênio ligado ao anel aromático da tirosina foram atribuídos ao oxigênio esterificado do fenilfosfato. As seguintes cargas parciais foram atribuídas: $q_m = 0.895 e$ para o fósforo, $q_m = -0.82 e$ para os oxigênios não ligantes, $q_m = -0.90 e$ para o enxofre ionizado e $q_m = -0.22 e$ para o C_β da CYS no P-loop. Estas cargas parciais são bastante próximas àsquelas usadas em outros estudos e simulações de PTPs com fenilfosfato dianiônico^[110, 114, 119, 120], e são consistentes com uma análise populacional^[124, p. 151 e 152] derivada do potencial eletrostático gerado por compostos modelo e calculado no nível MP2/6-311+G(2df,2p) (resultado não mostrado aqui). A construção dos hidrogênios e otimização das coordenadas foram realizadas pela biblioteca de rotinas DYNAMO^[179].

As estruturas finais das duas proteínas complexadas com fenilfosfato são equivalentes aos respectivos MCs no vácuo (figura 10). Nas etapas seguintes de construção do modelo, o tratamento dado às duas proteínas foi idêntico. Os parâmetros do campo de força descritos acima e atribuídos para CDC25B foram mantidos e os mesmos parâmetros foram usados na VHR. Todos os resíduos ARG, LYS e HIS foram protonados em ambas proteínas, em acordo com o estado de protonação adotado em outras simulações de PTPs^[110, 112, 114]. A protonação da HIS pode ser discutida, porque o pK_a da cadeia lateral deste aminoácido livre é ~ 6.0 . Todavia, medidas experimentais (espectroscópicas e cinéticas, §1.3.3) indi-

cam que os resíduos de HIS, como aquele que forma o P-loop na VHR^[257] ou em outras PTPs^[42, 258], ou mesmo resíduos de HIS sem função catalítica definida e não conservados na família das PTPs têm seu pK_a aumentado^[259] e devem estar protonados no pH ótimo. Este pH é ≈ 5.5 para VHR^[103], portanto, a maior parte dos resíduos de HIS que não tiveram seu pK_a alterado também deve estar protonada. Mutações da HIS do P-loop (como a HIS123 na VHR)^[257] resultam em diminuições de menos de 10 vezes em k_{cat} e o KIE medido para a reação catalisada pelas PTPs não sofre alteração com mudança do pH (p. 40). Portanto, parece-nos que a protonação dos resíduos de HIS é adequada e não influencia significativamente as propriedades catalíticas das PTPs. Os resíduos GLU e ASP foram desprotonados, exceto o ASP92 na VHR, que funciona como ácido geral. Na CDC25B, os resíduos GLU474 e GLU478 foram modelados na forma desprotonada, exceto nas simulações em que cada um dos grupos foi aproximado ao grupo de saída do substrato ou funcionou como um ácido geral (simulações 11 e 14, tabela 27). Estes estados de protonação são idênticos àqueles adotados por Asthagiri et al.^[114] e similares aos estados adotados por Åqvist et al.^[112, 113, 120] na simulação da catálise de uma LM-PTP. A carga total de ambas proteínas (excluindo a carga do substrato) foi de +5 *e*.

Para solvatar os MCs, cada proteína foi sobreposta a uma caixa de água cúbica pré-equilibrada, com dimensões de 55.92 Å de lado, contendo 5832 moléculas de água, representadas pelo potencial flexível TIP3P^[247]. Estes valores correspondem a uma densidade de ≈ 1.0 g/L para água. A caixa foi construída a partir de replicações e translações de outra caixa pré-equilibrada, contendo 729 moléculas de água e 27.96 Å de lado, e foi equilibrada a 300 K por 200 ps de dinâmica molecular segundo um algoritmo de velocidade de Verlet-Langevin com intervalo $\tau = 1$ fs e coeficiente de fricção $\gamma = 10 \text{ ps}^{-1}$ ($\gamma_i = \gamma$, qualquer que seja *i*, p. 69), usando condições de contorno periódicas e um esquema de truncagem para as interações não ligantes com mudança da força com $r_{on} = 8$ Å e $r_{off} = 12$ Å (§1.6.2). A proteína foi centralizada na caixa de solvente e todas as moléculas de água sobrepostas ou localizadas a menos que 2.8 Å de algum átomo do MC foram removidas. Como a VHR e a CDC25B têm dimensões de cerca de 50 por 40 por 32 Å e 51 por 36 por 28 Å, respectivamente, toda proteína foi solvatada. Uma otimização de geometria por gradiente conjugado (até a média quadrática do gradiente ser inferior a 0.1 kJ mol⁻¹ Å⁻¹) e uma curta dinâmica molecular (usando o campo de força e as condições para a dinâmica da caixa de água descritos acima) de 500 fs, com os átomos do MC fixos, foram realizadas para relaxar contatos energeticamente desfavoráveis entre o solvente e a proteína. O processo de sobreposição e relaxação das moléculas de água foi repetido até que nenhuma molécula de água pudesse ser adicionada, o que ocorreu após dois ciclos

de sobreposição para ambas enzimas. O sistema solvatado para VHR apresenta 17197 átomos, sendo 178 resíduos protéicos, 1 substrato (16 átomos) e 4804 moléculas de água. Para CDC25B, o sistema final contém 17212 átomos, sendo 177 resíduos, 1 substrato e 4756 moléculas de água.

Em seguida, as posições da proteína e do solvente foram gradualmente relaxadas e termalizadas por simulações de dinâmica molecular (usando as mesmas condições descritas acima para o campo de força e para a dinâmica). As posições da proteína e do solvente foram restritas às coordenadas cartesianas iniciais por potenciais harmônicos (opção “TETHER”^[24, 179], como as equações 1.61 ou 1.78). Entre 10 a 12 simulações de 1 ps foram realizadas, em que as constantes de força usadas nestes potenciais harmônicos foram gradualmente reduzidas de $1000 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ até zero. O RMSD da temperatura (em relação à $T = 300 \text{ K}$) foi de 2 K, sugerindo que o sistema estava equilibrado após as simulações. Procedimentos similares de equilibração foram adotados na construção dos modelos macromoleculares solvatados usados em outros estudos de reatividade enzimática^[120, 260, 261]

O sistema solvatado foi centralizado no ponto médio entre as coordenadas do fósforo do fenilfosfato e do enxofre da CYS nucleofílica (§1.3). As posições dos resíduos da proteína ou das moléculas de água com pelo menos um de seus átomos localizados a mais de $14 \text{ \AA}$ deste ponto médio foram fixadas (usando restrições “duras” ao invés de potenciais harmônicos, isto é, o gradiente em relação às coordenadas destas espécies foi zerado em cada passo de simulação). A distância dos átomos que se movem (ou seja, dentro da esfera de $14 \text{ \AA}$ de raio, figura 33) aos átomos congelados no exterior da caixa de solvatação (ou seja, próximos à lateral da caixa cúbica idealizada, vide supra) foi sempre maior que $13 \text{ \AA}$ (correspondente ao r_{off} usado nas simulações para obtenção dos PMFs, §4.1.2). Embora os átomos da parte fixa não se movam, eles interagem normalmente (por interações covalentes, eletrostáticas e de LJ) com os átomos da região central móvel. O congelamento das posições dos centros distantes do sítio ativo é usado para diminuir o custo computacional e também foi empregado em estudos com PTPs^[110, 112, 115, 120] e outras enzimas^[260, 261].

A região central móvel foi re-equilibrada a 300 K após o congelamento da parte exterior (figura 33) em 5 simulações de dinâmica molecular de 2 ps (vide supra). Restrições harmônicas foram impostas às coordenadas cartesianas e gradativamente relaxadas (reduzindo as constantes de força de $400 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ até zero). Finalmente, 10 ps de dinâmica molecular foram computados sem restrições às coordenadas da parte móvel (figura 33). O

Figura 33: Representação das diferentes regiões usadas nas simulações enzimáticas. Para VHR, as partes fixa e móvel contêm 16310 e 887 átomos, respectivamente. Para CDC25B, as partes fixa e móvel contêm 16424 e 728 átomos, respectivamente.

RMSD da temperatura nestes 10 ps foi de 8 K. Este desvio é maior que aquele observado acima em virtude do congelamento de parte das posições atômicas (figura 33). Para a VHR, o RMSD das coordenadas cartesianas obtidas ao final da etapa de construção e equilíbrio em comparação às coordenadas obtidas da estrutura original do PDB, incluindo as cadeias laterais dos aminoácidos e o substrato é de 0.68 Å. Para CDC25B, este RMSD é de 0.55 Å. Estas foram as estruturas iniciais usadas em cada umas das simulações apresentadas na §4.2. Na simulação de mutantes ou de diferentes substratos (por exemplo, simulação 2 ou 3, tabela 27), as posições dos novos centros foram introduzidas ou substituídas (por exemplo, o $C_\beta H_3$ da cadeia lateral da ALA foi colocado nas posições do $C_\beta H_2 C$ da cadeia lateral do ASP92 e os oxigênios do grupo carboxilato foram apagados na simulação 3) e uma otimização de geometria por gradiente conjugado (até a que média quadrática do gradiente fosse inferior a $0.1 \text{ kJ mol}^{-1} \text{ Å}^{-1}$) foi realizada apenas para os novos centros, com o restante do sistema restrito por um potencial harmônico ($k_m = 500 \text{ kJ mol}^{-1} \text{ Å}^{-2}$) às coordenadas cartesianas originais.

4.1.2 Potencial híbrido e dinâmica molecular

Nas simulações com o potencial híbrido (§1.6.4), a partição das regiões quântica (QM) e clássica (MM) foi alterada de acordo com o sistema ou processo estudado (figura 34 e tabela 27). Os átomos do substrato pertenceram à região QM em todas simulações realizadas (exceto a simulação 11, tabela 27). Na simulação 2, o substrato foi metilfosfato

Figura 34: Esquema das distâncias e grupos reativos usados nas simulações enzimáticas.

Tabela 27: Grupos reativos ou catalíticos e coordenadas de reação usadas nas simulações enzimáticas

Simulação	X ^a	Y ^a	Observação ^b	QM ^c	ξ ^d
VHR					
1a	S ⁻	O ⁻		29	d(PO) – d(SP)
1b	S ⁻	O ⁻		29	d(PO)
1c	S ⁻	O ⁻		29	d(PO) – d(OH)
1d	S ⁻	O ⁻		29	d(OH)
1e	S ⁻	O ⁻		29	d(PO) – d(SP) – d(OH)
2	S ⁻	O ⁻	lg = metil	22	d(PO) – d(SP)
3	S ⁻	O ⁻	ASP→ALA	21	d(PO) – d(SP)
4	S ⁻	O ⁻	ARG→ALA	29	d(PO) – d(SP)
5	S ⁻	OH		30	d(PO) – d(SP)
6	SH	O ⁻		30	d(PO) – d(SP)
7	O ⁻	O ⁻	CYS→SER	29	d(PO) – d(OP)
8	O ⁻	OH	CYS→SER	30	d(PO) – d(OP)
9	OH	O ⁻	CYS→SER	30	d(PO) – d(OP)
10	OH	OH	CYS→SER	31	d(PO) – d(OP)
CDC25B					
11	S ⁻	O ⁻	Ác. geral = GLU	0	d(OH)
12	S ⁻	O ⁻	Sem Ác. geral	21	d(PO) – d(SP)
13	S ⁻	OH	Sem Ác. geral	22	d(PO) – d(SP)
14a	S ⁻	O ⁻	Ác. geral = GLU	29	d(PO) – d(SP)
14b	S ⁻	O ⁻	Ác. geral = GLU	29	d(PO) – d(OH)
14c	S ⁻	O ⁻	Ác. geral = GLU	29	d(PO) – d(SP) – d(OH)
14d	S ⁻	O ⁻	Ác. geral = GLU	29	d(OH)

^a indicam os grupos X e Y da figura 34. ^b indica mudanças do grupo de saída (lg = fenil para todas simulações exceto a 2, figura 34), mutações (por exemplo, CYS→SER) ou ácido geral usado, no caso da CDC25B. ^c indica o número de átomos que pertenceram a região QM. ^d coordenada de reação (veja a figura 34 para definição das distâncias de ligação usadas).

dianiônico e, nas outras simulações, o substrato foi fenilfosfato mono ou dianiônico (tabela 27). Metilfosfato foi usado como modelo da fosfoserina ou fosfotreonina de proteínas fosforiladas, que também são os substratos naturais das DSPs e CDC25s (§1.3.1). O nucleófilo enzimático corresponde à cadeia lateral da cisteína (CYS124 para VHR e CYS473 para CDC25B, tabela 27) ou da serina para o mutante CYS124→SER. Assim, os átomos $C_\beta H_2-X^-$ ou $C_\beta H_2-XH$, onde $X = S$ ou O , respectivamente para CYS ou SER, foram representados na região QM. Na VHR, o ácido geral enzimático correspondeu à cadeia lateral do ASP92 e os átomos $C_\beta H_2-COOH$ estiveram na região QM (figura 34). Na CDC25B, a cadeia lateral dos resíduos GLU474 e GLU478 foram testadas como possíveis ácidos gerais (§1.3.3) e os átomos $C_\gamma H_2-COOH$ estiveram na região QM. Na simulação 3 (tabela 27), a cadeia lateral da alanina mutante pertenceu à região MM. A carga total da região QM foi -3 e nas simulações 1 a 4, 7, 12 e 14; -2 e nas simulações 5, 6, 8, 9 e 13; e -1 e na simulação 9. Todos os átomos foram descritos pelo potencial MM na simulação 11.

Os outros átomos do MC solvatado foram representados na região MM. Portanto, excluindo-se os átomos da cadeia lateral relacionados acima, o restante do aminoácido que teve função de nucleófilo ou de ácido geral foi representado na região MM. Na VHR, a fronteira entre as regiões QM e MM foi definida entre a ligação $C_\alpha-C_\beta$ para o nucleófilo e para o ácido geral. Na CDC25B, a fronteira foi definida entre a ligação $C_\alpha-C_\beta$ para o nucleófilo e na ligação $C_\beta-C_\gamma$ para o ácido geral. O esquema de átomo fictício, como descrito em §1.6.4.2 e §3.1.4, foi empregado para saturar a valência da região QM. Os átomos fictícios necessários (um para o nucleófilo enzimático e outro para o ácido geral) foram posicionados automaticamente (equação 3.15), em cada passo da dinâmica molecular (vide infra).

Assim, as moléculas que formaram a região QM nas simulações foram idênticas ou muito semelhantes às moléculas usadas na calibração específica da hamiltoniana semiempírica MNDO+G/CHOPS (§3.1.2). Por exemplo, a parte QM na simulação 1a (tabela 27) é composta por $H_3C-S^- + PhOP^{2-}$, que pode ser comparada à reação F (cap. 2), e por $H_3C-COOH$, cuja transferência de H^+ para o grupo de saída é idêntica à reação 2.1 (p. 103).

A hamiltoniana semiempírica calibrada MNDO+G/CHOPS (§3.2.1) foi usada para descrição da parte QM, o campo de força OPLS-AA foi usado para a descrição da parte MM da proteína e o potencial TIP3P foi usado para a água. A descrição geral do potencial híbrido é aquela apresentada em §1.6.4. A interação eletrostática QM/MM segundo a

proposta de Field et al. foi usada (§1.6.4.1 e §3.2.3). Os parâmetros de LJ do OPLS-AA (e do AMBER para fósforo do grupo fosfato) originais foram usados para os átomos QM (§3.2.3). As interações covalentes e não ligantes na fronteira foram idênticas às aquelas definidas em §3.1.4.

A biblioteca de rotinas DYNAMO foi usada em todas as simulações enzimáticas. Embora os parâmetros MNDO+G/CHOPS tenham sido calibrados utilizando o programa GEOMOP (§3.1.2), cálculos de energia e otimizações de geometria com a hamiltoniana semiempírica (QM pura) para espécies usadas no conjunto de treinamento da calibração demonstraram (resultados não mostrados aqui) que os valores obtidos com o conjunto MNDO+G/CHOPS e a biblioteca DYNAMO são essencialmente idênticos aos resultados dos mesmos cálculos realizados com o programa GEOMOP (as diferenças nas energias calculadas não ultrapassaram 10^{-5} au e podem ser atribuídas a diferenças no arredondamento das constantes ou da precisão numérica usada em cada programa).

Dinâmica molecular segundo um algoritmo de velocidade de Verlet-Langevin foi usada para propagar as coordenadas do sistema enzimático no tempo (§1.6.2). Um intervalo de tempo $\tau = 1$ fs e um coeficiente de fricção $\gamma = 10$ ps $^{-1}$ foram usados em todas as simulações. O esquema de truncagem com mudança da força com $r_{on} = 10$ Å e $r_{off} = 13$ Å foi usado para as interações não ligantes (tanto para as interações entre dois centros MM, como entre centros MM e QM, §1.6.1 e apêndice A). Testes em outros sistemas mostraram que os resultados obtidos com este esquema de truncagem são equivalentes aos resultados obtidos sem truncagem das interações não ligantes^[260, 261] (§4.3.1). Nenhuma condição de contorno periódica foi usada, já que as moléculas no exterior da caixa de solvatação tiveram suas coordenadas fixadas (§4.1.1).

4.1.3 Potencial de força média

Os perfis de energia livre ao longo de diferentes coordenadas de reação são os PMFs calculados pelo método de amostragem “guarda-chuva” (§1.6.3).

As coordenadas de reação escolhidas, ξ , foram distâncias interatômicas, isto é, $\xi = d(XY)$, ou combinações de distâncias, $\xi = d(XY) - d(YZ)$, onde X, Y e Z são centros atômicos do sistema reativo. As diferentes coordenadas de reação escolhidas estão relacionadas na tabela 27. Um potencial guarda-chuva harmônico (equação 1.78) é introduzido para ampliar a amostragem da coordenada de reação numa região ou janela próxima a uma referência constante ξ_i (§1.6.3). Os valores usados para k_{umb} foram 1500 e 2000 kJ mol $^{-1}$ Å $^{-2}$ e os valores de referência ξ_i foram igualmente espaçados de 0.10 ou 0.15 Å ao

longo da coordenada de reação de interesse.

Uma série de simulações de dinâmica molecular (segundo as condições descritas na §4.1.2) foi realizada na presença do potencial guarda-chuva com diferentes valores de referência ξ_i . As simulações foram sequenciais e partiram da estrutura do MC obtido na §4.1.1. A estrutura final da dinâmica molecular em uma janela (definida por ξ_i) foi usada como estrutura inicial na janela adjacente ($\xi_{i\pm1}$). A dinâmica molecular em cada janela foi evoluída por 4 ps na fase de equilíbrio e 10 ps na fase de coleta de dados. Estes tempos mostraram-se suficientes para obtenção de PMFs convergidos (§4.3.1). O objetivo destas simulações foi amostrar o intervalo completo de ξ pelo qual o perfil de energia livre era procurado^[260, 261]. O PMF foi determinado pelo método WHAM (§1.6.3), a partir dos dados de ocorrência de ξ durante a dinâmica molecular, agrupando-se as janelas e retirando-se a influência do potencial guarda-chuva. Apenas PMFs unidimensionais $[\mathcal{W}(\xi)]$ foram determinados.

Todos PMFs foram obtidos com a biblioteca DYNAMO e os computadores PCs Pentium IV, usando o sistema operacional GNU/Linux Debian^[209, 210]. Nestas máquinas, 1 ps de dinâmica molecular, por exemplo, para simulação 1a (tabela 27), dura cerca de 15 minutos de processamento.

4.2 Resultados

Todas as simulações referem-se à primeira etapa (equação 1.9) observada na reação de desfosforilação catalisada por PTPs. A denominação usada para as simulações (por exemplo, simulação 1b, simulação 10, etc.) foi definida na tabela 27 e usada na legenda das figuras e no texto abaixo. O progresso da reação ocorre na direção da esquerda para direita em todas figuras de perfil de energia livre apresentadas, exceto nos casos indicados.

Os reagentes foram definidos como os mínimos dos perfis de energia livre na região do MC entre a enzima e o substrato. Estes mínimos são os zeros de energia livre relativa. Os produtos, quando obtidos, correspondem aos mínimos dos perfis de energia livre na região do complexo do grupo de saída com o sítio ativo da PTP fosforilada. Os TSs foram definidos como os máximos dos perfis de energia livre entre os reagentes e produtos.

Os PMFs das reações enzimáticas foram avaliados por uma amostragem do espaço de configurações, portanto, os reagentes, TSs e produtos obtidos não são definidos por uma única posição $\mathbf{R}$ das coordenadas nucleares do sistema. Os valores de distância descritos abaixo são valores médios (ou aproximados) das distâncias observadas em uma trajetória

(ou conjunto de configurações) numa região ou janela da simulação de dinâmica molecular.

4.2.1 Simulações da fosfatase de proteínas VHR

4.2.1.1 Possíveis coordenadas de reação

Já que a escolha de ξ não é única (§4.3.2)^[262, 263], diferentes propostas foram testadas para reação de fenilfosfato dianiônico catalisada pela VHR (figura 35). Esta reação, equivalente à simulação 1 (tabela 27), foi escolhida, porque representa o estado de protonação tradicionalmente aceito para os grupos reativos durante a catálise nas PTPs (§1.3).

Embora as posições iniciais usadas em cada simulação sejam as mesmas (§4.1.1) e os reagentes sejam todos MC, as posições dos átomos nos reagentes são ligeiramente diferentes para cada definição de ξ , já que os mínimos no perfil de energia livre são diferentes. A reação foi acompanhada até o aumento da energia livre no lado dos produtos somente nas simulações 1a e 1d. As coordenadas de reação observadas para os produtos na reação 1a e 1d são respectivamente $\xi_{prod} = 2.33 \text{ \AA}$ e $\xi_{prod} = 1.02 \text{ \AA}$. Em todas simulações 1, a região dos produtos (algumas janelas após o TS, nas regiões em que a energia livre está diminuindo) corresponde ao fenol desfosforilado, protonado e complexado ao sítio ativo, e à enzima tiofosforilada com o ASP92 ionizado (H^+ foi doado para o fenol).

Se a $d(OH)$ (figura 34) é incluída em ξ (simulações 1c, 1d e 1e), o H^+ do ASP92 é transferido inicialmente (em $\xi < \xi_{TS}$, tabela 30) para um dos oxigênios não ligantes do grupo fosfato e depois para o oxigênio do grupo de saída. Nas simulações 1a e 1b, $d(OH)$ não compõe ξ e o H^+ é transferido direta e espontaneamente para o oxigênio do grupo de saída na região do TS ($\xi \approx \xi_{TS}$). Somente um máximo (ou TS) no perfil de energia livre é obtido em cada uma das simulações. Portanto, o mecanismo observado é sempre A_nD_n (figura 3), sem formação de fosforana ou metafosfato estáveis. Nas simulações 1c e 1e, o mecanismo tem maior caráter associativo, ou seja, ocorre maior participação do nucleófilo e menor quebra da ligação $P-O$ [$d(SP) \sim 2.5 \text{ \AA}$ em ξ_{TS}]. Nas simulações 1a, 1b e 1e, a participação do nucleófilo é reduzida, o TS tem caráter dissociativo, como metafosfato, e $d(SP) > 2.7 \text{ \AA}$ em ξ_{TS} .

As energias de ativação calculadas variam em até 15 kcal/mol (tabela 28). A menor barreira corresponde à $\xi = d(PO) - d(SP)$ (simulação 1a), portanto (§4.3.2), esta coordenada foi usada nas simulações apresentadas nas próximas seções.

Figura 35: Perfil de energia livre para diferentes escolhas da coordenada de reação para desfosforilação de PhOP^{2-} catalisada pela VHR. A legenda mostra as simulações realizadas usando as definições indicadas na tabela 27. A reação de desfosforilação do substrato corresponde à direção da esquerda para direita em todos perfis, exceto o perfil 1d, onde a desfosforilação corresponde à direção da direita para esquerda.

Tabela 28: Energia livre de ativação e coordenadas de reação dos reagentes e TSs calculadas para as diferentes escolhas de ξ^a

Simulação	ξ_{reac}	$\Delta G^\ddagger$	ξ_{TS}
1a	-1.86	16.4	-0.36
1b	1.75	18.7	2.46
1c	-2.22	26.2	0.73
1d	4.26	31.7	1.32
1e	-5.61	26.6	-1.71

^a coordenada de reação, veja tabela 27 para definições. ξ_{reac} é a coordenada de reação dos reagentes e ξ_{TS} é a coordenada de reação dos TSs. $\Delta G^\ddagger$ em kcal/mol e ξ em Å.

4.2.1.2 Fenilfosfato ou metilfosfato como substrato

A VHR é uma enzima de dupla especificidade e, logo, desfosforila resíduos de aril e alquilderivados de fosfato. Assim, a desfosforilação de metilfosfato dianiônico catalisada pela VHR foi simulada (2, tabela 27). A coordenada de reação usada foi $\xi = d(\text{PO}) - d(\text{SP})$ e o perfil de energia livre mostrado para a reação de PhOP^{2-} na figura 36 é o mesmo daquele mostrado na figura 35 (simulação 1a).

Para ambos substratos, os oxigênios não ligantes do grupo fosfato formam ligações de hidrogênio bidentadas com a cadeia lateral da ARG130 (figura 10) ao longo de praticamente toda coordenada de reação. O grupo de saída na reação de fenilfosfato é protonado pelo ASP92 na região do TS e, na reação de metilfosfato, a protonação ocorre após a região do TS (em $\xi > \xi_{TS}$), mas $d(\text{OH}) \sim 3 \text{ \AA}$ em ξ_{TS} . O produto de ambas reações é a VHR tiofosforilada com o grupo de saída protonado e complexado ao sítio ativo. As distâncias médias observadas para o MC (reagente), complexo dos produtos e TS da reação de PhOP^{2-} catalisada pela VHR (simulação 1a), foram calculadas a partir das trajetórias de dinâmica molecular nas respectivas janelas do PMF e são apresentadas na tabela 29.

O TS obtido na reação de PhOP^{2-} catalisada pela VHR (figura 37) pode ser comparado àquele observado para a tiólise de ésteres de fosfato na fase gasosa (§2.2), pois tem estrutura de uma TBP (figura 4), em que os grupos axiais são os átomos de enxofre da CYS124 e o oxigênio do grupo de saída (O1, o ângulo S–P–O1 é $168.3 \pm 4.4^\circ$), e os grupos equatoriais são os oxigênios não ligantes do fosfato (O2, O3 e O4, figura 37). Os oxigênios não ligantes e o fósforo estão aproximadamente coplanares, como no metafosfato livre (§2.2.2). Na figura 37, mostramos uma configuração de dinâmica molecular na região do TS após a transferência de H^+ .

A altura da barreira calculada para reação de fenilfosfato (simulação 1a) é 16.4 kcal/mol (tabela 30), bastante próxima ao valor medido experimentalmente, 15.5 kcal/mol, para a primeira etapa de reação catalisada pela VHR (equação 1.9). A energia de ativação foi obtida pela equação 1.7 [com $\kappa(T) = 1$], a partir da constante de velocidade k_2 medida para catálise pela VHR (tabela 2). A 30 °C, $k_{cat} = 5.8 \pm 0.2 \text{ s}^{-1}$ para pNPP como substrato, $k_{cat} = 5.9 \pm 0.4 \text{ s}^{-1}$ para fosfotirosina e $k_{cat} = 6.7 \pm 0.1 \text{ s}^{-1}$ para fenilfosfato^[98], indicando que k_{cat} é independente do $\text{p}K_a$ do grupo de saída (p. 41). A constante k_2 é dada apenas para reação de pNPP, mas como k_3 é o mesmo para ambos substratos (deve ser independente do grupo de saída) e os valores de k_{cat} dados acima são praticamente os mesmos, podemos assumir que k_2 (e a energia de ativação) para reação de fenilfosfato

Tabela 29: Distâncias médias calculadas para reação de PhOP^{2-} catalisada pela VHR^a

Centros ^b				MC	TS	produtos
S	(CYS124)	P	(PhOP^{2-})	3.58 ± 0.05	2.74 ± 0.10	2.20 ± 0.05
S	(CYS124)	HN	(SER131)	2.26 ± 0.16	2.50 ± 0.21	2.69 ± 0.24
S	(CYS124)	HO	(SER131)	1.95 ± 0.09	2.02 ± 0.11	2.14 ± 0.14
O1	(PhOP^{2-})	P	(PhOP^{2-})	1.77 ± 0.04	2.40 ± 0.14	4.45 ± 0.06
O2	(PhOP^{2-})	P	(PhOP^{2-})	1.54 ± 0.02	1.52 ± 0.02	1.54 ± 0.02
O3	(PhOP^{2-})	P	(PhOP^{2-})	1.53 ± 0.02	1.51 ± 0.02	1.53 ± 0.02
O4	(PhOP^{2-})	P	(PhOP^{2-})	1.55 ± 0.02	1.52 ± 0.02	1.53 ± 0.02
OH	(ASP92)	HO	(ASP92)	0.99 ± 0.02	1.57 ± 0.62	4.34 ± 0.33
O1	(PhOP^{2-})	HO	(ASP92)	4.09 ± 0.05	2.46 ± 1.48	0.98 ± 0.02
O2	(PhOP^{2-})	HN2	(ARG130)	3.18 ± 0.13	1.75 ± 0.24	1.63 ± 0.10
O2	(PhOP^{2-})	HN	(GLU126)	2.16 ± 0.21	2.21 ± 0.22	2.67 ± 0.21
O2	(PhOP^{2-})	HN	(ARG125)	1.90 ± 0.13	1.79 ± 0.13	1.89 ± 0.15
O3	(PhOP^{2-})	HN	(ARG130)	2.47 ± 0.22	1.98 ± 0.16	1.84 ± 0.13
O3	(PhOP^{2-})	HN1	(ARG130)	1.67 ± 0.11	1.65 ± 0.10	1.65 ± 0.10
O4	(PhOP^{2-})	HN	(SER129)	1.95 ± 0.17	1.88 ± 0.16	1.77 ± 0.12
O4	(PhOP^{2-})	HN	(TYR128)	2.28 ± 0.23	2.12 ± 0.24	1.91 ± 0.16
O1	(PhOP^{2-})	H1	(H_2O 1729)	2.05 ± 0.22	2.64 ± 0.23	8.15 ± 0.46
O1	(PhOP^{2-})	H2	(H_2O 1729)	2.74 ± 0.19	2.20 ± 0.25	7.86 ± 0.68

^a distâncias e desvios padrão em Å. ^b indica entre quais centros a distância foi calculada. HN indica o hidrogênio ligado à cadeia principal do aminoácido mostrado entre parênteses e HO indica o hidrogênio ligado ao oxigênio da cadeia lateral. H1 e H2 indicam os hidrogênios da molécula de água. Veja a legenda da figura 37 para o restante das definições. Os valores foram calculados pela análise das trajetórias de dinâmica molecular por 10 ps (pré-equilibradas durante 4 ps, §4.1.3) em que as distâncias foram amostradas com intervalo de 20 fs (ou seja, 500 valores de distância foram acumulados). As trajetórias corresponderam à janela do PMF com $\xi_i = -1.8$ Å para o MC e $\xi_i = 2.25$ Å para os produtos, e às janelas $\xi_i = -0.3$ e $\xi_i = -0.4$ Å para o TS (portanto, 1000 valores de distância foram acumulados para o TS).

Tabela 30: Energia livre de reação e de ativação, e coordenadas de reação dos reagentes, TSs e produtos calculadas para as reações catalisadas pela VHR^a

Simulação	ξ_{reac}	ΔG_{reac}	ξ_{prod}	$\Delta G^\ddagger$	ξ_{TS}
1a	-1.86	-9.2	2.33	16.4	-0.36
2	-2.09	-3.7	2.54	22.4	-0.38
3	-1.87	5.3	3.00	21.8	-0.12
4	-1.84	15.1	1.42	33.7	-0.14
5 (O2)	-2.03			55.9	0.07
5 (O3)	-2.04			36.6	-0.16
5 (O4)	-2.03			39.9	-0.25
6	-2.06			31.5	-0.20

^a ξ_{reac} é a coordenada de reação dos reagentes, ξ_{TS} dos TSs e ξ_{prod} dos produtos. ΔG em kcal/mol e ξ em Å. Na simulação 5, O2, O3 e O4 indicam qual oxigênio não ligante do grupo fosfato foi protonado.

Figura 36: Perfil de energia livre para reação de PhOP^{2-} (1a, tabela 27) e MeOP^{2-} (2) catalisada pela VHR.

Figura 37: Estrutura do sítio ativo obtida de uma configuração da dinâmica molecular na região do TS da reação de PhOP^{2-} catalisada pela VHR. Parte dos átomos das cadeias laterais dos aminoácidos, cadeia principal e substrato foram apagados para facilitar a visualização. O ângulo de visão é similar àquele adotado na figura 10 para o MC da VHR com fenilfosfato. Legenda para os tipos atômicos: H(branco), C(cinza), C_α (linhas diagonais cruzadas), N(preto), O(linhas diagonais), P(ondulado) e S(linhas cruzadas).

é igual a k_2 para reação de pNPP. Não conhecemos medidas experimentais para desfosforilação de metilfosfato catalisada por uma PTP. Todavia, a constante de velocidade pode ser estimada em $k_2^{MeOP^{2-}} \approx k_2^{pSER} = 0.16 \text{ s}^{-1}$, pelo k_{cat} medido para a reação de desfosforilação fosfoserina (pSER) catalisada pela VHR^[98], assumindo $k_3^{pSER} \equiv k_3^{pNPP}$ e usando a equação 1.11. Esta constante de velocidade estimada corresponde a uma energia de ativação de 18.7 kcal/mol que é 3.7 kcal/mol mais baixa que a barreira calculada para reação de metilfosfato (2, tabela 30).

A reação de ambos os substratos é exoérgica (figura 36 e tabela 30). A energia livre de reação calculada para fenilfosfato é 5.5 kcal/mol mais baixa do que a reação de metilfosfato. Não conhecemos medidas experimentais da energia de reação para estes substratos. O ΔG_{reac} calculado (tabela 30) refere-se à diferença de energia entre os complexos do reagente e do produto com a enzima, portanto, pode ser diferente do ΔG_{reac} medido experimentalmente. O mecanismo de reação observado para os dois substratos é A_nD_n com um TS dissociativo, com caráter de metafosfato, isto é, com considerável quebra da ligação P–O e pouca participação do nucleófilo na formação da ligação P–S.

4.2.1.3 Mutantes ASP→ALA e ARG→ALA

As reações de PhOP^{2-} em que o ácido geral ASP92 e a ARG130 foram mudados para ALA também foram simuladas (3 e 4, figura 38).

Para o mutante ASP92→ALA (simulação 3), os oxigênios não ligantes do grupo fosfato formam um ligante bidentado com a cadeia lateral da ARG130. O grupo hidroxila da cadeia lateral da SER129, que na enzima selvagem liga-se ao oxigênio desprotonado da cadeia lateral do ASP92, liga-se no mutante a uma molécula de água e a um oxigênio não ligante do substrato (PhOP^{2-}). Estas alterações estruturais em relação ao MC da VHR selvagem não são significativas. Por exemplo, a ξ_{reac} é praticamente a mesma da enzima selvagem (tabela 30). Na região do produto, o grupo de saída está ionizado (fenóxido), ligado a uma molécula de água e complexado à cadeia lateral da ARG130, resultando num produto com maior $d(\text{PO})$ (e, logo, maior ξ_{prod}) do que na enzima selvagem. Na região do TS, o perfil de energia livre é mais plano do que nas outras simulações (figuras 36 e 38). As distâncias relevantes são $d(\text{SP}) = 2.53 \pm 0.09 \text{ \AA}$ e $d(\text{PO}) = 2.41 \pm 0.10 \text{ \AA}$, e as distâncias entre o fósforo e os oxigênios não ligantes do grupo fosfato são $1.52 \pm 0.03 \text{ \AA}$. A energia de reação é endoérgica e 14.5 kcal/mol mais alta do que a energia de reação na enzima selvagem. O mecanismo de reação é A_nD_n com um TS associativo, com maior participação do nucleófilo. A altura da barreira calculada é 21.8 kcal/mol, 5.4 kcal/mol

Figura 38: Perfil de energia livre para reação de PhOP^{2-} catalisada pelos mutantes $\text{ASP92} \rightarrow \text{ALA}$ (3, tabela 27) e $\text{ARG130} \rightarrow \text{ALA}$ (4) da VHR.

mais alta que a barreira na enzima selvagem (tabela 30).

Não conhecemos medidas experimentais de k_2 (equação 1.9) para desfosforilação de fenilfosfato pela VHR mutante sem o ácido geral. Porém, a reação catalisada pela mutante $\text{ASP92} \rightarrow \text{ASN}$, apresenta $k_{\text{cat}} = 5 \cdot 10^{-2} \text{ s}^{-1}$ se pNPP é o substrato^[95, 103] e $k_{\text{cat}} = 2 \cdot 10^{-3} \text{ s}^{-1}$ se PhOP^{2-} for o substrato^[95], indicando que, na ausência do ácido geral, a constante de velocidade para desfosforilação é dependente do $\text{p}K_a$ do grupo de saída (p. 41, §1.3)^[95, 264]. A mutação do ácido geral em membros de outras subfamílias de PTPs provoca diminuições de k_{cat} de 200 a 10^5 vezes^[104, 265, 266]. Por exemplo, no mutante $\text{ASP} \rightarrow \text{ALA}$ de uma LM-PTP, k_2 e k_3 foram reduzidos respectivamente 10^4 e 40 vezes em relação às constantes observadas com a enzima selvagem e a etapa determinante da reação foi mudada para tiólise do substrato (equação 1.9, §1.3.2)^[265]. Para VHR mutante, a etapa determinante também foi mudada para a tiólise do substrato^[95, 103]. Assim, podemos supor que as mutações na VHR e na LM-PTP têm o mesmo efeito sobre k_2 e k_3 , e, a partir da equação 1.11, temos que $k_2 \approx k_{\text{cat}} \Leftrightarrow k_2 \ll k_3$. A energia de ativação para desfosforilação de PhOP^{2-} pela VHR mutante $\text{ASP} \rightarrow \text{ALA}$ pode ser estimada em 21.3 kcal/mol, usando k_{cat} dada acima para o mutante $\text{ASP} \rightarrow \text{ASN}$. Para LM-PTP^[265], k_{cat} medido para o mutante $\text{ASP} \rightarrow \text{ALA}$ é metade do k_{cat} medido para o mutante $\text{ASP} \rightarrow \text{ASN}$, logo, podemos esperar que a estimativa de energia de ativação para VHR mutante esteja subestimada (em cerca

de 0.5 kcal/mol). De qualquer forma, a barreira calculada (21.8 kcal/mol, tabela 30) e a observada experimentalmente são muito próximas.

Para o mutante ARG→ALA (simulação 4), as alterações estruturais do MC em relação à VHR selvagem são muito significativas. O grupo fosfato do substrato (PhOP^{2-}) encontra-se inclinado em relação ao plano formado pelo P-loop, (na VHR selvagem, o grupo fosfato é aproximadamente coplanar ao P-loop, figura 10). A cadeia lateral da CYS124 está torcida na direção contrária ao substrato, formando uma ligação de hidrogênio com a cadeia lateral da SER131. Outra importante alteração ocorre no $\text{p}K_a$ dos grupos reativos, pois o hidrogênio ácido do ASP92 é transferido para um dos oxigênios não ligantes do grupo fosfato já no reagente. Em $\xi < -0.4 \text{ \AA}$, o hidrogênio transferido para o grupo fosfato fica coordenado ao grupo tiol da CYS124; em $\xi \sim -0.4 \text{ \AA}$, o hidrogênio torciona-se na direção do grupo de saída e, na região do TS, é transferido para O1. A estabilização do grupo de saída no produto é menor e a $d(\text{PO})$ é reduzida em relação ao complexo do produto na VHR selvagem (figura 38). A energia livre de reação é endoérgica, cerca de 24 kcal/mol mais alta do que a energia de reação da enzima selvagem e a altura de barreira é cerca de 17 kcal/mol mais alta do que na VHR selvagem (tabela 30). O mecanismo de reação é A_nD_n , com um TS bastante associativo, comparável às simulações 5 e 6 (figura 39, vide infra), já que o H^+ do ASP92 é transferido para o grupo fosfato no MC. Não conhecemos medidas cinéticas da VHR mutante ARG→ALA, mas o aumento da energia de ativação da primeira etapa irreversível (a reação de tiólise do substrato, no caso das PTPs, §1.3) da catálise por uma PTP clássica mutante ARG→ALA pode ser estimado em cerca de 7.5 kcal/mol^[101]. A elevação da barreira na simulação é 10 kcal/mol maior que esta estimativa (tabela 30), portanto, sugerimos que a estrutura usada para simulação 4 é muito diferente ou incompatível com a estrutura observada experimentalmente para o mutante.

4.2.1.4 Protonação do substrato ou do nucleófilo

Os oxigênios não ligantes do grupo fosfato não são equivalentes em razão da complexação com diferentes grupos no P-loop e da orientação do grupo de saída. Assim, a simulação 5 foi realizada com três configurações iniciais em que cada um dos três oxigênios não ligantes foi protonado (figura 39). A protonação do grupo tiol da CYS124 foi testada na simulação 6, mas, na primeira janela desta simulação, o H^+ é transferido da CYS124 para o oxigênio (O2) do grupo fosfato (veja a numeração dos oxigênios na figura 37). A coordenada de reação (tabela 30) e a estrutura do MC são semelhantes para as qua-

Figura 39: Perfil de energia livre para reação catalisada pela VHR de PhOP^- (fenilfosfato monoprotonado, simulação 5, onde O2, O3 e O4 indicam qual dos oxigênios não ligantes no grupo fosfato foi protonado) ou de PhOP^{2-} com a CYS124 protonada (6).

tro simulações. A principal diferença entre os MCs é a ligação com o grupo guanidina da ARG130, pois o oxigênio protonado do grupo fosfato não deve formar um ligação de hidrogênio tão estável quanto os oxigênios desprotonados. Se O2 está inicialmente protonado, o grupo fosfato forma um ligante monodentado com o grupo guanidina da ARG130. Nas outras simulações (inclusive a simulação 6), o grupo fosfato forma um ligante bidentado.

Nas primeiras janelas das quatro simulações ($\xi_{\text{reac}} \leq \xi < -1 \text{ Å}$), o H^+ ligado ao grupo fosfato também forma uma ligação de hidrogênio com o grupo tiolato da CYS124. Para as simulações 5 (O3) e 6, os perfis de energia livre para desfosforilação do monoânion são bimodais, ou seja, indicam dois processos diferentes. O primeiro processo corresponde à mudança de inclinação dos perfis em $\xi \approx -0.8 \text{ Å}$ para simulação 5 (O3) e em $\xi \approx -1.0 \text{ Å}$ para simulação 6, e equivale à quebra da ligação entre o hidrogênio do grupo fosfato e o grupo tiolato da CYS124, resultando na torção do H^+ em direção ao grupo de saída. O segundo processo é a transferência de fosfato do grupo de saída para o nucleófilo (em ξ_{TS} , tabela 30). Este comportamento bimodal também foi observado na simulação da desfosforilação de fenilfosfato monoprotonado catalisada por uma LM-PTP (§4.3.3.1)^[110]. Nas simulações 5 (O3) e 6, o grupo de saída retira o H^+ ligado ao grupo fosfato no

TS, e, na região dos produtos, a enzima tiofosforilada está desprotonada e o ASP92, protonada. As alturas de barreira para estas simulações são as mais baixas entre as reações de fenilfosfato monoprotonado (figura 39). Na simulação 5 (O2), o H^+ ligado ao grupo fosfato permanece complexado ao grupo tiolato da CYS124 até a dissociação do grupo de saída no TS, resultando em ξ_{TS} e $\Delta G^\ddagger$ significativamente maiores que nas outras simulações de fenilfosfato monoprotonado (tabela 30). Na simulação 5 com O4 protonado, o hidrogênio ácido do ASP92 é transferido ao oxigênio do grupo de saída e na região dos produtos, o oxigênio protonado do grupo fosfato e o grupo carboxila ionizado do ASP92 formam uma ligação de hidrogênio. A transferência de H^+ do ASP92 é observada somente na simulação 5 (O4), porque, nas outras simulações, a solvatação da cadeia lateral do ASP92 impede o contato entre o oxigênio do grupo de saída e o hidrogênio ácido.

Nas simulações 5 e 6, o mecanismo calculado foi A_nD_n , com um TS de caráter associativo. As energias livres de ativação são de 15 a 40 kcal/mol mais altas do que a barreira na simulação 1a e do que a barreira medida experimentalmente para desfosforilação de fenilfosfato (§4.2.1.2). Estas simulações não foram executadas até o aumento da energia na região dos produtos, assim, não obtivemos energias de reação.

4.2.1.5 Mutante CYS→SER

O mutante CYS124→SER da VHR foi simulado em diferentes estados de protonação (figura 40). Na simulação 7, a cadeia lateral da SER e o substrato ($PhOP^{2-}$) estão desprotonados. A coordenada de reação do reagente e a $d(OP) \sim 3.8 \text{ \AA}$ (veja a figura 34 e a tabela 27 para definição desta distância) são maiores que ξ_{reac} e $d(SP)$ no MC da VHR selvagem (simulação 1a), respectivamente (tabela 31). O grupo fosfato forma um ligante bidentado com a cadeia lateral da ARG130. A quantidade e as distâncias das ligações de hidrogênio no MC (tabela 32) são similares às aquelas observadas para a VHR selvagem (tabela 29) e indicam que a mutação altera pouco a estrutura do MC (o grupo fosfato sofreu uma rotação no mutante, assim, a posição ocupada por O2 na enzima selvagem é aproximadamente ocupada por O4 no mutante). Na região dos produtos, a cadeia lateral da SER124 é fosforilada e fenóxido está complexado a uma molécula de água e ao grupo guanidina da ARG130. A ASP92 está protonada, portanto, não ocorre transferência de H^+ para o grupo de saída. Na região do TS, a $d(PO) = 1.90 \pm 0.06 \text{ \AA}$ e a $d(OP) = 2.37 \pm 0.07 \text{ \AA}$ indicam que o mecanismo é A_nD_n com um TS associativo, que apresenta pouca formação da ligação entre o grupo fosfato e o oxigênio (O_γ) da cadeia lateral da SER124, mas pouca dissociação do grupo de saída. A $d(OH) = 3.97 \pm 0.23 \text{ \AA}$ na região

Tabela 31: Energia livre de ativação e coordenada de reação calculadas para catálise pelo mutante CYS124→SER da VHR^a

Simulação	ξ_{reac}	ξ_{TS}	$\Delta G^\ddagger$
7	-2.11	-0.47	14.1
8 (O2)	-1.64	0.73	61.7
8 (O3)	-1.71	0.24	40.4
8 (O4)	-1.64	0.65	61.0
9	-1.70	0.39	48.5
10 (O2)	-1.55	0.56	55.0
10 (O4)	-1.64	0.78	73.2

^a ξ_{reac} é a coordenada de reação dos reagentes e ξ_{TS} dos TSs. ΔG em kcal/mol e ξ em Å. Nas simulações 8 e 10, O2, O3 e O4 indicam qual oxigênio não ligante do grupo fosfato foi protonado.

do TS é um pouco maior que a mesma distância [$d(OH) = 3.70 \pm 0.30$ Å] observada antes da transferência de H^+ na simulação 1a. A região do TS contém um intervalo de ξ maior que na reação da VHR selvagem, já que o perfil de energia livre é mais plano nesta região (figura 40). Todavia, a ξ_{TS} para o mutante e para enzima selvagem são similares. A altura de barreira para reação de $PhOP^{2-}$ catalisada pela VHR mutante CYS124→SER é 14 kcal/mol (tabela 31), indicando que o mutante seria ativo (pelo menos na primeira etapa, equação 1.9) na desfosforilação do substrato.

A protonação dos três oxigênios não ligantes do grupo fosfato (não equivalentes, §4.2.1.4) foi testada (figura 40). A cadeia lateral da SER124 esteve inicialmente ionizada, no entanto, na região dos reagentes ($\xi < -0.5$ Å), as distâncias entre O_γ da SER124 e o hidrogênio do substrato foram ~ 1.1 Å e as distâncias entre este mesmo hidrogênio e o oxigênio do grupo fosfato foram ~ 1.2 Å. Portanto, o hidrogênio ficou compartilhado entre o O_γ e o grupo fosfato. Nas simulações com O2 e O4 protonados, o substrato forma um ligante bidentado com a cadeia lateral da ARG130, mas, na simulação com O3 protonado, o ligante é monodentado. Em $\xi \approx -0.2$, uma espécie pentacoordenada com geometria de TBP é formada nas três simulações, com $d(PO) \sim 1.9$ Å e $d(OP) \sim 1.8$ Å. Se O3 está protonado, o H^+ da cadeia lateral do ASP92 é transferido para o oxigênio do grupo de saída que se dissocia como fenol em ξ_{TS} (tabela 31). Nos outros dois casos, o TS é atingido em ξ e altura de barreira significativamente maiores (tabela 31). O grupo de saída é protonado pelo H^+ ligado ao grupo fosfato e o ASP92 permanece protonada. A espécie pentacoordenada é um mínimo no perfil de energia livre em $\xi \approx 0.0$ Å para as simulações com O2 e O4 inicialmente protonados, mas o mínimo é instável (as barreiras para reação inversa, na direção dos reagentes, são menores que 1 kcal/mol) e, portanto, não podemos dizer que a espécie pentacoordenada caracteriza uma fosforana (p. 21).

Figura 40: Perfil de energia livre para reação da VHR mutante CYS124→SER ionizada com substrato PhOP^{2-} (7, tabela 27) ou PhOP^- (8, onde O2, O3 e O4 indicam qual oxigênio não ligante do grupo fosfato está protonado).

Assim, o mecanismo em ambas simulações é A_nD_n , com um TS altamente associativo, com caráter de fosforana, ou seja, com significativa formação da ligação com o nucleófilo, e pouca quebra da ligação com o grupo de saída. Na região dos produtos, fenol está complexado ao sítio ativo e a enzima está fosforilada. As barreiras são maiores que 40 kcal/mol para as três simulações, indicando que não ocorre catálise na VHR mutante se o substrato é um monoânion.

As reações em que O_γ da SER124 está protonado e o substrato é PhOP^{2-} (simulação 9) ou PhOP^- (simulação 10) também foram testadas (figura 41). Na simulação 9, a estrutura do MC e ξ_{reac} (tabela 31) são similares ao MC observado para enzima selvagem (tabela 30). Um ligante bidentado é inicialmente formado e o hidrogênio ligado ao O_γ da SER124 (H_δ) está compartilhado com O2, com $d(\text{O}_\gamma\text{H}_\delta) = 1.05 \pm 0.04$ Å e $d(\text{O2H}_\delta) = 1.44 \pm 0.10$ Å. O hidrogênio ácido do ASP92 permanece a ~ 2.5 Å do oxigênio do grupo de saída durante toda simulação. Em $\xi = -0.34$ Å, uma espécie pentacoordenada com geometria de TBP é formada e o oxigênio do grupo de saída e O_γ da SER124 assumem as posições axiais na TBP. A quantidade e as distâncias das ligações de hidrogênio no MC (tabela 32) são similares àsquelas observadas para a VHR selvagem (tabela 29) e para simulação 7. No TS, $d(\text{PO}) = 1.96 \pm 0.05$ Å, $d(\text{OP}) = 1.69 \pm 0.03$ Å e o grupo de saída é protonado pelo

Figura 41: Perfil de energia livre para reação da VHR mutante CYS124→SER protonada com PhOP^{2-} (9, tabela 27) e PhOP^- (10, onde O2 e O4 indicam qual oxigênio não ligante do grupo fosfato está protonado).

hidrogênio ligado ao O2. A espécie pentacoordenada formada é instável (a decomposição na direção dos reagentes tem uma barreira menor que 2 kcal/mol) e, portanto, o mecanismo é A_nD_n , altamente associativo e com um TS tardio, ou seja, com completa formação da ligação entre fosfato e o nucleófilo e pouca quebra da ligação com o grupo de saída (§1.2). A distância entre os oxigênios não ligantes e o fósforo aumenta ligeiramente (0.03 Å) no TS em relação às distâncias no MC. Este comportamento é o inverso daquele observado para os mecanismos dissociativos, mas é semelhante ao comportamento das distâncias entre fósforo e os oxigênios equatoriais nas reações em fase gasosa A_n+D_n de tri e monoésteres de fosfato (§2.2). Na região dos produtos, fenol está complexado ao P-loop, o ASP92 está protonada e a SER124, fosforilada. A altura de barreira para a simulação 9 é maior que 40 kcal/mol (tabela 31), indicando que também não ocorre catálise, se a SER124 está protonada.

Na simulação 10 (figura 41), a carga total da região QM é -1 e a estrutura dos MCs está razoavelmente alterada, pois o grupo fosfato está inclinado em relação ao plano definido pelo P-loop, principalmente se O4 é protonado. Para ambas simulações (a protonação de O3 não foi testada), uma espécie pentacoordenada com geometria de TBP é formada em $\xi \sim -0.35$ Å e um mínimo no perfil de energia livre é atingido em $\xi \approx -0.05$

Tabela 32: Distâncias médias calculadas para reação de PhOP^{2-} catalisada pela VHR mutante $\text{CYS124} \rightarrow \text{SER}^a$

Centros ^b		MC	TS
simulação 7			
O (SER124)	HN (SER131)	2.66 ± 0.35	3.15 ± 0.17
O (SER124)	HO (SER131)	1.78 ± 0.13	2.46 ± 0.22
O2(PhOP^{2-})	HN (TYR128)	1.85 ± 0.15	1.85 ± 0.15
O2(PhOP^{2-})	HN (SER129)	2.18 ± 0.21	1.97 ± 0.13
O3(PhOP^{2-})	HN (ARG125)	1.89 ± 0.12	1.66 ± 0.09
O3(PhOP^{2-})	HN (GLU126)	2.27 ± 0.22	2.15 ± 0.18
O3(PhOP^{2-})	HN2(ARG130)	1.62 ± 0.10	1.59 ± 0.09
O4(PhOP^{2-})	HN (ARG130)	2.24 ± 0.17	1.83 ± 0.12
O4(PhOP^{2-})	HN1(ARG130)	1.66 ± 0.11	1.63 ± 0.10
O4(PhOP^{2-})	HN2(ARG130)	2.70 ± 0.14	2.83 ± 0.16
simulação 9			
O (SER124)	HN (SER131)	2.75 ± 0.24	2.91 ± 0.20
O (SER124)	HO (SER131)	1.94 ± 0.14	2.04 ± 0.15
O2(PhOP^{2-})	HN (TYR128)	1.92 ± 0.12	1.89 ± 0.16
O2(PhOP^{2-})	HN (SER129)	1.87 ± 0.12	1.75 ± 0.10
O3(PhOP^{2-})	HN (ARG125)	1.85 ± 0.11	1.78 ± 0.13
O3(PhOP^{2-})	HN (GLU126)	2.29 ± 0.22	1.71 ± 0.10
O3(PhOP^{2-})	HN (GLY127)	3.32 ± 0.26	1.90 ± 0.18
O3(PhOP^{2-})	HN2(ARG130)	1.73 ± 0.14	3.42 ± 0.19
O4(PhOP^{2-})	HN1(ARG130)	1.86 ± 0.17	1.81 ± 0.10
O4(PhOP^{2-})	HN2(ARG130)	2.68 ± 0.20	1.93 ± 0.15

^a distâncias e desvios padrão em Å. ^b indica entre quais centros a distância foi calculada. Veja a tabela 29 para definições dos centros e de como os valores foram calculados. Na simulação 7, as trajetórias corresponderam à janela do PMF com $\xi_i = -2.11$ Å para o MC e $\xi_i = -0.47$ Å para o TS. Na simulação 9, $\xi_i = -1.70$ e $\xi_i = 0.39$ Å.

Å (figura 41). Esta espécie é estável (a barreira para decomposição em direção aos reagentes é de 5 ou 9 kcal/mol, se O2 ou O4 estão protonados, respectivamente), caracterizando uma fosforana e o mecanismo de reação é $A_n + D_n$. As posições equatoriais da TBP são ocupadas por O_γ e pelo oxigênio do grupo de saída nas duas simulações, além de O4 (protonado) na simulação 10 (O4) ou O3 (desprotonado) na simulação 10 (O2). Estas configurações devem sofrer pseudorrotações para ligação do nucleófilo e dissociação do grupo de saída. O TS para as duas reações corresponde à dissociação do grupo de saída. Na simulação 10 com O2 protonado, fenol é o grupo de saída, pois recebe o H^+ ligado ao grupo fosfato. Na simulação 10 com O4 protonado, a barreira é quase 20 kcal/mol mais alta (tabela 31), porque o grupo de saída no TS é fenóxido e o arranjo dos grupos na fosforana parece desfavorável. O hidrogênio ligado ao O4 está coordenado à cadeia lateral do ASP92, cujo hidrogênio ácido está torcido na direção oposta e exposto ao solvente. O

H^+ que inicialmente estava ligado ao O_γ na SER124, foi transferido para O_2 na formação da fosforana e realiza ligações de hidrogênio com a SER131 no TS. Assim, os três hidrogênios ácidos não estão disponíveis para transferência no TS. Nas janelas posteriores, fenóxido retira o H^+ ligado ao O_2 . Para as duas simulações, a altura de barreira total é maior que 50 kcal/mol, indicando que não ocorre catálise se o substrato e o nucleófilo estiverem protonados. Os produtos de reação correspondem à fenol complexado ao sítio ativo, à ASP92 protonada e à SER124 fosforilada, com um dos oxigênios não ligantes do grupo fosfato protonado.

Nenhuma simulação da VHR mutante CYS124→SER foi executada até o aumento do perfil de energia livre na região dos produtos, assim, não obtivemos energias de reação para o mutante.

4.2.2 Simulações da fosfatase de proteínas CDC25B

4.2.2.1 Estado de protonação do substrato

Embora sua presença seja postulada, o resíduo que deve funcionar como ácido geral nas CDC25s ainda não foi identificado (p. 39, §1.3). Assim, nas simulações 12 e 13 (tabela 27), calculamos o perfil de energia livre para desfosforilação de fenilfosfato catalisada pela CDC25B sem que um resíduo protéico funcionasse como ácido geral (figura 42). Na simulação 13, cada um dos oxigênios não ligantes do grupo fosfato foi protonado para testar a reatividade do substrato monoaniônico, como já proposto para outras fosfatases (§1.3.3)^[115] e testado para VHR (§4.2.1.4). Recentemente, o substrato monoprotonado também foi proposto como um possível doador de H^+ para o grupo de saída na reação catalisada pela CDC25B^[22], portanto, a simulação 13 também é um teste desta hipótese.

Uma diferença importante entre a VHR e a CDC25B é que esta última tem um sítio ativo bem mais exposto ao solvente. Moléculas de água cobrem o sítio ativo e o substrato, e solvatam o grupo de saída e os oxigênios não ligantes do grupo fosfato (de 3 a 4 moléculas de água estão em contato direto com o substrato). No entanto, nenhuma molécula de água penetra no interior do sítio ou se localiza entre o nucleófilo e o substrato.

Na simulação 12, o reagente é formado numa coordenada de reação comparável à ξ_{reac} na VHR (simulação 1a, tabelas 30 e 33). As distâncias $d(SP)$ no reagente também são próximas e o grupo fosfato encontra-se no plano definido pelo P-loop, indicando que a estrutura do MC é similar em ambas enzimas. O substrato dianiônico forma um ligante monodentado com o grupo guanidina da ARG479 durante toda simulação (tabela 34).

Figura 42: Perfil de energia livre para reação de PhOP^{2-} (12, tabela 27) ou PhOP^- (13, onde O2, O3 e O4 indicam qual oxigênio não ligante do grupo fosfato foi protonado) catalisada pela CDC25B.

Tabela 33: Energia livre de reação e de ativação, e coordenada de reação calculadas para os reagentes, TSs e produtos das reações catalisadas pela CDC25B^a

Simulação	Observação ^b	ξ_{reac}	ΔG_{reac}	ξ_{prod}	$\Delta G^\ddagger$	ξ_{TS}
12		-1.95	-7.9	2.75	16.6	-0.25
13	(O2)	-2.02			34.9	-0.34
13	(O3)	-1.98			48.9	-0.07
13	(O4)	-1.91			38.5	0.05
14b	GLU474 d(OH) = 1.9 Å				26.1	1.61
14b	GLU474 d(OH) = 4.0 Å				38.4	1.95
14b	GLU478 d(OH) = 2.3 Å				30.5	0.60
14b	GLU478 d(OH) = 4.1 Å	-1.26			31.3	0.50

^a ξ_{reac} é a coordenada de reação dos reagentes, ξ_{TS} dos TSs e ξ_{prod} dos produtos. ΔG em kcal/mol e ξ em Å. ^b na simulação 13, O2, O3 e O4 indicam qual oxigênio não ligante do grupo fosfato foi protonado e, na simulação 14b, indica qual o resíduo foi usado como ácido geral e qual a d(OH) inicial (p. 199 e figura 43).

O TS tem geometria de uma TBP, comparável à geometria do TS na VHR (figura 37, mas sem o H^+ ligado ao grupo de saída, sem o ASP92 e com o grupo fosfato formando um ligante monodentado com a ARG conservada), com os oxigênios não ligantes em posição equatorial e o oxigênio do grupo de saída e o enxofre da CYS473 em posição axial (quase colineares, pois o ângulo S–P–O1 é $172.2 \pm 4.5^\circ$). Os oxigênios equatoriais e o fósforo estão coplanares, como no metafosfato livre (§2.2.2). A quebra da ligação P–O é avançada e semelhante à distância observada no TS da simulação 1a (tabelas 29 e 34). A $d(SP)$ é um pouco menor na CDC25B, indicando maior participação do nucleófilo no TS da CDC25B (tabela 34). O mecanismo de reação é A_nD_n com um TS dissociativo, ainda com caráter de metafosfato. Obviamente, não ocorre protonação do grupo de saída, mas o fenóxido é solvatado por moléculas de água (O1 realiza cerca de 3 ligações hidrogênio no TS e nos produtos, tabela 34). A coordenada de reação dos produtos é maior que ξ_{prod} na simulação 1a, porque $d(PO)$ é cerca de 0.5 Å maior, novamente em consequência da maior exposição ao solvente. A energia de reação é -7.9 kcal/mol (tabela 33), 1.3 kcal/mol mais alta que a energia de reação da simulação 1a.

A altura de barreira calculada, 16.6 kcal/mol (tabela 33), também é próxima à altura na VHR. Não conhecemos medidas experimentais de k_2 (equação 1.9) para desfosforilação de fenilfosfato pela CDC25B. Esta constante é conhecida para reação de pNPP catalisada pela CDC25B (tabela 2) e equivale a uma energia de ativação de 18.0 kcal/mol. Contudo, a atividade das CDC25s apresenta uma estranha dependência com o pK_a do grupo de saída (p. 41)^[105] e a constante de velocidade observada para pNPP não pode ser extrapolada para $PhOP^{2-}$ (ao contrário da reação catalisada pela VHR, §4.2.1.2). Uma melhor estimativa da energia de ativação experimental pode ser obtida a partir do $k_{cat}/K_M = 1.2$ s⁻¹ M⁻¹ medido em pH = 7 e a 30 °C para desfosforilação de fenilfosfato, catalisada pela CDC25A^[105] e assumindo que $K_M^{PhOP^{2-}} \approx K_M^{pNPP} = 2 \cdot 10^{-2}$ M, e que $k_2^{PhOP^{2-}} \ll k_3$ (equação 1.10). Assim, $k_2^{PhOP^{2-}} \approx k_{cat}^{PhOP^{2-}} = 2.4 \cdot 10^{-2}$ s⁻¹ (equação 1.11), que equivale a uma energia de ativação de 19.8 kcal/mol para primeira etapa (tiólise de fenilfosfato) catalisada pela CDC25A. A k_{cat} medida a 30 °C para reação de pNPP catalisada pela CDC25A é metade da mesma constante medida para CDC25B^[99] a 20 °C, portanto, a energia de ativação para reação de fenilfosfato catalisada pela CDC25B deve ser um pouco menor que 19.8 kcal/mol, mas é certamente superior à 18 kcal/mol. As duas aproximações usadas para calcular $k_2^{PhOP^{2-}}$ são razoáveis, porque K_M não deve ser significativamente alterado pelo grupo em posição *para* no anel aromático do substrato (exposto ao solvente e sem contatos com a CDC25) e, porque, para reação catalisada pela CDC25B, $k_2^{pNPP} = 0.46$ s⁻¹, $k_3 = 1.5$ s⁻¹ ($3k_2 \simeq k_3$) e $k_{cat}^{pNPP} = 0.35$ s⁻¹. Espera-se que k_2 seja

Tabela 34: Distâncias médias calculadas para reação de PhOP^{2-} catalisada pela CDC25B^a

Centros ^b			MC	TS	produtos
S (CYS473)	P	(PhOP ²⁻)	3.64±0.05	2.59±0.09	2.19±0.04
S (CYS473)	HN	(GLY480)	1.98±0.11	2.16±0.14	2.17±0.14
O1(PhOP ²⁻)	P	(PhOP ²⁻)	1.74±0.03	2.35±0.10	5.10±0.06
O2(PhOP ²⁻)	P	(PhOP ²⁻)	1.53±0.02	1.53±0.02	1.54±0.02
O3(PhOP ²⁻)	P	(PhOP ²⁻)	1.55±0.02	1.52±0.02	1.55±0.02
O4(PhOP ²⁻)	P	(PhOP ²⁻)	1.55±0.02	1.53±0.02	1.54±0.02
O2(PhOP ²⁻)	HN	(PHE475)	1.87±0.16	2.05±0.20	1.93±0.16
O2(PhOP ²⁻)	HN	(SER476)	2.58±0.24	2.57±0.29	2.29±0.25
O2(PhOP ²⁻)	HN	(SER477)	2.24±0.19	2.15±0.22	1.97±0.16
O3(PhOP ²⁻)	HN	(GLU474)	2.81±0.27	2.09±0.19	2.03±0.18
O3(PhOP ²⁻)	HN1	(ARG479)	1.89±0.22	1.89±0.18	1.93±0.17
O3(PhOP ²⁻)	HN2	(ARG479)	2.04±0.24	1.90±0.18	1.90±0.16
O4(PhOP ²⁻)	HN	(GLU478)	1.82±0.15	1.95±0.17	1.96±0.16
O4(PhOP ²⁻)	HN	(ARG479)	2.10±0.21	1.81±0.12	1.93±0.14
O1(PhOP ²⁻)	H2	(H ₂ O190)	4.60±0.53	2.04±0.29	4.55±0.31
O1(PhOP ²⁻)	H1	(H ₂ O1447)	3.10±0.27	1.94±0.37	3.05±0.15
O1(PhOP ²⁻)	H2	(H ₂ O1447)	3.33±0.31	2.70±0.20	1.75±0.12
O1(PhOP ²⁻)	H1	(H ₂ O1556)	2.81±0.21	1.79±0.23	1.71±0.11
O2(PhOP ²⁻)	H1	(H ₂ O190)	7.03±0.70	2.01±0.32	3.11±0.24
O3(PhOP ²⁻)	H1	(H ₂ O1308)	1.77±0.16	4.29±0.77	2.95±0.20
O4(PhOP ²⁻)	H1	(H ₂ O1447)	1.84±0.23	1.88±0.18	2.11±0.48
O4(PhOP ²⁻)	H1	(H ₂ O1556)	2.09±0.29	2.33±0.38	3.12±0.16

^a distâncias e desvios padrão em Å. ^b indica entre quais centros a distância foi calculada. Veja a tabela 29 para definições dos centros e de como os valores foram calculados. As trajetórias corresponderam à janela do PMF com $\xi_i = -1.9$ Å para o MC, $\xi_i = -0.25$ Å para o TS e $\xi_i = 2.9$ Å para os produtos.

ainda menor ($\beta_{lg} < 0$, p. 41) que $\frac{1}{3}$ de k_3 para fenilfosfato. Portanto, a barreira calculada é cerca de 3 kcal/mol mais baixa que a energia de ativação experimental.

A reação de fenilfosfato monoprotonado foi testada na simulação 13, em que cada um dos três oxigênios não ligantes do grupo fosfato foi protonado. Para as três simulações 13, a estrutura do MC e a coordenada de reação dos reagentes são próximas à estrutura do MC e à ξ_{reac} na simulação 12 (tabela 33). O substrato forma uma ligação monodentada com a ARG479 durante todo progresso das reações e o hidrogênio ligado ao grupo fosfato realiza uma ligação de hidrogênio com o enxofre do grupo tiolato da CYS473. Em $\xi \approx -0.7$ Å para simulação 13 com O2 protonado, e $\xi \approx -0.4$ Å para simulação 13 com O4 protonado, observamos uma mudança de inclinação do perfil de energia livre que corresponde à quebra da ligação de hidrogênio e à torção deste átomo na direção do grupo de saída. Na simulação 13 (O3), a ligação de hidrogênio com o grupo tiolato é mantida até $\xi \sim \xi_{TS}$. Uma espécie pentacoordenada é formada nas três simulações em $\xi \sim -0.6$ Å, com $d(\text{SP}) \sim 2.3$ Å

e $d(\text{PO}) \sim 1.9 \text{ \AA}$. No TS, a ligação P–O é rompida e o H^+ ligado ao grupo fosfato é transferido para o grupo de saída, resultando na fosforilação da enzima e na dissociação de fenol que é solvatado e fica parcialmente complexado ao sítio ativo na região dos produtos. O mecanismo para as três simulações é A_nD_n , com um único TS altamente associativo. Os $\Delta G^\ddagger \geq 35 \text{ kcal/mol}$, indicam que pouca ou nenhuma catálise seria observada se o substrato da CDC25B fosse PhOP^- (tabela 33). A simulação 13 (O3) apresenta uma barreira mais elevada, porque a ligação de hidrogênio com o grupo tiolato é rompida somente na região do TS, diminuindo a estabilização e dificultando a formação do TS. O O3 está coordenado ao grupo guanidina da ARG479, impedindo a torção do hidrogênio ligado ao O3 para direção do grupo de saída.

4.2.2.2 Ácido glutâmico do P-loop como ácido geral

Nesta seção, apresentamos as simulações em que cada um dos resíduos GLU no P-loop (GLU474 e GLU478, §4.1.1) foram testados como possíveis doadores de H^+ para o grupo de saída na reação de PhOP^{2-} catalisada pela CDC25B (§1.3.3). A cadeia lateral do GLU474 está localizada a cerca de $6 \text{ \AA}$ do substrato na estrutura solvatada derivada do PDB (§4.1.1) e forma ligações de hidrogênio com o grupo hidroxila da TYR528 e a cadeia principal do resíduo MET531. Já a cadeia lateral do GLU478 está exposta ao solvente, mas a cerca de $8 \text{ \AA}$ do substrato. Estas distâncias são inapropriadas para transferência de H^+ para o grupo de saída, portanto, alguma mudança conformacional deve ser observada para que a cadeia lateral de um destes resíduos aproxime-se do substrato. Nas outras subfamílias de PTPs, a ligação do substrato resulta numa mudança conformacional do WPD-loop que posiciona um resíduo ASP para catálise ácido-base geral. Na PTP livre, o resíduo ASP encontra-se de 7 a $10 \text{ \AA}$ distante do substrato (§1.3).

Como a estrutura do P-loop deve ser mantida para catálise, a única mudança conformacional possível nas CDC25s é a torção da cadeia lateral do GLU. O PMF para aproximação do hidrogênio ácido construído na cadeia lateral de cada GLU ao oxigênio esterificado do substrato foi obtido com todo o sistema descrito pelo campo de força clássico (simulação 11, tabela 27). As interações da cadeia lateral do GLU474 e do GLU478 na estrutura inicial (vide supra) são facilmente rompidas e o aumento de energia livre até $\xi \sim 4 \text{ \AA}$ é de cerca de 4 a 5 kcal/mol (figura 43). Esta $d(\text{OH})$ é observada no MC para a reação catalisada pela VHR selvagem (tabela 29). A maior aproximação do GLU478 resulta em considerável aumento da energia livre e cerca de 12 kcal/mol devem ser fornecidas ao sistema para que $\xi \sim 2.3 \text{ \AA}$. Para o GLU474, a aproximação até cerca de $2 \text{ \AA}$

Figura 43: Perfil de energia livre para aproximação do hidrogênio ácido da cadeia lateral do GLU474 e do GLU478 ao oxigênio (O1) do grupo de saída de PhOP^{2-} na CDC25B. O GLU aproximado e a simulação (tabela 27) estão indicados na legenda. A aproximação ocorre na direção da direita para esquerda no gráfico. As distâncias assinaladas com flechas indicam a região de onde as estruturas usadas em outras simulações foram retiradas. O ΔG da estrutura inicial (obtido na simulação 11) foi adicionado ao respectivo PMF na simulação 14d.

tem uma barreira de cerca de 5 kcal/mol, porém, para $\xi < 2$ Å, a energia livre aumenta rapidamente.

Quatro pontos estão indicados na figura 43. Estruturas retiradas destas regiões de coordenadas de reação foram usadas como configurações iniciais nas simulações 14 (vide infra). Se dinâmica molecular é evoluída por 5 ps a partir destas estruturas, usando o potencial híbrido e na ausência de um potencial guarda-chuva, a $d(\text{OH})$ assume valores maiores que 5 Å, exceto para a estrutura com o GLU474 protonado e $d(\text{OH}) = 4.3$ Å. Neste caso, a $d(\text{OH})$ não muda significativamente durante a dinâmica molecular, a estrutura obtida no final dos 5 ps foi usada na simulações 14b e 14c, e indicada por $d(\text{OH}) = 4.0$ Å (vide infra). Nas estruturas retiradas da simulação 11 e na estrutura com $d(\text{OH}) = 4.0$ Å, a $d(\text{SP}) \sim 5$ Å (cerca de 1.5 Å maior que a distância observada no MC da simulação 12, tabela 34), pois o substrato é parcialmente deslocado do P-loop e solvatado em maior grau (em relação à estrutura construída em §4.1.1) por moléculas de água cujas ligações de hidrogênio substituem parte das interações dos oxigênios não ligantes do grupo fosfato com a cadeia protéica.

A influência do potencial híbrido no PMF para diminuição da $d(\text{OH})$ (simulação 11) foi testada, pois a transferência de H^+ poderia ocorrer durante a aproximação do hidrogênio ácido e O1 (simulação 14d, tabela 27). As estruturas iniciais usadas foram retiradas da simulação 11 [$d(\text{OH}) = 4.3 \text{ \AA}$ para o GLU474 protonado e $d(\text{OH}) = 4.1 \text{ \AA}$ para o GLU478 protonado, vide supra], já que a reação de transferência não ocorreria em $d(\text{OH})$ elevadas. Se o GLU474 é protonado, o perfil de energia livre calculado indica que a transferência de H^+ para o oxigênio do grupo de saída é bastante desfavorável (figura 43). Para simulação com GLU478 protonado, os perfis de energia livre calculados na simulação 11 e 14d são significativamente diferentes para $\xi < 3 \text{ \AA}$. Nesta região, o perfil obtido com o potencial híbrido é constante em cerca de 9 kcal/mol, indicando que a transferência de H^+ pode ocorrer.

Para as simulações 14d, a protonação do substrato em $\xi = d(\text{OH}) \sim 1.0$ (diretamente de O1) resulta na formação de um zwitterión, $\text{PhO}^+(\text{H})\text{PO}_3^{2-}$ (§1.2.1), sem a dissociação de fenol [$d(\text{PO}) \sim 1.8 \text{ \AA}$]. Todavia, a transferência de H^+ ocorre somente se o potencial guarda-chuva restringe a coordenada de reação. Se este potencial é desligado durante a simulação, o H^+ é imediatamente transferido de volta para a cadeia lateral do GLU. Como podemos observar na figura 43, não existe uma barreira no perfil de energia livre para a reação inversa, ou seja, na direção da esquerda para direita, correspondente ao aumento de ξ . A maior exposição ao solvente não é suficiente para estabilizar a formação do zwitterión e do carboxilato no GLU.

A transferência de H^+ do ácido geral para o oxigênio esterificado também foi testada para VHR na simulação 1d. A transferência ocorre inicialmente para um dos oxigênios não ligantes do grupo fosfato e, em seguida, para o oxigênio esterificado (formando um zwitterión transiente que se dissocia, quebrando a ligação P–O), mas a reação inversa (aumento de ξ) apresenta uma barreira de 5 kcal/mol (figura 35). Por outro lado, a transferência de H^+ do ácido geral na simulação 1a ocorre espontaneamente na região do TS (§4.2.1.2).

As simulações 14a foram realizadas a partir da configuração construída na §4.1.1, das configurações extraídas da simulação 11 [$d(\text{OH}) = 4.3 \text{ \AA}$ para o GLU474 protonado e $d(\text{OH}) = 4.1 \text{ \AA}$ para o GLU478 protonado, vide supra] e das configurações finais da simulação 14d [em $d(\text{OH}) = 1 \text{ \AA}$, em que o H^+ foi transferido para O1 e o substrato é um zwitterión]. A definição adotada para coordenada de reação foi $\xi = d(\text{PO}) - d(\text{SP})$ e GLU474 ou GLU478 foram protonados (tabela 27). Nenhum perfil de energia livre para simulação 14a é mostrado aqui. Se a configuração construída na §4.1.1 é usada, o perfil de energia

Figura 44: Perfil de energia livre para simulação 14b (tabela 27 e figura 34) da reação de PhOP^{2-} catalisada pela CDC25B com um GLU no P-loop atuando como ácido geral. A legenda indica qual GLU foi protonado e qual a configuração inicial usada nas simulações (p. 199).

livre obtido é equivalente ao perfil da simulação 12 (figura 42). Não ocorre protonação do grupo de saída e os TSs e produtos apresentam a mesma geometria daquelas observadas na simulação 12. Se a configuração inicial parte daquelas obtidas na simulação 11, a desfosforilação de PhOP^{2-} é observada, mas as alturas de barreira são maiores que 40 kcal/mol e também não ocorre transferência de H^+ para O1. Se a configuração inicial parte do substrato zwitteriônico, as alturas de barreira para desfosforilação são maiores que 35 kcal/mol. O H^+ é transferido de volta para o GLU no início da simulação (na primeira janela da região dos reagentes) e os TSs e produtos observados apresentam ao grupo de saída desprotonado.

As simulações 14b e 14c foram realizadas a partir das configurações extraídas da simulação 11 (vide supra). Como discutido na §4.3.2, apresentaremos aqui apenas os resultados da simulação 14b, que são semelhantes aos resultados da simulação 14c.

Nas simulações em que o GLU474 está protonado (figura 44), os reagentes (p. 179), correspondem à janela inicial da determinação PMF. Não observamos elevação da energia livre na região dos reagentes (direita para esquerda no perfil da figura 44), porque este processo corresponde ao aumento da $d(\text{OH})$ e à diminuição da energia livre (figura 43). As

simulações são dominadas inicialmente ($\xi < 1.5 \text{ \AA}$) pela diminuição de $d(\text{OH})$. Nos MCs, o substrato está parcialmente deslocado do sítio ativo, o grupo fosfato não é coplanar ao P-loop e os oxigênios não ligantes e esterificado são parcialmente solvatados por moléculas de água, que chegam a penetrar no interior do sítio ativo e coordenam-se ao grupo tiolato da CYS473. Em $\xi \sim 1$, o H^+ ácido fica compartilhado entre a cadeia lateral do GLU474 e O1. O perfil de energia livre eleva-se até uma região quase plana (em $\xi > 1.5 \text{ \AA}$) e o TS observado corresponde à transferência de H^+ para O1. Em ξ_{TS} (tabela 33), $d(\text{PO}) \approx 2.7 \text{ \AA}$ e a ligação P–S não é formada [$d(\text{SP}) > 4 \text{ \AA}$], assim, mesmo antes do TS, a ligação P–O já está praticamente rompida e metafosfato é formado. Em $\xi > \xi_{TS}$, fenol e metafosfato permanecem solvatados por moléculas de água e coordenados ao P-loop. As simulações não foram executadas até o aumento da energia livre na região dos produtos. Todavia, como observado na §1.2.1, metafosfato é uma espécie altamente reativa e deve reagir com o solvente. O potencial híbrido utilizado não contém moléculas de água na região QM, portanto, o ataque nucleofílico da água sobre metafosfato não pode ser observado (§4.3.3). Para as simulações 14c, a distância da ligação S–P é incluída na coordenada de reação, mas, em ξ_{TS} , a $d(\text{SP}) \sim 3.2 \text{ \AA}$, a $d(\text{PO}) \sim 2.7 \text{ \AA}$ e metafosfato solvatado também é formado.

As simulações com GLU478 protonado também são inicialmente dominadas pela diminuição de $d(\text{OH})$ (figura 44). No MC, o deslocamento do substrato para fora do P-loop é menor, principalmente na simulação que parte de $d(\text{OH}) = 4.1 \text{ \AA}$. Nas janelas iniciais ($\xi < \xi_{TS}$), o hidrogênio ácido do GLU478 é transferido para um dos oxigênios não ligantes do fosfato. Para simulação partindo de $d(\text{OH}) = 2.3 \text{ \AA}$, O2 é protonado e o substrato já está bastante deslocado para fora do P-loop, com uma molécula de água penetrando no interior do sítio ativo e coordenando-se ao grupo tiolato da CYS473. Para simulação partindo de $d(\text{OH}) = 4.1 \text{ \AA}$, O4 é protonado, mas o grupo fosfato ainda apresenta de 5 a 6 ligações de hidrogênio com o P-loop (na simulação 13, por exemplo, cerca de 7 ligações de hidrogênio entre o fosfato e o P-loop são observadas no MC). Em ξ_{TS} (tabela 33), o H^+ é transferido para O1 através de um anel de 4 membros, formando um zwitterión $[\text{PhO}^+(\text{H})\text{PO}_3^{2-}]$ com $d(\text{PO}) \approx 1.9 \text{ \AA}$ e $d(\text{SP}) > 4 \text{ \AA}$. Este zwitterión é estável, formando um mínimo no perfil de energia livre em $\xi \approx 0.8 \text{ \AA}$, e sua decomposição, pela quebra da ligação P–O ou pela transferência de H^+ de volta ao oxigênio não ligante, apresenta barreiras de 5 a 10 kcal/mol (figura 44). Em $\xi > 1.5 \text{ \AA}$, a ligação P–O é rompida, produzindo fenol e metafosfato, que são totalmente solvatados por moléculas de água. A energia livre calculada (figura 44) é sempre crescente em $\xi > 0.9 \text{ \AA}$ e um TS para quebra da ligação P–O não foi determinado (vide supra e §4.3.3). Portanto, o TS caracterizado (tabela 33)

corresponde ao TS para formação do zwitterion e a altura da barreira total para reação de desfosforilação será igual ou maior do que a altura deste TS. As simulações com GLU478 protonado também não foram executadas até a caracterização de um mínimo de ΔG na região dos produtos, já que a energia livre teve um aumento constante até $\xi > 2 \text{ \AA}$.

Para a protonação de GLU474, as barreiras apresentadas na tabela 33 para simulação 14b devem ser acrescidas por 5.4 ou 3.5 kcal/mol, se a configuração inicial apresentou $d(\text{OH}) = 1.9 \text{ \AA}$ ou $d(\text{OH}) = 4.0 \text{ \AA}$, respectivamente. Para a protonação de GLU478, as barreiras são acrescidas por 12.3 e 3.8 kcal/mol, se a configuração inicial apresentou $d(\text{OH}) = 2.3 \text{ \AA}$ ou $d(\text{OH}) = 4.3 \text{ \AA}$, respectivamente. Estas energias livres adicionais correspondem à energia relativa estimada na simulação 11 (figura 43) para aproximar o ácido geral ao substrato até as $d(\text{OH})$ indicadas, a partir da configuração construída na §4.1.1. Assim, as barreiras calculadas em todas as simulações da CDC25B usam a mesma estrutura como referência ou zero de energia livre.

4.3 Discussão

4.3.1 Incertezas nas simulações

O objetivo desta seção é avaliar as possíveis incertezas presentes nas simulações e nos perfis de energia livre calculados na §4.2. As possíveis fontes de desvio são diversas e provenientes das aproximações e das metodologias empregadas. Estas fontes podem super ou subestimar as interações e as contribuições para energia livre, e os respectivos erros podem se cancelar parcialmente, dificultando ou impossibilitando uma estimativa segura de um erro ou desvio total. Portanto, não definimos ou estimamos uma faixa de precisão para as quantidades obtidas neste capítulo. O principal teste dos valores calculados é a comparação com observáveis experimentais. A comparação com diferentes propriedades pode sugerir qual a precisão das quantidades calculadas. A energia de ativação das reações catalisadas pode ser obtida (pela teoria do estado de transição, apêndice B) das constantes de velocidade de reação medidas e comparada à altura de barreira dos perfis de energia livre calculada nas simulações. Uma medida da energia livre de reação também poderia ser comparada aos valores calculados (§4.2), mas estes últimos são alterados pela complexação do reagente e do produto à enzima. Todavia, desconhecemos valores experimentais de energia de reação. A estrutura do TS calculado pode ser comparada à interpretação de medidas de KIE e o mecanismo de reação obtido na simulação também pode ser comparado à interpretação de LFER, entre outras medidas experimentais (§1.2.1).

As seguintes fontes de incerteza podem influir na determinação dos perfis de energia livre:

- Hamiltoniana QM semiempírica ($\hat{\mathcal{H}}_{QM}$);
- Interação entre as regiões QM e MM ($\hat{\mathcal{H}}_{QM/MM}$);
- Fronteira covalente entre as regiões QM e MM;
- Campo de força empírico ($\mathcal{V}_{tot}$);
- Condições de contorno e truncagem das interações não ligantes;
- Construção do modelo tridimensional protéico;
- Escolha da coordenada de reação;
- Efeitos quânticos nas propriedades dinâmicas e de equilíbrio;
- Erro estatístico e erro sistemático no cálculo do PMF;
- Amostragem do espaço de configurações.

Os seis primeiros itens compõem o potencial e o modelo usados para descrever as interações no sistema e, portanto, influenciam criticamente a determinação das energias livres. Os últimos quatro itens são inerentes à metodologia de simulação computacional empregada e, em alguns casos, suas contribuições para o erro total podem ser facilmente avaliadas.

A hamiltoniana semiempírica MNDO+G/CHOPS foi calibrada especificamente para reações de tiólise e alcoólise de ésteres de fosfato (§3.2.1), e foi amplamente testada na §3.2.2. O RMSD para a energia relativa (ΔE_{opt}) das espécies obtidas nas reações de fosfatos (reagentes, produtos, intermediários e TSs) é menor que 3 kcal/mol. Este desvio é bastante reduzido (compare, por exemplo, com os desvios desta mesma grandeza obtidos para as parametrizações semiempíricas originais, tabela 24) e mostra que a incerteza nas energias QM calculadas com a hamiltoniana calibrada são pequenas e não alteram nossas conclusões. Mesmo os desvios máximos, observados para reações de triésteres de fosfato (p. 157), não ultrapassam 7.5 kcal/mol. Para os triésteres, a tiólise está artificialmente favorecida e a alcoólise, desfavorecida. A comparação entre a catálise pela VHR selvagem e o mutante CYS124→SER pode ser influenciada por estes desvios. Todavia, o desvio observado para todas as espécies de reações de monoésteres é inferior a 3 kcal/mol. Assim,

podemos esperar que a incerteza relativa à descrição QM da desfosforilação dos substratos (monoésteres de fosfato) pelas enzimas selvagens ou pelas mutantes serão menores que 3 kcal/mol.

A transferência de H^+ entre as espécies reativas também foi calibrada e o RMSD observado para energia de reações modelo (§3.1.2) é menor que 4 kcal/mol (tabela 24). Os principais artifícios observados foram o aumento da basicidade de fenóxido e o desbalanço energético da equação 3.8, envolvendo fenilfosfato (p. 156). No caso do fenóxido, a maior basicidade não deve alterar a altura de barreira calculada na simulação 1a, porque esta espécie é neutralizada por uma transferência de H^+ do ácido geral na VHR. Medidas experimentais mostram que o β_{lg} é próximo de zero para desfosforilação catalisada pela VHR (§1.3), indicando que a basicidade do grupo de saída não influencia a catálise. Para a CDC25B, o fenóxido deve ser desestabilizado na ausência do ácido geral. Uma possível consequência é que a diferença entre as barreiras obtidas na simulação 12 e nas simulações 13 ou 14 estaria subestimada (§4.2.2). Por outro lado, o sítio ativo da CDC25B é exposto ao solvente e o fenóxido pode ser solvatado e estabilizado por interações com moléculas de água (§4.3.7). O desbalanço energético da equação 3.8 não deve influenciar as energias de transferência de H^+ entre fenilfosfato e as outras espécies reativas, já que o desvio observado na energia de reação (ΔE_{pK_a} , §3.1.2) para as equações 2.3, 3.9 e 3.11 é menor que 1 kcal/mol.

Outra possível fonte de erro na descrição QM é a diferença entre a carga total desta região em algumas simulações (por exemplo, nas simulações 1 a 4, 12 e 14, a carga da região QM foi $-3 e$) e a carga total das espécies do conjunto de treinamento na calibração da hamiltoniana semiempírica ($-1 e$ nas reações descritas na §2.2 ou $-1 e$ e $-2 e$ nas espécies descritas na §3.1.2). Contudo, no sítio ativo das PTPs, parte da carga dos grupos reativos na região QM é blindada pelas fortes interações eletrostáticas, por exemplo, com o grupo guanidina da ARG conservada. A protonação do grupo de saída ou a sua solvatação também podem contribuir para blindagem e deslocalização da carga negativa na região QM. Os TSs formados são geralmente dissociativos, assim, as reações (como o ataque nucleofílico pelo grupo tiol da CYS conservada) ocorrem essencialmente entre espécies com carga total $-2 e$.

A interação entre as regiões QM e MM também foi testada (§3.2.3). A tentativa de calibração desta interação pela otimização dos parâmetros de LJ não foi bem sucedida (§3.2.3.2) em razão da limitação da forma funcional para interação de van der Waals ($\mathcal{H}_{QM/MM}^{vdW}$). A interação QM/MM (usando os parâmetros de LJ originais do OPLS-AA

para os átomos QM) que mimetiza a ligação de hidrogênio entre o grupo fosfato e modelos da cadeia protéica principal e lateral da ARG é subestimada, em média, em cerca de 5 kcal/mol (§3.2.3), sugerindo que a interação do grupo fosfato com os resíduos no P-loop nas simulações também pode estar subestimada. Por outro lado, a interação QM/MM entre fenol e água, e entre ácido acético e água é superestimada em cerca de 2 kcal/mol (§3.2.3), principalmente em distâncias maiores que a distância de mínimo ($R > R_e$). Estas energias de interação QM/MM foram determinadas em complexos bimoleculares (§3.1.3.1), dificultando a estimativa do desvio em uma interação de muitos-corpos, como é observado no sítio ativo das PTPs (por exemplo, o mesmo oxigênio do grupo fosfato pode realizar até três ligações de hidrogênio com a cadeia protéica, tabela 29). Porém, o desvio na interação de muitos-corpos não é aditivo e deve ocorrer cancelamento de erro. Podemos especular que a barreira calculada para a VHR selvagem está superestimada (para metil e fenilfosfato como substratos, §4.2.1.2), porque a energia de interação das ligações de hidrogênio do grupo fosfato com a cadeia protéica está subestimada. A barreira para desfosforilação da PhOP^{2-} catalisada pela CDC25B está subestimada (§4.2.2), porque a energia de interação entre fenol e água está superestimada.

A transferência e a polarização eletrônica da região MM em resposta a mudanças na estrutura eletrônica da região QM ou por alterações da geometria da região MM não foram incluídas no potencial híbrido. Medidas de KIE sugerem que a interação do substrato com a cadeia protéica é predominantemente eletrostática (p. 215). Portanto, a contribuição da transferência e da polarização eletrônica na região MM para interação com o substrato deve ser relativamente menor e sua negligência não deve aumentar significativamente para a incerteza das energias livres calculadas.

Como discutido no final da §3.2.4, o erro associado à fronteira covalente entre as regiões QM e MM deve ser pequeno para as reações de ataque nucleofílico ao fosfato (dianiônico) e para transferência de H^+ do ácido geral (ASP ou GLU). No entanto, a protonação da cadeia lateral da CYS e da SER podem apresentar desvios mais significativos, superestimando as barreiras para transferência de H^+ , envolvendo estes resíduos. Este processo não é importante para as conclusões obtidas aqui. No sítio ativo, o pK_a da CYS é alterado de modo que esta espécie não estará protonada, nem mesmo no MC (§4.3.5).

A sensibilidade das diferenças de energia livre aos parâmetros do campo de força usado na descrição da região MM é pequena em razão do cancelamento de desvios^[180]. No entanto, se um campo de força for usado para cálculo de energias livres, propriedades termodinâmicas devem ser usadas na sua parametrização^[267]. Embora os campos de força

geralmente utilizados, como AMBER e CHARMM, sejam calibrados usando critérios estruturais, o OPLS-AA, usado nesta tese (§4.1.2), foi ajustado para reproduzir calores de vaporização e a densidade de líquidos simples e, portanto, é mais indicado e deve apresentar melhor desempenho na determinação de energias livres^[174].

O congelamento das posições atômicas na região exterior da macromolécula solvatada (figura 33) pode ser comparado a uma parede rígida e estendida que confina o volume do sistema simulado. Os efeitos deste tipo de condição de contorno no cálculo de energias livres não foram adequadamente analisados^[180], mas testes (não mostrados aqui) em que a região móvel é ampliada para o sistema estudado nesta tese e para outras enzimas^[260, 261] indicam que a influência desta aproximação no PMF calculado é pequena (menor que 1 kcal/mol). A truncagem das interações não ligantes pelo esquema de mudança da força (§1.6.2) apresenta resultados superiores em relação a outros esquemas de truncagem^{[179][24, §9.2]}. Comparações em outros sistemas indicaram que os perfis de energia livre obtidos com raios de corte similares aos usados nesta tese (§4.1.2) diferem muito pouco dos perfis obtidos sem truncagem^[179, 260, 261].

Os modelos estruturais obtidos após as etapas de construção e equilibração das enzimas VHR e CDC25B solvatadas (§4.1.1) devem ser bastante próximos às estruturas experimentais dos complexos das respectivas enzimas com fenilfosfato. Todavia, as estruturas iniciais usadas na simulação de metilfosfato como substrato e dos mutantes da VHR foram derivadas da estrutura equilibrada para a enzima selvagem sem nenhuma construção de moléculas de solvente ou equilibração adicionais (§4.1.1). Assim, a estrutura usada nestas simulações pode ser diferente da estrutura experimental. Esta diferença é proporcional à perturbação que a mutação ou o metilfosfato causam na estrutura do MC com PhOP²⁻ selvagem, ou seja, se a mutação for conservativa, o modelo estrutural usado na simulação do mutante será adequado (próximo à estrutura experimental). O mesmo raciocínio pode ser aplicado para mudança do substrato. Caso contrário, as estruturas simuladas e observadas experimentalmente serão diferentes. Longas simulações de dinâmica molecular a partir da estrutura selvagem seriam necessárias para permitir que mudanças conformacionais adequadas para a simulação do mutante sejam observadas.

A escolha da coordenada de reação é arbitrária e introduz incertezas na determinação dos PMFs, pois a coordenada escolhida pode ser diferente da coordenada ideal para descrição da reação (§4.3.2) ou, em casos extremos, representar um processo que leve a um produto diferente daquele de interesse. No primeiro caso, as alturas de barreira calculadas serão maiores do que a barreira obtida pela coordenada de reação ideal. Contudo,

os testes da §4.2.1.1 e a proximidade entre a altura de barreira calculada para simulação 1a e a energia de ativação experimental indicam que a escolha da coordenada de reação, $d(\text{PO}) - d(\text{SP})$, foi correta e a incerteza associada não é significativa.

O comportamento mecânico-clássico (§1.6.2) foi adotado para descrição dos sistemas simulados. No entanto, efeitos quânticos podem ser importantes para determinação de propriedades dinâmicas (§1.1.2.9) e de equilíbrio (como alturas de barreira)^[268] em alguns fenômenos, como transferências de H^+ (§1.1.2.9). Uma maneira de estimar efeitos quânticos em propriedades de equilíbrio é a integral de trajetória de Feynman^[269]. Resultados obtidos utilizando esta metodologia e potenciais QM puros^[179] ou híbridos^[268], indicaram que os efeitos quânticos podem diminuir em cerca de 30% as alturas de barreira de reações em que a transferência de H^+ é a principal componente da coordenada de reação. Para a reação de desfosforilação catalisada pelas PTPs, uma transferência de H^+ pouco contribui para coordenada de reação (§4.3.2) e, portanto, os efeitos quânticos devem estar minimizados. A comparação da catálise pela VHR selvagem e pela mutante ASP92→ALA (§4.2.1) indica que a contribuição da protonação do grupo de saída para o abaixamento da barreira é cerca de 5 kcal/mol (§4.3.4). Assim, mesmo que efeitos quânticos sejam importantes para esta transferência, o erro observado em consequência da descrição clássica deve ser menor que 1.5 kcal/mol.

Erros chamados de estatísticos e de sistemáticos são inerentes a simulações computacionais e a algoritmos para determinação de energia livre. O erro estatístico é a diferença da energia livre determinada para o mesmo processo em simulações diferentes (por exemplo, que partam de condições iniciais diferentes) e pode ser corrigido por uma média sobre várias simulações (com diferentes condições iniciais)^[270, 271]. Uma fonte de erros estatísticos é a falta de ou a incompleta equilibração da simulação em cada uma das janelas definidas pelo potencial guarda-chuva para determinação do PMF. Os perfis de energia livre obtidos para as simulações 1a e 12, usando uma etapa de equilibração de 15 ps de dinâmica molecular, foram essencialmente idênticos aos perfis obtidos com 4 ps de equilibração (§4.1.3). Tempos de equilibração de 3 a 5 ps são normalmente usados em outras simulações enzimáticas^[260, 261, 272] e, logo, consideramos que a equilibração das janelas usadas para determinação dos PMFs da §4.2 foi adequada e não introduz incertezas. Outra fonte de desvios que pode ser considerada um erro estatístico é a incapacidade de simulações computacionais em amostrar todo o espaço de configurações^[180, 271]. Isto pode ser corrigido, a princípio, testando a convergência da energia livre calculada em relação ao tempo de simulação (coleta de dados, §4.1.3), realizando uma média sobre diferentes simulações (vide supra) ou usando correções estatísticas (o que é dificultado pela alta

correlação entre as configurações calculadas)^[180, 271]. Não realizamos um conjunto de simulações para cada um dos sistemas ou definições dados na tabela 27. No entanto, o perfil de energia livre para as simulações 1a e 12 obtido com 30 ps de dinâmica molecular durante a coleta de dados também foram similares (as diferenças não ultrapassam 0.5 kcal/mol) aos PMFs mostrados para estas simulações nas figuras 36 e 42. Portanto, concluímos que o erro estatístico, mesmo usando apenas uma determinação (e não uma média) para cada simulação, é bastante reduzido. O algoritmo de dinâmica molecular Verlet-Langevin possibilita uma melhor amostragem do espaço de configurações do que outros algoritmos usualmente empregados em simulações moleculares (§1.6.2). O erro sistemático é atribuído à metodologia ou ao algoritmo usado na determinação da energia livre do processo simulado, por exemplo, o uso de um potencial guarda-chuva (§1.6.3) pode influenciar a determinação do PMF^[180]. Métodos para correção do erro sistemático foram propostos^[270]. Todavia, este tipo de desvio deve ser reduzido se a influência dos potenciais guarda-chuva é corretamente retirada na determinação do PMF^[182]. O método WHAM (§4.1.3) foi usado para esta tarefa e as janelas vizinhas se recobriram no espaço de configurações para todas simulações, como pode ser observado na figura 45, por exemplo, para simulação 1a (distribuições semelhantes foram observadas nas outras simulações).

Finalmente, o ensemble canônico (§1.6.3) foi usado nas determinações dos PMFs, portanto, as energias livres calculadas são de Helmholtz que, para as reações em fase condensada estudadas, devem ser muito próximas às energias livres de Gibbs (que são as quantidades mostradas nas figuras de perfil de energia livre e obtidas das medidas experimentais)^[12].

Já que parte de nossas conclusões (por exemplo, as respostas às perguntas da p. 44) são tiradas de comparações entre sistemas semelhantes, parte dos desvios observados também serão os mesmos. Por exemplo, a diferença entre as simulações 1a e 7 (tabela 27) é apenas a substituição do átomo de enxofre por um de oxigênio, assim, excetuando-se desvios na hamiltoniana QM, os erros provenientes de outras fontes devem ser essencialmente os mesmos nas duas simulações.

4.3.2 Escolha da coordenada de reação

As coordenadas de reação são definidas como as posições cuja flutuação resulta em transições entre estados estáveis do sistema, como, por exemplo, reagentes e produtos de uma reação química^[262]. Para sistemas simples, com até uma ou duas dezenas de átomos, a coordenada de reação e o mecanismo de interesse (que leva dos reagentes aos produtos

Figura 45: Distribuição das ocorrências de um dado valor de coordenada de reação para simulação 1a alterada pelo potencial guarda-chuva. A ordenada é o número de ocorrências de um valor ξ . As janelas adjacentes estão mostradas em linha cheia e tracejada para facilitar a visualização do recobrimento de cada janela.

da reação estudada) podem ser determinados, negligenciando-se os efeitos entrópicos e analisando-se a PES do sistema, especialmente as regiões de TSs (que devem corresponder ao gargalo dinâmico da transição, pois são regiões de alta energia em relação aos reagentes)^[262]. O uso da intuição química sobre quais deslocamentos ou flutuações devem contribuir para a transição entre as regiões de mínimo da PES também é importante para determinação da coordenada e do mecanismo de reação. Este foi o esquema empregado na determinação dos mecanismos das reações na fase gasosa estudadas no cap. 2.

Para sistemas complexos, como macromoléculas solvatadas, a situação é bem diferente. Embora os estados estáveis (agora, necessariamente mínimos de uma *superfície de energia livre*) ainda sejam separados por barreiras energéticas apreciáveis (em relação a $k_B T$), uma coleção (muitas vezes inumerável) de TSs existe e a escolha de uma coordenada de reação (e um TS) que conecte os estados de interesse pode ser muito difícil, já que estas coordenadas podem ser variáveis coletivas num espaço multidimensional. Por exemplo, para transferências de elétrons, a coordenada relevante é a diferença de energia entre os estados, que depende de várias posições atômicas das moléculas reativas e do

solvente^[71, 262]. Desta maneira, a escolha *a priori* de uma coordenada de reação introduz incertezas na determinação dos perfis de energia livre (§4.3.1) e requer ainda maior uso e confiança na intuição química.

Na §4.2.1.1, apresentamos perfis de energia livre obtidos para diferentes coordenadas de reação que são combinações de distâncias de ligações que devem se formar ou se romper durante a reação de desfosforilação catalisada por PTPs (§1.3). Todas as propostas de ξ conectam os reagentes (MC) aos produtos (substrato desfosforilado e enzima tiofosforilada, §4.2.1.1). A melhor escolha é aquela que apresenta menor energia de ativação, da mesma forma que um TS (ou a coleção de TSs que divide as regiões de reagentes e produtos em sistemas complexos, chamada de *superfície separatriz*)^[262] é determinado pela teoria variacional do estado de transição^[35, 263].

A menor barreira corresponde à $\xi = d(\text{PO}) - d(\text{SP})$ (simulação 1a, tabela 27) e é bastante próxima da energia de ativação experimental (§4.2.1.4), indicando que esta escolha de coordenada de reação é adequada. Esta é a mesma ξ usada para descrever tanto as reações de tiólise de fosfato apresentadas no cap. 2 como a reação estudada em outra simulação da catálise por PTPs^[110]. Uma vantagem desta coordenada é que a natureza associativa ou dissociativa da reação não é alterada, já que a restrição do potencial guarda-chuva, usado durante a obtenção do PMF (§4.1.3), é imposta à combinação das distâncias de ligação e não a cada distância separadamente^[110]. Simulações de outros sistemas indicaram que a IRC determinada na reação em fase gasosa pode ser usada como ξ para a reação equivalente em solução aquosa ou catalisada por uma enzima^[12]. A IRC determinada no cap. 2 para as reações de tiólise de arilderivados de mono e triésteres de fosfato segundo mecanismos associativos ($A_n + D_n$) é dominada pelas distâncias $d(\text{PO})$ e $d(\text{SP})$. A IRC para as mesmas reações segundo mecanismos $D_n + A_n$ é dominada pela transferência do H^+ , ou seja, pela distância do H^+ ao seu doador e ao seu acceptor (§2.3.2).

A comparação dos perfis de energia livre e dos mecanismos observados nas simulações 1a e 1b sugere que a principal componente da ξ para descrever a reação de desfosforilação é a distância da ligação P–O que se rompe. As alturas de barreira calculadas nestas simulações diferem menos que 2 kcal/mol (tabela 28).

A inclusão de $d(\text{OH})$ em ξ não é necessária para que a transferência de H^+ do ASP92 para o grupo de saída aconteça na VHR. Na simulação 1a (§4.2.1.1), a transferência de H^+ ocorre diretamente para o grupo de saída na região do TS e simultaneamente à quebra da ligação P–O. Esta observação é equivalente à interpretação de medidas experimentais (§4.3.3). Por outro lado, a inclusão de $d(\text{OH})$ em ξ implica em mudanças do mecanismo

da reação, uma vez que a transferência de H^+ ocorre inicialmente para um oxigênio não ligante do grupo fosfato. O pK_a dos oxigênios não ligantes deve ser maior que o pK_a do oxigênio esterificado do substrato. A protonação do grupo fosfato reduz a repulsão com a CYS nucleofílica e a tiólise ocorre com maior participação do nucleófilo. No entanto, as barreiras observadas são no mínimo 10 kcal/mol mais altas do que a barreira observada quando a transferência de H^+ ocorre diretamente para o grupo de saída (simulação 1a). A interpretação de medidas experimentais, como KIE, indica que a transferência de H^+ não envolve os oxigênios não ligantes na etapa determinante da velocidade de reação (§4.3.5). Na simulação 1d, a protonação preserva o caráter dissociativo do mecanismo, porque as distâncias $d(PO)$ e $d(SP)$ não fazem parte de ξ , mas na etapa de quebra da ligação P–O, $d(SP) > 3.0 \text{ \AA}$.

A comparação entre as simulações 1c e 1e sugere que a inclusão de $d(SP)$ em ξ não é necessária, já que os mecanismos observados nas duas simulações são equivalentes, assim como as barreiras calculadas (tabela 28). Portanto, apenas o perfil referente à simulação 14b é mostrado na §4.2.2. O perfil e o mecanismo calculados para a simulação 14c não sugerem conclusões diferentes do que a simulação 14b.

Também usamos diferentes coordenadas de reação para a catálise pela CDC25B (§4.2.2), principalmente na reação em que os resíduos GLU do P-loop foram testados como doadores de H^+ (§4.2.2.2). A introdução de $d(OH)$ em ξ nas simulações 11, 14b ou 14d resulta na transferência de H^+ diretamente para o oxigênio esterificado do substrato se o GLU474 está protonado, ao invés do H^+ ser transferido inicialmente para um dos oxigênios não ligantes, como observado na VHR ou para GLU478 protonado na CDC25B. As barreiras são elevadas em ambos os casos, indicando que a transferência de H^+ direta ou indireta não é favorecida na CDC25B.

A reação de desfosforilação acompanhada de catálise ácido geral pode ocorrer segundo três mecanismos: com as duas transferências (do grupo fosfato e do H^+) ocorrendo simultaneamente e um único TS (como observado na simulação 1a para a VHR), ou em dois passos distintos, apresentando dois TSs separados por um mínimo na superfície de energia livre. No caso de dois passos, a transferência de H^+ pode ocorrer antes da transferência de fosfato, ou vice-versa. Nas simulações 14a, cujas configurações iniciais foram obtidas dos produtos da simulação 14d (em que o substrato era um zwitterion, §4.2.2.2), o processo simulado correspondeu a reação em dois passos, com a transferência de H^+ na primeira etapa. O H^+ é transferido de volta ao GLU no início da simulação 14a e, mesmo se um potencial harmônico é usado para restringir a ligação O–H entre o oxigênio esterificado

(resultados não mostrados aqui), as barreiras para o passo de transferência de fosfato são elevadas. Nas simulações 14a cujas configurações iniciais são retiradas da simulação 11 (§4.2.2.2) ou da estrutura construída na §4.1.1, a primeira etapa corresponde à desfosforilação, mas nenhuma transferência de H^+ é observada posteriormente. O mecanismo simultâneo poderia ser observado nas simulações 14a ou nas simulações 14b e 14c. No entanto, nestas últimas simulações, somente a transferência de fosfato é observada. Uma maneira mais adequada de avaliar a viabilidade do mecanismo simultâneo seria a avaliação de PMFs bidimensionais^[182], $\mathcal{W}(\xi^{(1)}, \xi^{(2)})$, em que $\xi^{(1)}$ indicaria a transferência de H^+ e $\xi^{(2)}$ indicaria a transferência de fosfato. Contudo, a determinação de PMFs bidimensionais não está implementada na biblioteca de rotinas DYNAMO^[179], usada para as computações dos perfis de energia livre.

4.3.3 Mecanismo da reação no sítio ativo das PTPs

Nesta seção, detalhes do mecanismo de reação de fenilfosfato dianiônico calculado no sítio ativo da VHR e da CDC25B, e sua relação com observações experimentais será discutida.

Para a reação catalisada pela VHR, a $d(SP)$ (figura 34) diminui de 3.58 ± 0.05 Å no MC até 2.74 ± 0.10 Å no TS, caracterizando um TS com pouca (no entanto, diferente de zero) formação da ligação P–S (tabela 29). A soma do desvio padrão da distância $d(SP)$ (0.10 Å no TS) com o $RMSD_{max}$ observado para cálculo das distâncias com o conjunto semiempírico MNDO+G/CHOPS (0.08 Å, tabela 24) indica uma incerteza de até 0.18 Å na $d(SP)$ obtida na simulação enzimática. Na reação F (tiólise de $PhOP^-$ determinada *ab initio*, §2.2.2), $d(SP) = 2.79$ Å no TS do mecanismo $A_n + D_n$ e $d(SP) = 2.98$ Å no TS do mecanismo $D_n + A_n$. Assim, a $d(SP)$ na enzima é próxima à $d(SP)$ para as reações associativa ou dissociativa de um monoânion na fase gasosa. O TS na enzima é dissociativo e o grupo de saída tem parte de sua carga negativa neutralizada, como na reação em fase gasosa. A $d(PO)$ (figura 34) aumenta gradativamente de 1.77 ± 0.04 Å no MC (distância característica do fenilfosfato em equilíbrio), até 2.40 ± 0.14 Å no TS (tabela 29), caracterizando um TS com substancial ou avançada quebra da ligação P–O do grupo de saída. Na reação de dissociação de $PhOP^-$ em fase gasosa (§2.2.2), a $d(PO) = 2.38$ Å, próximo ao valor observado na enzima. Nas reações dissociativas em fase gasosa, uma transferência de H^+ do oxigênio não esterificado do grupo fosfato participa da coordenada de reação (§2.2.2) e, portanto, as comparações das distâncias de ligação no vácuo e na enzima devem ser cautelosas. As distâncias dos oxigênios não ligantes (equatoriais) do

grupo fosfato (O2, O3 e O4, figura 37) sofrem pequena diminuição na região do TS da reação enzimática (tabela 29). Este comportamento é característico de mecanismos dissociativos, em que o TS tem caráter de metafosfato. Na reação de dissociação de PhOP^- em fase gasosa (§2.2.2), a distância entre o fósforo e o oxigênio não ligante é 1.49 Å no TS e 1.51 Å no reagente.

Os valores médios de $d(\text{OH})$ e $d(\text{AcOH})$, e os elevados desvios padrão para estas distâncias (tabelas 27 e 29) indicam que o H^+ é transferido do ácido geral para o grupo de saída no TS. Esta transferência é espontânea, ou seja, ocorre naturalmente durante a dinâmica molecular, sem que $d(\text{OH})$ ou $d(\text{AcOH})$ sejam incluídas na coordenada de reação. A protonação do grupo de saída anula a carga negativa formada durante a dissociação. Esta observação é equivalente à interpretação das medidas de KIE [$^{15}(\text{V}/K)$ e $^{18}(\text{V}/K)_{\text{bridge}}$, vide infra].

Diversas ligações de hidrogênio estabilizam as espécies ao longo da reação. O enxofre da cadeia lateral da CYS124 forma ligações com os hidrogênios da cadeia principal (HN) e lateral (HO) da SER131. As distâncias de ambas ligações aumentam com o progresso da reação (tabela 29), sugerindo que as ligações de hidrogênio são enfraquecidas e que menos estabilização do grupo tiol é necessária à medida que a ligação P–S é formada (por outro lado, estas ligações de hidrogênio devem ser importantes para estabilização do grupo tiol ionizado na CYS124). A ligação de hidrogênio do enxofre da CYS124 com a cadeia lateral da SER131 parece não ter um papel catalítico, pois k_2 (equação 1.9) e KIE (tabela 3) não são alterados com a mutação SER131→ALA (§1.3). Uma ligação bidentada entre a cadeia lateral da ARG130 (HN1 e HN2, tabela 29) e os oxigênios não ligantes do grupo fosfato (O3 e O2, respectivamente) é formada logo após o MC (no MC, a distância entre O2 e HN2 ainda é grande, mas diminui sensivelmente nas janelas seguintes). Os oxigênios não ligantes O3 e O4 fazem duas ligações de hidrogênio cada, e O2 faz três ligações. A distância de ligação entre O3 e HN1 (cadeia lateral da ARG130, figura 37) se mantém constante durante todo progresso da reação, sugerindo que esta ligação é bastante estável. A ligação entre O2 e HN da ARG125 atinge uma distância mínima no TS e a distância da maioria das ligações de hidrogênio contendo o grupo fosfato e os resíduos no P-loop diminui com o progresso da reação, sugerindo que estas ligações de hidrogênio são uma das principais fontes de catálise nas PTPs (§1.1.2.6).

O KIE foi usado para investigar a estrutura do TS e o mecanismo da reação de ésteres de fosfato em solução e catalisada por PTPs (§1.2.1 e §1.3.3)^[273]. O $^{15}(\text{V}/K)$ medido para reação catalisada pela VHR (tabela 3) é unitário e indica que o grupo de saída está

protonado no TS, como observado na simulação 1a (figura 37). O $^{18}(V/K)_{bridge}$ medido é um compromisso entre a quebra da ligação P–O do grupo de saída (torna o KIE maior que um) e a transferência de H^+ para o grupo de saída (torna o KIE menor que um). Portanto, o $^{18}(V/K)_{bridge}$ para VHR é menor que o valor observado em solução para a reação do diânion (em que não ocorre protonação do grupo de saída, tabela 1), mas ainda indica substancial quebra da ligação P–O no TS. Esta observação experimental é equivalente à estrutura do TS calculada na simulação da reação 1a. O $^{18}(V/K)_{nonbridge}$ medido para VHR também é unitário indicando que o TS tem caráter de metafosfato, como observado na reação de pNPP dianiônico em solução (tabela 1) e na simulação (vide supra). O $^{18}(V/K)_{nonbridge}$ pode ser sensível à ligação do substrato ao sítio ativo, caso a interação (principalmente as ligações de hidrogênio dos oxigênios não ligantes do grupo fosfato com o grupo guanidina da ARG130) tenha caráter covalente ou apresente transferência ou polarização eletrônica acentuadas^[273]. No entanto, os $^{18}(V/K)_{nonbridge}$ medidos são próximos aos $^{18}k_{nonbridge}$ para a reação do diânion (tabela 1), sugerindo que as ligações de hidrogênio devem ter caráter eletrostático^[20, 273] e que o uso dos potenciais híbridos segundo o particionamento adotado para as regiões QM e MM é adequado.

As comparações feitas (vide supra) entre o mecanismo de reação na VHR com a tiólise de $PhOP^-$ na fase gasosa (reação F, §2.2.2) são igualmente válidas para a reação na CDC25B. Ambos os mecanismos de reação catalisada são A_nD_n com TSs dissociativos, com caráter de metafosfato. A $d(SP)$ no TS da CDC25B é ligeiramente menor que no TS da VHR (tabelas 29 e 34), indicando maior participação do nucleófilo na primeira enzima, o que pode ser justificado pela necessidade de expulsão (via interação eletrostática entre a carga negativa no tiolato e no fenóxido) do grupo de saída na ausência de protonação. Na CDC25B, a distância entre os oxigênios não ligantes e o fósforo também é diminuída no TS em relação aos reagentes e produtos, como é característico de mecanismos dissociativos (vide supra).

A ligação de hidrogênio entre o tiolato e a cadeia protéica do GLY480 na CDC25B parece ligeiramente enfraquecida com o progresso da reação e a fosforilação do grupo tiol (tabela 34). Novamente, as distâncias das ligações de hidrogênio realizadas entre o P-loop e o grupo fosfato diminuem com o progresso da reação e, em alguns casos (por exemplo, na ligação entre O4 e o HN da ARG479), apresentam um valor mínimo no TS, sugerindo que estas ligações são responsáveis por parte da catálise também na CDC25B. Uma importante diferença entre a VHR e a CDC25B é a exposição ao solvente. Para VHR, apenas uma molécula de água faz uma ligação de hidrogênio com O1 (oxigênio do grupo de saída) no TS (tabela 29), a segunda molécula de água mais próxima está a

cerca de 3.5 Å de O1. Para a CDC25B, três ligações de hidrogênio são realizadas entre moléculas de água e O1 no TS, e outras três ligações com o solvente envolvem os oxigênios não ligantes (tabela 34). Portanto, a maior exposição ao solvente na CDC25B e as ligações de hidrogênio com a água devem estabilizar a forma ionizada do grupo de saída.

Os $^{15}(V/K)$ e $^{18}(V/K)_{bridge}$ medidos para a reação de pNPP catalisada pela CDC25A são bem maiores que os mesmos KIEs medidos para catálise pela VHR (tabela 3) e comparáveis aos KIEs medidos em solução aquosa para a reação de pNPP dianiônico (tabela 1), sugerindo que o grupo de saída está desprotonado e que o mecanismo é dissociativo, com elevada quebra da ligação P–O, exatamente como observamos na simulação 12 (§4.2.2 e §4.3.7). O $^{18}(V/K)_{bridge}$ para CDC25A é um pouco maior que o KIE medido para a VHR mutante ASP92→ALA, em que a catálise ácido-base geral não ocorre. A diferença pode ser consequência da maior solvatação do grupo de saída na CDC25A, já que a formação da ligação com o nucleófilo no TS calculado nas duas enzimas (CDC25B e VHR mutante) é semelhante e ligeiramente maior que a participação do nucleófilo na VHR selvagem (§4.2.1). O $^{18}(V/K)_{nonbridge}$ medido para CDC25A também indica um mecanismo dissociativo, com caráter de metafosfato, mas é um pouco menor que o KIE observado em solução para o monoéster dianiônico e para VHR mutante ASP92→ALA (tabela 3), o que levou Hengge, Zhang et al.^[105] a propor que “CDC25s empregam uma estratégia para estabilização do TS diferente daquela usada pelas PTPs selvagens e dos mutantes com o ácido geral removido”. No entanto, a diferença entre os KIEs é reduzida e pode ser atribuída a uma combinação de interações dos oxigênios não ligantes com a cadeia protéica e com moléculas de água, que estão ausentes ou diminuídas nas reações em solução aquosa e na VHR mutante, respectivamente.

Nas simulações 14b, metafosfato é formado, parcialmente complexado ao P-loop e solvatado por moléculas de água. Esta espécie é altamente reativa e a reação com uma molécula de solvente deve ser bastante rápida (como observado em solução aquosa, §1.2.1). Não simulamos o ataque nucleofílico de água sobre o metafosfato, porque o potencial híbrido utilizado não foi calibrado para esta reação. No entanto, evidências experimentais indicam que metafosfato livre não é formado, ou seja, o grupo fosfato é transferido para o tiolato da CYS conservada nas PTPs (§1.3) e nunca diretamente para uma molécula de solvente. Por exemplo, o intermediário tiofosforilado da VHR foi identificado^[96] e a enzima mutante CYS124→SER não tem atividade catalítica. Caso metafosfato livre fosse formado, a participação do nucleófilo seria irrelevante e alguma atividade seria detectada neste mutante (§4.3.6). Outra evidência para presença de metafosfato livre é a inversão de configuração de um fosfato quiral (§1.2.1). Todavia, total retenção de configuração é

observada na reação catalisada por PTPs, pois este é o resultado estereoquímico de uma dupla inversão de configuração (em virtude das duas transferências nucleofílicas, equações 1.9 e 1.10)^[66].

Finalmente, quais são os mecanismos moleculares de catálise enzimática (§1.1) usados pelas PTPs para acelerar a velocidade de desfosforilação do substrato? Uma resposta quantitativa requer a simulação da reação de referência em solução aquosa e a avaliação das interações ou forças microscópicas como discutido na §1.1. Estas simulações não foram realizadas (p. 170). Todavia, podemos especular sobre quais são os mecanismos moleculares e suas relativas contribuições para catálise.

A inspeção da estrutura do MC na VHR já indica que a interação eletrostática (§1.1.2.6) é essencial para atividade das PTPs. Todos hidrogênios da cadeia protéica principal no P-loop estão direcionados e coordenados ao substrato. As distâncias da maioria das ligações de hidrogênio diminuem e se aproximam dos valores de mínimo (R_e) observados nos complexos bimoleculares modelo (§3.2.3.2) com o progresso da reação enzimática, sugerindo que as interações são fortalecidas, ou seja, que o TS e o produto de reação são estabilizados. O substrato é uma molécula dianiônica e sua interação com o grupo catiônico da ARG conservada, no ambiente livre de solvente aquoso do P-loop, também é essencial para catálise. A mutação desta ARG para uma ALA (cuja cadeia lateral é neutra e apolar) resulta em uma elevação das energias livres calculadas para a barreira e para a reação em 17 a 26 kcal/mol, respectivamente. Estas diferenças, embora não sejam comparáveis às observações experimentais em razão da provável inadequação da estrutura enzimática usada na simulação (§4.3.1), indicam a magnitude da estabilização eletrostática promovida pelas PTPs. A estabilização eletrostática também deve ser responsável pela blindagem da carga negativa do TS com caráter de metafosfato, que facilita o ataque nucleofílico pela cadeia lateral da CYS124 e a formação do intermediário tiofosforilado.

A catálise ácido-base geral (§1.1.2.5) também contribui para a aceleração da reação de desfosforilação. O mutante ASP92→ALA na VHR tem sua atividade significativamente reduzida e apresenta aumento de 5 kcal/mol na energia de ativação. Portanto, contribuições desta magnitude podem ser atribuídas à catálise ácido-base geral na primeira etapa de reação enzimática. Na CDC25B, em que a catálise ocorre sem a participação de um ácido geral (§4.3.7), a energia de ativação para tiólise de fenilfosfato é maior que na VHR (tabela 2). No entanto, a exposição do sítio ativo e, principalmente, do grupo de saída ao solvente nas CDC25s pode estabilizar a carga negativa presente no grupo

de saída no TS. Esta estratégia usada pelas CDC25s é diferente daquela observada nas outras PTPs.

A catálise covalente (§1.1.2.4) também deve ser um mecanismo de catálise importante para atividade das PTPs, principalmente para a reatividade do intermediário tiofosforilado, porque a hidrólise da ligação P–S ocorre mais facilmente do que a hidrólise do substrato natural (ligação P–O)^[28] e de um possível intermediário fosforilado do mutante CYS→SER (novamente, quebra de uma ligação P–O, §4.3.6). Assim, a reação na VHR deve seguir pelo menos as duas primeiras condições listadas na p. 11 (para que a catálise covalente seja efetiva). Em solução aquosa, a energia de ativação para hidrólise de tioéster de fosfato é cerca de 5 a 10 kcal/mol mais baixa do que a hidrólise do oxiéster correspondente^[27, 28] e, portanto, contribuições desta magnitude podem ser atribuídas à catálise covalente. O ataque nucleofílico ao éster de fosfato na primeira etapa de catálise (equação 1.9) não depende da natureza do nucleófilo enzimático, caso este esteja ionizado, já que a altura de barreira para a reação do mutante CYS124→SER com a cadeia lateral ionizada é similar à altura de barreira na reação da enzima selvagem (§4.2.1.5). Como discutido na §4.3.6, a escolha de CYS também parece estar ligada ao menor pK_a do grupo tiol em relação ao hidroxila (e, portanto, a maior facilidade para ionização do tiol).

Cabe aqui uma crítica à observação feita por Zhang que “PTPs não alteram o estado de transição para hidrólise de fosfatos e sua eficiência catalítica é alcançada estabilizando o TS da reação em solução”^[86]. Esta visão é no mínimo simplista, porque a catálise pelas PTPs ocorre em duas reações desacopladas, tiólise do éster de fosfato e hidrólise do tioéster formado (equações 1.9 e 1.10), e não ocorre diretamente pela hidrólise do éster de fosfato, mesmo que os produtos formados nos dois processos (enzimático e em solução) sejam os mesmos. Assim, a comparação entre o TS (por exemplo, usando um ciclo termodinâmico como a equação 1.5) da reação catalisada e em solução deve ser feita com o TS de tiólise de fosfato (no caso da primeira reação enzimática, 1.9). Contudo, a tiólise de monoéster de fosfato não é observada em solução (na verdade, nucleófilos centrados em enxofre são extremamente reticentes frente a centros eletrofílicos de fósforo, §1.2.1) e, portanto, não faz sentido dizer que a enzima estabiliza um TS que, até hoje, não foi identificado. Acreditamos que as PTPs estabilizam um TS dissociativo, com caráter de metafosfato, para reação de tiólise segundo um mecanismo A_nD_n . Uma simulação da tiólise em solução aquosa, na ausência da enzima, pode indicar se o mecanismo e o TS com barreira mais baixa são semelhantes àqueles calculados na enzima e, neste caso, poderíamos dizer que a enzima estabiliza o TS da reação em solução (por um ciclo como a equação 1.5). Caso contrário, ou seja, se o TS e o mecanismo em solução forem diferentes, devemos dizer que

a enzima utiliza um mecanismo alternativo e que os TSs em solução e na enzima não são diretamente comparáveis.

4.3.3.1 Simulações da catálise por uma LM-PTP

Três estudos computacionais, utilizando diferentes metodologias, simularam a primeira etapa da catálise de desfosforilação de fenilfosfato dianiônico (equação 1.9), usando a mesma LM-PTP como modelo protéico^[110, 112, 114]. Kolmodin e Åqvist^[112], e Gao et al.^[110] obtiveram $\Delta G^\ddagger = 14$ kcal/mol, que coincide com a energia de ativação determinada experimentalmente para esta enzima, 14.3 kcal/mol (tabela 2)^[110]. Kolmodin e Åqvist obtiveram $\Delta G_{reac} = -13$ kcal/mol e Gao et al. não apresentaram energias de reação, mas pode-se deduzir pelo PMF mostrado^[110] que a reação calculada é exoérgica. Estas diferenças de energia livre são comparáveis àquelas obtidas na §4.2.1 para simulação da VHR.

Contudo, as seguintes importantes diferenças são observadas entre os resultados e suas interpretações dadas por Kolmodin e Åqvist, Gao et al. e aquelas apresentadas nesta tese:

1. Na simulação de Gao et al., o H^+ do ácido geral não é transferido para o grupo de saída que é fenóxido nos produtos;
2. O mecanismo calculado por Gao et al. é associativo, pois, na região do TS ($\xi_{TS} = -0.14$ Å), $d(SP) = 2.28 \pm 0.07$ Å e $d(PO) = 2.15 \pm 0.08$ Å;
3. Kolmodin e Åqvist e Gao et al. obtêm $\Delta G^\ddagger < 16$ kcal/mol para a reação de fenilfosfato *monoaniônico*.

Todos estes resultados são diferentes das tradicionais interpretações de observações experimentais feitas para a catálise em LM-PTPs (§1.3) e dos resultados obtidos aqui para VHR (§4.2.1).

Nas PTPs selvagens (excetuando-se as CDC25s), o KIE medido (tabela 3) indica que o grupo de saída é protonado e que o TS é dissociativo (vide supra), ou seja, com $d(SP)$ e $d(PO)$ significativamente maiores que as distâncias de equilíbrio [tipicamente, $d(SP) = 2.20$ Å e $d(PO) = 1.75$ Å]. Se a energia de ativação para a reação catalisada do fenilfosfato monoaniônico é menor que 16 kcal/mol, a reação do substrato neste estado de protonação seria observada experimentalmente (§4.3.5). Novamente, uma quantidade apreciável de resultados experimentais indica que o substrato das PTPs liga-se à enzima e reage como um diânion (§1.3 e §4.3.5).

Kolmodin e Åqvist^[112] criticaram os KIEs [$^{18}(V/K)_{nonbridge}$] obtidos para reações enzimáticas, porque um estado de protonação tem que ser postulado (para corrigir os efeitos de fracionamento isotópico)^[20, 240] para que o $^{18}(V/K)_{nonbridge}$ seja determinado e, portanto, este efeito não pode ser usado para provar qual o estado de protonação. No entanto, se o monoânion for postulado como substrato, as medidas experimentais corresponderiam a valores de $^{18}(V/K)_{nonbridge}$ na faixa de 1.015^[240], que são distantes dos $^{18}k_{nonbridge}$ para hidrólise do monoânion (tabela 1). Se o diânion é postulado como substrato, os valores de $^{18}(V/K)_{nonbridge}$ medidos são bastante próximos ao KIE observado na hidrólise do diânion em solução (§4.3.3). Kolmodin e Åqvist também estimaram que a diferença entre a energia livre de ligação (ΔG_B) do substrato monoaniônico à LM-PTP seria 16 kcal/mol mais baixa que o ΔG_B do diânion e, portanto, argumentam que o diânion não se ligaria a enzima (ou se ligaria caso a CYS nucleofílica estivesse protonada, com uma subsequente transferência de H^+ para o grupo fosfato)^[115]. Este $\Delta\Delta G_B$ foi avaliado pelo método de perturbação de energia livre (§1.6.3) a partir do MC equilibrado para o monoânion como substrato. Neste complexo, o hidrogênio ligado ao oxigênio equatorial do grupo fosfato e o grupo tiolato da CYS nucleofílica formam uma ligação de hidrogênio (assim como nas simulações apresentadas na §4.2.1.4). A energia de uma ligação de hidrogênio é em geral cerca de 10 a 15 kcal/mol. Assim, a estimativa de Kolmodin e Åqvist para $\Delta\Delta G_B$ do substrato não implica que o diânion não deve se ligar, mas apenas que a interação do monoânion no sítio ativo apresenta um contato (uma ligação de hidrogênio) favorável. Reunidas, as observações de Åqvist et al. (estimativa de $\Delta\Delta G_B$, crítica a observações experimentais, como KIE ou perfis de k_{cat} versus pH, §1.3.3, e, principalmente, o item 3 listado acima) levaram estes autores a sugerir que a reação catalisada pelas PTPs é a desfosforilação do substrato *monoaniônico*^[115]. Os resultados apresentados nesta tese mostram que a hipótese de Åqvist et al. está equivocada (§4.3.3 e §4.3.5).

Claramente, as simulações realizadas por Åqvist et al. e Gao et al. apresentam incertezas consideráveis. Quais são elas? Estes estudos^[110, 112, 115] apresentam perfis de energia livre avaliados por técnicas semelhantes àsquelas usadas aqui (§1.6.3). Portanto, os erros estatísticos e sistemáticos inerentes devem ser similares (§4.3.1). A amostragem do espaço de configurações e o tipo de campo de força usado nestes estudos e nesta tese também foram similares. A sensibilidade aos parâmetros e as diferenças de ΔG calculadas por campos de força do tipo apresentado na §1.6.1 também são pequenas (§4.3.1). Assim, a maior fonte de incerteza nos trabalhos de Åqvist et al. e Gao et al. é o potencial híbrido, principalmente a hamiltoniana QM.

A mistura de parâmetros semiempíricos usada por Gao et al.^[110] para descrição da

região QM apresenta RMSD de cerca de 15 kcal/mol no cálculo de energias de espécies ao longo do caminho de reações de ésteres de fosfato (tabela 24 e §3.2.2). Portanto, esta descrição da região QM não é adequada para simulação da catálise por PTPs e deve ser responsável pelas importantes diferenças enumeradas acima entre as observações experimentais e os cálculos de Gao et al.^[110].

O método EVB, usado por Åqvist et al.^[19] nas simulações das LM-PTP, depende fundamentalmente de um procedimento de calibração que usa *medidas experimentais de reações não catalisadas em água*, mas equivalentes à reação enzimática. Os parâmetros determinados a partir deste procedimento empírico são usados nas simulações enzimáticas. Não vamos detalhar o procedimento nem as medidas experimentais usadas por Åqvist et al.^[112, 113, 115, 120], mas, como apresentado na §1.2.1, *não existem medidas cinéticas ou termodinâmicas para reações de tiólise de monoésteres de fosfato* que possam ser diretamente usados no procedimento de calibração. Assim, Åqvist et al. usaram uma série de aproximações e dados de outras reações para estimar indiretamente ΔG_{reac} e $\Delta G^\ddagger$ para reação de tiólise de fenilfosfato mono e dianiônico em solução aquosa. Por exemplo, $\Delta G^\ddagger = 22$ kcal/mol foi usado para a tiólise de fenilfosfato monoaniônico em solução aquosa^[112]. Este valor subestima a energia de ativação para esta reação em, no mínimo, 10 kcal/mol (§2.3.2) e, assim, as barreiras simuladas para a reação enzimática também serão subestimadas.

Outros autores^[9] também demonstraram surpresa ao constatar que metodologias bastante diferentes para descrição da região QM, como aquelas usadas por Åqvist et al. e Gao et al., foram capazes de calcular alturas de barreiras próximas entre si e coincidentes às medidas experimentais (vide supra).

Embora a descrição da região QM no estudo de Asthagiri et al.^[114] tenha sido feita por DFT (que, a princípio, deve apresentar incertezas menores que os métodos aproximados usados aqui e nos estudos de Åqvist et al. e Gao et al.), apenas otimizações de geometria e determinação do caminho de reação de fenilfosfato na presença da matriz enzimática, com o solvente representado por um dielétrico contínuo, foram realizados. Energias livres foram estimadas a partir das propriedades termodinâmicas calculadas em torno dos pontos estacionários (apêndice B) e apenas a posição de duas dezenas de átomos (reativos, localizados na região QM) foram otimizadas, o restante da enzima (praticamente toda cadeia protéica da LM-PTP) foi fixado. Para sistemas complexos, como macromoléculas solvatadas, este método introduz incertezas consideráveis (§4.3.1 e §4.3.2). Não obstante, o TS obtido é altamente dissociativo [$d(\text{PO}) = 2.70$ Å e $d(\text{SP}) = 3.02$ Å], com o grupo de

saída protonado pelo ácido geral. O $\Delta G^\ddagger = 8$ kcal/mol foi calculado para a desfosforilação de fenilfosfato dianiônico (vide supra), 6 kcal/mol mais baixo do que a energia de ativação experimental. Portanto, embora a geometria do TS seja comparável às observações experimentais, as energias calculadas apresentam erros significativos e os resultados obtidos por Asthagiri et al.^[114] devem ser vistos com cautela.

A subfamília das LM-PTPs tem estrutura terciária bastante diversa das outras PTPs e apenas parte da estrutura do P-loop é conservada (a LM-PTP não apresenta HIS no P-loop, §1.3). Assim, consideramos que a generalização do mecanismo observado nas LM-PTPs é inadequada. Também é possível, mas improvável, que as LM-PTPs utilizem um mecanismo diferente das PTPs clássicas ou DSPs. A estrutura terciária das DSPs e das PTPs clássicas são semelhantes. A principal diferença estrutural entre as duas subfamílias é a profundidade do sítio ativo, que foi sugerida como causa da seletividade por resíduos de tirosina nas PTPs clássicas (§1.3.1), porém, estas enzimas também desfosforilam alquilfosfatos com cadeia carbônica alongada^[89, 90]. Portanto, os mecanismos de catálise nas duas famílias devem ser equivalentes e os resultados observados nas simulações com a VHR podem ser extrapolados para a atividade das PTPs clássicas.

4.3.4 Considerações das reações nos mutantes e de metilfosfato

O reagente e o produto da reação de metilfosfato catalisada pela VHR selvagem são formados em ξ maiores em módulo do que a coordenada de reação para as mesmas espécies na reação de fenilfosfato (tabela 30), pois a $d(\text{SP})$ no reagente e a $d(\text{PO})$ no produto da reação de MeOP^{2-} são maiores. Em razão do menor volume do grupo metil em relação ao grupo fenil, as moléculas de água penetram mais na região do sítio ativo e solvatam em maior grau os grupos ou centros reativos, como o ASP92 e o oxigênio grupo de saída. A transferência de H^+ do ácido geral para o grupo de saída acontece após o TS da quebra da ligação P–O, possivelmente em virtude desta maior solvatação. Assim, a formação de metóxido não é estabilizada da mesma forma que a formação de fenol na simulação 1a (certamente, a transferência de H^+ para o oxigênio do grupo de saída é mais eficaz na estabilização do TS do que a solvatação aquosa, como pode ser concluído da maior reatividade de monoésteres de fosfato monoprotônicos em relação a monoésteres desprotonados em solução aquosa, §1.2.1) e a barreira calculada para a reação de MeOP^{2-} catalisada pela VHR é cerca de 4 kcal/mol mais alta do que a barreira estimada de medidas experimentais (§4.2.1.2).

Consideramos que a mudança do grupo de saída alterou e desequilibrou a solvatação na

entrada do sítio ativo. Portanto, o procedimento de construção do modelo estrutural para reação de MeOP^{2-} a partir do MC para PhOP^{2-} (p. 175) deve ser alterado para incluir a construção adicional de moléculas de água (que preencham o volume antes ocupado pelo grupo fenol) e permitir a equilibração do solvente. Podemos especular que o desvio da barreira calculada em relação à experimental seria reduzido, se uma média fosse realizada entre diferentes simulações utilizando o mesmo modelo estrutural da simulação 2, mas partindo de diferentes condições iniciais (§4.3.1). Outra possível causa da diferença entre o valor calculado e o medido é a ausência de dados experimentais para reação de MeOP^{2-} , já que a energia de ativação experimental foi estimada indiretamente pela reação de fosfoserina.

Na região do TS da reação de PhOP^{2-} catalisada pela VHR mutante $\text{ASP92} \rightarrow \text{ALA}$, a $d(\text{SP})$ é cerca de 0.2 Å menor que a mesma distância na reação catalisada pela VHR selvagem (§4.2.1.3), indicando que o TS é mais associativo na ausência do ácido geral e que a formação da ligação P–S é considerável. O $^{15}(V/K)$ e $^{18}(V/K)_{\text{bridge}}$ medidos na catálise pela VHR mutante são maiores do que na enzima selvagem (tabela 3), indicando que o grupo de saída não é protonado. Para VHR mutante, a comparação de $^{18}(V/K)_{\text{bridge}}$ com $^{18}k_{\text{bridge}}$ observado em solução aquosa (tabela 1) sugere que a quebra da ligação P–O na enzima é maior ou mais avançada do que em solução. A $d(\text{PO})$ para o TS obtido na simulação da VHR mutante é a mesma da distância obtida para enzima selvagem, confirmando que ambas as reações apresentam significativa quebra da ligação P–O no TS. Embora as distâncias entre o fósforo e os oxigênios não ligantes calculadas para o TS da reação da VHR mutante e selvagem sejam equivalentes, o $^{18}(V/K)_{\text{nonbridge}}$ medido indica que o mecanismo é mais associativo (tabela 3). Esta medida pode ser justificada pela considerável formação da ligação P–S no TS da reação calculada na VHR mutante.

A interpretação do $^{18}(V/K)_{\text{nonbridge}}$ medido para catálise pelo mutante $\text{ASP92} \rightarrow \text{ALA}$ dada por Hengge et al.^[273] é, no mínimo, confusa. Estes autores sugeriram que o KIE poderia estar aumentado em razão da maior participação do nucleófilo ou da interação dos oxigênios não ligantes com o P-loop (em especial o grupo guanidina da ARG130, p. 215). No artigo publicado em 1996, os autores concluem que as interações eletrostáticas são a causa da alteração do KIE, mas em 2002, o mesmo autor diz que “o TS apresenta um maior grau de participação do nucleófilo” para justificar o $^{18}(V/K)_{\text{nonbridge}}$ medido^[20]. A simulação apresentada para VHR mutante soluciona esta dúvida, pois mostra maior participação ou formação da ligação do nucleófilo no TS (§4.2.1.3).

A simulação realizada para o mutante $\text{ARG130} \rightarrow \text{ALA}$ utilizou um modelo da estru-

tura protéica que claramente não é adequado ou comparável à estrutura experimental. Portanto, as observáveis obtidas na simulação não são comparáveis às medidas experimentais. Contudo, a simulação indica que a mutação da ARG130 altera significativamente o posicionamento do substrato e do grupo tiol da CYS124 no MC, e, principalmente, altera o pK_a dos grupos reativos, elevando o pK_a dos oxigênios não ligantes do grupo fosfato ou reduzindo o pK_a do ácido geral (ASP92). Estas alterações podem ser atribuídas à ausência da estabilização eletrostática realizada pela carga positiva do grupo guanidina da ARG130. O TS calculado para este mutante é mais associativo, mas não podemos afirmar que esta alteração é consequência da ausência de estabilização eletrostática sobre o TS ou da reação ocorrer para o fenilfosfato monoprotonado (§4.2.1.4), como observado nas simulações 5 e 6.

Pela análise dos KIEs medidos para PTPs mutantes ARG→ALA, Hengge^[20] e Hoff et al.^[100] concluíram que esta mutação não altera o mecanismo de reação diretamente pela mudança da interação eletrostática, e sim desativa a catálise ácido-base geral. A ARG conservada no P-loop também é responsável pelo correto posicionamento do WPD-loop e, portanto, do resíduo ASP que funciona como ácido geral (§1.3). Na ausência da ARG, o resíduo ASP não estaria corretamente posicionado para catálise^[20, 100]. A simulação 4 sugere que a alteração dos KIEs para a VHR mutante (tabela 3) pode ser consequência da alteração dos pK_a s citados acima, que resulta na transferência do H^+ do ASP92 para os oxigênios não ligantes logo na região dos reagentes.

4.3.5 Qual é o estado de protonação do substrato?

Nas simulações 5, em que o substrato fenilfosfato é protonado, ou na simulação 6, em que o grupo tiol da CYS124 é inicialmente protonado, o H^+ ficou ligado covalentemente a um dos oxigênios não ligantes do fosfato no MC. Portanto, os pK_a s destes oxigênios são maiores do que o pK_a do grupo tiol. Em solução aquosa, os pK_a s são cerca de 6 e 8 unidades, respectivamente, mas no sítio ativo da VHR a ordem está trocada, pois o pK_a da CYS124 é abaixado (§1.2.1 e §1.3.3). A cadeia lateral do ASP92 também permaneceu protonada no MC.

Para as reações de $PhOP^{2-}$ e $MeOP^{2-}$, estas espécies permaneceram desprotonadas e o ASP92, protonado no MC. Assim, o pK_a do ácido geral no MC da VHR é maior que o pK_a dos oxigênios não ligantes do grupo fosfato. Esta observação é contrária aos cálculos realizados por Hart et al.^[274] que sugeriram que o estado de protonação no MC das PTPs seria equivalente à desprotonação da cadeia lateral do ASP (ASP^-) e à protonação de um

dos oxigênios não esterificados do grupo fosfato. Usando um potencial híbrido semelhante àquele empregado nesta tese (§1.6.4), mas com a região QM descrita pela hamiltoniana PM3 original, os autores calcularam que o MC com ASP^- e substrato monoprotonado é 21 kcal/mol mais estável que o estado equivalente à ASP protonada e o substrato desprotonado. Na §3.2.2, mostramos que a hamiltoniana PM3 calcula energias de reação de transferência de H^+ (ΔE_{pK_a} , tabela 24) com RMSD de mais de 16 kcal/mol e com um desvio máximo de até 29 kcal/mol, em relação a uma referência *ab initio*. Portanto, concluímos que os resultados de Hart et al.^[274] apresentam um desvio bastante elevado e que o estado de protonação proposto por estes autores está incorreto.

A simulação 5 com O2 protonado apresenta uma barreira cerca de 15 a 25 kcal/mol mais alta que a barreira das outras simulações com o substrato protonado. Esta diferença deve-se em parte à interação monodentada (§4.2.1.4) com o grupo guanidina da ARG130 que estabiliza o TS em menor grau do que observado nas outras simulações com o substrato protonado. Por outro lado, a ligação de hidrogênio entre o tiolato da CYS124 e o hidrogênio ligado ao grupo fosfato parece ser bastante estável, pois é preservada até a dissociação do grupo de saída (nas outras simulações com o substrato protonado, esta ligação de hidrogênio é rompida em $\xi < -0.7 \text{ \AA} < \xi_{TS}$). As energias livres de reação não foram calculadas para as reações de PhOP^- (§4.2.1.4), mas a diferença de energia livre entre a simulação 5 (O2) e os outros perfis (figura 39) na região dos TSs ($\xi \sim 0 \text{ \AA}$) e dos produtos ($\xi \sim 1 \text{ \AA}$) é semelhante. Portanto, a ligação de hidrogênio estável entre o tiolato e o hidrogênio de O2 deve ser responsável por uma maior estabilização do MC na simulação 5 (O2) em relação aos complexos formados nas outras simulações, conseqüentemente, elevando a altura de barreira e a energia livre de reação (§1.1.2.1).

As barreiras para as simulações 6 e 5 (O3 e O4 protonados) são mais próximas e variam entre si menos que 8.5 kcal/mol. A média da altura de barreira entre estas três simulações é 36.0 ± 3.7 kcal/mol, onde a incerteza é o desvio padrão das três medidas. Assim, a barreira calculada para desfosforilação de fenilfosfato monoprotonado catalisada pela VHR é cerca de 20 kcal/mol mais alta que a barreira calculada para a reação do substrato desprotonado (PhOP^{2-}). Este resultado indica que a VHR não catalisa a desfosforilação de fenilfosfato monoprotonado, ao contrário do que foi proposto por Åqvist et al.^[115] a partir da simulação da atividade de uma LM-PTP (§4.3.3.1).

Assim, concluímos que o substrato enzimático das DSPs (VHR) e CDC25s é um *diânion* e que o *nucleófilo enzimático está ionizado*. Os seguintes argumentos também apóiam esta conclusão:

- A altura de barreira calculada para reação de fenilfosfato dianiônico e a energia de ativação experimental para a catálise pela VHR (§4.2.1.2) e pela CDC25B (§4.2.2) são bastante próximas;
- As alturas de barreira calculadas para reação de MeOP^{2-} catalisada pela VHR selvagem e PhOP^{2-} catalisada pela VHR mutante $\text{ASP92} \rightarrow \text{ALA}$ também são bastante próximas às energias de ativação experimentais (§4.2.1.3);
- O mecanismo de reação e a estrutura do TS calculados para VHR e CDC25B concordam com as interpretações de KIEs medidos (§4.3.3 e §4.3.4);
- A interpretação de diferentes observáveis experimentais (apresentadas na §1.3.3).

Os KIEs medidos para a reação de desfosforilação catalisada por PTPs também indicam que o substrato reage como um diânion. A hipótese levantada por Åqvist et al.^[115], propondo que a interpretação das medidas de KIE também está de acordo com um substrato monoprotonado (p. 220), é incorreta^[240]. Excluindo-se as simulações apresentadas por Åqvist et al. e Gao et al. (§4.3.3.1), não existem evidências que suportem a possível reação de um monoânion.

4.3.6 Por que o mutante $\text{CYS} \rightarrow \text{SER}$ é inativo?

As PTPs mutantes $\text{CYS} \rightarrow \text{SER}$ não desfosforilam ésteres de fosfato, porém, em solução aquosa, a alcoólise destes compostos é até 10^7 vezes mais rápida que a tiólise (p. 39, §1.3.3). As seguintes causas podem ser atribuídas para a inatividade dos mutantes (§2.3.3):

- Alterações estruturais do P-loop ou do WPD-loop no mutante inviabilizariam a catálise ou a estabilização do TS^[111, 275, 276];
- O_γ da SER no mutante estaria muito distante do grupo fosfato, já que as distâncias das ligações C—O e O—P são, cada uma, cerca de 0.4 Å menores do que as distâncias C—S e S—P (§2.2)^[276];
- Ao contrário da enzima selvagem, a cadeia lateral do nucleófilo no mutante estaria protonada, impedindo o ataque ao fosfato^[111, 115];
- A nucleofilicidade do enxofre na CYS da enzima selvagem estaria alterada^[273], diminuindo sua reticência frente a centros de fósforo, mas este aumento na nucleofilicidade não seria observado na mesma escala pelo mutante.

Uma combinação destas hipóteses podem ser responsáveis pela inatividade. No entanto, parece-nos claro que a resposta definitiva para pergunta desta seção só pode ser dada pela comparação dos perfis de energia livre para as reações no sítio ativo da VHR selvagem e da mutante, e da estrutura do sítio ativo ligado ao substrato no TS, porque é nesta região do caminho de reação que as interações e, portanto, a estabilização, devem ser diferenciadas (§1.1.2.1). Os mutantes retêm a capacidade de ligar substratos (§1.3.3) com energias de ligação (ΔG_B) muito próximas àsquelas observadas para as enzimas selvagens^[277]. Assim, comparações entre a enzima selvagem e a mutante nas formas livres (não ligadas ao substrato)^[275] ou de MCs por medidas espectrofotométricas^[276] ou por simulações de dinâmica molecular^[111] não podem fornecer uma resposta definitiva para esta questão.

Os perfis de energia livre calculados para o mutante em diferentes estados de protonação (figuras 40 e 41) indicam que a catálise da desfosforilação de fenilfosfato ocorreria se o O_γ da SER124 e o substrato estivessem ionizados. Atividade catalítica não é detectada para este mutante^[42], portanto, concluímos que este estado de protonação não é observado experimentalmente. As simulações em que o O_γ ou o substrato (PhOP^-) são protonados têm perfis de energia livre semelhantes, porque o H^+ adicionado ao sistema fica compartilhado entre o O_γ e um dos oxigênios não ligantes do grupo fosfato no MC (§4.2.1.5). As barreiras observadas para estas simulações (8 e 9) variam entre 40 e 60 kcal/mol, indicando que nenhuma catálise enzimática seria observada (as energias de ativação para a hidrólise não catalisada de monoésteres de fosfato em solução aquosa são desta magnitude, §1.2.1). Finalmente, a simulação 10 também apresenta barreiras elevadas, maiores que 50 kcal/mol.

Portanto, pelo menos um H^+ deve estar em excesso para que a VHR mutante seja inativa. Contudo, concluímos na §4.3.5 que o substrato da VHR selvagem é fenilfosfato dianiônico. A mutação CYS124→SER não deve alterar o estado de protonação do substrato ligado à enzima, pois esta mutação é conservativa e não deve alterar a estrutura do MC ou o $\text{p}K_a$ dos oxigênios não ligantes do substrato. As estruturas determinadas por cristalografia de raios-X para MCs de PTPs selvagens e mutantes não apresentam alterações perceptíveis (que sejam maiores que a resolução destes experimentos, tipicamente da ordem de 10^{-1} Å)^[102, 276, 278]. Medidas espectroscópicas e o resultado das simulações 7 e 9 (tabela 32 e §4.2.1.5) também sugerem que somente pequenas alterações são observadas no MC da enzima mutante em relação ao MC selvagem^[276]. A cadeia lateral da SER livre tem um $\text{p}K_a \sim 6$ unidades mais alto que a cadeia lateral da CYS livre, logo, mesmo que o $\text{p}K_a$ esteja abaixado na enzima (este abaixamento não é maior que 4 unidades para a CYS124, §1.3.3), é bastante improvável que O_γ da SER esteja ionizado no sítio ativo

da enzima mutante. Portanto, o substrato da VHR, selvagem ou mutante, é sempre um diânion e o mutante *CYS124*→*SER* é inativo, porque o O_γ da *SER124* está protonado.

O perfil energético da alcoólise de monoésteres de fosfato na fase gasosa (§2.2.2) sugere que estas reações dependem significativamente da transferência de H^+ . Na enzima mutante, a alcoólise do grupo fosfato somente ocorreria com a transferência do hidrogênio ligado ao grupo hidroxila do nucleófilo para uma base geral (algum resíduo da proteína, o substrato ou o solvente). Todavia, é improvável que exista algum grupo pré-posicionado nas enzimas que catalise esta transferência de H^+ . A barreira para reação do mutante protonado (simulação 9) é cerca de 10 kcal/mol mais alta que a barreira para a VHR selvagem com a *CYS124* protonada (simulação 6, figura 39). Esta diferença pode ser atribuída ao maior pK_a da cadeia lateral da SER.

Assim, a natureza escolheu CYS como o nucleófilo nas PTPs, porque seu grupo lateral tem um pK_a mais baixo, que possibilita a ionização mais facilmente. Todavia, por que a evolução não modificou a enzima mutante para estabilizar a ionização da cadeia lateral da SER, como, por exemplo, algumas peptidases ou a fosfatase alcalina^[279]? Acreditamos que parte da vantagem energética em consequência da catálise covalente (§1.1.2.4) seria perdida, já que a hidrólise da ligação P–S (presente no intermediário fosforilado da PTP selvagem) é mais rápida que a hidrólise da ligação P–O (§4.3.3). Este possível intermediário fosforilado da PTP mutante não respeitaria as condições sugeridas para a catálise covalente na p. 11, pois, o intermediário formado seria menos ou tanto reativo quanto o substrato (já que alquilfosfatos são menos reativos que os arilfosfatos, §2.2). A reação catalisada pela VHR mutante é mais exoérgica que a enzima selvagem (compare a energia livre na região dos produtos nos perfis das simulações 1a e 7, figuras 36 e 40) e, portanto, o intermediário de alcoólise pode acumular-se (caso este seja mais estável que o reagente (MC) e que o produto da segunda etapa, equação 1.9).

Hengge, Denu e Dixon^[273] sugeriram que, caso a enzima aumentasse a nucleofilicidade do grupo tiol na *CYS124*, o mecanismo na reação catalisada seria mais associativo. Porém, observações experimentais, como, por exemplo, KIE, indicam que o mecanismo de reação catalisada é dissociativo e semelhante àquele observado em solução. Estes autores concluem que não ocorre alteração de nucleofilicidade do grupo tiol na enzima^[273]. Assim, o mutante não é inativado, porque sua nucleofilicidade não está amplificada. A conclusão parece-nos correta, mas discordamos da justificativa dada. O aumento de nucleofilicidade não leva necessariamente a uma maior participação do nucleófilo na etapa de dissociação do grupo de saída. No caso de um mecanismo D_n+A_n , a adição nucleofílica ao meta-

fosfato poderia ser facilitada pela enzima (por exemplo, nas reações $D_n + A_n$ F e H dos monoésteres em fase gasosa, a etapa determinante é justamente a adição do nucleófilo (A_n) ao metafosfato, §2.2.2). No entanto, acreditamos que a nucleofilicidade no sítio ativo não é relevante nem está significativamente alterada, para VHR selvagem e para o mutante, porque as simulações 1a e 7 mostram que a catálise ocorre independentemente do nucleófilo, se este estiver ionizado. A barreira para alcoólise é ligeiramente menor do que a barreira de tiólise, mas a transferência de H^+ e protonação do grupo de saída não ocorre no TS do mutante (§4.2.1.5), portanto, a altura de barreira simulada para alcoólise na simulação 7 pode estar superestimada em até 5 kcal/mol.

Como discutido acima, as mudanças estruturais na enzima mutante CYS124→SER são pequenas^[276] e podem ser observadas durante a dinâmica molecular no período de equilíbrio (ao contrário das mutações não conservativas, como ARG130→ALA, §4.3.1 e §4.3.4). Zhang e Wu^[276] sugeriram que pequenas alterações estruturais são responsáveis pela inativação do mutante CYS→SER, baseados em medidas de fluorescência (principalmente de um resíduo de triptofano, TRP354 na YOP51 de *Yersinia*, uma PTP clássica, §1.3.1). No entanto, como os próprios autores reconhecem, a fluorescência pode ser seletivamente extinguida pela CYS na enzima selvagem, mas não ser afetada na enzima mutante. Outras medidas indicam que a estabilidade da enzima mutante é menor que a selvagem^[275, 276]. As distâncias das ligações de hidrogênio do nucleófilo enzimático com a SER131 (tabelas 29 e 32) estão aumentadas se o mutante está protonado no MC, sugerindo uma interação mais fraca que pode resultar na menor estabilidade. A comparação das distâncias envolvendo os oxigênios não ligantes do substrato e o P-loop observadas na simulação 1a, 7 e 9 (tabelas 29 e 32) indica que 7 ou 8 ligações de hidrogênio são realizadas em todos TSs, portanto, a mutação não deve alterar a estabilização do TS. Zhang e Wu também discutem que, em razão do mecanismo de reação dissociativo, a desfosforilação do substrato deveria ser independente da participação do nucleófilo e alguma atividade seria observada experimentalmente para o mutante (mesmo que o intermediário fosforilado não fosse observado). O mecanismo calculado na simulação 1a é $A_n D_n$, com um TS de caráter dissociativo, porém, metafosfato não é formado. A ligação P–S é formada minimamente (§4.2.1.2) e, logo, a reação não é independente da participação do nucleófilo ou de um intermediário covalente.

Finalmente, embora as ligações C–O e O–P no mutante sejam menores que a ligação C–S e S–P na enzima selvagem, a desfosforilação pode ocorrer normalmente se a SER estiver ionizada (simulação 7), pois o grupo fosfato tem suficiente mobilidade para se aproximar do nucleófilo enzimático. A distância da ligação P–O no $MeOP^-$ é 0.45 Å

menor que a distância da ligação P–S no MeSP^- (§2.2.2) e a diferença entre as distâncias de ligação do nucleófilo com o grupo fosfato no TS da simulação 1a [$d(\text{SP})$] e da simulação 7 [$d(\text{OP})$] é de 0.37 ± 0.17 Å (tabelas 29 e 32), sugerindo que a ordem da ligação entre o grupo fosfato e o nucleófilo enzimático é semelhante nos dois TSs.

4.3.7 Qual é o ácido geral na CDC25B?

Diferentes observações experimentais sugerem que a catálise ácido-base geral não ocorre nas CDC25s (§1.3). Além dos KIEs (comentados na §4.3.3) que indicam que o grupo de saída e os oxigênios não ligantes do grupo fosfato estão desprotonados no TS, os perfis de k_{cat} e k_{cat}/K_M versus pH para desfosforilação de pNPP pelas CDC25s não apresentam a forma típica de sino, observada nas outras PTPs^[93]. A atividade enzimática é independente do pH acima de 6.5, ou seja, a inclinação descendente no lado básico, normalmente atribuída a ionização do ácido geral, não é observada. A velocidade de desfosforilação catalisada pelas CDC25s depende do pK_a do grupo de saída, assim como as PTPs mutantes ASP→ALA (p. 41), indicando que não ocorre protonação do grupo de saída na etapa determinante da velocidade de reação.

A barreira calculada para a reação de PhOP^{2-} catalisada pela CDC25B é cerca de 3 kcal/mol mais baixa que a energia de ativação experimental estimada para a mesma reação na §4.2.2. As outras barreiras calculadas, em que o substrato reage como um monoânion ou que um H^+ é doado de um dos resíduos GLU no P-loop são no mínimo 10 kcal/mol mais altas que a energia de ativação experimental (vide infra e §4.2.2). Portanto, concluímos que o substrato das CDC25s é um diânion e que *não ocorre catálise ácido-base geral* na primeira etapa de reação (equação 1.9).

A proximidade entre as energias de reação e de ativação calculadas para a catálise pela CDC25B e pela VHR indica que a solvatação do grupo de saída na primeira enzima é energeticamente comparável à protonação do mesmo grupo na segunda enzima. Como já foi apontado por outros autores^[280], a eficiente solvatação do grupo de saída por moléculas de água ao invés da protonação pode ser suficiente para catálise nas CDC25s.

Podemos especular que o desvio observado na interação QM/MM entre, por exemplo, água (MM) e fenol (QM, §3.2.3), sugerido como possível fonte de incertezas (§4.3.1), deve estar amplificado nas simulações 14 que utilizaram as configurações iniciais retiradas da simulação 11 ou da simulação 14d (§4.2.2.2). A exposição dos grupos reativos ao solvente nestas simulações é ainda maior que as simulações que partiram da configuração construída na §4.1.1, porque o substrato é parcialmente deslocado para fora do sítio ativo.

Assim, as barreiras calculadas na §4.2.2.2 estariam subestimadas.

Para as reações de ésteres de fosfato monoprotonados em solução aquosa, a transferência de H^+ via um anel de 6 membros, com participação de uma molécula de água (figura 5) parece contribuir para um abaixamento de até 10 kcal/mol na barreira de desfosforilação (§1.2.1). Este mecanismo não foi testado aqui para reação catalisada pela CDC25B, porque a participação de uma molécula de água na região QM e sua reação (doação de H^+) não foi calibrada ou testada. Contudo, mesmo que um anel de 6 membros também facilitasse a transferência de H^+ no sítio ativo da CDC25B, a desfosforilação a partir da espécie zwitteriônica (formada após a transferência) apresenta uma barreira mais alta que 30 kcal/mol (§4.2.2.2). Em solução aquosa, a existência de uma espécie zwitteriônica para reação de monoésteres de fosfato monoprotonados foi proposta a partir de observações experimentais e cálculos teóricos (§1.2.1). Nas simulações 13, a transferência de H^+ é simultânea à quebra da ligação P–O e a estabilização por um anel de 6 membros não deve reduzir a barreira calculada significativamente (tabela 33). Para a VHR, a protonação do grupo de saída pelo ASP92 é simultânea à quebra da ligação P–O e diminui a altura de barreira em 5 kcal/mol. A participação de uma molécula de água na transferência de H^+ na VHR é menos provável, porque, nesta enzima, o substrato está menos exposto ao solvente.

As diferenças entre as alturas de barreira e as energias livres na região dos produtos para as simulações 13 (§4.2.2) podem ser interpretadas da mesma maneira que nas simulações 5 (§4.3.5). Ou seja, a ligação de hidrogênio entre o tiolato da CYS473 e o hidrogênio do grupo fosfato deve ser responsável por uma maior estabilização do MC na simulação 13 (O3) em relação aos complexos formados se O2 ou O4 são protonados, conseqüentemente, elevando a altura de barreira e a energia livre de reação (§1.1.2.1).

Utilizando os mesmos dados que mostraram a independência da atividade catalítica da CDC25B em $pH > 6.5$ ^[93], Rudolph^[22] reajustou os perfis de k_{cat} e k_{cat}/K_M versus pH para desfosforilação de pNPP (para dois pK_a s, ao invés de três^[93]) e sugeriu que o substrato reage como um monoânion. O H^+ ligado ao fosfato no MC seria transferido para o grupo de saída no TS com auxílio do GLU474^[22], ou seja, o substrato funcionaria como um ácido geral e o GLU474, como uma “ponte” para transferência de H^+ . As simulações 13 mostram que a reação do substrato monoaniônico não é catalisada, pois apresenta barreiras de no mínimo 35 kcal/mol. As simulações 14d e 14b com GLU474 protonado indicam que este resíduo não transfere H^+ de ou para um oxigênio não ligante e, portanto, não deve funcionar como uma ponte. A inspeção da estrutura do MC da CDC25B mostra que não

existe uma conformação da cadeia lateral do GLU474 que aproxime suficientemente (para uma transferência de H^+) o grupo carboxilato aos oxigênios não ligantes do substrato sem que este seja removido do sítio ativo. Assim, qualquer transferência de H^+ que envolva o GLU474 no MC acontecerá diretamente para o oxigênio esterificado (O1, §4.2.2.2). Rudolph^[22] também sugere que $PhOP^{2-}$ seria um *inibidor competitivo* da CDC25B! Mais uma vez, as simulações apresentadas na §4.2.2 desqualificam esta hipótese.

Embora a desfosforilação de pNPP seja independente do pH em meio básico, outros substratos cujo grupo de saída tem pK_a (pK_a^{lg}) mais elevado, como fenilfosfato ($pK_a^{lg} = 9.99$) e 4-acetilfenil fosfato ($pK_a^{lg} = 8.05$), apresentaram um perfil de k_{cat}/K_M versus pH descendente em $pH > 6.5$ para reação catalisada pela CDC25A^[105]. O k_{cat}/K_M medido para diferentes substratos em $pH = 7$ é dependente de pK_a^{lg} , se este é menor que 8.0 ($\beta_{lg} = -0.8$) e quase independente do pK_a^{lg} , se este é maior que 8.0 ($\beta_{lg} = 0.08$). Para a CDC25A mutante GLU431→GLN (os resíduos GLU431 e GLU435 na CDC25A são equivalentes aos resíduos GLU474 e GLU478 na CDC25B), o β_{lg} e o perfil de k_{cat}/K_M versus pH para os substratos com $pK_a^{lg} \leq 8$ foram mantidos, mas a inclinação descendente observada no perfil de pH na desfosforilação de 4-acetilfenil fosfato foi suprimida e a desfosforilação de fenilfosfato e outros substratos com $pK_a^{lg} > 8$ não pode ser medida (o sinal era baixo e irreproduzível^[105]). Os valores de k_{cat}/K_M e os perfis de pH não foram alterados pela mutação GLU435→GLN na CDC25A. Assim, Hengge, Zhang et al.^[105] sugeriram três diferentes modelos para o mecanismo de reação catalisada pela CDC25A: em dois deles, o GLU431 participaria como um ácido geral (numa hipótese com e na outra sem uma mudança conformacional), doando um H^+ para os substratos somente com $pK_a^{lg} > 8.0$; no terceiro modelo, o GLU431 teria apenas uma função estrutural, devendo estar protonado para que uma mudança conformacional, limitante da velocidade de reação para os substratos com $pK_a^{lg} > 8.0$, ocorra na proteína. Os próprios autores reconhecem que estes três possíveis modelos de mecanismo são igualmente possíveis e não podem ser distinguidos pelos experimentos apresentados^[105].

No entanto, a existência de diferentes mecanismos para a reação catalisada, que sejam dependentes do pK_a^{lg} do substrato, parece-nos estranha e pouco provável. As simulações da desfosforilação de fenilfosfato com transferência de H^+ do GLU474 (equivalente a GLU431 na CDC25A) para o grupo de saída apresentam barreiras pelo menos 10 kcal/mol mais altas que a energia de ativação experimental e não resultam nos produtos esperados, fenol e intermediário tiofosforilado (§4.3.3).

A desfosforilação de fosfotreonina do substrato natural (uma proteína CDK, §1.3.1)

da CDC25B também apresenta um perfil em forma de sino, com k_{cat}/K_M versus pH descendente em $\text{pH} > 6.5$ ^[93]. Mutações nos resíduos GLU474 e GLU478 não suprimem a inclinação descendente em $\text{pH} > 6.5$, indicando que estes resíduos não funcionam como ácidos gerais^[93]. Todavia, o k_{cat}/K_M máximo para o substrato natural é diminuído de 100 vezes no mutante GLU474→GLN, em relação a CDC25B selvagem. A manutenção do perfil em forma de sino para reação do substrato natural na CDC25B mutante não pode ser explicada pelos três modelos propostos por Hengge, Zhang et al.^[105], que implicam que a catálise pelo mutante com substratos de $\text{p}K_a^{lg} > 8.0$ (como a fosfotreonina) seria independente do pH em meio básico.

Para a reação de PhOP^{2-} sem participação de um ácido geral (simulação 12, figura 42), o GLU474 foi modelado na forma desprotonada e nenhuma mudança estrutural significativa é observada neste grupo ou em outros resíduos do P-loop. Como sugerimos acima, a catálise sem ácido geral implica que a solvatação do grupo de saída deve ser bastante eficiente. Logo, especulamos que a dependência da catálise com o pH do meio e com o $\text{p}K_a^{lg}$ ^[93, 105] podem ser explicadas se assumirmos que o GLU474 na CDC25B e o GLU431 na CDC25A têm função estrutural, ligada ao correto posicionamento das moléculas de água para solvatação do grupo de saída. Resíduos estruturalmente importantes, mas não envolvidos diretamente com a catálise, podem afetar as inclinações dos perfis de pH (p. 42). A análise das trajetórias de dinâmica molecular na região do TS e dos produtos na simulação 12 mostram que GLU474 forma duas ligações de hidrogênio com moléculas de água que participam direta e indiretamente (estão na primeira e segunda camadas de solvatação) da solvatação do fenóxido e do grupo fosfato. Nos substratos com maior $\text{p}K_a^{lg}$, a solvatação do grupo de saída deve ser mais eficiente do que nos grupos de saída menos básicos para que a catálise ocorra. Assim, a desprotonação de GLU474 (com aumento do pH) pode alterar a conformação deste resíduo e dificultar a solvatação do grupo de saída, pois a coordenação da cadeia lateral com moléculas de água não seria ideal. A mutação de GLU474 na CDC25B (ou GLU431 na CDC25A) eliminaria a dependência com o pH nos modelos de substrato (como fenilfosfato ou 4-acetilfenil fosfato) e diminuiria a atividade da enzima (vide supra). Contudo, no substrato natural, resíduos (tituláveis como GLU, ASP ou HIS) da própria proteína fosforilada poderiam estar envolvidos com o correto posicionamento de moléculas de água para solvatação do grupo de saída, já que a atividade da CDC25B na desfosforilação do substrato natural é 10^3 vezes maior do que nos modelos de substrato^[93].

5 Conclusões

Apresentamos nesta seção as principais conclusões feitas nas discussões dos resultados obtidos nos capítulos anteriores.

O principal objetivo nesta tese foi responder algumas perguntas (p. 44) sobre o mecanismo de reação das proteínas tirosina-fosfatases. A metodologia adotada, simulação computacional das reações enzimáticas, apresenta uma série de aproximações que tiveram que ser testadas e calibradas para assegurar que os resultados das simulações tenham sentido físico (p. 131). Como dados experimentais que pudessem ser usados para teste e calibração da metodologia computacional são escassos, utilizamos cálculos de estrutura eletrônica *ab initio* (que por sua vez também foram testados) como valores de referência (§1.4). A partir destes valores, um potencial híbrido, capaz de simular reações enzimáticas, foi construído, calibrado e usado para obtenção de perfis de energia livre para reações catalisadas, que, finalmente, foram usados para alcançar o objetivo principal.

Como a alcoólise e tiólise de ésteres de fosfato não são facilmente observadas experimentalmente (§1.2), as reações (A a H, figuras 7 e 8) estudadas por métodos de cálculo *ab initio* no cap. 2 são uma referência para a análise da reatividade intrínseca de ésteres de fosfato e modelos dos possíveis mecanismos de catálise em PTPs.

A metodologia de cálculo *ab initio* utilizada no cap. 2 foi calibrada por comparações com as medidas experimentais disponíveis. Energia de ativação ou de reação para a tiólise e alcoólise de ésteres de fosfato em fase gasosa não são conhecidas. Assim, energias de reação de transferência de H^+ e diferentes propriedades como geometrias moleculares, momentos de dipolo, frequências vibracionais e potenciais de ionização eletrônica de produtos e reagentes das reações modelo foram usadas. Em geral, as propriedades calculadas foram bastante próximas às medidas experimentais, especialmente as energias de transferência de H^+ , indicando que o nível MP2/6-311+G(2df,2p) utilizado foi adequado.

O estudo da convergência da energia eletrônica com o aumento do conjunto base (§2.2.3) mostrou que as energias relativas obtidas com o conjunto 6-311+G(2df,2p) estavam aparentemente convergidas (diferiam menos que 1 kcal/mol das energias obtidas

com o maior conjunto, TZVPP). Melhorias significativas do conjunto base ou do tratamento N -eletrônico para resgate da correlação eletrônica resultariam em grande aumento do custo computacional e são inviáveis para sistemas do tamanho daqueles estudados no cap. 2. Portanto, o nível de cálculo *ab initio* utilizado é o melhor que podemos empregar para o estudo das reações modelo.

As reações de ataque nucleofílico sobre os ésteres de fosfato em fase gasosa podem ocorrer idealmente segundo dois mecanismos: um associativo, denominado A_n+D_n ; e outro dissociativo, D_n+A_n . Nas reações de tri e monoésteres associativas, a IRC é dominada pela formação (na etapa de ataque) e quebra (na etapa de dissociação) da ligação com o fósforo. Nas reações dissociativas, a IRC foi dominada pela transferência de H^+ . Triésteres reagem sempre segundo mecanismos A_n+D_n e monoésteres podem reagir segundo ambos mecanismos, dependendo do nucleófilo e do grupo de saída envolvidos (§2.2). Intermediários estáveis, com geometria de TBP, são formados nos mecanismos associativos, mas sua geometria não segue as regras empíricas propostas (p. 21) e a pseudorrotação, que reposiciona os ligantes na TBP, não é cineticamente importante para as reações de tiólise e alcoólise.

Os principais fatores que controlam a termodinâmica, a cinética e o mecanismo das reações modelo são, nesta ordem: o estado de alquilação dos ésteres, a basicidade do nucleófilo e do grupo de saída e a possibilidade de transferências de H^+ (de espécies ácidas, §2.3.2). As barreiras e energias de reação calculadas para as reações de triésteres de fosfato são bastante diferentes das reações de monoésteres. As reações destes últimos são sempre endotérmicas, mas consomem no máximo 10 kcal/mol. Por outro lado, as energias de reação de triésteres podem variar de -36 kcal/mol até 34 kcal/mol. Excetuando-se a reação de $TMP + CHS^-$ (reação A, figura 7), as barreiras para as reações de triésteres não ultrapassam 16 kcal/mol e são menores que as barreiras para reação de monoésteres (sempre maiores que 28 kcal/mol), independentemente do mecanismo. As espécies e as reações envolvendo metóxido têm sua geometria e os perfis de energia total alterados em razão da maior basicidade do metóxido. Fenol, menos básico, distorce a geometria das espécies em menor grau e é um melhor grupo de saída do que metóxido.

Pela análise das evidências experimentais para reações de ésteres de fosfato (§1.2) e dos resultados computados no cap. 2 e em outros estudos (§1.2.1), especulamos que a tiólise de ésteres de fosfato, até hoje observada apenas para reação intramolecular de um diéster (§1.2.1), pode ser observada em um meio não-aquoso, com solvente pouco nucleofílico e impedido estericamente (§2.3.2). Efeitos de solvatação específica, como a

participação de uma molécula de solvente na transferência intramolecular de H^+ (figura 5), parecem importantes para reatividade de ésteres de fosfato e, portanto, devem ser incluídos em estudos computacionais destas espécies em solução aquosa.

Também analisamos as reações estudadas no cap. 2 como modelos da atividade das PTPs selvagens e seus mutantes (§2.3.3). A contribuição da repulsão eletrostática entre nucleófilo e o substrato dianiônico para a altura de barreira de ataque nucleofílico foi estimada em cerca de 10 kcal/mol e sua diminuição é uma possível fonte de catálise nas enzimas. A formação da ligação covalente entre o nucleófilo enzimático (aniônico) e o grupo fosfato do substrato (neutro, ou seja, sem a repulsão eletrostática descrita acima) apresenta barreiras iguais ou menores que 16 kcal/mol (§2.2). Esta é a energia de ativação para a etapa de tiólise de fenilfosfato catalisada pela VHR (tabela 2) e, logo, a formação da ligação P–S não necessita de catálise. A formação de um intermediário estável, seja ele uma fosforana (num mecanismo A_n+D_n) ou um metafosfato (num mecanismo D_n+A_n) deve ser catalisada pelas PTPs, pois todos estes intermediários de reações de monoésteres têm energia de formação maiores que 24 kcal/mol. Talvez, por este motivo, o mecanismo utilizado pelas enzimas é A_nD_n e a tiólise ocorre em uma única etapa, sem a formação de intermediários estáveis (§4.3.3). Outra possível fonte de catálise identificada seria a protonação ou a estabilização da dissociação de um alcoóxido, que é bastante desfavorável em fase gasosa. Na VHR e PTPs clássicas (§1.3), observamos catálise ácido-base geral, que diminui a barreira para dissociação.

As reações modelo também permitiram a sugestão de algumas causas para a inatividade das PTPs mutantes CYS→SER, como a distância elevada entre O_γ (na SER mutante) e o substrato ou a maior basicidade da cadeia lateral da SER, que estaria protonada ou alteraria a estrutura do sítio ativo. Estas hipóteses foram testadas nas simulações enzimáticas (§4.3.6).

O capítulo 3, em que o potencial híbrido foi testado e calibrado, é uma contribuição importante e original desta tese, pois possibilitou o aumento da precisão e da exatidão nos resultados das simulações enzimáticas. Acreditamos que o potencial híbrido desenvolvido, principalmente a hamiltoniana QM semiempírica, representa um dos mais confiáveis potenciais para simulação de reações enzimáticas existentes. Calibrações ou testes, como executados aqui, não são comuns em estudos de enzimologia computacional^[3, 9, 184].

No entanto, o potencial desenvolvido é restrito a reações de tiólise e alcoólise de ésteres de fosfato e sua calibração foi trabalhosa. Outras soluções, como a correção das deficiências na hamiltoniana QM *a posteriori*^[260, 261], são menos trabalhosas, mas não as-

seguram (principalmente em sistemas com elevado número de graus de liberdade) que o espaço de configurações ou as energias livres sejam corretamente avaliadas, já que a energia obtida da simulação, ao invés do potencial, é corrigida. Outra possível crítica às parametrizações específicas é que “a precisão em outras áreas (da PES) pode ser mudada inadvertidamente”^[260]. Como observamos em §3.1.2 e §3.2.1, o potencial ou a hamiltoniana reparametrizados devem ser testados durante a calibração, porque artifícios (ou “mudanças inadvertidas”) podem ser introduzidos. Estes artifícios podem ser facilmente corrigidos (§3.2.1).

Portanto, construímos um modelo de calibração que é mais completo do que a otimização de uma função erro simples (equação 3.1)^[157], pois inclui diferentes propriedades de diversas famílias de reações e suas espécies.

Identificamos que a causa dos desvios elevados observados no cálculo semiempírico de propriedades de moléculas contendo elementos hipervalentes do segundo período não é a falta de funções com simetria “d” no conjunto base, e sim o uso de conjuntos de treinamento na parametrização contendo compostos com o elemento em diferentes estados de oxidação (§3.2.1). Observamos compensações no conjunto calibrado (MNDO+G/CHOPS) que, por exemplo, supriam a falta de funções de polarização (como as de simetria “d”), aumentando o caráter iônico de determinados parâmetros. Todavia, a complicada interdependência dos parâmetros semiempíricos não foi racionalizada e, assim, não podemos tirar conclusões mais fundamentadas sobre qual a influência de algumas aproximações e do procedimento de calibração sobre os parâmetros. Este tipo de análise seria útil para o desenho de modelos mais eficientes para calibração de hamiltonianas tipo NDDO. Embora seja uma possível fonte de erros nas PES semiempíricas (p. 59), a introdução de funções gaussianas (de baixa magnitude, tabela 23) em f_{core} (equação 1.49) foi necessária para calibração dos parâmetros.

A diminuição dos desvios em *todas* propriedades calculadas com o conjunto de parâmetros calibrado MNDO+G/CHOPS em relação aos parâmetros originais foi bastante significativa (§3.2.2). Em especial, o RMSD da energia obtida para as espécies com geometria otimizada pelo conjunto calibrado em relação a referência *ab initio* é menor que 3 kcal/mol e, para as reações de monoésteres de fosfato (como os substratos enzimáticos), o erro máximo observado é menor que 3 kcal/mol, indicando que a precisão da hamiltoniana semiempírica calibrada é no mínimo semiquantitativa.

Na hamiltoniana de interação QM/MM (§1.6.4), a descrição da parte eletrostática segundo a proposta de Field et al.^[179, 187] apresenta resultados de boa qualidade em com-

parações com energias de interação *ab initio* para complexos bimoleculares (§3.2.3.1). A proposta do grupo de Nancy apresenta uma repulsão excessiva e seu uso deve ser evitado. A calibração específica da interação QM/MM não eletrostática ($\mathcal{H}_{QM/MM}^{vdW}$), pela otimização dos parâmetros de LJ (equação 1.65) designados aos centros QM, novamente usando energias de interação *ab initio* para complexos bimoleculares como referência (§3.2.3.2), foi bem sucedida apenas para complexos de van der Waals, mas não funcionou para os complexos que realizavam ligações de hidrogênio, mesmo se as cargas parciais designadas aos centros MM que realizavam a ligação de hidrogênio também fossem otimizadas. Concluímos que a forma do potencial de LJ não é apropriada para descrição de $\mathcal{H}_{QM/MM}^{vdW}$. Como discutido na §4.3.1, acreditamos que esta seja a maior fonte de desvios nas simulações enzimáticas do capítulo 4. A fronteira covalente entre as regiões QM e MM também foi testada e não deve introduzir desvios significativos nos processos de interesse para a simulação das PTPs.

Simulações da primeira etapa de catálise em duas fosfatases de dupla especificidade, a VHR e a CDC25B, utilizando o potencial híbrido calibrado, foram apresentadas no cap. 4. Esta etapa corresponde à tiólise de fenilfosfato e resulta na desfosforilação do substrato, produzindo a enzima tiofosforilada e fenol na VHR ou fenóxido na CDC25B (equação 1.9).

Possíveis fontes de incertezas nas simulações enzimáticas foram consideradas. No entanto, é praticamente impossível estimar uma faixa de precisão segura para as energias livres calculadas e o principal teste dos resultados calculados é a comparação com observáveis experimentais (§4.3.1). Diferentes coordenadas de reação foram testadas (§4.2.1.1 e §4.3.2). A $\xi = d(\text{PO}) - d(\text{SP})$ foi escolhida, porque apresenta a menor altura de barreira calculada que é bastante próxima a energia de ativação experimental para primeira etapa de catálise na VHR.

O mecanismo da reação catalisada pela VHR é A_nD_n , ou seja, a adição do nucleófilo e a eliminação do grupo de saída ocorrem simultaneamente, em uma única etapa, sem formação de um intermediário estável. Assim, somente um TS separa os reagentes dos produtos. O TS obtido é dissociativo, com geometria de TBP e caráter de metafosfato, apresentando elevada quebra da ligação P–O com o grupo de saída e pouca (mas diferente de zero) formação da ligação P–S com o nucleófilo enzimático. O grupo de saída é protonado no TS pela ASP92 na VHR. Para CDC25B, o mecanismo é similar, A_nD_n , com um TS dissociativo, porém a participação do nucleófilo é um pouco maior que na VHR e não ocorre protonação do grupo de saída. Na CDC25B, a maior exposição do sítio

ativo ao solvente e a eficiente solvatação do fenóxido por moléculas de água é suficiente para estabilizar o grupo de saída ionizado. Os mecanismos calculados nas duas enzimas são suportados por diversas observações experimentais. As energias de ativação medidas são bastante próximas às alturas de barreira calculadas. A interpretação de todos KIEs medidos na VHR e na CDC25B também suportam a estrutura do TS calculado (§4.3.3).

As ligações de hidrogênio entre o substrato e as cadeias protéica principal e lateral da ARG conservada no P-loop são aparentemente a principal fonte de catálise em ambas PTPs estudadas. Ao lado da interação com a carga positiva do grupo guanidina, as ligações de hidrogênio indicam que o TS é estabilizado (§1.1.2.1) principalmente por forças eletrostáticas (§1.1.2.6). A catálise ácido-base geral (§1.1.2.5) também está operante na VHR selvagem e deve ser responsável pela estabilização do TS por até 5 kcal/mol. O terceiro mecanismo microscópico usado é a catálise covalente, que deve ser importante na segunda etapa de reação em que a ligação P–S é hidrolisada (§4.3.3). Como a reação de tiólise de monoéster de fosfato nunca foi observada em solução aquosa, não podemos afirmar que a enzima estabiliza o TS da reação em solução ou que a enzima utiliza um mecanismo alternativo. A resposta para esta dúvida pode ser dada se a reação de referência for simulada no solvente aquoso, na ausência do catalisador (§1.1.2).

Outras simulações computacionais da catálise por uma LM-PTP apresentam elevadas incertezas em virtude de descrições aproximadas e não calibradas (corretamente) para parte QM do potencial ou de imprecisões na avaliação da energia livre (§4.3.3.1). O mecanismo de reação catalisado pela LM-PTP não deve ser generalizado para as outras subfamílias de PTPs, porque a estrutura terciária desta enzima é bastante diferente. O mecanismo calculado na VHR é um melhor exemplo da reatividade nas DSPs e PTPs clássicas.

A altura de barreira e a estrutura do TS calculado na simulação da VHR mutante ASP92→ALA estão bastante próximos respectivamente da energia de ativação e da interpretação dos KIEs medidos (§4.3.4). A simulação da desfosforilação de metilfosfato pela VHR selvagem também apresentou uma altura de barreira compatível com a energia de ativação para desfosforilação de alquilfosfatos. A proximidade entre as medidas experimentais e os resultados das simulações para o mutante e para metilfosfato são evidências adicionais de que as simulações com o potencial híbrido são confiáveis e podem ser usadas para nossos objetivos.

No entanto, o aumento da barreira calculada na simulação da VHR mutante ARG130→ALA é bem maior que a elevação estimada experimentalmente, indicando que a estru-

tura tridimensional da proteína usada na simulação (derivada da VHR selvagem) não é apropriada. Assim, o procedimento usado para construção dos modelos estruturais para as simulações (§4.1.1) deve ser alterado para construção de estruturas de mutantes não conservativos ou de MCs que sofram considerável perturbação em relação a estrutura nativa. Sugerimos que a construção dos modelos estruturais dos mutantes, como a VHR ARG130→ALA, seja derivada de estruturas cristalográficas dos respectivos mutantes, ao invés da estrutura da enzima selvagem.

Os resultados das simulações para VHR e CDC25B nativas mostra que o substrato e o nucleófilo enzimático estão desprotonados (§4.3.5). As simulações de outros estados de protonação (substrato monoprotonado ou nucleófilo protonado) apresentam barreiras significativamente mais altas (15 kcal/mol, no mínimo) que as energias de ativação experimentais para as duas enzimas. A interpretação de diferentes medidas experimentais (KIE, LFER, perfis de k_{cat} versus pH, etc.) também indica que os substratos e o nucleófilo enzimático estão desprotonados (§1.3.3).

A inatividade das PTPs mutantes CYS→SER foi investigada, porque, em solução, a reação de tiólise intramolecular de um diéster de fosfato é 10^7 vezes mais lenta do que a alcoólise (§1.2.1)! Concluímos que o mutante não realiza catálise, porque está protonado (§4.3.6). A evolução não alterou as PTPs para estabilizar a ionização da cadeia lateral da SER nem para acomodar um nucleófilo protonado. A CYS deve ser usada, porque o pK_a da sua cadeia lateral é cerca de 6 unidades mais baixo do que a SER e a vantagem energética pela catálise covalente (cerca de 5 a 10 kcal/mol) seria perdida se o nucleófilo fosse SER. Outras hipóteses levantadas para explicar a inatividade do mutante, como alterações estruturais do sítio ativo, maior distância entre O_γ na SER mutante e o substrato, ou alterações na nucleofilicidade da SER não parecem importantes, pois a desfosforilação catalisada pelo mutante com a SER ionizada apresenta barreiras menores que 15 kcal/mol e, portanto, seria observada.

A CDC25B não utiliza um ácido geral para catalisar a desfosforilação de fenilfosfato e o grupo de saída está ionizado na região dos produtos (§4.3.7). Portanto, o mecanismo de catálise nas CDC25s é diferente das outras PTPs e estas enzimas devem ser classificadas numa (sub) família diferente das outras DSPs e PTPs. Simulações com um substrato monoprotonado apresentam alturas de barreira maiores ou iguais a 35 kcal/mol, indicando que nenhuma catálise seria observada com o substrato neste estado de protonação. As simulações com doação de H^+ dos dois resíduos GLU no P-loop da CDC25B apresentam barreiras no mínimo 10 kcal/mol mais altas que as energias de ativação experimentais e

não levam a formação dos produtos esperados (intermediário enzimático tiofosforilado).

Os resultados e as interpretações das simulações também foram usados para solucionar ou refutar algumas dúvidas ou especulações feitas anteriormente. Kolmodin e Åqvist^[115] sugerem que o substrato das PTPs é um monoânion monoprotonado. No entanto, nossos resultados mostram que os substratos da VHR (selvagem e mutante sem o ácido geral) e da CDC25B são diânions desprotonados. Hengge^[20] argumenta que a mutação da ARG conservada no sítio ativo impede que a doação de H^+ pela ASP92 ocorra. A simulação desta VHR mutante sugere que os pK_a s dos grupos reativos (ácido geral e substrato) estão alterados e que o H^+ da ASP é transferido para um dos oxigênios equatoriais do grupo fosfato do substrato. A interpretação dada por Hengge, Denu e Dixon^[273] para o $^{18}(V/K)_{nonbridge}$ medido na VHR mutante ASP92→ALA é confusa. A estrutura do TS obtida na simulação sugere que o $^{18}(V/K)_{nonbridge}$ está alterado no mutante em razão da maior formação da ligação P–S no TS. Cálculos realizados por Hart et al.^[274] sugeriram que o estado de protonação no MC das PTPs equivale à desprotonação da cadeia lateral da ASP (ASP^-) e à protonação de um dos oxigênios não esterificados do grupo fosfato. Todos MCs observados nas simulações da VHR selvagem, ligada a fenilfosfato ou metilfosfato, apresentam a ASP protonada e o substrato desprotonado. Rudolph^[22] propôs que o substrato na forma de um diânion seria um inibidor competitivo na CDC25B. As simulações desta enzima indicam que o substrato reage como um diânion e que, na verdade, o monoânion seria um inibidor competitivo, já que a barreira calculada para reação desta espécie é elevada.

APÊNDICE A – Integrais semiempíricas de repulsão eletrônica

As integrais semiempíricas de repulsão eletrônica em 2-centros ($\mu\nu|\lambda\sigma$) são calculadas por uma expansão multipolar que é forçada a ter o comportamento correto quando a distância interatômica $R_{AB} \rightarrow \infty$ (ou seja, convergir aos valores clássicos das interações multipolares) e quando $R_{AB} \rightarrow 0$ (ou seja, convergir aos valores semiempíricos para as integrais de 1-centro, $g_{\mu\nu}$ e $h_{\mu\nu}$)^[143, 193]. A expansão multipolar é truncada após o termo quadrupolar e, assim, cada elemento possui configurações de monopolo, dipolo, quadrupolo linear e quadrupolo quadrangular que representam respectivamente as distribuições de carga dos orbitais ss , sp_η , $p_\eta p_\eta$ e $p_\eta p_\tau$, onde η e τ são os eixos x, y e z.

O comportamento correto da expansão é forçado pela expressão de Klopman-Ohno^[142]:

$$f(R_{ij}) = [R_{ij}^2 + (\rho_a^A + \rho_b^B)^2]^{-1/2} \quad (\text{A.1})$$

onde i e j representam as cargas pontuais referentes a cada n -polo atômico usado na expansão multipolar nos centros A e B , e $\rho_a^A, a=0, 1$ ou 2 são termos aditivos obtidos para cada elemento A (vide infra). A distância R_{ij} é calculada a partir de R_{AB} e das distâncias de separação das cargas pontuais, D_a , nas configurações de cada n -polo^[193]. Por exemplo, a integral $(ss|sp_z)$ se reduz a interação entre um monopolo em A (q^A) e um dipolo em B ao longo do eixo z (μ_z^B) (figura 46), já que todos outros momentos multipolares se anulam:

$$\begin{aligned} (ss|sp_z) = [q^A|\mu_z^B] &= -(e^2/2)[(R_{AB} - D_1^B)^2 + (\rho_0^A + \rho_1^B)^2]^{-1/2} + \\ &\quad (e^2/2)[(R_{AB} + D_1^B)^2 + (\rho_0^A + \rho_1^B)^2]^{-1/2} \end{aligned} \quad (\text{A.2})$$

As separações de carga, D_a , são obtidas ao se igualar o momento multipolar de cada configuração a correspondente distribuição de carga^[193], assim:

$$D_0 = 0 \quad (\text{A.3})$$

Figura 46: Configuração de carga pontual usada na expansão da integral semiempírica de repulsão eletrônica $(ss|sp_z) = [q^A|\mu_z^B]$, onde η é o eixo x ou y.

$$D_1 = \frac{2n+1}{\sqrt{3}} \frac{(4\zeta_{ns}\zeta_{np})^{n+1/2}}{(\zeta_{ns} + \zeta_{np})^{2n+2}} \quad (\text{A.4})$$

$$D_2 = \sqrt{\frac{(2n+1)(2n+2)}{20}} \zeta_{np}^{-1} \quad (\text{A.5})$$

onde n é o número quântico principal do orbital atômico de valência ($n = 2$ para os átomos do primeiro período e $n = 3$ para átomos do segundo período). Os termos aditivos são obtidos solucionando-se as expressões das expansões multipolares no limite $R_{AB} \rightarrow 0$ para o caso homonuclear^[193]. O valor de $\rho_0 = e^2/(2g_{ss})$ é dado por uma expressão fechada mas os valores de ρ_1 e ρ_2 só podem ser obtidos numericamente, iterando-se a expressão^[147]:

$$\Omega^{(n)} = \Omega^{(n-2)} + (\Omega^{(n-1)} - \Omega^{(n-2)}) \left(\frac{\vartheta - \omega^{(n-2)}}{\omega^{(n-1)} - \omega^{(n-2)}} \right) \quad (\text{A.6})$$

que converge a $\Omega^{(conv)}$ em cerca de $n = 5$ passos e $\rho_a = (2\Omega^{(conv)})^{-1}$. Para $a = 1$, a expressão A.6 contém os seguintes termos:

$$\vartheta = h_{sp} \quad (\text{A.7})$$

$$\Omega^{(1)} = \frac{1}{2} \left[\frac{\vartheta}{e^2(D_1)^2} \right]^{(1/3)} \quad (\text{A.8})$$

$$\Omega^{(2)} = \Omega^{(1)} \left(\frac{\vartheta}{\omega^{(1)}} \right) \quad (\text{A.9})$$

$$\omega^{(n)} = \frac{1}{2}\Omega^{(n)} - \frac{1}{2}[4(D_1)^2 + (\Omega^{(n)})^{-2}]^{-1/2} \quad (\text{A.10})$$

e para $a = 2$, a expressão A.6 contém:

$$\vartheta = \frac{1}{2}(g_{pp'} - g_{p2}) \quad (\text{A.11})$$

$$\Omega^{(1)} = \frac{1}{2} \left[\frac{\vartheta}{e^2 3(D_2)^4} \right]^{(1/5)} \quad (\text{A.12})$$

$$\Omega^{(2)} = \Omega^{(1)} \left(\frac{\vartheta}{\omega^{(1)}} \right) \quad (\text{A.13})$$

$$\omega^{(n)} = \frac{1}{4}\Omega^{(n)} - \frac{1}{2}[4(D_2)^2 + (\Omega^{(n)})^{-2}]^{-1/2} + \frac{1}{4}[8(D_2)^2 + (\Omega^{(n)})^{-2}]^{-1/2} \quad (\text{A.14})$$

As expressões são repetidas aqui explicitamente, porque ao conhecimento do autor apenas uma referência bibliográfica^[147] apresenta estas equações e as expressões dadas para $\omega_{(n)}$ estão incorretas.

Nos potenciais híbridos (§1.6.4) cuja parte quântica é descrita por uma hamiltoniana semiempírica e cujas interações eletrostáticas entre os centros quânticos e clássicos estão sujeitas a truncagem, o esquema proposto na p. 71 não pode ser diretamente utilizado. A função de corte $S(r)$ (equação 1.74) tem que ser multiplicada às derivadas de $(\mu\nu|\lambda\sigma)$ e a integração deste produto resulta nas expressões truncadas usadas nas equações 1.83 e 1.84^[179].

APÊNDICE B – Cálculo de propriedades termodinâmicas

A função de partição z é uma quantidade fundamental na mecânica estatística e pode ser pensada como uma medida do número de estados efetivamente acessíveis ao sistema^[201]. Podemos assumir que a função de partição pode ser separada em contribuições eletrônicas, translacionais, rotacionais e vibracionais:

$$z = z_{elet} z_{trans} z_{rot} z_{vib} \quad (\text{B.1})$$

Na maioria dos casos, a parte eletrônica é $z_{elet} = 1$, pois não existe degenerescência do estado fundamental e os estados eletrônicos excitados praticamente não contribuem para z_{elet} , pois suas energias são muito maiores que a do estado fundamental.

A parte translacional é tratada classicamente e as partes rotacionais e vibracionais podem ser separadas e calculadas pelas aproximações de rotor rígido e oscilador harmônico, respectivamente. Assim, para uma molécula não linear, temos^{[201, cap. 8][24, §7.6]}:

$$z_{trans} = \left(\frac{2\pi M k_B T}{h^2} \right)^{3/2} V \quad (\text{B.2})$$

$$z_{rot} = \frac{\sqrt{\pi}}{\varpi} \sum_{y=A,B,C} \left(\frac{8\pi^2 \mathcal{I}_y k_B T}{h^2} \right)^{1/2} \quad (\text{B.3})$$

$$z_{vib} = \prod_{i=1}^{3N-6} \frac{\exp[-h\nu_i/(2k_B T)]}{1 - \exp[-h\nu_i/(k_B T)]} \quad (\text{B.4})$$

onde M é a massa, V o volume ocupado e ϖ o número de simetria da molécula ($\varpi = 1$ para moléculas assimétricas). Os momentos de inércia $\mathcal{I}_y$, ($y = A, B, C$) são facilmente calculados a partir da estrutura e das massas atômicas da molécula. As $3N - 6$ frequências vibracionais ν_i podem ser calculadas dos autovalores ς_i da hessiana (equação 1.53) ponderada por massa, em coordenadas internas e projetada (dos modos translacionais e rotacionais)^{[171, cap. 2][200]}:

$$\nu_i = \frac{\sqrt{\varsigma_i}}{2\pi} \quad (\text{B.5})$$

A hessiana pode ser obtida com uma boa precisão por cálculos de estrutura eletrônica (§1.4.1 e §1.5.1).

A energia do ponto-zero é a contribuição energética do estado vibracional fundamental ($n = 0$) e pode ser facilmente calculada por:

$$ZPE = \sum_i \frac{1}{2} h\nu_i \quad (\text{B.6})$$

onde a somatória corre sobre todos os i modos vibracionais da molécula.

Assumindo que as moléculas se comportam como um gás ideal (como é o caso para moléculas isoladas na fase gasosa), podemos obter as seguintes propriedades termodinâmicas a partir da função de partição:

$$E = RT^2 \left(\frac{\partial \ln z}{\partial T} \right)_V \quad (\text{B.7})$$

$$S = R \ln z + RT \left(\frac{\partial \ln z}{\partial T} \right)_V \quad (\text{B.8})$$

$$H = E + PV \quad (\text{B.9})$$

$$A = E - TS \quad (\text{B.10})$$

$$G = H - TS \quad (\text{B.11})$$

onde E é a energia interna, R é a constante dos gases perfeitos, S a entropia, H a entalpia, P a pressão, A a energia livre de Helmholtz e G a energia livre de Gibbs. A notação $()_V$ indica que a derivação foi efetuada a volume constante.

Se considerarmos um equilíbrio entre dois estados, $A \rightleftharpoons B$, a constante de equilíbrio é:

$$K_{eq} = \exp \left[- \left(\frac{G_B - G_A}{RT} \right) \right] \quad (\text{B.12})$$

A constante de velocidade de reações, $k(T)$, pode ser obtida na chamada teoria do estado de transição canônica^[7, cap. 6], se assumirmos que um equilíbrio entre reagente e um complexo ativado leva ao produto de reação, $A \rightleftharpoons A^\ddagger \rightarrow B$. A constante $k(T)$ é dada por uma expressão como 1.7, onde $\kappa(T) = 1$ e $\Delta G^\ddagger = G_{A^\ddagger} - G_A$.

Referências Bibliográficas

- 1 LAD, C.; WILLIAMS, N. H.; WOLFENDEN, R. The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases. *Proc. Natl. Acad. Sci. USA*, v. 100, p. 5607–5610, 2003.
- 2 MENDER, F. M. Analysis of ground-state and transition-state effects in enzyme catalysis. *Biochemistry*, v. 31, p. 5368–5373, 1992.
- 3 FIELD, M. J. Simulating enzyme reactions: Challenges and perspectives. *J. Comp. Chem.*, v. 23, p. 48–58, 2002.
- 4 WARSHEL, A.; FLORIÁN, J. Computer simulation of enzyme catalysis: Finding out what has been optimized by evolution. *Proc. Natl. Acad. Sci. USA*, v. 95, p. 5950–5955, 1998.
- 5 FERSHT, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. 1st. ed. New York: W. H. Freeman, 1999.
- 6 VOET, D.; VOET, J. *Biochemistry*. 2nd. ed. New York: Wiley, 1995.
- 7 BILLING, G. D.; MIKKELSEN, V. *Introduction to Molecular Dynamics and Chemical Kinetics*. 1st. ed. New York: Wiley, 1996.
- 8 JENCKS, W. P. *Catalysis in Chemistry and Enzymology*. 1st. ed. New York: Dover, 1987.
- 9 KOLLMAN, P. A.; KUHN, B.; PERAKYLA, M. Computational studies of enzyme-catalyzed reactions: Where are we in predicting mechanisms and in understanding the nature of enzyme catalysis? *J. Phys. Chem. B*, v. 106, p. 1537–1542, 2002.
- 10 WARSHEL, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. 1st. ed. New York: Wiley, 1991.
- 11 KOLLMAN, P. A. et al. Elucidating the nature of enzyme catalysis utilizing a new twist on an old methodology: Quantum mechanical-free energy calculations on chemical reactions in enzymes and in aqueous solution. *Acc. Chem. Res.*, v. 34, p. 72–79, 2001.
- 12 ROCA, M. et al. Theoretical modeling of enzyme catalytic power: Analysis of “cratic” and electrostatic factors in catechol o-methyltransferase. *J. Am. Chem. Soc.*, v. 125, p. 7726–7737, 2003.
- 13 WARSHEL, A. Electrostatic origin of the catalytic power of the enzymes and the role of the preorganized active site. *J. Biol. Chem.*, v. 273, p. 27035–27038, 1998.
- 14 BENKOVIC, S. J.; HAMMES-SCHIFFER, S. A perspective on enzyme catalysis. *Science*, v. 301, p. 1196–1202, 2003.

- 15 PAULING, L. Nature of forces between large molecules of biological interest. *Nature*, v. 161, p. 707–709, 1948.
- 16 GARCIA-VILOCA, M. et al. How enzymes work: Analysis by modern rate theory and computer simulations. *Science*, v. 303, p. 186–195, 2004.
- 17 MENGER, F. M. On the source of intramolecular and enzymatic reactivity. *Acc. Chem. Res.*, v. 18, p. 128–134, 1985.
- 18 BRUICE, T. C. A view at the millennium: the efficiency of enzymatic catalysis. *Acc. Chem. Res.*, v. 35, p. 139–148, 2002.
- 19 ÅQVIST, J.; WARSHEL, A. Simulation of enzyme-reactions using valence-bond force-fields and other hybrid quantum-classical approaches. *Chem. Rev.*, v. 93, p. 2523–2544, 1993.
- 20 HENGGE, A. C. Isotope effects in the study of phosphoryl and sulfonyl transfer reactions. *Acc. Chem. Res.*, v. 35, p. 105–112, 2002.
- 21 LEATHERBARROW, R. J.; FERSHT, A. R.; WINTER, G. Transition-state stabilization in the mechanism of tyrosyl-tRNA synthetase revealed by protein engineering. *Proc. Natl. Acad. Sci. USA*, v. 82, p. 7840–7844, 1985.
- 22 RUDOLPH, J. Catalytic mechanism of CDC25. *Biochemistry*, v. 41, p. 14613–14623, 2002.
- 23 BRUICE, T. C.; LIGHTSTONE, F. C. Ground state and transition state contributions to the rates of intramolecular and enzymatic reactions. *Acc. Chem. Res.*, v. 32, p. 127–136, 1999.
- 24 FIELD, M. J. *A Practical Introduction to the Simulation of Molecular Systems*. 1st. ed. Cambridge: Cambridge University Press, 1999.
- 25 VILLÀ, J.; WARSHEL, A. Energetics and dynamics of enzymatic reactions. *J. Phys. Chem. B*, v. 105, p. 7887–7907, 2001.
- 26 SHURKI, A. et al. How much do enzymes really gain by restraining their reacting fragments? *J. Am. Chem. Soc.*, v. 124, p. 4097–4107, 2002.
- 27 GREGERSEN, B. A.; LOPEZ, X.; YORK, D. M. Hybrid QM/MM study of thio effects in transphosphorylation reactions. *J. Am. Chem. Soc.*, v. 125, p. 7178–7179, 2003.
- 28 THAIN, E. M. Reactions of esters of phosphorothioic acid. Part I. A comparison of the hydrolyses of triethyl phosphorotrithiolate and triethyl phosphate. *J. Chem. Soc.*, p. 4694–4699, 1957.
- 29 HAMMES, G. G. Multiple conformation changes in enzyme catalysis. *Biochemistry*, v. 41, p. 8221–8228, 2002.
- 30 KOHEN, A.; KLINMAN, J. P. Enzyme catalysis: Beyond classical paradigms. *Acc. Chem. Res.*, v. 31, p. 397–404, 1998.
- 31 AYALA, P.; SCHLEGEL, H. B. Identification and treatment of internal rotation in normal mode vibrational analysis. *J. Chem. Phys.*, v. 108, p. 2314–2325, 1998.

- 32 VILLÁ, J. et al. How important are entropic contributions to enzyme catalysis? *Proc. Natl. Acad. Sci. USA*, v. 97, p. 11899–11904, 2000.
- 33 YU, Y. B. Comment on “Computational studies of enzyme-catalyzed reactions: Where are we in predicting mechanisms and in understanding the nature of enzyme catalysis”. *J. Phys. Chem. B*, v. 107, p. 1721, 2003.
- 34 ANTONIOU, D. et al. Barrier passage and protein dynamics in enzymatically catalyzed reactions. *Eur. J. Biochem.*, v. 269, p. 3103–3112, 2002.
- 35 TRUHLAR, D. G.; GARRETT, B. C.; KLIPPENSTEIN, S. J. Current status of transition-state theory. *J. Phys. Chem.*, v. 100, p. 12771–12800, 1996.
- 36 WESTHEIMER, F. H. Why nature chose phosphates. *Science*, v. 235, p. 1173–1178, 1987.
- 37 COX, J. R.; RAMSAY, O. B. Mechanisms of nucleophilic substitution in phosphate esters. *Chem. Rev.*, v. 64, n. 4, p. 317–351, 1964.
- 38 BUNTON, C. A. Hydrolysis of orthophosphate esters. *Acc. Chem. Res.*, v. 3, p. 257–265, 1970.
- 39 THATCHER, G. R. J.; KLUGER, R. Mechanism and catalysis of nucleophilic-substitution in phosphate-esters. *Adv. Phys. Org. Chem.*, v. 25, p. 99–265, 1989.
- 40 LUM, R. C.; GRABOWSKI, J. J. Trimethyl phosphate - the intrinsic reactivity of carbon versus phosphorus sites with anionic nucleophiles. *J. Am. Chem. Soc.*, v. 114, n. 22, p. 8619–8627, 1992.
- 41 HODGES, R. V.; SULLIVAN, S. A.; BEAUCHAMP, J. L. Nucleophilic reactions of anions with trimethyl phosphate in the gas-phase by ion-cyclotron resonance spectroscopy. *J. Am. Chem. Soc.*, v. 102, n. 3, p. 935–938, 1980.
- 42 JACKSON, M. D.; DENU, J. M. Molecular reaction of protein phosphatases - insights from structure and chemistry. *Chem. Rev.*, v. 101, n. 8, p. 2313–2340, 2001.
- 43 BARFORD, D.; DAS, A. K.; EGLOFF, M. P. The structure and mechanism of protein phosphatase: Insights into catalysis and regulation. *Annu. Rev. Biophys. Biomol. Struct.*, v. 27, p. 133–164, 1998.
- 44 GUTHRIE, R. D.; JENCKS, W. P. IUPAC recommendations for the representation of reaction mechanisms. *Acc. Chem. Res.*, v. 22, p. 343–349, 1989.
- 45 CLELAND, W. W.; HENGGE, A. C. Mechanisms of phosphoryl and acyl transfer. *FASEB J.*, v. 9, p. 1585–1594, 1995.
- 46 PARR, R. G.; YANG, W. *Density-Functional Theory of Atoms and Molecules*. 1st. ed. Oxford: Oxford University Press, 1996.
- 47 EDWARDS, J.; PEARSON, R. The factors determining nucleophilic reactivities. *J. Am. Chem. Soc.*, v. 84, p. 16–24, 1962.
- 48 WOLFE, S. et al. Theoretical studies of S_N2 transition states. The alpha effect. *Tetrahedron Lett.*, v. 23, p. 615–618, 1982.

- 49 PROSS, A. *Theoretical and Physical Principles of Organic Reactivity*. 1st. ed. New York: Wiley, 1995.
- 50 KHAN, S.; KIRBY, A. The reactivity of phosphate esters. Multiple structure-reactivity correlations for the reactions of triesters with nucleophiles. *J. Chem. Soc. B*, p. 1172–1182, 1970.
- 51 HOFF, R. H.; HENGGE, A. C. Entropy and enthalpy contributions to solvent effects in phosphate monoester solvolysis. The importance of entropy effects in the dissociative transition state. *J. Org. Chem.*, v. 63, p. 6680–6688, 1998.
- 52 ÅQVIST, J. et al. Mechanistic alternatives in phosphate monoester hydrolysis: what conclusions can be drawn from available experimental data? *Chem. Biol.*, v. 6, p. R71–R80, 1999.
- 53 WESTHEIMER, F. H. Pseudo-rotation in hydrolysis of phosphate esters. *Acc. Chem. Res.*, v. 1, p. 70–78, 1968.
- 54 CHANG, N. Y.; LIM, C. Factors governing the enhanced reactivity of five-membered cyclic phosphate esters. *J. Am. Chem. Soc.*, v. 120, p. 2156–2167, 1998.
- 55 DUBOURG, A. et al. Phosphoranes with a P-OH bond - crystal-structure and acidity. *J. Chem. Res. (S)*, v. 7, p. 180–181, 1982.
- 56 LAHIRI, S. D. et al. The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science*, v. 299, p. 2067–2071, 2003.
- 57 YLINIEMEIA, A. et al. Do pentacoordinate oxyphosphorane intermediates always exist? *J. Am. Chem. Soc.*, v. 115, p. 3032–3033, 1993.
- 58 LIM, C.; KARPLUS, M. Nonexistence of dianionic pentacovalent intermediates in an ab initio study of the base-catalyzed hydrolysis of ethylene phosphate. *J. Am. Chem. Soc.*, v. 112, p. 5872–5873, 1990.
- 59 ISAACS, N. S. *Physical Organic Chemistry*. 2nd. ed. New York: Longman, 1995.
- 60 PURICH, D. L. (Ed.). *Contemporary Enzyme Kinetics and Mechanism*. San Diego: Academic Press, 1996.
- 61 GUTHRIE, J. Hydration and dehydration of phosphoric acid derivatives: Free energies of formation of the pentacoordinate intermediates for phosphate ester hydrolysis and of monomeric metaphosphate. *J. Am. Chem. Soc.*, v. 99, p. 3991–4001, 1977.
- 62 KIRBY, A.; VARVOGLIS, A. The reactivity of phosphate monoesters: Monoester hydrolysis. *J. Am. Chem. Soc.*, v. 89, p. 415–423, 1968.
- 63 WOLFENDEN, R.; RIDGWAY, C.; YOUNG, G. Spontaneous hydrolysis of ionized phosphate monoesters and diesters and the proficiencies of phosphatases and phosphodiesterases as catalysts. *J. Am. Chem. Soc.*, v. 120, p. 833–834, 1998.
- 64 KIRBY, A.; VARVOGLIS, A. The reactivity of phosphate esters: Reactions of monoesters with nucleophiles. Nucleophilicity independent of basicity in a bimolecular substitution reaction. *J. Chem. Soc. (B)*, p. 135–141, 1968.

- 65 KIRBY, A.; JENCKS, W. The reactivity of nucleophilic reagents toward the p-nitrophenyl phosphate dianion. *J. Am. Chem. Soc.*, v. 87, p. 3209–3216, 1965.
- 66 FRIEDMAN, J.; FREEMAN, S.; KNOWLES, J. The quest for free metaphosphate in solution. Racemization at phosphorus in the transfer of the phospho group from aryl phosphate monoesters to tert-butyl alcohol in acetonitrile or in tert-butyl alcohol. *J. Am. Chem. Soc.*, v. 110, p. 1268–1275, 1988.
- 67 GRZYŹSKA, P. et al. Generality of solvation effects on the hydrolysis rates of phosphate monoesters and their possible relevance to enzymatic catalysis. *J. Org. Chem.*, v. 67, p. 1214–1220, 2002.
- 68 BIANCIOTTO, M.; BARTHELAT, J.-C.; VIGROUX, A. Reactivity of phosphate monoester monoanions in aqueous solution. 1. Quantum mechanical calculations support the intermediate in the dissociative hydrolysis of the methyl phosphate anion. *J. Am. Chem. Soc.*, v. 124, p. 7573–7587, 2002.
- 69 BIANCIOTTO, M.; BARTHELAT, J.-C.; VIGROUX, A. Reactivity of phosphate monoester monoanions in aqueous solution. 2. A theoretical study of the elusive zwitterion intermediates $\text{RO}^+(\text{H})\text{PO}_3^{2-}$. *J. Phys. Chem. A*, v. 106, p. 6521–6526, 2002.
- 70 FLORIÁN, J.; WARSHEL, A. A fundamental assumption about OH^- attack in phosphate ester hydrolysis is not fully justified. *J. Am. Chem. Soc.*, v. 119, p. 5473–5474, 1997.
- 71 DONAHUE, N. M. Reaction barriers: Origin and evolution. *Chem. Rev.*, v. 103, p. 4593–4604, 2003.
- 72 FLORIÁN, J.; ÅQVIST, J.; WARSHEL, A. On the reactivity of phosphate monoester dianions in aqueous solution: Bronsted linear free-energy relationships do not have a unique mechanistic interpretation. *J. Am. Chem. Soc.*, v. 120, p. 11524–11525, 1998.
- 73 FLORIÁN, J.; WARSHEL, A. Phosphate ester hydrolysis in aqueous solution: Associative versus dissociative mechanisms. *J. Phys. Chem. B*, v. 102, p. 719–734, 1998.
- 74 HU, C.; BRINCK, T. Theoretical studies of the hydrolysis of the methyl phosphate anion. *J. Phys. Chem. A*, v. 103, p. 5379–5386, 1999.
- 75 GLENNON, T. M.; VILLÁ, J.; WARSHEL, A. How does GAP catalyze the GTPase reaction of Ras?: A computer simulation study. *Biochemistry*, v. 39, p. 9641–9651, 2000.
- 76 DANTZMAN, C. L.; KIESSLING, L. L. Reactivity of a 2'-thio nucleotide analog. *J. Am. Chem. Soc.*, v. 118, p. 11715–11719, 1996.
- 77 BUNTON, C. A.; FENDLER, E. J.; FENDLER, J. H. The hydrolysis of dinitrophenyl phosphates. *J. Am. Chem. Soc.*, v. 89, p. 1221–1230, 1967.
- 78 MILLER, B. Phosphorothioates. II. The effect of the nucleophilicity of the attacking anion upon rates of displacement on O,O-diphenyl phosphorochloridothioate. *J. Am. Chem. Soc.*, v. 84, p. 403–409, 1962.
- 79 SHAIK, S. S.; SCHLEGEL, H. B.; WOLFE, S. *Theoretical Aspects of Physical Organic Chemistry: The $\text{S}_\text{N}2$ Mechanism*. 1st. ed. New York: Wiley, 1992.

- 80 MORRIS, R. et al. The gas-phase acidity of H_3PO_4 . *J. Chem. Phys.*, v. 106, p. 3545–3547, 1997.
- 81 ASUBIOJO, O. I.; BRAUMAN, J. I.; LEVIN, R. H. Gas phase hydrolysis of phosphorous esters. *J. Am. Chem. Soc.*, v. 99, n. 23, p. 7707–7708, 1977.
- 82 CHERMETTE, H. Chemical reactivity indexes in density funtional theory. *J. Comp. Chem.*, v. 20, p. 129–154, 1999.
- 83 GOLDSTEIN, B. J. *Tyrosine Phospoprotein Phosphatases*. 2nd. ed. Oxford: Oxford University Press, 1998.
- 84 NEEL, B. G.; TONKS, N. K. Protein tyrosine phosphatases in signal transduction. *Curr. Opin. Cell. Biol.*, v. 9, p. 193–204, 1997.
- 85 PANNIFER, A. D. B. et al. Visualization of the cysteinyl-phosphate intermediate of a protein-tyrosine phosphatase by X-ray crystallography. *J. Bio. Chem.*, v. 273, p. 10454–10462, 1998.
- 86 ZHANG, Z. Y. Chemical and mechanistic approaches to the study of protein tyrosine phosphatases. *Acc. Chem. Res.*, v. 36, p. 385–392, 2003.
- 87 XIE, L. et al. Cellular effects of small molecule PTP1B inhibitors on insulin signaling. *Biochemistry*, v. 42, p. 12792–12804, 2003.
- 88 YUVANIYAMA, J. et al. Crystal structure of the dual specificity protein phosphatase VHR. *Science*, v. 272, p. 1328–1331, 1996.
- 89 DUNN, D. et al. The active site specificity of the *yersinia* protein-tyrosine phosphatase. *J. Biol. Chem.*, v. 271, p. 168–173, 1996.
- 90 ZHANG, Z.-Y. Are protein-tyrosine phosphatases specific for phosphotyrosine? *J. Biol. Chem.*, v. 270, p. 16052–16055, 1995.
- 91 CIRRI, P. et al. The role of CYS12, CYS17 and ARG18 in the catalytic mechanism of low- M_r cytosolic phosphotyrosine protein phosphatase. *Eur. J. Biochem.*, v. 214, p. 647–657, 1993.
- 92 TADDEI, N. et al. Aspartic-129 is an essential residue in the catalytic mechanism of the low M_r phosphotyrosine protein phosphatase. *FEBS Lett.*, v. 350, p. 328–332, 1994.
- 93 CHEN, W.; WILBORN, M.; RUDOLPH, J. Dual-specific CDC25B phosphatase: In search of the catalytic acid. *Biochemistry*, v. 39, p. 10781–10789, 2000.
- 94 FORLANI, F.; CARPEN, A.; PAGANI, S. Evidence that elongation of the catalytic loop of the azotobacter vinelandii rhodanese changed selectivity from sulfur- to phosphate-containing substrates. *Protein Eng.*, v. 16, p. 515–519, 2003.
- 95 DENU, J. et al. Visualization of the intermediate and transition-state structures in protein-tyrosine phosphatase catalysis. *Proc. Natl. Acad. Sci. USA*, v. 93, p. 2493–2498, 1996.
- 96 ZHOU, G. et al. The catalytic role of CYS in the dual specificity phosphatase VHR. *J. Biol. Chem.*, v. 269, p. 28084–28090, 1994.

- 97 ZHANG, Z. Kinetic and mechanistic characterization of a mammalian protein-tyrosine phosphatase, PTP1. *J. Biol. Chem.*, v. 270, p. 11199–11204, 1995.
- 98 ZHANG, Z.-Y.; WU, L.; CHEN, L. Transition state and rate-limiting step of the reaction catalyzed by the human dual-specificity phosphatase, VHR. *Biochemistry*, v. 34, p. 16088–16096, 1995.
- 99 GOTTLIN, E. B. et al. Kinetic analysis of the catalytic domain of human CDC25B. *J. Biol. Chem.*, v. 271, p. 27445–27449, 1996.
- 100 HOFF, R. H. et al. Does positive charge at the active sites of phosphatases cause a change in mechanism? The effect of the conserved arginine on the transition state for phosphoryl transfer in the protein-tyrosine phosphatase from *yersinia*. *J. Am. Chem. Soc.*, v. 121, p. 9514–9521, 1999.
- 101 ZHANG, Z.-Y. et al. The CYS(X)₅ARG catalytic motif in phosphoester hydrolysis. *Biochemistry*, v. 33, p. 15266–15270, 1994.
- 102 BARFORD, D.; FLINT, A.; TONKS, N. Crystal structure of the human protein tyrosine phosphatase 1B. *Science*, v. 263, p. 1397–1404, 1994.
- 103 DENU, J. M. et al. The catalytic role of aspartic acid-92 in a human dual-specific protein-tyrosine-phosphatase. *Biochemistry*, v. 34, p. 3396–3403, 1995.
- 104 LOHSE, D. L. et al. Roles of aspartic acid-181 and serine-222 in intermediate formation and hydrolysis of the mammalian protein-tyrosine-phosphatase PTP1. *Biochemistry*, v. 36, p. 4568–4575, 1997.
- 105 MCCAIN, D. F. et al. The catalytic mechanism of CDC25A phosphatase. *J. Biol. Chem.*, v. 277, p. 11190–11200, 2002.
- 106 FAUMAN, E. B. et al. Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. *Cell*, v. 93, p. 617–625, 1998.
- 107 REYNOLDS, R. A. et al. Crystal structure of the catalytic subunit of CDC25B required for G2/M phase transition of the cell cycle. *J. Mol. Biol.*, v. 293, p. 559–568, 1999.
- 108 KIM, J.-H. et al. Mutational and kinetic evaluation of conserved HIS-123 in dual specificity protein-tyrosine phosphatase *vaccinia* H1-related phosphatase. *J. Biol. Chem.*, v. 276, p. 27568–27574, 2001.
- 109 DENU, J. M.; DIXON, J. E. A catalytic mechanism for the dual-specific phosphatases. *Proc. Natl. Acad. Sci. USA*, v. 92, p. 5910–5914, 1996.
- 110 ALHAMBRA, C. et al. Walden-inversion-enforced transition-state stabilization in a protein tyrosine phosphatase. *J. Am. Chem. Soc.*, v. 120, p. 3858–3866, 1998.
- 111 ALHAMBRA, C.; GAO, J. Hydrogen-bonding interactions in the active site of a low molecular weight protein-tyrosine phosphatase. *J. Comp. Chem.*, v. 21, p. 1192–1203, 2000.

- 112 KOLMODIN, K.; ÅQVIST, J. Computational modeling of catalysis and binding in low-molecular-weight protein tyrosine phosphatase. *Int. J. Quant. Chem.*, v. 73, p. 147–159, 1999.
- 113 KOLMODIN, K.; ÅQVIST, J. Computational modeling of the rate limiting step in low molecular weight protein tyrosine phosphatase. *FEBS Lett.*, v. 456, p. 301–305, 1999.
- 114 ASTHAGIRI, D. et al. Density functional study of the mechanism of a tyrosine phosphatase: I. Intermediate formation. *J. Am. Chem. Soc.*, v. 124, p. 10225–10235, 2002.
- 115 KOLMODIN, K.; ÅQVIST, J. The catalytic mechanism of protein tyrosine phosphatases revisited. *FEBS Lett.*, v. 498, p. 208–213, 2001.
- 116 ZHANG, Z.-Y.; WANG, Y.; DIXON, J. Dissecting the catalytic mechanism of protein-tyrosine phosphatase. *Proc. Natl. Acad. Sci. USA*, v. 91, p. 1624–1627, 1994.
- 117 EVANS, B. et al. Site-directed mutagenesis, kinetic, and spectroscopic studies of the P-loop residues in a low molecular weight protein tyrosine phosphatase. *Biochemistry*, v. 35, p. 13609–13617, 1996.
- 118 PETERS, G.; FRIMURER, T.; OLSEN, O. Electrostatic evaluation of the signature motif (H/V)CX₅R(S/T) in protein-tyrosine phosphatase. *Biochemistry*, v. 37, p. 5383–5393, 1998.
- 119 DILLET, V.; VAN ETTEN, R. L.; BASHFORD, D. Stabilization of charges and protonation states in the active site of the protein tyrosine phosphatases: A computational study. *J. Phys. Chem. B*, v. 104, p. 11321–11333, 2000.
- 120 HANSSON, T.; NORDLUND, P.; ÅQVIST, J. Energetics of nucleophile activation in a protein tyrosin phosphatase. *J. Mol. Biol.*, v. 265, p. 118–127, 1997.
- 121 HELGAKER, T.; JØRGENSEN, P.; OLSEN, J. *Molecular Electronic-Structure Theory*. 1st. ed. New York: Wiley, 2000.
- 122 HIRSCHFELDER, J. O.; MEATH, W. J. The nature of intermolecular forces. *Adv. Chem. Phys.*, v. 12, p. 3–106, 1967.
- 123 SAKURAI, J. J. *Modern Quantum Mechanics*. 1st. ed. Menlo Park, USA: Benjamin/Cummings, 1985.
- 124 SZABO, A.; OSTLUND, N. *Modern Quantum Chemistry*. 1st. ed. New York: Dover, 1989.
- 125 ROOHTAAN, C. C. J. New developments in molecular orbital theory. *Rev. Mod. Phys.*, v. 23, p. 69–89, 1951.
- 126 SLATER, J. C. *Quantum Theory of Atomic Structure*. 1st. ed. New York: McGraw-Hill, 1960.
- 127 MØLLER, C.; PLESSET, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, v. 46, p. 618, 1934.

- 128 HEAD-GORDON, M.; POPLE, J. A.; FRISCH, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.*, v. 153, p. 503–506, 1988.
- 129 RASSOLOV, V. A. et al. The definition of core electrons. *Chem. Phys. Lett.*, v. 350, p. 573–576, 2001.
- 130 KOHN, W.; BECKE, A. D.; PARR, R. G. Density functional theory of electronic structure. *J. Phys. Chem.*, v. 100, p. 12974–12980, 1996.
- 131 BECKE, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.*, v. 98, p. 1372–1377, 1993.
- 132 BECKE, A. D. Density functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, v. 98, p. 5648–5652, 1993.
- 133 BECKE, A. D. Exchange-correlation approximations in density functional theory. In: YARKONY, D. R. (Ed.). *Modern Electronic Structure Theory*. Singapore: World Scientific, 1995. v. 2, p. 1022–1046.
- 134 DITCHFIELD, R.; HEHRE, W.; POPLE, J. A. Self-consistent molecular-orbital methods. IX. An extended gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.*, v. 54, p. 724–728, 1971.
- 135 KRISHNAN, R. et al. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.*, v. 72, p. 650–654, 1980.
- 136 SCHAEFER, A.; HORN, H.; AHLRICHS, R. Fully optimized contracted gaussian basis sets for atoms Li to K. *J. Chem. Phys.*, v. 97, p. 2571–2577, 1992.
- 137 HAMERMESH, M. *Group Theory and Its Application to Physical Problems*. 1st. ed. New York: Dover, 1989.
- 138 TAYLOR, P. R. Accurate calculations and calibration. In: ROOS, B. O. (Ed.). *Lectures Notes in Quantum Chemistry - European Summer School in Quantum Chemistry*. Berlin: Springer-Verlag, 1992. v. 1, p. 325–412.
- 139 FRISCH, M. J.; POPLE, J. A. Self-consistent molecular orbital methods 25. supplementary functions for gaussian basis sets. *J. Chem. Phys.*, v. 80, p. 3265–3269, 1984.
- 140 VAN DUIJNEVELDT, F. B.; VAN DE RIJDT, J. G. C. M.; VAN LENTHE, J. H. State of the art in counterpoise theory. *Chem. Rev.*, v. 94, p. 1873–1885, 1994.
- 141 LENDVAY, G.; MAYER, I. Some difficulties in computing BSSE-corrected potential surfaces of chemical reactions. *Chem. Phys. Lett.*, v. 297, p. 365–373, 1998.
- 142 KLOPMAN, G.; EVANS, R. C. The neglect-of-differential-overlap methods of molecular orbital theory. In: SEGAL, G. A. (Ed.). *Semiempirical Methods of Electronic Structure Calculation*. New York: Plenum, 1977. v. 7, p. 29–67.
- 143 DEWAR, M.; THIEL, W. Ground states of molecules. 38. The MNDO method. Approximation and parameters. *J. Am. Chem. Soc.*, v. 99, p. 4899–4907, 1977.

- 144 DEWAR, M. et al. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, v. 107, p. 3902–3909, 1985.
- 145 STEWART, J. Optimization of parameters for semiempirical methods .1. Method. *J. Comp. Chem.*, v. 10, p. 209–220, 1989.
- 146 ROBY, K. R. Fundamentals of an orthonormal basis set molecular orbital theory. *Chem. Phys. Lett.*, v. 12, p. 579–582, 1972.
- 147 STEWART, J. MOPAC - a semiempirical molecular-orbital program. *J. Comput. Aid. Mol. Des.*, v. 4, p. 1–45, 1990.
- 148 THIEL, W. Semiempirical methods. In: GROTENDORST, J. (Ed.). *Modern Methods and Algorithms of Quantum Chemistry*. Jülich: John von Neumann Institute for Computing, 2001. v. 3, p. 261–283.
- 149 SLATER, J. C. *Quantum Theory of Atomic Structure*. 1st. ed. New York: McGraw-Hill, 1960.
- 150 DEWAR, M.; LO, D. Ground states of σ -bonded molecules. XVII. Fluorine compounds. *J. Am. Chem. Soc.*, v. 94, p. 5296–5301, 1972.
- 151 OLEARI, L.; SIPIO, L. D.; MICHELIS, G. D. The evaluation of the one-center integrals in the semi-empirical molecular orbital theory. *Mol. Phys.*, v. 10, p. 97–109, 1966.
- 152 DEWAR, M.; JIE, C. AM1 parameters for phosphorus. *THEOCHEM*, v. 187, p. 1–13, 1989.
- 153 DEWAR, M.; YUAN, Y. AM1 parameters for sulfur. *Inorg. Chem.*, v. 29, p. 3881–3890, 1990.
- 154 PRESS, W. et al. *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*. 2nd. ed. Cambridge: Cambridge University Press, 1992.
- 155 GOLDBERG, D. *Genetic Algorithms in Search, Optimization and Machine Learning*. 1st. ed. Reading: Addison-Wesley, 1989.
- 156 JUDSON, R. Genetic algorithms and their use in chemistry. In: LIPKOVITZ, K.; BOYD, D. (Ed.). *Reviews in Computational Chemistry*. New York: VHC, 1997. v. 10, p. 1–74.
- 157 ROSSI, I.; TRUHLAR, D. Parametrization of NDDO wavefunctions using genetic algorithms. An evolutionary approach to parametrizing potential energy surfaces and direct dynamics calculations for organic reactions. *Chem. Phys. Lett.*, v. 233, p. 231–236, 1995.
- 158 REPASKY, M. P.; CHANDRASEKHAR, J.; JORGENSEN, W. L. PDDG/PM3 and PDDG/MNDO: Improved semiempirical methods. *J. Comput. Chem.*, v. 23, p. 1601–1622, 2002.
- 159 DEWAR, M. J. S.; O'CONNOR, B. M. Testing abinitio procedures ; the 6-31G* model. *Chem. Phys. Lett.*, v. 138, p. 141–145, 1987.

- 160 CURTISS, L. A. et al. Gaussian-2 (G2) theory: Reduced basis set requirements. *J. Chem. Phys.*, v. 104, p. 5148–5152, 1996.
- 161 ANH, N. T. et al. Some difficulties encountered with AM1 and PM3 calculations. *Tetrahedron*, v. 54, p. 12841–12852, 1998.
- 162 REED, A. E.; SCHLEYER, P. v. R. Chemical bonding in hypervalent molecules. The dominance of ionic bonding and negative hyperconjugation over d-orbital participation. *J. Am. Chem. Soc.*, v. 113, p. 1434–1445, 1990.
- 163 AHLWEDE, B.; JUG, K. Consistent modifications of SINDO1: I. Approximations and parameters. *J. Comput. Chem.*, v. 20, p. 563–571, 1999.
- 164 THIEL, W.; VOITYUK, A. Extension of the MNDO formalism to d orbitals: Integral approximations and preliminary numerical results. *Theor. Chim. Acta*, v. 81, p. 391–404, 1992.
- 165 HOGGAN, P.; RINALDI, D. Calculation of coulomb integrals in molecules over an spd basis of STO. *Theor. Chim. Acta*, v. 72, p. 47–55, 1987.
- 166 BERNAL-URUCHRTU, M.; RUIZ-LOPEZ, M. Basic ideas for the correction of semiempirical methods describing H-bonded systems. *Chem. Phys. Lett.*, v. 330, p. 118–124, 2000.
- 167 BASH, P. A. et al. Progress toward chemical accuracy in the computer simulation of condensed phase reactions. *Proc. Natl. Acad. Sci. USA*, v. 93, p. 3698–3703, 1996.
- 168 LOPEZ, X.; YORK, D. M. Parameterization of semiempirical methods to treat nucleophilic attacks to biological phosphates: AM1/d parameters for phosphorus. *Theor. Chem. Acc.*, v. 109, p. 149–159, 2003.
- 169 HELGAKER, T. Optimization of minima and saddle points. In: ROOS, B. O. (Ed.). *Lectures Notes in Quantum Chemistry - European Summer School in Quantum Chemistry*. Berlin: Springer-Verlag, 1992. v. 1, p. 295–324.
- 170 SCHLEGEL, H. B. Optimization of equilibrium geometries and transition structures. *Adv. Chem. Phys.*, v. 67, p. 249–286, 1987.
- 171 WILSON JR., E. B.; DECIUS, J. C.; CROSS, P. C. *Molecular Vibrations. The Theory of Infrared and Raman Vibrational Spectra*. 1st. ed. Ney York: Dover, 1980.
- 172 JEZIORSKI, B.; MOSZYNSKI, R.; SZALEWICZ, K. Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem. Rev.*, v. 94, p. 1887–1930, 1994.
- 173 CORNELL, W. D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, v. 117, p. 5179–5197, 1995.
- 174 JORGENSEN, W. L.; MAXWELL, D. S.; TIRADO-RIVES, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, v. 118, p. 11225–11236, 1996.

- 175 CHALMET, S.; RUIZ-LÓPEZ, M. F. New approaches to the description of short-range repulsion interactions in hybrid quantum/classical systems. *Chem. Phys. Lett.*, v. 329, p. 154–159, 2000.
- 176 ALLEN, M.; TILDESLEY, D. *Computer Simulation of Liquids*. 1st. ed. New York: Oxford University Press, 1987.
- 177 KOTTALAM, J.; CASE, D. A. Dynamics of ligand escape from the heme pocket of myoglobin. *J. Am. Chem. Soc.*, v. 110, p. 7690–7697, 1988.
- 178 VAN KAMPEN, N. G. *Stochastic Processes in Physics and Chemistry*. 1st. ed. Amsterdam: North Holland Publishing, 1981.
- 179 FIELD, M. J. et al. The DYNAMO library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.*, v. 21, p. 1088–1100, 2000.
- 180 STRAATSMA, T. P.; MCCAMMON, J. A. Computational alchemy. *Annu. Rev. Phys. Chem.*, v. 43, p. 407–435, 1992.
- 181 JORGENSEN, W. J. Free energy calculations: A breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.*, v. 22, p. 184–189, 1989.
- 182 ROUX, B. The calculation of the potential of mean force using computer simulations. *Comp. Phys. Comm.*, v. 91, p. 275–282, 1995.
- 183 TORRIE, G. M.; VALLEAU, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.*, v. 28, p. 578–581, 1974.
- 184 GAO, J. Methods and applications of combined quantum mechanical and molecular mechanical potentials. In: LIPKOVITZ, K.; BOYD, D. (Ed.). *Reviews in Computational Chemistry*. New York: VHC, 1995. v. 7, p. 119–185.
- 185 SHERWOOD, P. Hybrid quantum mechanics/molecular mechanics approaches. In: GROTEENDORST, J. (Ed.). *Modern Methods and Algorithms of Quantum Chemistry*. Jülich: John von Neumann Institute for Computing, 2001. v. 3, p. 257–277.
- 186 MONARD, G.; MERZ JR., K. M. Combined quantum mechanical / molecular mechanical methodologies applied to biomolecular systems. *Acc. Chem. Res.*, v. 32, p. 904–911, 1999.
- 187 FIELD, M. J.; BASH, P. A.; KARPLUS, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.*, v. 11, p. 700–733, 1990.
- 188 CUMMINS, P. L.; GREASY, J. E. Coupled semiempirical quantum mechanics and molecular mechanics (QM/MM) calculations on the aqueous solvation free energies of ionized molecules. *J. Comput. Chem.*, v. 20, p. 1028–1038, 1999.
- 189 TU, Y.; LAAKSONEN, A. On the effect of Lennard-Jones parameters on the quantum mechanical and molecular mechanical coupling in a hybrid molecular dynamics simulation of liquid water. *J. Chem. Phys.*, v. 111, p. 7519–7525, 1999.

- 190 LUQUE, F. J. et al. Calibration of the quantum/classical Hamiltonian in semiempirical QM/MM AM1 and PM3 methods. *J. Phys. Chem. A*, v. 104, p. 10923–10931, 2000.
- 191 THOMPSON, M. A. QM/MMpol: A consistent model for solute/solvent polarization. Application to the aqueous solvation and spectroscopy of formaldehyde, acetaldehyde, and acetone. *J. Phys. Chem.*, v. 100, p. 14492–14507, 1996.
- 192 AMARA, P.; FIELD, M. J. Combined quantum mechanical and molecular mechanical potentials. In: SCHLEYER, P. v. R. et al. (Ed.). *Encyclopedia of Computational Chemistry*. 1st. ed. Chichester: Wiley, 1998. v. 1, p. 431–436.
- 193 DEWAR, M. J. S.; THIEL, W. A semiempirical model for the two-center repulsion integrals in the NDDO approximation. *Theor. Chim. Acta*, v. 46, p. 89–104, 1977.
- 194 RINALDI, D. et al. *GEOMOP*. 2001. Nancy, França.
- 195 REUTER, N. et al. Frontier bonds in QM/MM methods: A comparison of different approaches. *J. Phys. Chem. A*, v. 104, p. 1720–1735, 2000.
- 196 MONARD, G. et al. Hybrid classical quantum force field for modeling very large molecules. *Int. J. Quant. Chem.*, v. 58, p. 153–159, 1996.
- 197 ASSFELD, X.; RIVAIL, J.-L. Quantum chemical computations on parts of large molecules: The ab initio local self consistent field method. *Chem. Phys. Lett.*, v. 263, p. 100–106, 1996.
- 198 GAO, J. et al. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A*, v. 102, p. 4714–4712, 1998.
- 199 STECKLER, R.; TRUHLAR, D. G. Reaction-path power-series analysis of NH_3 inversion. *J. Chem. Phys.*, v. 93, p. 6570–6577, 1990.
- 200 OCHTERSKI, J. W. *White paper: Vibrational analysis in Gaussian*. 1999. Gaussian Inc. <http://www.gaussian.com>.
- 201 MCQUARRIE, D. A. *Statistical Mechanics*. 1st. ed. New York: Harper and Row, 1976.
- 202 HEHRE, W. J. et al. *Ab Initio Molecular Orbital Theory*. 1st. ed. New York: Wiley, 1986.
- 203 GEORGE, L.; VISWANATHAN, K. S.; SINGH, S. Ab initio study of trimethyl phosphate: Conformer analysis, dipole moments, vibrational frequencies and barriers for conformer interconversion. *J. Phys. Chem. A*, v. 101, p. 2459–2464, 1997.
- 204 FUKUI, K. The path of chemical-reactions - the IRC approach. *Acc. Chem. Res.*, v. 14, p. 363–368, 1981.
- 205 GONZALEZ, C.; SCHLEGEL, H. B. Reaction path following in mass-weighted internal coordinates. *J. Phys. Chem.*, v. 94, p. 5523–5527, 1990.

- 206 FRISCH, M. J. et al. *Gaussian 98 Rev. A.9*. 1998. Gaussian Inc. Pittsburgh.
- 207 DUNNING, T. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, v. 90, p. 1007–1023, 1989.
- 208 AHLRICHS, R. et al. Electronic structure calculations on workstation computers: The program system Turbomole. *Chem. Phys. Lett.*, p. 165–169, 1989.
- 209 DEBIAN Project. 2004. <http://www.debian.org>.
- 210 GNU Project. 2004. <http://www.gnu.org>.
- 211 MENEGON, G.; LOOS, M.; CHAIMOVICH, H. Ab initio study on the thiolysis of trimethyl phosphate ester in the gas-phase. *J. Phys. Chem. A*, v. 106, p. 9078–9084, 2002.
- 212 CHANG, N.; LIM, C. Ab initio study of nucleophilic attack at trimethyl phosphate: Factors influencing site reactivity. *J. Phys. Chem. A*, v. 101, p. 8706–8713, 1997.
- 213 CUMMING, J.; KEBARLE, P. Summary of gas phase measurements involving acids AH. Entropy changes in proton transfer reactions involving negative ions. Bond dissociation energies $D(A-H)$ and electron affinities $EA(A)$. *Can. J. Chem.*, v. 56, p. 1–9, 1978.
- 214 OSBORN, D. et al. Photoelectron spectroscopy of CH_3O^- and CD_3O^- . *Chem. Phys. Lett.*, v. 292, p. 651–655, 1998.
- 215 GUNION, R. et al. Ultraviolet photoelectron spectroscopy of the phenide, benzyl, and phenoxide anions. *Int. J. Mass Spectrom. Ion Proc.*, v. 117, p. 601–620, 1992.
- 216 IVASH, E.; DENNINSON, D. The methyl alcohol molecule and its microwave spectrum. *J. Chem. Phys.*, v. 21, p. 1804–1816, 1953.
- 217 WEISS, S.; LEROY, G. Direct observation of the infrared torsional spectrum of C_2H_6 , CH_3CD_3 and C_2D_4 . *J. Chem. Phys.*, v. 48, p. 962–967, 1968.
- 218 GITANY, R.; MCEWEN, R. Molecular conformation of S-4-nitrophenyl O,O-diphenyl thiophosphate: X-ray crystal structure analysis. *J. Chem. Soc. Perkin 2*, p. 57–59, 1975.
- 219 KARLE, J.; KARLE, I. Structure of the radiation protection agent s-2-(3-aminopropylamino)ethylphosphorothioic acid (wr 2721). *Acta Crystallogr Sect. C Cryst. Struct. Commun.*, v. 44, p. 135, 1988.
- 220 SANTORO, E. The fragmentation of some alkyl thiophosphate esters by electron-impact. *Org. Mass Spectrom.*, v. 7, p. 589–600, 1973.
- 221 GEORGE, L. et al. Matrix-isolation infrared-spectroscopy of organic-phosphates. *Appl. Spectrosc.*, v. 48, p. 7–12, 1994.
- 222 VIDYA, V.; SANKARAN, K.; VISWANATHAN, K. S. Matrix isolation-supersonic jet infrared spectroscopy: Conformational cooling in trimethyl phosphate. *Chem. Phys. Lett.*, v. 258, p. 113–117, 1996.

- 223 BALLARD, R. E. et al. He(i) photoelectron studies of liquids and gases. *Chem. Phys. Lett.*, v. 137, p. 125–129, 1987.
- 224 NELSON, R.; LIDE, D.; MARYOTT, A. Selected values of electric dipole moments for molecules in the gas phase. *Natl. Stand. Ref. Data Ser.*, v. 10, p. 1, 1967.
- 225 ROBIN, M.; KUEBLER, N. Excited electronic states of the simple alcohols. *J. Electron Spectrosc. Relat. Phenom.*, v. 1, p. 13–28, 1972.
- 226 LESS, R.; BAKER, J. Torsion-vibration-rotation interactions in methanol. I. Millimeter wave spectrum. *J. Chem. Phys.*, v. 48, p. 5299–5318, 1968.
- 227 LIAS, S. et al. Gas-phase ion and neutral thermochemistry. *J. Phys. Chem. Ref. Data*, v. 17, p. 1–816, 1988. Suplemento 1.
- 228 HARMONY, M. et al. Molecular structures of gas-phase polyatomic molecules determined by spectroscopic methods. *J. Phys. Chem. Ref. Data*, v. 8, p. 619–722, 1979.
- 229 DERISSEN, J. A reinvestigation of the molecular structure of acetic acid monomer and dimer by gas electron diffraction. *J. Mol. Struct.*, v. 7, p. 67–80, 1971.
- 230 MERCERO, J. M. et al. Quantum mechanical calculations on phosphate hydrolysis reactions. *J. Comp. Chem.*, v. 21, p. 43–51, 2000.
- 231 HERZBERG, G. *Molecular Spectra and Molecular Structure III. Electronic Spectra and Eletronic Structure of Polyatomic Molecules*. 1st. ed. New York: Van Nostrand, 1966.
- 232 ZVEREV, V.; VILLEM, Y. Ionization potentials of phosphoryl compounds. *J. Struct. Chem.*, v. 21, p. 22–26, 1980.
- 233 STRECK, R. et al. Conformational behaviour of trimethyl phosphate studied by infrared spectroscopy. *J. Mol. Strut.*, v. 376, p. 277–287, 1996.
- 234 ROOS, B. O. The multiconfigurational (MC) self-consistent field (SCF) theory. In: ROOS, B. O. (Ed.). *Lectures Notes in Quantum Chemistry - European Summer School in Quantum Chemistry*. Berlin: Springer-Verlag, 1992. v. 1, p. 177–254.
- 235 PULAY, P.; HAMILTON, T. UHF natural orbitals for defining and starting MC-SCF calculations. *J. Chem. Phys.*, v. 88, p. 4926–4933, 1988.
- 236 TOMASI, J.; PERSICO, M. Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent. *Chem. Rev.*, v. 94, p. 2027 – 2094, 1994.
- 237 ROUX, B.; SIMONSON, T. Implicit solvent models. *Biol. Chem.*, v. 78, p. 1–20, 1999.
- 238 BOLDYREV, A.; GUTOWSKI, M.; SIMONS, J. Small multiply charged anions as building blocks in chemistry. *Acc. Chem. Res.*, v. 29, p. 497–502, 1996.
- 239 WOLFENDEN, R. Analog approaches to the structure of the transition state in enzyme reactions. *Acc. Chem. Res.*, v. 5, p. 10–18, 1972.

- 240 CZYRYCA, P. G.; HENGGE, A. C. The mechanism of the phosphoryl transfer catalyzed by *yersinia* protein-tyrosine phosphatase: a computational and isotope effect study. *Biochimica et Biophysica Acta*, v. 1547, p. 245–253, 2001.
- 241 LAU, E. Y. et al. The importance of reactant positioning in enzyme catalysis: A hybrid quantum mechanics / molecular mechanics study of haloalkane dehalogenase. *Proc. Natl. Acad. Sci. USA*, v. 97, p. 9937–9942, 2000.
- 242 MENEGON, G. et al. Parameterization of the electronegativity equalization method based on the charge model 1. *Phys. Chem. Chem. Phys.*, v. 4, p. 5933–5936, 2002.
- 243 CARROLL, D. L. *Genetic algorithm driver, version 1.70*. 1998. University of Illinois. E-mail: carroll@uiuc.edu.
- 244 PAGE, C. G. *Professional Programmer's Guide to FORTRAN77*. University of Leicester, 1995.
- 245 DEWAR, M.; REYNOLDS, C. H. An improved set of MNDO parameters for sulfur. *J. Comput. Chem.*, v. 7, p. 140–143, 1986.
- 246 DEWAR, M.; MCKEE, M.; RZEPA, H. MNDO parameters for third period elements. *J. Am. Chem. Soc.*, v. 100, p. 3607, 1978.
- 247 JORGENSEN, W. L. et al. Comparison of simple potential functions for simulating liquid water. *J. Phys. Chem.*, v. 79, p. 926–935, 1983.
- 248 STEFANOVICH, E. V. et al. Ab initio study of the stabilization of multiply charged anions in water. *J. Phys. Chem. B*, v. 102, p. 4205–4208, 1998.
- 249 SOMMERFELD, T. A fresh look at aromatic dianions. *J. Am. Chem. Soc.*, v. 124, p. 1119–1124, 2002.
- 250 MAGNUSSON, E. Hypercoordinate molecules of second-row elements: d functions or d orbitals? *J. Am. Chem. Soc.*, v. 112, p. 7940–7951, 1990.
- 251 REED, A. E.; WEINSTOCK, R. B.; WEINHOLD, F. Natural population analysis. *J. Chem. Phys.*, v. 83, p. 735–746, 1985.
- 252 REED, A. E.; WEINHOLD, F. On the role of d orbitals in SF₆. *J. Am. Chem. Soc.*, v. 108, p. 3587–3593, 1986.
- 253 FIELD, M. J. Hybrid QM/MM fluctuating charge models for condensed phase simulations. *Mol. Phys.*, v. 91, p. 835–845, 1997.
- 254 ANTONCZAK, S. et al. Modeling of peptide hydrolysis by thermolysin. A semiempirical and QM/MM study. *J. Am. Chem. Soc.*, v. 120, p. 8825–8833, 1998.
- 255 The Protein Data Bank. 2002. Royal Society Protein Data Bank. <http://www.pdb.org>.
- 256 MACKERELL JR., A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, v. 102, p. 3586–3616, 1998.

- 257 KIM, J.-H. et al. Mutational and kinetic evaluation of conserved His-123 in dual specificity protein-tyrosine phosphatase vaccinia h1-related phosphatase. *J. Biol. Chem.*, v. 276, p. 27568–27574, 2001.
- 258 ZHANG, Z. Y.; DIXON, J. E. Active-site labeling of the *yersinia* protein-tyrosine-phosphatase - the determination of the pK(a) of the active-site cysteine and the function of the conserved histidine-402. *Biochemistry*, v. 32, n. 36, p. 9340–9345, 1993.
- 259 TISHMACK, P. A. et al. Use of H NMR spectroscopy and computer simulations to analyze histidine pKa changes in a protein tyrosine phosphatase: Experimental and theoretical determination of electrostatic properties in a small protein. *Biochemistry*, v. 36, p. 11984–11994, 1997.
- 260 MARTIN, F. P.-D.; DUMAS, R.; FIELD, M. J. A hybrid-potential free-energy study of the isomerization step of the acetohydroxy acid isomeroreductase reaction. *J. Am. Chem. Soc.*, v. 122, p. 7688–7697, 2000.
- 261 THOMAS, A. et al. Is there a covalent intermediate in the viral neuraminidase reaction? a hybrid potential free-energy study. *J. Am. Chem. Soc.*, v. 121, p. 9693–9702, 1999.
- 262 DELLAGO, C.; BOLHUIS, P. G.; GEISLER, P. L. Transition path sampling. *Adv. Phys. Chem.*, v. 123, p. 1–86, 2002.
- 263 CHANDLER, D. Barrier crossings: classical theory of rare but important events. In: BERNE, B.; CICCOTTI, G.; COKER, D. (Ed.). *Computer Simulation of Rare Events and Dynamics of Classical and Quantum Condensed-Phase Systems – Classical and Quantum Dynamics in Condensed Phase Simulations*. Singapore: World Scientific, 1998. v. 1, p. 3–23.
- 264 ZHANG, Z.; HARMS, E.; VAN ETTEN, R. L. Asp129 of low molecular weight protein tyrosine phosphatase is involved in leaving group protonation. *J. Biol. Chem.*, v. 269, p. 25947–25950, 1994.
- 265 WU, L.; ZHANG, Z.-Y. Probing the function of Asp128 in the low molecular weight protein-tyrosine phosphatase-catalyzed reaction. A pre-steady-state and steady-state kinetic investigation. *Biochemistry*, v. 35, p. 5426–5434, 1996.
- 266 FLINT, A. J. et al. Development of “substrate-trapping” mutants to identify physiological substrates of protein tyrosine phosphatases. *Proc. Natl. Acad. Sci. USA*, v. 94, p. 1680–1685, 1997.
- 267 MARK, A. E. Free energy perturbation calculations. In: SCHLEYER, P. v. R. et al. (Ed.). *Encyclopedia of Computational Chemistry*. 1st. ed. Chichester: Wiley, 1998. v. 2, p. 1070–1083.
- 268 HINSEN, K.; ROUX, B. Potential of mean force and reaction rates for proton transfer in acetylacetone. *J. Chem. Phys.*, v. 106, p. 3567–3577, 1997.
- 269 FEYNMAN, R. P.; HIBBS, A. R. *Quantum Mechanics and Path Integrals*. 1st. ed. New York: McGraw-Hill, 1965.

- 270 JARZYNSKI, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, v. 78, p. 2690–2693, 1997.
- 271 PEARLMAN, D. A.; RAO, G. Free energy calculations: Methods and applications. In: SCHLEYER, P. v. R. et al. (Ed.). *Encyclopedia of Computational Chemistry*. 1st. ed. Chichester: Wiley, 1998. v. 2, p. 1036–1061.
- 272 SORIANO, A.; SILLA, E.; TUNÓN, I. Internal rotation of 1,2-dichloroethane in haloalkane dehalogenase. A test case for analyzing electrostatic effects in enzymes. *J. Phys. Chem. B*, v. 107, p. 6234–6238, 2003.
- 273 HENGGE, A. C.; DENU, J. M.; DIXON, J. E. Transition-state structures for the native dual-specific phosphatase VHR and D92N and S131A mutants. Contributions to the driving force for catalysis. *Biochemistry*, v. 35, p. 7084–7092, 1996.
- 274 HART, J. C. et al. An alternative role for the conserved ASP residue in phosphoryl transfer reactions. *J. Am. Chem. Soc.*, v. 120, p. 13535–13536, 1998.
- 275 SCAPIN, G. et al. The structure of apo protein-tyrosine phosphatase 1B C215S mutant: More than just an S → O change. *Protein Sci.*, v. 10, p. 1596–1605, 2001.
- 276 ZHANG, Z.; WU, L. The single sulfur to oxygen substitution in the active site nucleophile of the *yersinia* protein-tyrosine phosphatase leads to substantial structural and functional perturbations. *Biochemistry*, v. 36, p. 1362–1369, 1997.
- 277 ZHANG, Y. et al. Thermodynamic study of ligand binding to protein-tyrosine phosphatase 1b and its substrate-trapping mutants. *J. Biol. Chem.*, v. 275, p. 34205–34212, 2000.
- 278 JIA, Z. et al. Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science*, v. 268, p. 1754–1758, 1995.
- 279 HAN, R.; COLEMAN, J. E. Dependence of the phosphorylation of alkaline phosphatase by phosphate monoesters on the pK_a of the leaving group. *Biochemistry*, v. 34, p. 4238–4245, 1995.
- 280 KOLMODIN, K.; ÅQVIST, J. Prediction of a ligand-induced conformational change in the catalytic core of cdc25A. *FEBS Letters*, v. 465, p. 8–11, 2000.