

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA**

Programa de Pós Graduação em Ciências Biológicas (Bioquímica)

André Sabeça Anversa Reis

**Fosfatase Cdc25B: Interações transientes intramoleculares
e complexação com pequenos ligantes**

Versão corrigida da tese defendida

SÃO PAULO

Data de depósito na SPG

26/05/2022

André Sabença Anversa Reis

Fosfatase Cdc25B: Interações transientes intramoleculares e
complexação com pequenos ligantes

*Tese apresentada ao Instituto de Química da
Universidade de São Paulo para obtenção do
título de Doutor em Ciências (Bioquímica)*

Área de Concentração: Bioquímica

Orientador: Dr. Guilherme Menegon Arantes

São Paulo

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha Catalográfica elaborada eletronicamente pelo autor, utilizando o programa desenvolvido pela Seção Técnica de Informática do ICMC/USP e adaptado para a Divisão de Biblioteca e Documentação do Conjunto das Químicas da USP

Bibliotecária responsável pela orientação de catalogação da publicação:
Marlene Aparecida Vieira - CRB - 8/5562

R375f Reis, André Sabença Anversa
 Fosfatase Cdc25B: Interações transientes
 intramoleculares e complexação com pequenos ligantes
 / André Sabença Anversa Reis. - São Paulo, 2022.
 153 p.

Tese (doutorado) - Instituto de Química da
Universidade de São Paulo. Departamento de
Bioquímica.
Orientador: Arantes, Guilherme Menegon

1. Simulação de dinâmica molecular. 2. Cdc25B. 3.
Ancoragem molecular. 4. Interações cátion- π . 5.
Proteínas intrinsecamente desordenadas. I. T. II.
Arantes, Guilherme Menegon, orientador.



Universidade de São Paulo
Instituto de Química

"Fosfatase Cdc25B: Interações transientes intramoleculares e complexação com pequenos ligantes"

ANDRÉ SABENÇA ANVERSA REIS

Tese de Doutorado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Doutor em Ciências obtido no Programa Ciências Biológicas (Bioquímica) - Área de Concentração: Bioquímica.

Prof. Dr. Guilherme Menegon Arantes
(Orientador e Presidente)

APROVADO(A) POR:

Prof. Dr. Roberto Kopke Salinas (por videoconferência)
IQ - USP

Prof. Dr. Gustavo Henrique Goulart Trossini
FCF - USP

Profa. Dra. Ana Lígia Barbour Scott
UFABC

SÃO PAULO
15 de setembro de 2022

Aos meus pais, aos meus irmãos e à minha esposa, minha família amada.

AGRADECIMENTOS

Agradeço à minha família, sem a qual esta tese não seria possível. Seja pelo suporte financeiro, logístico ou emocional, e, sobretudo, pelo carinho e amor dedicados ao longo de décadas, obrigado José, Christian, Rafael, Lucas e Gabriela, vocês são as pessoas mais importantes e é um prazer e uma honra poder dividir este Mundo com vocês.

Ao professor Dr. Guilherme Menegon Arantes, que me ensinou muito e demonstrou uma paciência quase infinita ao longo destes mais de cinco anos. Saio desta experiência com uma grande aprendizagem pela nossa convivência.

Aos amigos de vida Bruna Rodrigues, Cláudio Colli, Lilian Anversa, Nicolli Colli e Guilherme Raposo Medeiros. Obrigado por tornarem a vida mais leve.

Aos colegas de trabalho Ariane, Murilo, Felipe e Sofia, que dividiram o laboratório por agradáveis anos e me auxiliaram muito ao longo deste projeto.

Aos professores Dr. Sandro Roberto Marana e Dr. Bayardo Baptista Torres, sempre gentis. Levarei muitos ensinamentos pela vida.

Aos funcionários do Instituto de Química da Universidade de São Paulo, concursados ou terceirizados, tão importantes na manutenção de um ambiente seguro e confortável, além da responsabilidade no andamento burocrático da instituição.

Ao Instituto de Química cujo firmamento, paredes e teto me abrigaram por tantos dias, tardes, noites e madrugadas. Um local fundado na ciência, casa de grandes mentes e espíritos.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro, por meio da bolsa de doutorado nº 141683/2016-3.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo suporte financeiro à pesquisa do laboratório.

EPÍGRAFE

O objetivo da educação é que sejamos capazes de nos deleitar e sofrer com aquilo que nos deve causar deleite e sofrimento.

Platão

RESUMO

Reis, A. S. A. **Fosfatase Cdc25B: Interações transientes intramoleculares e complexação com pequenos ligantes.** 2022. 153 p. Tese de Doutorado - Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

Proteínas tirosina-fosfatase (PTPs) possuem papel fundamental na regulação da transdução de sinais e estão envolvidas em diversos processos fundamentais do ciclo celular. As Cdc25 (*Cell Division Cycle 25*) são fosfatases duais encontradas em todos os organismos eucarióticos e atuam em *checkpoints* do ciclo celular, permitindo ou inibindo o prosseguimento deste. Este grupo de proteínas pertence à classe de PTPs com atividade baseada em cisteína, apresenta domínio catalítico altamente conservado assim como o motivo catalítico, *P-loop*. Devido sua função, as Cdc25 são consideradas possíveis alvos terapêuticos para tratamento de câncer e sua interação com pequenas moléculas e inibidores tem sido investigada de forma que análises estruturais e de ligação das Cdc25 com inibidores podem elucidar aspectos importantes do mecanismo de ação destes além de direcionar para o desenho racional de fármacos.

Interações cátion- π são interações intra ou intermoleculares não-covalentes que ocorrem entre uma espécie química catiônica, como o grupo guanidino de argininas, e uma das faces de um sistema π rico em elétrons, como dos anéis indólicos de triptofanos. Apesar de pouco discutidas na literatura, quando em comparação às interações não-covalentes mais convencionais, do ponto de vista energético as interações cátion- π são tão importantes na estruturação de proteínas quanto às ligações de hidrogênio ou pontes salinas. De fato estas interações são observadas com frequência em estruturas proteicas resolvidas. O domínio catalítico da Cdc25B possui diversas argininas expostas em sua superfície e um único resíduo de triptofano localizado na região C-terminal flexível, muito próximo do sítio catalítico da proteína. A flexibilidade de proteínas ou de regiões proteicas apresenta importante papel no reconhecimento entre biomoléculas participantes de vias de sinalização e tem sido muito estudada atualmente.

Aqui, simulações de dinâmica molecular, experimentos de ^1H - ^{15}N HSQC RMN, ensaios de cinética de inibição e de ancoragem molecular, evidenciam a existência de contatos cátion- π transientes na superfície de um importante membro da família das Cdc25, a Cdc25B, e de sítios de interação entre inibidores testados e a proteína com destaque a sítios na proximidades do *P-loop*, região próxima ao C-terminal desordenado, onde se demonstra estabilidade da interação com os pequenos ligantes.

Palavras-chave: Simulação de dinâmica molecular; Cdc25B; Ancoragem molecular; RMN; Interações cátion- π ; Proteínas intrinsecamente desordenadas.

ABSTRACT

Reis, A. S. A. **Cdc25B phosphatase: Transient intramolecular interactions and small molecules complexation.** 2022. 153 p. PhD Thesis – Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Protein tyrosine phosphatase (PTPs) play a fundamental role in the regulation of signal transduction and are involved in several fundamental processes of the cell cycle. Cdc25 (Cell Division Cycle 25) are dual phosphatases found in all eukaryotic organisms and act at checkpoints of the cell cycle, allowing or inhibiting its progression. This group of proteins belongs to the class of PTPs with cysteine-based activity, presenting a highly conserved catalytic domain as well as the catalytic motif, P-loop. Due to their function, Cdc25 are considered possible therapeutic targets for cancer treatment and their interaction with small molecules and inhibitors has been investigated so that structural and binding analyzes of Cdc25 with inhibitors can elucidate important aspects of their mechanism of action besides directing to rational drug design.

Cation- π interactions are non-covalent intra- or intermolecular interactions that occur between a cationic chemical species, such as the guanidino group of arginines, and one of the faces of an electron-rich π system, such as the indole rings of tryptophans. Although little discussed in the literature, when compared to more conventional non-covalent interactions, from the energetic point of view, cation- π interactions are as important in the structuring of proteins as hydrogen bonds or salt bridges. In fact, these interactions are frequently observed in solved protein structures. The catalytic domain of Cdc25B has several arginines exposed on its surface and a single tryptophan residue located in the flexible C-terminal region, very close to the catalytic site of the protein. The flexibility of proteins or protein regions plays an important role in the recognition between biomolecules participating in signaling pathways and has been extensively studied today.

Here, molecular dynamics simulations, ^1H - ^{15}N HSQC NMR experiments, inhibition kinetics and molecular anchoring assays, evidence the existence of transient cation- π contacts on the surface of an important member of the Cdc25 family, Cdc25B, and of sites of interaction between tested inhibitors and the protein, with emphasis on sites in the vicinity of the P-loop, a region close to the disordered C-terminus, where stability of the interaction with the small ligands is demonstrated.

Keywords: Molecular dynamics simulation; Cdc25B; Molecular docking; NMR; Cation- π interactions; Intrinsically disordered proteins.

ABREVIATURA E SIGLAS

$\Delta\delta$	Variação de deslocamento químico
Abs	Absorbância (em dado comprimento de onda)
BMER	2-mercaptoetanol
BMRB	<i>Biological magnetic resonance bank</i> - Banco de dados de RMN de amostras biológicas
BSA	Albumina sérica bovina
C1-17N resíduos	Construção 1 (selvagem) do DC da Cdc25B sem os 17 primeiros resíduos
C-terminal	Carboxi-terminal proteico
Cak	Quinase ativadora de Cdk
Cdc	Família de proteínas de controle do ciclo celular
Cdk	Proteína quinase dependente de ciclina
CDM	Centro de massas
CSP	Perturbação de deslocamento químico
DC	Domínio catalítico
DNA	Ácido desoxirribonucleico
DO	Densidade óptica
DR	Domínio regulatório
FPLC	Cromatografia líquida rápida de proteínas
G ₀	Fase adicional de intervalo do ciclo celular (<i>Gap 0</i>)
G ₁	Primeira fase de intervalo do ciclo celular (<i>Gap 1</i>)
G ₂	Segunda fase de intervalo do ciclo celular (<i>Gap 2</i>)
HSQC	Correlação quântica única heteronuclear
I1	Inibidor 1 - CAS 289913-88-8
I2	Inibidor 2 - CAS 7330-46-3

I3	Inibidor 3 - CAS 5466-99-9
IPTG	Isopropil- β -D-1-tiogalactopiranosídeo
K_i	Constante de inibição
K_m	Constante de Michaelis-Menten
LB	Meio de cultura Luria-Bertani
M	Fase de mitose do ciclo celular
MES	Ácido 2-(N-morfolino)etanosulfônico
MW	Peso molecular
MWCO	<i>Cutoff</i> de peso molecular
N-terminal	Amino-terminal proteico
OMFP	Fosfato de 3-O-metil fluoresceína
P-loop	Motivo proteico catalítico “ <i>loop</i> de fosfato”
PAGE	Eletroforese de proteínas em gel de poliacrilamida
PDB	Banco de Dados de Proteínas online (<i>Protein Data Bank</i>)
pH	Cologaritmo da concentração hidrogeniônica
pI	Ponto isoelétrico
pKa	Cologaritmo da constante ácida
PID	Proteína intrinsecamente desordenada
PMSF	Fluoreto de fenilmetilsulfonila
pNP	para-Nitrofenol
pNPP	para-Nitrofenol fosfato
PRID	Proteína com região intrinsecamente desordenada
PTP	Proteína tirosina-fosfatase
R482A	Mutante com troca da arginina 482 do DC da Cdc25B por alanina
R544A	Mutante com troca da arginina 544 do DC da Cdc25B por alanina

RMN	Ressonância magnética nuclear
RMSD	<i>Root mean square deviation</i> - Desvio quadrático médio
RNA	Ácido ribonucleico
S	Fase de síntese do ciclo celular
S1	Sítio de ancoragem 1
S2	Sítio de ancoragem 2
S3	Sítio de ancoragem 3
S ²	Parâmetros de ordem
SDS	Dodecil sulfato de sódio
T-loop	Motivo proteico em laço ou em forma de "T"
US	<i>Umbrella Sampling</i> - Amostragem por potencial guarda-chuva
V _{máx}	Velocidade máxima de geração de produtos enzimática
W550A	Mutante com troca do triptofano 550 do DC da Cdc25B por alanina
WHAM	<i>Weighted histograms analysis method</i> - Método de análise por histogramas ponderados

SUMÁRIO

Introdução	25
1.1 Proteínas como reguladoras das atividades celulares e do ciclo celular	27
1.2 Estrutura de proteínas e interações cátion- π	31
1.3 Dinâmica conformacional de proteínas e proteínas desordenadas ou com regiões desordenadas	35
1.4 A Cdc25B	36
1.5 Dinâmica molecular e potenciais clássicos	42
1.6 Objetivos	45
Identificação de contatos cátion-π transientes na superfície do domínio catalítico da Cdc25B	46
2.1. Metodologia e desenvolvimento	46
2.1.1. Cálculos quânticos, simulações de dinâmica molecular e geração de perfis de energia livre	46
2.1.2. Amplificação de vetores de expressão do domínio catalítico da Cdc25B, análise de integridade e geração de mutantes pontuais	48
2.1.3. Expressão, extração e purificação da forma selvagem e dos mutantes do domínio catalítico da Cdc25B	52
2.1.3. Expressão, extração e purificação da forma selvagem e dos mutantes do domínio catalítico da Cdc25B	63
2.2. Resultados e discussão	65
2.2.1. Validação da estimativa energética da interação cátion- π por campos de força clássicos	65
2.2.2 Escolha de resíduos candidatos e geração de perfis de energia livre	66
2.2.3 Validação de contatos por perturbação de deslocamento químico em ressonância magnética nuclear	69
2.2.4 Dinâmica de parceiros e orientação dos contatos cátion- π	77
2.2.5 Conclusão	79
Reconhecimento molecular pela fosfatase Cdc25B	80
3.1. Metodologia e desenvolvimento	80
3.1.1. Expressão, extração e purificação da forma selvagem do domínio catalítico da Cdc25B	80
3.1.2. Cinética de inibição enzimática	80
3.1.3. Aquisição de espectros de ^1H - ^{15}N - HSQC RMN e análises de perturbação de deslocamento químico na presença dos inibidores testados	81
3.1.4. Desenho dos inibidores e ensaios de ancoragem molecular	82
3.1.5. Parametrização de campo de força e conferência da parametrização dos pequenos ligantes	83
3.1.6. Simulações de dinâmica molecular	83
3.2. Resultados e discussão	84
3.2.1. Cinética de inibição enzimática	84
3.2.2. Perturbação de deslocamento químico	86

3.2.3. Ancoragem e dinâmica molecular	88
3.2.3.1. Sítios S1, S2 e S3	90
3.2.3.2. Sítio catalítico: Cátion- π R482 x W550	95
3.2.3.3. Sítio catalítico: Cátion- π R544 x W550	103
3.2.3.4. Estabilidade do contato cátion- π R544 x W550 na estrutura utilizada para os ensaios de ancoragem molecular	109
3.3. Conclusão	110
Conclusão	112
Referências	114
ANEXO A - Sequenciamentos do domínio catalítico da Cdc25B selvagem e dos mutantes gerados	126
ANEXO B - Deslocamentos químicos do domínio catalítico da Cdc25B selvagem e dos mutantes gerados	132
ANEXO C - Artigo publicado: Combining free energy simulations and NMR chemical-shift perturbation to identify transient cation-π contacts in proteins.	141

Introdução

Seria difícil encontrar uma palavra para definir a beleza da complexidade existente nos seres vivos. Um sistema intrincado de processos regulados em diversas camadas baseadas na atuação de várias classes de biomoléculas com atividade definida pela adição e remoção de grupos químicos simples está presente desde a mais singela forma de vida. Surpreendentemente, vias metabólicas são comumente encontradas organizadas de forma semelhante, em termos de metabólitos, enzimas e reguladores presentes, nos organismos mais diversos.

É provável que, mesmo com todo o desenvolvimento atual da bioquímica, muitos ainda enxerguem com exacerbada reverência a nucleína de Miescher [1], uma posição dificilmente questionável. A biomolécula que há mais de 150 anos “surta” das ataduras da clínica de Tübingen para 75 anos mais tarde ter sua função desvendada por Avery, MacLeod e McCarty [2], e no início dos anos 1950 sua estrutura elucidada graças aos trabalhos de Watson, Crick e Franklin [3] guarda a honrosa função de definir o destino de cada componente celular a partir das longas frases escritas com seu breve alfabeto de quatro letras.

Se por um lado o sequenciamento do genoma humano [4] e o estudo de suas extensas regiões não transcritas ou não traduzidas [5] trouxeram grande avanço no entendimento estrutural, organizacional e da regulação da expressão gênica num dos organismos mais complexos que se conhece, por outro lado, avanços recentes no estudo de biomoléculas menos conhecidas do grande público demonstraram camadas regulatórias adicionais impressionantes no fluxo da informação somática. A imensa possibilidade de arranjo entre as moléculas de glicídios e o endereçamento de outras biomoléculas por parte dessas [6], a capacidade das proteínas de atuarem nas mais diversas funções e a flexibilidade estrutural de domínios de reconhecimento proteicos frente a diferentes parceiros são alguns dos exemplos de como moléculas biológicas já bem estudadas têm surpreendido a comunidade científica com seu dinamismo [7].

Dessa forma, se os códigos presentes nos ácidos nucleicos com sua regulação básica genética e epigenética, complementada pelos seus “novos” RNAs não codificadores de proteínas de tamanhos e funções diversas jamais deixarão de se apresentar majestosos por expressarem a linguagem da vida escrita por bilhões de anos de evolução, se os carboidratos, quase renegados pelo grande público

como agentes ativos na célula mas tão abundantes na superfície terrestre, tão importantes para as vias energéticas simples de todos os organismos e ainda tão misteriosos jamais deixarão de suscitar grande respeito pelos bioquímicos, as proteínas, por sua vez, nunca perderão seu posto de grandes responsáveis pela regulação celular geral, as maiores atuadoras na linha de frente da regulação da vida.

As proteínas são, certamente, as biomoléculas com maior diversidade de função nos organismos: proteínas fibrosas atuam como componentes estruturais celulares e da matriz extracelular; os principais agentes do sistema imune, os anticorpos, são proteínas altamente especializadas na ligação a epítopos; o transporte de substâncias diversas, que vão do oxigênio molecular a lipídios variados, pela corrente sanguínea, é muitas vezes feito por proteínas transportadoras; a nível celular a regulação da entrada e saída de substâncias incapazes de se difundir pela membrana da célula é realizado por proteínas transportadoras integrantes de membrana. Finalmente, as proteínas são importantes agentes em diversas etapas da regulação da expressão gênica e são as grandes responsáveis pela viabilidade cinética das reações bioquímicas atuando como fantásticos catalisadores biológicos capazes de alcançar, em alguns casos, a perfeição catalítica [8].

A partir do desenvolvimento das técnicas de resolução das estruturas de proteínas ao longo do século passado, inicialmente com a difração de raios-X, definiu-se o paradigma central da biologia estrutural que relaciona a estrutura tridimensional de uma determinada proteína à sua função [9]. Com a enorme quantidade de estruturas cristalográficas resolvidas e depositadas em bancos de dados especializados e pelo fato de tal técnica fornecer um modelo rígido para as estruturas proteicas estudadas, permitiu-se surgir uma visão precipitada sobre a realidade da estruturação de proteínas, imóvel, que acabou sendo adotada por muitos estudantes e interessados na área, tendo sido mantida durante um longo período.

Apesar de se poder observar a evolução no entendimento da dinâmica estrutural de proteínas ao longo dos séculos nos próprios modelos propostos para a complexação enzima-substrato, com a modificação do consagrado modelo chave-fechadura de Fischer (1894) [10], absolutamente rígido, para o modelo do ajuste induzido de Koshland (1958) [11], que previa mudanças conformacionais

tanto na enzima quanto em seu substrato, é provável que somente com o desenvolvimento da simulação computacional de dinâmica molecular de proteínas a noção de quanto as estruturas de proteínas podem se apresentar flexíveis tenha sido entendida e difundida de maneira considerável. De fato, atualmente entende-se que um dos principais fatores por trás do reconhecimento molecular é a chamada “seleção conformacional” [12], que assume que determinados ligantes reconhecem estados específicos de proteínas dentro de um *pool* de possibilidades, ignorando aquelas que encontram-se em conformações muito diferentes das ideais, em termos de ligação e catálise, por exemplo.

Assim, as proteínas se mostram como as biomoléculas mais versáteis presentes nos organismos, capazes de atuar nas mais variadas funções celulares sendo protagonistas de diversos processos, com destaque a função de reguladoras da atividade celular. Além disso, uma visão estática das proteínas, talvez definida a partir do paradigma estrutura - função, aliado aos modelos estruturais não dinâmicos de proteínas, dominantes ao longo da maior parte do século passado, foi superada com auxílio de novas técnicas de resolução estrutural e simulação computacional de biomoléculas sendo atualmente reconhecido o dinamismo estrutural proteico inclusive considerando-se este responsável por importantes papéis em reconhecimento molecular nos organismos.

1.1 Proteínas como reguladoras das atividades celulares e do ciclo celular

As proteínas estão direta ou indiretamente relacionadas à regulação de praticamente todos os processos celulares. Se a expressão gênica se inicia com a transcrição de um gene codificado em uma das fitas de DNA de um determinado cromossomo é porque um fator de transcrição proteico, a RNA polimerase e outras proteínas (e biomoléculas acessórias) atuaram no processo [13]. Mas antes disso, o DNA, enrolado em octâmeros de histonas, deve estar acessível e, para isto, estas proteínas globulares altamente básicas devem ter sua afinidade pela dupla fita diminuída, o que ocorre pela adição de grupos químicos a partir da atividade de outras proteínas [14].

Talvez a forma mais conhecida de atuação das proteínas seja enzimática, permitindo a adequação cinética das reações bioquímicas, que ocorrem na célula, à vida. Proteínas pertencentes às vias metabólicas de produção de energia devem

estar entre as enzimas mais bem conhecidas e estudadas e são também, muitas vezes, encontradas atuando em rotas diversas do metabolismo. Mas se a manutenção homeostática de uma célula quiescente já é um trabalho complexo [15], a divisão celular, com todas as suas etapas altamente reguladas e sistemas redundantes de checagem que devem acontecer enquanto a célula passa por processos totalmente diversos dos usuais, é ainda mais.

A regulação do ciclo celular pode ser considerada uma das funções mais importantes dentro do rol de atividades de uma célula. Enquanto para organismos pluricelulares o ciclo celular representa o desenvolvimento de um novo indivíduo, crescimento de um tecido para reposição de células mortas e regeneração, para organismos unicelulares, eucarióticos ou procarióticos, ele representa a origem de um novo ser completo. Dessa forma, uma das funções mais nobres que uma proteína pode possuir é a de reguladora do ciclo celular, responsável, diretamente, por um dos mecanismos-chave da manutenção da vida.

Tradicionalmente divide-se o ciclo celular eucariótico em quatro fases. Três dessas fases, G_1 , S e G_2 , sequenciais, nomeadas em conjunto de intérfase, representam quase 95% do tempo gasto para o ciclo celular se completar enquanto a quarta fase, a fase M, em que ocorre a divisão celular de fato, é consideravelmente mais curta. As denominações G_1 e G_2 da intérfase vêm do termo *Gap* (intervalo) e representam momentos em que as células começam ou finalizam processos que permitem seu crescimento e, muito além disso, garantem que a célula possa monitorar se as condições para o prosseguimento do ciclo celular estão adequadas. Se as condições externas ou internas não permitirem que a célula prossiga o fluxo normal de seu ciclo celular ela pode retardar o mesmo ou entrar em uma fase de intervalo prolongado, denominada G_0 , até que as condições se alterem novamente. Diversos tipos celulares possuem fases de intervalo adicionais ao longo de seu ciclo. Entre as fases de G_1 e G_2 , a fase S (síntese), especialmente importante, permite a duplicação do material genético ocupando cerca de metade do tempo necessário para um ciclo celular completo [16].

Já a fase M (mitose) do ciclo celular é subdividida em outras cinco fases: prófase, em que ocorre a condensação cromossomal e a carioteca e o nucléolo são desfeitos; prometáfase, em que os cromossomos se ligam ao fuso mitótico e iniciam seu deslocamento para o plano equatorial; metáfase, em que ocorre a máxima condensação cromossomal e forma-se a placa equatorial (ou metafásica), que

permitirá a separação dos cromossomos; anáfase, com a separação, propriamente dita, das cromátides, e; telófase, em que os cromossomos se descondensam e são novamente formados a carioteca e o nucléolo [16].

Os principais momentos de monitoramento do ciclo celular, já mencionado, se dão nas transições entre as fases G_1/S , G_2/M e, dentro da fase M, na transição metáfase/ anáfase (met./ anáf.). Aqui, destaca-se a checagem na transição G_2/M , responsável por verificar se o material genético celular foi completamente duplicado e se o mesmo encontra-se íntegro, ou seja, se eventuais danos ao DNA foram devidamente reparados. Dentro desse contexto é importante notar que danos não reparados no DNA tornam-se mutações e o destino da célula e, eventualmente, do organismo inteiro, pode ser prejudicado se a mutação apresentar efeitos negativos e ocorrer sobre um gene importante de alguma via metabólica essencial. Dessa forma, se forem identificadas alterações no código genético a célula é destinada para o processo de morte celular, a apoptose [16].

A família das proteínas quinase dependentes de ciclina (Cdk) é considerada a mais importante família proteica em termos de controle do ciclo celular. As Cdk são proteínas que atuam em conjunto com proteínas da família das ciclinas e, quando associadas a estas fosforilam proteínas-alvo envolvidas em processos do ciclo celular, ativando estas e, conseqüentemente, permitindo o início de tais processos. Na fase G_2/M o complexo então ativo, Cdk1-CiclinaB, promove a ativação de proteínas relacionadas à condensação dos cromossomos e rompimento da carioteca, primeiros processos da fase M (durante a prófase) do ciclo celular [17, 18].

A regulação do complexo Cdk1-CiclinaB, por sua vez, é um processo elegante que envolve proteínas diversas. No momento em que a proteína Cdk se liga à Ciclina ocorrem mudanças conformacionais que tornam o sítio ativo da Cdk mais exposto. O complexo se ativa totalmente a partir do estado de fosforilação da Cdk, podendo apresentar atividade intermediária dependendo de como este se encontra. A ativação do complexo é máxima na fosforilação de um resíduo de treonina no motivo chamado *T-loop* (o que ocorre por atuação de uma quinase ativadora de Cdk - Cak) e na ausência de grupos fosfato em resíduos vizinhos de tirosina e treonina (posições 14 e 15) no N-terminal da proteína. Esses últimos são fosforilados por uma proteína denominada Wee1 e desfosforilados por proteínas da família Cdc25 [19].

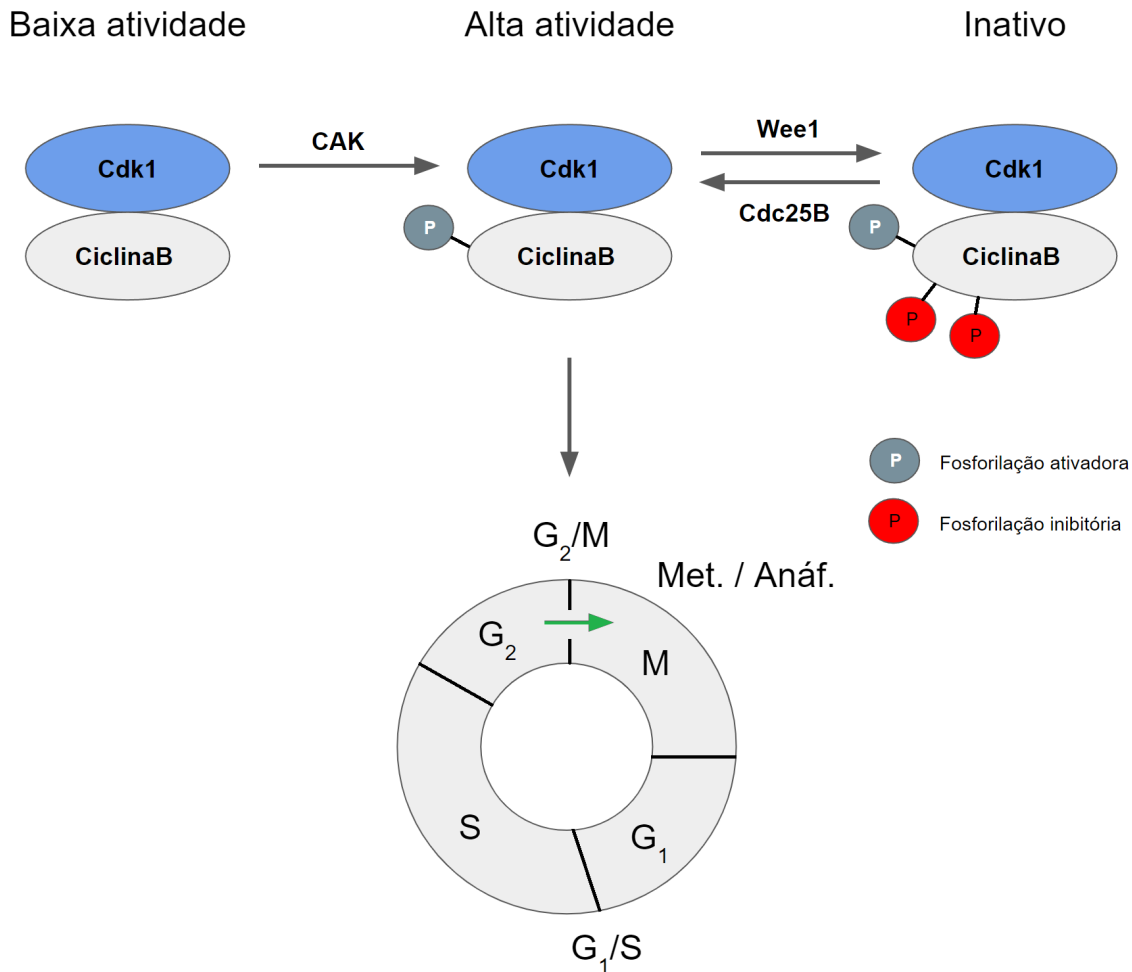


Figura 1. Esquema simplificado da atuação do complexo Cdk1-CiclinaB na progressão do ciclo celular (no *checkpoint* G₂/ M), mostrando o papel da Cdc25B na ativação plena deste complexo proteico.

A participação em uma etapa tão importante do ciclo celular pelas Cdc25 colocou esta família proteica em destaque levando à publicação de diversos estudos ao longo das últimas décadas, seja por questões relacionadas à atuação destas proteínas [19], seus inibidores [20], pelo interesse em seu mecanismo catalítico [21, 22] ou estrutural [23, 24], levando a importantes descobertas na área e a observações de questões científicas atuais importantes. Esta família proteica, em especial um de seus membros, a Cdc25B, será melhor discutida à diante.

1.2 Estrutura de proteínas e interações cátion- π

O termo estrutura de proteínas se refere aos arranjos tridimensionais que dadas proteínas (polímeros biológicos nos quais as unidades monoméricas são aminoácidos) possuem. Apesar da questão já discutida sobre como a ideia de rigidez na estrutura de proteínas é equivocada, aqui irá se desconsiderar a flexibilidade proteica assumindo, para efeitos de simplificação, que a estrutura de uma proteína possa ser representada por um único modelo, definido com cada átomo constituinte do polipeptídeo na “posição média” que ele ocupa, ao longo do tempo, em condições fisiológicas.

Estruturas de proteínas são altamente dependentes da sequência dos aminoácidos que as compõem (sequência esta também chamada de estrutura proteica primária). Cada resíduo de aminoácido possui uma tendência intrínseca a formar certos arranjos especiais, as estruturas secundárias, baseado na sua natureza química (carregado positivamente, carregado negativamente, neutro e hidrofílico, hidrofóbico) e em sua estereoquímica [25]. Certamente as hélices- α e folhas- β são as estruturas secundárias mais comuns, mas diversas estruturas menos conhecidas estão presentes nas proteínas como hélices 3_{10} , hélices- π e voltas. Dentre estas ainda existem estruturas “de mão direita” e “de mão esquerda” e folhas- β podem se orientar de maneira paralela ou antiparalela, o que permite aumentar consideravelmente a possibilidade de organização no segundo nível de complexidade estrutural [8, 13].

Da mesma forma que ocorre para as estruturas secundárias, que são uma consequência direta da estrutura primária proteica, as estruturas terciárias são uma consequência da sucessão de estruturas secundárias pois estas irão interagir entre si formando um arranjo espacial mais complexo favorecido pela tendência de cada estrutura secundária, na ordem em que se encontra, se organizar com as demais presentes. Finalmente, estruturas quaternárias são definidas pela interação entre duas ou mais estruturas terciárias utilizando, essencialmente, os mesmos tipos de interações que as utilizadas na estruturação terciária [8].

Uma das questões mais interessantes na estruturação de proteínas é o fato de que as estruturas de ordem superior são formadas e mantidas por interações fracas, com valores de energia de ligação de, no máximo, poucas dezenas de $\text{kJ}\cdot\text{mol}^{-1}$ frente às centenas de $\text{kJ}\cdot\text{mol}^{-1}$ de energia das ligações covalentes.

Os livros didáticos trazem, normalmente, como exemplos das interações fracas responsáveis pela manutenção da estrutura proteica as ligações de hidrogênio, pontes salinas, interações do tipo dipolo-permanente - dipolo-permanente e efeito hidrofóbico, raramente abordando tipos diferentes de interações ou aprofundando estas [8, 13].

Uma interação fraca pouco mencionada mas surpreendentemente comum e com energia de interação superior à da maior parte das demais interações fracas é a cátion- π , que, como o nome sugere, ocorre entre um grupo químico carregado positivamente, catiônico, e um sistema π , ou seja, no caso de moléculas orgânicas típicas, um sistema de átomos com orbitais p disponíveis, não hibridizados, conectados e que compartilham elétrons [26]. Apesar de qualquer sistema orgânico conjugado poder atuar em interações cátion- π (no caso de moléculas como etileno, por exemplo), para todos os propósitos do presente texto serão considerados sistemas π apenas as cadeias laterais dos aminoácidos aromáticos: fenilalanina, tirosina e triptofano. A figura 2 apresenta alguns exemplos de interações cátion- π possíveis entre cadeias laterais de aminoácidos.

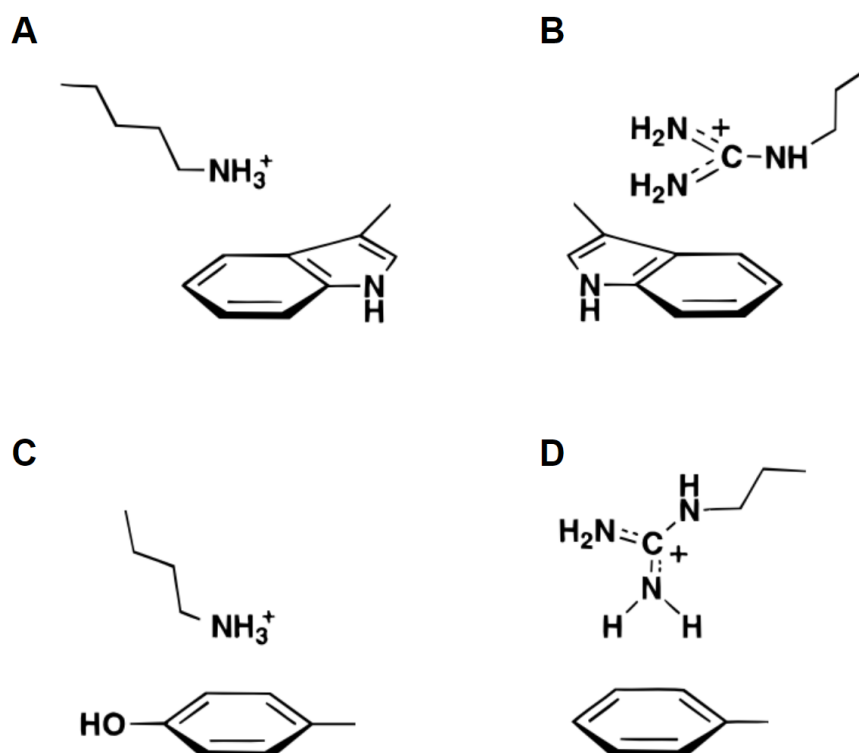


Figura 2. Representações de interações cátion- π entre cadeias laterais de: (A) lisina e triptofano, (B) arginina e triptofano, (C) lisina e tirosina e (D) arginina e fenilalanina.

Para efeito de comparação, alguns dos contatos mais fortes que podem ocorrer em proteínas são os entre pares iônicos no interior proteico hidrofóbico, onde não existe competição dos grupos carregados por solvatação. Pares iônicos podem alcançar valores de energia de interação de até $-80\text{kJ}\cdot\text{mol}^{-1}$ apesar de apresentarem, na maioria das vezes, energias de interação menos intensas; interações do tipo dipolo-permanente - dipolo-permanente, entre resíduos polares não carregados, tendem a apresentar até $-10\text{kJ}\cdot\text{mol}^{-1}$ de energia; forças de dispersão de London apresentam, normalmente, energia de até $-8\text{kJ}\cdot\text{mol}^{-1}$; as famosas ligações de hidrogênio apresentam energia variável, a depender dos átomos participantes e ângulo de interação, tipicamente com valores entre -12 e $-40\text{kJ}\cdot\text{mol}^{-1}$. As interações cátion- π apresentam valores de energia que normalmente vão de -9 a $-21\text{kJ}\cdot\text{mol}^{-1}$, superiores, em alguns casos, aos da maioria das interações anteriormente descritas, inclusive ligações de hidrogênio [8, 13, 27].

É importante considerar que em ambientes expostos ao solvente a dessolvatação de cátions irá gerar uma penalidade entrópica ao sistema, muitas vezes substancial e capaz de definir se uma determinada interação cátion- π será ou não favorecida energeticamente. Ainda assim, a energia de interação somada a outros efeitos (como polarizabilidade local, por exemplo) torna a formação de contatos cátion- π , transientes ou estáveis, não somente possível como importante, no contexto biológico, mesmo na superfície de proteínas [27].

Conforme já mencionado, interações cátion- π são bastante comuns. Uma análise com quase 600 estruturas de proteínas em alta-resolução depositadas no *Protein Data Bank* (PDB) [28] apontou uma interação cátion- π ocorrendo a cada 77 resíduos de aminoácidos, com um quarto dos resíduos de triptofano realizando pelo menos uma interação energeticamente significativa com argininas ou lisinas [26]. Outra análise estatística mostrou que cerca de metade dos complexos proteicos e um terço das estruturas homodiméricas depositadas possuem pelo menos uma interação cátion- π importante para a manutenção estrutural [29].

Descrições energéticas completas da interação cátion- π devem levar grande quantidade de fatores em conta: momento de quadrupolo, momento de dipolo, efeitos de transferência de carga, forças de dispersão e efeito hidrofóbico [30]. Para interações com sistemas aromáticos típicos, no entanto, o principal componente que define a energia de interação é o eletrostático [31] entre a espécie catiônica e o sistema π do anel, com densidade negativa de carga (figura 3).

1.3 Dinâmica conformacional de proteínas e proteínas desordenadas ou com regiões desordenadas

A evolução no entendimento estrutural e da dinâmica conformacional das proteínas está diretamente ligada à evolução na compreensão das funções destas macromoléculas biológicas nos organismos. Se no passado prevalecia a imagem estática de proteínas atuando nas mais diversas tarefas celulares hoje existe a clara ideia de que proteínas em solução são macromoléculas flexíveis com vibrações de ligações, torções e reorientação de cadeias laterais de aminoácidos, pequenos segmentos de estruturas secundárias e até mesmo grandes domínios [35].

Se para uma proteína imaginária, perfeitamente imóvel em seu estado energético mais estável, um único modelo poderia ser considerado uma representação estrutural adequada, à medida que outra proteína se apresenta flexível ou apresenta regiões flexíveis, um conjunto de estruturas se torna necessário para representá-la. Aqui, técnicas de simulação computacional de proteínas se mostram indispensáveis em atribuir conjuntos de estruturas prováveis para proteínas flexíveis.

Indo ainda além, técnicas de resolução estrutural de macromoléculas que necessitam de condições extremas, distantes da realidade de uma proteína em seu meio, como a cristalografia de raios-X, podem, em alguns casos, produzir estruturas que refletem pouco da realidade estrutural de partes de uma proteína em solução, como em certos domínios flexíveis, levando a equívocos no entendimento de trechos da estrutura protéica por causa de vieses introduzidos pela própria natureza da técnica experimental. Novamente a simulação computacional de biomoléculas representa uma ferramenta importante que aqui pode ser utilizada em conjunto com outras técnicas buscando sustentar observações ou sugerir investigações mais profundas.

Apesar de estudadas desde o século XIX, os primeiros estudos de proteínas com estrutura flexível ou desordenada só seriam publicados em 1952, com a descoberta da caseína [36]. As primeiras referências à estruturação protéica intrinsecamente desordenada viriam somente cerca de 15 anos depois [37] destas proteínas serem mencionadas em publicações científicas e somente no final da década de 1990 as proteínas intrinsecamente desordenadas (PIDs) ou com regiões intrinsecamente desordenadas (PRIDs) ganhariam destaque no meio acadêmico.

Os motivos que levam uma proteína, ou um segmento dentro de uma proteína, a ser flexível são diversos mas normalmente observa-se sequências de aminoácidos com baixa complexidade e pequena proporção de aminoácidos hidrofóbicos em relação aos aminoácidos hidrofílicos ou carregados [7]. O termo "intrinsecamente", para tratar destas regiões desordenadas, aqui utilizado como sinônimo de flexível, como em "proteínas intrinsecamente desordenadas", a denominação mais popular, vem do fato de que a sequência primária da região desordenada a torna assim, flexível, por uma questão intrínseca a sua natureza.

PIDs e PRIDs são bastante comuns em organismos eucarióticos e estão presentes em processos de regulação de vias de sinalização, transcrição, tradução, ciclo celular, entre outros [7]. A ausência de uma estrutura rígida nas PIDs e PRIDs permite que essas proteínas ou regiões proteicas interajam de maneira promíscua com diversos parceiros e em diferentes ocasiões permitindo, ainda, fácil dissociação da interação com a terminação do processo até então sustentado [9]. A promiscuidade em termos de parceiros de interação ainda permite às proteínas com estas características atuarem em *hubs* de sinalização e sua preferência por determinados parceiros dentro das vias em que se encontra pode ser definida por modificações pós-traducionais [38, 39].

Assim, a presença de regiões desordenadas em proteínas permite uma camada adicional regulatória nos processos celulares, colocando certas proteínas no centro de grandes *hubs* de sinalização e possibilitando o reconhecimento e interação com diferentes parceiros em diferentes condições, uma característica essencial em organismos superiores, mais complexos. O estudo de PIDs e PRIDs necessita de ferramentas adicionais e se aproveita grandemente da simulação computacional, sendo essencial ao pleno entendimento dos processos celulares e sua regulação.

1.4 A Cdc25B

É uma tarefa difícil descobrir as primeiras referências sobre a Cdc25B na literatura. É certo que a proteína foi descoberta em estudos do organismo modelo eucariótico, *Schizosaccharomyces pombe*, também conhecida como a "levedura de fissão", próximo ao ano de 1970 [40, 41, 42]. Tal afirmação, no entanto, deve se

aproximar de uma verdade para a maioria dos genes e proteínas de organismos eucariotos.

Descrições sobre a atuação de proteínas da superfamília da Cdc no controle do tamanho celular e da divisão nuclear, ainda exclusivamente em *S. pombe*, surgiram na metade da década de 70 [41, 42], de onde é derivado o nome da família proteica Cdc - *Cell Division Cycle* - Ciclo de Divisão Celular. Uma série de trabalhos complexos com proteínas mutantes que apresentam atividade sensível a pequenas variações de temperatura foi responsável pela elegante elucidação da função de diversas proteínas da família das Cdc. Antes dos anos 1980 já era bem estabelecida a importância das Cdc no controle do ciclo celular em leveduras, inclusive da Cdc25 [43], na época chamada Cdc25+, a qual demonstrou um papel central como indutor mitótico em meados da década seguinte, de onde deriva outra denominação desta proteína - *M-phase Inducer* - Indutora de fase "M" (fase de mitose do ciclo celular) [44].

No caso de humanos, as primeiras proteínas Cdc25, as Cdc25A e Cdc25B, foram isoladas em 1991 [45], já sendo, à época, conhecido um terceiro membro da mesma família, a Cdc25C. No mesmo trabalho em que estas proteínas foram isoladas pela primeira vez a atividade de tirosina fosfatase e atuação conjunta com ciclinas foi elucidada e surgiram evidências que suportavam a ação análoga à das proteínas homólogas de *S. pombe*, verificando-se a capacidade de interrupção do ciclo celular de células com alto potencial de proliferação pela microinjeção de anticorpos anti-Cdc25A.

Ainda em 1991, as primeiras associações entre a superexpressão de genes da família Cdc25 e alguns tipos de câncer começaram a surgir [46, 47] um tema que se tornaria recorrente durante os anos seguintes e colocaria esta família de proteínas em evidência como potenciais alvos terapêuticos para drogas anti-tumorais por um longo período de tempo [48].

A Cdc25A teve a estrutura cristalográfica de seu domínio catalítico resolvida no ano de 1998 [49] (a estrutura encontra-se atualmente depositada no PDB sob o código 1C25). No ano seguinte o domínio catalítico da Cdc25B também foi resolvido por cristalografia de raios-X [23] (PDB: 1QB0) tendo esta proteína uma quantidade significativa de depósitos adicionais nos anos seguintes. A Cdc25C não possui estrutura depositada no PDB no presente momento.

Ainda existe discussão na literatura sobre o papel exato de cada isoforma e se as Cdc25 A, B e C apresentam capacidade de atuar umas nas funções típicas das outras [50]. Indiscutível é que as três proteínas Cdc25 (A, B e C) são consideravelmente semelhantes em seus domínios catalíticos apresentando maiores diferenças em seu domínio regulatório. As Cdc25 também são fosfatases duais, capazes de remover grupos fosfato de resíduos de tirosina ou serina/treonina. Tal característica é devida a sua morfologia característica com um “bolsão” pouco profundo no sítio catalítico que não permite discriminar os grupos de p-serina/p-treonina dos de p-tirosina, maiores [51].

O domínio catalítico de fosfatases duais, também conhecido como *P-loop*, é conservado, seguindo a sequência consenso His-Cys-(X)₅-Arg-Ser, com a cisteína catalítica, X representando resíduos de aminoácidos quaisquer e a serina atuando normalmente em conjunto com um resíduo de aspartato próximo na remoção do fosfato que se liga à cisteína durante sua remoção [51].

À princípio, a Cdc25A atua na transição de fase G₁/S do ciclo celular, mas também pode apresentar papel regulatório em momentos finais do ciclo celular. A Cdc25A regula dois complexos de quinases dependentes de ciclina, Cdk2 - CiclinaE e Cdk2 - CiclinaA. Existem evidências de que a Cdc25A seja capaz de ativar a principal Cdk mitótica, a Cdk1 (também conhecida como Cdc2), no complexo Cdk1 - CiclinaB, promovendo a transição G₂/M, sendo, no entanto, esta transição de principal responsabilidade das Cdc25 B e C [19, 52, 53].

Diversos estudos discutem a localização celular das Cdc25 ao longo do processo de mitose, assim como a translocação entre os compartimentos celulares por parte destas enzimas. Tanto a Cdc25A quanto a Cdc25B, por exemplo, são encontradas no núcleo e no citoplasma em diferentes fases do ciclo celular e, aparentemente, sinais (sequências) de exportação ligadas ao domínio regulatório das proteínas são os responsáveis pela definição da localização destas durante sua atividade, explicando o motivo das diferenças consideráveis entre os domínios regulatórios das três Cdc25 [54, 55].

O gene da Cdc25B se localiza no braço curto do cromossomo 20, na região p13 e são conhecidas pelo menos quatro isoformas desta proteína (B1, B2, B3 e B4). A expressão diferencial de isoformas específicas da Cdc25B já foi relacionada com agressividade de tumores e também existem evidências de funções não

relacionadas ao controle do ciclo celular em células germinativas [56]. A sequência completa da Cdc25B é apresentada na figura 4:

 DR	10	20	30	40	50
	GSALSPARVL	GGIQRPRHLS	VFEFESDGFL	GSPEPTASSS	PVTTLTQTMH
	60	70	80	90	100
	NLAGLGSEPP	KAQVGSLSFQ	NRLADLSLSR	RTSECSLSSE	SSESSDAGLC
	110	120	130	140	150
	MDSPSPVDPQ	MAERTFEQAI	QAASRVIQNE	QFTIKRFRSL	PVRLLEHSPV
	160	170	180	190	200
	LQSITNSRAL	DSWRKTEAGY	RAAANSPGED	KENDGYIFKM	PQELPHSSSA
	210	220	230	240	250
	QALAEWVSRR	QAFTQRPSSA	PDLMCLTTEW	KMEVEELSPV	AQSSSLTPVE
	260	270	280	290	300
	RASEEDDGFV	DILESDLKDD	EKVPAGMENL	ISAPLVKKLD	KEEEQDLIMF
	310	320	330	340	350
	SKCQRLFRSP	SMPCSVIRPI	LKRLERPQDR	DVPVQSKRRK	SVTPLEEQQL
	360	370	DR DC	380	390
	400	410	420	430	440
	EEPKARVFRS	KSLCHEIENI	LDSHRGLIG	DYSKAFLLQT	VDGKHQDLKY
	450	460	470	480	490
	ISPETMVALI	TGKFSNIVEK	FVIVDCRYPY	EYEGGH IKNA	VNLPL ERDAE
	500	510	520	530	540
	TFLI QRPIMP	CSLDK R IILI	FHCEFS ERG	PRMCR FIRER	DRAAN DYPSL
	550	560	570	580	590
	YYPE MYIL KG	GYKE FFQHP	NFCEP QDYRP	MNH EAFR DEL	RNFRLK TRSW
	600	610	620	630	640
	AGERSR RELC	SRLQDQ			

DR - Domínio regulatório

DC - Domínio catalítico

Hélice α

Folha β

P-loop

Região desordenada

Figura 4. Sequência primária da Cdc25B. São indicados o domínio regulatório (DR), o domínio catalítico (DR) e algumas estruturas e regiões importantes da proteína conforme a legenda apresentada.

A isoforma Cdc25B1, que aqui será referida apenas como Cdc25B (sua denominação anterior à descoberta das demais isoformas desta proteína), possui 566 resíduos de aminoácidos sendo os primeiros 372 primeiros definidos como do domínio regulatório e os últimos 194 do domínio catalítico. A proteína completa apresenta peso molecular de aproximadamente 62,3kDa com o domínio regulatório possuindo cerca de 40,9kDa e o domínio catalítico 21,3kDa.

Em termos estruturais, o domínio catalítico da Cdc25B se apresenta como o de uma proteína globular clássica, com o motivo típico α/β , com grande equilíbrio na quantidade de hélices- α e folhas- β . A sequência, formada pelos 194 resíduos de aminoácidos contém 13 resíduos de lisina e 17 resíduos de arginina, totalizando 30 resíduos de aminoácidos carregados positivamente, um valor bastante superior ao esperado para uma distribuição aleatória (que apresentaria cerca de 19 resíduos). O ponto isoelétrico teórico calculado pela ferramenta *get pI/ MW* do portal de bioinformática *ExPASy* [57] é de 8,22, o que classifica o domínio catalítico da Cdc25B como básico.

A figura 5 apresenta a estrutura cristalográfica 1QB0 do domínio catalítico da Cdc25B. Na figura é dado destaque às hélices- α representadas na forma de cilindros e nomeadas conforme o uso dado ao longo do presente texto. A disposição da proteína será mantida aqui, com a representação do amino-terminal (N-terminal) à esquerda e do carboxi-terminal (C-terminal) à direita. As estruturas secundárias também são, posteriormente, posicionadas de maneira análoga à da figura 5, para permitir a identificação, exceto quando mencionado. Especialmente importantes são a hélice D, onde se encontram seis resíduos de arginina (com destaque às R479 e R482) e um sítio de ligação à fosfato, e a hélice F, onde se localiza outra importante arginina (R544), em posição anterior ao único triptofano presente no polipeptídeo. A hélice F também precede o C-terminal flexível da proteína, não representado na imagem devido à baixa aquisição de sinal na região pela técnica de cristalografia de raios-X, justamente pela natureza desordenada da mesma. Esta é a região da proteína onde se localiza o resíduo de triptofano que atua em importantes contatos cátion- π com pelo menos duas argininas, R482 e R544 [58].

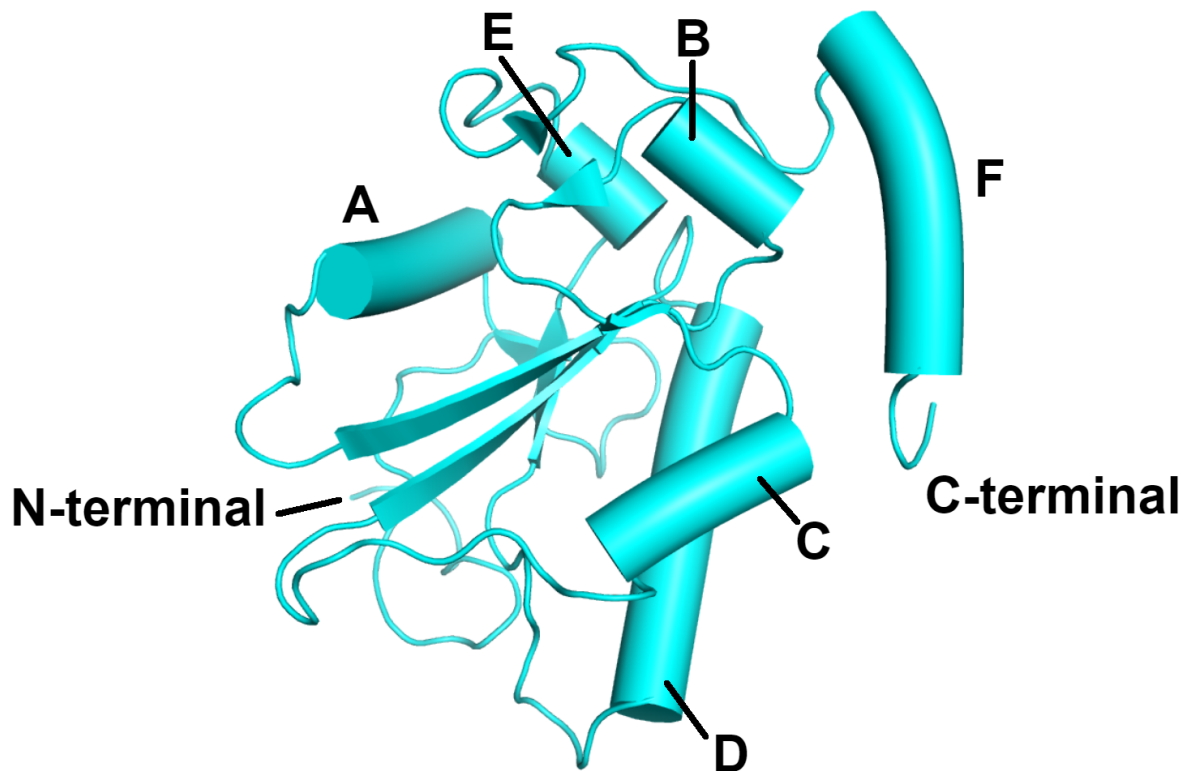


Figura 5. Estrutura tridimensional da Cdc25B representada em forma de *cartoon* gerado a partir da estrutura cristalográfica depositada no PDB sob código 1QB0. Hélices são apresentadas como cilindros e folhas como setas, voltas e regiões não estruturadas são apresentadas como tubos. É dado destaque ao amino-terminal, carboxi-terminal e à nomenclatura das hélices da proteína.

Já a figura 6 apresenta o domínio catalítico da Cdc25B, conforme a estrutura cristalográfica, após adição dos 16 resíduos não identificados originalmente pela cristalografia e simulação curta para relaxamento do sistema e estruturação do C-terminal. Agora é dado enfoque às 17 argininas que compõem o polipeptídeo e ao triptofano (átomos de C representados em vermelho e de N em azul), ao sítio catalítico enzimático (*P-loop*, em amarelo) e à região C-terminal desordenada (em verde). Hélices- α são mostradas na representação usual de *cartoon* para permitir a visualização de resíduos localizados na parte posterior da proteína. O modelo apresentado à esquerda se encontra na representação definida previamente (N-terminal à esquerda e C-terminal à direita) enquanto o modelo à direita é obtido após um giro de 180° em relação a um eixo imaginário perpendicular ao eixo N-terminal - C-terminal da proteína.

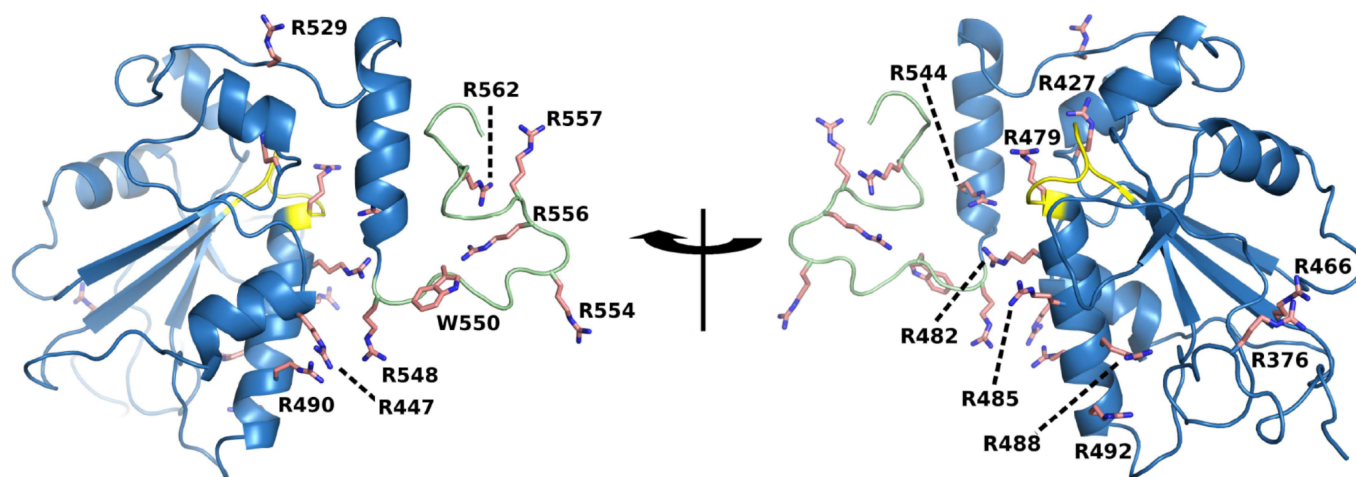


Figura 6. Representação em *cartoon* da estrutura cristalográfica da Cdc25B. É dado destaque para o *P-loop* catalítico (em amarelo), para o C-terminal desordenado (em verde), e para as diversas argininas e o único triptofano presentes na estrutura proteica (em vermelho e azul).

Assim, a Cdc25B é uma proteína essencial na regulação do ciclo celular que apresenta uma característica interessante do ponto de vista estrutural: uma região desordenada com um resíduo de triptofano capaz de interagir com resíduos de arginina distribuídos superficialmente na estrutura do peptídeo.

Estudos da Cdc25B podem lançar luz a questões estruturais desconhecidas como na importância de uma dinâmica de contatos transientes cátion- π em regiões desordenadas de proteínas além de indicar como se dá a complexação de inibidores desta proteína, considerada um alvo terapêutico promissor para diversos tipo de cânceres.

1.5 Dinâmica molecular e potenciais clássicos

Dinâmica molecular computacional é uma técnica de simulação que descreve a trajetória das partículas que compõem um sistema em estudo ao longo do tempo [59]. Aqui é importante destacar que o modelo assumido para as simulações de dinâmica molecular considera os átomos do sistema como esferas rígidas que podem se encontrar neutras ou eletricamente carregadas. O método consiste em integrar a equação de Newton (equação 1) a partir de potenciais ou funções de energia que descrevam as interações entre cada par de partículas presente no sistema modelado (ϵ_{mm} , equação 2). As forças são usadas na integração da equação de Newton (equação 1) para que sejam geradas novas posições e velocidades para os átomos do sistema em curtos períodos, re-calculando-se os

parâmetros do sistema a cada intervalo de tempo definido. Para sistemas compostos por biomoléculas, por exemplo, o intervalo típico utilizado tem duração ao redor de 1 femtossegundo, ordem de grandeza do tempo de estiramento de uma ligação C - H, a alteração mais rápida observada no sistema simulado. Na equação 1 tem-se “m” a massa e “r” a coordenada espacial do iésimo átomo, com “t” o tempo. A integração da função gera novas posições (R) e assim uma trajetória consecutiva pode ser obtida.

Equação 1
$$F_i = m_i \frac{\partial^2 r_i}{\partial t^2}$$

Equação 2
$$F_i = - \frac{\partial \varepsilon_{mm}}{\partial r}$$

A ε_{mm} , acima mencionada (equação 2) é composta pela soma das energias referentes às ligações covalentes (ε_{cov}) e às interações não-covalentes (ε_{ncov}) presentes no sistema.

Equação 3
$$\varepsilon_{mm} = \varepsilon_{cov} + \varepsilon_{ncov}$$

O termo covalente da equação acima mantém a estrutura interna de cada molécula e é composto pela soma das energias de estiramento das ligações (ε_{lig}), ângulos de ligação (ε_{ang}) e ângulos diedrais próprios e impróprios formados entre os átomos covalentemente ligados (ε_{died}). Ângulos diedrais próprios são definidos a partir de quatro átomos ligados sequencialmente, como o que ocorre ao longo dos átomos H - C - C - H de uma molécula de etano, por exemplo. Ângulos diedrais impróprios são definidos a partir de quatro átomos ligados não sequencialmente, como o que ocorre entre os átomos de H - N(H) - H, na molécula de amônia, por exemplo.

O termo não covalente é composto pela soma das interações eletrostáticas (ε_{elet}) e pela função de Lennard-Jones (ε_{LJ}), que fornece um valor aproximado para a contribuição das forças de van der Waals.

A energia de ligação (ε_{lig} - equação 4) e a energia dos ângulos de ligação (ε_{ang} - equação 5) são definidas tratando-se estas entidades químicas como molas, de

forma que o potencial se altera com o quadrado do deslocamento da distância “R” ou do ângulo “ Θ ” em relação aos valores de equilíbrio definido (“ R_0 ” e “ Θ_0 ”). Os parâmetros “ k_m ” e “ k_e ” são constantes de força, análogos às constantes de molas observadas em cálculos de energia potencial elástica da física clássica.

Equação 4
$$\epsilon_{lig}(R) = \sum_{lig} \frac{1}{2} k_m (R - R_0)^2$$

Equação 5
$$\epsilon_{ang}(\Theta) = \sum_{ang} \frac{1}{2} k_e (\Theta - \Theta_0)^2$$

Já a contribuição da energia para torção dos ângulos diedrais (ϵ_{died} - equação 6) é descrita de por uma função periódica em que “n” é a periodicidade do ângulo diedral, “ K_f ” a constante de força e “ δ ” é a fase.

Equação 6
$$\epsilon_{died}(\Theta) = \sum_{died} \sum_{i=1}^4 \frac{1}{2} k_f ([1 + \cos(n\Phi - \delta)])$$

O termo não-covalente (equação 3) engloba as interações entre átomos de diferentes moléculas ou entre átomos da mesma molécula que estão distantes, separados por três ou mais ligações covalentes. Neste termo, estão incluídas contribuições eletrostáticas (ϵ_{elet} - equação 7), descritas pela lei de Coulomb, enquanto a função de energia de Lennard-Jones (ϵ_{LJ} - equação 8) é utilizada para se obter o valor aproximado das forças de van der Waals descritas no sistema modelado. Abaixo “i” e “j” representam pares de átomos com cargas parciais “q” localizadas a uma distância “ R_{ij} ”. “ e_{ij} ” é a profundidade do potencial de Lennard-Jones e “ σ_{ij} ” a posição no fundo do potencial.

Equação 7
$$\epsilon_{elet}(R) = \sum_{i < j} \frac{q_i q_j}{R_{ij}}$$

Equação 8
$$\epsilon_{LJ}(R) = \sum_{i < j} e_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}} \right)^6 \right]$$

Após 45 anos do primeiro estudo lançado na área [60] o uso de simulações de dinâmica molecular vem sendo cada vez mais utilizado na pesquisa científica, assim como técnicas computacionais diversas. Dentre as áreas de pesquisa que se aproveitam da química computacional destacam-se o estudo de estrutura eletrônica [61, 62], estabilidade e estrutura de macromoléculas [63], ligação de pequenas moléculas a proteínas [64, 65], entre outras.

Dessa forma, a simulação computacional de biomoléculas se mostra uma importante e poderosa ferramenta para o estudo da Cdc25B, suas características estruturais e a dinâmica de interação com pequenos ligantes.

Esta tese foi organizada de forma que o capítulo seguinte corresponde ao desenvolvimento do primeiro objetivo apresentado na sequência desta introdução, abordando a possibilidade de contatos transientes cátion- π na superfície do domínio catalítico da Cdc25B. O último capítulo aborda o segundo objetivo, avaliando mecanismos de inibição e reconhecimento molecular de inibidores da Cdc25B. Em seguida são apresentadas as conclusões deste trabalho, as referências, e por fim, os anexos.

1.6 Objetivos

Avaliar a estabilidade de contatos transientes entre resíduos localizados no domínio enovelado e na região desordenada da proteína Cdc25B por medidas de ressonância magnética nuclear (^1H - ^{15}N HSQC RMN) e simulações de dinâmica molecular.

Determinar mecanismos, sítios de inibição e resíduos importantes na complexação da Cdc25B com inibidores previamente testados a partir de simulações de ancoragem e dinâmica molecular.

Identificação de contatos cátion- π transientes na superfície do domínio catalítico da Cdc25B

2.1. Metodologia e desenvolvimento

Os cálculos quânticos, simulações de dinâmica molecular e geração dos perfis de energia livre relatados nas duas próximas subseções foram realizados pelo professor Dr. Guilherme Menegon Arantes anteriormente ao início do projeto de doutorado aqui descrito. A metodologia utilizada e os resultados serão brevemente apresentados devido à importância dos mesmos, responsáveis pelo direcionamento da pesquisa desenvolvida.

2.1.1. Cálculos quânticos, simulações de dinâmica molecular e geração de perfis de energia livre

Sistemas-modelo compostos por argininas bloqueadas (bArg) e triptofanos bloqueados (bTrp), isolados e em fase gasosa, foram utilizados para avaliar a descrição das interações cátion- π , entre as cadeias laterais destes dois resíduos de aminoácidos modificados, por campos de força. As energias de interação entre ambos resíduos, estruturados em duas orientações diferentes e relevantes para o presente estudo (nomeadas “sanduíche” e “T” ou “perpendicular”), foram medidas a partir de cálculos quânticos realizados com uso do programa GAUSSIAN 09 rev. A [66] pelo método/ base MP2/6-311++G** [67] corrigido para superposição de bases, e comparadas os valores fornecidos em cálculos realizados para os mesmos sistemas com os campos de força CHARMM36 [68] e AMBER99-ILDN [69, 70].

Simulações de dinâmica molecular do domínio catalítico completo da Cdc25B foram realizadas a partir da estrutura depositada no *Protein Data Bank* (PDB) [30] sob o código 1QB0, após a adição manual de 16 resíduos da porção C-terminal (A551 - Q566), originalmente ausentes, e 7 resíduos na porção N-terminal (“G366” a “Q372”), não naturais da proteína mas presentes na construção utilizada experimentalmente. Moléculas de água cristalográficas e de 2-mercaptoetanol (BMER) foram removidas, átomos de hidrogênio foram adicionados e os estados de protonação das cadeias laterais dos resíduos de aminoácidos foram definidos conforme o esperado para pH = 7. Dois íons fosfato inorgânico (HPO_4^{2-}) foram adicionados manualmente ao sistema, um ao sítio catalítico (*P-loop*, ligado à R479)

e outro ao sítio de ligação de fosfato (ligado à R488), o qual se localiza entre as R488 e R492, na hélice D do polipeptídeo. O modelo foi solvatado em uma caixa de simulação dodecaédrica com 13234 moléculas de água, 4 íons Na^+ e 2 íons Cl^- para mimetizar as condições experimentais de aquisição de espectros de RMN, relatadas à frente.

O sistema foi relaxado e equilibrado durante simulações livres de dinâmica molecular com 100ns de duração. Foi utilizado o campo de força CHARMM36 [68] para a proteína e íons presentes e o modelo TIP3P para as moléculas de água [71]. As simulações foram realizadas utilizando-se passo de integração de 2fs, 300K de temperatura e pressão de 1bar (*ensemble* NPT), termostato de Bussi [72] e barostato de Parrinello-Rahman [73], ambos com período de 0,5ps. Para o método de PME foi utilizado *cutoff* de espaço real de 1,2nm e espaçamento de Fourier de 0,12nm para interações eletrostáticas de longo alcance. Todos os cálculos com campos de força clássicos, simulações de dinâmica molecular e análises foram realizados com o programa GROMACS 4.6.7 [74].

A escolha dos resíduos candidatos para a análise preliminar da capacidade de participação em contatos cátion- π com o resíduo de triptofano da proteína foi feita excluindo-se as lisinas (que possuem cadeias laterais relativamente curtas) e selecionando-se, dentre as 17 argininas presentes no polipeptídeo, aquelas cujo centro de massas (CDM) do grupo guanidino visitasse uma distância menor que 8 Å em relação ao CDM do anel indólico do triptofano ao longo de uma simulação de dinâmica molecular longa (6 μ s). Resíduos enterrados ou que participam de contatos com grupos fosfato no sistema montado (observação esta feita a partir de análises de ensaios de titulação com fosfato realizadas em trabalhos anteriores do grupo de pesquisa do professor Dr. Guilherme Menegon Arantes) também foram descartados.

Perfis de energia livre para a interação entre as cadeias laterais de argininas e do triptofano foram estimados pela técnica de *Umbrella Sampling* (US) [75]. As configurações iniciais para cada janela de US foram obtidas após a equilibração dos modelos gerados. Sete janelas de US com coordenadas de reação de referência separadas por 0,1nm, variando de 0,4 a 1,0nm, foram utilizadas para estimar os perfis para cada par Triptofano - Arginina. Foi utilizado um potencial harmônico com constante de força de 2000kJ.mol⁻¹nm⁻².

Cada janela de US foi amostrada por 500ns com uma amostra coletada a cada picosegundo. Potenciais de força média foram obtidos a partir da técnica de

WHAM [75] e a incerteza estatística foi estimada dentro de um intervalo de 95% de confiança por análise pela técnica de *Bootstrap* [76] com 50 passos de reamostragem. Os 50ns iniciais de cada janela foram descartados para permitir equilíbrio dos graus de liberdade ortogonais e a sobreposição da distribuição das coordenadas de reação entre janelas adjacentes foi confirmada.

2.1.2. Amplificação de vetores de expressão do domínio catalítico da Cdc25B, análise de integridade e geração de mutantes pontuais

Vetores pET-28a (+) (*AddGene* - 2565), disponíveis no laboratório do professor Dr. Sandro Roberto Marana e que continham a sequência relativa ao domínio catalítico da Cdc25B, foram amplificados para possibilitar a análise da integridade da sequência original da enzima e dispor de material para a posterior geração de mutantes desta. Esses vetores foram originalmente produzidos pelo Dr. Raphael SantaRosa Sayegh durante o desenvolvimento de seu projeto de doutorado, a partir da clonagem da sequência do polipeptídeo (excluídos os 17 primeiros resíduos da porção N-terminal), inicialmente presente em vetores pUC18 (*OriGene* - SDC126594), no sítio múltiplo de clonagem de vetores pET-28a (+) adquiridos comercialmente.

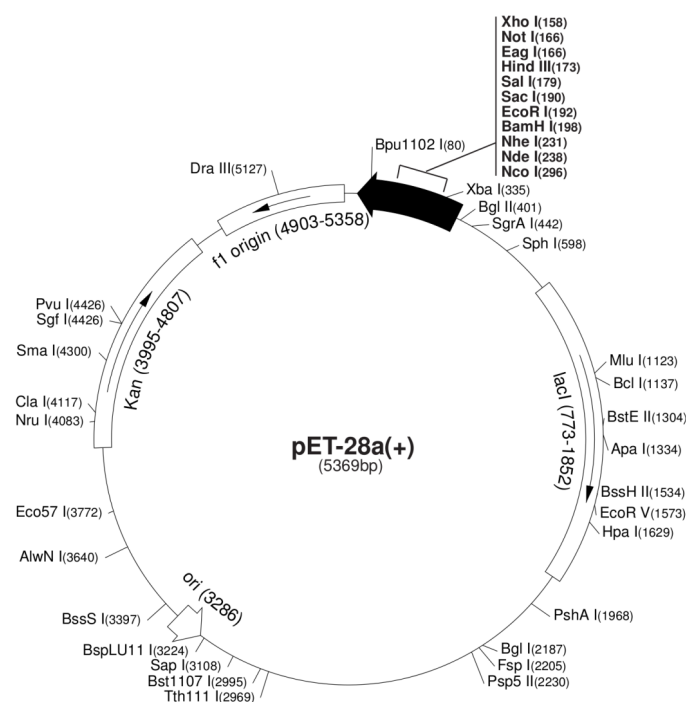


Figura 7. Mapa do plasmídeo pET-28a(+), utilizado como vetor para expressão heteróloga da Cdc25B selvagem e síntese/ expressão dos mutantes pontuais produzidos. A sequência relativa ao domínio catalítico da proteína selvagem foi inserida fusionada a sequências de cauda de histidina e de sítio de clivagem de trombina no sítio múltiplo de clonagem do vetor.

Essa construção inicial presente no vetor, será aqui referida como C1-17N e possui um gene de resistência à canamicina, para seleção de bactérias transformadas, e o trecho relativo aos resíduos de aminoácidos S373 a Q566 do domínio catalítico da Cdc25B selvagem fusionado, em sua extremidade N-terminal, a uma cauda de histidina/ *his-tag* (MGSSHHHHHSSG) para posterior purificação por cromatografia de afinidade em resina de agarose com níquel imobilizado. Entre a cauda de histidina e a sequência da enzima há um sítio de clivagem por trombina (LVPRGS) para a remoção da *his-tag* após a etapa inicial de purificação da proteína recombinante. A sequência da região expressa pode ser verificada na figura 8 com destaque a regiões importantes da enzima, referidas na legenda.

373

MGSSHHHHHSSG **LVPRGS**HMEFQSDHRELIGDYSKAFLQLQTVDGKHQDLKYISPETMVALL

TGKFSNIIVDKFVIIVDCRYPYEYEGGHIKTAVNLPLERDAESFLLKSPIAPCSLDKRVILIFH

473

482

CEFSSE **R**GPRMCRFIRERDRAVNDYPSLYPEMYILKGGYKEFFPQHPNFCEPQDYRPMNHE

544

550

566

AFKDELKTRFLKTRSWAGERSRRELCSRLQDQ **Stop**

MGSSHHHHHSSG - Cauda de Histidina.

LVPRGS - Sítio de clivagem de Trombina.

GSHMEFQ - Segmento extra pós-clivagem.

FHCEFSSE - *P-Loop*.

Figura 8. Sequência presente nos vetores pET-28a(+) C1-17N, produzidos pelo Dr. Raphael SantaRosa Sayegh e disponíveis no laboratório do professor Dr. Sandro Roberto Marana. São destacados a *his-tag*, o sítio de clivagem de trombina, o segmento extra pós-clivagem, o sítio catalítico enzimático (*P-loop*) e resíduos importantes para a geração dos mutantes (argininas 482 e 544 e triptofano 550).

Utilizando plasmídeos dos vetores pET-28a (+) com a construção C1-17N foram realizadas transformações de bactérias competentes *E. coli XL1-Blue (Agilent - 200249)* a partir da técnica de choque térmico sendo todo o procedimento realizado em ambiente estéril e utilizando-se de micropipetador e ponteiras autoclavadas. Inicialmente foram misturados 1µL de solução do plasmídeo

($1\text{ng}\cdot\mu\text{L}^{-1}$) a uma alíquota de $200\mu\text{L}$ de bactérias competentes *XL1-Blue*, a mistura foi mantida no gelo por 30 minutos e foi então realizado o choque térmico em banho-maria a 42°C por 40 segundos, seguido de banho de gelo por mais 3 minutos. Foram, então, adicionados $500\mu\text{L}$ de meio de cultura líquido Luria-Bertani (LB/ *LB Broth*) autoclavado e sem antibióticos e a nova mistura foi incubada em estufa, a 37°C , por 1 hora, invertendo-se o tubo que a continha a cada 20 minutos.

O meio de cultura com as bactérias transformadas e recuperadas foi utilizado para semeadura/ plaqueamento de placas de Petri estéreis com LB-ágar e canamicina ($5\mu\text{g}\cdot\text{mL}^{-1}$), além de placas controle (sem antibiótico), e todas foram incubadas *overnight* para crescimento de colônias a serem selecionadas no dia seguinte. As colônias selecionadas foram utilizadas para o preparo de inóculos em tubos do tipo *Falcon* (50mL) contendo 5mL de meio líquido LB autoclavado, ao qual foi adicionado o antibiótico canamicina ($5\mu\text{g}\cdot\text{mL}^{-1}$), previamente esterilizado por filtração com filtros de porosidade $0.22\mu\text{m}$. Após incubação em *shaker*, a 37°C e 150rpm *overnight*, a densidade óptica da mistura em comprimento de onda 600nm (DO_{600}) foi medida para garantir a presença de uma população de células bacterianas em estado adequado ($\text{DO}_{600} < 2,4$) e estas foram coletadas por centrifugação a 4.000g e 4°C por 30 minutos. O sobrenadante foi descartado.

Proseguiu-se com lise das células de bactérias para isolamento e purificação do DNA plasmidial utilizando-se o kit comercial *Pure Yield™ Plasmid Miniprep System* (Promega - PRA6810) seguindo-se as orientações do fabricante. Os plasmídeos isolados foram submetidos a reação de preparo para sequenciamento com o kit comercial *Big Dye™ Direct Cycle Sequencing Kit* (ThermoFisher - 4458687) de acordo com as orientações do fabricante com exceção da substituição do tampão do kit original pelo tampão *Save Money*, prática usual nas reações de preparo para os sequenciamentos realizados na Central Analítica do Instituto de Química da Universidade de São Paulo (CA - IQUSP). Os iniciadores utilizados foram os das sequências complementares às das regiões T7 promotora (5' - TAA TAC GAC TCA CTA TAG GG - 3') e SP6 (5' - TAT TTA GGT GAC ACT ATA G - 3') presentes no vetor pET-28a (+), flanqueando o sítio múltiplo de clonagem.

O DNA sintetizado foi então submetido a sequenciamento fluorescente automatizado por sequenciador capilar pelo método de Sanger modificado, em sequenciador ABI 3000® (Applied Biosystems), disponível na CA - IQUSP.

Após o sequenciamento, os resultados fornecidos foram analisados e as sequências finais montadas (Anexo A) levando-se em conta as indicações de qualidade fornecidas pelos cromatogramas gerados, visualizados com o uso dos programas *ApE* [77] e *SnapGene Viewer* [78]. A sequência final obtida foi, então, traduzida com uso do programa *Expassy Translate Tool* [57] e comparada com a sequência esperada para a enzima selvagem com os trechos fusionados e a correspondência obtida foi perfeita (Anexo A).

Uma vez confirmada a identidade da sequência do domínio catalítico da proteína selvagem clonada prosseguiu-se com a construção dos mutantes pontuais da Cdc25B. As construções desenhadas para os ensaios contêm substituições dos resíduos de arginina 482 (R482) , arginina 544 (R544) ou triptofano 550 (W550) por alanina (respectivamente R482A, R544A e W550A). Os mutantes foram gerados a partir de reações de mutação sítio dirigidas em plasmídeos da C1-17N, isolados e purificados, utilizando-se o kit comercial *QuickChange Lightning Site-Directed Mutagenesis Kit* (Agilent Technologies - 210518) conforme as instruções do fabricante. Os iniciadores mutagênicos utilizados apresentam as seguintes sequências:

R482A - *Primer Forward*

5' - TGA GCG TGG GCC CGC CAT GTG CCG TTT C - 3'

R482A - *Primer Reverse*

5' - GAA ACG GCA CAT GGC GGG CCC ACG CTC A - 3'

R544A - *Primer Forward*

5' - TGA GCT AAA GAC CTT CGC CCT CAA GAC TCG CAG C - 3'

R544A - *Primer Reverse*

5' - GCT GCG AGT CTT GAG GGC GAA GGT CTT TAG CTC A - 3'

W550A - *Primer Forward*

5' - CAA GAC TCG CAG CGC GGC TGG GGA GCG G - 3'

W550A - *Primer Reverse*

5' - CCG CTC CCC AGC CGC GCT GCG AGT CTT G - 3'

As concentrações dos iniciadores adquiridos comercialmente foram ajustadas para $50\text{ng}\cdot\mu\text{L}^{-1}$ antes do preparo dos *mixes* para reação. Os termocicladores foram ajustados para o maior número de ciclos indicado nas instruções, haja visto a reconhecida baixa eficiência deste tipo de reação (de polimerase não do tipo “amplificação em cadeia”). O conteúdo de cada uma das amostras resultantes, as construções R482A-17N, R544A-17N e W550A-17N, foram totalmente utilizados para transformar bactérias *XL1-Blue* competentes, seguindo o mesmo protocolo descrito anteriormente para transformação das bactérias com o vetor contendo a sequência selvagem (construção C1-17N).

Uma vez transformadas com cada vetor mutado, as novas bactérias foram submetidas aos mesmos procedimentos de preparo de inóculo, isolamento, extração, purificação de DNA plasmidial e sequenciamento anteriormente descritos. Os resultados das análises dos sequenciamentos realizados indicaram êxito no preparo de todos os mutantes e as montagens podem ser conferidas no anexo A.

2.1.3. Expressão, extração e purificação da forma selvagem e dos mutantes do domínio catalítico da Cdc25B

Confirmadas as identidades da proteína selvagem e dos mutantes, seguiu-se para a expressão e purificação das quatro construções da Cdc25B. Os vetores isolados foram antes utilizados para transformação de bactérias *E. coli BL21(GOLD) - DE3*, bactérias super competentes capazes de expressar maiores quantidades de proteínas heterólogas e que não possuem proteases OmpT e Lon, respectivamente presentes na membrana celular e matriz mitocondrial de diversas cepas de bactérias *E. coli*, inclusive *XL1-Blue*. Tais proteases são capazes de degradar quantidades consideráveis de proteínas expressas, durante a lise celular bacteriana, sendo importante sua ausência. A técnica para transformação das bactérias da cepa *BL21(GOLD) - DE3* foi a de choque térmico e o protocolo utilizado foi o mesmo anteriormente descrito para a transformação das bactérias da cepa *XL1-Blue*.

Foram realizados pré-inóculos em meio líquido LB com antibiótico canamicina ($5\mu\text{g}\cdot\text{mL}^{-1}$) para as bactérias transformadas com os vetores de cada construção após a seleção de colônias em placas de Petri com LB-ágar e canamicina ($5\mu\text{g}\cdot\text{mL}^{-1}$), da mesma forma como mencionado na seção anterior. Nesse momento

foram retiradas alíquotas de 1mL do meio de cultura homogeneizado e estas foram separadas para análises posteriores em géis de poliacrilamida [79].

Soluções de CaCl_2 e MgSO_4 de concentração $1,0\text{mol.L}^{-1}$, D-glicose ($0,2\text{g.mL}^{-1}$), canamicina ($30\mu\text{g.}\mu\text{L}^{-1}$) e IPTG (200mmol.L^{-1}) foram preparadas, filtradas com filtros de porosidade $0,22\mu\text{m}$ e armazenadas em frascos esterilizados. Os pré-inóculos anteriormente produzidos foram centrifugados em 4.000g , a 25°C por 30 minutos e utilizados para o preparo de inóculos que foram crescidos em meio mínimo M9 marcado com ^{15}N ($6,0\text{g.L}^{-1}$ Na_2HPO_4 , $3,0\text{g.L}^{-1}$ KH_2PO_4 , $1,0\text{g.L}^{-1}$ $^{15}\text{NH}_4\text{Cl}$ e $0,5\text{g.L}^{-1}$ NaCl), preparados para volume final de 1,0L em frascos do tipo *Erlenmeyer*. Aos frascos com os inóculos adicionou-se 2mL da solução de MgSO_4 , $100\mu\text{L}$ da solução de CaCl_2 , 10mL da solução de glicose e 1mL da solução de canamicina (para uma concentração final de $30\mu\text{g.mL}^{-1}$).

Os inóculos foram incubados em *shaker* a 37°C e 150rpm por aproximadamente 6 horas, quando sua DO_{600} alcançou um valor próximo a 0,6 (fase de crescimento exponencial ou fase log). Nesse momento adicionou-se 2mL de solução estéril de isopropil- β -D-1-tiogalactopiranosida (IPTG) 200mmol.L^{-1} para que a concentração final no meio de cultura do inóculo fosse de $0,4\text{mmol.L}^{-1}$. Novamente os inóculos foram incubados, agora em *shaker* a 20°C e 150rpm por 18 horas e após remoção de uma alíquota de 1mL de cada inóculo, para análise posterior em gel, as bactérias foram coletadas por centrifugação em 4.000g , a 4°C por 30 minutos. Com as bactérias sedimentadas o sobrenadante foi descartado e os *pellets* não imediatamente processados foram mantidos em congelador a -20°C .

Foi realizada “corrida” de eletroforese em gel de poliacrilamida com dodecil sulfato de sódio (SDS-PAGE) [79] utilizando-se as alíquotas retiradas dos pré-inóculos e inóculos, respectivamente antes e depois da indução da expressão da proteína selvagem e dos mutantes gerados. Foi utilizada cuba *BioRad* (BioRad - 1704487) com aproximadamente 600mL de tampão de corrida, voltagem de 150V e amperagem entre 10 e 40mA (a depender do momento da corrida). O tempo total de corrida foi próximo de 65 minutos e todos os ensaios de SDS-PAGE aqui relatados seguiram o mesmo protocolo.

Terminada a corrida, o gel foi corado com solução de *Coomassie R* 0,5% (*PhastGel*®, Sigma-Aldrich - 4921) em H_2O *milli-Q*/ metanol (em proporção de três volumes para quatro) e ácido acético 20% (misturado em igual volume à solução

anterior) por aproximadamente 20 minutos. O gel foi posteriormente descorado em solução de ácido acético 10% e, em seguida, solução de etanol 40%. A imagem do gel foi capturada em fotodocumentador e esta pode ser conferida na figura 9. É possível observar que, nas amostras em que não havia sido realizada a indução da expressão das proteínas não aparecem bandas com quantidades apreciáveis de proteínas nas regiões relativas ao peso molecular da Cdc25B e dos mutantes gerados (aproximadamente 24kDa), enquanto nas amostras em que houve a indução da expressão das proteínas estão presentes bandas bastante visíveis nesta posição.

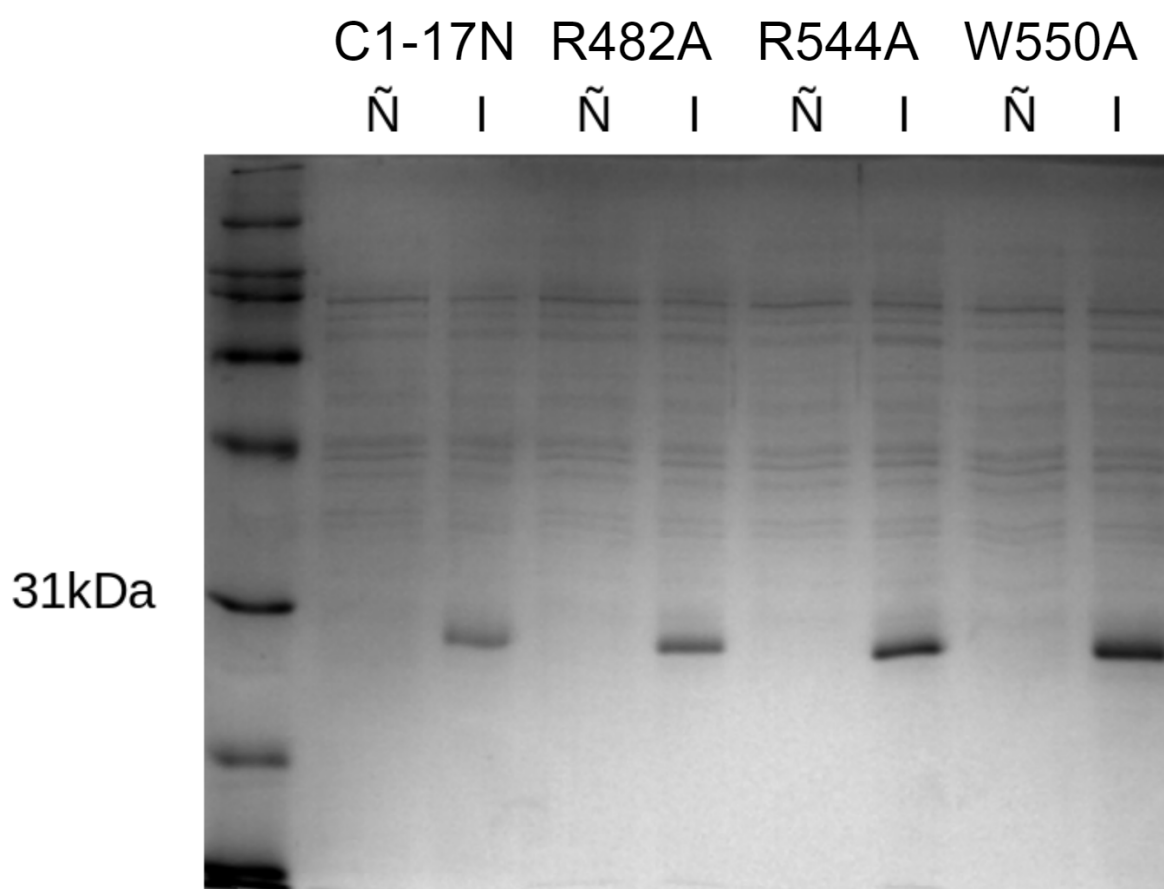


Figura 9. Gel de indução da C1-17N e mutantes gerados: O gel mostra as amostras não induzidas (Ñ) e induzidas (I) para cada um dos inóculos transformados com os plasmídeos das construções da Cdc25B e seus mutantes (C1-17N, R482A-17N, R544A-17N e W550A-17N). É possível verificar que a indução da expressão das proteínas (de massa aproximada de 24kDa com a cauda de histidina) ocorreu com sucesso em todos os casos.

Uma vez expressas as construções da forma selvagem e dos mutantes da Cdc25B seguiu-se com a purificação destas. Para romper as células das bactérias os *pellets* coletados foram ressuspensos em 5mL de tampão de lise (50mmol.L⁻¹ trisaminometano - HCL, 500mmol.L⁻¹ NaCl, 10% glicerol, 5mmol.L⁻¹

2-mercaptoetanol [BMER], 1mmol.L^{-1} fenilmetil sulfonil fluoreto [PMSF], $\text{pH} = 7,4$) e a mistura foi submetida a sonicação por 10 pulsos de 10 segundos, com potência de 75W e intervalo de 2 minutos entre os pulsos (durante os quais as amostras foram mantidas em gelo).

Os lisados de bactérias foram então centrifugados em $4.000g$, a 4°C por 30 minutos para a sedimentação de *debris* celulares e os sobrenadantes foram coletados e separados. Colunas cromatográficas de vidro foram manualmente empacotadas com 1mL de resina Ni-NTA agarose (Qiagen - 70666) e lavadas com 5mL de tampão de lavagem (50mmol.L^{-1} trisaminometano - HCL, 500mmol.L^{-1} NaCl, 20mmol.L^{-1} imidazol, 5mmol.L^{-1} BMER, $\text{pH} = 7,4$). Após completa eluição do tampão de lavagem os sobrenadantes (lisados) previamente coletados foram adicionados às colunas preparadas e incubados à temperatura ambiente por 20 minutos. Alíquotas de amostras das misturas foram retiradas em diversas etapas do processo para que fossem realizadas corridas de conferência por SDS-PAGE. Com as colunas em posição vertical os lisados foram esgotados e foram realizadas quatro lavagens com 5mL de tampão de lavagem cada. Finalmente foi feita a eluição das proteínas recombinantes com 4mL de tampão de eluição (50mmol.L^{-1} trisaminometano - HCL, 500mmol.L^{-1} NaCl, 500mmol.L^{-1} imidazol, 5mmol.L^{-1} BMER, $\text{pH} = 7,4$). A figura 10 refere-se ao gel de SDS-PAGE com alíquotas das amostras retiradas ao longo do processo de extração e das etapas iniciais de purificação da construção C1- 17N.

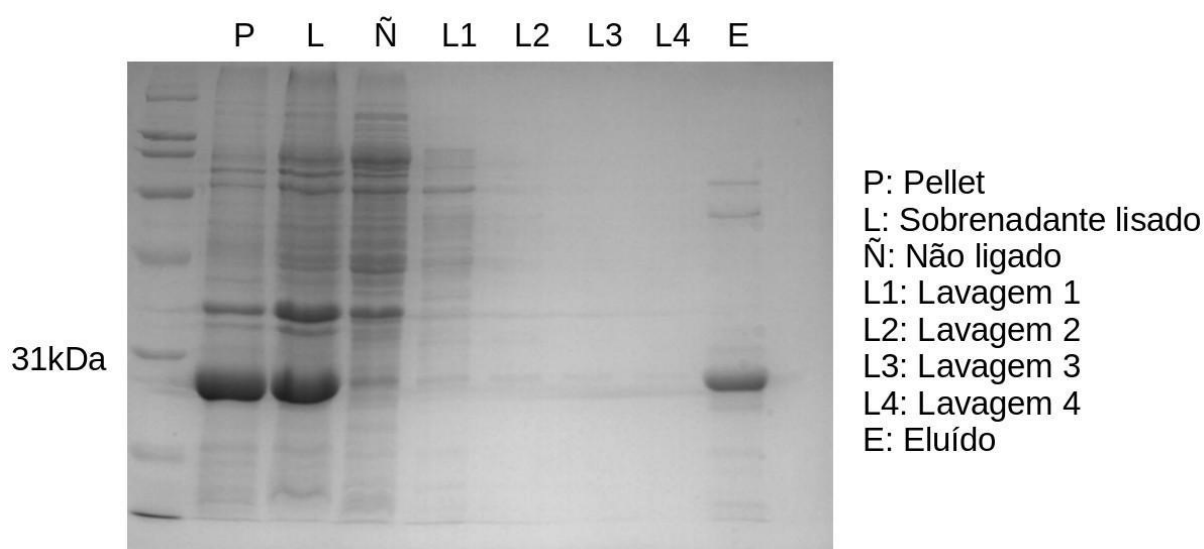


Figura 10. Purificação C1-17N – Coluna de Ni^{2+} : O gel acima é composto por amostras do *pellet* resultante da centrifugação do lisado (P), sobrenadante do lisado centrifugado (L), da fração de proteínas não ligadas na primeira incubação com a coluna de níquel (Ñ), das frações resultantes de quatro lavagens com baixa concentração de imidazol (L1, L2, L3 e L4) e da fração eluída com alta concentração de imidazol (E).

É possível observar a grande quantidade de proteína expressa (bandas próximas ao marcador de peso molecular relativo à 31kDa), as baixas perdas da proteína ao longo das lavagens e a baixa contaminação na fração eluída. Géis com frações equivalentes das proteínas mutantes são mostrados à frente.

Os tampões utilizados para eluição das proteínas foram então trocados utilizando-se colunas de troca de tampão *PD MiniTrap G-25* (GE Healthcare - 28918007). Para isso as colunas foram previamente equilibradas com 4 lavagens com 2mL de tampão de clivagem (50mmol.L^{-1} Na_2HPO_4 , 100mmol.L^{-1} NaCl , 5mmol.L^{-1} BMER, pH = 7,4) e alternaram-se adições de 0,5mL da fração eluída, 1,0mL de tampão de clivagem (fração descartada) e 2,0mL de tampão de clivagem (fração coletada) até que todo o volume originalmente eluído de cada amostra fosse passado pelas colunas.

Foram então realizadas incubações com 50 μL de solução de trombina ($1\text{U.}10\mu\text{L}^{-1}$) com albumina de soro bovino (BSA) a 0,1% para clivagem das caudas de histidina sendo realizados testes em intervalos de 1 hora (até 4 horas) ou *overnight* para incubação visando observar a eficiência do processo. A figura 11 mostra o gel de SDS-PAGE corrido para analisar as alíquotas retiradas para cada teste realizado. Verificou-se que após 1 hora de incubação virtualmente toda enzima presente apresentou-se clivada (aproximadamente 3mg). Como questão de segurança, entretanto, foi utilizado um período mínimo de incubação de 2 horas para todos os ensaios posteriores, inclusive com os mutantes gerados. Os demais ensaios de clivagem de cauda de histidina dos mutantes aqui apresentados foram incubados *overnight*.

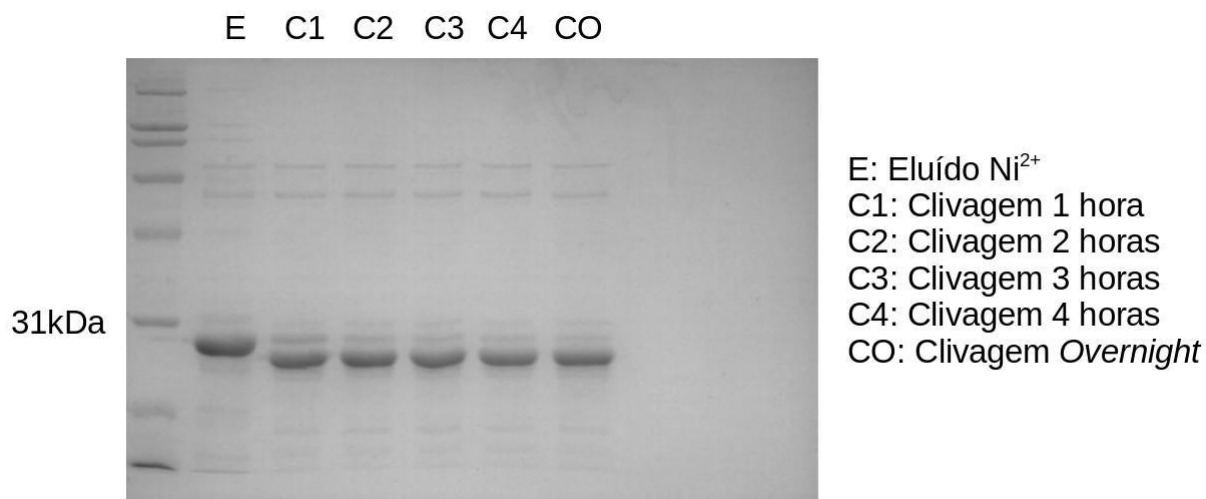


Figura 11. Purificação C1-17N – Clivagem da cauda de Histidinas: O gel acima é composto por amostras da fração eluída da coluna de Ni²⁺ (E), e de frações desta amostra eluída após tratamento com 5U de trombina/ BSA 0,1% por 1h (C1), 2h (C2), 3h (C3), 4h (C4) e *overnight* (CO), ou, aproximadamente, 18 horas.

Após a clivagem das caudas de histidinas as amostras obtidas foram concentradas com filtros MWCO de 3,0kDa *Centriplus YM-3 - Regenerated Cellulose* (Sigma-Aldrich - 4420) até que o volume total da solução de proteína (neste momento com cerca de 8,0mL) chegasse a cerca de 2,0mL. Um cromatógrafo AKTA FPLC (UPC-900) foi então preparado com coluna de cromatografia de exclusão de tamanho/ gel filtração *Superdex™ 75 10/300 GL* (Sigma-Aldrich - GE17-5174-01) a qual foi lavada com dois volumes de H₂O *milli-Q* e equilibrada com dois volumes de tampão de cromatografia (50mmol.L⁻¹ Na₂HPO₄, 500mmol.L⁻¹ NaCl, 1mmol.L⁻¹ EDTA, 5mmol.L⁻¹ BMER, pH = 7,4) ambos filtrados com filtros de porosidade 0.22µm e desaerados. Após o procedimento de cromatografia as frações com os maiores valores de Abs₂₈₀ detectados pelo aparelho (para a proteína selvagem: B10, B11, B12, C1, C2 e C3) foram selecionadas para corrida em gel de SDS-PAGE. A figura 12 mostra o gel obtido desta corrida e permite verificar que os contaminantes de alto peso molecular foram completamente removidos das frações selecionadas após a gel-filtração. Quantidades muito pequenas de contaminantes de menor peso molecular podem ser observadas, algumas possivelmente proteínas que interagem com a Cdc25B. Considerando as pequenas quantidades e manutenção destes contaminantes nas demais amostras (mutantes da Cdc25B) a presença dos mesmos não foi considerada suficiente para interferir nas observações dos ensaios de ¹H-¹⁵N - HSQC RMN e não foram realizadas etapas posteriores de purificação.

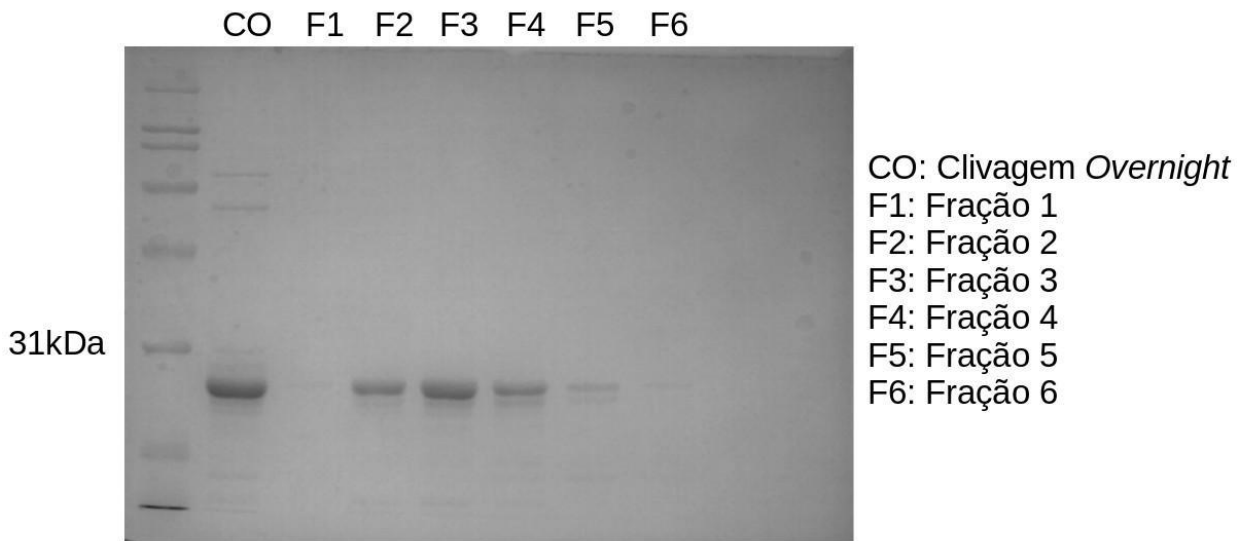


Figura 12. Purificação C1-17N – Gel-filtração: O gel acima é composto por amostras da enzima incubada *overnight* com trombina e pelas frações B10, B11, B12, C1, C2 e C3 eluídas da cromatografia de exclusão de tamanho (nomeadas respectivamente F1, F2, F3, F4, F5 e F6), as frações que apresentaram os maiores valores de abs_{280} medidos pelo aparelho AKTA FPLC UC-900.

Os géis com os resultados das cromatografias de afinidade, clivagem de cauda de histidina e gel-filtração dos mutantes seguem abaixo. Todos os procedimentos foram realizados da mesma forma que para a construção da proteína selvagem, utilizando-se os mesmos preparos de tampão, no mesmo dia.

Mutante R482A-17N:

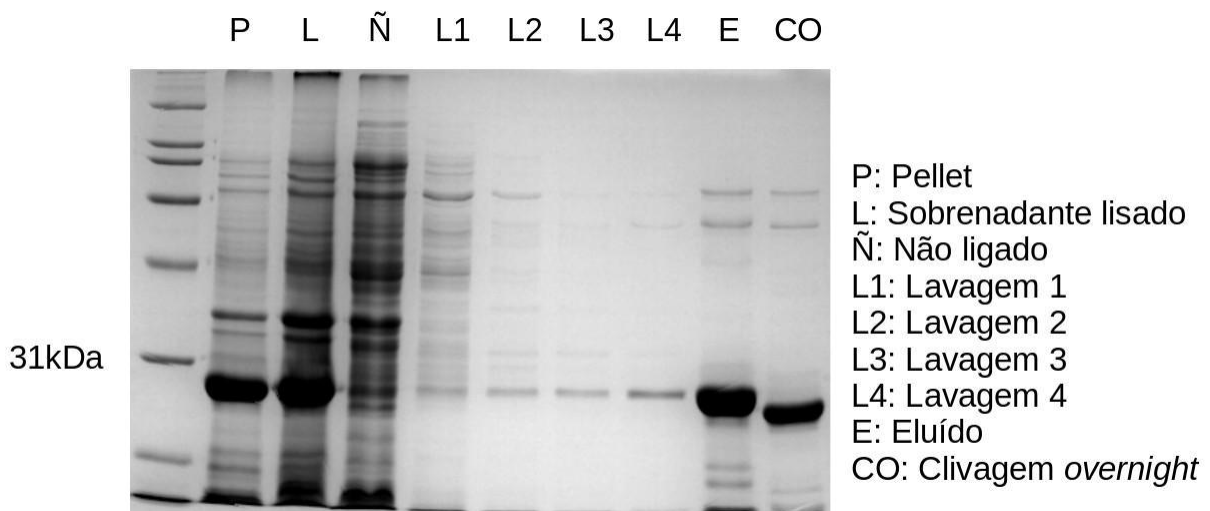


Figura 13. Purificação R482A - 17N – Coluna de Ni^{2+} : Amostras: *Pellet* resultante da centrifugação do lisado (P), sobrenadante do lisado centrifugado (L), da fração de proteínas não ligadas na primeira incubação com a coluna de níquel (Ñ), das frações resultantes de quatro lavagens com baixa concentração de imidazol (L1, L2, L3 e L4) e da fração eluída com alta concentração de imidazol (E). Proteína resultante da clivagem com trombina *overnight* (CO).

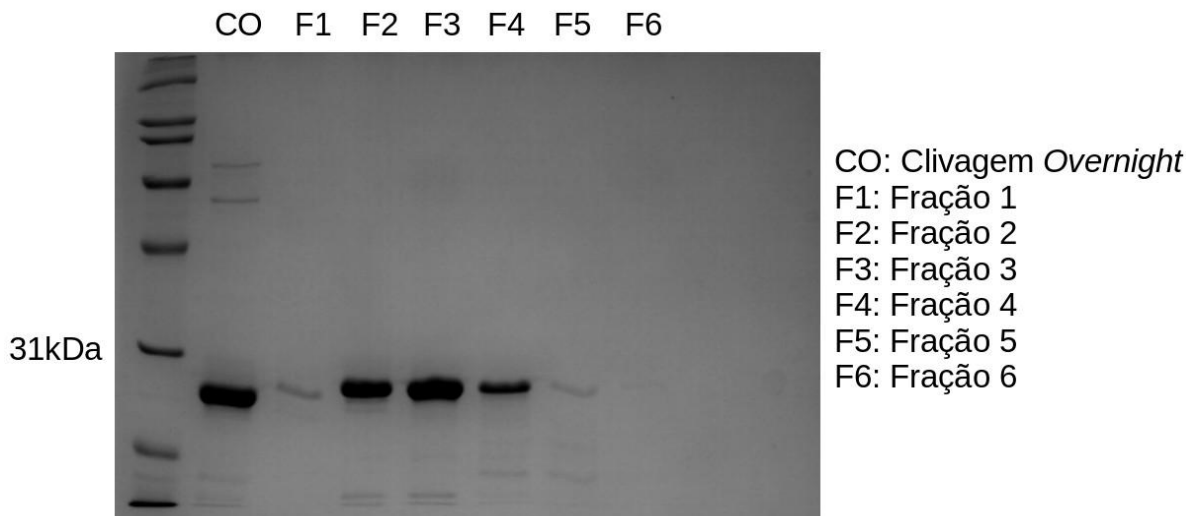


Figura 14. Purificação R482A - 17N – Gel-filtração: O gel acima é composto por amostras da enzima incubada *overnight* com trombina e pelas frações B10, B11, B12, C1, C2 e C3 da cromatografia de exclusão de tamanho (nomeadas respectivamente F1, F2, F3, F4, F5 e F6), as frações que apresentaram os maiores valores de abs_{280} medidos pelo aparelho AKTA FPLC UC-900.

Mutante R544A-17N:

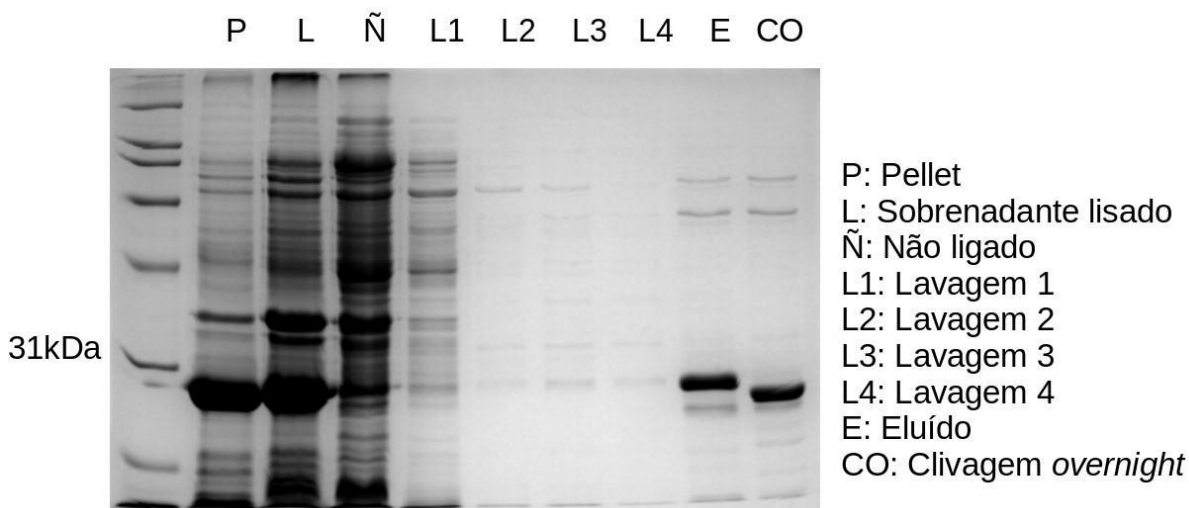


Figura 15. Purificação R544A - 17N – Coluna de Ni^{2+} : Amostras: *Pellet* resultante da centrifugação do lisado (P), sobrenadante do lisado centrifugado (L), da fração de proteínas não ligadas na primeira incubação com a coluna de níquel (Ñ), das frações resultantes de quatro lavagens com baixa concentração de imidazol (L1, L2, L3 e L4) e da fração eluída com alta concentração de imidazol (E). Proteína resultante da clivagem com trombina *overnight* (CO).

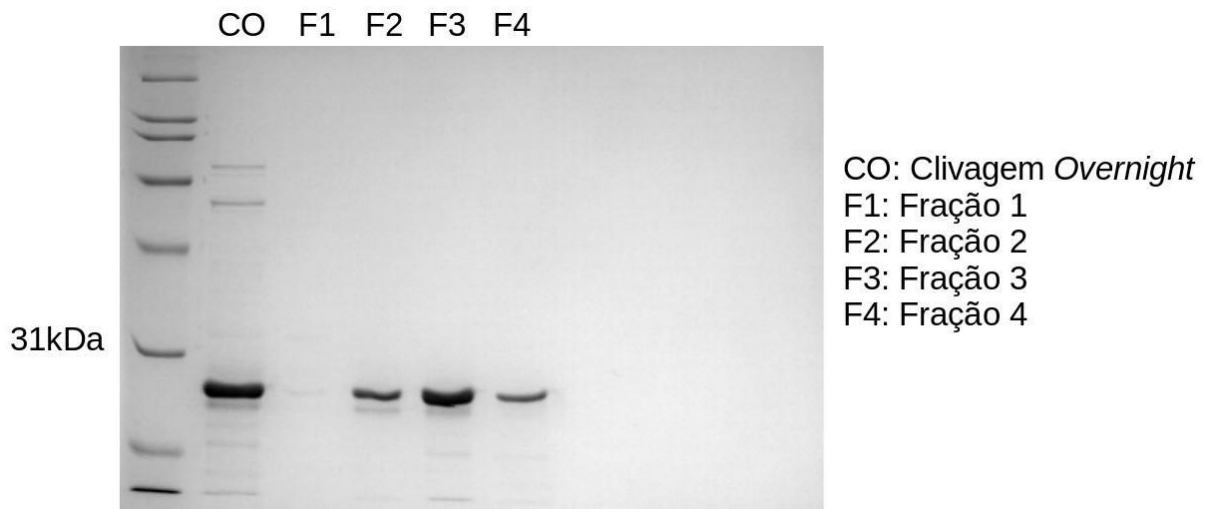


Figura 16. Purificação R544A - 17N – Gel-filtração: O gel acima é composto por amostras da enzima incubada *overnight* com trombina e pelas frações B10, B11, B12 e C1 da cromatografia de exclusão de tamanho (nomeadas respectivamente F1, F2, F3 e F4), as frações que apresentaram os maiores valores de abs_{280} medidos pelo aparelho AKTA FPLC UC-900.

Mutante W550A-17N:

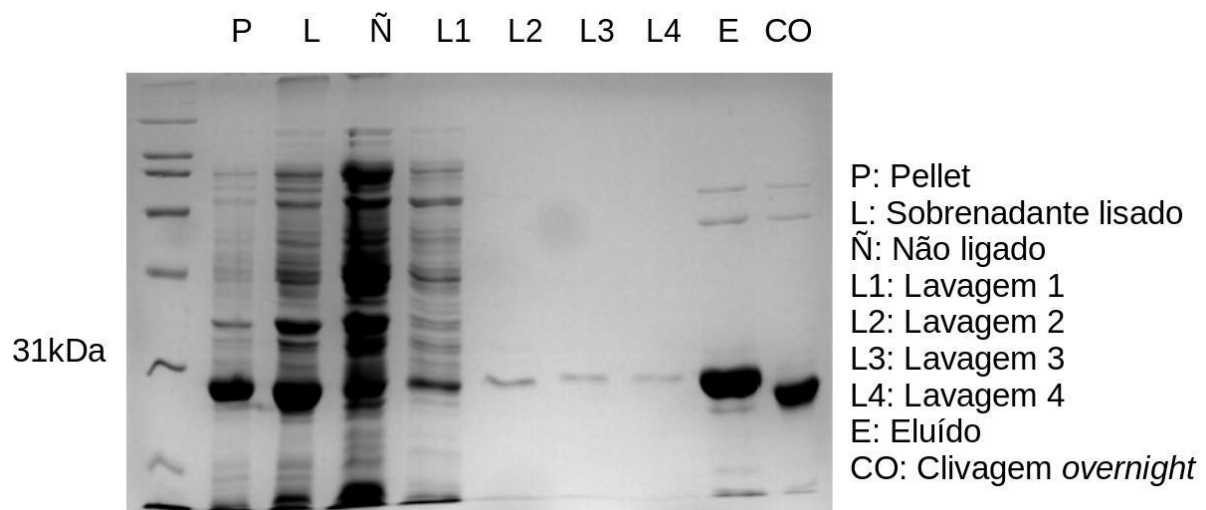


Figura 17. Purificação W550A - 17N – Coluna de Ni^{2+} : Amostras: *Pellet* resultante da centrifugação do lisado (P), sobrenadante do lisado centrifugado (L), da fração de proteínas não ligadas na primeira incubação com a coluna de níquel (Ñ), das frações resultantes de quatro lavagens com baixa concentração de imidazol (L1, L2, L3 e L4) e da fração eluída com alta concentração de imidazol (E). Proteína resultante da clivagem com trombina *overnight* (CO).

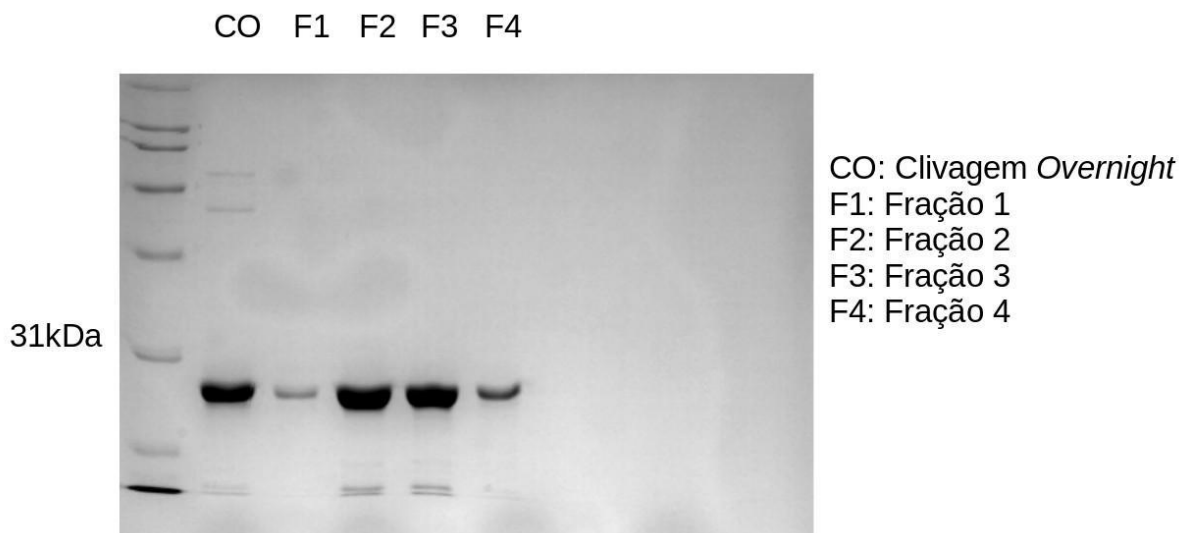


Figura 18. Purificação W550A - 17N – Gel-filtração: O gel acima é composto por amostras da enzima incubada *overnight* com trombina e pelas frações B10, B11, B12 e C1 da cromatografia de exclusão de tamanho (nomeadas respectivamente F1, F2, F3 e F4), as frações que apresentaram os maiores valores de abs_{280} medidos pelo aparelho AKTA FPLC UC-900.

As figuras de cada gel obtido após as corridas com as alíquotas retiradas durante o processo de expressão, extração e purificação dos mutantes pontuais gerados mostram, essencialmente, o mesmo que para os géis com as amostras de alíquotas retiradas durante as etapas equivalentes para a construção da proteína selvagem: a expressão dos mutantes ocorreu conforme o esperado e em quantidade significativa, a extração das proteínas foi adequada e durante as primeiras fases do processo de purificação a cromatografia de afinidade forneceu amostras consideravelmente puras, sendo que as lavagens posteriores garantiram a remoção da maior parte dos contaminantes, fornecendo amostras com grau de pureza consideravelmente superior. A cauda de histidinas utilizada para a cromatografia de afinidade foi removida com sucesso e a cromatografia por exclusão de tamanho permitiu obter amostras ainda mais puras removendo, sobretudo, contaminantes de maior peso molecular que restavam na amostra. Contaminantes de baixo peso molecular foram mantidos em pequenas quantidades, sobretudo comparando-se às quantidades expressas de cada construção.

Para confirmar a atividade da construção da proteína selvagem e testar a atividade dos mutantes gerados, foram realizados ensaios de atividade enzimática com o substrato p-Nitrofenil fosfato (pNPP) em concentração $0,2\text{mmol.L}^{-1}$ em tampão fosfato (50mmol.L^{-1} Na_2HPO_4 , 100mmol.L^{-1} NaCl , $0,2\text{mmol.L}^{-1}$ pNPP, 5mmol.L^{-1} BMER, $\text{pH} = 7,2$). O ensaio de atividade enzimática baseia-se na seguinte

sequência de reações químicas, catalisada por proteínas tirosina-fosfatases, como a Cdc25B:

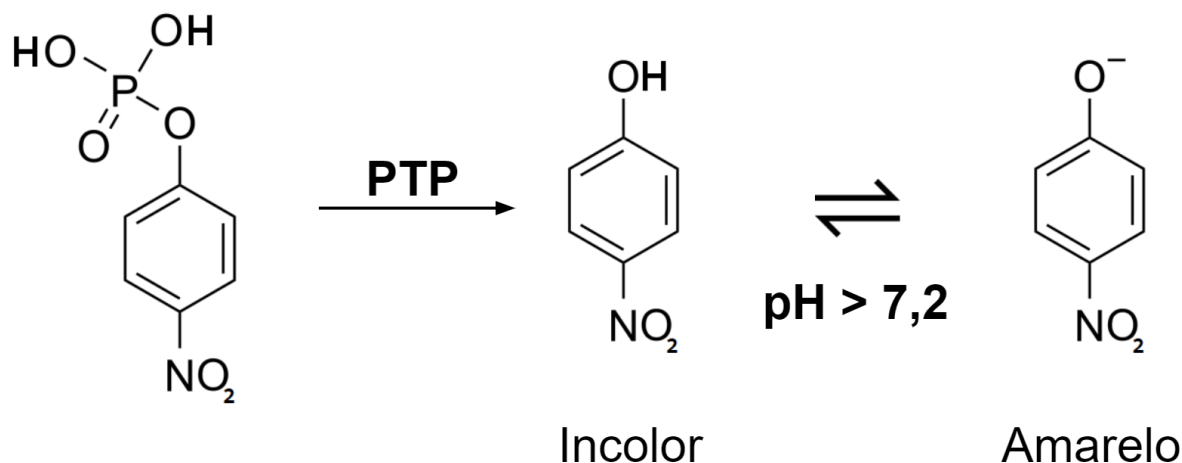


Figura 19. Simplificação da sequência de reações químicas ocorridas ao longo do ensaio de atividade para proteínas fosfatase realizado. O reagente para-Nitrofenol Fosfato (pNPP - à esquerda) tem o grupo fosfato removido pela proteína tirosina fosfatase (PTP) formando o para-Nitrofenol (pNP). Em meio alcalino, o pNP, incolor, é convertido ao para-Nitrofenolato, de coloração amarela, permitindo a observação da reação química anterior.

A série de reações químicas mostra o para-Nitrofenol Fosfato (pNPP) tendo seu grupo fosfato removido por ação de uma proteína tirosina fosfatase (PTP) formando o para-Nitrofenol (pNP), incolor. A adição de Na₂CO₃ ao meio após a incubação inicial das amostras das proteínas com o pNPP converte eventuais moléculas de pNP presentes ao fenolato equivalente, o para-Nitrofenolato, de coloração amarela, o que permite inferir a ocorrência da reação anterior.

O ensaio demonstrou atividade nas frações eluídas e selecionadas por sua Abs₂₈₀ durante a cromatografia de exclusão de tamanho. Medidas de absorvância em comprimento de onda 420nm (Abs₄₂₀) das amostras após o ensaio de atividade enzimática mostraram correlação aparente com a Abs₂₈₀ da curva indicada no AKTA, conforme esperado, assim como pode ser visualmente verificado pelos tamanhos das bandas relativas nos géis de SDS-PAGE corridos. Estes ensaios mostram que as mutações realizadas não afetam a capacidade da enzima de converter substratos simples em produtos sem fornecer informações, entretanto a respeito de questões cinéticas, não avaliadas. Como última observação é possível inferir que as proteínas expressas, assim como os mutantes gerados,

apresentam-se com estruturação nativa, necessária para atuação enzimática das mesmas.

Tabela 1. Valores de absorvância em comprimento de onda 420nm (Abs_{420}) após ensaio de atividade de alíquotas retiradas das frações obtidas a partir da cromatografia de exclusão de tamanho na última etapa de purificação das 4 construções da Cdc25B (proteína selvagem e mutantes gerados). A observação retirada daqui é meramente qualitativa e não foram realizados ensaios de cinética enzimática com os mutantes permitindo-se apenas observar a atividade de todas as construções e que frações que apresentaram maiores bandas nos géis de SDS-PAGE corridos apresentaram maiores valores de Abs_{420} .

	C1-17N		R482A		R544A		W550A	
Fração/ Abs₄₂₀	B10	0,1836	B10	0,1917	B10	0,4620	B10	0,2355
	B11	0,7266	B11	1,2209	B11	0,5792	B11	1,5841
	B12	1,2975	B12	1,6169	B12	1,1758	B12	1,7350
	C1	0,7653	C1	0,5956	C1	0,7279	C1	0,5293
	C2	0,3254	C2	0,3104	-	-	-	-
	C3	0,1935	C3	0,1412	-	-	-	-

2.1.4. Aquisição de espectros de 1H - ^{15}N - HSQC RMN e análises de perturbação de deslocamento químico

Para aquisição dos espectros de 1H - ^{15}N HSQC RMN as amostras foram preparadas com concentração de $0,2\text{mmol.L}^{-1}$ da proteína de interesse em tampão fosfato redutor e água deuterada (20mmol.L^{-1} Na_2HPO_4 , 50mmol.L^{-1} NaCl , 5mmol.L^{-1} BMER, 2mmol.L^{-1} ditiotretitol, 5% $^2\text{H}_2\text{O}$, pH = 6,7). Foi utilizado aparelho de ressonância magnética nuclear Bruker Ascend III de 800MHz com criosonda TCI de 4 canais e console Avance III. Os espectros foram adquiridos na CA - IQUSP.

O processamento inicial dos espectros foi realizado com o programa NMRPipe [80] e a transferência de assinalamento para os picos obtidos foi realizada utilizando-se o programa CCPNMR Analysis 2.4 [81], inicialmente a partir do acessório de seleção de área e posteriormente com marcação manual. A partir dessa abordagem, sabendo-se que dentre os 194 resíduos do domínio catalítico da proteína 150 foram assinalados pelo trabalho anterior do Dr. Raphael SantaRosa Sayegh [23], quase todos os resíduos tiveram seu assinalamento transferido para as

novas construções (146 para a proteína selvagem, 146 para o mutante R482A, 144 para o mutante R544A e 144 para o mutante W550A - Anexo B).

Para definição da perturbação de deslocamento químico (*Chemical Shift Perturbation* - CSP) foi utilizada a seguinte fórmula:

$$\text{Equação 9} \quad \text{CSP} = \sqrt{\frac{1}{j_{\alpha}} (\Delta\delta^H)^2 + \frac{1}{k_{\alpha}} (\Delta\delta^{15N})^2}$$

Em que os valores para $\Delta\delta$ representam as variações de deslocamento químico pós-mutação nas frequências de ^1H ou ^{15}N e os valores de j_{α} e k_{α} são fatores de ponderação para cada resíduo analisado, dependendo do desvio padrão observado nas distribuições normais dos valores depositados para estas frequências no banco de dados RefDB [82]. A correção individual para diferentes resíduos e entre cadeia principal e lateral se fez necessária devido ao fato de a análise de CSP com uma fórmula genérica (normalmente ponderando a frequência de ^1H como “1,00” e de ^{15}N como “0,15”) subestimar perturbações de deslocamento químico pequenas que podem ser representativas. Por exemplo, enquanto os resíduos apresentam desvio padrão médio para a frequência de próton de 0,670, para a cadeia lateral do triptofano este valor é de 0,490. Para as frequências de nitrogênio o desvio padrão médio é de 4,350 frente 1,880 para a cadeia lateral do triptofano. Os valores de j e k utilizados o encontram-se na tabela 2.

Tabela 2. Valores dos fatores j e k utilizados na fórmula para o cálculo da perturbação de deslocamento químico (CSP) dos resíduos de aminoácidos dos mutantes da Cdc25B em relação à proteína selvagem. Para o resíduo de triptofano foram definidos valores para o grupo N - H da cadeia principal (*bb* - *backbone*) e da cadeia lateral do resíduo (*sc* - *sidechain*).

	j_{α}	k_{α}		j_{α}	k_{α}		j_{α}	k_{α}
Ala	0,660	4,040	Arg	0,640	3,960	Asn	0,600	4,070
Asp	0,710	4,460	Cys	0,730	4,331	Gln	0,620	3,730
Glu	0,640	3,670	Gly	0,760	3,730	His	0,730	4,090
Ile	0,750	4,460	Leu	0,700	4,131	Lys	0,680	3,851
Met	0,650	3,660	Phe	0,770	4,230	Pro	-	-
Ser	0,650	3,810	Thr	0,650	4,941	Trp <i>bb</i>	0,860	4,649
Trp <i>sc</i>	0,491	1,884	Tyr	0,790	4,550	Val	0,720	4,850

A análise estatística foi realizada a partir da remoção iterada de valores de CSP *outliers* definidos como aqueles acima da média dos valores presentes (\bar{x}) somados 3 desvios padrão do conjunto de valores (σ) até que não restassem valores fora deste limite ($\bar{x} + 3\sigma$). Os valores de CSP superiores a $\bar{x}' + \sigma'$ (com \bar{x}' e σ' a média e o desvio padrão para o último grupo de valores de CSP definido) foram considerados significativamente alterados [83]. Os deslocamentos químicos da proteína selvagem e dos mutantes R482A, R544A e W550A foram depositados no BMRB *Databank* [84] com códigos de acesso respectivos entre 50030 a 50033.

2.2. Resultados e discussão

2.2.1. Validação da estimativa energética da interação cátion- π por campos de força clássicos

Os cálculos MP2/6-311++G** para a energia de interação entre as cadeias laterais de resíduos de arginina e triptofano (figura 20) indicam que ambas as orientações analisadas (“sanduíche” e “T”) apresentam mínimos de energia de aproximadamente $-10\text{kJ}\cdot\text{mol}^{-1}$. Para a orientação sanduíche o mínimo se dá quando a distância entre CDMs de ambas cadeias laterais é de, aproximadamente, 0,41nm enquanto para a orientação T o mínimo se dá quando a distância entre os CDMs é de, aproximadamente, 0,59nm.

A comparação entre as energias de interação calculadas pelo campo de força e a referência quântica apresenta diferenças inferiores a $0,5\text{kJ}\cdot\text{mol}^{-1}$ para a maioria das distâncias em ambas orientações utilizando-se o CHARMM36 enquanto o AMBER99-ILDN indica valores ligeiramente mais distantes para a orientação sanduíche (com diferenças de até $3,0\text{kJ}\cdot\text{mol}^{-1}$) e sequer indica uma interação energeticamente favorável entre os grupos na orientação T.

A figura 20 reúne os resultados obtidos e mostra que o campo de força CHARMM36 descreve com excelente precisão a energia da interação cátion- π entre a cadeia lateral de uma arginina e o anel indólico de um triptofano sendo adequado para ser utilizado em simulações de dinâmica molecular com o objetivo de estudar este tipo de contato. O campo de força AMBER99-ILDN foi desconsiderado para esse mesmo propósito.

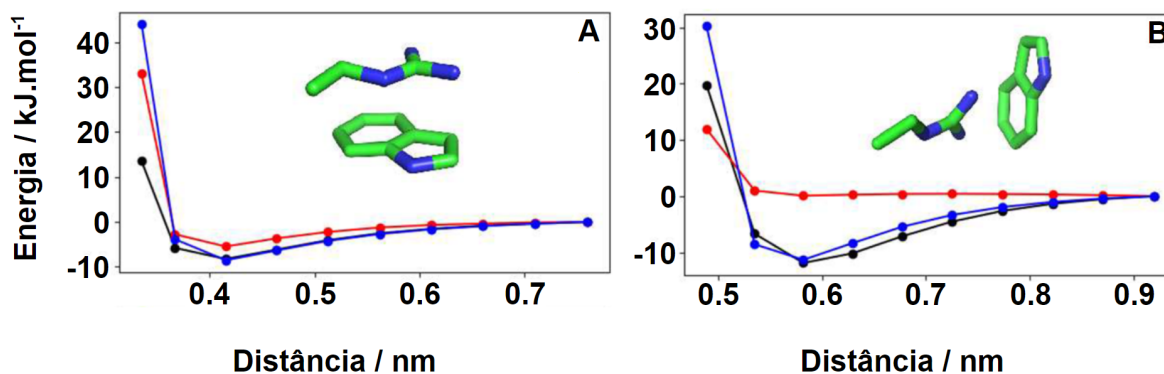


Figura 20. Energia potencial de interação entre as cadeias laterais de resíduos modificados (bloqueados) de arginina e triptofano nas orientações “sanduíche” (A) ou em “T” (B). Pontos em preto representam os valores obtidos a partir da referência quântica (MP2/6-311++G**) e os pontos em azul e vermelho representam, respectivamente, os valores obtidos pelos campos de força testados, CHARMM36 e AMBER99-ILDN.

2.2.2 Escolha de resíduos candidatos e geração de perfis de energia livre

As análises de distância entre os CDMs das cadeias laterais das argininas e do triptofano presentes no domínio catalítico da Cdc25B ao longo da simulação de dinâmica molecular longa (6 μ s) demonstraram grandes variações na posição relativa de resíduos específicos em relação ao definido pelo cristal da proteína (1QB0) [23]. Dentre os resíduos de arginina inicialmente mais distantes do triptofano, na estrutura cristalográfica depositada, alguns chegam a se aproximar mais de 1,90nm deste último, alcançando uma distância que torna possível, por exemplo, a formação de um contato cátion- π em qualquer uma das orientações já mencionadas.

A tabela abaixo apresenta as distâncias entre os CDMs das cadeias laterais das argininas numeradas e do W550 medidas no cristal depositado (d_{RX}) e a distância mínima observada ao longo da simulação longa realizada (d_{MD}). Valores em negrito encontram-se abaixo do *cutoff* estabelecido para consideração do resíduo como candidato em interações com o triptofano ($d_{MD} \leq 0,8\text{nm}$). Os resíduos com d_{RX} anotados com “-” estão localizados no C-terminal desordenado da proteína e, portanto, não se encontram presentes na estrutura cristalográfica 1QB0 [23].

Tabela 3. Distâncias entre os CDMs de cadeias laterais de argininas específicas (# Arg) e o CDM do anel indólico do W550 para o cristal depositado da proteína (d_{RX}) e no momento em que estes se encontram mais próximos durante uma dinâmica molecular longa simulada (dm_{MD}). Valores apresentados em negrito encontram-se abaixo do *cutoff* estabelecido para definir um resíduo de arginina como candidato em interações com o W550.

# Arg	d_{RX}	dm_{MD}	# Arg	d_{RX}	dm_{MD}	# Arg	d_{RX}	dm_{MD}
376	3,71	1,85	427	2,07	1,27	447	1,25	0,37
466	3,81	2,26	479	1,47	0,54	482	1,19	0,29
485	2,10	0,33	488	2,62	0,66	490	2,02	0,31
492	2,84	0,46	529	2,32	2,07	544	0,83	0,52
548	0,97	0,30	554	-	0,30	556	-	0,40
557	-	0,80	562	-	0,36	-	-	-

Dessa forma, oito resíduos de arginina foram selecionados para geração de perfis de energia livre: R447, R482, R485, R490, R544, R554, R556 e R562. Dentre os resíduos citados, os últimos quatro não estão presentes na estrutura cristalina da proteína fazendo parte da hélice C-terminal desordenada [24], conforme já mencionado. A R479 atua no ciclo catalítico e foi desconsiderada como candidata. Os resíduo de R488 e R492 representam um sítio de ligação a fosfato e, dessa forma, não foram gerados perfis de energia livre para estes. O resíduo de R548 foi excluído da análise devido sua grande proximidade com o W550. É importante notar, também, que o CDM de nenhum grupo guanidino se encontrava inicialmente em uma distância do CDM do anel indólico do W550 inferior ao valor de *cutoff* definido para a escolha dos candidatos (ou seja, não existe valor de $d_{RX} \leq 0,8\text{nm}$).

A figura 21 reúne os perfis de energia livre gerados para as interações cátion- π entre as cadeias laterais das argininas candidatas e do W550, obtidos a partir das simulações realizadas com a técnica de US e análise pelo método de WHAM:

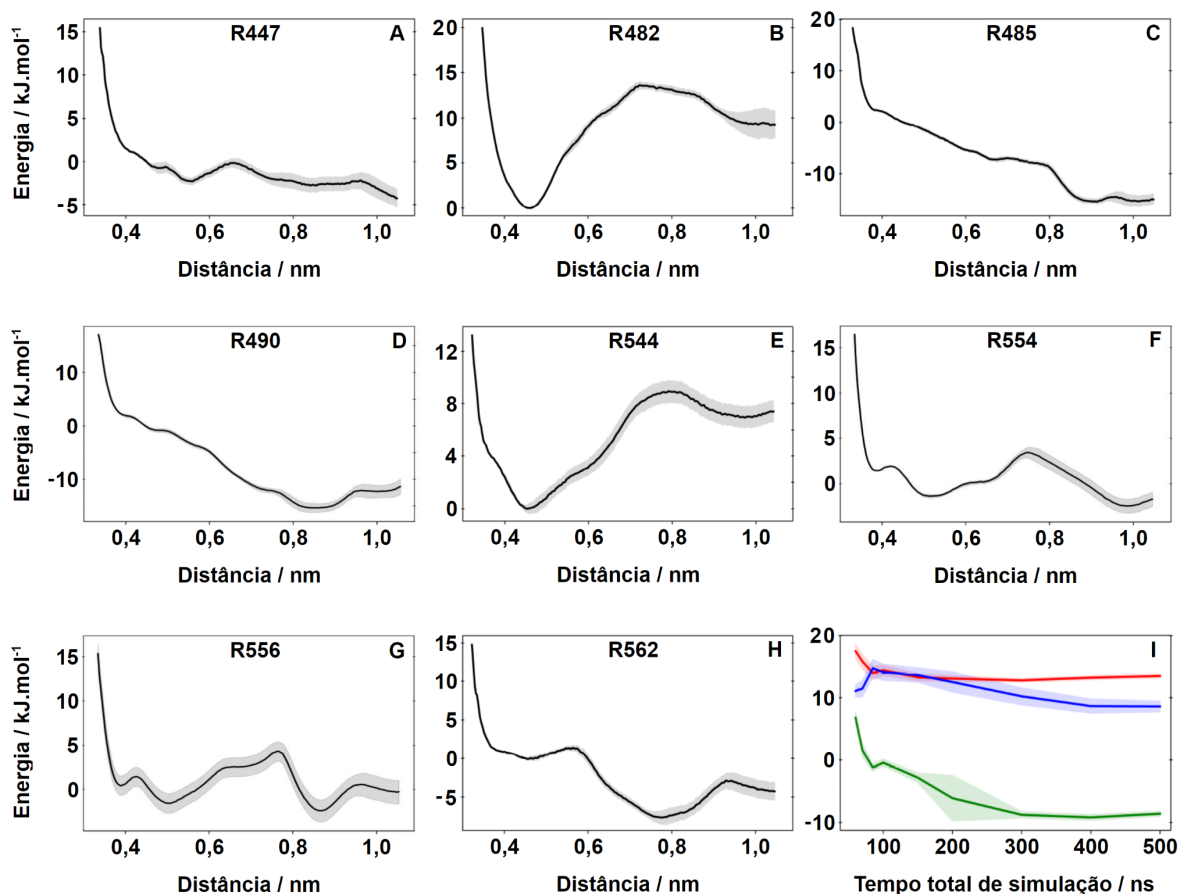


Figura 21. Perfis de energia livre para a interação cátion- π entre o anel indólico do W550 e os grupos guanidino das argininas: R447 (A), R482 (B), R485 (C), R490 (D), R544 (E), R554, (F), R556 (G), R562 (H). É mostrada a incerteza estatística em sombreado e a convergência das simulações com diferença de energia livre entre complexos cátion- π com distância entre 0,45 e 0,80nm para R482 (vermelho), R485 (verde) e R544 (azul).

Dentre os perfis de energia livre gerados, apresentados na figura acima, é possível observar três conjuntos característicos: o de perfis com aparência plana, o de perfis dissociativos e o de perfis que indicam ligação. Os perfis representando a energia de interação entre as cadeias laterais das R447, R554, R556 e R562 e do W550 mostram-se, essencialmente, planos. Nesses, são observados valores de mínimos de energia livre de interação em distâncias de ligação (entre 0,4 e 0,8nm) mas estes valores são pouco pronunciados, indicando que os contatos que ocorrem entre as cadeias laterais destas argininas e do W550 devem ser pouco estáveis, apresentando curto tempo de vida.

Os perfis para as R485 e R490 apresentam-se do tipo dissociativo, com mínimos de energia livre de interação ocorrendo mais distantes do que o necessário para permitir a formação de contatos cátion- π entre seus grupos guanidino e o anel

indólico do triptofano, indicando que não existem interações estáveis por partes destes grupos.

Finalmente, as R482 e R544 apresentaram perfis de energia livre de ligação, indicando interação favorável entre suas cadeias laterais e a do W550. Os mínimos de energia livre de interação são consideráveis (respectivamente $14\text{kJ}\cdot\text{mol}^{-1}$ e $9\text{kJ}\cdot\text{mol}^{-1}$), próximos ao valor observado pelos cálculos quânticos realizados, e localizados em distâncias adequadas para a formação de contatos cátion- π (ambos a cerca de $0,45\text{nm}$). com barreiras de dissociação para estes mesmos contatos.

A partir destas observações iniciou-se a tentativa de validar experimentalmente as interações entre as cadeias laterais das R482 e R544 e do W550, utilizando-se a técnica de ^1H - ^{15}N - HSQC RMN.

2.2.3 Validação de contatos por perturbação de deslocamento químico em ressonância magnética nuclear

Com as análises de ^1H - ^{15}N - HSQC RMN para os mutantes da Cdc25B buscou-se encontrar resíduos com deslocamento químico perturbado em decorrência da modificação do ambiente químico local pelo rompimento de um contato cátion- π entre as R482 e R544 e o W550 ou da alteração entre a dinâmica dos contatos em que o W550 participa (caso alguma arginina, por exemplo, além das duas mutadas, que participe de contatos transientes com o triptofano, passasse a interagir por períodos mais longos ou com maior frequência com este).

A tabela 4 apresenta as CSP para as todas as argininas assinaladas na proteína e o para o triptofano (*backbone* - W550 - e cadeia lateral - W550 ϵ_1) nos três mutantes gerados em comparação à proteína selvagem. É possível observar que dentre as 15 argininas mostradas (das 17 presentes no polipeptídeo), oito sofreram perturbações significativas com a troca do W550 por um resíduo de alanina. Apesar do grande número de resíduos perturbados, é importante notar que a metade destes (R548, R554, R556 e R557) se encontra na região C-terminal flexível, e estão muito próximos do resíduo W550 sendo que a troca deste aminoácido por uma alanina altera consideravelmente a estereoquímica local (pela substituição de um grupo bastante volumoso - anel indólico - por uma metila), e, dessa forma, já era esperado que resíduos próximos apresentassem este

comportamento em termos de CSP. Assim, estes resíduos foram desconsiderados como parceiros de interação do triptofano em contatos cátion- π .

Tabela 4. Perturbação de deslocamento químico para argininas e triptofano nos mutantes pontuais da Cdc25B gerados: W550A, R482A e R544A. É possível observar o CSP anotado para o anel indólico do W550 (W550 ϵ_1) e a característica da curva do perfil de energia livre.

Resíduo	Mutante			Perfil de energia livre
	W550A	R482A	R544A	
W550	-	0,103	0,105	-
W550 ϵ_1	-	0,045	0,027	-
R376	0,011	0,014	0,018	-
R427	0,007	0,036	0,042	-
R447	0,087	0,093	0,011	Plano
R466	-	-	-	-
R479	-	-	-	-
R482	0,016	-	0,029	Ligação
R485	0,009	0,062	0,013	Dissociativo
R488	0,002	0,022	0,026	-
R490	0,004	0,033	0,006	Dissociativo
R492	0,004	0,021	0,006	-
R529	0,016	0,012	0,041	-
R544	0,167	0,249	-	Ligação
R548	0,170	0,083	0,100	-
R554	0,409	0,019	0,028	Plano
R556	0,082	0,037	0,027	Plano
R557	0,016	0,049	0,015	-
R562	0,008	0,014	0,070	Plano
$\bar{x} + \sigma'$	0,012	0,030	0,026	-

Dentre os demais resíduos de argininas significativamente perturbados para o mutante W550A dois foram previamente desconsiderados como potenciais parceiros do triptofano em contatos com anel indólico: a R529 não se aproxima nem mesmo a 2,0nm deste durante a simulação longa, distância muito superior à necessária para uma interação cátion- π em qualquer orientação, e a R447 apresentou um perfil de energia livre “plano”, indicando que possíveis contatos com o triptofano não devem ser estáveis e, portanto, estes resíduos também foram desconsiderados como parceiros importantes do W550 em nas interações procuradas.

Os motivos para serem observados valores significativos de CSP nas R447 e R529 podem ser indiretos e derivados de perturbações causadas a longa distância pela mutação do triptofano. É uma observação experimental que estes resíduos se localizam em regiões consideravelmente flexíveis (parâmetros de ordem obtidos em experimentos de RMN $S^2 \leq 0,95$ para o segmento D448 - L453 e $S^2 \leq 0,90$ para o segmento D527 - M531 da proteína, onde se localizam as R447 e R529) [23] e a remoção de um resíduo volumoso de uma região próxima, ainda mais candidato a interações com outros resíduos, alguns localizados em hélices distantes (como R482, localizado na hélice D) pode alterar a dinâmica regional e, conseqüentemente, o ambiente químico para ambos resíduos, gerando perturbações que acabam se propagando.

Finalmente, os resíduos de R482 e R544, os únicos resíduos, dentre os analisados pelos perfis de energia livre de ligação para a interação com o triptofano, apresentaram valores de CSP significativamente alterados, uma evidência experimental da formação dos contatos cátion- π .

É importante ressaltar que a inexistência de CSP significativo para sete argininas, R376, R427, R485, R488, R490, R492 e R562, representa uma evidência experimental da não participação em contato cátion- π com o triptofano. É difícil imaginar uma situação em que houvesse tão importante alteração no ambiente químico em que um destes resíduos se encontra e esta fosse, de alguma forma, compensada a ponto de não expressar perturbação no deslocamento químico medido. Deve-se lembrar, também, que dentre os resíduos deste grupo para os quais foram traçados perfis de energia livre (R485, R490 e R562) todos se apresentaram dissociativos ou planos de forma que a evidência experimental concorda com o observado nas simulações.

A figura 22 apresenta uma montagem com o espectro obtido para o mutante W550A com destaque para os picos deslocados de alguns resíduos importantes (A) e representa a estrutura da Cdc25B com um mapa relativo aos valores de CSP observados (B):

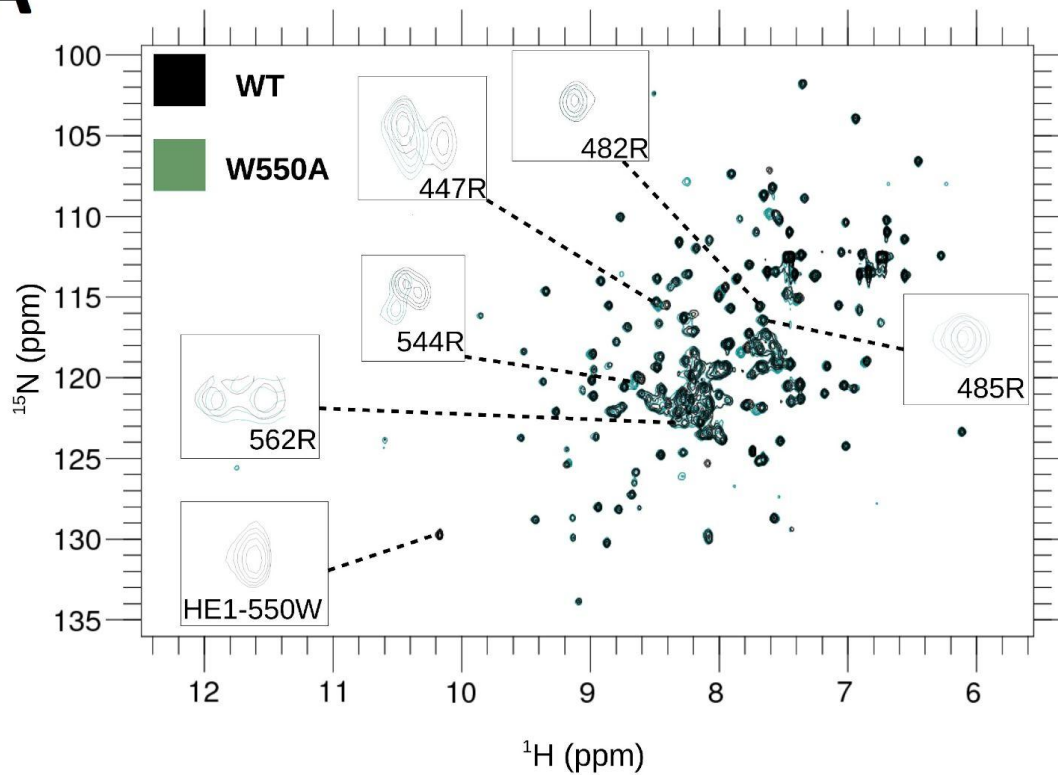
A**B**

Figura 22. (A) Espectro de ^1H - ^{15}N HSQC RMN do domínio catalítico da Cdc25B e CSP após mutação W550A. (B) Projeção de CSP ao longo da estrutura da Cdc25B, de acordo com legenda ao lado. Trechos em cinza não foram assinalados.

Conforme pode ser observado na figura 22 uma grande quantidade de regiões significativamente perturbadas são verificadas ao longo da estrutura da Cdc25B. Como esperado, diversos resíduos na região C-terminal da proteína (à direita na imagem), tanto na região flexível quanto na hélice alfa imediatamente vizinha a esta região, apresentam altos valores de CSP mas as alterações não se resumem a estes segmentos muito próximos do resíduo mutado. Resíduos consideravelmente distantes do W550, alguns localizados na região N-terminal da proteína, apresentaram-se significativamente perturbados e é difícil encontrar uma resposta simples para esta observação.

Explicações para as perturbações significativas observadas em regiões distantes do resíduo inicialmente mutado no peptídeo devem seguir o mesmo raciocínio especulativo que o utilizado para as argininas que não apresentaram perfis de energia livre de ligação: é possível que se tratem de perturbações indiretas a longa distância ou que alterações na dinâmica de contatos de resíduos da região C-terminal tenham causado perturbações indiretas, gerando as CSP significativas verificadas nestas regiões.

É difícil prever quais resíduos dentro da estrutura de uma proteína são os principais responsáveis pela estruturação de uma determinada região quando se foge de análises simples como da estrutura de hélices- α ou folhas- β . Em termos de estrutura tridimensional, alterações próximas a resíduos conhecidos como “resíduos *hubs*” são conhecidos por gerar CSP em longas distâncias [85] e este fato poderia, facilmente, causar as observações verificadas como nas grandes variações de CSP em áreas distantes do triptofano mutado no domínio catalítico da Cdc25B.

Este tipo de argumentação, no entanto, pode levar ao questionamento sobre as perturbações observadas nos próprios resíduos candidatos com perfis de energia livre de ligação. Dessa forma, com o intuito de obter uma contra-prova para as observações do espectro obtido para o mutante W550A foram realizados procedimentos idênticos com alteração dos resíduos de R482 e R544 (respectivamente mutantes R482A e R544A). Assim como na figura 22 as figuras 23 e 24 apresentam uma montagem com os espectros obtidos para os mutantes de arginina e apresentam a estrutura tridimensional do domínio catalítico da Cdc25B com mapas relativos aos valores de CSP observados.

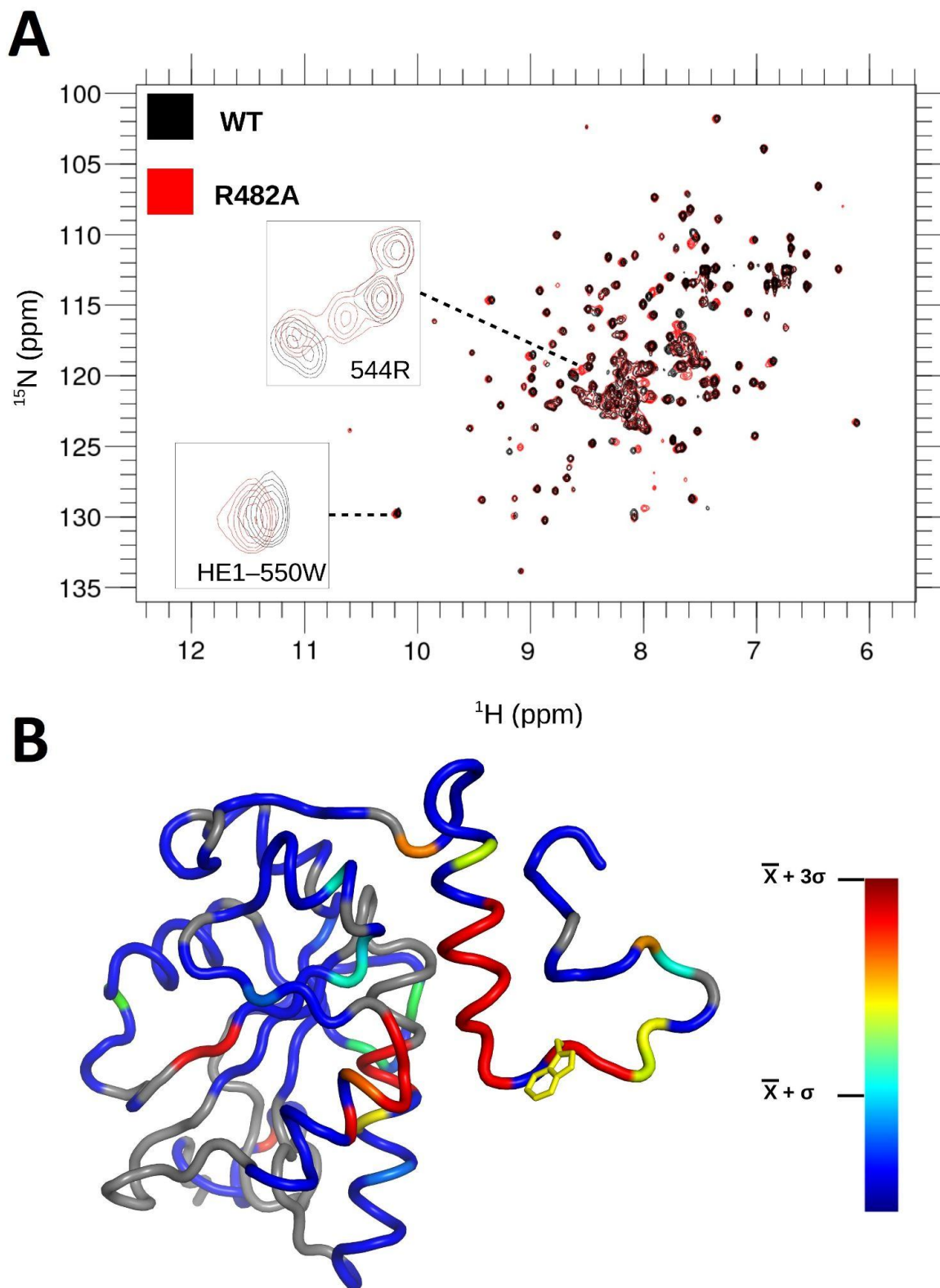


Figura 23. (A) Espectro de ^1H - ^{15}N HSQC RMN do domínio catalítico da Cdc25B e CSP após mutação R482A. (B) Projeção de CSP ao longo da estrutura da Cdc25B, de acordo com legenda ao lado. Trechos em cinza não foram assinalados.

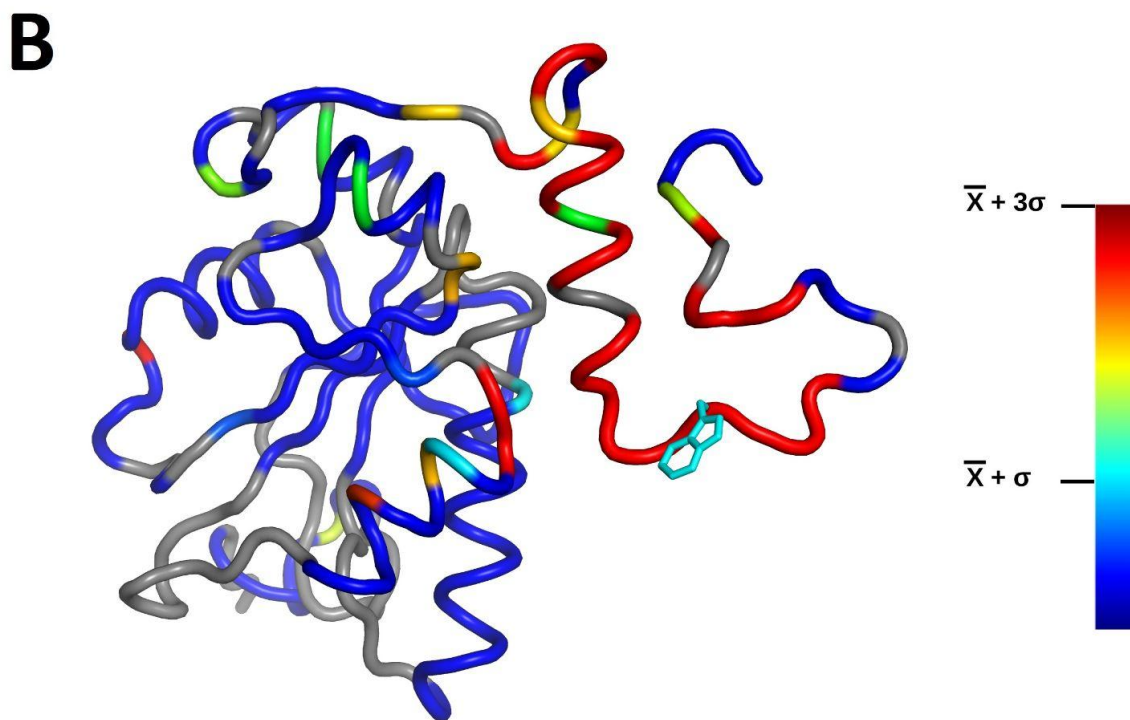
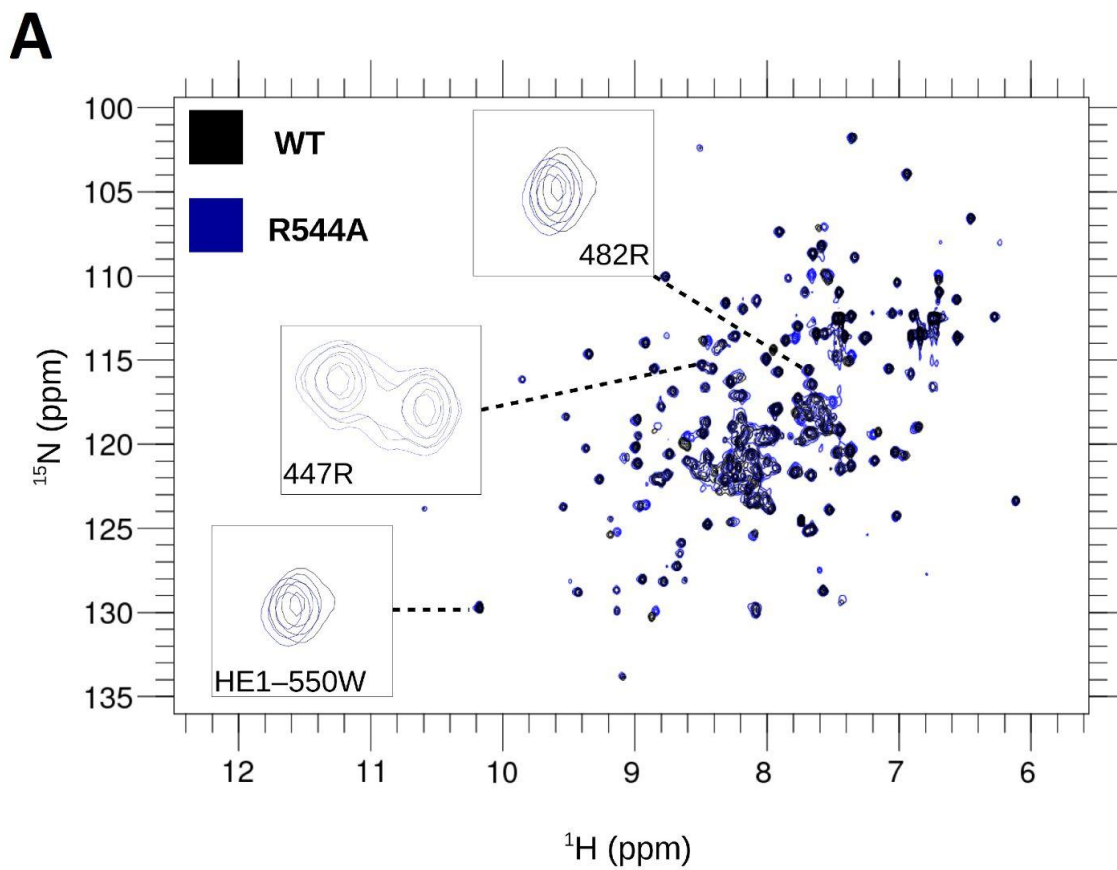


Figura 23. (A) Espectro de ^1H - ^{15}N HSQC RMN do domínio catalítico da Cdc25B e CSP após mutação R544A. (B) Projeção de CSP ao longo da estrutura da Cdc25B, de acordo com legenda ao lado. Trechos em cinza não foram assinalados.

Para ambos mutantes, R482A e R544A, são novamente observadas diversas regiões distantes dos sítios de mutação que se encontram significativamente perturbadas. As possíveis explicações seguem o mesmo raciocínio definido para o mutante W550A. Primeiramente, é importante notar que para ambos mutantes, tanto o sinal de cadeia principal (W550, na tabela 4) do triptofano quanto do anel indólico (W550 ϵ_1 , na tabela 4) apresentam-se significativamente perturbados, um indicativo de alteração do ambiente químico local que, considerando-se todos os dados apresentados até o momento, é evidência adicional da formação do contato cátion- π entre as R482 e R544 e o W550.

Interessantemente, os padrões de perturbação observados em ambos mutantes de argininas (R482A e R544A) se assemelha consideravelmente ao mutante gerado para o triptofano, com grande quantidade de resíduos significativamente perturbados principalmente ao longo da porção C-terminal da proteína, assim como em alguns resíduos da hélice F e até mesmo do N-terminal, distantes dos resíduos de R482 e R544. Tais observações servem como mais uma evidência de interação entre os resíduos estudados (R482/W550 e R544/W550) uma vez que é de se esperar perfis de perturbação semelhantes com o impedimento da formação de um contato específico pela remoção de qualquer um dos resíduos envolvidos neste contato.

Finalmente, verifica-se que o sinal para a R544 encontra-se significativamente perturbado no mutante R482A e que o sinal para a R482 encontra-se perturbado no mutante R544A. Novamente, tal observação é esperada considerando-se a expectativa de contatos transientes entre os resíduos de arginina citados e o W550, uma vez que a remoção de um dos pares de interação do triptofano iria alterar a dinâmica de contatos deste, modificando o ambiente químico percebido pelo outro resíduo de arginina par da interação. Não foi analisada CSP para cadeias laterais das R482 e R544 uma vez que os sinais para estes grupos não foram assinalados no trabalho original do Dr. Raphael Santa Rosa Sayegh [86], tampouco no presente projeto.

Dessa forma, os experimentos de ^1H - ^{15}N HSQC RMN com os mutantes da Cdc25B, para o W550 e para as R482 e R544, apresentaram a confirmação experimental da existência de contatos cátion- π transientes entre estes resíduos, conforme as observações retiradas das simulações de dinâmica molecular. Os resultados sugerem também uma dinâmica de contatos que se modifica na ausência

de um ou outro parceiro desta interação, como verificado pela perturbação de deslocamento químico significativamente alterada para a arginina parceira não mutada nos espectros dos contra-mutantes (R482A e R544A).

2.2.4 Dinâmica de parceiros e orientação dos contatos cátion- π

Com a confirmação experimental dos contatos entre as cadeias laterais das R482 e R544 e do W550 restou voltar-se para as simulações e compreender melhor a dinâmica das interações cátion- π entre os resíduos estudados. Analisando-se as simulações de US em janelas com distâncias longas o suficiente para não permitir contatos entre as R482 ou R544 e o W550, buscou-se por outros parceiros de interação do triptofano durante as dinâmicas moleculares.

A figura 25 mostra que, ao longo das simulações realizadas, na exclusão da possibilidade de contato do W550 com a R482 (A) as R544, R548 e R562 se revezam em interações cátion- π com o triptofano. Da mesma forma, quando se mantém o resíduo de R544 suficientemente distante do W550, as R482 e R548 passam a formar interações cátion- π com aquele resíduo.

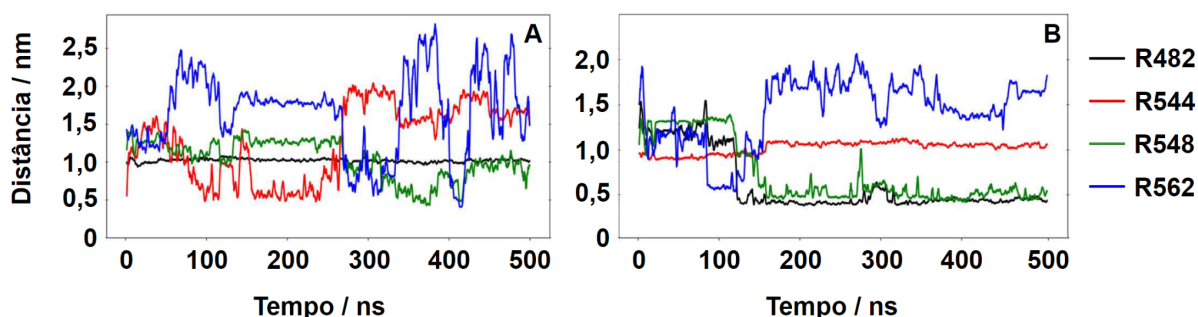


Figura 25. Trajetórias das distâncias entre o CDM da cadeia lateral do W550 e o CDM das cadeia lateral das R482 (preto), R544 (vermelho), R548 (verde) e R562 (azul), obtidas com coordenadas de referência de US em 1,0nm para as interações entre as cadeias laterais do W550 e a R482 (A) ou R544 (B).

É importante lembrar que para a R562 o perfil de energia livre traçado sugere a possibilidade de contato com o triptofano, mas sua aparência plana mostra baixos valores de energia de interação, indicando contatos pouco estáveis, com tempo de vida curto. Análises de perturbação de deslocamento químico para um mutante R562A, no entanto, com o intuito de confirmar essas interações, seriam inviáveis devido a localização deste resíduo na porção C-terminal flexível da proteína.

Com relação aos contatos observados para a R548, são esperados com frequência e possivelmente até mesmo com tempos de vida mais longos, lembrando-se, entretanto, que esse resíduo foi excluído das análises pela geração de perfis de energia livre devido à proximidade com o W550. Além disso, a mutação deste resíduo com o intuito de criar uma contra-prova para este possível contato seria, também, pouco reveladora, uma vez que a já mencionada proximidade com o W550 iria, certamente, gerar perturbações significativas de deslocamento químico neste último, as quais não poderiam ser unicamente atribuídas à extinção de contatos cátion- π com a R548 devido à óbvia alteração de ambiente químico sentida pelo triptofano ao ter um resíduo vizinho, positivamente carregado, removido.

Finalmente, foram analisadas as distribuições dos ângulos formados entre os planos definidos pelos átomos de nitrogênio dos grupos guanidino das argininas pelo anel indólico do triptofano. Nas ligações cátion- π , ângulos entre estes planos próximos a 0° representam a orientação sanduíche enquanto valores próximos a 90° representam a orientação T.

A figura 26 mostra que, conforme esperado, a curtas distâncias as R482 e R544 interagem na orientação sanduíche com o W550 enquanto com o aumento de distância a orientação da interação tende a se tornar mais próxima de “T”, porém com uma angulação variável, com máximo de frequência próximo aos 60° .

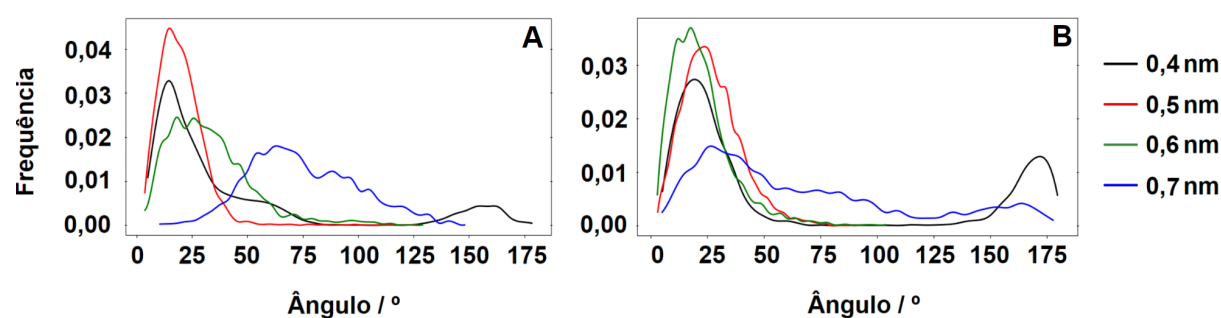


Figura 26. Distribuição dos ângulos formados entre os planos definidos pelo anel indólico do W550 e o grupo guanidino das R482 (A) e R544 (A) ao longo das simulações de US com coordenada de referência em 0,4nm (preto), 0,5nm (vermelho), 0,6nm (verde) e 0,7nm (azul).

As observações realizadas mostram um dinamismo nos contatos do triptofano, que interage, em diferentes momentos, com diferentes parceiros e, eventualmente, até mesmo com mais de uma arginina ao mesmo tempo. A duração dos contatos também varia, podendo se apresentar maior ou menor para cada parceiro, conforme indicado pelos perfis de energia livre. Essas verificações abrem

espaço para especulação da possibilidade de contatos com argininas superficiais e o papel que tal dinâmica apresenta para a proteína.

Uma das possibilidades aventadas é que o triptofano apresente uma função no auxílio da liberação de grupos fosfato removidos do substrato protéico, que apresentam grande afinidade pelos grupos guanidinos das argininas assim como pelas cadeias laterais das lisinas. Devido ao grande número de resíduos positivamente carregados presentes, uma forma de facilitar a saída do fosfato removido, com alta densidade de carga negativa, seria através da competição da interação das cadeias laterais das argininas e lisinas, não somente pelo fosfato mas também pelo anel indólico do triptofano, por exemplo, o qual estando localizado numa região flexível da proteína possui liberdade para atuar de maneira promíscua com diversos destes resíduos.

2.2.5 Conclusão

A partir do exposto, conclui-se que o campo de força CHARMM36 estima corretamente a energia de interações do tipo cátion- π entre as cadeias laterais de argininas e triptofanos.

Simulações de dinâmica molecular indicam interações favoráveis entre as cadeias laterais das R482 e R544 e do W550 presentes no domínio catalítico da Cdc25B. Mutantes W550A, R482A, R544A foram gerados, expressos e purificados com sucesso e ensaios de ^1H - ^{15}N HSQC RMN com os mutantes produzidos confirmaram as interações cátion- π previstas nas simulações computacionais.

O rompimento das interações entre os grupos guanidinos das R482 ou R544 com o anel indólico do W550 provoca alterações de ambiente químico em diversos resíduos do domínio catalítico da Cdc25B, sobretudo grande parte dos localizados na hélice F e na região C-terminal desordenada da proteína, ocorrendo também perturbações a longa distância.

Os resíduos de R482 e R544 participam da uma dinâmica de interações com o W550 que deve ser consideravelmente mais complexa envolvendo, provavelmente, resíduos como a R548 e outras argininas em interações de curta duração. A orientação de interação mais provável entre as R482 e R544 e o W550 é do tipo sanduíche, favorecida a curtas distâncias.

Reconhecimento molecular pela fosfatase Cdc25B

3.1. Metodologia e desenvolvimento

A expressão, extração e purificação da forma selvagem do domínio catalítico da Cdc25B, assim como os ensaios de cinética de inibição e de perturbação de deslocamento químico (CSP) na presença dos inibidores (relatados nas duas próximas seções) foram realizados pelo Dr. Raphael Santa Rosa Sayegh. A metodologia utilizada e os resultados serão brevemente apresentados devido à importância dos mesmos, responsáveis pelo direcionamento desta parte da pesquisa desenvolvida.

3.1.1. Expressão, extração e purificação da forma selvagem do domínio catalítico da Cdc25B

A expressão, extração e purificação da forma selvagem do domínio catalítico da Cdc25B foi realizada de forma análoga à mencionada na seção 2.1.3 do presente documento. Os relatos originais, imagens de géis e demais informações podem ser consultadas na tese de doutorado do Dr. Raphael Santa Rosa Sayegh [86].

3.1.2. Cinética de inibição enzimática

Ensaio de cinética de inibição enzimática do domínio catalítico da Cdc25B foram realizados utilizando-se três inibidores identificados na literatura [87]. Os inibidores podem ser classificados como benzoatos e são todos aniônicos, apresentando cargas formais -1, -2 e -3 em pH fisiológico. Aqui, os inibidores serão mencionados em analogia a sua carga de forma que o ânion monovalente será referido como inibidor 1 (CAS 289913-88-8), o ânion bivalente será referido como inibidor 2 (CAS 7330-46-3) e o ânion trivalente será referido como inibidor 3 (CAS5466-99-9) (figura 27).

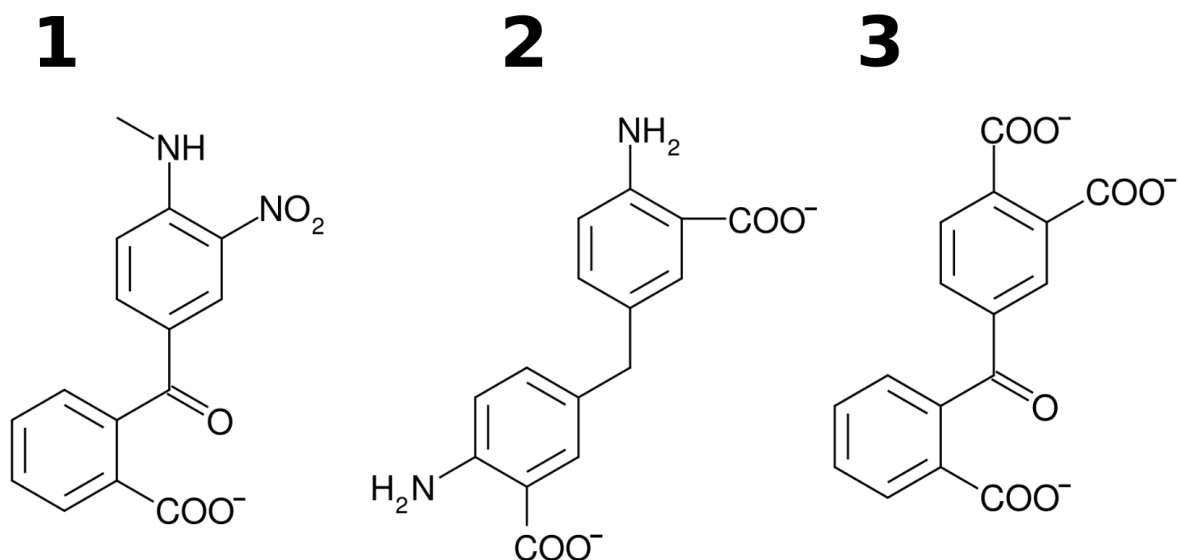


Figura 27. Inibidores utilizados nos ensaios de cinética de inibição enzimática: I1 (1 - CAS289913-88-8). I2 (2 - CAS7330-46-3) e I3 (3 - CAS5466-99-9).

Os ensaios foram realizados em tampão MES (20mmol.L⁻¹ ácido 2-(N-morfolino)etanosulfônico - MES, 50mmol.L⁻¹ NaCl, 0,005% Tween-20, 1mmol.L⁻¹ ditioneitol, pH = 6,7) com substrato fluorescente fosfato de 3-O-metil fluoresceína (OMFP) em concentrações variáveis (0,94, 1,88, 3,76, 7,50, 15,0 e 30,0 mmol.L⁻¹) por período de 60 minutos. As concentrações dos inibidores também foram variadas a depender do ensaio realizado. Medidas de fluorescência em comprimento de onda 485nm foram realizadas a cada 5 minutos durante o ensaio e os valores foram coletados e guardados para posterior análise.

3.1.3. Aquisição de espectros de ¹H-¹⁵N - HSQC RMN e análises de perturbação de deslocamento químico na presença dos inibidores testados

A aquisição de espectros de ¹H-¹⁵N - HSQC RMN na presença dos inibidores testados foi realizada de forma análoga à mencionada na seção 2.1.4 do presente documento, com exceção da presença dos inibidores I1, I2 e I3 na solução analisada. As quantidades dos inibidores utilizados foi sempre consideravelmente superior aos valores de K_i estimados de forma que estas fossem suficientes para garantir a máxima formação de complexos enzima-inibidor ao longo do ensaio.

Para definição da perturbação de deslocamento químico (*Chemical Shift Perturbation* - CSP) foi utilizada a seguinte fórmula:

Equação 10:
$$CSP = \sqrt{(\Delta\delta H)^2 + 0,15(\Delta\delta^{15}N)^2}$$

Esta fórmula é uma fórmula genérica, comumente utilizada para análise de perturbação de deslocamento químico em espectros de ^1H - ^{15}N - HSQC RMN [88] e, diferentemente da fórmula expressa na equação 9, pode ser utilizada sem prejuízo da presente análise uma vez que nesta não são consideradas cadeias laterais de resíduos de aminoácidos mas apenas sinais amídicos de *backbone* proteico, com perturbação bem identificada a partir da equação 10.

3.1.4. Desenho dos inibidores e ensaios de ancoragem molecular

Os inibidores I1, I2 e I3 foram desenhados utilizando-se o programa Molden [89] e salvos no formato .pdb [90]. Os estados de protonação dos grupos químicos protonáveis foram definidos por analogia aos de moléculas com valores de pKa definidos na literatura. Para os inibidores 1 e 3 definiu-se o grupo benzoato desprotonado a partir do pKa do ácido benzóico (pKa = 4,19 [91]). O grupo amino do inibidor 1 foi definido como desprotonado a partir do pKa da 2-nitroanilina (pKa = -0,28 [92]). Para o inibidor 2 definiu-se o grupo benzoato e a amina desprotonados a partir dos pKas dos grupos equivalentes no ácido antranílico (respectivamente, pKa = 2,17 e pKa = 4,85 [93]). As estruturas dos inibidores desenhados foram otimizadas utilizando-se o método/ base MP2/6-311+G(d) [67] e o programa Gaussian 09 rev. A [66].

Para os ensaios de ancoragem molecular utilizou-se duas estruturas distintas da Cdc25B, cada uma com um contato cátion- π considerado importante, conforme discutido no capítulo anterior: cátion- π entre o W550 e as R482 e cátion- π entre o W550 e as R544. Ambas estruturas foram retiradas da simulação longa (6 μ s) realizada para efeito de definição das argininas candidatas em interação com o triptofano, conforme discutido nas seções 2.1.1 e 2.2.2 sendo a estrutura com o contato cátion- π W550 - R482 bastante semelhante à estrutura cristalográfica 1QB0 depositada [23], a qual não foi utilizada para os ensaios de *docking* molecular devido a ausência dos 16 últimos resíduos da porção C-terminal da proteína.

Os ensaios de ancoragem molecular foram realizados para os três inibidores utilizando-se o programa AutoDock Vina 4.2 [95]. Para o processamento inicial das estruturas utilizou-se a interface gráfica AutoDock Tools (ADT) [96] e nesta etapa definiu-se a flexibilidade de cada ligante a partir da estrutura desenhada, previamente fornecida, foram adicionados os átomos de hidrogênio às estruturas da Cdc25B utilizadas e definidos os *grids* de ancoragem (foram utilizados *grids* grandes, de 20x20x20Å).

3.1.5. Parametrização de campo de força e conferência da parametrização dos pequenos ligantes

Os parâmetros de campo de força para os pequenos ligantes foram definidos utilizando-se o programa CHARMM-GUI [97], a partir da disponibilização dos arquivos .pdb previamente gerados.

A conferência da parametrização foi feita observando-se os valores de “penalidade de parâmetro” fornecidos no documento gerado pelo CGenFF/CHARMM-GUI e utilizando estes como guia para geração de dímeros ligante - H₂O, utilizando os grupos com alta penalidade como aceptores e/ ou doadores de H em ligação de hidrogênio.

As distâncias entre os átomos envolvidos na ligação de H foram otimizadas utilizando-se o método/ base HF/6-31G(d) mantendo-se os átomos do ligante congelados. Com a distância ótima definida foram ajustados comprimentos de ligação H - O e o0 ângulo de ligação H - O - H na molécula de água do dímero de acordo com os parâmetros para o modelo TIP3P [64] (0,9572Å e 104,52°). Foram então realizados cálculos de energia (*single point*) para o ligante isolado, para a molécula de água TIP3P e para os dímeros e as energias de interação calculadas utilizando-se o método/ base HF/6-31G(d) e o campo de força CHARMM36 com os parâmetros fornecidos para que, então os resultados fossem comparados.

3.1.6. Simulações de dinâmica molecular

Simulações de dinâmica molecular do domínio catalítico completo da Cdc25B foram realizadas a partir das estruturas utilizadas para os ensaios de ancoragem molecular (cátion- π W550 - R482 e cátion- π W550 - R544). Os ligantes retirados

dos ensaios de ancoragem tiverem os átomos de hidrogênio adicionados na posição em que se encontravam, o modelo foi solvatado em uma caixa de simulação dodecaédrica com 12612 moléculas de água e íons Na^+ e Cl^- foram adicionados para mimetizar as condições experimentais de aquisição de espectros de RMN em proporção adequada, considerando-se a neutralidade do sistema e a presença dos inibidores aniônicos.

O sistema foi relaxado e equilibrado durante rampas de aquecimento com 2ns de duração (a 5, 50, 150 e 250K por 500ps cada). Foi utilizado o campo de força CHARMM36 [68] para a proteína e íons presentes e o modelo TIP3P para as moléculas de água [71]. As simulações foram realizadas utilizando-se passo de integração de 2fs, 300K de temperatura e pressão de 1bar (*ensemble* NPT), termostato de Bussi [72] e barostato de Parrinello-Rahman [73], ambos com período de 0,5ps. Para o método de PME foi utilizado *cutoff* de espaço real de 1,2nm e espaçamento de Fourier de 0,12nm para interações eletrostáticas de longo alcance. Todos os cálculos com campos de força clássicos, simulações de dinâmica molecular e análises foram realizados com o programa GROMACS 4.6.7 [74].

3.2. Resultados e discussão

3.2.1. Cinética de inibição enzimática

As análises dos mecanismos de inibição e das constantes de inibição (K_i) aqui apresentadas foram realizadas com auxílio do professor Dr. Sandro Roberto Marana. Os resultados dos experimentos de inibição executados pelo Dr. Raphael Santa Rosa Sayegh e analisados pelo autor da tese com auxílio do professor Dr. Sandro Roberto Marana são apresentados na figura 28, linearizados num gráfico do tipo duplo-recíproco de Lineweaver-Burk.

A partir do formato dos gráficos, com intersecção das curvas com o eixo das ordenadas em pontos diferentes para I1 e em pontos muito próximos para I2 e I3, é possível observar mecanismo de inibição não-competitivo para I1 e competitivo para I2 e I3. Os valores calculados para K_m e $V_{\text{máx}}$ para a enzima livre foram de, respectivamente, $5,4\mu\text{mol.L}^{-1}$ e $1,6\mu\text{mol.min}^{-1}$. Os inibidores apresentaram valores de K_i de $2,8\mu\text{mol.L}^{-1}$ (I1), $8,0\mu\text{mol.L}^{-1}$ (I2) e $2,2\mu\text{mol.L}^{-1}$ (I3).

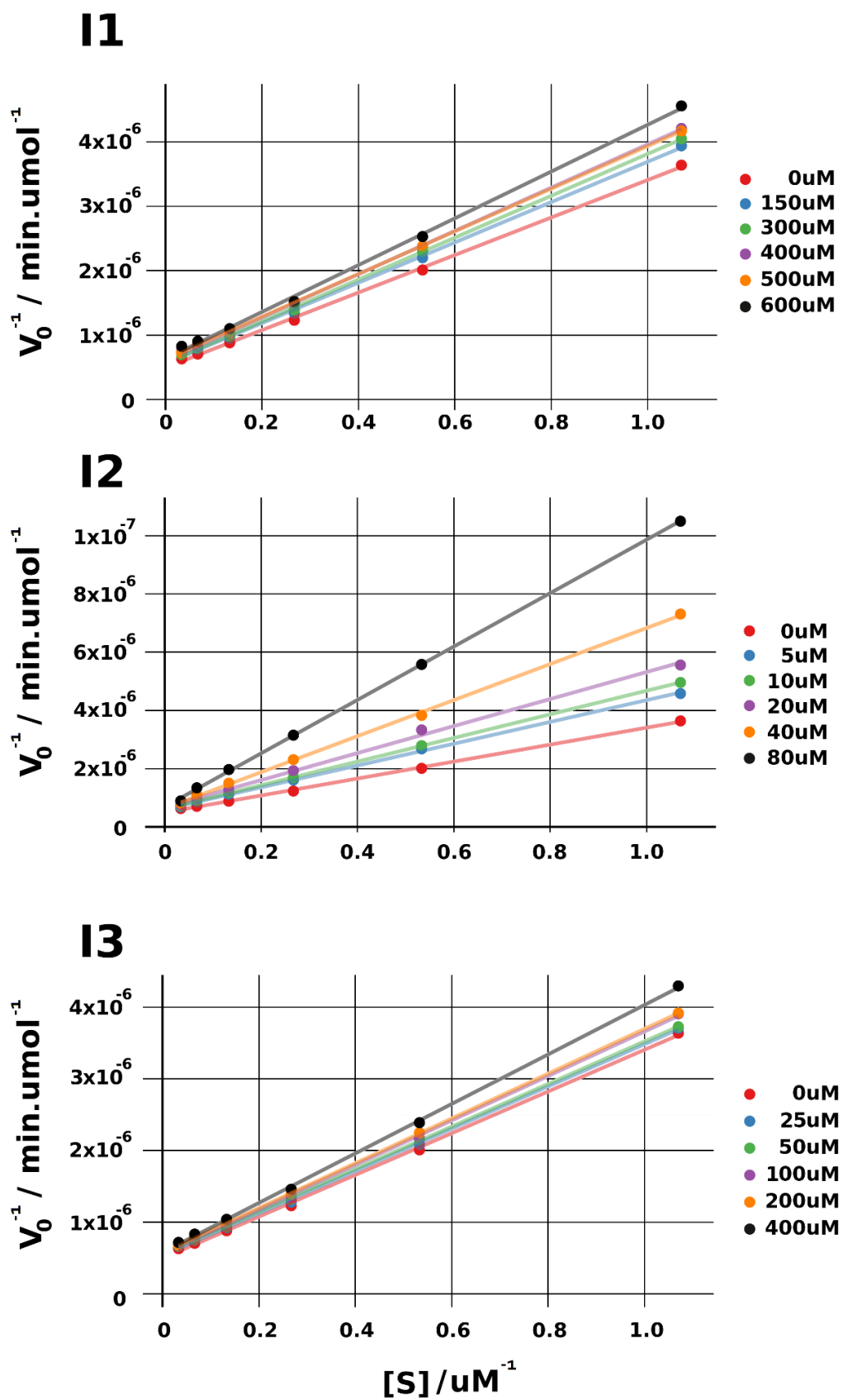


Figura 28. Gráfico duplo-recíproco para os dados obtidos durante os ensaios de cinética enzimática dos inibidores I1, I2 e I3, da Cdc25B.

Do ponto de vista molecular, as justificativas para os mecanismos de inibição observados podem ser dadas a partir da análise das cargas dos inibidores. Os inibidores 2 e 3 são ânions, respectivamente, bi e trivalentes, ou seja, com considerável densidade de carga. As argininas 427, 479, 482 e 544 e as lisinas 509 e 512 são todos resíduos positivamente carregados localizados muito próximos à cisteína catalítica da Cdc25B e podem interagir com qualquer um dos ligantes aniônicos, estabilizando-os, sendo que esta tendência deve ser maior para os inibidores que apresentam maior carga líquida. Dessa forma, não é de se estranhar que os inibidores aniônicos possuam interação com resíduos próximos do *P-loop*, ocluindo o sítio catalítico enzimático em algum momento. De fato é sabido que nas fosfatases a grande quantidade de resíduos positivamente carregados auxiliam na estabilização de grupos fosfatos no sítio catalítico e em suas proximidades e a interação com os inibidores testados é um reflexo da natureza química do bolsão catalítico.

O inibidor 1, por sua vez, pode interagir preferencialmente com resíduos carregados positivamente localizados na superfície da Cdc25B que, lembrando, possui mais de 20 outras argininas e lisinas distantes do sítio catalítico. Conforme discutido no capítulo anterior, sabe-se que a formação ou rompimento de interações entre resíduos da proteína podem transmitir alterações estruturais à distância, capazes de modificar a capacidade catalítica de uma enzima. Uma explicação mais adequada para o mecanismo real de inibição pode envolver outras possibilidades e até mesmo uma combinação das possibilidades aqui apresentadas para ambos mecanismos com diferentes pesos para cada um.

O exposto até o momento permite assumir que o inibidor 1 deve interagir com resíduos em regiões diversas da proteína, atuando como um inibidor não competitivo, enquanto os inibidores 2 e 3 devem interagir, sobretudo, com resíduos de lisina e arginina localizados próximos do sítio catalítico enzimático, ocluindo este e atuando como inibidores competitivos.

3.2.2. Perturbação de deslocamento químico

Os ensaios de ^1H - ^{15}N HSQC RMN e as análises realizadas apontaram perturbação de deslocamento químico (CSP) significativa em diversos resíduos do polipeptídeo, distribuídos ao longo de várias regiões, distantes entre si. Os três

perfis de perturbação observados, para os três inibidores testados, são semelhantes, o que, apesar da diferença entre as formas de inibição verificadas para o inibidor 1 (não-competitiva) e os inibidores 2 e 3 (competitiva) não é surpreendente considerando-se a natureza química semelhante dos três inibidores.

A figura 29 mostra uma representação da estrutura da proteína com um mapa de CSP indicando regiões significativamente perturbadas nos ensaios de RMN na presença dos inibidores. Pode-se observar CSP significativas próximas ao N-terminal da proteína, nas hélices A e em suas proximidades e nas hélices D e F (hélice C-terminal). A figura 29 apresenta a proteína invertida em relação ao sentido N - C - terminal utilizado até o momento neste texto, tal representação foi escolhida para facilitar as visualizações dos ensaios de ancoragem mostrados adiante.

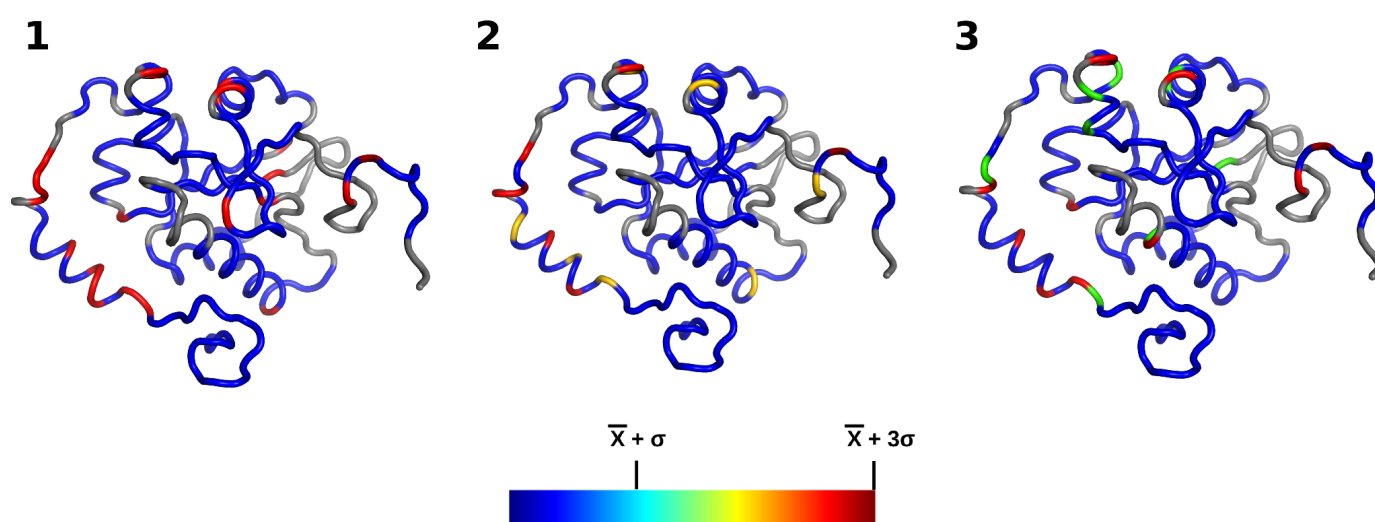


Figura 29. Representação de CSP na Cdc25B para os ensaios de ^1H - ^{15}N HSQC RMN com os inibidores I1 (1), I2 (2) e I3 (3).

Tabela 5. Resíduos significativamente perturbados para cada ensaio de ^1H - ^{15}N HSQC RMN com o respectivo inibidor (I1, I2 ou I3) em relação à proteína na ausência de qualquer inibidor. São indicados por código o N-terminal, **Lisinas/Argininas**, hélice A, **hélice D** e **hélice F**.

	Resíduos significativamente perturbados
I1	369, 370, 373, 375, 377, 394 , 396, 398, 404, 407, 408, 419, 428, 430, 433, 436, 437, 446, 468, 471, 493, 505, 509 , 511, 513 , 515, 516, 518, 519, 521, 531, 532, 533, 537 , 538 , 540 , 543 , 544 , 545 , 547, 548 , 549, 551, 562
I2	370, 371, 375, 377, 394 , 404, 407, 408, 428, 430, 482 , 492 , 513 , 516, 518, 519, 521, 522, 531, 532, 533, 534 , 537 , 540 , 543 , 544 , 545 , 547, 548
I3	370, 371, 375, 377, 394 , 404, 407, 418, 428, 430, 468, 473 , 482 , 483 , 492 , 513 , 514, 516, 518, 519, 521, 522, 532, 533, 538 , 540 , 543 , 545 , 547, 548 , 550, 551, 552

Interessantemente as análises de CSP para os ensaios com os inibidores demonstram padrões que corroboram com as observações dos ensaios de cinética de inibição enzimática. O ensaio com o I1, o inibidor que apresentou cinética de inibição não competitiva, foi o ensaio que revelou o maior número de resíduos perturbados, uma quantidade quase 35% superior à do ensaio que ocupa a segunda posição (44 resíduos para I1 frente 33 para I3) e mais de 50% superior ao ensaio que menos apresentou resíduos significativamente perturbados (I2, com 29 resíduos). Esta observação sugere falta de especificidade de interação entre o inibidor 1 e a proteína, aparentando poder haver complexação entre ambos de maneira inespecífica, em diversas regiões da proteína.

Os I2 e I3 apresentaram menos resíduos significativamente perturbados mas uma proporção muito maior que do I1 para resíduos próximos ao sítio catalítico enzimático, onde espera-se que estes inibidores competitivos estejam complexados, também indo ao encontro do esperado para o mecanismo de inibição sugerido para estes inibidores.

É importante observar, também, que uma parte considerável dos resíduos positivamente carregados da proteína apresentaram alterações significativas de deslocamento químico, sendo que dos 25 resíduos de lisina ou arginina anotados na proteína, para os inibidores I1, I2 e I3, respectivamente, 7, 7 e 5 resíduos mostraram CSP significativa. O ensaio com o I3, o inibidor com a maior carga negativa, foi o que menos resultou em resíduos básicos significativamente perturbados mas deve-se observar que o ensaio com o I3 foi o que resultou no maior número de resíduos perturbados na região próxima ao *P-loop*, novamente em acordo com a observação do mecanismo de inibição competitiva.

3.2.3. Ancoragem e dinâmica molecular

A partir da análise dos resultados dos ensaios de cinética de inibição enzimática e CSP na presença dos inibidores, foram definidos sítios candidatos de ligação, para as pequenas moléculas, no domínio catalítico da Cdc25B. Considerando-se que dois dos inibidores (I2 e I3) apresentaram mecanismo de inibição competitiva e o terceiro inibidor (I1) é bastante semelhante a estes, o sítio catalítico enzimático foi definido como o primeiro sítio de interesse para os ensaios de ancoragem molecular e posterior simulação de dinâmica molecular.

Sendo assim, a partir da cisteína catalítica presente no *P-loop* proteico, definiu-se uma região esférica com raio de 8Å (o comprimento médio dos inibidores testados) e os resíduos que se apresentaram significativamente perturbados dentro desta região foram excluídos para fins de definição de novos sítios de interesse para realização do ensaio de ancoragem e simulação molecular. Dessa forma, três outras regiões da proteína, distantes entre si e do sítio catalítico, apresentaram-se com resíduos significativamente perturbados e foram consideradas para os ensaios seguintes.

As novas regiões definidas foram nomeadas S1 (sítio 1), S2 (sítio 2) e S3 (sítio 3) de forma que a ordem apresentada na numeração dos sítios corresponde à posição em que os resíduos que compõem estas regiões, no sentido N-Terminal para o C-Terminal, aparecem na sequência proteica. A figura 30 apresenta a estrutura da proteína conforme utilizada na presente seção, os sítios 1, 2 e 3 definidos e destaca algumas argininas importantes, o W550, o C-terminal flexível e o sítio catalítico enzimático. É importante mencionar que, considerando-se a realização de dinâmicas moleculares posteriores, os ensaios de *docking* visaram principalmente apresentar poses iniciais para as simulações realizadas.

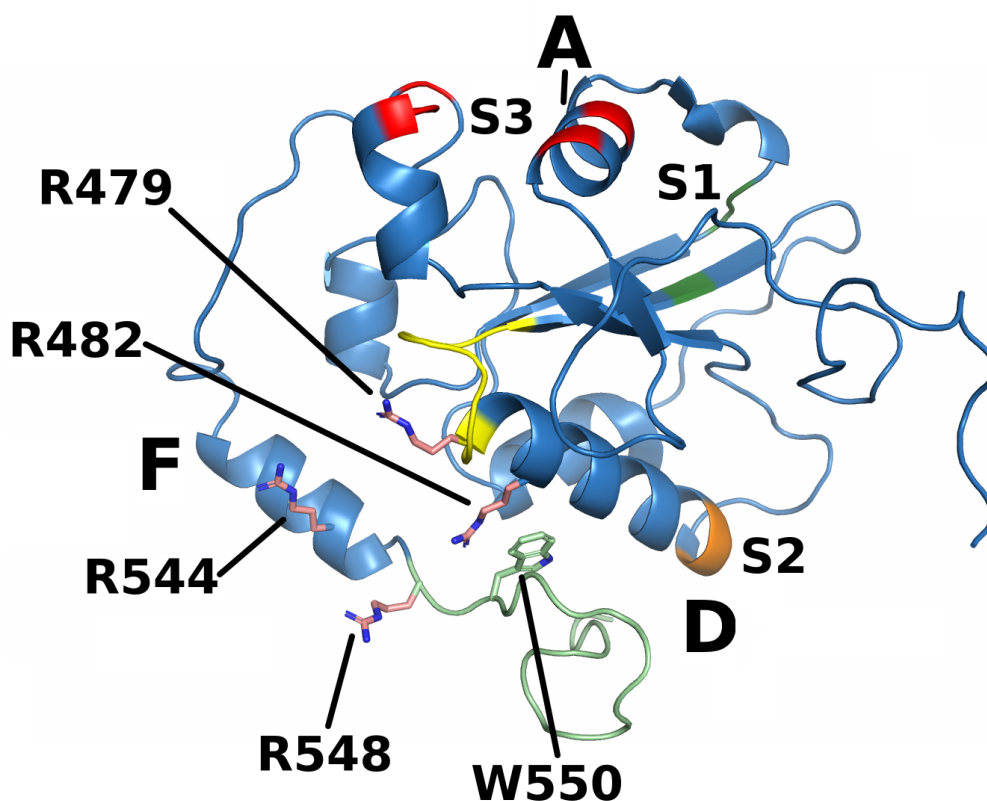


Figura 30. Sítios definidos para os ensaios de *docking* molecular, S1, S2 e S3 e sítio catalítico (em amarelo). São também mostrados resíduos importantes e a região C-terminal flexível (em verde). As hélices A, D e F são destacadas.

Dessa forma, quatro regiões foram definidas para realização dos ensaios de ancoragem molecular: as S1, S2 e S3 e o sítio catalítico enzimático. Os *grids* de ancoragem foram centralizados no ponto médio definido pelos resíduos que compõem os sítios 1 a 3 e no átomo de enxofre da cisteína catalítica para o sítio catalítico.

3.2.3.1. Sítios S1, S2 e S3

Para os sítios S1, S2 e S3 foram somente analisados o resultado mais bem ranqueado dentro dos *dockings* fornecidos pelo AutoDock Vina [95]. A figura 31, e as seguintes, apresentam montagens em que os quadros superiores mostram a localização do inibidor (I1, I2 ou I3), em verde, posicionado na estrutura da proteína colorida de acordo com os códigos utilizados até o presente momento porém no sentido inverso do utilizado até então ao longo do presente texto (aqui o C-terminal se encontra à esquerda).

A alteração do sentido da proteína para apresentação das poses obtidas no S1 se fez necessária devido ao posicionamento deste sítio, localizado na parte posterior da proteína quando observado no sentido até aqui utilizado. Os quadros inferiores foram montados a partir de modificações da análise de contatos realizada pelo programa LigPlot+ [97] e auxiliam na observação dos contatos proteína-ligante: resíduos indicados como círculos vermelhos indicam proximidade e interação por efeito hidrofóbico enquanto os resíduos mostrados na forma de bolas e bastões indicam interações de fato com a pequena molécula.

Os ensaios de *docking* molecular para os inibidores testados no S1 mostram o I1 ancorado numa região próxima aos resíduos de S415, V418, F421 e V422, o I2 próximo aos resíduos de G412, S415, F421 e V422 e o I3 próximo aos resíduos de G412, F414, S415, V418, F421 e T439. Nenhum dos resíduos, entretanto, foi indicado como interagindo diretamente com qualquer um dos inibidores por contatos específicos (como através de ligações de hidrogênio ou de pontes salinas, por exemplo).

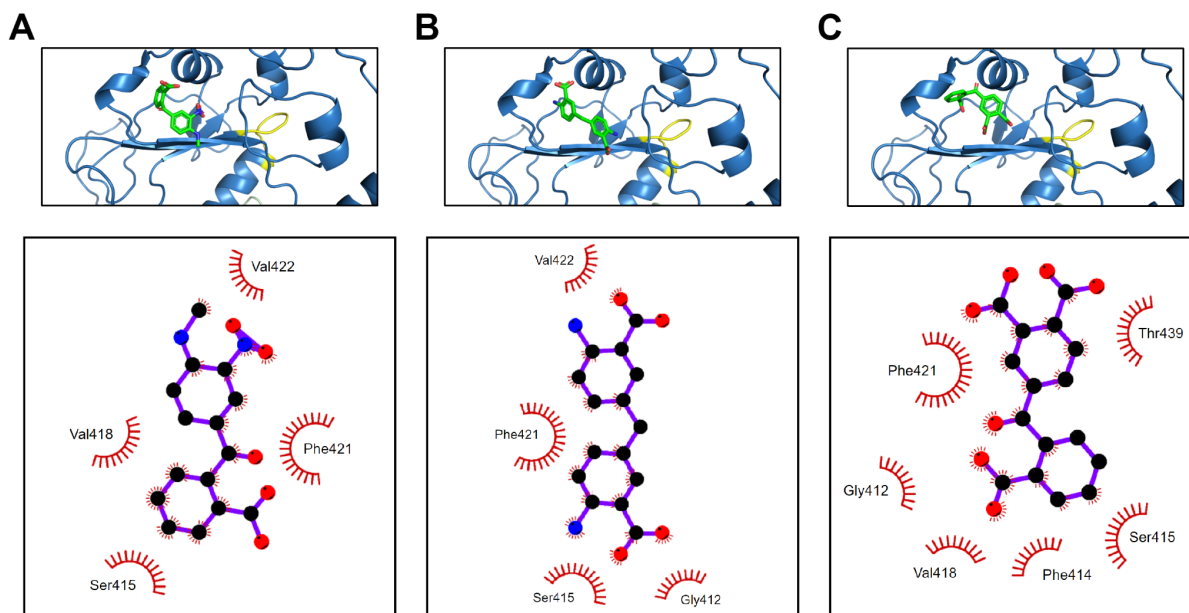


Figura 31. Poses (quadros superiores) e contatos (quadros inferiores) obtidos a partir dos ensaios de *docking* no S1 para os inibidores I1 (A), I2 (B) e I3 (C).

Para os inibidores I1 e I2 não ocorrem resíduos apresentados próximos aos ligantes que se encontram significativamente perturbados nos ensaios de ^1H - ^{15}N HSQC RMN. Para o I3 um único resíduo, próximo do inibidor no sítio de ancoragem, apresentou CSP significativa, a V418.

Já os ensaios realizados para o S2 mostram o I1 ancorado numa região próxima aos resíduos de R488, E489, R492, A493 e V494, com interações diretas entre o grupo guanidino da R492 e o grupo nitro do I1 e entre o N amídico do *backbone* do resíduo de A493 e o carboxilato do inibidor. O I2 é ancorado próximo aos resíduos de R482, R488 e D489 com interações diretas entre as argininas 488 e 492 e um dos carboxilatos desta pequena molécula. O I3 é ancorado próximo aos resíduos de L398, C484, R485, R488, D489, R492 e M505 apresentando interações entre as três argininas, 485, 488 e 492, e dois dos três carboxilatos presentes em lados opostos do inibidor, formando uma ponte entre a proteína e este.

Para o I1, dentre os resíduos apresentados como próximos ou interagindo com o inibidor, a A493 foi verificada com CSP significativa. Os I2 e I3 são indicados como interagindo com a R492, também observada com variação significativa de CSP.

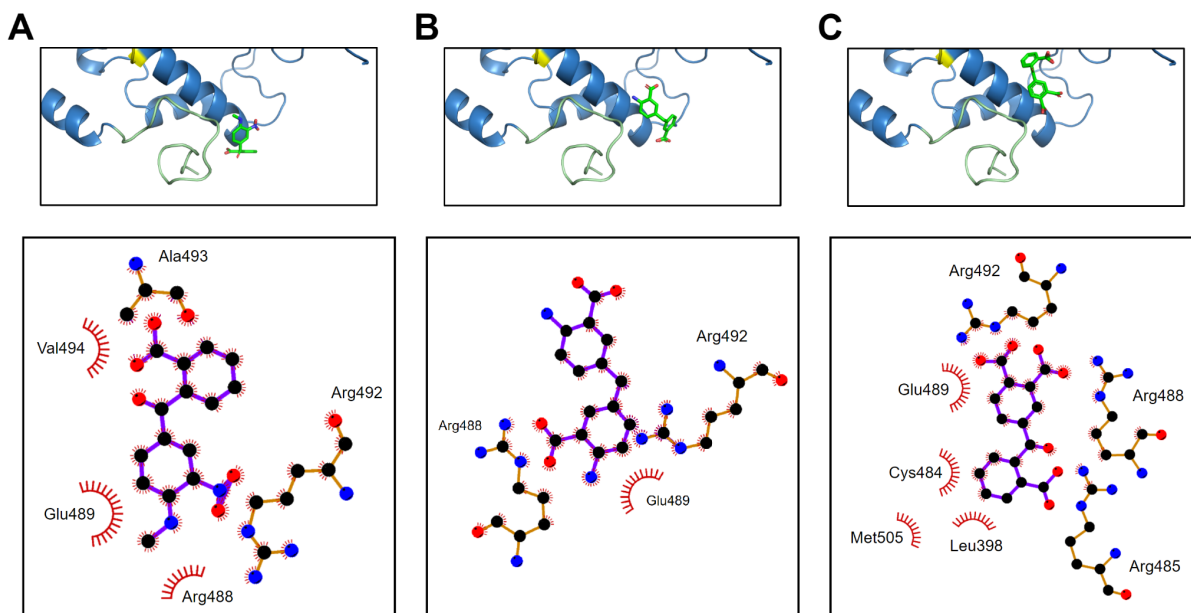


Figura 32. Poses (quadros superiores) e contatos (quadros inferiores) obtidos a partir dos ensaios de *docking* no S2 para os inibidores I1 (A), I2 (B) e I3 (C).

Finalmente os ensaios realizados para o S3 mostram o I1 ancorado numa região próxima aos resíduos de P403, E404, F515, P517, Q518 e H519, com interações diretas entre a carbonila amídica da Q518 e o grupo nitro do I1 e entre átomos do anel imidazólico da H519 e o carboxilato do ligante. O I2 é ancorado próximo aos resíduos de K513, F516, P517, Q518 e P520 com interações diretas entre as carbonilas amídicas da P517 e da K513 e os grupos amina nos anéis opostos do inibidor. O I3 é ancorado próximo aos resíduos de P403, E404, Q518, H519 e P520, apresentando interações entre o anel imidazólico da H519 e o carboxilato solitário do inibidor e entre os grupos NH e CO amídicos da Q518 e um dos carboxilatos do lado oposto da molécula.

Este conjunto de poses de ancoragem apresentou grande quantidade de resíduos indicados como próximos aos inibidores, ou interagindo com estes, e significativamente perturbados nos ensaios de RMN. Para o I1, o D404 e as F515, Q518 e H519 foram verificadas com CSP significativa. Para o I2 as K513, F516 e Q518 demonstraram CSP significativa. Finalmente, para o I3 as CSP do E404 e das Q518 e H519 se mostraram significativas.

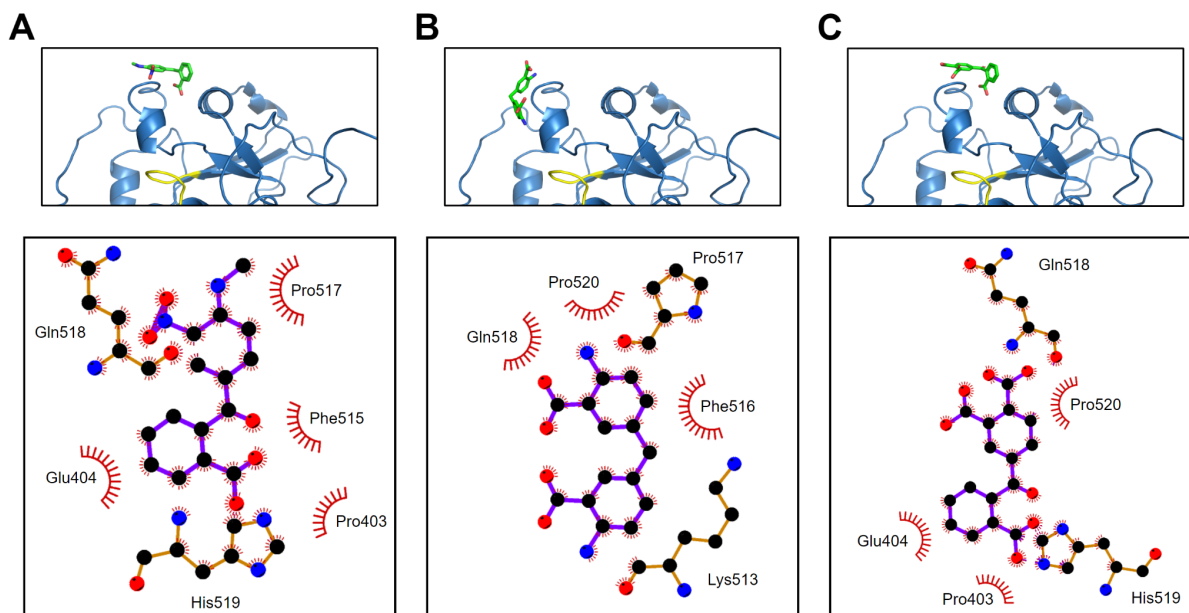


Figura 33. Poses (quadros superiores) e contatos (quadros inferiores) obtidos a partir dos ensaios de *docking* no S3 para os inibidores I1 (A), I2 (B) e I3 (C).

Assim, pode-se observar que os ensaios de *docking* molecular apontaram sítios de ancoragem em que os inibidores aparecem próximos ou interagindo, na maioria dos casos, com resíduos de aminoácidos que se apresentaram significativamente perturbados nos ensaios de ^1H - ^{15}N HSQC RMN. Algumas das interações que ocorrem podem ser consideradas inespecíficas, dando-se entre átomos do grupo amídico da cadeia principal de aminoácidos, independentemente da natureza química do aminoácido em questão, e os inibidores estudados. No caso do S2, as argininas 488 e 492 se mostraram muito importantes para os contatos e no caso do S3 os resíduos de Q518 e H519 também sendo que, com exceção da R488, todos estes resíduos apresentaram CSP significativo para os ensaios de RMN com os inibidores.

Uma vez definidas as poses iniciais para os ligantes e montados os sistemas foram realizadas simulações curtas (50ns) para cada um. As simulações foram realizadas em triplicatas, variando-se a velocidade inicial das partículas que compõem cada sistema. A figura 34 indica os inibidores nos sítios de ligação apresentados com as cores que definiram cada sítio na figura 30. Abaixo das imagens representando a proteína e os ligantes ancorados são mostradas as análises de *Root Mean Square Deviation* (RMSD) para os ligantes ao longo das trajetórias com cada curva traçada em determinada cor indicando o inibidor em questão no sítio da cor correspondente. Os RMSDs mostrados são referentes a uma

única réplica mas representam a tendência geral das três réplicas simuladas para cada sistema.

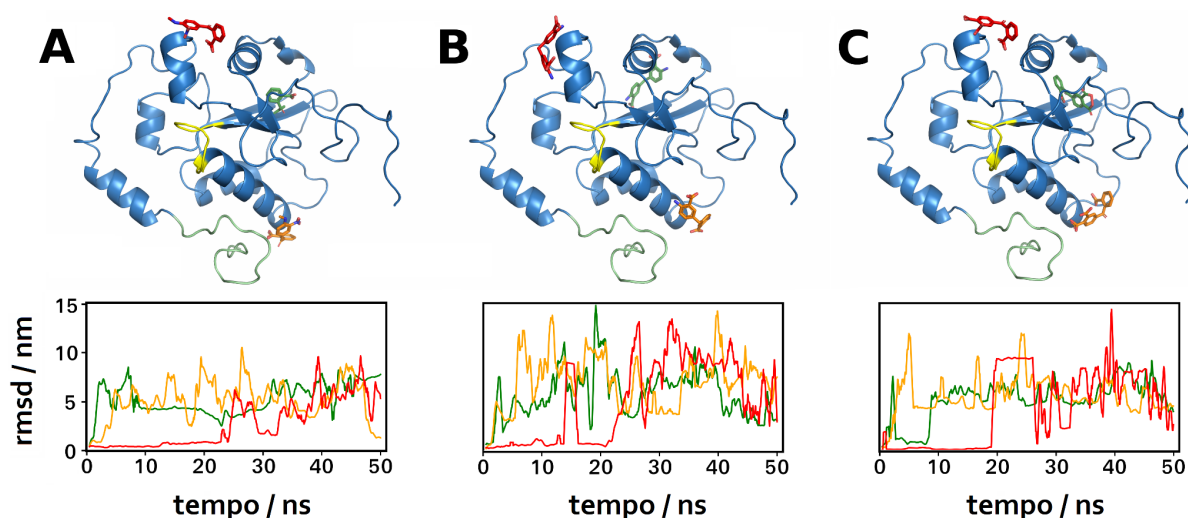


Figura 34. RMSDs para os inibidores ao longo das trajetórias simuladas nos sítios S1 (verde), S2 (laranja) e S3 (vermelho) para os inibidores I1 (A), I2 (B) e I3 (C).

As simulações curtas indicaram, em todos os casos, perfis dissociativos por parte dos inibidores em relação ao sítio de ancoragem. Com exceção das simulações iniciadas com os inibidores no S3, ocorre a dissociação destes logo nos primeiros instantes da trajetória. Dessa forma, apesar das perturbações observadas, muitas vezes condizentes com os resíduos apontados como pares de interação nos ensaios de ancoragem, os sítios de *docking* se apresentam como sítios não específicos, em que ocorrem interações de curta duração entre resíduos de aminoácidos locais e os inibidores testados. Essa observação corrobora os mecanismos observados para os inibidores: é esperado que os inibidores competitivos, I2 e I3, interajam de maneira estável apenas no sítio catalítico enzimático. Por outro lado, a possibilidade de interações pouco favoráveis, mesmo que curtas, entre o I1 e sítios de interação inespecíficos na superfície da proteína (aqui cabe lembrar que o I1 foi o inibidor que apresentou mais resíduos significativamente perturbados nos ensaios de ^1H - ^{15}N HSQC RMN) pode causar alterações estruturais consideráveis, capazes de diminuir a eficiência catalítica da Cdc25B.

3.2.3.2. Sítio catalítico: Cátion- π R482 x W550

Para o sítio catalítico enzimático foram analisadas as três poses com maior diferença de RMSD dentre as cinco primeiras ranqueadas (sempre incluindo a primeira pose) fornecidas pelo AutoDock Vina [95]. A figura 35, e as seguintes, apresentam montagens em que os quadros superiores mostram a localização do inibidor na pose melhor ranqueada (I1, I2 ou I3), em verde, posicionado na estrutura da proteína colorida de acordo com os códigos utilizados até o presente momento. Os quadros inferiores foram montados a partir de modificações da análise de contatos realizada pelo programa LigPlot+ [97] e auxiliam na observação dos contatos proteína-ligante: novamente, os resíduos indicados como círculos vermelhos indicam proximidade e efeito hidrofóbico enquanto os resíduos mostrados na forma de bolas e bastões indicam interações de fato com a pequena molécula.

Nesta seção os resultados serão mostrados em ordem diferente da seção anterior de forma que os ensaios de RMSD sempre virão juntos dos quadros com as poses e interações. A escolha da forma de apresentação destes resultados se dá pela maior quantidade de simulações e pelo fato de que a grande maioria das trajetórias revelaram os ligantes estáveis no sítio de ancoragem. São mostrados sempre os RMSD de uma das réplicas simuladas, quando esta representa a tendência geral observada, mas as simulações foram sempre realizadas em triplicatas para cada pose. A cor da linha representando o RMSD de uma das réplicas se refere ao número da pose analisada (1 - preto, 2 - vermelho, 3 - azul).

As poses escolhidas a partir dos ensaios de *docking* para o I1 no sítio catalítico da Cdc25B na estrutura com a interação cátion- π entre a R482 e o W550 mostram interações com a S477 e R479 e proximidade com a R482 (pose 1), interações com a R482 e T547 e proximidade com as R479, R544 e R548 (pose 2) e interações com as R479, R482 e R547 e com o W550 e proximidade com a R548 (pose 3). Dentre os resíduos mencionados, as R544 e R548 e a T547 aparecem significativamente perturbadas nos ensaios de RMN.

As análises de RMSD para o I1 indicam estabilidade deste em relação ao sítio de interação ao longo de toda trajetória simulada, apesar de ocorrerem alguns giros e trocas de grupos de interação entre o ligante e a proteína (observações visuais tiradas da trajetória).

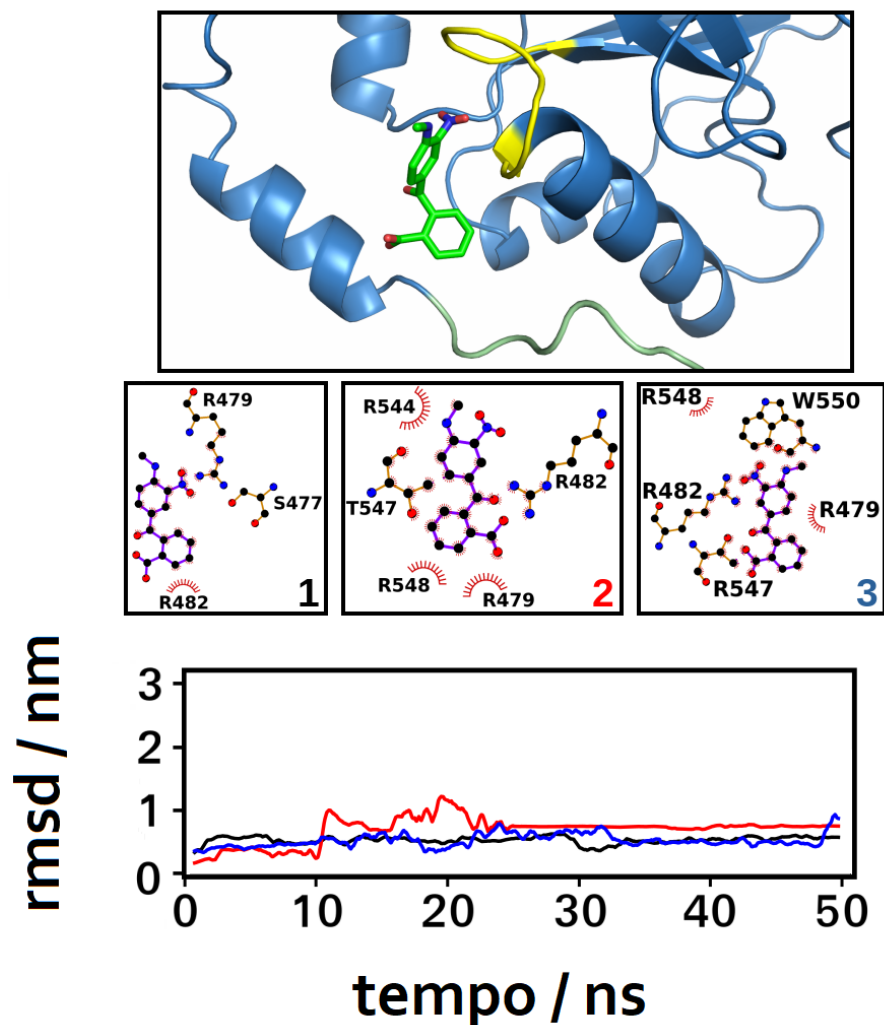


Figura 35. *Docking* molecular e simulações para o inibidor 1. É mostrada a pose mais bem ranqueada (quadro superior), contatos ligante - proteína (quadros centrais) para as poses 1, 2 e 3 e RMSD para o ligante nas simulações realizadas a partir dos sistemas montados com a pose 1 (preto), 2 (vermelho) ou 3 (azul). A estrutura inicial apresentava a interação cátion- π entre a R482 e o W550.

Foram realizadas análises de distância entre os centros de massa dos grupos químicos do ligante e das cadeias laterais dos resíduos de aminoácidos próximos do sítio de ancoragem. Em todas as análises de distâncias mostradas a seguir foram selecionados os resíduos observados em contatos mais importantes com o ligante. As análises de distância mostradas se referem sempre somente às réplicas simuladas para uma pose, o que foi definido devido ao grande número de simulações.

A figura 36 mostra as análises de distância para as três réplicas simuladas para a pose 1, aqui selecionada para ilustrar algumas das situações características observadas ao longo das réplicas produzidas para esta pose. É possível observar

que os principais contatos ao longo da trajetória não refletem os contatos identificados no ensaio de *docking*, por exemplo, a S477 e a R479 não representam exatamente interações importantes para a manutenção do ligante no sítio. Ao longo da réplica 1 é possível observar a formação e manutenção de um contato estável entre o grupo nitro do ligante e a R544 (quadro A, em verde), a qual é apenas apresentada próxima ao ligante em uma das poses selecionadas. A R479, que interage com o carboxilato do inibidor nesta pose, realiza contatos de curta duração diversas vezes ao longo da trajetória.

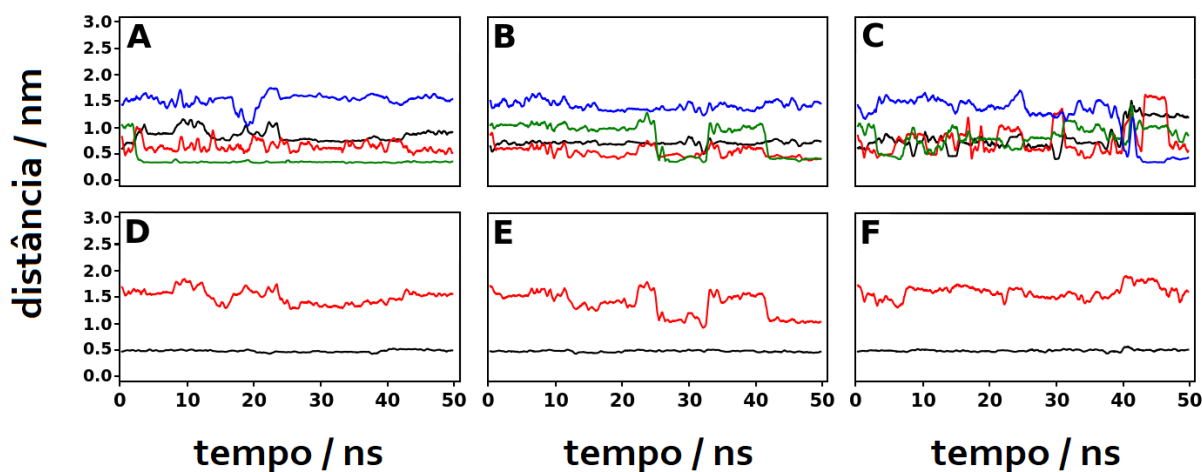


Figura 36. Análises de distâncias das trajetórias geradas a partir da pose 1 do inibidor 1 com o *docking* realizado sobre a estrutura com a interação cátion- π entre W550 e R482. Nos quadros acima têm-se: Preto: NOOxR479; Vermelho: COOxR479; Verde: NOOxR544; Azul: COOxR544; réplicas 1 (A), 2 (B) e 3 (C). Nos quadros abaixo têm-se: Preto: W550xR482; Vermelho: W550xR544; réplicas 1 (D), 2 (E) e 3 (F).

A trajetória simulada a partir da pose 2 mostra uma maior importância da interação entre a R479 e o grupo carboxilato na manutenção do inibidor no sítio catalítico. Nessa réplica é possível observar um giro do ligante seguido de um curto período de interação entre a R544 e o grupo nitro do inibidor que logo se acaba, sendo seguido de um novo giro e formação de um novo contato (período entre, aproximadamente, 25 a 30ns e 40 a 50ns de trajetória). Finalmente a trajetória simulada a partir da pose 3 é a mais caótica com a formação e quebra de diversas interações entre os grupos carboxilato e nitro e as R479 e R544.

Nos quadros inferiores da figura 36 é possível observar a manutenção do contato cátion- π entre R482 e W550 ao longo de todas as trajetórias simuladas. A distância entre a R544 e o W550 varia razoavelmente com destaque aos momentos de interação entre a R544 e o grupo nitro do inibidor durante a réplica 2.

Para os ensaios de *docking* para o I2 no sítio catalítico enzimático com a estrutura realizando a interação cátion- π entre a R482 e o W550 verificam-se novamente interações entre o ligante e a S477 e R479 e, dessa vez, também a R482 (pose 1), interação com o E478 e proximidade com as R479 e R482 (pose 2) e interação com o E446 e proximidade com as R479 e R482 (pose 3). Dentre os resíduos mencionados, apenas a R482 aparece significativamente perturbada nos ensaios de RMN.

As análises de RMSD para o I2 indicam estabilidade deste no sítio catalítico, ao longo de toda trajetória simulada, porém aqui ocorre um pequeno deslocamento do ligante em relação à posição inicial da ancoragem de forma que o ligante acaba interagindo a maior parte do tempo, em praticamente todas as réplicas simuladas, com as R482 e R544.

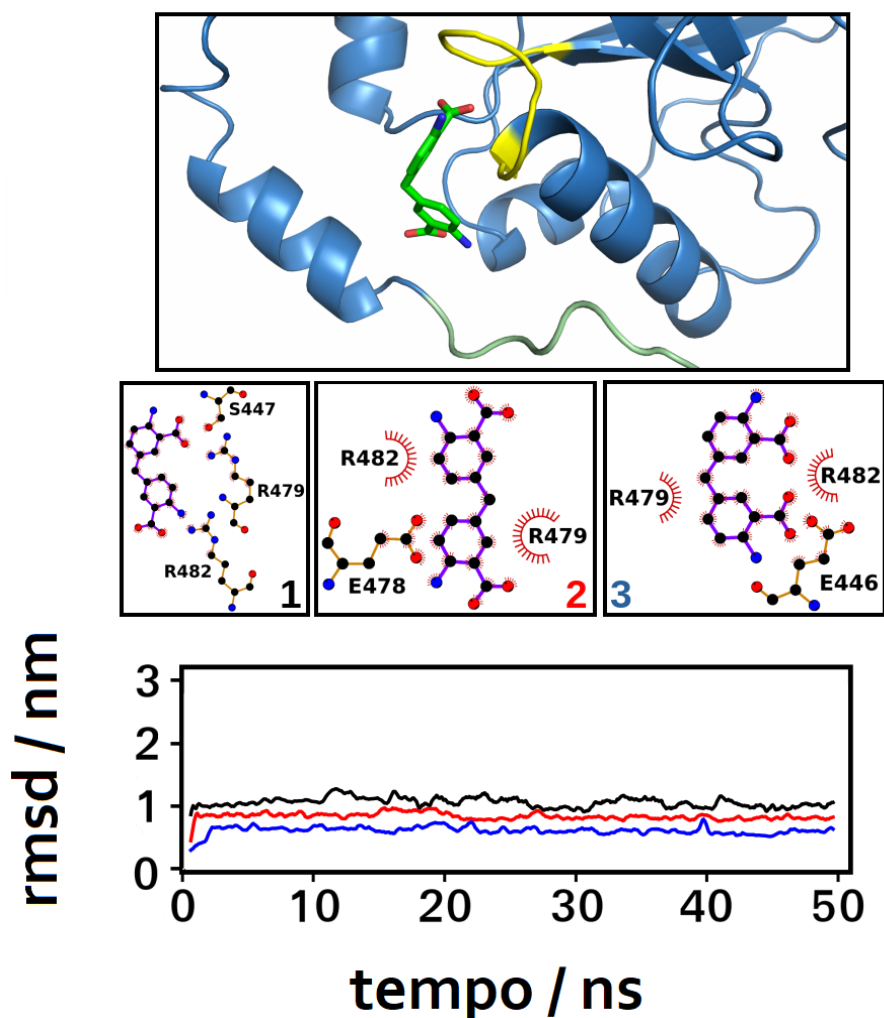


Figura 37. *Docking* molecular e simulações para o inibidor 2. É mostrada a pose mais bem ranqueada (quadro superior), contatos ligante - proteína (quadros centrais) para as poses 1, 2 e 3 e RMSD para o ligante nas simulações realizadas a partir dos sistemas montados com a pose 1 (preto), 2 (vermelho) ou 3 (azul). A estrutura inicial apresentava a interação cátion- π entre a R482 e o W550.

A figura 38 mostra as análises de distância para as três réplicas simuladas para a pose 2, selecionada por apresentar algumas das situações interessantes observadas ao longo das três réplicas produzidas para esta pose. Novamente os principais contatos ao longo da trajetória não refletem os contatos identificados no ensaio de *docking*, mas verifica-se a grande importância dos resíduos de R482 e R544, apenas o primeiro identificado como próximo da posição de ancoragem, para manutenção do ligante no sítio catalítico enzimático. Ao longo da réplica 1 é possível observar a troca de um contato estável entre um dos carboxilatos da molécula, que interage com a R544, e o carboxilato oposto que passa a interagir com este mesmo resíduo (quadro A, curvas em azul e verde).

A trajetória simulada a partir da pose 2 mostra a formação de uma ponte a partir do contato entre um dos carboxilatos e a R482 e entre o carboxilato oposto e a R544 (quadro B, curvas em azul e preto). Por um breve momento, durante esta trajetória, ocorre a formação de um terceiro contato entre a R544 e o carboxilato que originalmente interage com a R482, mesmo durante a interação da R544 com o carboxilato oposto. Essa interação sugere grande flexibilidade em torno do metileno da molécula, que une ambos benzoatos, o que pode ser possível devido ao fato de se tratar de um grupo com um átomo de carbono sp^3 realizando apenas ligações simples com livre rotação, mas parece improvável de ocorrer realmente. Finalmente, a terceira réplica apresenta contatos iniciais entre um dos carboxilatos e a R544 (quadro C, curva em azul) mas em seguida ocorre a troca dos pares de interação com formação simultânea de contatos entre o carboxilato oposto e as R482 e R544, que se mantêm até o fim da trajetória.

Novamente verifica-se a manutenção do contato cátion- π entre R482 e W550 ao longo de todas as trajetórias simuladas. A distância entre a R544 e o W550 varia pouco ao longo destas trajetórias, mas se encontra sempre em distâncias que impedem o contato entre estes resíduos.

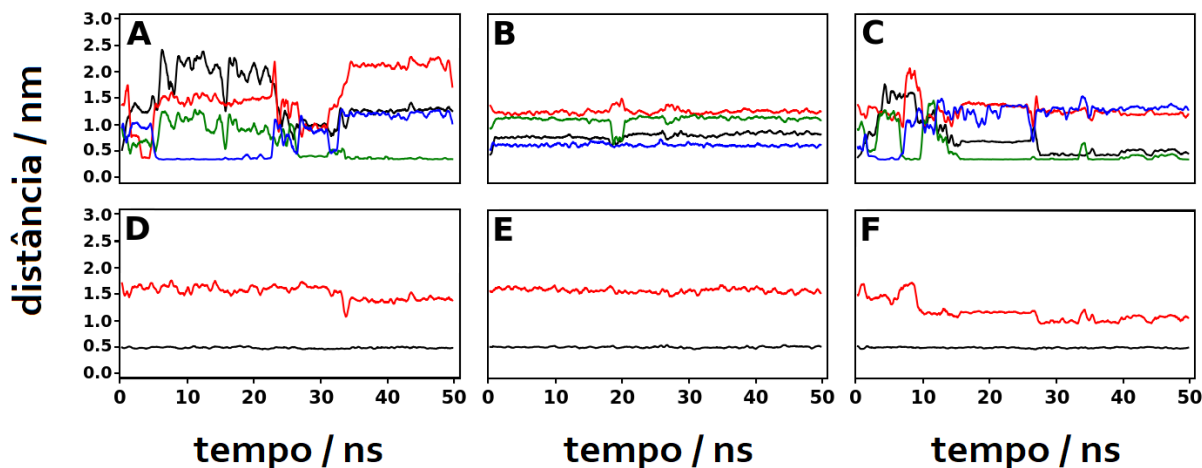


Figura 38. Análises de distâncias das trajetórias geradas a partir da pose 2 do inibidor 2 com o *docking* realizado sobre a estrutura com a interação cátion- π entre W550 e R482. Nos quadros acima têm-se: Preto: COO-1xR482; Vermelho: COO-2xR482; Verde: COO-1xR544; Azul: COO-2xR544; réplicas 1 (A), 2 (B) e 3 (C). Nos quadros abaixo têm-se: Preto: W550xR482; Vermelho: W550xR544; réplicas 1 (D), 2 (E) e 3 (F).

Os ensaios de *docking* molecular para o I3 no sítio catalítico enzimático com a estrutura realizando a interação cátion- π entre a R482 e o W550 apresentaram interação entre o ligante e o E446 e a R482 (pose 1), interação com a R482 (pose 2) e interações com o E446, R482 e T547 (pose 3). As três poses apresentaram o ligante numa região próxima aos resíduos de R479 e R548 sem realizar, no entanto, contato com estes. Dentre os resíduos mencionados, as R482, T547 e R548 aparecem significativamente perturbadas nos ensaios de RMN.

Dentre todas as simulações realizadas para os ligantes a partir da ancoragem no sítio catalítico enzimático, com a formação de ambos contatos cátion- π (R482xW550 ou R544xW550) as análises de RMSD para o I3 foram as únicas que apresentaram grande afastamento da pequena molécula da região de ancoragem e houve dissociação do inibidor da proteína em 2 réplicas das 9 simuladas.

A figura 40 mostra as análises de distância para as três réplicas simuladas para a pose 2, selecionada por representar o comportamento típico observado para as trajetórias deste inibidor. Aqui, um dos contatos importantes para a manutenção do ligante no sítio de ancoragem está identificado em todos os ensaios de *docking*, o contato entre a R482 e o carboxilato isolado. A R544, no entanto, resíduo que mais interage com os grupos carregados do ligante ao longo das simulações, não aparece nos ensaios de ancoragem como importante para a complexação do inibidor nas poses apresentadas.

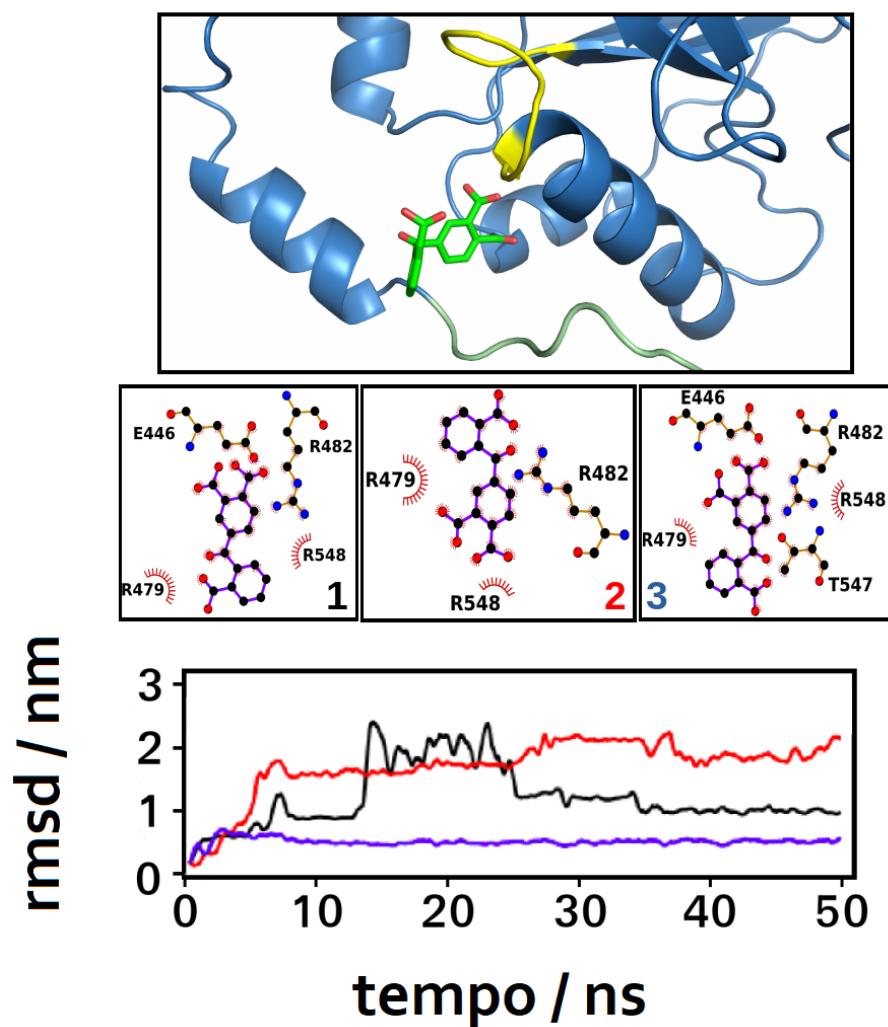


Figura 39. *Docking* molecular e simulações para o inibidor 3. É mostrada a pose mais bem ranqueada (quadro superior), contatos ligante - proteína (quadros centrais) para as poses 1, 2 e 3 e RMSD para o ligante nas simulações realizadas a partir dos sistemas montados com a pose 1 (preto), 2 (vermelho) ou 3 (azul). A estrutura inicial apresentava a interação cátion- π entre a R482 e o W550.

A primeira e a última réplica apresentam alternância entre os contatos das R482 e R544 e os carboxilatos de ambos os lados da molécula durante praticamente toda a trajetória sendo clara a formação de uma ponte entre as R482 e R544 e os carboxilatos opostos da molécula durante a trajetória da réplica 3 (quadro C, curvas em preto e azul). A réplica 2 (para a qual é mostrada a análise de RMSD acima, curva vermelha) apresenta um afastamento considerável do sítio inicial de interação por parte do ligante que passa a interagir com outros resíduos mais distantes do sítio de ancoragem. Para este ligante foi observada dissociação completa em duas réplicas simuladas para a terceira pose.

Para estas simulações também verifica-se a manutenção do contato cátion- π entre R482 e W550 ao longo de todas as trajetórias simuladas. A distância entre a R544 e o W550 também varia pouco ao longo destas trajetórias, mais uma vez, sempre em distâncias que impedem o contato entre estes resíduos.

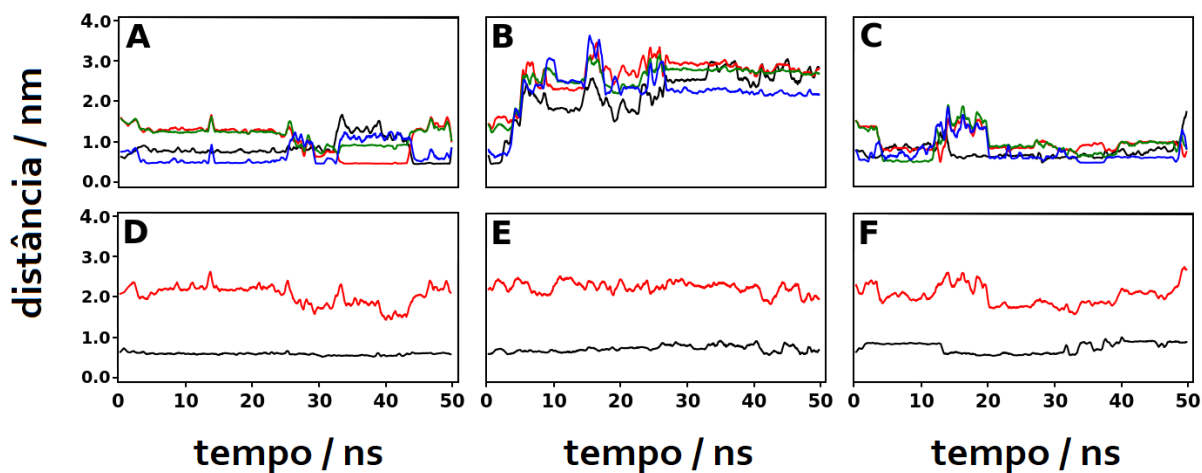


Figura 40. Análises de distâncias das trajetórias geradas a partir da pose 2 do inibidor 3 com o *docking* realizado sobre a estrutura com a interação cátion- π entre W550 e R482. Nos quadros acima têm-se: Preto: COO⁻¹xR482; Vermelho: COO⁻¹xR544; Verde: COO⁻²xR544; Azul: COO⁻³xR544; réplicas 1 (A), 2 (B) e 3 (C). Os carboxilatos 1 e 2 são os que se encontram no mesmo lado da molécula (vizinhos) enquanto o carboxilato 3 é o isolado. Nos quadros abaixo têm-se: Preto: W550xR482; Vermelho: W550xR544; réplicas 1 (D), 2 (E) e 3 (F).

Dessa forma observou-se que existe uma concordância razoável, mas não absoluta, entre os resíduos indicados como interagindo com os ligantes nos sítios de ancoragem e aqueles significativamente perturbados nos ensaios de RMN sendo que, de maneira geral, muitas vezes estes resíduos não correspondem com aqueles que mantêm as interações com os grupos químicos dos inibidores durante as simulações de dinâmica molecular. Além disso, verificou-se que os contatos responsáveis pela manutenção dos inibidores I1, I2 e I3 complexados no sítio catalítico enzimático ao longo das trajetórias ocorrem, principalmente, a partir da interação entre os carboxilatos dos inibidores e as R482 e R544. Para o I1 o contato entre o grupo nitro e a R479 também é importante para a estabilidade do inibidor. A interação R482 x W550 é estável durante as simulações.

3.2.3.3. Sítio catalítico: Cátion- π R544 x W550

Os resultados para os ensaios de *docking* e as simulações realizadas no sítio catalítico a partir da estrutura da proteína com a interação cátion- π formada entre a R544 e o W550 serão mostrados num padrão semelhante ao utilizado acima.

As poses escolhidas a partir dos ensaios de *docking* molecular para o I1 no sítio catalítico da Cdc25B na estrutura com a interação cátion- π entre a R544 e o W550 mostram interações com a Y428 e a R479 e proximidade com as R482 e 544 (pose 1), proximidade com as R479, R544 e R548 (pose 2) e interações com o grupo amídico do *backbone* da F475 e proximidade com a R479. Dentre os resíduos mencionados, as Y428, R544 e R548 aparecem significativamente perturbadas nos ensaios de RMN.

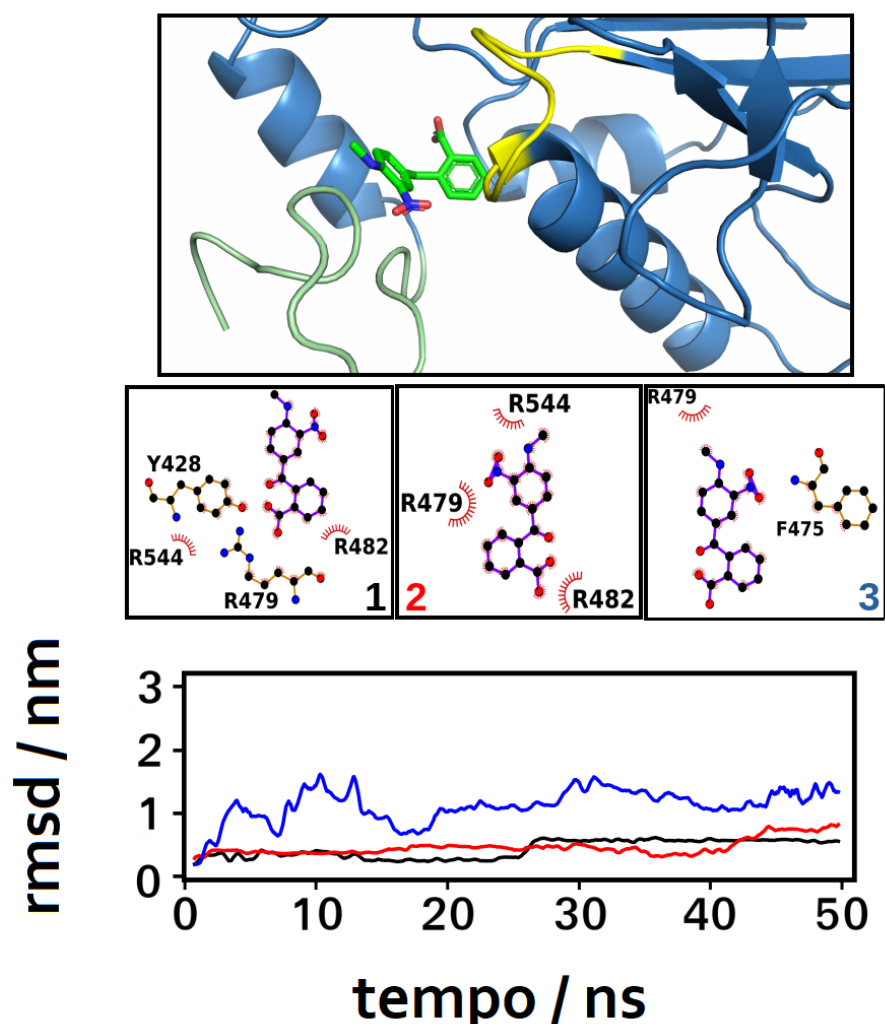


Figura 41. *Docking* molecular e simulações para o inibidor 1. É mostrada a pose mais bem ranqueada (quadro superior), contatos ligante - proteína (quadros centrais) para as poses 1, 2 e 3 e RMSD para o ligante nas simulações realizadas a partir dos sistemas montados com a pose 1 (preto), 2 (vermelho) ou 3 (azul). A estrutura inicial apresentava a interação cátion- π entre a R544 e o W550.

As análises de RMSD para o I1 indicam estabilidade deste em relação ao sítio de interação ao longo de toda trajetória simulada para as poses 1 e 2 enquanto a simulação a partir da pose 3 apresenta um distanciamento do sítio inicial, sem dissociação do ligante.

A figura 42 mostra as análises de distância para as três réplicas simuladas para a pose 3, selecionada para esta pose. Assim como na seção anterior, os principais contatos ao longo da trajetória não refletem os contatos identificados no ensaio de *docking*: aqui a R482 é responsável pela manutenção do ligante no sítio catalítico enzimático, numa nova região mais abaixo na hélice D. Os demais resíduos apontados como importantes para a complexação do ligante não participam de mais contatos estáveis durante a trajetória.

A trajetória simulada a partir da pose 2 mostra a formação de uma ponte a partir do contato entre o grupo nitro do I1 e a R544 e o carboxilato do inibidor e a R479 (quadro B, curvas em verde e vermelho). A trajetória simulada a partir da pose mostra grande importância do contato entre o carboxilato do ligante e a R479 na manutenção do inibidor no sítio com formação de um contato do mesmo carboxilato com a R544 durante parte considerável da simulação.

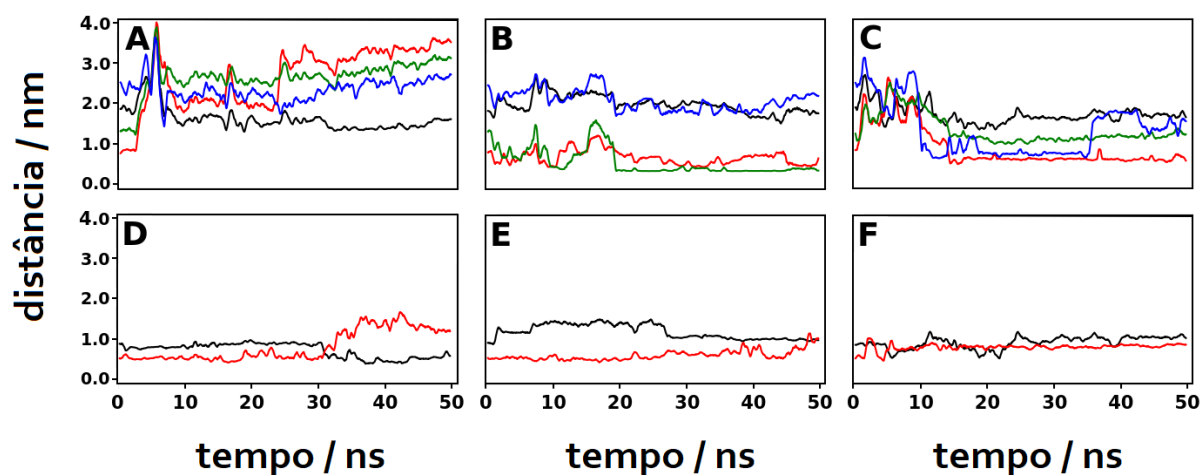


Figura 42. Análises de distâncias das trajetórias geradas a partir da pose 1 do inibidor 1 com o *docking* realizado sobre a estrutura com a interação cátion- π entre W550 e R544. Nos quadros acima têm-se: Preto: NOOxR479; Vermelho: COOxR479; Verde: NOOxR544; Azul: COOxR544; réplicas 1 (A), 2 (B) e 3 (C). Nos quadros abaixo têm-se: Preto: W550xR482; Vermelho: W550xR544; réplicas 1 (D), 2 (E) e 3 (F).

Aqui verifica-se uma situação que se repete ao longo das trajetórias simuladas a partir da estrutura da proteína com a interação cátion- π entre a R544 e o W550: durante a trajetória simulada a partir da pose 1 observa-se uma troca entre

o contato do W550 para formação da interação cátion- π com a R482. Na simulação a partir da pose 3 o cátion- π com a R544 também é desfeito com breves contatos com a R482.

Para os ensaios de *docking* molecular com o I2 verificam-se interações com a L443 e as R482 e R544 e proximidade com a R479 (pose 1), interações com o E478 e a R479 e proximidade com a R482 (pose 2) e interações com o E474, F475, S477 e R479 (pose 3). Dentre os resíduos mencionados, as R482 e R544 aparecem significativamente perturbadas nos ensaios de RMN.

As análises de RMSD para o I2 indicam estabilidade deste em relação ao sítio de interação ao longo de toda trajetória simulada para as poses 1 e 2 enquanto a simulação a partir da pose 3 apresenta um distanciamento do sítio inicial, sem dissociação do ligante.

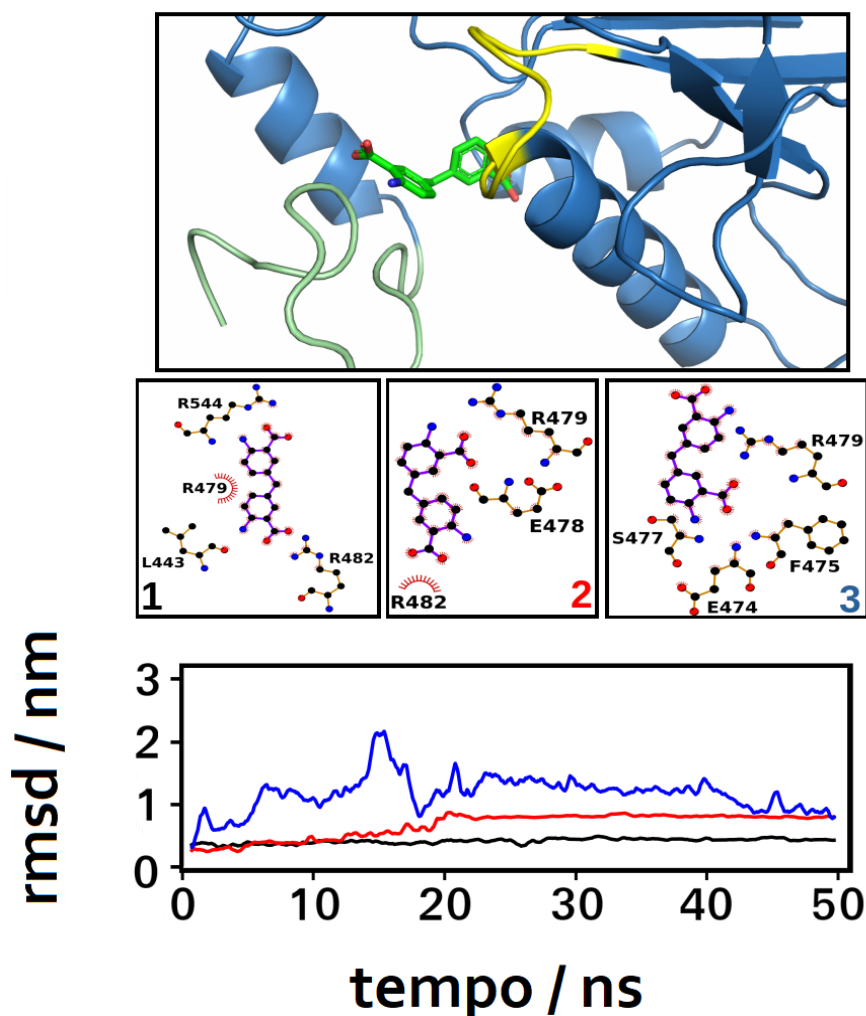


Figura 43. *Docking* molecular e simulações para o inibidor 2. É mostrada a pose mais bem ranqueada (quadro superior), contatos ligante - proteína (quadros centrais) para as poses 1, 2 e 3 e RMSD para o ligante nas simulações realizadas a partir dos sistemas montados com a pose 1 (preto), 2 (vermelho) ou 3 (azul). A estrutura inicial apresentava a interação cátion- π entre a R544 e o W550.

As análises de distância para as três réplicas simuladas para o I2 a partir da pose 3 são mostradas na figura 44. Aqui o resíduo que apresenta as interações mais importantes com o ligante é a R544, interagindo a todo instante com pelo menos um dos carboxilatos do inibidor, em alguns momentos com ambos. A R482 chega a fazer contatos curtos com o outro carboxilato. Este padrão é observado para as três réplicas da pose analisada.

Novamente a interação cátion- π inicial entre a R544 e o W550 se desfaz, dessa vez ao longo de todas as trajetórias sendo que, com exceção da segunda réplica, isto é observado bem cedo na simulação. A interação do ligante com a R544, inclusive com contatos dos dois carboxilatos ao mesmo tempo (quadros A e C, curvas em azul e verde) é o motivo para essa troca: com a arginina interagindo fortemente com os grupos aniônicos do ligante o W550 prefere interagir com a R482, próxima. Talvez por este motivo sejam vistos poucos contatos posteriores com essa última arginina.

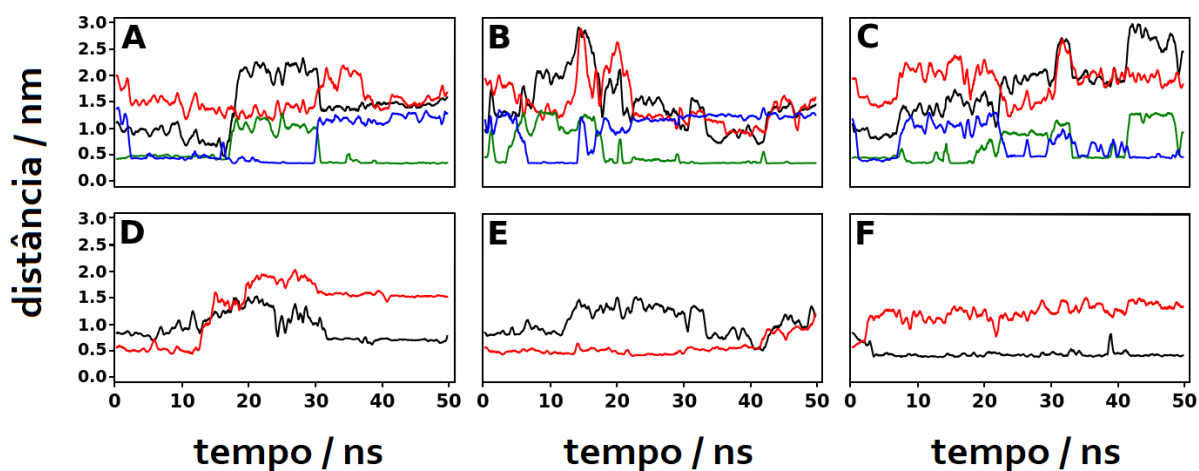


Figura 44. Análises de distâncias das trajetórias geradas a partir da pose 3 do inibidor 2 com o *docking* realizado sobre a estrutura com a interação cátion- π entre W550 e R544. Nos quadros acima têm-se: Preto: COO-1xR482; Vermelho: COO-2xR482; Verde: COO-1xR544; Azul: COO-2xR544; réplicas 1 (A), 2 (B) e 3 (C). Nos quadros abaixo têm-se: Preto: W550xR482; Vermelho: W550xR544; réplicas 1 (D), 2 (E) e 3 (F).

Finalmente os ensaios de ancoragem com o I3 apresentaram interação com a R544 e proximidade com as R479 e R482 (pose 1), interações com o E478 e as R479 e R544 (pose 2) e interações com o E474, S477, E478 e R479 (pose 3). Dentre os resíduos mencionados, as R482 e R544 aparecem significativamente perturbadas nos ensaios de RMN.

As análises de RMSD para o I3 indicam estabilidade deste em relação ao sítio de interação ao longo de toda trajetória simulada para as poses 1 e 2 enquanto a simulação a partir da pose 3 apresenta, novamente, um distanciamento do sítio inicial, sem dissociação do ligante.

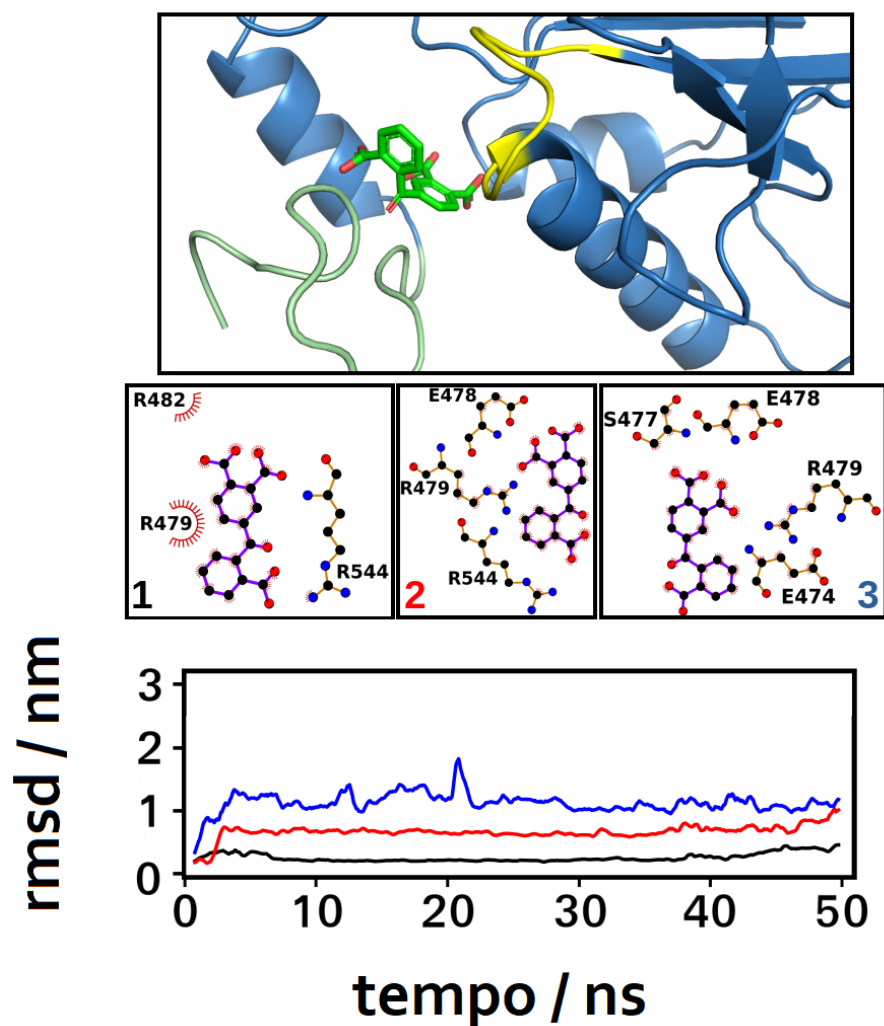


Figura 45. *Docking* molecular e simulações para o inibidor 3. É mostrada a pose mais bem ranqueada (quadro superior), contatos ligante - proteína (quadros centrais) para as poses 1, 2 e 3 e RMSD para o ligante nas simulações realizadas a partir dos sistemas montados com a pose 1 (preto), 2 (vermelho) ou 3 (azul). A estrutura inicial apresentava a interação cátion- π entre a R544 e o W550.

As análises de distância para as três réplicas simuladas para o I3 a partir da pose 1 são mostradas na figura 46. Aqui o resíduo que apresenta as interações mais importantes com o ligante é, novamente, a R544, mais uma vez interagindo a todo instante com pelo menos um dos carboxilatos do inibidor, e em alguns

momentos com os três. Durante a primeira réplica alternam-se nos contatos os carboxilatos vizinhos; durante a segunda réplica estes carboxilatos interagem a quase todo instante com a R544 e após os instantes iniciais a molécula se dobra ligeiramente de forma que todos os carboxilatos ficam próximos deste resíduo; durante a terceira réplica ocorrem alternâncias entre os contatos da R544 com os carboxilatos opostos do inibidor.

A interação cátion- π inicial entre a R544 e o W550 se desfaz logo no início das simulações para a primeira e para a terceira réplica, permitindo o contato do triptofano com a R482. Interessantemente durante a segunda réplica ocorre uma espécie de sanduíche W550 / R544 / inibidor, de forma que o contato cátion- π , após ser brevemente interrompido, é restabelecido ao longo do restante da simulação. Mais uma vez as interações do ligante com a R544 parecem justificar a quebra do contato R544 - W550.

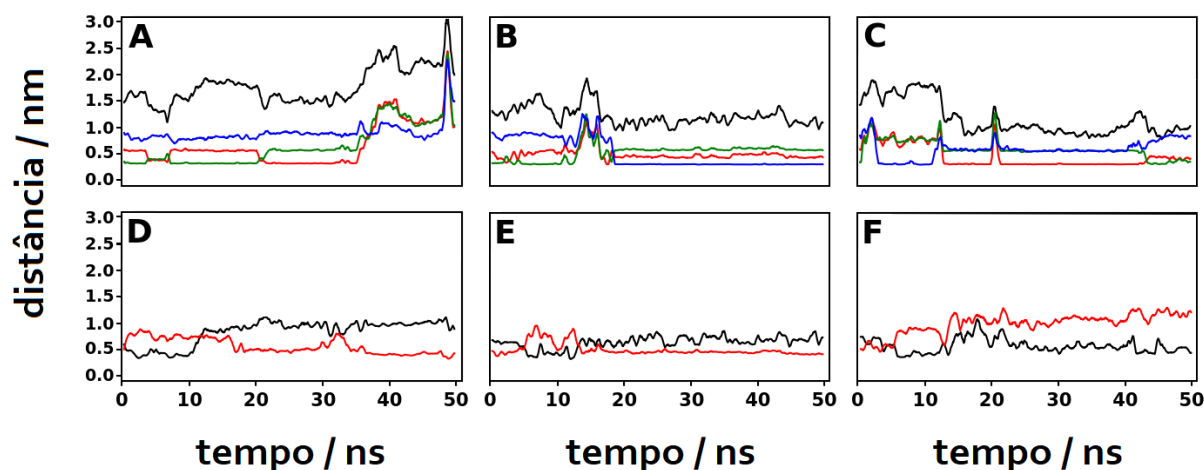


Figura 46. Análises de distâncias das trajetórias geradas a partir da pose 3 do inibidor 3 com o *docking* realizado sobre a estrutura com a interação cátion- π entre W550 e R544. Nos quadros acima têm-se: Preto: COO⁻1xR482; Vermelho: COO⁻1xR544; Verde: COO⁻2xR544; Azul: COO⁻3xR544; réplicas 1 (A), 2 (B) e 3 (C). Os carboxilatos 1 e 2 são os que se encontram no mesmo lado da molécula (vizinhos) enquanto o carboxilato 3 é o isolado. Nos quadros abaixo têm-se: Preto: W550xR482; Vermelho: W550xR544; réplicas 1 (D), 2 (E) e 3 (F).

Assim, novamente foi observada uma concordância razoável, mas não absoluta, entre os resíduos indicados como interagindo com os ligantes nos sítios de ancoragem e as perturbações verificadas nos ensaios de RMN. O resíduo de R544 foi observado muitas vezes interagindo com os ligantes ou próximo a estes nos ensaios de ancoragem e se apresentou como o principal responsável pela manutenção dos inibidores no sítio catalítico.

Observou-se uma troca no par do triptofano na interação cátion- π da estrutura inicial na maioria das simulações alterando o contato com a R544 para um contato com a R482. O provável motivo para esta mudança é a importante participação da R544 na manutenção dos ligantes complexados no sítio catalítico enzimático.

3.2.3.4. Estabilidade do contato cátion- π R544 x W550 na estrutura utilizada para os ensaios de ancoragem molecular

A partir das observações realizadas sobre a baixa estabilidade do contato cátion- π ao longo das trajetórias mostradas na seção 3.2.4.3, com o intuito de responder à questão de se o motivo de tal instabilidade é devido à presença dos inibidores complexados ou inerente à estrutura escolhida (caso houvesse sido, por exemplo, escolhida uma estrutura particularmente instável em termos de manutenção do contato entre a R544 e o W550), foram realizadas novas simulações de dinâmica molecular.

Para isso partiu-se da exata estrutura obtida para a proteína no sistema após a rampa de aquecimento realizada, mas removeram-se os inibidores do sistema de forma que todas as posições dos componentes do sistema se encontravam iguais, com exceção dos pequenos ligantes complexados. Foram realizadas simulações em triplicatas para cada um dos sistemas originais (agora sem o I1, sem o I2 e sem o I3).

A figura 47 apresenta as distâncias entre os centros de massa das cadeias laterais da R544 e do W550 ao longo das trajetórias simuladas para cada novo sistema (A, B e C) com as cores representando cada réplica: preto, vermelho e azul respectivamente para as réplicas 1, 2 e 3. É possível observar que, com exceção de uma pequena porção da trajetória de uma réplica simulada, os contatos entre as cadeias laterais da R544 e do W550 são mantidos, indicando estabilidade na interação.

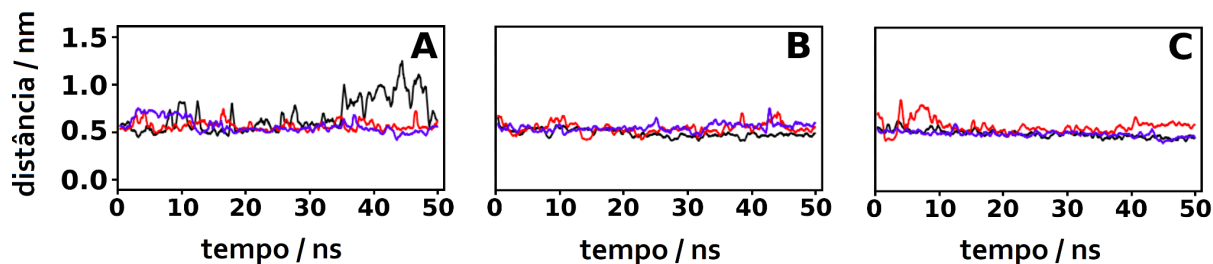


Figura 47. Distância entre os centros de massa das cadeias laterais da R544 e do W550 ao longo das novas simulações realizadas. As cores indicam diferentes réplicas para sistemas montadas na ausência dos inibidores I1 (A), I2 (B) e I3 (C).

Dessa forma, a observação da seção anterior sobre a possibilidade de quebra dos contatos entre a R544 e o W550 ao longo das simulações se darem pela competição dos inibidores pela arginina parece se confirmar assim como a ausência de um viés na estrutura inicial que pudesse favorecer a quebra da interação cátion- π .

3.3. Conclusão

Os ensaios de ancoragem molecular realizados apresentam resíduos importantes para a complexação dos ligantes para as quais em vários momentos existe concordância com a evidência experimental (CSP) e as simulações, mas este padrão não é absoluto. No geral estes resíduos não se mostraram essenciais para a manutenção dos ligantes nos sítios de ancoragem ao longo das simulações.

As simulações com os ligantes nos sítios S1, S2 e S3 indicaram instabilidade nas interações destes com estas regiões da proteína, no geral, concordando com os mecanismos de inibição observados (não competitivo para I1 e competitivo para I2 e I3). As simulações com os ligantes no sítio catalítico apresentaram estabilidade dos ligantes nesta região, apesar de muitas vezes estes se manterem no sítio catalítico complexados com resíduos diferentes daqueles apontados pela ancoragem como importantes para a ligação destes com a proteína. Assim, tem-se novamente uma evidência do mecanismo sugerido para os inibidores I2 e I3. Esta observação não suporta o mecanismo sugerido para I1, que seria esperado não se manter estável no sítio catalítico, de forma que parece mais provável que este atue de maneira inespecífica em diversos sítios ao longo da superfície, sendo pouco provável sua complexação no sítio catalítico enzimático ao longo do tempo e/ ou que o mecanismo de inibição observado seja, na realidade, um mecanismo misto com

ponderação diversa para as interações com o sítio catalítico enzimático e os demais sítios de ligação.

As argininas 482 e 544 se apresentaram como os resíduos mais importantes para a manutenção dos ligantes no sítio catalítico ao longo das trajetórias simuladas e a arginina 479 se mostra importante em contatos com o inibidor 1, uma situação especialmente interessante uma vez que esta arginina se encontra no *P-loop* e este inibidor não apresentou cinética de inibição enzimática compatível com mecanismo competitivo.

Pode-se observar que a carga do inibidor (consequência direta dos carboxilatos presentes em cada um) permite maior variação de contatos. Enquanto o inibidor 1 tende a manter contatos entre o carboxilato ou o grupo nitro e as R479 e R544 sem muitas trocas ao longo da trajetória, os inibidores 2 e 3 trocam constantemente as argininas parceiras (R482 e R544, tipicamente) dos carboxilatos presentes nos anéis opostos. Este fato pode justificar uma maior estabilidade destes inibidores no sítio catalítico que poderia ser observada com simulações mais longas ou a partir de outras técnicas de simulação computacional.

Nas simulações realizadas com a estrutura inicial da Cdc25B realizando o contato cátion- π entre o W550 e a R544, diferentemente do contato W550 - R482, esse se mostrou instável devido à importância da R544 na complexação dos inibidores estudados, o que se demonstrou com simulações adicionais que confirmaram a estabilidade do contato na ausência dos ligantes.

Conclusão

Simulações de dinâmica molecular indicaram interações favoráveis entre as cadeias laterais das R482 e R544 e do W550 presentes no domínio catalítico da Cdc25B. Com a produção dos mutantes W550A, R482A, R544A e ensaios de ^1H - ^{15}N HSQC RMN com os mutantes produzidos, interações cátion- π previstas nas simulações computacionais foram confirmadas. As análises do conjunto de dados gerados sugerem que os contatos observados sejam, de fato, transientes e é ainda possível que mais parceiros interajam com o W550.

Interessantemente, o rompimento das interações entre os grupos guanidinos das R482 ou R544 com o anel indólico do W550 provoca alterações de ambiente químico em diversos resíduos do domínio catalítico da Cdc25B, sobretudo grande parte dos localizados na hélice F e na região C-terminal desordenada da proteína, ocorrendo também perturbações a longa distância.

Ensaio de cinética de inibição enzimática sugerem mecanismo de inibição não-competitivo para o I1 e competitivo para os I2 e I3. Os ensaios de ancoragem molecular realizados apresentam resíduos importantes para a complexação dos ligantes sendo que em vários momentos existe concordância com a evidência experimental (CSP) e as simulações, mas este padrão não é absoluto. No geral estes resíduos não se mostraram essenciais para a manutenção dos ligantes nos sítios de ancoragem ao longo das simulações.

As simulações com os ligantes nos sítios S1, S2 e S3 indicaram instabilidade nas interações destes com estas regiões da proteína, no geral, concordando com os mecanismos de inibição observados (não competitivo para I1 e competitivo para I2 e I3). As simulações com os ligantes no sítio catalítico apresentaram estabilidade dos ligantes nesta região, inclusive o I1 (para o qual não seria esperada esta estabilidade).

As argininas 482 e 544 se apresentaram como os resíduos mais importantes para a manutenção dos ligantes no sítio catalítico ao longo das trajetórias simuladas e a arginina 479 se mostra importante em contatos com o inibidor 1. Como este resíduo se encontra no *P-loop* tem-se novamente uma discordância com o esperado para o mecanismo de inibição constatado para o I1.

Pode-se observar que a carga do inibidor (consequência direta dos carboxilatos presentes em cada um) permite maior variação de contatos

relacionados à estabilidade dos ligantes no sítio catalítico enzimático. Este fato pode justificar uma maior estabilidade dos inibidores 2 e 3 no sítio catalítico em comparação com o inibidor 1, o que poderia ser observado em simulações mais longas ou a partir de outras técnicas de simulação computacional.

Nas simulações realizadas com a estrutura inicial da Cdc25B realizando o contato cátion- π entre o W550 e a R544, diferentemente do contato W550 - R482, esse se mostrou instável devido à importância da R544 na complexação dos inibidores estudados, o que se demonstrou com simulações adicionais que confirmaram a estabilidade do contato na ausência dos ligantes.

Referências

[1] Dahm, R. **Friedrich Miescher and the discovery of DNA.** *Developmental Biology* (2005) 278: 274 - 288.

[2] Avery, OT; MacLeod, CM; McCarty, M. **Studies of the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III.** *Journal of Experimental Medicine* (1944) 79: 137 - 158.

[3] Watson, JD; Crick, FHC. **A structure for deoxyribose nucleic acid.** *Nature* (1953) 171: 737 - 738.

[4] Abdellah, Z; Ahmadi, A; Ahmed, S; Aimable, M; Ainscough, R; Almeida, J; Almond, C; Ambler, A; Ambrose, K; Ambrose, K; Andrew, R; Andrews, D; Andrews, N; Andrews, D; Apweiler, E; Arbery, H; Archer, B; Ash, G; Ashcroft, K; Ashurst, J; Ashwell, R; Atkin, D; Atkinson, A; Atkinson, B; Attwood, J; Aubin, K; Auger, K; Avis, K; Babbage, A; Babbage, S; Bacon, J; Bagguley, C; Bailey, J; Baker, A; Banerjee, R; Bardill, S; Barker, D; Barker, G; Barker, D; Barlow, K; Baron, L; Barrett, A; Bartlett, R; Basham, D; Basham, V; Bateman, A; Bates, K; Baynes, C; Beard, L; Beard, S; Beare, D; Beasley, A; Beasley, H; Beasley, O; Beck, S; Bell, E; Bellerby, D; Bellerby, T; Bemrose, R; Bennett, J; Bentley, D; Bentley, A; Berks, M; Berks, M; Bethel, G; Bird, C; Birney, E; Bissell, H; Blackburne-Maze, S; Blakey, S; Bolton, C; Bonfield, J; Bonnett, R; Border, R; Bradley, A; Brady, N; Bray, J; Bray-Allen, S; Bridgeman, A; Brook, J; Brooking, S; Brown, A; Brown, C; Brown, J; Brown, M; Brown, M; Bruskiwich, R; Bryant, J; Buck, D; Buckle, V; Budd, C; Buller, S; Burberry, J; Burford, D; Burgess, J; Burrill, W; Burrows, C; Burton, J; Burton, C; Butcher, P; Butler, A; Cairns, M; Camm, N; Campbell, C; Canning, B; Carder, C; Carder, P; Carter, N; Cavanna, T; Chalk, S; Chan, K; Chapman, J; Charles, R; Chillingworth, N; Chothia, T; Chui, C; Clack, R; Clamp, M; Clark, A; Clark, G; Clark, K; Clark, S; Clark, S; Clark, R; Clarke, B; Clarke, E; Clarke, K; Clarke, A; Clarke, L; Clee, C; Clegg, S; Clifford, K; Coates, J; Cogley, V; Coffey, A; Coggill, P; Cole, L; Collier, R; Collings, S; Collins, J; Collins, P; Colman, L; Connolly, A; Connor, R; Conquer, J; Conroy, D; Constance, D; Cook, L; Cooper, J; Cooper, R; Cooper, R; Coppola, M; Copsey, T; Corby, N; Cornell, L; Cornell, R; Cornell, C; Cottage, A; Coulson, A; Coville, G; Cox, A; Cox, T; Coxhill, R; Craig, M; Crane, T; Crawley, M; Crew, V; Cuff, J; Culley, K; Cummings, A; Cummings, K; Cummings, P; Curran, A; Curwen, V; Cutts, J; Daniels, R; Davidson, L; Davies, J; Davies, J; Davies, N; Davies, R; Davis, J; Davis, J; Davis, M; Dawson, E; Deadman, R; Dean, P; Dear, S; Dearden, F; Delgado, M; Deloukas, P; Dennis, J; Dhimi, J; Dibling, C; Dobbs, R; Dobson, R; Dockree, C; Doddington, D; Dodsworth, S; Doggett, N; Down, T; Dunham, A; Dunham, I; Dunn, A; Dunn, M; Durbin, R; Durham, J; Dutta, I; Dwyer, R; Dyer, L; Earthrowl, M; Eastham, T; Eastham, E; Edwards, C; Edwards, K; Ellington, A; Elliott, D; Ellwood, M; Emberson,

B; Errington, H; Evans, G; Evans, J; Evans, K; Evans, R; Eyraş, E; Faulkner, L; Fellingham, C; Feltwell, T; Fennell, S; Finn, R; Flack, T; Felming, C; Fleming, K; Flint, J; Flint, M; Floyd, Y; Footman, S; Fowler, J; Frame, D; Francis, M; Francis, S; Frankish, A; Frankland, J; Fraser, A; Fraser, D; French, L; Fricker, D; Frost, D; Frost, J; Frost, L; Frost, C; Fuller, L; Fullerton, K; Gardner, A; Garner, P; Garnett, J; Gatland, L; Gatland, L; Ghorı, J; Gibbs, B; Gibson, D; Gibson, E; Gilbert, J; Gilby, L; Gillson, C; Glithero, R; Gooderham, A; Gorton, M; Grafham, D; Grant, M; Grant, S; Gray, I; Gray, E; Green, L; Greenhalgh, J; Greenhill, J; Griffiths-Jones, S; Gregg, P; Gregory, S; Gribble, S; Griffiths, C; Griffiths, E; Griffiths, M; Grocock, M; Guthrie, I; Gwilliam, R; Hall, R; Halls, K; Hall-Tamly, G; Hamlett, J; Hamlett, J; Hammond, S; Hancock, J; Harding, A; Harley, J; Harper, D; Harper, G; Harper, P; Harradence, G; Harrison, C; Harrison, E; Harrison, R; Hart, E; Hassan, D; Hawkins, N; Hawley, K; Hayes, K; Heath, P; Heathcott, R; Hembry, C; Henderson, C; Herd, T; Hewitt, S; Higgs, D; Hillyard, G; Hinkins, R; Ho, S; Hodgson, D; Hoffs, M; Holden, J; Holdgate, J; Holloway, E; Holmes, I; Holmes, S; Holroyd, S; Hooper, A; Hopewell, L; Hopkins, B; Hornett, G; Hornsby, G; Hornsby, T; Horsley, S; Horton, R; Howard, P; Howden, P; Howe, K; Howell, G; Hubbard, T; Huckle, E; Hughes, J; Hughes, J; Hull, L; Hummeric, H; Humphray, S; Humphries, M; Hunt, A; Hunt, P; Hunt, S; Hunter, G; Hyde, D; Ince, M; Isherwood, J; Iyer, V; Izatt, J; Izmajlowicz, M; Jareborg, N; Jassal, B; Jeffery, G; Jeffery, K; Jeffrey, C; Jekosch, K; Jenkins, L; Johansen, T; Johnson, C; Johnson, C; Johnson, D; Jolley, K; Jones, A; Jones, C; Jones, J; Jones, M; Jones, M; Jones, S; Joseph, S; Joy, A; Joy, L; Joy, V; Joyce, G; Jubb, M; Karunaratne, K; Kay, M; Kaye, D; Kearney, L; Keenan, S; Kelley, S; Kershaw, J; Kettleborough, R; Kidd, C; Kierstan, P; Kimberley, A; King, A; Kingsley, S; Kingswood, C; Kingle, G; Knights, A; Korf, I; Krogh, A; Lad, H; Laidlaw, P; Laing, M; Laird, G; Lambart, C; Lambie, R; Langford, C; Larke, B; Lau, T; Lawlor, S; Leather, S; Lehvaslaiho, M; Lemke, J; Leonard, S; Leongamornlert, D; Leversha, M; Lightning, J; Lindsay, S; Line, M; Linsdell, S; Little, P; Lloyd, C; Lloyd, D; Lock, V; Lock, W; Lodziak, A; Longden, I; Loraine, H; Lord, R; Loveland, J; Lovell, J; Lye, G; Marriott, N; Marrone, A; Marsden, P; Marsh, V; Martin, M; Martin, S; Maslen, G; Mason, D; Matthews, L; Matthews, P; Matthews, N; Maynard, M; McCann, O; McClay, J; McCollum, C; McConnachie, L; McDonald, B; McDonald, L; McDowall, J; McGuire, S; McKeown, S; McLaren, S; McLay, K; McLean, J; McMurdo, J; McMurray, A; McMurray, D; McWilliams, N; Mehta, N; Meidl, P; Menuge, N; Mercer, S; Miah, A; Micklem, G; Miles, S; Milne, S; Mistry, D; Mistry, S; Mitchell, J; Mitchell, J; Mohammadi, M; Molina, C; Mooney, P; Moore, M; Moreland, A; Mortimore, B; Mott, R; Mullenger, I; Mullikin, J; Munday, B; Munday, E; Mungall, A; Murnane, C; Murrell, K; Myers, A; Negus, D; Ng, B; Niblett, D; Nicholson, J; Nickerson, T; Nijjar, S; Ning, Z; Nisbet, J; Novik, K; Odell, C; O'Donovan, D; Ogbighele, F; Oinn, T; Oliver, H; Oliver, K; Orbell, H; Osborn, A; Osborne, J; Overton-Larty, E; Palmer, S; Pandian, R; Parker, A; Parkin, C; Parkin, K; Parry-Brown, G; Patel, D; Patel, R; Pearce, A; Pearson, D; Peck, A; Peck, R; Peden, J; Pelan, S; Percy, C; Perito, A; Perrault, I; Peters, A; Pettett, R; Phillimore, B; Phillips, K; Phillips, S; Platt, D; Playford, E; Plumb, B; Pocock, M; Porter, K; Porter, T; Potter, C; Potter, S; Powell, D; Prathalingham, R;

Prigmore, E; Quail, M; Quarrie, H; Quince, C; Qureshi, M; Ramsay, H; Ramsey, Y; Ranby, S; Rance, R; Rand, V; Ratford, J; Ratford, L; Read, D; Redhead, D; Redon, R; Rees, C; Reid, M; Reinhardt, A; Rice, A; Rice, C; Rice, P; Richard, S; Richardson, S; Ridler, K; Riethoven, L; Rigby, R; Robinson, M; Rochford, R; Rogers, J; Rogers, L; Ross, H; Ross, M; Rule, A; Rule, J; Russell, B; Rutter, J; Safdar, K; Salter, N; Santoyo-Lopez, J; Saunders, D; Scott, C; Scott, D; Scott, I; Seager, F; Searle, M; Searle, P; Searle, S; Sehra, H; Shakespeare, J; Shardelow, J; Sharp, G; Shaw, T; Shaw-Smith, C; Shearing, J; Sheppard, K; Sheppard, R; Sheridan, E; Shownkeen, R; Silk, R; Sims, M; Sims, S; Sivadasan, S; Skuce, C; Smink, L; Smith, A; Smith, L; Smith, L; Smith, M; Smith, R; Smith, S; Smith, S; Sneath, H; Soderlund, C; Solovyev, V; Sonhammer, E; Sotheran, E; Spooner, W; Spraggon, L; Squares, J; Squares, S; Stables, M; Stalker, J; Stamford, S; Stammers, M; Steingruber, H; Stephens, Y; Steward, C; Stewart, A; Stewart, M; Still, I; Stock, M; Stoppard, L; Storey, P; Storey, R; Stowe, S; Strachan, C; Strachan, G; Stribling, C; Sturdy, J; Sulston, J; Swainson, C; Swann, M; Swarbreck, D; Sycamore, N; Tagney, M; Tan, S; Tarling, E; Taylor, A; Taylor, G; Taylor, K; Taylor, R; Taylor, S; Taylor, S; Tee, L; Tester, J; Theaker, A; Thexton, D; Thomas, C; Thomas, D; Thomas, K; Thomas, R; Thommai, R; Thorpe, A; Thorpe, K; Threadgold, G; Thurston, S; Tinsley, E; Tracey, A; Travers, J; Tromans, A; Tubby, B; Tufarelli, C; Turney, K; Upson, D; Helmond, ZV; Vaudin, M; Viknaraja, R; Vine, W; Voak, P; Vollenhoven, D; Walker, S; Wall, M; Wallis, J; Wallis, M; Warren, G; Warry, G; Watson, A; Watt, N; Webb, A; Webb, J; Weir, J; Wells, A; Wells, S; Welton, R; West, P; West, T; Wheatley, A; Wheatley, C; Wheeler, G; Whitaker, H; White, A; White, A; White, B; White, J; White, S; Whitehead, S; Whiteley, M; Whittaker, P; Whittaker, A; Widaa, S; Wild, A; Wilkinson, J; Wilkinson, P; Willey, D; Williams, A; Williams, B; Williams, L; Williamson, H; Wilmer, T; Wilming, L; Wilson, B; Wilson, G; Wilson, M; Wilson, N; Wilson, S; Wilson, W; Wilson, P; Window, P; Winster, J; Winzar, C; Witt, J; Wobus, F; Wood, E; Wood, J; Woodeson, S; Woodfine, K; Woodhouse, R; Woodmansey, R; Wooster, R; Wray, M; Wray, P; Wright, C; Wright, K; Wright, D; Wyatt, J; Xie, J; Young, L; Young, S; Younger, R; Zhao, S.

Finishing the euchromatic sequence of the human genome. *Nature* (2004) 431: 931 - 945.

[5] Birney, EJA; Stamatoyannopoulos, A; Dutta, R; Guigo, TR; Gingeras, EH; Margulies, Z; Weng, M; Snyder, ET; Dermitzakis, RE; Thurman, MS; Kuehn, CM; Taylor, S; Neph, CM; Koch, S; Asthana, A; Malhotra, I; Adzhubei, JA; Greenbaum, RM; Andrews, P; Flicek, PJ; Boyle, H; Cao, NP; Carter, GK; Clelland, S; Davis, N; Day, P; Dhami, SC; Dillon, MO; Dorschner, H; Fiegler, PG; Giresi, J; Goldy, M. Hawrylycz, A; Haydock, R; Humbert, KD; James, BE; Johnson, EM; Johnson, TT; Frum, ER; Rosenzweig, N; Karnani, K; Lee, GC; Lefebvre, PA; Navas, F; Neri, SC; Parker, PJ; Sabo, R; Sandstrom, A; Shafer, D; Vetrici, M; Weaver, S; Wilcox, M; Yu, FS; Collins, J; Dekker, JD; Lieb, TD; Tullius, GE; Crawford, S; Sunyaev, WS; Noble, I; Dunham, F; Denoeud, A; Reymond, P; Kapranov, J; Rozowsky, D; Zheng, R; Castelo, A; Frankish, J; Harrow, S; Ghosh, A; Sandelin, IL; Hofacker, R; Baertsch, D; Keefe, S; Dike, J; Cheng, HA; Hirsch, EA; Sekinger, J; Lagarde, JF; Abril, A;

Shahab, C; Flamm, C; Fried, J; Hackermuller, J; Hertel, M; Lindemeyer, K; Missal, A; Tanzer, S; Washietl, J; Korbel, O; Emanuelsson, JS; Pedersen, N; Holroyd, R; Taylor, D; Swarbreck, N; Matthews, MC; Dickson, DJ; Thomas, MT; Weirauch, J; Gilbert, J; Drenkow, I; Bell, X; Zhao, KG; Srinivasan, WK; Sung, HS; Ooi, KP; Chiu, S; Foissac, T; Alioto, M; Brent, L; Pachter, ML; Tress, A; Valencia, SW; Choo, CY; Choo, C; Ucla, C; Manzano, C; Wyss, E; Cheung, TG; Clark, JB; Brown, M; Ganesh, S; Patel, H; Tammana, J; Chrast, CN; Henrichsen, C; Kai, J; Kawai, U; Nagalakshmi, J; Wu, Z; Lian, J; Lian, P; Newburger, X; Zhang, P; Bickel, JS; Mattick, P; Carninci, Y; Hayashizaki, S; Weissman, T; Hubbard, RM; Myers, J; Rogers, PF; Stadler, TM; Lowe, CL; Wei, Y; Ruan, K; Struhl, M; Gerstein, SE; Antonarakis, Y; Fu, ED; Green, U; Karaoz, A; Siepel, J; Taylor, LA; Liefer, KA; Wetterstrand, PJ; Good, EA; Feingold, MS; Guyer, GM; Cooper, G; Asimenos, CN; Dewey, M; Hou, S; Nikolaev, JI; Montoya-Burgos, A; Loytynoja, S; Whelan, F; Pardi, T; Massingham, H; Huang, NR; Zhang, I; Holmes, JC; Mullikin, A; Ureta-Vidal, B; Paten, M; Sringhaus, D; Church, K; Rosenbloom, WJ; Kent, EA; Stone, S; Batzoglou, N; Goldman, RC; Hardison, D; Haussler, W; Miller, A; Sidow, ND; Trinklein, ZD; Zhang, L; Barrera, R; Stuart, DC; King, A; Ameer, S; Enroth, MC; Bieda, J; Kim, AA; Bhinge, N; Jiang, J; Liu, F; Yao, VB; Vega, CW; Lee, PNA; Yang, Z; Moqtaderi, Z; Zhu, X; Xu, S; Squazzo, MJ; Oberley, D; Inman, MA; Singer, TA; Richmond, KJ; Munn, A; Rada-Iglesias, O; Wallerman, J; Komorowski, JC; Fowler, P; Couttet, AW; Bruce, OM; Dovey, PD; Ellis, CF; Langford, DA; Nix, G; Euskirchen, S; Hartman, AE; Urban, P; Kraus, S; Van Calcar, N; Heintzman, TH; Kim, K; Wang, C; Qu, G; Hon, R; Luna, CK; Glass, MG; Rosenfeld, SF; Aldred, SJ; Cooper, A; Halees, JM; Lin, HP; Shulha, M; Xu, JN; Haidar, Y; Yu, VR; Iyer, RD; Green, C; Wadelius, PJ; Farnham, B; Ren, RA; Harte, AS; Hinrichs, H; Trumbower, H; Clawson, J; Hillman-Jackson, AS; Zweig, K; Smith, A; Thakkapallayil, G; Barber, RM; Kuhn, D; Karolchik, L; Armengol, CP; Bird, PI; Bakker, AD; Kern, N; Lopez-Bigas, JD; Martin, BE; Stranger, A; Woodroffe, E; Davydov, A; Dimas, E; Eyraes, IB; Hallgrimsdottir, J; Huppert, MC; Zody, GR; Abecasis, X; Estivill, GG; Bouffard, X; Guan, NF; Hansen, JR; Idol, VV; Maduro, B; Maskeri, JC; McDowell, M; Park, PJ; Thomas, AC; Young, RW; Blakesley, DM; Muzny, E; Sodergren, DA; Wheeler, KC; Worley, H; Jiang, GM; Weinstock, RA; Gibbs, T; Graves, R; Fulton, ER; Mardis, RK; Wilson, M; Clamp, J; Cuff, S; Gnerre, DB; Jaffe, JL; Chang, K; Lindblad-Toh, ES; Lander, M; Koriabine, M; Nefedov, K; Osoegawa, Y; Yoshinaga, B; Zhu, PJJ. **Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project.** *Nature* (2007) 447(7146): 799 - 816.

[6] Gabius, HJ. **The sugar code: Why glycans are so important.** *BioSystems* (2018) 16: 102 - 111.

[7] Wright, PE; Dyson, HJ. **Intrinsically disordered proteins in cellular signalling and regulation.** *Molecular Cell Biology* (2015) 16: 18 - 29.

[8] Voet, D; Voet, JG; **Biochemistry - 4th edition.** Wiley, Hoboken, NJ, 2011.

- [9] Wright, PE; Dyson, HJ. **Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm.** *Journal of Molecular Biology* (1999) 293: 321 - 331.
- [10] Fischer, E. **Einfluss der configuration auf die wirkung der enzyme.** *Berichte der Deutschen Chemischen Gesellschaft* (1894) 27(3): 2985 - 2993.
- [11] Koshland, DE. **Application of a theory of enzyme specificity to protein synthesis.** *Proceedings of the National Academy of Sciences* (1958) (44) 2: 98 - 104.
- [12] Boehr, DD; Nussinov, R; Wright, PE. **The role of dynamic conformational ensembles in biomolecular recognition.** *Natural Chemical Biology* (2009) (5) 11: 789 - 796.
- [13] Nelson, D; Cox, M. **Lehninger principles of biochemistry - 6th edition.** *W. H. Freeman*, New York, NY, 2012.
- [14] Zaha, A; Ferreira, HB; Passaglia, LMP. **Biologia molecular básica - 5ª edição.** *Artmed*, Porto Alegre, RS, 2014.
- [15] Marescal, O; Cheeseman, IM. **Cellular mechanisms and regulation of quiescence.** *Developmental Cell* (2020) 55: 259 - 271.
- [16] Alberts, B; Johnson, A; Lewis, J; Raff, M; Roberts, K; Walter, P. **Molecular biology of the cell - 5th edition.** *Garland Science*, New York, NY, 2008.
- [17] Pavletich, NP. **Mechanisms of cyclin-dependent kinase regulation: structure of Cdks, their cyclin activators, and Cip and INK4 inhibitors.** *Journal of Molecular Biology* (1999) 287(5): 821 - 828.
- [18] Obaya, A; Sedivy, J. **Regulation of cyclin-Cdk activity in mammalian cells.** *Cellular and Molecular Life Sciences* (2002): 59(1): 126 - 142.
- [19] Aressy, B; Ducommun, B. **Cell cycle control by the Cdc25 phosphatases.** *Anti-Cancer in Medicinal Chemistry* (2008) 8: 818 - 824.
- [20] Lavecchia, A; di Giovanni, C; Novellino, E. **Cdc25B phosphatase inhibitors: An update.** *Mini-Reviews in Medicinal Chemistry* (2012) 12: 62 - 73.
- [21] Arantes, GM. **Free-energy profiles for catalysis by dual-specificity phosphatases.** *Biochemistry Journal* (2006) 399: 343 - 350.
- [22] Arantes, GM. **The catalytic acid in the dephosphorylation of the Cdk2-pTpY/CycA protein complex by Cdc25B phosphatase.** *Journal of Physical Chemistry B* (2008) 112: 15244 - 15247.
- [23] Reynolds, RA; Yem, AW; Wolfe, CL; Deibel Jr, MR; Chidester, CG; Watenpaugh, KD. **Crystal structure of the catalytic subunit of Cdc25B required for G₂/M**

phase transition of the cell cycle. *Journal of Molecular Biology* (1999) 293: 559 - 568.

[24] Sayegh, RSR; Tamaki, FK; Marana, SR; Salinas, RK; Arantes, GM. **Conformational flexibility of the complete catalytic domain of Cdc25B phosphatases.** *Proteins: Structure, Function, and Bioinformatics* (2016) 84: 1567 - 1575.

[25] Hollingsworth, SA; Karplus, PA. **A fresh look at the Ramachandran plot and the occurrence of the standard structures in proteins.** *Biomolecular Concepts* (2010) 1(3-4): 271 - 283.

[26] Dougherty, DA. **The cation- π interaction.** *Accounts of Chemical Research* (2013) 46(4): 885 - 893.

[27] Dougherty, DA; Ma, JC. **The cation- π interaction.** *Chemical Reviews* (1997) 97: 1303 - 1324.

[28] Berman, HM; Henrick, K; Nakamura, H. **Announcing the worldwide Protein Data Bank.** *Nature Structural Biology* (2003) 10(12): 980.

[29] Crowley, PB; Golovin, A. **Cation- π interactions in protein-protein interfaces.** *Proteins: Structure, Function, and Bioinformatics* (2005): 59: 231 - 239.

[30] Mecozzi, S; West Jr, AP; Dougherty, DA. **Cation- π interactions in aromatics of biological and medicinal interest: Electrostatic potential surfaces as a useful qualitative guide.** *Proceedings of National Academy of Sciences* (1996) 93: 10566 - 10571.

[31] Mecozzi, S; West Jr, AP; Dougherty, DA. **Cation- π interactions in simple aromatics: Electrostatics provide a predictive tool.** *Journal of American Chemical Society* (1996) 118: 2307 - 2308.

[32] Sussman, JL; Harel, M; Frolow, F; Oefner, C; Goldman, A; Toker, L; Silman, I. **Atomic structure of acetylcholinesterase from *Torpedo californica*: A prototypic acetylcholine-binding protein.** *Science* (1991) 253: 873 - 879.

[33] Lim, LW; Shamala, N; Mathews, FS. **Three-dimensional structure of the iron-sulfur flavoprotein trimethylamine dehydrogenase at 2,4-Å resolution.** *The Journal of Biological Chemistry* (1986) 261(32): 15140 - 14146.

[34] Quioco, FA; Hu, G; Gershon, PD. **Structural basis of mRNA cap recognition by proteins.** *Current Opinions in Structural Biology* (2000) 10: 78 - 86.

[35] Henzler-Wildman, KA; Lei, M; Thai, V; Kerns, SJ; Karplus, M; Kern, D. **A hierarchy of timescales in protein dynamics is linked to enzyme catalysis.** *Nature* (2007) 450: 913 - 918.

- [36] Oldfield, JC; Dunker, AK. **Intrinsically disordered proteins and intrinsically disordered protein regions.** *Annual Reviews in Biochemistry* (2014) 85: 553 - 584.
- [37] Jirgensons, B. **Classification of proteins according to conformation.** *Die Makromolekulare Chemie* (1966) 91: 74 - 86.
- [38] Stein, A; Pache, RA; Bernadó, P; Pons, M; Aloy, P. **Dynamic interactions of proteins in complex networks: A more structured view.** *The FEBS Journal* (2009) 276: 5390 - 5405.
- [39] Gsponer, J; Babu, MM. **The rules of disorder or why disorder rules.** *Progress in Biophysics and Molecular Biology* (2009) 99: 94 - 103.
- [40] Mitchison, JM; Creanor, J. **Further measurements of DNA synthesis and enzyme potential during cell cycle of fission yeast *Schizosaccharomyces pombe*.** *Experimental Cell Research* (1971) 69: 244 - 247.
- [41] Nurse, P. **Genetic control of cell size at cell division in yeast.** *Nature* (1975) 256: 547 - 551.
- [42] Nurse, P; Thuriaux, P; Nasmyth, K. **Genetic control of the cell division cycle in the fission yeast *Schizosaccharomyces pombe*.** *Molecular and General Genetics* (1976) 146: 167 - 178.
- [43] Fantes, P. **Epistatic gene interactions in the control of division in fission yeast.** *Nature* (1979) 279: 428 - 430.
- [44] Russel, P; Nurse, P. **Cdc25+ functions as an inducer in the mitotic control of fission yeast.** *Cell* (1986) 45: 145 - 153.
- [45] Galaktionov, K; Beach, D. **Specific activation of Cdc25 tyrosine phosphatases by B-type cyclins: Evidence for multiple roles of mitotic cyclins.** *Cell* (1991) 67: 1181 - 1194.
- [46] Nagata, A; Igarashi, M; Jinno, S; Suto, K; Okayama, H. **An additional homolog of the fission yeast cdc25+ gene occurs in humans and is highly expressed in some cancer cells.** *Nature New Biology* (1991) 10: 959 - 968.
- [47] Chen, YC; Hsieh, HH; Chang, HC; Wang, HC; Lin, WJ; Lin, JJ. **Cdc25B induces cellular senescence and correlates with tumor suppression in a p53-dependent manner.** *Journal of Biological Chemistry* (2021) 296: 100564.
- [48] Boutros, R; Labjois, V; Ducommun, B. **Cdc25B phosphatases in cancer cells: Key players? Good targets?** *Nature reviews - Cancer* (2007) 7: 495 - 507.
- [49] Fauman, EB; Cogswell, JP; Lovejoy, B; Rocque, WJ; Holmes, W; Montana, VG; Piwnicka-Worms, H; Rink, MJ; Saper, MA. **Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A.** *Cell* (1998) 93: 617 - 625.

- [50] Nilsson, I; Hoffmann, I. **Cell cycle regulation by the Cdc25 phosphatase family.** *Progress in Cell Cycle Research* (2000) 4: 107 - 114.
- [51] Alonso, A; Pulido, R. **The extended human ptpome: A growing tyrosine phosphatase family.** *The FEBS Journal* (2015) 283: 1404 - 1429.
- [52] Shen, T; Huang, S. **The role of Cdc25A in the regulation of cell proliferation and apoptosis.** *Anticancer Agents in Medicinal Chemistry* (2012) 12(6): 631 - 639.
- [53] Karlsson, C; Katich, S; Hagting, A; Hoffmann, I; Pines, J. **Cdc25B and Cdc25C differ markedly in their properties as initiators of mitosis.** *The Journal of Cell Biology* (1999) 146(3): 573 - 583.
- [54] Källström, H; Lindqvist, A; Pospisil, V; Lundgren, A; Rosenthal, CK. **Cdc25A localisation and shuttling: Characterisation of sequences mediating nuclear export and import.** *Experimental Cell Research* (2005) 303: 89 - 100.
- [55] Davezac, N; Baldin, V; Gabrielli, B; Forrest, A; Theis-Febvre, N; Yashida, M; Ducommun. **Regulation of Cdc25B phosphatases subcellular localization.** *Nature* (2000) 19: 2179 - 2185.
- [56] Teng, YN; Chung, CL; Lin, YM; Pan, HA; Liao, RW; Kuo, PL. **Expression of various Cdc25B isoforms in human spermatozoa.** *Fertility and Sterility* (2007) 88(2): 379 - 382.
- [57] Gasteiger, E; Gattiker, A; Hoogland, C; Ivanyi, I; Appel, RD; Bairoch, A. **ExPASy: the proteomics server for in-depth protein knowledge and analysis.** *Nucleic Acids Research* (2003) 31 (13): 3784 - 3788.
- [58] Reis, AOR; Sayegh, RSR; Marana, SR; Arantes, GM. **Combining free energy simulations and NMR chemical-shift perturbation to identify transient cation- π contacts in proteins.** *Journal of Chemical Information and Modeling* (2020) 60: 890 - 897.
- [59] Field, MJ. **A practical introduction to the simulation of molecular systems - 2nd edition.** Cambridge University Press, Cambridge, UK, 2007.
- [60] McCammon, JA; Gelin, BR; Karplus, M. **Dynamics of folded proteins.** *Nature* (1977) 267: 585 - 590.
- [61] Curtolo, F; Arantes, GM. **Mechanisms for flavin-mediated oxidation: Hydride or hydrogen-atom transfer?** *Journal of Chemical Information and Modeling* (2020) 60: 6282 - 6287.
- [62] Teixeira, MH; Curtolo, F; Camilo, SG; Field, MJ; Zheng, P; Li H; Arantes, GM. **Modeling the hydrolysis of iron-sulfur clusters.** *Journal of Chemical Information and Modeling* (2020) 60: 653 - 660.

[63] Nunes-Alves, A; Arantes, GM. **Mechanical unfolding of macromolecules coupled to bond dissociation.** *Journal of Chemical Theory and Computation* (2018) 14: 282 - 290.

[64] Nunes-Alves, A; Zuckerman, DM; Arantes, GM. **Escape of a small molecule from inside T4 lysozyme by multiple pathways.** *Biophysical Journal* (2018) 114: 1058 - 1066.

[65] Nunes-Alves, A; Arantes, GM. **Ligand-receptor affinities computed by an adapted linear interaction model for continuum electrostatics and by protein conformational averaging.** *Journal of Chemical Information and Modeling* (2014) 54: 2309 - 2319.

[66] Frisch, MJ; Trucks, GW; Schlegel, HB; Scuseria, GE; Robb, MA; Cheeseman, JR; Scalmani, G; Barone, V; Mennucci, B; Petersson, GA; Nakatsuji, H; Caricato, M; Li, X; Hratchian, HP; Izmaylov, AF; Bloino, J; Zheng, G; Sonnenberg, JL; Hada, M; Ehara, M; Toyota, K; Fukuda, R; Hasegawa, J; Ishida, M; Nakajima, T; Honda, Y; Kitao, O; Nakai, H; Vreven, T; Montgomery, JA; Peralta Jr, JE; Ogliaro, F; Bearpark, M; Heyd, JJ; Brothers, E; Kudin, KN; Staroverov, VN; Kobayashi, R; Normand, J; Raghavachari, K; Rendell, A; Burant, JC; Iyengar, SS; Tomasi, J; Cossi, M; Rega, N; Millam, JM; Klene, M; Knox, JE; Cross, JB; Bakken, V; Adamo, C; Jaramillo, J; Gomperts, R; Stratmann, RE; Yazyev, O; Austin, AJ; Cammi, R; Promelli, C; Ochterski, JW; Martin, RL; Morokuma, K; Zakrzewski, VG; Voth, GA; Salvador, P; Dannenberg, JJ; Dapprich, S; Daniels, AD; Farkas, A; Foresman, JB; Ortiz, JV; Cioslowski, J; Fox, DJ. **Gaussian 09, Revision A.** Gaussian Inc. Wallingford, CT, 2009.

[67] Møller, C; Plesset, MS. **Note on an approximation treatment for many-electron systems.** *Physical Review* (1934) 46: 618 - 622.

[68] Vanommeslaeghe, K; Hatcher, E; Acharya, C; Kundu, S; Zhong, S; Shim, J; Darian, E; Guvench, O; Lopes, P; Vorobyov, I; Mackerell Jr, AD. **CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.** *Journal of Computational Chemistry* (2009) 31: 671 - 690.

[69] Hornak, V; Abel, R; Okur, A; Strockbine, B; Roitberg, A; Simmerling, C. **Comparison of multiple Amber force fields and development of improved protein backbone parameters.** *Proteins: Structure, Function and Bioinformatics* (2006) 65: 712 - 725.

[70] Lindorff-Larsen, K; Piana, S; Palmo, K; Maragakis, P; Klepeis, JL; Dror, RO; Shaw, DE. **Improved side-chain torsion potentials for the Amber ff99SB protein force field.** *Proteins: Structure, Function and Bioinformatics* (2010) 78: 1950 - 1958.

- [71] Jorgensen, WL; Chandrasekhar, J; Madura, JD; Impey, RW; Klein, ML. **Comparison of simple potential functions for simulating liquid water.** *Journal of Chemical Physics* (1983) 79: 926 - 935.
- [72] Bussi, G; Donadio, D; Parrinello, M. **Canonical sampling through velocity rescaling.** *Journal of Chemical Physics* (2007) 126: 014101.
- [73] Parrinello, M; Rahman, A. **Polymorphic transitions in single crystals: A new molecular dynamics method.** *Journal of Applied Physics* (1981) 52: 7182 - 7190.
- [74] Hess, B; Kutzner, C; van der Spoel, D; Lindahl, E. **GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation.** *Journal of Chemical Theory and Computation* (2008) 4: 435 - 447.
- [75] Roux, B. **The calculation of the potential of mean force using computer simulations.** *Computer Physics Communications* (1995) 91: 275 - 282.
- [76] Johnson, RW. **An introduction to the bootstrap.** *Teaching Statistics* (2001) 23: 49 - 54.
- [77] Davis, MW; Jorgensen, EM. **ApE, a plasmid editor: A freely available DNA manipulation and visualization program.** *Frontiers in Bioinformatics* (2022) 2: 818619.
- [78] **SnapGene software**, disponível em (<http://www.snapgene.com>).
- [79] Laemmli, UK. **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* (1970) 227: 680 - 685.
- [80] Delaglio, F; Grzesiek, S; Vuister, GW; Zhu, G; Pfeifer, J; Bax, A. **NMRPipe: A multidimensional spectral processing system based on UNIX pipes.** *Journal of Biomolecular NMR*(1995) 6: 277 - 293.
- [81] Vranken, WF; Boucher, W; Stevens, TJ; Fogh, RH; Pajon, A; Llinas, M; Ulrich, EL; Markley, JL; Ionides, J; Laue, ED. **The CCPN data model for NMR spectroscopy: Development of a software pipeline.** *Proteins: Structure, Function, and Bioinformatics* (2005) 59: 687 - 696.
- [82] Zhang, H; Neal, S; Wishart, DS. **RefDB: A database of uniformly referenced protein chemical shifts.** *Journal of Biomolecular NMR* (2003) 25: 173 - 195.
- [83] Schumann, F; Riepl, H; Maurer, T; Gronwald, W; Neidig, KP; Kalbitzer, H. **Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions.** *Journal of Biomolecular NMR* (2007) 39: 275 - 289.
- [84] Ulrich, EL; Akutsu, H; Doreleijers, JF; Harano, Y; Ioannidis, YE; Lin, J; Livny, M; Mading, S; Maziuk, D; Miller, Z; Nakatani, E; Schulte, CF; Tolmie, DE; Wenger, RK;

Yao, H; Markley, JL. **BioMagResBank**. *Nucleic Acids Research* (2008) 36: D402 - D408.

[85] Souza, VP; Ikegami, CM; Arantes, GM; Marana, SR. **Mutations close to a hub residue affect the distant active site of a GH1 β -glucosidase**. *Plos One* (2018) 13 - 6: e0198696.

[86] Sayegh, RSR. **Flexibilidade conformacional do domínio catalítico da fosfatase Cdc25B**. Tese de doutorado (2016), Instituto de Química da Universidade de São Paulo, São Paulo, Brasil.

[87] Park, H; Li, M; Choi, J; Cho, H; Ham, SW. **Structure-based virtual screening approach to identify novel classes of Cdc25B phosphatase inhibitors**. *Bioorganic & Medicinal Chemistry Letters* (2009) 3: 250 - 265.

[88] Williamson, MP. **Using chemical shift perturbation to characterize ligand binding**. *Progress in Nuclear Magnetic Resonance Spectroscopy* (2013) 73: 1 - 16.

[89] Schaftenaar, G; Noordik, JH. **Molden: a pre- and post-processing program for molecular and electronic structures**. *Journal of Computer-Aided Molecular Design* (2000) 14: 123 - 134.

[90] Adams, PD; Afonine, PV; Baskaran, K; Berman, HM; Berrisford, J; Bricogne, G; Brown, DG; Burley, SK; Chen, M; Feng, Z; Flensburg, C; Gutmanas, A; Hoch, JC; Ikegawa, Y; Kengaku, Y; Krissinel, E; Kurisu, G; Liang, Y; Liebschner, D; Mak, L; Markley, JL; Moriarty, NW; Murshudov, GB; Noble, M; Peisach, E; Persikova, I; Poon, BK; Sobolev, OV; Ulrich, EL; Velankar, S; Vonrhein, C; Westbrook, J; Wojdyr, M; Yokochi, M; Young, JY. **Announcing mandatory submission of PDBx/ mmCIF format files for crystallographic depositions to the Protein Data Bank (PDB)**. *Acta Crystallography* (2019) D75: 451 - 454.

[91] Settimo, L; Bellman, K; Knegt, RMA. **Comparison of the accuracy of experimental and predicted pKa values of basic and acidic compounds**. *Pharmaceutical Research* (2014) 31(4): 1082 - 1095.

[92] Dean, JA. **Lange's Handbook of Chemistry - 13th edition**. Hill Book, New York, NY, 1985.

[93] Haynes, WM. **CRC Handbook of Chemistry and Physics 97th edition**. CRC Press, Boca Raton, FL, 2016.

[94] Trott, O; Olson, AJ. **Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading**. *Journal of Computational Chemistry* (2009) 31(2): 455 - 461.

- [95] Morris, GM. Huey, R; Lindstrom, W; Sanner, MF; Belew, RK; Goodsell, DS; Olson, AJ. **AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.** *Journal of Computational Chemistry* (2009) 30(16): 2785 - 2791.
- [96] Jo, S; Kim, T; Iyer, VG; Im, W. **CHARMM-GUI: A web-based graphical user interface for CHARMM.** *Journal of Computational Chemistry* (2008) 29: 1859 - 1865.
- [97] Laskowski, RA; Swindells, MB. **LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery.** *Journal of Chemical Information and Modeling* (2011) 51: 2778 - 2786.

Sequência esperada - Construção C1-N17

MGSSHHHHHHSSGLVPRGSHMEFQSDHRELIGDYSKAFLLQTVDGKHQDLKYISPETMVALL
 TGKFSNIVDKFVIVDCRYPYEYEGGHIKTAVNLPLERDAESFLLKSPIAPCSLDKRVILIFH
 CEFSSERGPRMCRFIRERDRAVNDYPSLYPEMYILKGGYKEFFPQHHPNFCEPQDYRPMNHE
 AFKDELKTFRLKTRSWAGERSRRELCSRLQDQ Stop

Sequência obtida- Construção C1-N17

MGSSHHHHHHSSGLVPRGSHMEFQSDHRELIGDYSKAFLLQTVDGKHQDLKYISPETMVALL
 TGKFSNIVDKFVIVDCRYPYEYEGGHIKTAVNLPLERDAESFLLKSPIAPCSLDKRVILIFH
 CEFSSERGPRMCRFIRERDRAVNDYPSLYPEMYILKGGYKEFFPQHHPNFCEPQDYRPMNHE
 AFKDELKTFRLKTRSWAGERSRRELCSRLQDQ Stop

Comparação ExPASy:

Range 1: 1 to 218 Graphics				▼ Next Match	▲ Previous Match
NW Score	Identities	Positives	Gaps		
1186	218/218(100%)	218/218(100%)	0/218(0%)		
Query 1	MGSSHHHHHHSSGLVPRGSHMEFQSDHRELIGDYSKAFLLQTVDGKHQDLKYISPETMVA			60	
Sbjct 1	MGSSHHHHHHSSGLVPRGSHMEFQSDHRELIGDYSKAFLLQTVDGKHQDLKYISPETMVA			60	
Query 61	LLTGKFSNIVDKFVIVDCRYPYEYEGGHIKTAVNLPLERDAESFLLKSPIAPCSLDKRVILIFH			120	
Sbjct 61	LLTGKFSNIVDKFVIVDCRYPYEYEGGHIKTAVNLPLERDAESFLLKSPIAPCSLDKRVILIFH			120	
Query 121	LIFHCEFSSERGPRMCRFIRERDRAVNDYPSLYPEMYILKGGYKEFFPQHHPNFCEPQDY			180	
Sbjct 121	LIFHCEFSSERGPRMCRFIRERDRAVNDYPSLYPEMYILKGGYKEFFPQHHPNFCEPQDY			180	
Query 181	RPMNHEAFKDELKTFRLKTRSWAGERSRRELCSRLQDQ	218			
Sbjct 181	RPMNHEAFKDELKTFRLKTRSWAGERSRRELCSRLQDQ	218			

R482A-17N:

>A_24NOV2016_B1_R117A_1_T7_PROMOTER_B01_003.ab1

TAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATGAAATCATCATCAC
 AGCAGCGGCCTGGTGCCGCGCGGCTGTTATATGGAATTCAGAGTGACCACCGAGAGCTGAT
 TGGAGATTACTCTAAGGCCTTCCCTACAGACAGTAGACGGAAAGCACCAAGACCTCAAGT
 ACATCTCACCAGAAACGATGGTGGCCCTATTGACGGGCAAGTTCAGCAACATCGTGGATAAG
 TTTGTGATTGTAGACTGCAGATACCCCTATGAATATGAAGGCGGGCACATCAAGACTGCGGT
 GAACTTGCCCCTGGAACGCGACGCCGAGAGCTTCCCTACTGAAGAGCCCCATCGCGCCCTGTA
 GCCTGGACAAGAGAGTCATCCTCATTTTCCACTGTGAATTCTCATCTGAGCGTGGGCCCGCC
 ATGTGCCGTTTCATCAGGGAACGAGACCGTGCTGTCAACGACTACCCAGCCTCTACTACCC
 TGAGATGTATATCCTGAAAGGCGGCTACAAGGAGTCTTCCCTCAGCACCCGAACTTCTGTG
 AACCCAGGACTACCGGCCCATGAACCACGAGGCCCTTCAAGGATGAGCTAAAGACCTCCGC
 CTCAAGACTCGCAGCTGGGCTGGGGAGCGGAGCCGGCGGGAGCTCTGTAGCCGGCTGCAGGA

Sequence ID: Query_208625 Length: 218 Number of Matches: 1

Range 1: 1 to 218 [Graphics](#)

▼ Next Match ▲ Previous Match

NW Score	Identities	Positives	Gaps
1180	217/218(99%)	217/218(99%)	0/218(0%)
Query 1	MGSSHHHHHSSGLVPRGSHMEFQSDHRELIGDYSKAFLLQTVDGKHQDLKYISPETMVA		60
Sbjct 1	MGSSHHHHHSSGLVPRGSHMEFQSDHRELIGDYSKAFLLQTVDGKHQDLKYISPETMVA		60
Query 61	LLTGKFSNIVDKFVIVDCRYPYEEYEGGHIKTAVNPLERDAESFLLKSPIAPCSLDKRVI		120
Sbjct 61	LLTGKFSNIVDKFVIVDCRYPYEEYEGGHIKTAVNPLERDAESFLLKSPIAPCSLDKRVI		120
Query 121	LIFHCEFSSSERGPRMCRFIRERDRAVNDYPSLYPEMYILKGGYKEFFPQHPNFCEPQDY		180
Sbjct 121	LIFHCEFSSSERGPRMCRFIRERDRAVNDYPSLYPEMYILKGGYKEFFPQHPNFCEPQDY		180
Query 181	RPMNHEAFKDELKTFRLKTRSWAGERSRRELCSRLQDQ	218	
Sbjct 181	RPMNHEAFKDELKTFRLKTRSWAGERSRRELCSRLQDQ	218	

R544A-17N:**>A_24NOV2016_D1_R179A_2_T7_PROMOTER_C01_005.ab1**

TAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATGAACGCATCATCAC
AGCAGCGGCCTGGTGCCGCGGCAGTAAATGGAATTCAGAGTGACCACCGAGAGCTGATT
GGAGATTACTCTAAGGCCTTCCTCCTACAGACAGTAGACGGAAAGCACCAAGACCTCAAGTA
CATCTCACCAGAAACGATGGTGGCCCTATTGACGGGCAAGTTCAGCAACATCGTGGATAAGT
TTGTGATTGTAGACTGCAGATACCCCTATGAATATGAAGGCGGGCACATCAAGACTGCGGTG
AACTTGCCCCTGGAACGCGACGCCGAGAGCTTCCTACTGAAGAGCCCCATCGCGCCCTGTAG
CCTGGACAAGAGAGTCATCCTCATTTTCCTCTGTGAATTCTCATCTGAGCGTGGGCCCCGCA
TGTGCCGTTTCATCAGGGAACGAGACCGTGCTGTCAACGACTACCCAGCCTCTACTACCCT
GAGATGTATATCCTGAAAGGCGGCTACAAGGAGTTCCTCCCTCAGCACCCGAACTTCTGTGA
ACCCAGGACTACCGGCCCATGAACCACGAGGCCTTCAAGGATGAGCTAAAGACCTTCGCCC
TCAAGACTCGCAGCTGGGCTGGGGAGCGGAGCCGGCGGGAGCTCTGTAGCCGGCTGCAGGAC
CAGTGACGGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCA
CCACCACCACTGAGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGCTGAGTTGGCTGCTGCCA
CCGCTGAGCATACTAGCATAACCCT

>A_24NOV2016_C2_R179A_1_T7_TERMINATOR_C02_006.ab1

GTGGTGGTGGTGCTCGAGTGCGGTGACGCTTGTGACGGAGCTCGAATTCGGATCCCGTCAC
TGGCAAAGGCAGCCGGCTACAGAGCTCCCGCCGGCTCCGCTCCCCAGCCCAGCTGCGAGTCT
TGAGGGCGAAGGTCTAAAGATCATCCTTGAAGGCCTCGTGGTTTCATGGGCCGGTAGTCTTGG
GGTTCACAGAAGTTCGGGTGCTGAGGGAAGAACTCCTTGTAGCCGCCTTTCAGGATATACAT
CTCAGGGTAGTAGAGGCTGGGGTAGTCGTTGACAGCACGGTCTCGTTCCTGATGAAACGGC
ACATGCGGGGCCACGCTCAGATGAGAATTCACAGTGGAAAATGAGGATGACTCTCTTGTCC
AGGCTACAGGGCGCGATGGGGCTCTTCAAGTAGGAAGCTCTCGGCGTCGCGTTCAGGGGCAA
GTTACCCGAGTCTTGATGTGCCCGCCTTCATATTCATAGGGGTATCTGCAGTCTACAATCA
CAAACCTATCCACGATGTTGCTGAACTTGCCCGTCAATAGGGCCACCATCGTTTCTGGTGAG
ATGTACTTGAGGTCTTGGTGCTTTCCGTCTACTGTCTGTAGGAGGAAGGCCTTAGAGTAATC
TCCAATCAGCTCTCGGTGGTCACTCTGGAATTCATATGGCTGCCGCGCGGCACCAGGCCG

TGCTGTGATGATGATGATGATGGCTGCTGCCCATGGTATATCTCCTTCTTAAAGTTAAACAA
AAT

Sequência esperada:

MGSSHHHHHHSSGLVPRGSHMEFQSDHRELIGDYSKAFLLQTVDGKHQDLKYISPETMVALL
TGKFSNIVDKFVIVDCRYPYEYEGGHIKTAVNLPLERDAESFLLKSPIAPCSLDRVILIFH
CEFSSERGPRMCRFIRERDRAVNDYPSLYPEMYILKGGYKEFFPQHPNFCEPQDYRPMNHE
AFKDELKTFALKTRSWAGERSRRELCSRLQDQ Stop

Sequência construída:

MGSSHHHHHHSSGLVPRGSHMEFQSDHRELIGDYSKAFLLQTVDGKHQDLKYISPETMVALL
TGKFSNIVDKFVIVDCRYPYEYEGGHIKTAVNLPLERDAESFLLKSPIAPCSLDRVILIFH
CEFSSERGPRMCRFIRERDRAVNDYPSLYPEMYILKGGYKEFFPQHPNFCEPQDYRPMNHE
AFKDELKTFALKTRSWAGERSRRELCSRLQDQ Stop

Comparação ExPASy (em relação à sequência selvagem):

Sequence ID: Query_204645 Length: 218 Number of Matches: 1

Range 1: 1 to 218 [Graphics](#)

▼ Next Match ▲ Previous Match

NW Score	Identities	Positives	Gaps
1180	217/218(99%)	217/218(99%)	0/218(0%)
Query 1	MGSSHHHHHHSSGLVPRGSHMEFQSDHRELIGDYSKAFLLQTVDGKHQDLKYISPETMVA		60
Sbjct 1	MGSSHHHHHHSSGLVPRGSHMEFQSDHRELIGDYSKAFLLQTVDGKHQDLKYISPETMVA		60
Query 61	LLTGKFSNIVDKFVIVDCRYPYEYEGGHIKTAVNLPLERDAESFLLKSPIAPCSLDRVI		120
Sbjct 61	LLTGKFSNIVDKFVIVDCRYPYEYEGGHIKTAVNLPLERDAESFLLKSPIAPCSLDRVI		120
Query 121	LIFHCEFSSERGPRMCRFIRERDRAVNDYPSLYPEMYILKGGYKEFFPQHPNFCEPQDY		180
Sbjct 121	LIFHCEFSSERGPRMCRFIRERDRAVNDYPSLYPEMYILKGGYKEFFPQHPNFCEPQDY		180
Query 181	RPMNHEAFKDELKTFRLKTRSWAGERSRRELCSRLQDQ 218		
Sbjct 181	RPMNHEAFKDELKTF LKTRSWAGERSRRELCSRLQDQ 218		

W550A-17N:

>A_07Dez_2016_W185A_6_Lud_F_Mix_1_D01_007.ab1

GAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCA
TCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGAATTCCAGAGTGACCACCGAGAGC
TGATTGGAGATTACTCTAAGGCCTTCCTCCTACAGACAGTAGACGGAAAGCACCAAGACCTC
AAGTACATCTCACCAGAAACGATGGTGGCCCTATTGACGGGCAAGTTCAGCAACATCGTGGA
TAAGTTTGTGATTGTAGACTGCAGAT

>A_07Dez_2016_W185A_6_Lud_R_Mix_2_D02_008.ab1

ANEXO B - Deslocamentos químicos domínio catalítico Cdc25B e mutantes

São apresentados os valores de deslocamento químico obtidos para as frequências de ^1H e ^{15}N para cada construção produzida. São indicados os resíduos da sequência primária e se estes foram (S) ou não (N) anotados em Sayegh 2016 [24].

Resíduo	Anotado	C1 - 17N		R482A - 17N		R544A - 17N		R544A - 17N	
		$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$
369 – Met	S	-	-	8,1697	121,3542	-	-	8,1848	121,3567
370 – Glu	S	8,2365	121,9034	8,2391	121,9188	8,2436	121,9203	8,2415	121,9156
371 – Phe	S	8,1198	120,6237	8,1228	120,6267	8,1253	120,6295	8,1245	120,6275
372 – Gln	S	8,1842	121,8959	8,1855	121,9041	8,1878	121,8872	8,1856	121,8893
373 – Ser	S	8,2349	117,0991	8,2374	117,1064	8,2425	117,1050	8,2437	117,1293
374 – Asp	S	8,2356	121,8589	8,2354	121,8346	8,2445	121,8684	8,2351	121,8470
375 – His	S	8,2588	119,7385	8,2638	119,8475	8,2557	119,8164	8,2539	119,8162
376 – Arg	S	8,1183	121,4299	8,1278	121,4147	8,1319	121,4420	8,1255	121,4420
377 – Glu	S	8,6004	120,9768	8,5958	120,9727	8,6035	120,9992	8,6018	120,9887
378 – Leu	N	-	-	-	-	-	-	-	-
379 – Ile	N	-	-	-	-	-	-	-	-
380 – Gly	N	-	-	-	-	-	-	-	-
381 – Asp	N	-	-	-	-	-	-	-	-
382 – Tyr	N	-	-	-	-	-	-	-	-
383 – Ser	N	-	-	-	-	-	-	-	-
384 – Lys	N	-	-	-	-	-	-	-	-
385 – Ala	N	-	-	-	-	-	-	-	-
386 – Phe	N	-	-	-	-	-	-	-	-
387 – Leu	N	-	-	-	-	-	-	-	-
388 – Leu	N	-	-	-	-	-	-	-	-
389 – Gln	S	8,4438	120,7520	8,4459	120,7580	8,4467	120,7651	8,4456	120,7630

Resíduo	Anotado	C1 - 17N		R482A - 17N		R544A - 17N		R544A - 17N	
		$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$
390 – Thr	S	8,0039	114,9272	8,0061	114,9271	8,0073	114,9238	8,0049	114,9184
391 – Val	S	8,2384	113,5530	8,2411	113,5623	8,2419	113,5579	8,2426	113,5668
392 – Asp	S	8,2774	121,3497	8,2800	121,3351	8,2790	121,3565	8,2814	121,3614
393 – Gly	S	7,6530	108,6586	7,6527	108,6131	7,6537	108,6472	7,6556	108,6564
394 – Lys	S	7,6611	119,3468	7,6466	119,3384	7,6531	119,3610	7,6544	119,3681
395 – His	S	7,9073	115,6821	7,9235	115,7525	7,9167	115,6969	7,9107	115,6871
396 – Gln	S	8,9568	123,6670	8,9561	123,6854	8,9601	123,6758	8,9598	123,6699
397 – Asp	S	9,3436	114,6371	9,3689	114,6370	9,3510	114,6352	9,3487	114,6372
398 – Leu	S	7,3731	120,3990	7,3772	120,3209	7,3734	120,3439	7,3754	120,3991
399 – Lys	S	8,4571	118,7038	8,4596	118,6930	8,4629	118,6908	8,4619	118,7091
400 – Tyr	S	8,8284	122,0995	8,8223	122,0834	8,8355	122,0835	8,8338	122,0919
401 – Ile	S	8,7651	110,0245	8,7706	110,0245	8,7666	110,0217	8,7670	110,0230
402 – Ser	S	8,7114	116,8284	8,7171	116,8247	8,7148	116,8321	8,7158	116,8425
403 – Pro	N	-	-	-	-	-	-	-	-
404 – Glu	S	8,2739	116,2778	8,2780	116,2919	8,2771	116,2785	8,2742	116,2805
405 – Thr	S	7,7876	121,6844	7,7942	121,7105	7,7948	121,6881	7,7949	121,7240
406 – Met	S	7,6663	121,8524	7,6686	121,8530	7,6688	121,8563	7,6693	121,8625
407 – Val	S	8,0126	119,3114	8,0201	119,3234	8,0158	119,3081	8,0133	119,3105
408 – Ala	S	7,6593	125,0842	7,6591	125,0781	7,6611	125,0683	7,6608	125,0800
409 – Leu	S	7,6764	118,2914	7,6768	118,2610	7,6784	118,2855	7,6789	118,2980
410 – Leu	S	8,0253	119,6886	8,0297	119,6994	8,0323	119,7152	8,0269	119,6803
411 – Thr	S	7,9038	107,3404	7,9036	107,3567	7,9072	107,3675	7,9068	107,3679
412 – Gly	S	7,5855	108,1783	7,5873	108,1920	7,5877	108,1818	7,5865	108,1861
413 – Lys	S	8,1747	119,6003	8,1642	119,5227	8,1614	119,3213	8,1693	119,5387
414 – Phe	S	8,3367	114,0344	8,3423	114,0574	8,3408	114,0474	8,3399	114,0392

Resíduo	Anotado	C1 - 17N		R482A - 17N		R544A - 17N		R544A - 17N	
		$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$
415 – Ser	S	7,5853	117,9944	7,5900	117,9911	7,5913	118,0055	7,5870	118,0042
416 – Asn	S	8,6291	117,1882	8,6335	117,1291	8,6364	117,1506	8,6459	117,1279
417 – Ile	N	-	-	-	-	-	-	-	-
418 – Val	N	-	-	-	-	-	-	-	-
419 – Asp	S	9,1378	128,6921	9,1424	128,6882	9,1392	128,6684	9,1397	128,6835
420 – Lys	N	-	-	-	-	-	-	-	-
421 – Phe	S	7,9375	118,4843	7,9349	118,3398	7,9431	118,4265	7,9323	118,4306
422 – Val	S	7,9370	121,3877	7,9496	121,5472	7,9428	121,4344	7,9375	121,4016
423 – Ile	S	8,9399	128,0341	8,9403	128,0268	8,9440	128,0513	8,9401	128,0357
424 – Val	S	9,4246	128,8101	9,4332	128,7945	9,4325	128,8067	9,4264	128,8024
425 – Asp	S	8,6770	127,2415	8,6788	127,2695	8,6827	127,2691	8,6810	127,2587
426 – Cys	S	8,2800	124,6364	8,2541	124,6813	8,2592	124,6164	8,2809	124,6342
427 – Arg	S	7,6244	117,3884	7,5988	117,3574	7,6054	117,3190	7,6303	117,3879
428 – Tyr	S	7,5778	118,1335	7,5864	118,1961	7,5922	118,2159	7,5759	118,1391
429 – Phe	N	-	-	-	-	-	-	-	-
430 – Tyr	S	7,0160	110,3460	7,0354	110,3815	7,0147	110,3861	7,0201	110,3653
431 – Glu	S	6,1132	123,3528	6,1258	123,2888	6,1192	123,3790	6,1167	123,3543
432 – Tyr	S	6,9532	120,6831	6,9593	120,7182	6,9725	120,6736	6,9517	120,6835
433 – Glu	S	9,0603	120,7614	9,0578	120,7708	9,0833	120,7991	9,0660	120,7574
434 – Gly	S	6,9382	103,9080	6,9361	103,8852	6,9449	103,9039	6,9421	103,9135
435 – Gly	S	6,4532	106,5560	6,4558	106,5503	6,4601	106,5589	6,4561	106,5596
436 – His	S	7,5237	123,9217	7,5234	123,9263	7,5316	123,8908	7,5269	123,9186
437 – Ile	S	10,5982	123,8294	10,5997	123,8699	10,5948	123,8479	10,5974	123,8602
438 – Lys	N	-	-	-	-	-	-	-	-
439 – Thr	S	8,7512	113,5655	8,7611	113,5822	8,7589	113,5330	8,7584	113,5683

Resíduo	Anotado	C1 - 17N		R482A - 17N		R544A - 17N		R544A - 17N	
		$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$
465 – Lys	N	-	-	-	-	-	-	-	-
466 – Arg	N	-	-	-	-	-	-	-	-
467 – Val	N	-	-	-	-	-	-	-	-
468 – Ile	S	8,0089	123,3550	8,0142	123,3135	8,0151	123,3543	8,0105	123,3608
469 – Leu	S	7,4379	121,4815	7,4337	121,4888	7,4427	121,5082	7,4399	121,4862
470 – Ile	S	8,7974	122,1613	8,7975	122,1593	8,8039	122,1549	8,8026	122,1603
471 – Phe	S	9,5377	123,7449	9,5326	123,7206	9,5423	123,7315	9,5369	123,7290
472 – His	S	8,9189	113,9914	8,9174	113,9428	8,9199	113,9453	8,9202	113,9979
473 – Cys	N	-	-	-	-	-	-	-	-
474 – Glu	N	-	-	-	-	-	-	-	-
475 – Phe	N	-	-	-	-	-	-	-	-
476 – Ser	N	-	-	-	-	-	-	-	-
477 – Ser	N	-	-	-	-	-	-	-	-
478 – Glu	N	-	-	-	-	-	-	-	-
479 – Arg	N	-	-	-	-	-	-	-	-
480 – Gly	N	-	-	-	-	-	-	-	-
481 – Pro	N	-	-	-	-	-	-	-	-
482 – Arg	S	7,6860	115,5556	-	-	7,6952	115,6084	7,6883	115,5863
483 – Met	S	8,0011	121,0682	7,9927	120,5095	7,9896	121,0487	8,0038	121,0507
484 – Cys	S	7,7368	119,7838	7,6282	119,9826	7,7232	119,7986	7,7436	119,7916
485 – Arg	S	7,6568	116,4347	7,7048	116,4051	7,6666	116,4396	7,6642	116,4325
486 – Phe	S	8,0841	121,5584	8,0828	121,4934	8,0886	121,5123	8,0837	121,5726
487 – Ile	S	8,2589	118,9539	8,2531	118,8584	8,2554	118,9440	8,2637	118,9628
488 – Arg	S	7,5541	118,3932	7,5681	118,4206	7,5531	118,4445	7,5550	118,3906
489 – Glu	S	7,9865	119,6481	7,9943	119,6452	7,9871	119,6440	7,9889	119,6407

Resíduo	Anotado	C1 - 17N		R482A - 17N		R544A - 17N		R544A - 17N	
		$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$
490 – Arg	S	7,9496	119,3919	7,9433	119,3275	7,9538	119,3875	7,9517	119,3869
491 – Asp	S	8,7534	121,8108	8,7583	121,7798	8,7546	121,8180	8,7557	121,8231
492 – Arg	S	8,4665	116,6327	8,4572	116,5984	8,4681	116,6213	8,4700	116,6323
493 – Ala	S	7,5595	118,9951	7,5686	119,0248	7,5625	118,9780	7,5629	119,0113
494 – Val	S	7,3349	108,8658	7,3453	108,8997	7,3370	108,8722	7,3384	108,8568
495 – Asn	S	7,0284	120,4672	7,0271	120,4566	7,0300	120,4616	7,0309	120,4653
496 – Asp	S	8,7369	120,5663	8,7385	120,5720	8,7395	120,5763	8,7381	120,5626
497 – Tyr	S	8,7789	128,1642	8,7800	128,1738	8,7823	128,1765	8,7816	128,1680
498 – Pro	N	-	-	-	-	-	-	-	-
499 – Ser	N	-	-	-	-	-	-	-	-
500 – Leu	N	-	-	-	-	-	-	-	-
501 – Tyr	N	-	-	-	-	-	-	-	-
502 – Tyr	N	-	-	-	-	-	-	-	-
503 – Pro	N	-	-	-	-	-	-	-	-
504 – Glu	N	-	-	-	-	-	-	-	-
505 – Met	S	9,8508	116,1517	9,8519	116,1196	9,8515	116,1451	9,8548	116,1558
506 – Tyr	S	8,7974	117,7600	8,8089	117,7639	8,8017	117,7394	8,7999	117,7776
507 – Ile	S	9,3656	120,2414	9,3779	120,2386	9,3720	120,2395	9,3688	120,2416
508 – Leu	S	7,6915	125,1729	7,7045	125,1921	7,6970	125,1657	7,6935	125,1758
509 – Lys	S	9,6347	136,7302	9,6430	136,7911	9,6373	136,7796	9,6375	136,7351
510 – Gly	S	9,5174	118,3584	9,5202	118,3667	9,5216	118,3733	9,5170	118,3764
511 – Gly	S	7,3489	101,7651	7,3605	101,7705	7,3602	101,7690	7,3542	101,7771
512 – Tyr	N	-	-	-	-	-	-	-	-
513 – Lys	S	8,8557	115,5082	8,8565	115,5186	8,8493	115,4812	8,8577	115,5252
514 – Glu	S	6,5599	111,3952	6,5636	111,3740	6,5679	111,3852	6,5617	111,4026

Resíduo	Anotado	C1 - 17N		R482A - 17N		R544A - 17N		R544A - 17N	
		$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$
515 – Phe	S	7,3655	121,2954	7,3606	121,2645	7,3666	121,2603	7,3658	121,2924
516 – Phe	S	7,7143	117,8706	7,7190	117,9323	7,7206	117,8004	7,7184	117,8249
517 – Pro	N	-	-	-	-	-	-	-	-
518 – Gln	S	6,2762	112,4243	6,2762	112,4360	6,2784	112,4156	6,2786	112,4115
519 – His	S	7,0723	115,5031	7,0791	115,5162	7,0825	115,5032	7,0736	115,5036
520 – Pro	N	-	-	-	-	-	-	-	-
521 – Asn	S	8,3100	111,5656	8,3151	111,6077	8,3157	111,5879	8,3112	111,5702
522 – Phe	S	7,9772	117,8960	7,9822	117,9171	7,9880	117,9687	7,9816	117,9147
523 – Cys	S	7,6779	119,3141	7,6853	119,3074	7,6835	119,3106	7,6793	119,3199
524 – Glu	S	9,2649	122,0967	9,2638	122,0868	9,2679	122,0915	9,2675	122,1061
525 – Pro	N	-	-	-	-	-	-	-	-
526 – Gln	S	8,4829	119,3455	8,4856	119,3411	8,4853	119,3163	8,4843	119,3368
527 – Asp	S	7,9737	123,8353	7,9649	123,8094	7,9724	123,8380	7,9765	123,8351
528 – Tyr	S	8,9883	120,1504	8,9960	120,1522	9,0001	120,1877	8,9874	120,1445
529 – Arg	S	9,0891	133,8469	9,0801	133,8421	9,1009	133,7704	9,0920	133,8785
530 – Pro	N	-	-	-	-	-	-	-	-
531 – Met	S	8,3611	122,7962	8,3216	122,7783	8,3814	122,5184	8,3504	122,8172
532 – Asn	S	8,4801	113,8227	8,4775	113,8602	8,4463	113,8388	8,4821	113,8516
533 – His	S	7,4668	120,5070	7,4627	120,4738	7,4753	120,4633	7,4732	120,5071
534 – Glu	S	8,8727	130,2631	8,8754	130,2533	8,8414	129,9317	8,8745	130,2353
535 – Ala	S	10,5900	124,3510	10,5909	124,3272	10,5202	124,3756	10,6055	124,3341
536 – Phe	S	7,7671	117,2415	7,7680	117,2036	7,7683	117,3269	7,7736	117,2760
537 – Lys	S	7,1599	119,2764	7,1535	119,3175	7,1932	119,4161	7,1706	119,2746
538 – Asp	S	8,6342	119,9037	8,6257	119,8180	8,6039	119,5371	8,6464	119,9173
539 – Glu	S	8,0779	123,4413	8,0604	123,4634	8,0702	123,1948	8,0993	123,5129

Resíduo	Anotado	C1 - 17N		R482A - 17N		R544A - 17N		R544A - 17N	
		$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$
540 – Leu	S	8,3159	122,1216	8,3072	122,0862	8,3199	122,0509	8,3114	122,0460
541 – Lys	S	7,6870	118,0939	7,6780	117,8890	7,6817	117,7303	7,7034	118,1149
542 – Thr	S	7,9515	114,3642	7,9199	114,0816	7,7868	113,6511	7,9803	114,5714
543 – Phe	S	8,3988	121,6350	8,3458	121,6148	-	-	8,4389	121,7170
544 – Arg	S	8,6048	120,0818	8,5443	119,6105	-	-	8,6582	120,3873
545 – Leu	S	7,5254	119,0346	7,5032	118,8916	7,5047	117,4844	7,5461	119,3181
546 – Lys	S	7,7782	118,1103	7,6975	117,0380	7,7427	118,5228	7,8313	118,4431
547 – Thr	S	7,5620	109,8806	7,5810	110,6398	7,6620	109,9162	7,6121	109,7940
548 – Arg	S	7,7526	121,6818	7,8184	121,6697	7,7892	121,5044	7,7110	122,0038
549 – Ser	S	8,1995	116,0596	8,1983	116,0323	8,2295	115,8821	8,2414	116,7043
550 – Trp	S	8,0815	123,3007	8,0501	123,0902	8,0616	123,0782	-	-
	S	10,1694	129,7572	10,1882	129,8075	10,1799	129,7265	-	-
551 – Ala	S	8,0908	125,3069	8,0477	125,1885	8,1046	125,4589	8,2865	126,1019
552 – Gly	S	7,6084	107,1490	7,6184	107,0663	7,5680	107,0771	8,2511	107,8433
553 – Glu	S	8,1450	120,9866	8,1566	120,9047	8,1460	121,1051	8,1526	120,9365
554 – Arg	S	8,3203	121,4071	8,3115	121,3757	8,3198	121,3521	8,3342	122,2202
555 – Ser	N	-	-	-	-	-	-	-	-
556 – Arg	S	8,1889	122,5805	8,1840	122,6529	8,1953	122,6318	8,2279	122,7119
557 – Arg	S	8,1630	121,2606	8,1607	121,1640	8,1744	121,2694	8,1724	121,2818
558 – Glu	S	8,2808	121,0438	8,2816	121,0128	8,2690	120,6936	8,3024	121,2082
559 – Leu	S	8,0997	122,2593	8,1024	122,2509	8,0608	121,8651	8,1242	122,3745
560 – Cys	S	8,1567	119,3074	8,1616	119,3659	8,1214	119,0881	8,1617	119,3357
561 – Ser	N	-	-	-	-	-	-	-	-
562 – Arg	S	8,2688	122,7940	8,2606	122,7742	8,2336	122,6864	8,2745	122,7993
563 – Leu	S	8,1439	122,7235	8,1456	122,7206	8,1328	122,6506	8,1510	122,7329

Resíduo	Anotado	C1 - 17N		R482A - 17N		R544A - 17N		R544A - 17N	
		$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$	$\delta^1\text{H}$	$\delta^{15}\text{N}$
564 – Gln	S	8,2078	119,8584	8,2094	119,8814	8,2063	119,8607	8,2163	119,8877
565 – Asp	S	8,2148	121,4173	8,2182	121,4257	8,2164	121,4129	8,2189	121,4293
566 – Gln	S	7,7392	124,5620	7,7422	124,5757	7,7432	124,5768	7,7428	124,5721

ANEXO C - Artigo publicado

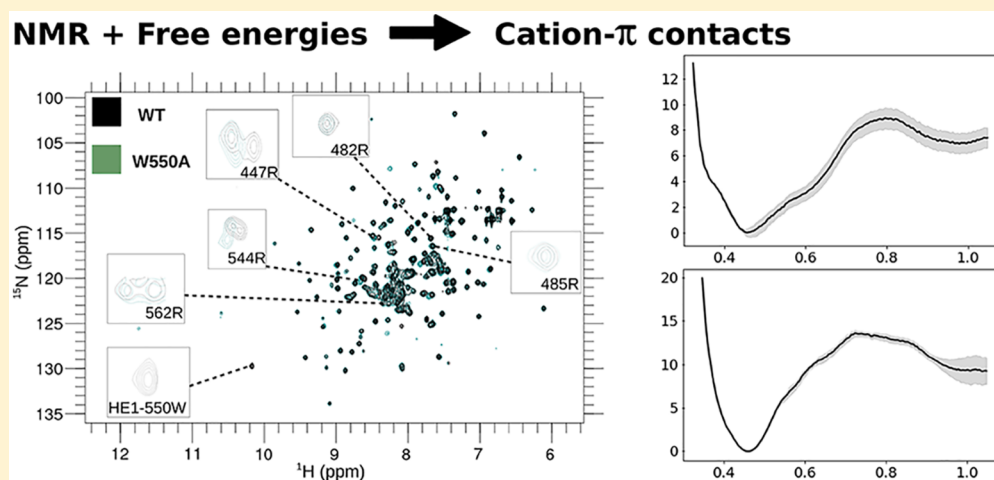
Reis, AOR; Sayegh, RSR; Marana, SR; Arantes, GM. **Combining free energy simulations and NMR chemical-shift perturbation to identify transient cation- π contacts in proteins.** *Journal of Chemical Information and Modeling* (2020) 60: 890 - 897.

Combining Free Energy Simulations and NMR Chemical-Shift Perturbation To Identify Transient Cation– π Contacts in Proteins

André A. O. Reis, Raphael S. R. Sayegh, Sandro R. Marana, and Guilherme M. Arantes*[†]

Department of Biochemistry, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, 05508-900 São Paulo, São Paulo, Brazil

S Supporting Information



ABSTRACT: Flexible protein regions containing cationic and aromatic side-chains exposed to solvent may form transient cation– π interactions with structural and functional roles. To evaluate their stability and identify important intramolecular cation– π contacts, a combination of free energy profiles estimated from umbrella sampling with molecular dynamics simulations and chemical shift perturbations (CSP) obtained from nuclear magnetic resonance (NMR) experiments is applied here to the complete catalytic domain of human phosphatase Cdc25B. This protein is a good model system for transient cation– π interactions as it contains only one Trp residue (W550) in the disordered C-terminal segment and a total of 17 Arg residues, many exposed to solvent. Eight putative Arg–Trp pairs were simulated here. Only R482 and R544 show bound profiles corresponding to important transient cation– π interactions, while the others have dissociative or almost flat profiles. These results are corroborated by CSP analysis of three Cdc25B point mutants (W550A, R482A, and R544A) disrupting cation– π contacts. The proposed validation of statistically representative molecular simulations by NMR spectroscopy could be applied to identify transient contacts of proteins in general but carefully, as NMR chemical shifts are sensitive to changes in both molecular contacts and conformational distributions.

1. INTRODUCTION

Proteins are stabilized by multiple noncovalent intramolecular contacts such as hydrogen bonds, hydrophobic contacts, and ion pairs. Cation– π interactions formed between cationic residues (in Arg and Lys) and aromatic rings (in Phe, Tyr and Trp) have comparable interaction energies mainly of electrostatic origin¹ and are easy to recognize when buried in folded proteins.² Cation– π contacts may also be observed between side-chains exposed to solvent due to the competition of aromatic rings with water in binding cations in aqueous environments.^{1,3–5}

The lifetime of a noncovalent contact in aqueous solution is determined by its stability or free energy, and weak contacts will survive briefly due to structural fluctuations and water competition.⁶ Flexible protein regions such as side-chains

exposed to solvent, loops, and intrinsically disordered segments should have their structure represented by a conformational ensemble in which the stability of a possible noncovalent contact is proportional to its population or statistical importance in the configurational distribution.⁷ However, ensembles commonly obtained from trajectories of molecular dynamics (MD) simulation often sample the underlying equilibrium distribution inappropriately or incompletely.⁸ Not seldom, a single observation of a transient contact during

Special Issue: Molecular Simulation in Latin America: Coming of Age

Received: September 27, 2019

Published: November 18, 2019

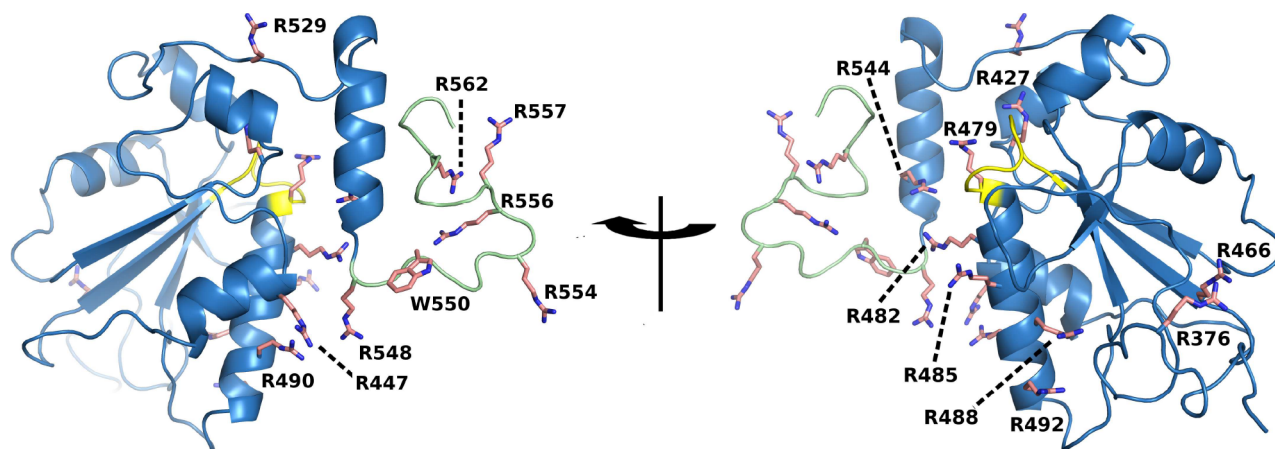


Figure 1. Structure of the complete catalytic domain of human phosphatase Cdc25B taken from a molecular dynamics snapshot and shown in cartoon with indicated Arg and Trp side-chains in pink sticks. The P-loop catalytic site is in yellow, and the disordered C-terminal is in pale green.

a simulation is taken as evidence of its structural or functional role^{9–11} but without statistical significance.¹²

Here, the statistical importance of cation– π interactions formed between protein side-chains is investigated with a combination of molecular simulation and experimental spectroscopy. Enhanced sampling free energy simulations are employed. In principle, these may lead to correct contact distributions if a meaningful collective or reaction coordinate describing the transient interaction is chosen and orthogonal degrees of freedom are properly sampled.⁸ But, given the approximate description of interaction energies used in biomolecular simulations and finite sampling available in practice, an experimental validation of putative transient contacts will help to confirm the computational predictions.

Nuclear magnetic resonance (NMR) is exquisitely sensitive to the nuclei chemical environment and is routinely used to determine intramolecular contacts of proteins in solution.¹³ For instance, cation– π interactions will shift Arg and Lys side-chain resonances upfield because of diamagnetic anisotropic shielding from a nearby indole ring in Trp.⁴ But NMR chemical shifts also report on the distribution of protein conformations. The nuclear Overhauser effect (NOE) is frequently used to discriminate shifts induced by molecular contact from those of conformational origin.¹³ Here, only standard ¹H–¹⁵N HSQC spectra were explored to avoid additional NMR experiments or extensive assignments of side-chain resonances,¹⁴ so that the proposed experimental validation is kept as simple and general as possible. It should be noted that other spectroscopical methods such as visible absorbance and fluorescence have recently been used to identify cation– π interactions in solution¹⁵ and could also be employed to validate simulations of transient contacts if enough spectral resolution is available.

The complete catalytic domain of human phosphatase Cdc25B was studied here as a model system.^{16,17} This enzyme catalyzes the dephosphorylation of pTyr and pThr from other peptides^{18–21} and has already been investigated by solution NMR.^{22,23} It contains only one Trp residue (W550) located in the disordered C-terminus^{23,24} as well as a total of 17 Arg (5 in the disordered C-terminus) and 13 Lys. Many of these Arg side-chains are exposed on the protein surface near the P-loop catalytic site and W550 (Figure 1). The large number of cationic residues in Cdc25B may be related to anionic substrate recognition and product release.¹⁷ Yet, multiple

Trp–Arg contacts were previously observed in a long MD simulation of Cdc25B.^{23,25}

In the following section the simulation and experimental methods are described with special care to ensure the approximate force-field used in the protein simulations reproduces the cation– π energetics accurately. Then, free energy profiles estimated from umbrella sampling (US) simulations are presented for the formation of eight putative cation– π contacts between side-chains of Trp–Arg pairs in Cdc25B. These results are validated by chemical shift perturbations obtained from NMR spectra of Cdc25B wild-type and point mutants disrupting putative cation– π contacts to identify important transient interactions in the protein conformational ensemble.

2. MATERIALS AND METHODS

2.1. Quantum Chemistry, Molecular Dynamics, and Free Energy Simulations. To evaluate the description of cation– π interactions by approximate force-fields, an isolated complex in the gas phase between blocked tryptophan (Ac-Trp-NHMe, denoted as bTrp) and blocked arginine (bArg) was used as a model system. The interaction energy with side-chains placed in two relevant orientations,¹ sandwich and T-shape (Figure 2), was evaluated quantum mechanically at the MP2/6-311++G** level corrected for the basis set superposition error with the counterpoise correction²⁶ and compared to CHARMM36²⁷ and AMBER99-ILDN^{28,29} force-fields in the same geometries. The GAUSSIAN 09 rev. A program was used for quantum calculations.³⁰

The reaction coordinate used here to describe cation– π interactions for both the bTrp...bArg complex and the Cdc25B free energy profiles below is the distance between the center-of-mass (COM) of all atoms in the Trp and in the Arg side-chains. These COMs roughly correspond to the position of Trp-C δ 2 and Arg-N ϵ atoms.

Molecular dynamics simulations of the complete catalytic domain of the human phosphatase Cdc25B were initiated from the structure in PDB code 1QB0. Sixteen residues (A551–Q566) lacking from the native C-terminal structure¹⁶ and seven residues in the N-terminal part of the experimental construction (see below) were manually added in an extended conformation.²³ Crystallographic water, ions, and β -mercaptoethanol (BMER) were removed. Hydrogen atoms were added, and a standard protonation state at pH = 7 was

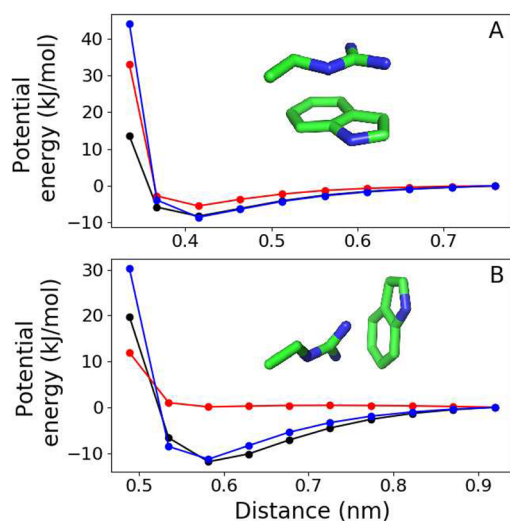


Figure 2. Potential energy for the interaction between Trp and Arg side-chains (bTrp···bArg complex) in sandwiched (panel A) and T-shaped (panel B) orientations, as shown in inserts. Black dots show the quantum chemical reference (MP2/6-311++G**), blue dots are the CHARMM36 force field, and red dots show the AMBER99-ILDN force field. Lines are guides to the eye.

assumed for side-chains. This is equivalent to the setup we used previously.²³ Two inorganic phosphate ions (HOPO_3^- form) were added manually to the catalytic P-loop (R479) and to the secondary phosphate binding site near R488 and R492 side-chains as previous data such as phosphate titrations strongly support that these sites will bind phosphate ions under our experimental conditions.^{17,23,25} This model was solvated in a dodecahedral simulation box of 13234 water molecules, with 4 Na^+ and 2 Cl^- ions added to neutralize the system and mimic the experimental salt concentration. This system was relaxed and equilibrated during a 100 ns free MD simulation. The CHARMM36 force-field for protein and ions and the TIP3P water model³¹ were used. Simulations were performed with a 2 fs integration step, 300 K temperature, and 1 bar pressure (NPT ensemble) using the Bussi thermostat³² and Parrinello–Rahman barostat both with a period of 0.5 ps. PME method with 1.2 nm real space cutoff and 0.12 nm Fourier spacing was used to treat long-range electrostatics. All classical force-field calculations, MD simulations, and analysis were performed with the GROMACS 4.6.7 program.³³

Free energy profiles were estimated with US.³⁴ Initial configurations for each window were obtained after model equilibration described above. Seven US windows with reference reaction coordinate separated by 0.1 nm, from 0.4 to 1.0 nm, were used to estimate the profiles for each of Trp–Arg pair. A harmonic potential with force constant $k_{\text{umb,dCOM}} = 2000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ was used. Each US window was sampled for 500 ns with a sample collected every ps. Thus, the total aggregate simulation time was 20 μs . Potentials of mean-force were obtained with WHAM,³⁴ and the statistical uncertainty was estimated as 95% confidence intervals by bootstrap analysis with 50 resampling steps.³⁵ The initial 50 ns of each window was discarded to allow equilibration of orthogonal degrees of freedom in all analysis presented. The overlap of reaction coordinate distribution between adjacent windows was confirmed.

2.2. Point Mutants, Protein Expression, and Purification. Cdc25B W550A, R482A, and R544A point mutants were

generated from pET-28a plasmids with the complete wild-type Cdc25B catalytic domain (S373–Q566) cloned and fused to His-tag and thrombin cleavage sites. The N-terminal sequence after cleavage was GSHMEFQ (373)SDHRELI···, where the first seven residues are not native. Point mutations were introduced by using the QuickChange Lightning Site-Directed Mutagenesis kit (Agilent Tech. #210518).

Expression of wild-type Cdc25B and point mutants was made after *Escherichia coli* BL21-Gold (DE3) transformation with the respective plasmids and growth in M9 minimal ($^{15}\text{NH}_4\text{Cl}$ -labeled) medium at 37 °C until $\text{OD}_{600\text{nm}} 0.6$ (exponential growth phase). At this time, 0.8 mM isopropyl β -thiogalactopyranoside (IPTG) solution was added for induction of protein expression, and temperature was maintained at 20 °C for 18 h when bacteria were then collected by centrifugation ($7000 \times g$ -20 min). Bacteria were lysed by sonication in lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 10% glycerol, 5 mM 2-mercaptoethanol (BMER), 1 mM phenylmethylsulfonylfluoride-PMSF, pH = 7.4). The lysate was subjected to centrifugation, and supernatant was subjected to column affinity (Ni-NTA) and size-exclusion (Superdex 75 HR 10/30) chromatographies to protein purification as previously described in detail.^{23,36}

2.3. NMR Spectra. All samples were prepared using 0.2 mM concentration of uniformly ^{15}N -labeled proteins in the same phosphate buffer preparation (20 mM NaH_2PO_4 , 50 mM NaCl, 5 mM BMER, 2 mM dithiothreitol, 5% $^2\text{H}_2\text{O}$, pH = 6.7) to obtain equivalent environmental conditions. NMR spectra were acquired on Bruker Ascend III 800 MHz NMR apparatus with 4-channel TCI cryoprobe and Avance III console. Initial processing of the spectra was done with NMRPipe program,³⁷ and assignment transfer for the main and side-chain resonances was performed with CCPNMR Analysis Assign 2.4 program.³⁸

The chemical shift perturbation (CSP) analyzed for each assigned N–H system was defined according to

$$\text{CSP} = \sqrt{\frac{1}{j_\alpha}(\Delta\delta^1\text{H})^2 + \frac{1}{k_\alpha}(\Delta\delta^{15}\text{N})^2}$$

where $\Delta\delta$ is the chemical shift change upon mutation, j_α and k_α are scaling factors (Table S1) that correspond to standard deviation of ^1H and ^{15}N chemical shifts of N–H system α (thus, unique to each amino acid type, main and side-chains) as observed in the RefDB database.³⁹ Statistical analysis was performed by the iterated removal of outliers defined as the values above the mean (\bar{x}) plus 3 standard deviations (σ).⁴⁰ CSP was considered significant when larger than $\bar{x} + \sigma$.⁴¹ Wild-type and mutant chemical shifts obtained here were deposited in the BMRB databank (accession numbers from 50030 to 50033).

3. RESULTS AND DISCUSSION

3.1. Benchmark of the Empirical Force Field. The accuracy of a molecular simulation will depend on both energetic description and conformational sampling. Before addressing the importance of transient cation– π interactions in the protein Cdc25B, we benchmarked the interactions between Trp and Arg side-chains by comparing two widely employed protein force fields to high-level quantum chemical calculations.

Figure 2 shows the CHARMM36 force field correctly describes the cation– π interaction in both sandwiched (or stacking, parallel) and T-shaped (or perpendicular) orienta-

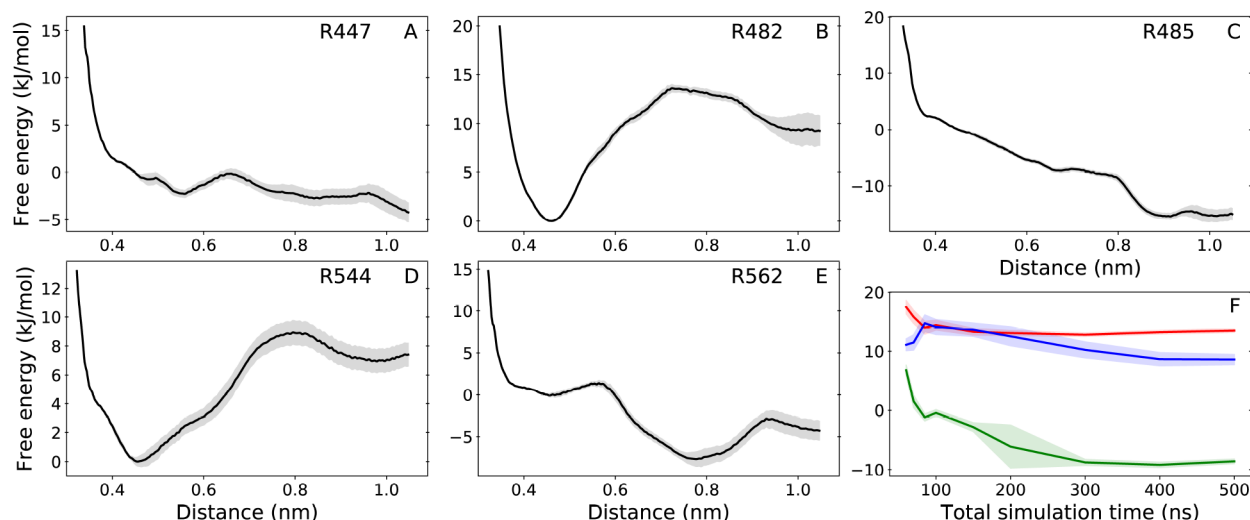


Figure 3. Free energy profiles for the cation- π interaction between W550 side-chain and R447 (panel A), R482 (B), R485 (C), R544 (D), and R562 (E) side-chains. Shadows around lines in all panels show the statistical uncertainty. (F) Convergence with total simulation time of the free energy difference between distance 0.45 and 0.80 nm for the cation- π complex with R482 in red line, R485 in green, and R544 in blue.

tions, with minimum energies within 0.5 kJ/mol of the MP2 energy reference. On the other hand, the AMBER99-ILDN energy function only agrees with the MP2 reference for the sandwiched orientation, with a difference of 3 kJ/mol for the minimum energy. Strikingly, the T-shaped orientation is fully nonbinding with the AMBER force field. Thus, CHARMM36 was employed for US simulations presented below, as this comparison suggests it should give a reliable energetic description for the mainly electrostatic interactions between Trp and Arg side-chains.¹

Discrepancies between quantum and classical energy descriptions in small distances (0.3 nm for sandwiched and 0.5 nm for T-shaped orientations, Figure 2) are due to the steep (and formally approximate) behavior of the repulsive component employed in force fields.⁴² This short-range discrepancy does not affect significantly the stability of cation- π interactions simulated here.

3.2. Free Energy Profiles for Trp-Arg Cation- π Interaction. Human Cdc25B catalytic domain contains 17 Arg, 13 Lys, and only one Trp residue. To reduce the number of simulations performed and simplify our analysis, we discarded Trp-Lys interactions and concentrated in Trp-Arg pairs, which interactions were shown above to be well described by the CHARMM36 force field. Arg residues visiting a distance of <8 Å from W550 during a previous long MD simulation (total of 6 μ s, Table S2)^{23,25} were considered. Partially buried residues (R479, R488 and R490) performing stable contacts with phosphate groups and R548 (next to W550 and impossible to check with the CSP analysis) were discarded. Thus, eight putative Trp-Arg pairs had the free energy for interaction between their side-chains estimated using US simulations. Results are shown in Figure 3 and Figure S1.

There are three kinds of free energy profiles. For R482 and R544 (Figure 3), the profile clearly shows a bound minimum at 0.45 nm distance and a dissociation barrier at 0.7–0.8 nm of 14 kJ/mol for R482 and 9 kJ/mol for R544, characterizing an important transient cation- π interaction with W550. For R485 (Figure 3) and R490 (Figure S1), the profile is downhill or dissociative, meaning that no stable interaction should be

observed. For R447, R562 (Figure 3), R554, and R556 (Figure S1), the profile shows very weak minima and is almost flat in bound distances (0.4–0.8 nm), characterizing possible but short-lived and statistically unimportant interactions with W550.

The free energy difference between distance 0.45 and 0.80 nm in the profiles of Figure 3, which corresponds to the dissociation barrier of bound profiles, converges to within one statistical uncertainty in <300 ns of total simulation time per US window (Figure 3F). Thus, free energy profiles were estimated here with twice more sampling time than necessary for barrier convergence. Note the pair between W550-R548 was not studied in detail. Their neighbor side-chains will trivially engage in a cation- π contact, but it is impossible to validate this contact with the experimental CSP analysis proposed below.

The minimum distance and dissociation barrier of bound profiles (Figure 3B,D) are in good agreement with the distance and interaction energy of the minima for the bTrp···bArg complex found in our benchmark calculations (Figure 2A). The distribution of the angle formed between the planes defined by the indole group of W550 and the guanidinium of R482 shown in Figure 4A spreads out and shifts to higher angle values as the US reference coordinate assumes longer distances. The distribution is narrower in short distances because the T-shaped orientation is sterically hindered. In agreement with previous studies,¹ a combination of mostly sandwiched but also T-shaped orientations is found for important transient cation- π interactions within the bound state (interval from 0.4 to 0.8 nm).

Given the large number of Arg residues in Cdc25B, many flanking W550 and the catalytic site, we asked which interactions would W550 perform when unbound from one of its important Arg partner. With R482 restricted to a long distance, Figure 4B shows that W550 side-chain spontaneously forms complexes with R544 and R548, with very short-lived contacts with other Arg residues (R562 in this case). Similarly, when R544 is held away, R482 and R548 form complexes with W550 (Figure S4). Thus, there is a competition of transient cation- π complexes with W550 in Cdc25B.

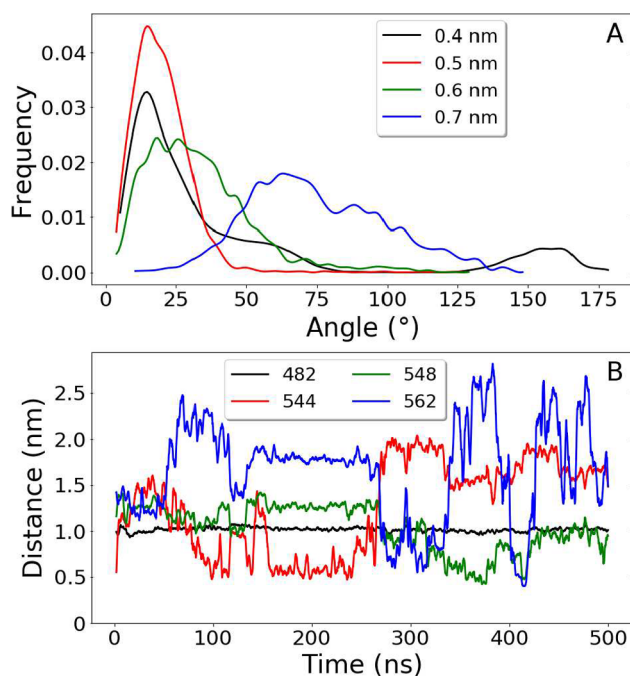


Figure 4. Analysis of Trp-Arg interactions during R482 US simulations. (A) Distribution of the angle formed between the planes defined by the indole group of W550 and the guanidinium of R482 with US reference coordinate centered in 0.4 (black), 0.5 (red), 0.6 (green), and 0.7 (blue) nm. Lower angles represent a sandwiched conformation and angles near 90 represent a T-shaped conformation, as in Figure 2. Data were smoothed with a spline interpolation. (B) Distance trajectories between the COM of W550 side-chain and the COM of R482 (black), R544 (red), R548 (green), and R562 (blue) side-chains, obtained with US reference coordinate in 1.0 nm.

3.3. Validation by NMR Chemical Shift Perturbation.

We attempted to experimentally validate the importance of transient cation- π interactions by comparing standard ^1H - ^{15}N HSQC spectra acquired for Cdc25B wild-type and point mutants W550A (Figure 5), R482A (Figure S2), and R544A (Figure S3) in solution. Our premise was that a

significant perturbation (i.e., larger than $\bar{x} + \sigma$) would be observed in the chemical shifts of residues performing important cation- π interactions by abolishing such contact upon a point mutation.

Main chain ^1H - ^{15}N and Trp $^{15}\text{N}\epsilon_1$ side-chain signals were analyzed here as they are generally available and easy to acquire. We avoided additional NMR experiments (for instance, 2D HE(NE)HGHH)¹⁴ that would be necessary to assign Arg and Lys side-chain resonances. Possible environmental perturbations were minimized by acquiring all spectra in the same batch and buffer solution. Assignments were transferred for 144 peaks out of the 150 originally assigned to wild-type Cdc25B,²³ and CSP was calculated for each point mutant, as shown in Table 1.

The CSP of cationic residues upon Trp mutation should be examined first. Insignificant perturbations found for R376, R427, R485, R488, R490, R492, and R562 in Table 1 imply that these residues do not perform important or long-lived cation- π interactions with W550. A similar conclusion is valid for most of the Lys residues (Table S3).

These observations are consistent with the Cdc25B crystal structure, where R376 and R427 are very distant from W550, but this flexible Trp residue approaches the other Arg located on the protein surface during a long MD simulation (Table S2). Why are their CSP insignificant? Side-chains of R488 and R492 make strong electrostatic interactions and remain buried within the secondary phosphate binding pocket.^{17,23} Side-chains of R485, R490, and R562 are fully exposed to solvent but do not perform a stable cation- π interaction with W550, as shown by the dissociative free energy profiles estimated here (Figures 3 and S1).

A significant CSP is observed for R482 and R544 in the W550A mutant, in line with their bound free energy profiles. However, a significant CSP is also observed for R447, R529, R554, R556, and R557, in apparent disagreement with the almost flat free energy profiles for R447, R554, and R556 and with the long distances to R529 and R557 (Table S2) that suggest no important interactions with W550 are formed. How can this be explained?

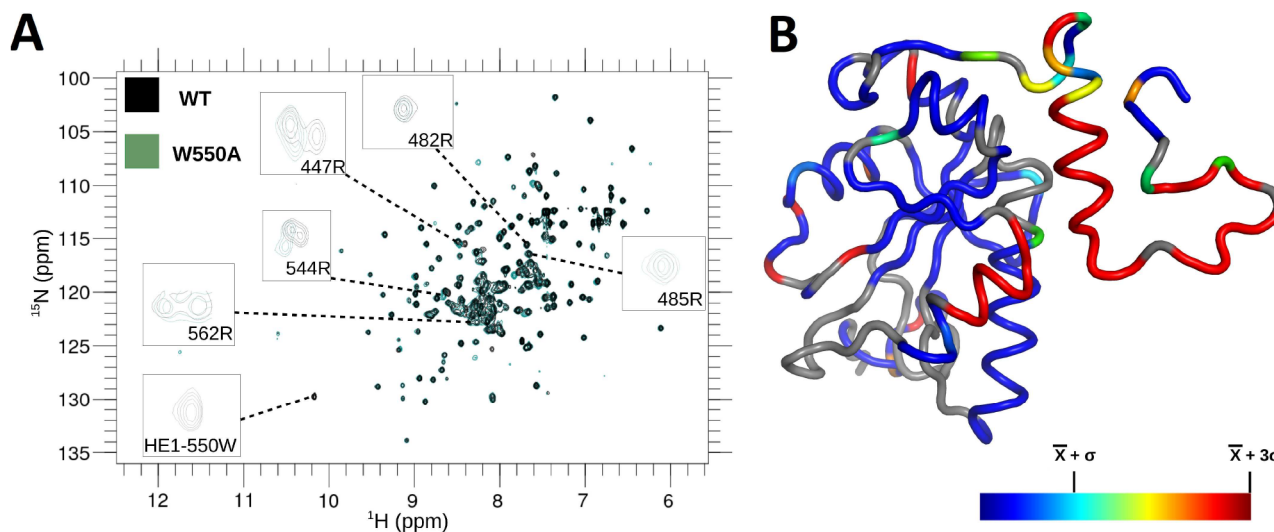


Figure 5. NMR spectra of Cdc25B catalytic domain and chemical shift perturbation upon mutation. (A) ^1H - ^{15}N HSQC for Cdc25B wild-type and W550A point mutant and (B) CSP projected according to colorbar over the same structure shown in Figure 1. Gray segments were not assigned.

Table 1. CSP of Arg and Trp Main Chain N–H Centers for W550A, R482A, and R544A Point Mutants of the Cdc25B Phosphatase^a

residue	mutant			PMF
	W550A	R482A	R544A	
Trp				
550	n.a.	0.103	0.105	
550e ₁	n.a.	0.045	0.027	
Arg				
376	0.011	0.014	0.018	
427	0.007	0.036	0.042	
447	0.087	0.093	0.011	F
466	n.a.	n.a.	n.a.	
479	n.a.	n.a.	n.a.	
482	0.016	n.a.	0.029	B
485	0.009	0.062	0.013	D
488	0.002	0.022	0.026	
490	0.004	0.033	0.006	D
492	0.004	0.021	0.006	
529	0.016	0.012	0.041	
544	0.167	0.249	n.a.	B
548	0.170	0.083	0.100	
554	0.409	0.019	0.028	F
556	0.082	0.037	0.027	F
557	0.016	0.049	0.015	
562	0.008	0.014	0.070	F
$\bar{x} + \sigma$	0.012	0.030	0.026	

^aThe Nε₁-H system in W550 side-chain is also shown. Unassigned centers are marked as n.a. Mean plus standard deviation ($\bar{x} + \sigma$) were calculated from around 115 CSP values for each point mutant after removal of outliers. Significant CSP are shown in bold. The PMF column codes whether the simulated free energy profile for the respective Arg-W550 cation- π interaction (Figures 3 and S1) shows a bound minimum for letter B, a dissociative profile for letter D, and an almost flat profile for letter F.

Chemical shifts are exquisitely sensitive to the distribution of protein conformations and to their interconversion rates.¹³ In Cdc25B, R554, R556, and R557 are close to W550 in sequence and placed in the same disordered C-terminal segment.²³ If the W550A mutation alters the distribution or dynamics of main-chain configurations, it will lead indirectly to significant CSP for these Arg even if they do not perform important cation- π contacts. Although R447 and R529 are placed in the folded Cdc25B protein core, their segments have considerable flexibility as indicated by NMR order parameters ($S^2 \leq 0.95$ for D448-L453 and $S^2 \leq 0.90$ for D527-M531) and exchange rates ($R_{\text{ex}} = 17 \text{ s}^{-1}$ for E534) previously measured.²³ Allosteric modulation of the flexibility along these segments by the W550A mutation may also lead indirectly to significant CSP.^{43,44}

These indirect effects may also account for the significant CSP observed in protein regions close to mutation sites of R482A and R544A constructions as well as for the CSP in R544 for the W550A mutant, although this Arg residue is in a stable α -helix. Thus, a significant CSP upon mutation of the π -system containing side-chain, here W550, is necessary but not sufficient to prove the importance of a transient cation- π contact.

Mutations of cationic residues may also be explored. The CSP for W550 main-chain N–H and ¹⁵Nε₁-H in the indole

side-chain was significant for both R482A and R544A mutants (Table 1). This may be a relatively easy and general probe of transient cation- π interactions in other proteins, as the Trp ¹⁵Nε₁ signal usually appears isolated and can be readily identified even for unassigned HSQC spectra.¹³

Given that a competition of transient cation- π interactions is present (Figure 4), abolishing one contact by mutation of the involved cationic residue would also perturb the molecular interactions (and the resonance signals) of the remaining Arg performing important contacts with the same π system. An indication of this competition in Cdc25B is the very high and localized CSP found for R544 upon R482A mutation (Figure S2), even though these residues are structurally distant (Figure 1). CSP is also significant for R548 and R556 in the R482A mutant and for R482, R548, and R556 (by only 0.001 CSP units) in the R544A mutant. On the other hand, insignificant perturbations are found for R529 and R554 in the R482A mutant and for R447 and R557 in the R544A mutant, indicating these four Arg do not perform important interactions with W550.

We may conclude that all CSP data (Table 1) and comparisons described here consistently suggest that only R482, R544, R548, and R556 may perform important transient cation- π interactions with W550 out of 15 assigned Arg residues in Cdc25B. Evidence is stronger for R482 and R544 placed in stable helices. Their mutation lead to significant CSP in W550 (Figures S2 and S3), presumably by abolishing a direct (side-chain) contact. The same analysis of the W550 CSP upon R548 and R556 mutations could not be carried out because these residues are placed in the same disordered C-terminus (pale green in Figure 1) and mutations would easily perturb the backbone configurational distribution.

Despite the various structural and dynamical sources of CSP upon point mutations which complicate the identification of transient contacts, another possible shortcoming of our NMR analysis is the non-unique formula used for calculation of CSP.⁴¹ Here we found it essential to iteratively remove outliers⁴⁰ and to scale chemical shift changes by their natural standard deviation obtained from a database.³⁹ This is particularly relevant to be able to compare main-chain to side-chain CSP, as the dispersion of chemical shift for side-chain Trp-¹⁵Nε₁ is much narrower than the dispersion of all main-chain shifts (Table S1). For the free energy simulations, recent force-field parametrizations^{45,46} and modified water models⁴⁷ could improve sampling of the backbone configurational distribution of the disordered C-terminal region in Cdc25B.

4. CONCLUSIONS

Free energy profiles estimated from US with MD simulations and perturbation of chemical shifts obtained from standard HSQC experiments in solution were combined here to identify transient cation- π interactions of solvent-exposed Arg-Trp pairs in proteins. The phosphatase Cdc25B, containing several Arg residues and only one Trp located in a disordered segment, was used as a model system.

Three kinds of free energy profiles were found: bound profiles (R482 and R544) with a clear dissociation barrier of 9–14 kJ/mol, characterizing an important transient cation- π interaction; dissociative profiles (R485 and R490), meaning that no cation- π interaction should be observed; and almost flat profiles (R447, R554, R556 and R562) with weak minima (≤ 6 kJ/mol) in bound distances, characterizing possible but

very short-lived and statistically unimportant cation- π interactions. The same three kinds of profiles should be found generally for transient cation- π interactions between solvent-exposed side-chains in other proteins.

The free energy profiles obtained for eight putative cation- π pairs in Cdc25B are in qualitative agreement with the CSP analysis, confirming that R482 and R544 perform important transient interactions with W550 in solution, even though their side-chains are separated by more than 8 Å in the crystal structure. The only exception is R556 which shows low but significant CSP and an almost flat free energy profile, with a shallow minimum. Although transient contacts of W550 with R447 and R485 were observed in previous MD simulations,²³ results presented here clearly show that these contacts are not important in the Cdc25B conformational ensemble. Thus, the statistical significance of transient contacts in solution should not be relied upon single observations over crystal structures or molecular simulations.

The experimental validation proposed here could be generally applied to other proteins. It is relatively simple to conduct once the protein constructs are purified and the (wild-type) ¹H-¹⁵N HSQC spectra is assigned. However, NMR chemical shifts are sensitive to changes in both molecular contacts and conformational distributions. Disruption of a cation- π interaction upon mutation will directly result in a significant CSP by abolishing this molecular contact. But a significant CSP may also appear due to indirect changes in the residue conformational distribution. Thus, the proposed CSP analysis should be applied carefully and in combination with other statistically representative data.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.9b00859>.

Three tables with CSP scaling factors and side-chain distances to Arg and to Lys residues. Four figures with additional free energy profiles, NMR spectra, and distance trajectories (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: garantes@iq.usp.br.

ORCID

Guilherme M. Arantes: 0000-0001-5356-7703

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Funding from FAPESP (scholarships 2012/00543-1 and 2016/24096-5 to R.S.R.S. and research grants 2016/22365-9, 2018/25952-8, and 2018/08311-9), CNPq (scholarship 141683/2016-3 to A.A.O.R.), CAPES (finance code 001), and computational resources from the SDumont cluster in the National Laboratory for Scientific Computing (LNCC/MCTI) are gratefully acknowledged.

■ REFERENCES

- (1) Ma, J. C.; Dougherty, D. A. The Cation- π Interaction. *Chem. Rev.* **1997**, *97*, 1303–1324.
- (2) Burley, S. K.; Petsko, G. A. Amino-aromatic interactions in proteins. *FEBS Lett.* **1986**, *203* (2), 139–143.

- (3) Dougherty, D. A. Cation- π interactions in chemistry and biology: A new view of benzene, Phe, Tyr and Trp. *Science* **1996**, *271*, 163–168.

- (4) Jing, W.; Demcoe, A. R.; Vogel, H. J. Conformation of a Bactericidal Domain of Puroindoline a: Structure and Mechanism of Action of a 13-Residue Antimicrobial Peptide. *J. Bacteriol.* **2003**, *185*, 4938–4947.

- (5) Qamar, S.; Wang, G.; Randle, S. J.; Ruggeri, F. S.; Varela, J. A.; Lin, J. Q.; Phillips, E. C.; Miyashita, A.; Williams, D.; Strohl, F.; Meadows, W.; Ferry, R.; Dardov, V. J.; Tartaglia, G. G.; Farrer, L. A.; Schierle, G. S. K.; Kaminski, C. F.; Holt, C. E.; Fraser, P. E.; Schmitt-Ulms, G.; Klenerman, D.; Knowles, T.; Vendruscolo, M.; St George-Hyslop, P. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation- π Interactions. *Cell* **2018**, *173*, 720–734.

- (6) Díaz-Moreno, I.; de la Rosa, M. Transient interactions between biomolecules. *Eur. Biophys. J.* **2011**, *40*, 1273–1274.

- (7) Zuckerman, D. M. *Statistical Physics of Biomolecules: An Introduction*, 1st ed.; CRC Press: Boca Raton, FL, 2010.

- (8) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annu. Rev. Biophys.* **2011**, *40*, 41–62.

- (9) Robustelli, P.; Stafford, K. A.; Palmer, A. G. Interpreting Protein Structural Dynamics from NMR Chemical Shifts. *J. Am. Chem. Soc.* **2012**, *134*, 6365–6374.

- (10) Gu, Y.; Li, D.-W.; Bruschweiler, R. NMR Order Parameter Determination from Long Molecular Dynamics Trajectories for Objective Comparison with Experiment. *J. Chem. Theory Comput.* **2014**, *10*, 2599–2607.

- (11) Sharma, V.; Belevich, G.; Gamiz-Hernandez, A. P.; Róg, T.; Vattulainen, I.; Verkhovskaya, M. L.; Wikström, M.; Hummer, G.; Kaila, V. R. I. Redox-induced activation of the proton pump in the respiratory complex I. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 11571–11576.

- (12) Knapp, B.; Ospina, L.; Deane, C. M. Avoiding false positive conclusions in molecular simulation: the importance of replicas. *J. Chem. Theory Comput.* **2018**, *14*, 6127–6138.

- (13) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. *J. Protein NMR Spectroscopy: Principles and Practice*, 2nd ed.; Academic Press: New York, 2007.

- (14) Pellecchia, M.; Wider, G.; Iwai, H.; Wuthrich, K. Arginine side chain assignments in uniformly ¹⁵N-labeled proteins using the novel 2D HE(NE)HGHH experiment. *J. Biomol. NMR* **1997**, *10*, 193–197.

- (15) Juszczak, L. J.; Eisenberg, A. S. The Color of Cation- π Interactions: Subtleties of Amine-Tryptophan Interaction Energetics Allow for Radical-like Visible Absorbance and Fluorescence. *J. Am. Chem. Soc.* **2017**, *139*, 8302–8311.

- (16) Reynolds, R. A.; Yem, A. W.; Wolfe, C. L.; Deibel, M. R.; Chidester, C. G.; Watenpaugh, K. D. Crystal structure of the catalytic subunit of Cdc25B required for G(2)/M phase transition of the cell cycle. *J. Mol. Biol.* **1999**, *293*, 559–568.

- (17) Rudolph, J. Cdc25 Phosphatases: Structure, Specificity, and Mechanism. *Biochemistry* **2007**, *46*, 3595–3604.

- (18) Rudolph, J. Catalytic mechanism of CDC25. *Biochemistry* **2002**, *41*, 14613–14623.

- (19) Arantes, G. M.; Loos, M. Specific Parametrisation of a Hybrid Potential to Simulate Reactions in Phosphatases. *Phys. Chem. Chem. Phys.* **2006**, *8*, 347–353.

- (20) Arantes, G. M. Free energy profiles for catalysis by Dual-specificity Phosphatases. *Biochem. J.* **2006**, *399*, 343–350.

- (21) Arantes, G. M. The Catalytic Acid in the Dephosphorylation of the Cdk2-pTpY/CycA Protein Complex by Cdc25B Phosphatase. *J. Phys. Chem. B* **2008**, *112*, 15244–15247.

- (22) Lund, G.; Cierpicki, T. Solution NMR studies reveal no global flexibility in the catalytic domain of CDC25B. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 2889–2895.

- (23) Sayegh, R. S. R.; Tamaki, F. K.; Marana, S. R.; Salinas, R. K.; Arantes, G. M. Conformational Flexibility of the Complete Catalytic Domain of Cdc25B Phosphatases. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 1567–1575.

- (24) Arantes, G. M. Flexibility and inhibitor binding in Cdc25 phosphatases. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 3017–3032.
- (25) Sayegh, R. S. R. Conformational flexibility of the catalytic domain of Cdc25B phosphatase. *Ph.D. Thesis*, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil, 2016.
- (26) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. State of the Art in Counterpoise Theory. *Chem. Rev.* **1994**, *94*, 1873–1885.
- (27) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2009**, *31*, 671–690.
- (28) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712–725.
- (29) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950–1958.
- (30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, A.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
- (31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (32) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (33) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (34) Roux, B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Comput. Phys. Commun.* **1995**, *91*, 275–282.
- (35) Johnson, R. W. An Introduction to the Bootstrap. *Teaching Statistics* **2001**, *23*, 49–54.
- (36) Souza, V. P.; Ikegami, C. M.; Arantes, G. M.; Marana, S. R. Protein thermal denaturation is modulated by central residues in the protein structure network. *FEBS J.* **2016**, *283*, 1124–1138.
- (37) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6*, 277–293.
- (38) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Struct., Funct., Genet.* **2005**, *59*, 687–696.
- (39) Zhang, H.; Neal, S.; Wishart, D. S. RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **2003**, *25*, 173–195.
- (40) Schumann, F.; Riepl, H.; Maurer, T.; Gronwald, W.; Neidig, K.-P.; Kalbitzer, H. Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J. Biomol. NMR* **2007**, *39*, 275–289.
- (41) Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16.
- (42) Leach, A. R. *Molecular modelling: principles and applications*, 2nd ed.; Prentice Hall: Harlow, 2001.
- (43) Tzeng, S.-R.; Kalodimos, C. G. Protein dynamics and allostery: an NMR view. *Curr. Opin. Struct. Biol.* **2011**, *21*, 62–67.
- (44) Lisi, G. P.; Loria, J. P. Allostery in enzyme catalysis. *Curr. Opin. Struct. Biol.* **2017**, *47*, 123–130.
- (45) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (46) Liu, H.; Song, D.; Zhang, Y.; Yang, S.; Luo, R.; Chen, H.-F. Extensive tests and evaluation of the CHARMM36IDPSFF force field for intrinsically disordered proteins and folded proteins. *Phys. Chem. Chem. Phys.* **2019**, *21*, 21918–21931.
- (47) Boonstra, S.; Onck, P. R.; van der Giessen, E. CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State. *J. Phys. Chem. B* **2016**, *120*, 3692–3698.

Supporting Information

Combining Free Energy Simulations and NMR Chemical-Shift Perturbation to Identify Transient Cation- π Contacts in Proteins

André A. O. Reis, Raphael S. R. Sayegh, Sandro R. Marana and Guilherme M. Arantes*

Department of Biochemistry, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, 05508-900, São Paulo, SP, Brazil

E-mail: garantes@iq.usp.br

Table S1: Standard deviation of ^1H and ^{15}N chemical shift (in ppm) for main chain N-H system of amino-acid type α obtained from the RefDB database.³⁹ Note the ratio between ^1H and ^{15}N scaling factors (j_α/k_α) is similar to 0.15, the standard value used in CSP studies,^{23,41} for all N-H systems except for Trp N ϵ 1-H.

Residue	Scaling factors	
	j_α (^1H)	k_α (^{15}N)
Ala	0.660	4.040
Arg	0.640	3.960
Asp	0.710	4.460
Asn	0.600	4.070
Cys	0.730	4.331
Gly	0.760	3.730
Glu	0.640	3.670
Gln	0.620	3.730
His	0.730	4.090
Iso	0.750	4.460
Leu	0.700	4.131
Lys	0.680	3.851
Met	0.650	3.660
Phe	0.770	4.230
Ser	0.650	3.810
Tyr	0.790	4.550
Thr	0.650	4.941
Val	0.720	4.850
Trp	0.860	4.649
Trp N ϵ 1-H	0.491	1.884

Table S2: Side-chain distance (in nm) to W550 in the crystallographic x-ray structure¹⁶ and minimum distance (in nm) observed along 6 μ s of MD simulation²³ for all Arg residues of the complete Cdc25B C-terminal domain.

Arg	X-ray	Long MD
376	3.71	1.85
427	2.07	1.27
447	1.25	0.37
466	3.81	2.26
479	1.47	0.54
482	1.19	0.29
485	2.10	0.33
488	2.62	0.66
490	2.02	0.31
492	2.84	0.46
529	2.32	2.07
544	0.83	0.52
548	0.97	0.30
554	–	0.30
556	–	0.40
557	–	0.80
562	–	0.36

Table S3: Side-chain distance (in nm) to W550 in the crystallographic x-ray structure,¹⁶ minimum distance (in nm) observed along 6 μ s of MD simulation,²³ and chemical shift perturbation for W550, R482 and R544 point mutants of Cdc25B phosphatase, for all Lys residues of the complete C-terminal domain. Unassigned is marked n.a.

Lys	x-ray	Long MD	W550A	R482A	R544A
384	3.882	1.96	n.a.	n.a.	n.a.
394	1.965	0.67	0.014	0.018	0.012
399	3.168	1.26	0.006	0.006	0.010
413	4.239	2.92	0.032	0.042	0.143
420	3.531	1.94	n.a.	n.a.	n.a.
438	3.661	2.40	n.a.	n.a.	n.a.
455	2.604	0.86	0.010	0.007	0.007
465	4.206	2.89	n.a.	n.a.	n.a.
509	3.016	1.48	0.004	0.033	0.025
513	2.794	1.30	0.009	0.005	0.016
537	2.079	1.14	0.013	0.022	0.082
541	1.333	0.60	0.023	0.105	0.185
546	1.360	0.53	0.181	0.555	0.215

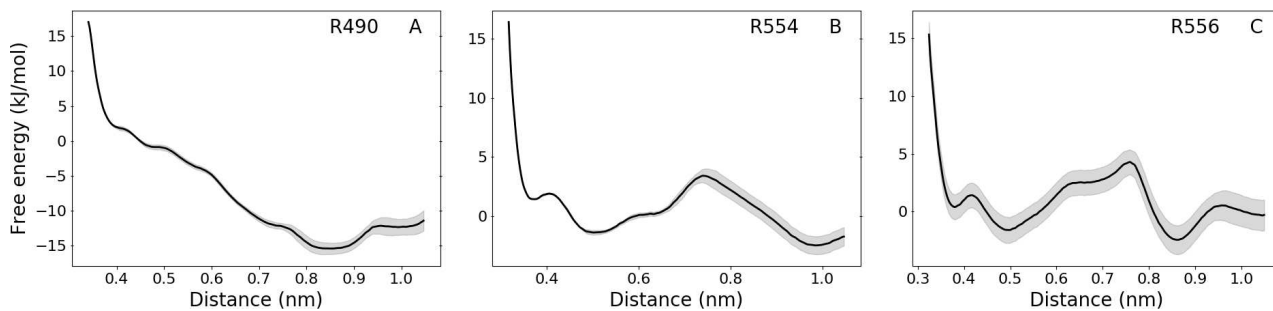


Figure S1: Free energy profiles for the cation- π interaction between W550 side-chain and Arg 490 (panel A), 554 (B) and 556 (C) side-chains. Shadows around lines in all panels show the statistical uncertainty. Profiles were obtained after 100 ns of total simulation time per US window and the first 30ns were discarded from analysis. These simulations were not carried out longer because the dissociative profiles shown and the low conservation of these Arg residues²³ suggested these are not important interactions for the Cdc25B conformational ensemble.

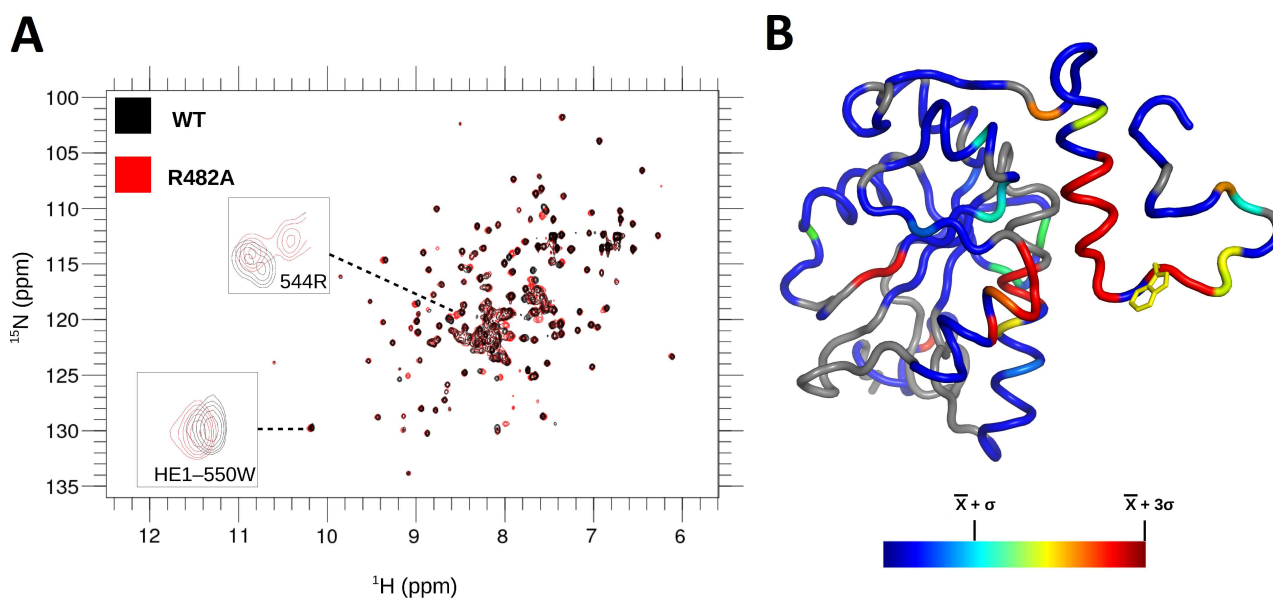


Figure S2: NMR spectra of Cdc25B catalytic domain and chemical shift perturbation upon mutation. (A) ^1H - ^{15}N HSQC for Cdc25B wild-type and R482A point mutant and (B) CSP projected according to colorbar over the same structure shown in Fig. 1. Gray segments were not assigned.

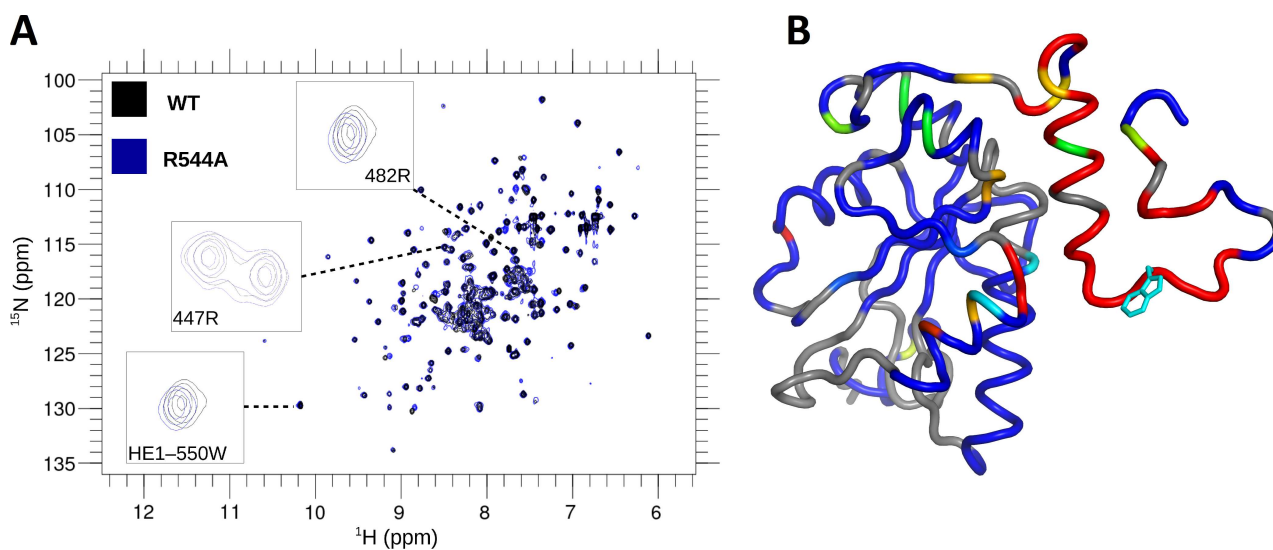


Figure S3: HSQC spectra and CSP as in Fig. S2 but for the R544A Cdc25B point mutant.

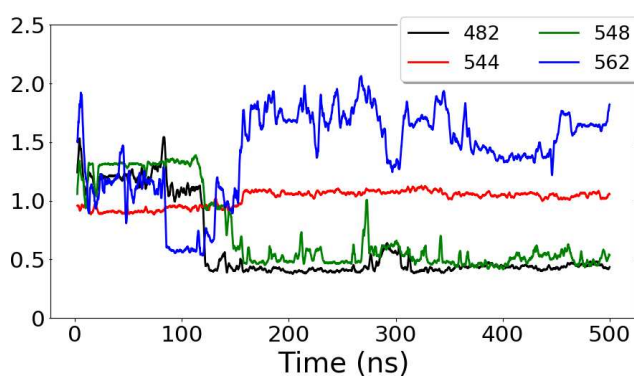


Figure S4: Distance trajectories between the COM of W550 side-chain and the COM of R482 (black), R544 (red), R548 (green) and R562 (blue) side-chains, obtained during R544 US simulations with reference coordinate in 1.0 nm.