LUÍS FERNANDO MERCIER FRANCO

ESTUDO DO EQUILÍBRIO SÓLIDO-LÍQUIDO DE SISTEMAS
CONTENDO AMINOÁCIDOS E PROTEÍNAS

São Paulo
2012

LUÍS FERNANDO MERCIER FRANCO

ESTUDO DO EQUILÍBRIO SÓLIDO-LÍQUIDO DE SISTEMAS
CONTENDO AMINOÁCIDOS E PROTEÍNAS

Dissertação apresentada à Escola
Politécnica da Universidade de São
Paulo para obtenção do título de
Mestre em Engenharia

São Paulo
2012

LUÍS FERNANDO MERCIER FRANCO

ESTUDO DO EQUILÍBRIO SÓLIDO-LÍQUIDO DE SISTEMAS
CONTENDO AMINOÁCIDOS E PROTEÍNAS

Dissertação apresentada à Escola
Politécnica da Universidade de São
Paulo para obtenção do título de
Mestre em Engenharia

Área de Concentração:
Engenharia Química

Orientador: Professor Livre-Docente
Pedro de Alcântara Pessoa Filho

São Paulo
2012

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 12 de abril de 2012.

Assinatura do autor

Assinatura do orientador

FICHA CATALOGRÁFICA

Franco, Luís Fernando Mercier

Estudo do equilíbrio sólido-líquido de sistemas contendo aminoácidos e proteínas / L.F.M. Franco. -- ed.rev. -- São Paulo, 2012.

126 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Química.

1. Termodinâmica química 2. Equilíbrio sólido-líquido 3. Proteínas 4. Aminoácidos I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Química II. t.

Aos meus avós Zenaide e Antônio Sérgio,
Marylande e Paulo Fernando (*in memoriam*);
a minha mãe, Priscila; ao papai,
Gilberto (*in memoriam*); e ao mano
Henrique dedico este trabalho.

AGRADECIMENTOS

É, além de mandatório que eu o faça, mandatório pela minha consciência mais do que pelo mero preenchimento de algumas linhas, um imenso prazer agradecer a quem de algum modo contribuiu comigo durante a execução deste trabalho.

O primeiro a quem devo e desejo agradecer é o meu orientador Professor Livre-Docente Pedro de Alcântara Pessoa Filho. É ele a quem devo tudo o que sei sobre Termodinâmica, desde os bancos acadêmicos do terceiro ano de engenharia quando aprendi os primeiros conceitos de equilíbrio de fases; época em que ainda eu não sabia que isso seria objeto de investigação do meu mestrado. É com imensa gratidão que agora me dirijo a ele dizendo-lhe que sem sua paciência e generosíssima disponibilidade nada do que este texto contém lograria tal êxito que agora surge a necessidade, mais do que o dever, de apresentar.

Também agradeço aos amigos e colegas do Grupo de Engenharia de Bioprocessos (GEnBio), antigo Laboratório de Engenharia Bioquímica (LEB), do Departamento de Engenharia Química da Escola Politécnica da Universidade de São Paulo cujas discussões incessantes, sobre os mais diversos conteúdos, estão dispersas ainda que de modo imperceptível nas entrelinhas deste trabalho. Em particular, agradeço a Kelly Cristina Nascimento Alves Cruvinel quem, além de me haver fornecido alguns dados experimentais que enriquecem a discussão dos modelos propostos neste trabalho, sempre me apoiou desde os tempos em que eu era aluno de iniciação científica.

Por fim, agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e ao Departamento de Engenharia Química da Escola Politécnica da Universidade de São Paulo, pois me concedendo uma bolsa para cursar meu mestrado permitiram que eu o fizesse dignamente.

“Deve-se escrever da mesma maneira como as lavadeiras lá de Alagoas fazem o seu ofício. Elas começam com uma primeira lavada, molham a roupa na beira da lagoa ou do riacho, torcem o pano, molham-no novamente, voltam a torcer. Colocam o anil, ensaboam e torcem uma, duas vezes. Depois enxáguam, dão mais uma molhada, agora jogando a água com a mão. Batem o pano na laje ou na pedra limpa, e dão mais uma torcida e mais outra, torcem até não pingar do pano uma só gota. Somente depois de feito tudo isso é que elas dependuram a roupa lavada na corda do varal, para secar. Pois quem se mete a escrever devia fazer a mesma coisa. A palavra não foi feita para enfeitar, brilhar como ouro falso; a palavra foi feita para dizer.”

(Graciliano Ramos)

RESUMO

A modelagem termodinâmica do equilíbrio sólido-líquido de sistemas que contêm aminoácidos e proteínas tem sido cada vez mais necessária para o desenvolvimento de equações que permitam um projeto mais racional e eficiente das operações unitárias encontradas nos processos de purificação destas moléculas, tais como a precipitação e a cristalização. Neste trabalho, é apresentada uma relação unívoca entre a solubilidade de proteínas e o pH; sendo necessários como parâmetros apenas informações sobre a estrutura primária e os valores das constantes de ionização de cada resíduo na cadeia protéica, considerando os diferentes estados de ionização que a proteína assume em solução. Esta relação foi aplicada a sistemas de insulina suína a três diferentes temperaturas, a sistemas de insulinas mutantes e a sistemas de β -lactoglobulina a diferentes concentrações de cloreto de sódio, resultando em uma bem sucedida correlação dos dados experimentais de solubilidade destes sistemas. Estudou-se também a não-idealidade de sistemas contendo aminoácidos, partindo de considerações sobre diferentes estados de ionização da molécula assim como feito para proteínas, mas adicionando um termo de não-idealidade, neste caso, um termo advindo do modelo de Pitzer para soluções eletrolíticas. Os sistemas, cujas curvas de solubilidade em função do pH foram estudadas, são: DL-Alanina, DL-Metionina, L-Isoleucina, L-Leucina, DL-Fenilalanina, também glicina e seus oligopeptídeos e moléculas com grupos β -lactâmicos tais como a ampicilina, que é um antibiótico, e o ácido 6-aminopenicilânico, que é um precursor na rota de produção da ampicilina. Uma interpretação através de termodinâmica estatística para o parâmetro de interação de Pitzer permitiu uma análise mais profunda dos resultados. Relações entre o segundo coeficiente virial osmótico, a solubilidade de proteínas e a concentração de agente precipitante foram desenvolvidas e aplicadas a sistemas de lisozima, ovalbumina e imunoglobulina humana. Os resultados desta aplicação mostraram quão profícua é a abordagem empregada.

Palavras-chave: Termodinâmica Química. Equilíbrio Sólido-Líquido. Proteínas. Aminoácidos.

ABSTRACT

The thermodynamic modeling of solid-liquid equilibrium of systems containing amino acids and proteins has become increasingly necessary for a more rational and efficient design of unit operations found in the downstream processing of these molecules, such as precipitation and crystallization. In this work, an unequivocal relationship between the protein solubility and the pH is presented. The resulting equation accounts for the different ionization states that protein molecules may assume in solution, and is written as a function of the primary structure and values of the ionization constant of each residue in the protein chain. The model was applied to the solid-liquid equilibrium of porcine insulin solutions at three different temperatures, solutions of mutant insulin molecules and solutions of β -lactoglobulin at different sodium chloride concentrations. A very successful correlation of the experimental solubility data was obtained in all cases. The solubility of amino acids and β -lactam compounds was also studied through a similar approach, but in this case the non-ideality of the solutions was accounted for by introducing Pitzer's model for electrolyte solutions. The compounds whose solubility curves as a function of the pH were modeled include the amino acids DL-Alanine, DL-Methionine, L-Isoleucine, L-Leucine, DL-Phenylalanine, Glycine and its oligopeptides, and β -lactam compounds such as ampicillin, which is an antibiotic, and 6-aminopenicillanic acid, which is a precursor in the ampicillin production route. The interpretation of the binary interaction parameter values through a statistical thermodynamic approach allowed a deeper analysis of the results. Finally, equations relating the osmotic second virial coefficient, the protein solubility and the concentration of precipitant agents such as salts were developed. The resulting equations were successfully applied to solutions containing lysozyme, ovalbumin and immunoglobulin, which shows that the considered approach is promising.

Keywords: Chemical Thermodynamics. Solid-Liquid Equilibrium. Proteins. Amino acids.

LISTA DE FIGURAS

Figura 3.1 - Estrutura primária da insulina suína: seqüência de aminoácidos e pontes de dissulfeto. Grupos ionizáveis em negrito e distribuição de carga da configuração neutra correspondente ao estado de ionização preponderante.....	47
Figura 3.2 - Fração de moléculas eletricamente neutras de insulina em função do pH. Dados de pK_A de Kaarsholm; Havelung e Hougaard (1990). Linha contínua, fração total. Linha pontilhada, fração preponderante.....	47
Figura 3.3 - Solubilidade da insulina suína em função do pH a 278,2 K em 10,0 mM NaHCO_3 . Dados experimentais ($\square$), Tashima et al., (2009). Modelo proposto com valores de pK_A de Kaarsholm; Havelung e Hougaard (1990), linha contínua; com valores de pK_A de Stryer (1995), linha pontilhada. Modelo de Tashima et al. (2009), linha semi-contínua.....	48
Figura 3.4 - Solubilidade da insulina suína em função do pH a 288,2 K em 10,0 mM NaHCO_3 . Dados experimentais ($\square$), Tashima et al., (2009). Modelo proposto com valores de pK_A de Kaarsholm; Havelung e Hougaard (1990), linha contínua; com valores de pK_A de Stryer (1995), linha pontilhada. Modelo de Tashima et al. (2009), linha semi-contínua.	49
Figura 3.5 - Solubilidade da insulina suína em função do pH a 298,2 K em 10,0 mM NaHCO_3 . Dados experimentais ($\square$), Tashima et al., (2009). Modelo proposto com valores de pK_A de Kaarsholm; Havelung e Hougaard (1990), linha contínua; com valores de pK_A de Stryer (1995), linha pontilhada. Modelo de Tashima et al. (2009), linha semi-contínua.....	49
Figura 3.6 - Curva de solubilidade da insulina mutante B13-Glu→Gln em função do pH em 0,1 M KCl a 23 °C. Dados experimentais ($\circ$), Kaarsholm; Havelung e Hougaard (1990). Equação (3 - 32), linha contínua.....	52
Figura 3.7 - Curva de solubilidade da insulina mutante B9-Ser→Asp em função do pH em 0,1 M KCl a 23 °C. Dados experimentais ($\circ$), Kaarsholm; Havelung e Hougaard (1990). Equação (3 - 32), linha contínua.....	53
Figura 3.8 - Curva de solubilidade da insulina mutante B25-Phe→His em função do pH em 0,1 M KCl a 23 °C. Dados experimentais ($\circ$), Kaarsholm; Havelung e Hougaard (1990). Equação (3 - 32), linha contínua.....	53
Figura 3.9 - Solubilidade da β -lactoglobulina bovina em solução aquosa contendo	

NaCl a 298,2 K. Dados experimentais (Grönwall, 1941): 0,001 mol L ⁻¹ NaCl (♦); 0,005 mol L ⁻¹ NaCl (○); 0,01 mol L ⁻¹ NaCl (▲); 0,02 mol L ⁻¹ NaCl (□). Modelagem: linha contínua, 0,001 mol L ⁻¹ ; linha pontilhada, 0,005 mol L ⁻¹ ; linha semi-contínua, 0,01 mol L ⁻¹ ; linha com travessões, 0,02 mol L ⁻¹	56
Figura 4.1 - Valores de solubilidade da DL-Alanina a 298,15 K em solução aquosa. Dados experimentais (○), Tseng et al. (2009); solução com NaOH ou HCl (●), Pradhan e Vera (1998); solução com KOH ou HNO ₃ (□), Pradhan e Vera (1998). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).	70
Figura 4.2 - Valores de solubilidade de DL-Metionina a 303,0 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (○), Fuchs et al. (2006). Modelagem: solução ideal, linha pontilhada; modelo proposto, linha contínua.....	70
Figura 4.3 - Valores de solubilidade da L-Isoleucina a 298,15 K em solução aquosa (com HCl ou NaOH). Dados experimentais (○), Tseng et al. (2009). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).	71
Figura 4.4 - Valores de solubilidade da L-Leucina a 298,15 K em solução aquosa (com HCl ou NaOH). Dados experimentais (○), Tseng et al. (2009). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).....	71
Figura 4.5 - Valores de solubilidade da DL-Fenilalanina a 298,15 K em solução aquosa (com HCl ou NaOH). Dados experimentais (○), Tseng et al. (2009). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).....	72
Figura 4.6 - Valores de solubilidade da L-Serina 298,15 K em solução aquosa (com HCl ou NaOH). Dados experimentais (○), Tseng et al. (2009). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).....	72
Figura 4.7 - Valores de solubilidade da tirosina 298,15 K em solução aquosa. Dados experimentais (○), Hitchcock (1924). Modelo considerando o grupo <i>R</i> ionizável (linha contínua). Modelo considerando o grupo <i>R</i> não-ionizável (linha pontilhada).....	75
Figura 4.8 - Curva de solubilidade em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais de glicina (○), de diglicina (■) e de triglicina (Δ), Lu et al. (2006). Modelagem - Equação (4 - 23) - com valores de λ da Tabela 4.4: glicina, linha contínua; diglicina, linha semi-contínua; triglicina, linha pontilhada.....	77
Figura 4.9 - Curva de solubilidade em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais de tetraglicina (○), de pentaglicina (■) e de hexaglicina (Δ), Lu et al. (2006). Modelagem - Equação (4 - 23) - com valores de λ da Tabela 4.4: tetraglicina, linha contínua; pentaglicina, linha semi-contínua;	

hexaglicina, linha pontilhada.....	77
Figura 4.10 - Estrutura da ampicilina.....	79
Figura 4.11 - Estrutura do ácido 6-aminopenicilânico.....	79
Figura 4.12 - Valores de solubilidade da ampicilina 283,06 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).....	80
Figura 4.13 - Valores de solubilidade da ampicilina 288,01 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).....	81
Figura 4.14 - Valores de solubilidade da ampicilina 292,95 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).....	81
Figura 4.15 - Valores de solubilidade da ampicilina 298,03 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).....	82
Figura 4.16 - Valores de solubilidade do ácido 6-aminopenicilânico 283,06 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).....	82
Figura 4.17 - Valores de solubilidade do ácido 6-aminopenicilânico 288,01 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).....	83
Figura 4.18 - Valores de solubilidade do ácido 6-aminopenicilânico 292,95 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).....	83
Figura 4.19 - Valores de solubilidade do ácido 6-aminopenicilânico 298,03 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).....	84
Figura 4.20 - Valores de solubilidade da ampicilina 298,15 K em solução aquosa em função do pH. Dados experimentais a diferentes frações mássicas de metanol: (o) 0,0 %, (■) 15,0 %, (Δ) 30,0 % e (♦) 50,0 %. Santana; Mattedi e Giordano (2010).	

Modelo proposto: 0,0 % (linha contínua), 15,0 % (linha com travessões), 30,0 % (linha semi-contínua) e 50,0 % (linha pontilhada).....	85
Figura 4.21 - Valores de solubilidade do ácido 6-aminopenicilânico 298,15 K em solução aquosa em função do pH. Dados experimentais a diferentes frações mássicas de metanol: (o) 0,0 %, (■) 15,0 %, (Δ) 30,0 % e (♦) 50,0 %. Santana; Mattedi e Giordano (2010). Modelo proposto: 0,0 % (linha contínua), 15,0 % (linha com travessões), 30,0 % (linha semi-contínua) e 50,0 % (linha pontilhada).....	86
Figura 5.1 - Solubilidade da lisozima da clara de ovo de galinha em função da molalidade de cloreto de sódio. Dados experimentais (○), Watanabe (2007). Equação de Cohn (linha contínua).....	95
Figura 5.2 - Valores do segundo coeficiente virial osmótico em função da molalidade de cloreto de sódio em pH = 4,5. Dados experimentais obtidos em Curtis et al. (2002). Lisozima nativa, modelo (linha contínua), dados experimentais (□). Lisozima D101F, modelo (linha pontilhada), dados experimentais (●).....	96
Figura 5.3 - Valores da solubilidade da lisozima em função do segundo coeficiente virial osmótico. Dados experimentais obtidos em Guo et al. (1999). Lisozima a 25 °C e pH = 4,2 com variações na concentração de cloreto de sódio (○). Lisozima a 18 °C e pH = 4,5 com variações na concentração de cloreto de amônio (□). Lisozima a 23 °C e pH = 7,8 com variações na concentração de brometo de magnésio (♦). Modelo de Haas; Drenth e Wilson (1999) com $z = 3$ e $A = 0.001$ (linha pontilhada). Modelo de Guo et al. (1999) com $\Delta\mu = 250 \text{ kJ mol}^{-1}$ (linha semi-contínua). Modelo de Ruppert; Sandler e Lenhoff (2001), (linha com travessões). Equação (5 – 23) (linha contínua).	97
Figura 5.4 - Solubilidade da ovalbumina em soluções de sulfato de amônio. Dados experimentais (○), Judge; Johns e White (1996). Equação de Cohn (linha contínua).	98
Figura 5.5 - Segundo coeficiente virial osmótico da ovalbumina em soluções de sulfato de amônio a pH = 4,0. Dados experimentais (○), Mehta; White e Litster (2011). Equação (5 – 21) com $K_s = 2,45 \text{ kg mol}^{-1}$, (linha contínua).....	99
Figura 5.6 - Segundo coeficiente virial osmótico da imunoglobulina humana (IgG) em soluções de cloreto de sódio. Dados experimentais (○), estes dados ainda não foram publicados, mas foram cedidos por Kelly Cristina do Nascimento Alves Cruvinel (comunicação pessoal). Equação (5 – 21), (linha contínua).....	100

LISTA DE TABELAS

Tabela 3.1 - Análise da função θ_{jk}	45
Tabela 3.2 - Valores de pK_A dos grupos ionizáveis da insulina humana.....	46
Tabela 3.3 - Valores de CQ [Equação (3 – 40)] e CQG [Equação (3 – 41)] entre os modelos aplicados e os dados experimentais de solubilidade de insulina suína a diferentes temperaturas.....	51
Tabela 3.4 - Valores de pK_A e de número de resíduo que se ionizam como cátions (n+) para insulinas mutantes. Dados de Kaarsholm; Havelung e Hougaard (1990)..	52
Tabela 3.5 - Valores de pK_A dos grupos ionizáveis da β -lactoglobulina bovina.....	55
Tabela 4.1 - Quadro-resumo de trabalhos sobre a não-idealidade de soluções de aminoácidos.....	60
Tabela 4.2 - Valores do parâmetro de interação λ e do CQ [Equação (3 – 40)] para diversos conjuntos de dados experimentais de solubilidade de aminoácidos em função do pH.....	69
Tabela 4.3 - Valores de CQ [Equação (3 – 40)] em porcentagem para os diferentes modelos aplicados aos conjuntos de dados de solubilidade de DL-Alanina, L-Isoleucina, L-Leucina, DL-Fenilalanina e L-Serina a 298,15 K. Dados experimentais de Tseng et al. (2009).	73
Tabela 4.4 - Valores do parâmetro de interação λ e do CQ (Equação 3 - 40) para diversos conjuntos de dados experimentais de solubilidade de glicina e seus oligopeptídeos em função do pH. Dados experimentais de Lu et al. (2006).....	76
Tabela 4.5 - Valores dos parâmetros σ , ε e CQG para os sistemas contendo ampicilina e ácido 6-aminopenicilânico. Dados experimentais de Santana et al. (2010).....	80
Tabela 5.1 - Comparação dos valores de CQ definido pela Equação (3 – 40) para o sistema contendo lisozima.....	97
Tabela A.1 - Valores de pK_A para aminoácidos, peptídeos e moléculas com grupos β -lactâmicos estudados no capítulo quatro.....	126

LISTA DE SÍMBOLOS E SIGLAS

Letras latinas

a	parâmetro universal do Modelo de Pitzer
a_{H^+}	atividade do íon H^+
a_i	atividade do componente i
a_{P0}	atividade da molécula de proteína eletricamente neutra
A	área superficial do soluto parâmetro do modelo de Debye-Hückel
b	parâmetro universal do modelo de Pitzer
B_{ij}	segundo coeficiente virial osmótico para interação entre os componentes i e j
B	parâmetro do modelo de Debye-Hückel
c_k	carga total de uma molécula de proteína em um estado de ionização k
c	carga média das moléculas de proteína
C	parâmetro do modelo de Debye-Hückel
$C1$	parâmetro ajustável para cálculo do segundo coeficiente virial do modelo de Pitzer
$C2$	parâmetro ajustável para cálculo do segundo coeficiente virial do modelo de Pitzer
d_s	densidade do solvente
D	parâmetro do modelo de Kirkwood
e	carga elétrica
f_i	fugacidade do componente i
f_i^0	fugacidade do componente i no estado de referência
g_k	restrição de igualdade devido às leis de conservação
g^{EX}	energia livre de Gibbs excedente
g_{ij}	parâmetro de energia característico da interação entre os componentes i e j
$g_{ij}(.)$	função de distribuição radial
G_{ij}	parâmetro do modelo NRTL
I	força iônica
k_B	constante de Boltzmann
K_s	constante de <i>salting-out</i>
K_k	constante do equilíbrio de ionização entre uma molécula de proteína neutra e uma molécula da mesma proteína em um estado de ionização k

$L(.)$	operador lagrangeano
m	número de componentes no sistema
m_i	molalidade do componente i
m_i^0	molalidade do componente i no estado de referência
m_0	molalidade de moléculas eletricamente neutras de proteínas
m_{sal}	molalidade do sal
M	cátion
	número de conjuntos de dados experimentais
M_s	massa molecular do solvente
n	número total de resíduos ionizáveis de uma dada proteína
n_A	número de resíduos carregados negativamente
n_B	número de resíduos negativos não carregados
n_C	número de resíduos carregados positivamente
n_D	número de resíduos positivos não carregados
n_+	número de resíduos positivos em uma dada proteína
N	quantidade de matéria do sistema
	número de dados experimentais
N_A	número de Avogadro
N_i	quantidade de matéria total do componente i no sistema
$N_i^{(\alpha)}$	quantidade de matéria do componente i na fase α
p	pressão
P_0	proteína eletricamente neutra
P_k^{+c}	proteína em um estado de ionização k com carga total $+c$
r	coordenada radial
R	constante real dos gases
S	entropia
	solubilidade
S_A	solubilidade de um dado aminoácido
T	temperatura
u_{ij}	potencial intermolecular entre os componentes i e j
U	energia interna do sistema
$\mathbf{U}$	energia interna total do sistema
$U^{(\alpha)}$	energia interna da fase α
V	volume do sistema
$\mathbf{V}$	volume total do sistema
$V^{(\alpha)}$	volume da fase α

w_s	massa do solvente
W_{ij}	potencial da força média entre os componentes i e j
x_i	fração molar do componente i em fase condensada
X	ânion
y_i	fração molar do componente i em fase gasosa
z_i	carga elétrica do componente i
z_+	carga elétrica do cátion M
z_-	carga elétrica do ânion X

Letras gregas

α_j	fração ionizada do resíduo j
α_{ij}	parâmetro de não-aleatoriedade na mistura
α_{+1}	fração de grupos amina ionizados em sistemas contendo aminoácidos
α_{-1}	fração de grupos ácido carboxílico ionizados em sistemas contendo aminoácidos
β	parâmetro da equação de Cohn
γ	tensão superficial do solvente
γ^0	tensão superficial da água pura
γ_i	coeficiente de atividade do componente i
γ_{AO}	coeficiente de atividade de um aminoácido eletricamente neutro
$\gamma_{\pm}$	coeficiente de atividade iônica médio
ε	parâmetro de interação atrativa de van der Waals
ε_0	permissividade do vácuo
ε_r	constante dielétrica
ε_{jk}	carga do resíduo j em uma molécula de proteína em um estado de ionização k
$\eta_j(.)$	função auxiliar que depende apenas do resíduo j
$\theta_{jk}(.)$	função auxiliar que depende tanto do resíduo j quanto do estado de ionização k
κ	inverso do comprimento de Debye
κ^e	termo de correção no cálculo da variação do potencial químico devida à energia envolvida na formação de uma cavidade para alocar a molécula na solução

λ_k	multiplicador de Lagrange relativo à restrição de igualdade g_k
λ_{ij}	segundo coeficiente virial do modelo de Pitzer para as interações entre os componentes i e j
λ_{-1}	segundo coeficiente virial do modelo de Pitzer entre as formas aniônica e zwitteriônica de um dado aminoácido
λ_{+1}	segundo coeficiente virial do modelo de Pitzer entre as formas catiônica e zwitteriônica de um dado aminoácido
λ_0	segundo coeficiente virial do modelo de Pitzer entre a forma zwitteriônica e a forma neutra não-zwitteriônica de um dado aminoácido
λ_{zwitt}	segundo coeficiente virial do modelo de Pitzer entre duas moléculas zwitteriônicas de um dado aminoácido
Λ	coeficiente de <i>salting-in</i>
Λ_{ijk}	terceiro coeficiente virial do modelo de Pitzer para as interações entre os componentes i, j e k
μ_d	momento de dipolo elétrico
μ_i	potencial químico do componente i
μ_i^L	potencial químico do componente i na fase líquida
μ_i^S	potencial químico do componente i na fase sólida
μ_i^0	potencial químico do componente i no estado de referência
$\Delta\mu^*$	potencial químico de transferência de uma molécula de proteína do estado de um gás hipotético para o interior de uma solução
$\Delta\mu_{cav}$	variação do potencial químico devida à energia envolvida na formação de uma cavidade para alocar a molécula na solução
$\Delta\mu_{elet}$	contribuição eletrostática à variação do potencial químico
$\Delta\mu_{vdW}$	variação sofrida pelo potencial químico devido as interações atrativas de van der Waals entre o solvente e o soluto.
ν_j	estado carregado do resíduo j
ν_+	coeficiente estequiométrico do cátion M
ν_-	coeficiente estequiométrico do ânion X
π	número de fases no sistema
Π	pressão osmótica
ρ_i	concentração do componente i
σ	raio de van der Waals médio
	incremento molal da tensão superficial

τ_{ij}	parâmetro do modelo NRTL
ϕ_{AO}	fração de moléculas de aminoácido eletricamente neutras
ϕ_k	fração de moléculas de proteína em um estado de ionização k coeficiente de fugacidade osmótico
ϕ_{zwitt}	fração de moléculas de aminoácido na forma zwitteriônica
ϕ_0	fração de moléculas de proteínas eletricamente neutras fração de moléculas de aminoácido eletricamente neutras na forma não-zwitteriônica
ϕ_{-1}	fração de moléculas de aminoácido na forma aniônica
ϕ_{+1}	fração de moléculas de aminoácido na forma catiônica
ψ	número de estados de ionização possíveis
ψ_j	potencial elétrico do componente j
Ω	razão entre o coeficiente de <i>salting-out</i> e o incremento molal da tensão superficial

Siglas

CQ	critério quantitativo
CQG	critério quantitativo global
NRTL	do inglês: <i>Non-Random Two Liquids</i>
pH	co-logaritmo decimal da atividade do íon H ⁺
pI	ponto isoelétrico
pK_A	co-logaritmo decimal da constante de equilíbrio de ionização

SUMÁRIO

1	Introdução.....	22
2	Elementos fundamentais da termodinâmica de equilíbrio de fases.....	25
2.1	A equação fundamental em sua representação entrópica.....	25
2.2	A equação de Gibbs-Duhem.....	26
2.3	Os critérios de equilíbrio de fases.....	27
2.4	Fugacidade, atividade e convenções.....	29
2.5	Modelos termodinâmicos de energia livre de Gibbs excedente.....	31
2.6	Modelos termodinâmicos para soluções eletrolíticas.....	32
3	Efeito do pH na solubilidade de proteínas.....	35
3.1	Introdução.....	35
3.2	Desenvolvimento teórico.....	36
3.2.1	Cálculo da fração de proteína eletricamente neutra.....	39
3.2.2	Extensão à teoria de Linderstrøm-Lang e Grönwall.....	41
3.2.3	Equivalência entre as duas abordagens.....	44
3.3	Resultados e Discussões.....	45
3.3.1	Solubilidade da insulina em função do pH e da temperatura.....	45
3.3.2	Solubilidade da β -lactoglobulina em função do pH e da força iônica.....	54
4	Não-idealidade de soluções de aminoácidos e compostos β -lactâmicos.....	57
4.1	Introdução.....	57
4.2	Desenvolvimento teórico.....	61
4.2.1	Cálculo das frações eletricamente neutras e carregadas	61
4.2.2	Modelo de Pitzer	63
4.2.3	Uma abordagem por termodinâmica estatística da não-idealidade.....	66
4.3	Resultados e Discussões.....	69
4.3.1	Solubilidade de aminoácidos em função do pH.....	69
4.3.2	Solubilidade da glicina e de seus oligopeptídeos em função do pH.....	76
4.3.3	Solubilidade de compostos β -lactâmicos em função do pH, da temperatura e do solvente.....	78
5	Segundo coeficiente virial osmótico de proteínas e suas relações.....	87
5.1	Introdução.....	87
5.2	Desenvolvimento teórico.....	89
5.2.1	Potencial químico de referência.....	90
5.2.2	Equação de Cohn.....	92
5.2.3	Modelo proposto.....	93

5.3 Resultados e Discussões.....	94
5.3.1 Resultados para a lisozima.....	94
5.3.2 Resultados para a ovalbumina.....	98
5.3.3 Resultados para a imunoglobulina humana.....	100
6 Conclusões.....	101
7 Referências bibliográficas.....	104
Apêndice A – Integração do perfil de carga em função do pH.....	110
Apêndice B – Código-fonte do programa de cálculo de solubilidade de proteínas. .	112
Apêndice C – Expressão para o coeficiente de atividade de um soluto no <i>ensemble</i> de McMillan-Mayer.....	117
Apêndice D – Código-fonte do programa de ajuste de parâmetros para o cálculo da solubilidade de aminoácidos e compostos β -lactâmicos.....	121
Anexo A – Valores de pK_A usados nos cálculos do quarto capítulo.....	126

1 Introdução

*“et si in tanta scriptorum turba mea fama in obscuro sit,
nobilitate ac magnitudine eorum me qui nomini officient
meo consoler”*

(Titus Livius, *in* História de Roma)

Apesar de muitos acreditarem ser a Termodinâmica uma área concluída do universo científico, prova-se constantemente que, justamente por estar a Termodinâmica consolidada do ponto de vista formal da lógica matemática e da interpretação física, ela pode contribuir de modo único para a descrição de fenômenos descritos apenas recentemente. A revolução biotecnológica, que marca desde a segunda metade do século XX até os dias correntes tanto os processos industriais quanto a pesquisa acadêmica, traz consigo desafios importantes para uma maior compreensão dos sistemas ditos biológicos, ou seja, sistemas que têm como alguns de seus componentes moléculas de algum modo originadas de seres vivos.

A importância mais óbvia da Termodinâmica reside no fato de ela poder fornecer uma imensa gama de informações sobre o sistema com a necessidade de poucos parâmetros. Este fato é tanto mais relevante quanto mais complexo for o sistema estudado. Na verdade, a Termodinâmica em sua forma mais primitiva prescinde inclusive da hipótese atômica, o que a garante tal generalidade que ainda em uma natureza na qual não houvesse átomos suas leis seriam válidas e poderiam ser aplicadas irrestritamente. Quanto a isso, notam-se na introdução de seu livro sobre Termodinâmica Estatística as observações de Terrell L. Hill:

[...] Although thermodynamics itself does not provide a molecular picture of nature, this is not always a disadvantage. Thus there are many complicated systems for which a molecular theory is not possible; but regardless of complications on the molecular level, thermodynamics can still be applied to such systems with confidence and exactness.[...]

(HILL, 1986, p. 1)

Aí reside a explicação para a comunhão entre a Termodinâmica e todas as outras ciências da natureza, sem considerar os aspectos filosóficos muitas vezes

apropriados das leis da Termodinâmica.

Não há mais espaço (se é que antes houve algum) para a pesquisa científica que se limite a uma única área do conhecimento. Dessarte, os profissionais de Termodinâmica têm na medida do possível servido a comunidade científica como um todo, com interpretações, modelos e mesmo dados experimentais de propriedades que auxiliam as pesquisas dedicadas às mais diversas áreas.

No caso de sistemas que contêm proteínas, há uma grande e crescente demanda por conhecimento devido às inúmeras aplicações, seja no setor farmacêutico, seja na indústria alimentícia. Aqui, estudar-se-ão alguns destes sistemas. A abordagem pretendida corresponde à exploração dos limites da termodinâmica clássica na descrição destes sistemas.

Moléculas de proteína apresentam, entre outras, uma característica que torna a sua modelagem termodinâmica um pouco mais complicada: a presença de íons na cadeia. Pode-se dizer que a teoria de soluções eletrolíticas nasce, ou pelo menos, ganha notoriedade com o trabalho de Petrus Debye e Erich Hückel, que no início da década de 1920, propuseram um método de cálculo da não-idealidade devida à presença de eletrólitos em solução através da solução linearizada e unidirecional da equação de Poisson-Boltzmann. Mesmo assim, as soluções contendo eletrólitos ainda hoje são alvo de intensa pesquisa. Até porque a própria teoria de líquidos ainda é um capítulo no qual há hoje muito esforço empregado. Quando se pensa em proteínas, que na verdade são polieletrólitos muito menos simples que polieletrólitos de polímeros sintéticos, a modelagem se apresenta repleta de simplificações, sejam de ordem geométrica (tal como considerar a proteína uma esfera), sejam de ordem de distribuição de cargas, dotando a proteína de uma carga líquida sem considerar os diferentes estados de ionização.

As principais intenções aqui pretendidas estão centradas em uma busca por uma descrição termodinâmica mais fiel na sua representação dos fenômenos estudados.

O segundo capítulo será dedicado a uma brevíssima revisão dos conceitos fundamentais de termodinâmica de equilíbrio de fases de modo a facilitar ao leitor, desacostumado à nomenclatura termodinâmica, a compressão dos capítulos seguintes. Diferentemente do segundo capítulo, os que o seguirão, com pouquíssimas exceções que serão postas em relevo a medida que aparecerem, integram os desenvolvimentos teóricos inéditos aqui propostos.

No terceiro capítulo será apresentado o desenvolvimento de uma relação teórica

para a predição da curva de solubilidade de proteínas em função do pH considerando que a fase sólida é única e só contém moléculas eletricamente neutras de proteínas pelo princípio de eletroneutralidade da fase em equilíbrio termodinâmico. Resultados para a insulina suína a diferentes temperaturas, algumas insulinas mutantes e para a β -lactoglobulina a diferentes valores de força iônica serão apresentados e discutidos também neste capítulo. Tanto a modelagem quanto os principais resultados deste capítulo foram publicados em um artigo recente (FRANCO; PESSÔA FILHO, 2011).

No quarto capítulo, a equação deduzida no capítulo terceiro será modificada de modo a ser aplicada a sistemas contendo aminoácidos e moléculas com grupos β -lactâmicos incluindo para tanto um termo de não-idealidade calculado pelo modelo de Pitzer para soluções eletrolíticas. Ao parâmetro de interação introduzido pelo modelo de Pitzer será dada uma interpretação a partir da termodinâmica estatística que permitirá uma discussão mais robusta do ponto de vista científico.

No quinto capítulo serão deduzidas relações entre o segundo coeficiente virial, a concentração de agente precipitante (tipicamente sais) e a solubilidade de proteínas. Serão estudados sistemas com lisozima, ovalbumina e imunoglobulina G humana. Outros modelos publicados na literatura serão comparados ao modelo proposto.

O sexto e derradeiro capítulo será dedicado exclusivamente a exposição das conclusões deste trabalho e a sugestão de trabalho futuros.

Todas estas contribuições visam a aprofundar e a discutir a físico-química envolvida em sistemas que contêm aminoácidos ou proteínas do ponto de vista clássico da termodinâmica de equilíbrio de fases. As relações aqui desenvolvidas são de utilidade não apenas da discussão científica, mas claramente da aplicação tecnológica.

Acaba-se, por fim, este introito com uma esperança sutil lançada pelo famoso historiador romano, Tito Lívio, e aqui registrada sob a forma de epígrafe: a que os trabalhos que a esse sucedam o superem, e este será o consolo deste que o escreve.

2 Elementos fundamentais da termodinâmica de equilíbrio de fases

“Ο γέγραφα, γέγραφα

(Pilatos, in Evangelho de São João)

O objetivo de despendar algumas páginas com aspectos introdutórios aos problemas de equilíbrio de fases é facilitar ao leitor desta dissertação alguns conceitos sem os quais a compreensão do trabalho que se segue se tornaria extremamente árdua e laboriosa. Não há, todavia, algum intento aqui senão o de expor exclusivamente o formalismo termodinâmico necessário às deduções que se seguirão. Reitera-se que nada do que neste capítulo está apresentado é novo, nem tão pouco fruto da pesquisa a qual esta dissertação justifica sua existência, todavia, estas linhas que custaram mais de cem anos de pesquisa em termodinâmica são indispensáveis ao entendimento das discussões que se seguirão.

A epígrafe, ainda que de alguma forma jocosa (em uma tradução livre: “o que escrevi, escrevi”), pretende ilustrar a robustez da termodinâmica quanto um ramo da ciência consistente e inviolável.

2.1 A equação fundamental em sua representação entrópica

Dado um sistema simples com m componentes, a equação fundamental descreve completamente todos os estados de equilíbrio. Em sua representação entrópica, a equação fundamental é dada explicitando a entropia, S , do sistema como função da energia interna, U , do volume, V , e da quantidade de matéria, N , de cada componente (MODELL; REID, 1983):

$$S = S(U, V, N_1, \dots, N_m) \quad (2 - 1)$$

A equação fundamental representa, assim, uma hipersuperfície em um espaço de $m+3$ dimensões cujos pontos representam estados de equilíbrio do dado sistema.

Sendo uma diferencial exata, a equação fundamental pode ser associada às duas primeiras leis da termodinâmica de modo que a variação infinitesimal de entropia do sistema seja dada por:

$$dS(U, V, N_1, \dots, N_m) = \frac{1}{T} dU + \frac{p}{T} dV - \sum_{i=1}^m \frac{\mu_i}{T} dN_i \quad (2 - 2)$$

Em que T é a temperatura termodinâmica:

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U} \right)_{V, N} \quad (2 - 3)$$

p é a pressão do sistema:

$$p = T \left(\frac{\partial S}{\partial V} \right)_{U, N} \quad (2 - 4)$$

E, μ_i é o potencial químico da espécie i :

$$\mu_i = T \left(\frac{\partial S}{\partial N_i} \right)_{U, V, N_j, j \neq i} \quad (2 - 5)$$

2.2 A equação de Gibbs-Duhem

A equação fundamental expressa em sua forma entrópica pela Equação (2 – 1) pode também ser, e assim comumente é, escrita em sua representação energética:

$$U = U(S, V, N_1, \dots, N_m) \quad (2 - 6)$$

Com sua diferencial exata dada por:

$$dU(S, V, N_1, \dots, N_m) = T dS - p dV + \sum_{i=1}^m \mu_i dN_i \quad (2 - 7)$$

Sendo a transformada de Legendre total de uma dada função

$f(\mathbf{x}): \mathbb{R}^n \rightarrow \mathbb{R} / \mathbf{x} = [x_1, \dots, x_n]$, tem-se que:

$$\mathcal{L}_T(f(\mathbf{x})) = f(\mathbf{x}) - \sum_{k=1}^n \left(\frac{\partial f(\mathbf{x})}{\partial x_k} \right) x_k \quad (2 - 8)$$

E, aplicando-se essa transformada de Legendre total na função definida na Equação (2 – 6), tem-se que:

$$\mathcal{L}_T(U(S, V, N_1, \dots, N_m)) = U(S, V, N_1, \dots, N_m) - TS + pV - \sum_{i=1}^m \mu_i N_i \quad (2 - 9)$$

Tomando a diferencial da Equação (2 – 9):

$$d\mathcal{L}_T(U(S, V, N_1, \dots, N_m)) = \dots$$

$$\dots dU(S, V, N_1, \dots, N_m) - TdS - SdT + pdV + Vdp - \sum_{i=1}^m \mu_i dN_i - \sum_{i=1}^m N_i d\mu_i \quad (2 - 10)$$

A equação de Gibbs-Duhem é simplesmente dada pela condição de que a variação infinitesimal da transformada de Legendre total da equação fundamental é nula, logo, observando-se a substituição da Equação (2 – 7) na Equação (2 – 10), tem-se:

$$SdT - Vdp + \sum_{i=1}^m N_i d\mu_i = 0 \quad (2 - 11)$$

Essa Equação (2 – 11) serve como uma restrição a ser respeitada por qualquer modelagem termodinâmica que se pretenda em equilíbrio de fases. Na verdade, ela é usada como garantia de consistência termodinâmica a um dado modelo.

2.3 Os critérios de equilíbrio de fases

Em geral, a apresentação que se faz dos critérios para os equilíbrio de fases perpassa caminhos matemáticos e propostas de exemplos que acabam por dificultar a compreensão.

Aqui, tentar-se-á um método pouco utilizado nas classes de termodinâmica, mas que segundo a percepção deste que escreve parece apresentar uma consistência matemática que deve por fim isentar dúvidas quanto a sua aplicação mais geral.

Tomemos para análise um sistema isolado com m componentes e π fases; de acordo com a segunda lei da termodinâmica a entropia deste sistema isolado sempre aumenta em uma dada transformação, logo no equilíbrio sua entropia deve ser máxima de modo que não haja nenhuma transformação que leve o sistema para outro estado que não seja o de equilíbrio.

Deve-se reescrever, então, a equação fundamental expressa pela equação (2 – 1) de modo a contemplar um sistema multifásico:

$$S = S(\mathbf{U}, \mathbf{V}, \mathbf{N}_1, \dots, \mathbf{N}_m) \quad (2 - 12)$$

Sendo os vetores definidos por coordenadas de cada fase do sistema:

$$\mathbf{U} = [U^{(1)}, U^{(2)}, \dots, U^{(\pi)}] \quad (2 - 13)$$

$$\mathbf{V} = [V^{(1)}, V^{(2)}, \dots, V^{(\pi)}] \quad (2 - 14)$$

$$\mathbf{N}_i = [N_i^{(1)}, N_i^{(2)}, \dots, N_i^{(\pi)}], \quad \text{com } i = 1, 2, \dots, m \quad (2 - 15)$$

Posto isto, tem-se que este é um problema de maximização de entropia sujeito às restrições de conservação energética, espacial e material:

$$\max S = S(\mathbf{U}, \mathbf{V}, \mathbf{N}_1, \dots, \mathbf{N}_m)$$

s. t. :

$$g_1(\mathbf{U}) = \sum_{\alpha=1}^{\pi} U^{(\alpha)} - U = 0$$

$$g_2(\mathbf{V}) = \sum_{\alpha=1}^{\pi} V^{(\alpha)} - V = 0$$

$$g_j(\mathbf{N}_i) = \sum_{\alpha=1}^{\pi} N_i^{(\alpha)} - N_i = 0, \quad i = 1, \dots, m \quad \text{e} \quad j = 3, \dots, m+2 \quad (2 - 16)$$

Assim, pode-se definir o lagrangeano da entropia como:

$$L(\mathbf{U}, \mathbf{V}, \mathbf{N}_1, \dots, \mathbf{N}_m, \lambda) = S(\mathbf{U}, \mathbf{V}, \mathbf{N}_1, \dots, \mathbf{N}_m) - \sum_{k=1}^{m+2} \lambda_k g_k \quad (2 - 17)$$

A condição de otimalidade é dada pela condição de gradiente do lagrangeano da entropia nulo:

$$\nabla L(\mathbf{U}, \mathbf{V}, \mathbf{N}_1, \dots, \mathbf{N}_m, \lambda) = 0 \quad (2 - 18)$$

Com isso, ter-se-ão $(m+2)(\pi+1)$ equações a serem satisfeitas:

$$\left\{ \begin{array}{l} \frac{\partial L(\mathbf{U}, \mathbf{V}, \mathbf{N}_1, \dots, \mathbf{N}_m, \lambda)}{\partial U^{(\alpha)}} = \frac{1}{T^{(\alpha)}} - \lambda_1 = 0, \quad \alpha = 1, \dots, \pi \\ \frac{\partial L(\mathbf{U}, \mathbf{V}, \mathbf{N}_1, \dots, \mathbf{N}_m, \lambda)}{\partial V^{(\alpha)}} = \frac{p^{(\alpha)}}{T^{(\alpha)}} - \lambda_2 = 0, \quad \alpha = 1, \dots, \pi \\ \frac{\partial L(\mathbf{U}, \mathbf{V}, \mathbf{N}_1, \dots, \mathbf{N}_m, \lambda)}{\partial N_i^{(\alpha)}} = -\frac{\mu_i^{(\alpha)}}{T^{(\alpha)}} - \lambda_{i+2} = 0, \quad \alpha = 1, \dots, \pi \quad \text{e} \quad i = 1, \dots, m \end{array} \right. \quad (2 - 19)$$

Obviamente para completar as $(m+2)(\pi+1)$ equações a serem satisfeitas, devem-se incluir as restrições de igualdade que reaparecem nas derivadas parciais do lagrangeano pelo multiplicadores de Lagrange.

Logo, tem-se que para maximizar a entropia de um sistema isolado com m componentes e π fases, devem-se obedecer as seguintes relações:

$$\begin{aligned}\frac{1}{\lambda_1} &= T^{(1)} = T^{(2)} = \dots = T^{(\pi)} \\ \frac{\lambda_2}{\lambda_1} &= p^{(1)} = p^{(2)} = \dots = p^{(\pi)} \\ -\frac{\lambda_{i+2}}{\lambda_1} &= \mu_i^{(1)} = \mu_i^{(2)} = \dots = \mu_i^{(\pi)}, \quad i=1, \dots, m\end{aligned}\tag{2-20}$$

As equações apresentadas em (2 – 20) representam os critérios do equilíbrio de fases: igualdade das temperaturas, das pressões e dos potenciais químicos de cada componente.

Do ponto de vista de fenômenos de transporte é possível, e até talvez mais fácil, imaginar que o equilíbrio é um caso particular do estado estacionário para o qual a força motriz da grandeza responsável pelo transporte é constante e nula, *id est*, o transporte só ocorre se há diferença de potencial, se o potencial térmico (temperatura) for o mesmo em todas as fases do sistema não há transporte de calor, se o potencial mecânico (pressão) for igual em todas as fases do sistema não há deslocamento de fronteira e, analogamente, se o potencial químico de cada componente for o mesmo em todas as fases não há transporte de massa. As equações fenomenológicas mostram claramente como o fluxo é dependente da força motriz de tal modo que se esta última for nula não há fluxo e o sistema está em equilíbrio. Ainda que se considerem acoplamentos, a situação de equilíbrio é aquela em que todas as forças motrizes são nulas; portanto, mesmo os efeitos de acoplamento não ocorrerão.

2.4 Fugacidade, atividade e convenções

O potencial químico tal como proposto por Gibbs na segunda metade do século XIX

(PRAUSNITZ; LICHTENTHALER; AZEVEDO, 1999) não é de fácil assimilação, porque não é possível sua percepção no mundo físico real de modo tão intuitivo quanto o potencial térmico (temperatura) e o mecânico (pressão). No entanto, a definição de fugacidade dá uma idéia mais sensível do fenômeno físico. A fugacidade, assim como definida por Lewis, substitui a pressão no cálculo do potencial químico de um processo isotérmico de um gás ideal de modo a contemplar os efeitos não considerados no modelo de gás ideal:

$$\frac{f_i}{f_i^0} = \exp\left(\frac{\mu_i - \mu_i^0}{RT}\right) \quad (2 - 21)$$

em que f_i é a fugacidade do componente i , f_i^0 é a fugacidade do componente i no estado padrão, μ_i é o potencial químico do componente i , μ_i^0 é o potencial químico do componente i no estado padrão, R é a constante real dos gases ($R = 8,314 \text{ J mol}^{-1} \text{ K}^{-1}$), T é a temperatura do sistema em K.

Pela sua definição é possível equivaler a igualdade de fugacidades de cada componente em todas as fases do sistema como um critério de equilíbrio alternativo ao da igualdade de potenciais químicos expressa na Equação (2 – 20) sem perda de generalidade.

Para um gás ideal puro, a fugacidade é igual a pressão e, para uma mistura de gases ideais, é igual a pressão multiplicada pela fração molar: $y_i p$. A definição de fugacidade é assim completada pelo limite observável a todos os sistemas puros ou misturados no qual o comportamento destes sistemas se aproxima do de gás ideal para pressões muito baixas:

$$\lim_{p \rightarrow 0} \frac{f_i}{y_i p} = 1 \quad (2 - 22)$$

em que y_i é a fração molar do componente i e p é a pressão do sistema.

Embora a definição de fugacidade ajude nos cálculos de equilíbrio de fases de gases, sua aplicação a sistemas condensados é um tanto quanto mais complicada. Isto se deve ao fato de ter-se relacionado tudo a diferença entre o comportamento do sistema real e o do gás ideal; e, portanto, isso exige dados experimentais a temperatura e composição constante sobre todo o intervalo de densidades não se excetuando a região bifásica. Como na maioria das vezes não se dispõe de tamanha quantidade de informações, usa-se alternativamente uma outra abordagem para cálculos da fase líquida. Definindo uma solução ideal e calculando os desvios entre

a solução real e essa solução ideal pelas propriedades de excesso, que no caso da energia livre de Gibbs de excesso – ou ainda, energia livre de Gibbs excedente – está relacionada ao coeficiente de atividade. Assim, cabe definir a atividade de um componente e o coeficiente de atividade deste componente. Da própria definição de fugacidade expressa pela Equação (2 – 21), define-se a atividade como:

$$a_i = \frac{f_i}{f_i^0} \quad (2 - 23)$$

De modo similar ao que está escrito em (2 – 22), surge a necessidade de se definir o coeficiente de atividade, γ_i , de um dado componente i como:

$$\gamma_i = \frac{a_i}{x_i} \quad (2 - 24)$$

Aqui cabe explicitar algo que será necessário: a convenção simétrica e a assimétrica. Esta necessidade emerge a medida em que se arbitra qual é o estado padrão (ou estado de referência) com o qual se executará o cálculo do potencial químico. Na convenção simétrica, o estado padrão é o mesmo para todos os componentes; já na convenção assimétrica, não.

2.5 Modelos termodinâmicos de energia livre de Gibbs excedente

O coeficiente de atividade está relacionado a energia livre de Gibbs excedente – ou seja, a diferença entre a energia livre de Gibbs de um sistema real e a energia livre de Gibbs de uma mistura ideal – pela seguinte equação:

$$\ln \gamma_i = \left[\frac{\partial \left(\frac{g^{EX}}{RT} \sum_{i=1}^m N_i \right)}{\partial N_i} \right]_{T, p, N_j} \quad (2 - 25)$$

A Equação (2 – 25) traz a possibilidade de ao se propor um modelo de energia livre de Gibbs excedente, possa-se avaliar a não-idealidade de um dado componente em solução. Um dos modelos mais aplicados nos últimos quarenta anos é o famoso NRTL (da sigla em inglês: *Non-Random Two Liquids*) que foi proposto em um artigo de Renon e Prausnitz (1968). Na sua forma para sistemas multicomponentes, a

energia livre de Gibbs excedente é dada por:

$$\frac{g^{EX}}{RT} = \sum_{i=1}^m x_i \frac{\sum_{j=1}^m \tau_{ji} G_{ji} x_j}{\sum_{l=1}^m G_{li} x_l} \quad (2-26)$$

em que, o parâmetro τ_{ij} é dado pela diferença entre os parâmetros de energia característicos da interação de i e j , g_{ij} :

$$\tau_{ji} = \frac{g_{ji} - g_{ii}}{RT} \quad (2-27)$$

E, o parâmetro G_{ij} é dado pela exponencial da multiplicação entre o parâmetro τ_{ij} e o parâmetro de não-aleatoriedade na mistura α_{ij} :

$$G_{ji} = \exp(-\alpha_{ji} \tau_{ji}), \quad \text{com } \alpha_{ji} = \alpha_{ij} \quad (2-28)$$

Em termos do logaritmo natural do coeficiente de atividade de um dado componente i , tem-se que:

$$\ln \gamma_i = \frac{\sum_{j=1}^m \tau_{ji} G_{ji} x_j}{\sum_{l=1}^m G_{li} x_l} + \sum_{j=1}^m \frac{x_j G_{ij}}{\sum_{l=1}^m G_{lj} x_l} \left(\tau_{ij} - \frac{\sum_{r=1}^m x_r \tau_{rj} G_{rj}}{\sum_{l=1}^m G_{li} x_l} \right) \quad (2-29)$$

Coube aqui apresentar mais pormenorizadamente este modelo em especial, para ajudar a discussão de uma particular aplicação a sistemas contendo aminoácidos apresentada no capítulo quarto. Evidentemente há outros tantos e tão importantes modelos que inclusive serão citados ao decorrer do texto, todavia como a aplicação do modelo NRTL será comparada a aplicação de um outro modelo neste trabalho, considerou-se por bem apresentar sua formulação.

2.6 Modelos termodinâmicos para soluções eletrolíticas

Enquanto a própria teoria de líquidos ainda é objeto de intensa pesquisa, não foi abandonada durante o século passado, e nem nos próximos o será, a modelagem de soluções que contenham eletrólitos. A natureza de tais soluções traz

complicações adicionais a este tipo de modelagem e restrições antes não consideradas tais como o princípio de eletroneutralidade da fase termodinâmica em equilíbrio estável.

A teoria de Debye-Hückel publicada no início dos anos de 1920 foi pioneira no tratamento de soluções eletrolíticas e até hoje constitui um trabalho de referência na maioria das publicações relacionadas a este tema. Esta teoria está baseada na solução linearizada e unidirecional da equação de Poisson-Boltzmann (HILL, 1986):

$$\nabla^2 \psi_j = \frac{-4\pi e}{\varepsilon} \sum_i z_i \rho_i \left(1 - \frac{z_i e \psi_i}{k_B T} \right) \quad (2 - 30)$$

em que ψ é o potencial da espécie j , e é a carga elétrica, ε é a permissividade do vácuo, z_i é a carga elétrica da espécie i , ρ_i é a concentração da espécie i , k_B é a constante de Boltzmann e T é a temperatura.

Pela solução da Equação (2 - 30), e por algumas considerações, chega-se ao coeficiente de atividade de uma espécie iônica no limite de altas diluições:

$$\ln \gamma_j = - \frac{z_j^2 e^2 \kappa}{2 \varepsilon k_B T} \quad (2 - 31)$$

em que κ é o inverso do comprimento de Debye¹:

$$\kappa^2 = \frac{4\pi e^2}{\varepsilon k_B T} \sum_i \rho_i z_i^2 \quad (2 - 32)$$

Muitas vezes esta Equação (2 - 32) é escrita em termos da força iônica, I , que é definida como:

$$I = \frac{1}{2} \sum_i m_i z_i^2 \quad (2 - 33)$$

em que m_i é a molalidade do componente i em mol por kilograma.

De fato este modelo é um modelo limite a concentrações de eletrólitos muito baixas, tipicamente limitado a força iônica da ordem de 0,01 mol kg⁻¹ (PRAUSNITZ; LICHTENTHALER; AZEVEDO, 1999). No decorrer do século passado outros modelos foram propostos com o intento de ampliar a faixa de aplicação a sistemas eletrolíticos; um dos mais exitosos nesta empresa é o modelo de Pitzer (PITZER, 1973). No fundo, este modelo pode ser dividido em dois termos principais: um de

¹ O comprimento de Debye é uma distância de interação característica associada ao potencial eletrostático blindado (PRAUSNITZ; LICHTENTHALER; AZEVEDO, 1999).

longo alcance que é devido as interações eletrostáticas análogas ao modelo de Debye-Hückel, e um de curto alcance que é uma expansão virial de modo a assumir a hipótese que as interações entre muitos corpos pode ser dividida em contribuições de interações de dois-a-dois, três-a-três e assim por diante. Este modelo também se baseia em uma proposição para a energia livre de Gibbs excedente:

$$\frac{g^{EX}}{RTw_s} = f(I) + \sum_i \sum_j m_i m_j \lambda_{ij}(I) + \sum_i \sum_j \sum_k m_i m_j m_k \Lambda_{ijk} + \dots \quad (2 - 34)$$

em que w_s é a massa de solvente, $f(I)$ é uma função dependente da força iônica, λ_{ij} é o parâmetro das interações binárias (dois-a-dois) e Λ_{ijk} é o parâmetro das interações ternárias (três-a-três).

Como as propriedades termodinâmicas são usualmente medidas para eletrólitos neutros e não espécies iônicas individuais, convencionou-se escrever os modelos para soluções eletrolíticas aplicando a restrição de eletroneutralidade em termos de um coeficiente de atividade iônica médio que é dado pela média geométrica dos coeficientes de atividade do cátion e o do ânion. Esta abordagem que para sistemas salinos, por exemplo, é muito interessante, não o é para polieletrólitos, para os quais algumas complicações aparecem. Em termos do coeficiente de atividade iônica médio, o modelo de Pitzer é dado por:

$$\begin{aligned} \ln \gamma_{\pm} = & -\frac{1}{3} |z_+ z_-| \left(\frac{e^2}{\epsilon_0 \epsilon_r k_B T} \right)^{3/2} \frac{(2 d_s N_A)^{1/2}}{8 \pi} \left[\frac{2}{b} \ln(1 + b\sqrt{I}) + \frac{\sqrt{I}}{1 + b\sqrt{I}} \right] + \dots \\ & \dots + m \left(\frac{2\nu_+ \nu_-}{\nu_+ + \nu_-} \right) \left\{ 2\beta^{(0)} + \frac{2\beta^{(1)}}{\alpha^2 I} \left[1 - \left(1 + \alpha I^{1/2} - \frac{\alpha^2 I}{2} \right) \exp(-\alpha I^{1/2}) \right] \right\} + \dots \\ & \dots + 9m^2 \left[\frac{(\nu_+ \nu_-)^2}{\nu_+ + \nu_-} \right] (\nu_+ \Lambda_{MMX} + \nu_- \Lambda_{MXX}) \end{aligned} \quad (2 - 35)$$

em que ν_+ e ν_- são respectivamente os coeficientes estequiométricos do cátion M e do ânion X . Os outros parâmetros serão definidos no quarto capítulo quando da aplicação do modelo de Pitzer.

3 Efeito do pH na solubilidade de proteínas

“Si l’on veut énoncer le principe dans toute sa généralité et en l’appliquant à l’Univers, on le voit pour ainsi dire s’évanouir et il ne reste plus que ceci: il y a quelque chose qui demeure constant.”

(Henri Poincaré, *in Thermodynamique*)

3.1 Introdução

Os primeiros trabalhos relacionados à modelagem da curva de solubilidade de proteínas em função do pH datam do início do século XX. Green (1931) propôs um modelo para a curva de solubilidade da carboxihemoglobina equina considerando que esta proteína se comportaria tanto como um ácido bivalente quanto uma base bivalente. Já Grönwall (1941) atribuiu a Linderstrøm-Lang o desenvolvimento de um modelo capaz de prever a curva de solubilidade a partir da integração do perfil de carga da proteína. Este trabalho de Grönwall, apesar de esquecido, é de extrema importância e relevância tanto para o trabalho aqui apresentado quanto, e aqui um motivo muito mais nobre, para o desenvolvimento de modelos a partir das ferramentas da termodinâmica clássica.

É digno de nota o trabalho de Fredericq e Neurath (1950) no qual uma proposta de equacionamento para uma correlação semi-empírica da curva de solubilidade de insulina é apresentada. Não há, de fato, algo novo em relação à modelagem do efeito do pH na curva de solubilidade neste trabalho; no entanto, apesar de não haver qualquer menção ou referência quanto à origem da equação proposta, nota-se que é uma aplicação direta da equação descrita por Grönwall (1941). Há uma hipótese implícita no trabalho de que o perfil de carga líquida da insulina em função do pH seja linear na região do ponto isoelétrico, o que é bastante plausível para muitas proteínas. Assim, pela integração da equação diferencial proposta por Grönwall (1941), uma correlação semi-empírica quadrática é obtida. Reforça-se o uso da expressão semi-empírica devido à natureza dual da equação: em parte advinda de um desenvolvimento teórico descrito por Grönwall (1941) e atribuído a Linderstrøm-Lang; e em parte pela necessidade de se obterem parâmetros a partir

de um ajuste de um conjunto de dados experimentais.

Contudo, os trabalhos mais recentes se concentram na modelagem da curva de solubilidade de aminoácidos em função do pH dando maior enfoque aos modelos termodinâmicos para soluções não-ideais, tanto através de equações de estado como de modelos de energia livre de Gibbs excedente.

Entretanto, o primeiro trabalho dedicado exclusivamente à modelagem e predição da curva de solubilidade de proteínas propriamente dita em função do pH foi publicado por Tjong e Zhou (2008). Neste trabalho, a solubilidade foi calculada a partir da energia livre de transferência da fase sólida para a fase líquida. Para a execução deste cálculo, a energia livre de transferência foi dividida em duas contribuições: uma de natureza eletrostática, calculada pela média de 100 conformações protéicas advindas da aplicação de dinâmica molecular a pH constante; e outra, de natureza não-eletrostática, baseada na área superficial acessível ao solvente. Porém, o esforço computacional requerido pelo modelo por eles proposto para calcular apenas alguns pontos da curva não apresenta uma aplicabilidade razoável. Além deste trabalho, Tashima et al. (2009) propuseram uma modelagem termodinâmica para prever a curva de solubilidade especificamente da insulina suína aplicando o modelo de Pitzer truncado após o segundo coeficiente virial.

Aqui uma relação analítica entre a solubilidade de proteínas em função do pH será desenvolvida sob a hipótese de que a atividade da molécula de proteína eletricamente neutra permanece constante com o pH. Assim se entende a epígrafe deste capítulo e que é encontrada não apenas neste mas em muitos desenvolvimentos teóricos: há alguma coisa que permanece constante.

3.2 Desenvolvimento teórico

No caso do equilíbrio sólido-líquido, no qual a fase sólida é única e contém apenas moléculas de um soluto i , a condição do equilíbrio de fases expressa na Equação (2 – 20) supondo a igualdade de pressões e temperaturas das fases *a priori* é dada simplesmente por:

$$\mu_i^L(T, p, \mathbf{x}) = \mu_i^S(T, p) \quad (3 - 1)$$

em que μ_i^L é o potencial químico do soluto na fase líquida e μ_i^S é o potencial químico do soluto na fase sólida. Desprezando-se os efeitos da pressão sobre o cálculo do potencial químico, hipótese comum a sistemas que não envolvam fase gasosa, tem-se que para um processo isotérmico não há variação do potencial químico da fase sólida já que sua composição é fixa; portanto, o potencial químico do soluto na fase líquida também não se altera em tal processo. Posto isto, conclui-se que mudanças nas condições da fase líquida tais como variação de pH não alteram o potencial químico do soluto nesta fase.

A solubilidade de um soluto em fase líquida é a concentração de equilíbrio deste componente nesta fase. Assim, se a solubilidade do soluto for suficientemente diminuta, para um processo isotérmico desprezando-se os efeitos da pressão, pode-se adotar a consideração de uma solução ideal segundo a lei de Henry e expressar o potencial químico deste soluto na fase líquida em termos de molalidade como:

$$\mu_i^L(\mathbf{m}) = \mu_i^0(\mathbf{m}^0) + RT \ln \frac{m_i}{m_i^0} \quad (3 - 2)$$

em que:

$$\mathbf{m}^0 = [0, 0, \dots, 0, m_i^0 = 1, 0, \dots, 0] \quad (3 - 3)$$

Sendo m_i a molalidade do soluto na fase líquida, ou ainda a solubilidade do soluto em termos de molalidade; e, $\mu_i^0(\mathbf{m}^0)$ o potencial químico de referência do soluto. Este estado de referência escolhido é o estado de referência de uma solução ideal de concentração unitária, ou seja, uma solução hipotética na qual o soluto tem uma molalidade $m_i^0 = 1,0 \text{ mol kg}^{-1}$, mas experimenta as mesmas interações que se estivesse em uma solução a infinita diluição.

Quando o soluto é uma proteína, outros pormenores devem ser levados em conta. A molécula de proteína é um polieletrólito biológico e, sendo assim, esta se ioniza em solução aquosa. Isto se deve ao fato de que alguns aminoácidos que compõem a proteína têm grupos polares que participam de equilíbrios químicos com o íon H^+ .

Para grupos alcalinos, o equilíbrio de ionização é:

Todavia, para grupos ácidos, o equilíbrio de ionização é:

A partir destes equilíbrios químicos é possível notar que as moléculas de proteína em uma solução têm diferentes estados de protonação, e, portanto, distintos valores de carga elétrica. Porém, a solubilidade determinada experimentalmente é a soma da concentração de todas as moléculas de proteínas.

Sendo a fase sólida composta apenas por proteína e sendo uma fase em equilíbrio estável, pelo princípio de eletroneutralidade nela deve haver somente moléculas eletricamente neutras. Isto significa que o equilíbrio de fases se estabelece apenas entre moléculas de proteína eletricamente neutras. Logo, a Equação (3 – 2) não é válida para qualquer molécula de proteína, mas apenas e tão-somente para as moléculas de proteína eletricamente neutras. E, por consequência do critério de equilíbrio de fases expresso na Equação (3 – 1), a molalidade das moléculas de proteína eletricamente neutras na fase líquida deve permanecer inalterada com a variação do pH da solução. A molalidade destas moléculas é uma fração da solubilidade da proteína nesta solução. Assim pode-se escrever:

$$m_0 = \phi_0(pH) S(pH) \quad (3 - 6)$$

em que m_0 é a molalidade das moléculas de proteína eletricamente neutras da fase líquida, ϕ_0 é a fração de moléculas de proteína eletricamente neutras da fase líquida e S é a solubilidade da proteína na fase líquida. Contudo, conforme afirmado anteriormente, m_0 deve ser constante com a variação do pH, assim ao se aplicar a Equação (3 – 6) no ponto isoelétrico, pode-se escrever que:

$$m_0 = \phi_0(pI) S(pI) \quad (3 - 7)$$

Logo, pela simples comparação entre as Equações (3 – 6) e (3 – 7), chega-se à seguinte relação:

$$S(pH) = \frac{\phi_0(pI) S(pI)}{\phi_0(pH)} \quad (3 - 8)$$

A Equação (3 – 8) descreve a curva de solubilidade de uma dada proteína em solução em função do pH. Apesar de ser uma equação simples, resta ainda a dificuldade de se calcular a fração de moléculas de proteína eletricamente neutras na fase líquida.

3.2.1 Cálculo da fração de proteína eletricamente neutra

Para o cálculo da fração de moléculas de proteína eletricamente neutras é indispensável considerar que todos os resíduos efetivamente ionizáveis da proteína, inclusive os grupos amina e ácido carboxílico terminais, podem estar eletricamente carregados ou não devido aos equilíbrios químicos expressos nas Equações (3 – 4) e (3 – 5). Sendo n o número total de resíduos ionizáveis de uma dada proteína, o número de estados de ionização possíveis, ψ , é:

$$\psi = \sum_{k=0}^n \frac{n!}{(n-k)!k!} = \sum_{k=0}^n \binom{n}{k} = 2^n \quad (3 - 9)$$

Dentre estes estados, alguns representam moléculas de proteína cuja carga total seja nula. A carga de uma molécula de proteína é dada pela soma da carga individual de cada resíduo j em um dado estado de ionização k :

$$C_k = \sum_{j=1}^n \varepsilon_{jk} \quad (3 - 10)$$

em que ε_{jk} é a carga individual de cada resíduo j em um dado estado de ionização k e pode ser expressa como:

$$\varepsilon_{jk} = \begin{cases} -1, & \text{para grupos ácidos carregados} \\ 0, & \text{para quaisquer grupos neutros} \\ +1, & \text{para grupos básicos carregados} \end{cases} \quad (3 - 11)$$

Considerando que a equação de Henderson-Hasselbach é válida para cada grupo j e que o equilíbrio químico estabelecido entre o grupo j e o íon H^+ é independente de qualquer outro equilíbrio químico desta natureza que ocorra nesta mesma molécula de proteína, tem-se que a fração ionizada deste grupo j é dada por:

$$\alpha_j(pH) = \left(\frac{1}{1 + 10^{pH - pK_{Aj}}} \right) \left[\frac{(1 + v_j)}{2} - \frac{(v_j - 1)}{2} 10^{pH - pK_{Aj}} \right] \quad (3 - 12)$$

em que pK_{Aj} é o co-logaritmo decimal da constante de equilíbrio químico do grupo j e v_j é o estado carregado do grupo j (-1, para grupos ácidos e, +1, para grupos alcalinos).

Assim, a fração de moléculas de uma dada proteína em um dado estado de ionização k é dada pelo produtório da fração do grupo j com a carga específica

associada ao estado de ionização k :

$$\phi_k(pH) = \prod_{j=1}^n \alpha'_j(pH) \quad (3-13)$$

Sendo que:

$$\alpha'_j(pH) = \begin{cases} \alpha_j(pH), & \text{se } \varepsilon_{jk} \neq 0 \\ 1 - \alpha_j(pH), & \text{se } \varepsilon_{jk} = 0 \end{cases} \quad (3-14)$$

Logo, substituindo a Equação (3-12) na Equação (3-14) e a resultante na Equação (3-13), ter-se-á que a fração de moléculas de proteína eletricamente neutras em um dado estado de ionização k é dada simplesmente por:

$$\phi_k(pH) = \prod_{j=1}^n \left\{ \left(\frac{1}{1 + 10^{pH - pK_{A_j}}} \right) \left[\left(\frac{1 + v_j}{2} - \varepsilon_{jk} \right) 10^{pH - pK_{A_j}} - \left(\frac{v_j - 1}{2} - \varepsilon_{jk} \right) \right] \right\} \quad (3-15)$$

Sujeita à seguinte restrição de eletroneutralidade:

$$\sum_{j=1}^n \varepsilon_{jk} = 0 \quad (3-16)$$

Apesar de a fração de moléculas de proteína eletricamente neutras de um dado estado de ionização k depender do pH, a razão entre esta fração e a fração de moléculas de proteína eletricamente neutras em outro estado de ionização qualquer l não depende do pH. Pode-se demonstrar por argumentos de equilíbrio químicos entre moléculas eletricamente neutras em diferentes estados de ionização que esta razão é dada por (FRANCO; PESSÔA FILHO, 2011) :

$$\log \frac{\phi_k(pH)}{\phi_l(pH)} = \sum_{j=1}^n pK_{A_j} (\varepsilon_{jk} - \varepsilon_{jl}) \quad (3-17)$$

Assim, é indiferente qual estado de ionização que resulte em uma molécula eletricamente neutra de proteína é utilizado para efetuar o cálculo da Equação (3-8). Combinando as Equações (3-8) e (3-15), tem-se que a curva de solubilidade de uma dada proteína em função do pH é dada por:

$$S(pH) = S(pI) \prod_{j=1}^n \left\{ \frac{\left(\frac{1 + 10^{pH - pK_{A_j}}}{1 + 10^{pI - pK_{A_j}}} \right) \left[\left(\frac{1 + v_j}{2} - \varepsilon_{jk} \right) 10^{pI - pK_{A_j}} - \left(\frac{v_j - 1}{2} - \varepsilon_{jk} \right) \right]}{\left[\left(\frac{1 + v_j}{2} - \varepsilon_{jk} \right) 10^{pH - pK_{A_j}} - \left(\frac{v_j - 1}{2} - \varepsilon_{jk} \right) \right]} \right\} \quad (3-18)$$

para qualquer estado de ionização k que respeite a restrição imposta na Equação (3 – 16). Basicamente a Equação (3 – 18) é uma expressão para a predição da curva de solubilidade de uma dada proteína em função do pH da solução e requer como parâmetros apenas a solubilidade no ponto isoelétrico e informações a respeito da estrutura primária da proteína, ou seja, a seqüência de aminoácidos que a compõe e as respectivas constantes de equilíbrio de ionização. Isto torna esta abordagem extremamente relevante e de fácil aplicação.

3.2.2 Extensão à teoria de Linderstrøm-Lang e Grönwall

É fato que as primeiras tentativas de se prever a curva de solubilidade de uma proteína, em função seja do pH seja da concentração de agente precipitante, datam da primeira metade do século passado, quando ainda muito poucas informações estruturais sobre proteínas eram conhecidas. Ainda assim, célebres trabalhos como o de Cohn (1925) e o de Green (1931) são ainda hoje referências abundantemente citadas e prepararam de certo modo o caminho trilhado neste quase um século de dedicação à pesquisa sobre a físico-química de sistemas protéicos. Infelizmente alguns trabalhos de não menor importância ficaram esquecidos. Este é o caso do trabalho de Grönwall (1941). Neste trabalho Grönwall atribui a Linderstrøm-Lang o desenvolvimento de uma teoria para o cálculo da solubilidade de proteínas em função do pH através da integração do perfil de carga líquida da proteína em função do pH.

As hipóteses desta teoria apresentada por Grönwall (1941) são: i) o equilíbrio sólido-líquido se estabelece com as moléculas eletricamente neutras de proteína em solução; ii) a variação do pH da solução não altera a atividade das moléculas eletricamente neutras de proteína em solução.

Assim, poder-se-ia reconstituir a formulação de Linderstrøm-Lang e Grönwall partindo do seguinte equilíbrio entre uma molécula de proteína eletricamente neutra e uma molécula ionizada em um dado estado de ionização k com uma carga total c :

cuja relação de equilíbrio é dada por:

$$K_k = \frac{a_{P_k^c}}{a_{P_0} a_{H^+}^c} \quad (3 - 20)$$

em que K_k é a constante de equilíbrio referente a Equação química (3 – 19), a_{P_k} é a atividade da molécula de proteína de carga total c em um estado de ionização k , a_{P_0} é a atividade da molécula de proteína eletricamente neutra e a_{H^+} é a atividade do íon H^+ . Todavia, a atividade de um componente é o produto entre sua concentração e o coeficiente de atividade deste componente (γ) – Equação (2 – 24). Assim, a molalidade da molécula de proteína com carga total c e em um estado de ionização k é dada por:

$$m_k = K_k \frac{a_{P_0} a_{H^+}^c}{\gamma_k} \quad (3 - 21)$$

Já a solubilidade da proteína é a soma das concentrações de todos os estados de ionização possíveis, mesmo de cargas totais distintas:

$$S = \sum_{k=1}^{\psi} m_k = \sum_{k=1}^{\psi} K_k \frac{a_{P_0} a_{H^+}^c}{\gamma_k} \quad (3 - 22)$$

Derivando a Equação (3 – 22) com respeito à atividade do íon H^+ e assumindo que a atividade da molécula de proteína eletricamente neutra permanece constante:

$$\frac{dS}{da_{H^+}} = a_{P_0} \sum_{k=1}^{\psi} c a_{H^+}^{c-1} \frac{K_k}{\gamma_k} - a_{P_0} \sum_{k=1}^{\psi} a_{H^+}^c \frac{K_k}{\gamma_k^2} \frac{d\gamma_k}{da_{H^+}} \quad (3 - 23)$$

O segundo termo do lado direito da Equação (3 – 23) pode ser simplificado do seguinte modo:

$$a_{P_0} \sum_{k=1}^{\psi} a_{H^+}^c \frac{K_k}{\gamma_k^2} \frac{d\gamma_k}{da_{H^+}} = a_{P_0} \sum_{k=1}^{\psi} a_{H^+}^c \frac{K_k}{\gamma_k} \frac{d \ln \gamma_k}{da_{H^+}} = \sum_{k=1}^{\psi} m_k \frac{d \ln \gamma_k}{da_{H^+}} \quad (3 - 24)$$

Considerando que os coeficientes de atividade da água e do íon H^+ são constantes a Equação (3 – 24) deve ser nula para que a equação de Gibbs-Duhem seja satisfeita (uma dedução para a equação de Gibbs-Duhem está apresentada no seção 2.2 do segundo capítulo). Assim, a Equação (3 – 23) pode ser reescrita como:

$$\frac{d \ln S}{d \ln a_{H^+}} = \left(\frac{a_{H^+}}{S} \right) \frac{dS}{da_{H^+}} = \frac{\sum_{k=1}^{\psi} c a_{H^+}^c \frac{K_k}{\gamma_k}}{\sum_{k=1}^{\psi} a_{H^+}^c \frac{K_k}{\gamma_k}} = \frac{\sum_{k=1}^{\psi} c m_k}{\sum_{k=1}^{\psi} m_k} = \bar{c} \quad (3 - 25)$$

em que $\bar{c}$ é a carga média das moléculas de proteína conforme proposta de Linderstrøm-Lang (1924). Reescrevendo a Equação (3 – 25) em termos de pH:

$$\frac{d \log S}{d pH} = -\bar{c} \quad (3 - 26)$$

A Equação (3 – 26) descreve a curva de solubilidade da proteína em função do pH pela integração da carga média da proteína. Esta equação foi primeiramente derivada por Grönwall (1941). Entretanto, o único método de aplicação desta equação encontrado na literatura era realizado pela aproximação linear do perfil de carga na região do ponto isoelétrico e a integração da reta resultante. Contudo, pode-se relacionar a carga média da proteína com as informações da estrutura primária da mesma através da seguinte equação:

$$\bar{c}(pH) = \sum_{j=1}^n v_j \alpha_j(pH) \quad (3 - 27)$$

Inserindo a Equação (3 – 12) na Equação (3 – 27), tem-se que:

$$\bar{c}(pH) = \frac{1}{2} \sum_{j=1}^n \left(\frac{1}{1 + 10^{pH - pK_{A_j}}} \right) [1 + v_j - (1 - v_j) 10^{pH - pK_{A_j}}] \quad (3 - 28)$$

Logo, substituindo a Equação (3 – 28) na Equação (3 – 26):

$$\frac{d \log S}{d pH} = -\frac{1}{2} \sum_{j=1}^n \left(\frac{1}{1 + 10^{pH - pK_{A_j}}} \right) [1 + v_j - (1 - v_j) 10^{pH - pK_{A_j}}] \quad (3 - 29)$$

A Equação (3 – 29) pode ser integrada de um pH de referência até o pH no qual se deseja calcular a solubilidade da proteína. Tomando como pH de referência o ponto isoelétrico, pI:

$$\int_{\log S(pI)}^{\log S(pH)} d \log S = -\frac{1}{2} \int_{pI}^{pH} \left\{ \sum_{j=1}^n \left(\frac{1}{1 + 10^{pH - pK_{A_j}}} \right) [1 + v_j - (1 - v_j) 10^{pH - pK_{A_j}}] \right\} d pH \quad (3 - 30)$$

Fazendo a integração como apresentado no Apêndice A, chega-se a:

$$\log \frac{S(pH)}{S(pI)} = \sum_{j=1}^n \left[\log \left(\frac{1 + 10^{pH - pK_{A_j}}}{1 + 10^{pI - pK_{A_j}}} \right) - \frac{(1 + v_j)}{2} (pH - pI) \right] \quad (3 - 31)$$

Ou ainda:

$$\log \frac{S(pH)}{S(pI)} = n_+(pI - pH) + \sum_{j=1}^n \left[\log \left(\frac{1 + 10^{pH - pK_{A_j}}}{1 + 10^{pI - pK_{A_j}}} \right) \right] \quad (3 - 32)$$

em que n_+ é o número total de grupos alcalinos na molécula de proteína.

A Equação (3 – 32) relaciona a solubilidade da proteína com a variação do pH da solução. A próxima subseção é dedicada a provar que tanto a Equação (3 – 18) quanto a Equação (3 – 32) são equivalentes.

3.2.3 Equivalência entre as duas abordagens

Apesar das formulações aparentemente distintas, tanto a integração do perfil de carga média da proteína quanto o desenvolvimento exposto em 3.2.1 são análogos.

Ao se tomar o logaritmo decimal da Equação (3 – 18), tem-se que:

$$\begin{aligned} \log \frac{S(pH)}{S(pI)} &= \sum_{j=1}^n \log \left(\frac{1 + 10^{pH - pK_{A_j}}}{1 + 10^{pI - pK_{A_j}}} \right) + \dots \\ &\dots + \sum_{j=1}^n \log \left\{ \frac{\left[\left(\frac{1 + v_j}{2} - \varepsilon_{jk} \right) 10^{pI - pK_{A_j}} - \left(\frac{v_j - 1}{2} - \varepsilon_{jk} \right) \right]}{\left[\left(\frac{1 + v_j}{2} - \varepsilon_{jk} \right) 10^{pH - pK_{A_j}} - \left(\frac{v_j - 1}{2} - \varepsilon_{jk} \right) \right]} \right\} \end{aligned} \quad (3 - 33)$$

Pelo bem da simplicidade, definir-se-ão duas funções:

$$\eta_j(pH) = \log \left(\frac{1 + 10^{pH - pK_{A_j}}}{1 + 10^{pI - pK_{A_j}}} \right) \quad (3 - 34)$$

$$\theta_{jk}(pH) = \log \left\{ \frac{\left[\left(\frac{1 + v_j}{2} - \varepsilon_{jk} \right) 10^{pI - pK_{A_j}} - \left(\frac{v_j - 1}{2} - \varepsilon_{jk} \right) \right]}{\left[\left(\frac{1 + v_j}{2} - \varepsilon_{jk} \right) 10^{pH - pK_{A_j}} - \left(\frac{v_j - 1}{2} - \varepsilon_{jk} \right) \right]} \right\} \quad (3 - 35)$$

Fica evidente pela observação das Equações (3 – 34) e (3 – 35) que a função η_j não depende do estado de ionização k , enquanto a função θ_{jk} depende. Assim, pode-se analisar cada possibilidade. Esta análise da função e θ_{jk} está na Tabela 3.1.

Tabela 3.1 - Análise da função θ_{jk} .

v_j	ε_{jk}	θ_{jk}	Número de resíduos
-1	-1	$pl - pH$	n_A
-1	0	0	n_B
+1	+1	0	n_C
+1	0	$pl - pH$	n_D

Pela Tabela 3.1, é possível inferir que a Equação (3 – 33) pode ser escrita como:

$$\log \frac{S(pH)}{S(pl)} = (n_A + n_D)(pl - pH) + \sum_{j=1}^n \eta_j(pH) \quad (3 - 36)$$

Entretanto, a Equação (3 – 36) deve também satisfazer a restrição expressa na Equação (3 – 16), assim:

$$\sum_{j=1}^n \varepsilon_{jk} = 0 \Rightarrow n_A = n_C \quad (3 - 37)$$

Portanto, de acordo com a restrição de eletroneutralidade advinda da Equação (3 – 37), tem-se que:

$$(n_A + n_D) = (n_C + n_D) = n_+ \quad (3 - 38)$$

Logo, a Equação (3 – 36) respeitando a restrição de eletroneutralidade pode ser escrita como:

$$\log \frac{S(pH)}{S(pl)} = n_+(pl - pH) + \sum_{j=1}^n \eta_j(pH) \quad (3 - 39)$$

Assim fica provada a identidade entre as Equações (3 – 18) e (3 – 32).

3.3 Resultados e Discussões

3.3.1 Solubilidade da insulina em função do pH e da temperatura

A Tabela 3.2 apresenta os valores de pK_A dos grupos ionizáveis da insulina humana para o cálculo da curva de solubilidade desta proteína em função do pH de acordo com a Equação (3 – 32). Há na Tabela 3.2 dois conjuntos de valores de pK_A : um chamado isolado que é o valor do pK_A do aminoácido livre obtido pela titulação

potenciométrica do aminoácido, e outro que é o valor do pK_A na cadeia protéica e é obtido pela titulação potenciométrica da proteína.

Tabela 3.2 - Valores de pK_A dos grupos ionizáveis da insulina humana.

Aminoácido	Grupo	Número	$pK_A(\text{isolado})^a$	$pK_A(\text{cadeia})^b$
Gly	-NH ₂ (terminal)	1	9,80	7,38
Glu	γ -COOH	4	4,30	4,37
Tyr	Fenol-OH	4	10,90	10,20
Asn	-COOH (terminal)	1	2,00	3,45
Phe	-NH ₂ (terminal)	1	9,10	7,38
His	Imidazol-NH ₂	2	6,00	5,98
Arg	Guanidina-NH ₂	1	12,50	11,10
Lys	ϵ -NH ₂	1	10,80	8,52
Ala	-COOH (terminal)	1	2,30	3,45

^a Stryer (1995).

^b Kaarsholm; Havelung e Hougaard (1990).

De fato, os dados que serão correlacionados aqui são de insulina suína; entretanto, a diferença entre a insulina suína e a insulina humana é de apenas um aminoácido em termos de estrutura primária, logo é bastante razoável considerar que os valores de pK_A na cadeia da insulina humana são praticamente iguais aos valores do pK_A na cadeia da insulina suína.

O código-fonte do programa de cálculo dos diversos estados de ionização bem como do cálculo da curva de solubilidade está apresentado no Apêndice B.

Como a insulina possui ao todo 16 grupos ionizáveis, o número de estados de ionização possíveis é: $\nu = 65536$. Destes, 8008 estados tem carga nula. A fração total é, pois, a soma das frações destes 8008 estados cuja carga total seja nula. Na Figura 3.1, ilustra-se a estrutura primária da insulina suína destacando-se os grupos ionizáveis e a distribuição de carga da molécula eletricamente neutra cuja fração é preponderante. Deve-se notar que o estado de ionização correspondente a molécula eletricamente neutra que prepondera é aquele no qual há maior possibilidade de resíduos eletricamente carregados que ainda resultam em uma molécula neutra, à exceção da tirosina (Tyr), estarem carregados positiva ou negativamente nesta configuração.

Figura 3.1 - Estrutura primária da insulina suína: sequência de aminoácidos e pontes de dissulfeto. Grupos ionizáveis em negrito e distribuição de carga da configuração neutra correspondente ao estado de ionização preponderante.

Na Figura 3.2, são apresentadas as curvas de fração de moléculas de insulina eletricamente neutras em função do pH na região do ponto isoelétrico. Há neste gráfico duas curvas: uma que é a soma de todas as frações neutras, chamada fração total; e, outra que representa o estado de ionização neutro preponderante da insulina.

Figura 3.2 - Fração de moléculas eletricamente neutras de insulina em função do pH. Dados de pK_a de Kaarsholm; Havelung e Hougaard (1990). Linha contínua, fração total. Linha pontilhada, fração preponderante.

Observa-se pela Figura 3.2 que as frações de moléculas de proteína eletricamente neutras alcançam um ponto máximo exatamente no ponto isoelétrico, ponto este que pode ser definido como pH no qual a carga líquida média das moléculas de proteína é nula. Ou seja, o ponto máximo da fração de moléculas neutras corresponde ao ponto de mínima solubilidade.

Já as Figuras 3.3, 3.4 e 3.5 apresentam o resultado da modelagem para três diferentes temperaturas para a curva de solubilidade da insulina suína em função do pH. Comparam-se o modelo proposto calculado com diferentes valores de pK_A – valores de pK_A isolados; e, valores de pK_A na cadeia – com o modelo proposto por Tashima et al. (2009).

Figura 3.3 - Solubilidade da insulina suína em função do pH a 278,2 K em 10,0 mM NaHCO₃. Dados experimentais (□), Tashima et al., (2009). Modelo proposto com valores de pK_A de Kaarsholm; Havelung e Hougaard (1990), linha contínua; com valores de pK_A de Stryer (1995), linha pontilhada. Modelo de Tashima et al. (2009), linha semi-contínua.

Figura 3.4 - Solubilidade da insulina suína em função do pH a 288,2 K em 10,0 mM NaHCO₃. Dados experimentais (□), Tashima et al., (2009). Modelo proposto com valores de pK_A de Kaarsholm; Havelung e Hougaard (1990), linha contínua; com valores de pK_A de Stryer (1995), linha pontilhada. Modelo de Tashima et al. (2009), linha semi-contínua.

Figura 3.5 - Solubilidade da insulina suína em função do pH a 298,2 K em 10,0 mM NaHCO₃. Dados experimentais (□), Tashima et al., (2009). Modelo proposto com valores de pK_A de Kaarsholm; Havelung e Hougaard (1990), linha contínua; com valores de pK_A de Stryer (1995), linha pontilhada. Modelo de Tashima et al. (2009), linha semi-contínua.

Deve-se notar que a validade do modelo é restrita a região próxima ao ponto isoelétrico; já que para valores de pH distantes deste ponto a curva de fração de moléculas neutras tende a zero, resultando pela Equação (3 – 8) em uma solubilidade infinita que é fisicamente inverossímil.

Para uma análise comparativa mais elaborada entre os modelos, um critério quantitativo é necessário. No entanto, a escolha de um critério é arbitrária e, por isso, um tanto quanto subjetiva, mas mesmo assim este deve permitir uma comparação equitativa entre os modelos. Aqui o critério escolhido é o seguinte:

$$CQ = 100 \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{S_i^{\text{exp}} - S_i^{\text{calc}}}{S_i^{\text{exp}}} \right)^2} \quad (3 - 40)$$

em que N é o número de pontos experimentais a uma dada temperatura.

Da forma como foi definido, o critério quantitativo (CQ) é adimensional e mede o desvio quadrático entre o valor da solubilidade experimental (S_i^{exp}) e o valor da solubilidade calculado por um dado modelo (S_i^{calc}) ponderando este desvio pelo inverso do quadrado do valor experimental da solubilidade. Este critério é interessante no caso em tela, pois dado o formato da curva de solubilidade de proteínas em função do pH, os pontos em valores de pH um pouco mais distantes do ponto isoelétrico – mas ainda dentro da região de validade do modelo – podem ser até uma ordem de grandeza maior do que os pontos próximos ao pl. Para de certa forma normalizar estes desvios é que o critério quantitativo foi definido assim de modo a atribuir menores pesos a valores de solubilidade absolutos maiores.

Como há dados a diferentes temperaturas, pode-se estender o critério quantitativo de modo a avaliar globalmente a eficiência dos modelos na correlação dos dados experimentais. Uma possibilidade de extensão que aqui será utilizada é a seguinte:

$$CQG = 100 \sqrt{\frac{1}{M} \sum_j^M \left[\frac{1}{N_j} \sum_{i=1}^{N_j} \left(\frac{S_{ij}^{\text{exp}} - S_{ij}^{\text{calc}}}{S_{ij}^{\text{exp}}} \right)^2 \right]} \quad (3 - 41)$$

em que M é número total de conjuntos de dados e j refere-se a um dado conjunto de dados a uma dada temperatura.

A Tabela 3.3 apresenta os valores obtidos do CQ e do CQG para os modelos aplicados aos sistemas contendo insulina.

Tabela 3.3 - Valores de CQ [Equação (3 – 40)] e CQG [Equação (3 – 41)] entre os modelos aplicados e os dados experimentais de solubilidade de insulina suína a diferentes temperaturas.

<i>T</i> / K	Modelo proposto com valores de pK_A de Kaarsholm; Havelung e Hougaard (1990)	Modelo proposto com valores de pK_A de Stryer (1995).	Modelo de Tashima et al. (2009)
278,2	25,1 %	32,5 %	26,7 %
288,2	19,8 %	27,8 %	26,9 %
298,2	27,4 %	23,9 %	15,4 %
Global	24,3 %	28,3 %	23,6 %

Pela Tabela 3.3 é possível afirmar que a abordagem aqui proposta é tão eficaz na correlação dos dados experimentais de solubilidade de insulina suína quanto o modelo de Tashima et al. (2009), no entanto, não exigiu nenhum ajuste de parâmetros. Outra importante menção deve ser feita no tangente a escolha do conjunto de dados de pK_A na execução do cálculo: embora os valores de pK_A na cadeia resultem em uma melhor descrição do fenômeno, não há perda significativa de precisão ao se utilizarem os dados de pK_A de aminoácidos isolados. Esse resultado é muito importante, porque raríssimos são os dados completos de pK_A em uma dada cadeia protéica.

Além da insulina suína que foi bastante bem correlacionada pelo modelo proposto, ao menos na região do ponto isoelétrico, há dados na literatura de solubilidade de três tipos de insulina mutante. Este dados foram publicados por Kaarsholm; Havelung e Hougaard (1990) junto aos valores de pK_A ajustados a partir da curva de titulação potenciométrica de cada tipo de insulina mutante.

A Tabela 3.4 apresenta os valores de pK_A para estes três tipos de insulina mutante. Nesta tabela também são apresentados os números de resíduos que se ionizam na forma catiônica, esses números são aqueles que corresponderam ao melhor ajuste da Equação (3 – 32) aos dados de solubilidade das insulinas mutantes. Apesar dos autores (KAARSHOLM; HAVELUNG; HOUGAARD, 1990) terem feito uma discussão a respeito da determinação dos valores de pK_A através do ajuste da curva de titulação potenciométrica de cada proteína mutante, não fica claro qual resíduo específico possui aquele valor de pK_A . Portanto, não é tão trivial especificar o número de resíduos que se ionizam na forma catiônica, até porque as formas mutantes podem ter estruturas terciárias distintas da insulina nativa o que favorece a blindagem de alguns resíduos antes expostos ao solvente.

Tabela 3.4 - Valores de pK_A e de número de resíduo que se ionizam como cátions (n_+) para insulinas mutantes. Dados de Kaarholm; Havelung e Hougaard (1990).

Grupos	B13-Glu→Gln			B9-Ser→Asp			B25-Phe→His		
	N	pK_A	n_+	N	pK_A	n_+	N	pK_A	n_+
1	2	3,40	5	3	3,52	5	3	3,63	6
2	3	4,66		2	4,28		1	3,99	
3	2	6,85		1	5,56		1	4,74	
4	1	7,85		3	6,82		2	6,18	
5	1	9,33		1	7,05		1	7,24	
6	1	9,83		1	8,23		2	7,61	
7	5	11,23		1	9,74		1	8,62	
8				1	10,12		1	10,00	
9				4	11,00		5	11,03	

A Figura 3.6 apresenta os resultados para a insulina mutante B13-Glu→Gln.

Figura 3.6 - Curva de solubilidade da insulina mutante B13-Glu→Gln em função do pH em 0,1 M KCl a 23 °C. Dados experimentais (○), Kaarholm; Havelung e Hougaard (1990). Equação (3 - 32), linha contínua.

As Figuras 3.7 e 3.8 apresentam os resultados para as insulinas mutantes B9-Ser→Asp e B25-Phe→His, respectivamente.

Figura 3.7 - Curva de solubilidade da insulina mutante B9-Ser→Asp em função do pH em 0,1 M KCl a 23 °C. Dados experimentais (○), Kaarsholm; Havelung e Hougaard (1990). Equação (3 - 32), linha contínua.

Figura 3.8 - Curva de solubilidade da insulina mutante B25-Phe→His em função do pH em 0,1 M KCl a 23 °C. Dados experimentais (○), Kaarsholm; Havelung e Hougaard (1990). Equação (3 - 32), linha contínua.

Há uma primeira observação que deve ser feita: a mudança de apenas um resíduo na cadeia protéica já produz efeitos notáveis tais como a mudança do ponto isoelétrico e o formato da curva de solubilidade. Dessarte é possível imaginar que uma pequena mudança da estrutura primária pode causar alterações significativas na estrutura terciária da proteína a ponto de alterar o acesso do solvente a um determinado resíduo, isso fica evidente ao se compararem os valores de pK_A da insulina nativa e suas respectivas mutantes.

Além disso, é importante notar que o modelo aqui proposto foi bastante exitoso tanto na correlação dos dados da insulina suína quanto nos dados das insulinas mutantes. Observam-se porém duas importantes constatações: se por um lado no caso da insulina mutante B13-Glu→Gln o modelo é menos preciso na região acima do ponto isoelétrico, por outro no caso da mutante B25-Phe→His o modelo é menos preciso na região abaixo do ponto isoelétrico. Provavelmente isso se deve aos erros associados a determinação dos valores de pK_A . Um modo de contornar esse problema é ajustar o valor de pK_A de um dos grupos. Aqui isso não será feito, pois não pertence ao escopo deste trabalho o ajuste de curvas, mas sim a proposição de um modelo e a verificação de suas limitações ao contrastá-lo aos dados experimentais.

Há que se explicar que a ordenada dos gráficos expostos nas Figuras 3.6, 3.7 e 3.8 não correspondem propriamente a solubilidade, mas sim a porcentagem de proteína em solução, que é relativa a quantidade de proteína inicialmente colocada no sistema no experimento de determinação de solubilidade. Assim, os patamares observados em valores de pH distantes do ponto isoelétrico e correspondentes a 100 % de proteína em solução não estão relacionados a solubilidade naquele valor de pH, significam apenas que toda a proteína introduzida ao sistema foi solubilizada, mas essa concentração não corresponde a concentração de equilíbrio para essa proteína nesse pH.

3.3.2 Solubilidade da β -lactoglobulina em função do pH e da força iônica

A Tabela 3.5 apresenta os dados de pK_A isolados para a molécula de β -lactoglobulina bovina. E a Figura 3.9 apresenta o resultado da modelagem das

curvas de solubilidade da β -lactoglobulina bovina em função do pH para quatro diferentes valores de concentração de cloreto de sódio. Observa-se que apesar do aumento da força iônica o modelo consegue descrever com grande precisão as curvas de solubilidade. Há que se dizer que os dados de pK_A isolados não reproduzem o valor experimental do ponto isoelétrico; deste modo, ajustou-se o valor do pK_A dos ácidos glutâmicos para efetuar o cálculo da curva de solubilidade. O valor ajustado foi de 5,14. A diferença entre o valor relatado na literatura e o ajustado é da mesma ordem de magnitude que as diferenças entre os valores de pK_A de aminoácidos isolados e na cadeia de insulina.

Tabela 3.5 - Valores de pK_A dos grupos ionizáveis da β -lactoglobulina bovina.

Aminoácido	Grupo	Número	$pK_A(\text{isolado})^a$
Leu	-NH ₂ (terminal)	1	9,60
Glu	γ -COOH	16	4,25
Tyr	Fenol-OH	4	10,07
Cys	β -SH	1	8,18
His	Imidazol-NH ₂	2	6,00
Arg	Guanidina-NH ₂	3	12,48
Asp	β -COOH	11	3,65
Lys	ϵ -NH ₂	15	10,53
Ile	-COOH (terminal)	1	2,36

^a Cox e Nelson (2003).

Pela Figura 3.9 é possível observar que o modelo descrito pela Equação (3 – 32) foi bastante eficiente na representação das curvas de solubilidade da β -lactoglobulina nas diferentes concentrações de cloreto de sódio.

Há que se mencionar que as curvas apresentadas na Figura 3.9 representam também um fenômeno aqui ainda não discutido e que será objeto de um breve estudo na capítulo cinco: o efeito da concentração de sal na solubilidade da proteína. Claramente, no caso destas curvas vê-se que o aumento da concentração de sal implicou no aumento da solubilidade da proteína, logo estas curvas se localizam na região se *salting-in* do diagrama de fases da β -lactoglobulina.

Figura 3.9 - Solubilidade da β -lactoglobulina bovina em solução aquosa contendo NaCl a 298,2 K. Dados experimentais (Grönwall, 1941): 0,001 mol L⁻¹ NaCl (◆); 0,005 mol L⁻¹ NaCl (○); 0,01 mol L⁻¹ NaCl (▲); 0,02 mol L⁻¹ NaCl (□). Modelagem: linha contínua, 0,001 mol L⁻¹; linha pontilhada, 0,005 mol L⁻¹; linha semi-contínua, 0,01 mol L⁻¹; linha com travessões, 0,02 mol L⁻¹.

Outra importante observação é que o incremento do agente precipitante, o cloreto de sódio neste caso, não alterou nem o ponto isoelétrico nem o formato da curva de solubilidade significativamente indicando que sua atuação reside apenas na mudança da interação entre a molécula de proteína e o solvente, mas não em um equilíbrio químico entre algum resíduo da cadeia protéica e um íon salino. Assim, no cálculo da Equação (3 – 32) apenas a solubilidade do ponto isoelétrico é alterada.

4 Não-idealidade de soluções de aminoácidos e compostos β -lactâmicos

“Erst kommt das Fressen, dann die Moral.”

(Bertolt Brecht, *in Dreigroschenoper*)

4.1 Introdução

Ao se modelar as curvas de solubilidade de aminoácidos ou compostos β -lactâmicos em função do pH, o aspecto mais importante é a modelagem da fase líquida, já que a fase sólida pode ser considerada única e constituída apenas por moléculas de soluto. Uma questão específica envolvendo aminoácidos é devido ao fato de eles se dissociarem em solução aquosa, de modo que um único aminoácido pode ser encontrado em três diferentes estados: como um cátion, como um ânion e como um zwitterion (a espécie neutra não-polar é virtualmente inexistente em solução). Dada a diferença de ordens de grandeza entre a solubilidade no ponto isoelétrico e a solubilidade em outros valores de pH, pode ser necessária em certos sistemas a consideração de não-idealidade da solução. Muitas estratégias têm sido propostas para calcular os coeficientes de atividade de aminoácidos na fase líquida: através da equação de Poisson-Boltzmann (KIRKWOOD, 1934), através de modelos de energia livre de Gibbs excedente (NASS, 1998; CHEN; ZHU; EVANS, 1989; RODRÍGUEZ-RAPOSO; FERNÁNDEZ-MÉRIDA; ESTESO, 1994; KHOSHKBARCHI; VERA, 1996b; PRADHAN; VERA, 1998; XU; PINHO; MACEDO, 2004 e TSENG et al., 2009), através de modelos de contribuição de grupos (GUPTA; HEIDEMANN, 1990 e PINHO; SILVA; MACEDO, 1994), através de abordagens via termodinâmica estatística (KHOSHKBARCHI; VERA, 1996c) e através de equações volumétricas de estado (PARK; YOO; LEE, 2003; FUCHS et al., 2006 e SEYFI et al., 2009).

O primeiro modelo, pelo mais que se sabe, a predizer a curva de solubilidade de um aminoácido em função do pH foi baseado inteiramente em considerações sobre dissociação química (HITCHCOCK, 1924), sem considerar a não-idealidade da fase líquida. Nesse trabalho, a curva de solubilidade da tirosina em solução aquosa em função do pH foi calculada com grande precisão quando comparada aos dados

experimentais. O trabalho de Kirkwood (1934) está entre um dos primeiros dedicados à modelagem de soluções zwitteriônicas. A abordagem foi baseada na solução analítica da equação linearizada de Poisson-Boltzmann considerando as três direções do operador laplaciano para um zwitterion esférico². O autor obteve uma expressão que relaciona o logaritmo do coeficiente de atividade de um íon a uma série matemática; o primeiro termo desta série é análogo à teoria de Debye-Hückel, sendo nulo para zwitterions, porque eles não têm carga. Com algumas simplificações, o autor também mostrou a aplicação desta abordagem para um zwitterion elipsoidal (KIRKWOOD, 1939).

Nass (1988) modelou a solubilidade de aminoácidos em função do pH assumindo que a energia livre de Gibbs excedente é dada pela soma de uma contribuição química e uma física. A contribuição química foi calculada através de considerações sobre a dissociação dos aminoácidos, de equações de balanço material e de condições de equilíbrio de fases, enquanto a contribuição física foi calculada pela equação de Wilson (WILSON, 1964) para a mistura binária de água e aminoácido. Embora os dados experimentais tenham sido bem correlacionados pelo modelo proposto, o número de parâmetros ajustáveis pode ser muito alto – até dez parâmetros foram usados para correlacionar uma única curva de solubilidade. Chen; Zhu e Evans (1989) consideraram que a energia livre de Gibbs excedente pode ser dividida em três termos: dois deles relacionados a interações de longo alcance, *videlicet* um termo de Pitzer-Debye-Hückel e um termo de Born, e um para levar em conta interações de curto alcance, o modelo NRTL. Para o coeficiente de atividade de zwitterions, a contribuição Pitzer-Debye-Hückel é nula. Os autores mostraram que os dados experimentais de solubilidade foram correlacionados satisfatoriamente com apenas um ou dois parâmetros ajustáveis para cada componente.

O trabalho de Gupta e Heidemann (1990) lançou as bases da maioria dos trabalhos que foram publicados *a posteriori*. A equação desenvolvida por eles, que relaciona a dissociação química de aminoácidos e a condição de equilíbrio sólido-líquido, tem sido usada por quase todos os autores para quem esta é a principal questão. A não-idealidade de soluções aquosas de aminoácidos nativos e modificados foi calculada através do modelo UNIFAC (do inglês: *UNiversal Functional Activity Coefficient*) modificado de Larsen; Rasmussen e Fredenslund (1987) com novos grupos tais

² Este é o principal avanço em relação à teoria de Debye-Hückel e apresentada na seção 2.6 porque considera de certa forma a anisotropia das interações.

como glicina e prolina. Seguindo este trabalho, Pinho; Silva e Macedo (1994) propuseram um modelo UNIFAC combinado com um termo de Debye-Hückel para calcular os coeficientes de atividade de aminoácidos; estes autores modelaram a solubilidade da glicina em função do pH em acordo com os dados experimentais. Entretanto, alguns autores questionam a introdução do termo de Debye-Hückel neste caso (KHOSHKBARCHI; VERA, 1996b, 1996c). Rodríguez-Raposo et al. (1994) ajustaram coeficientes osmóticos de glicina em solução aquosa usando a equação de Pitzer e calcularam o coeficiente de atividade da glicina em solução aquosa com excelente precisão quando comparados aos dados experimentais.

Khoshkbarchi e Vera (1996b), para calcular o coeficiente de atividade de aminoácidos em solução contendo um eletrólito, dividiram a energia livre de Gibbs excedente em contribuições de interações de curto e longo alcances. Eles testaram para a contribuição de interações de longo alcance os modelos de Bromley (BROMLEY, 1973) e o deles mesmos (KHOSHKBARCHI; VERA, 1996a), e para o termo de curto alcance o modelo NRTL, cuja formulação está explicitada no capítulo 2, e o modelo de Wilson (WILSON, 1994). Para qualquer modelo escolhido, foram necessários somente dois parâmetros, e foi possível correlacionar com precisão os dados experimentais de sistemas binários água-aminoácido e ternários água-eletrólito-aminoácido. Ainda que a equação básica para a predição da solubilidade de glicina em função do pH usada por Khoshkbarchi e Vera (1996c) siga a abordagem de Gupta e Heidemann (1990), eles introduziram o uso da teoria de perturbação para calcular a não-idealidade de aminoácidos em solução aquosa. Santana et al. (2010) mostraram que este modelo falha na predição quantitativa da solubilidade do antibiótico β -lactâmico ampicilina, da DL-fenilglicina e do ácido 6-aminopenicilânico.

Pradhan e Vera (1998) usaram uma equação similar àquela proposta por Gupta e Heidemann (1990) e aplicaram a equação NRTL para o cálculo da não-idealidade da fase líquida; assim, eles conseguiram calcular a solubilidade da DL-Alanina em função do pH em excelente acordo com os dados observados experimentalmente na região do ponto isoelétrico. Park; Yoo e Lee (2003) correlacionaram o coeficiente de atividade da valina, do ácido α -aminobutírico, da alanina, da glicina e da glicilglicina em solução aquosa pela aplicação da equação de estado de fluidos reticulados por pontes de hidrogênio sem considerar a dissociação química do aminoácido na água. Xu; Pinho e Macedo (2004) propuseram uma modificação no modelo de Wilson para

calcular a energia livre excedente de Gibbs de soluções de aminoácidos. Eles assumiram que todas as moléculas de aminoácidos estão na forma zwitteriônica em água pura, o que de algum modo restringe a aplicação do modelo deles ao cálculo da solubilidade de aminoácidos em valores de pH na região próxima ao ponto isoelétrico.

Usando a equação de estado PC-SAFT (do inglês: *Perturbed-Chain Statistical Associating Fluid Theory*) e levando em conta a dissociação de aminoácidos em solução, Fuchs et al. (2006) descreveram as curvas de solubilidade em função do pH para a glicina, para a DL-Alanina e para a DL-Metionina em solução aquosa com boa precisão quando comparadas aos dados experimentais. Do mesmo modo, Seyfi et al. (2009), seguindo Gupta e Heidemann (1990) e aplicando a equação de estado SAFT, predisseram a curva de solubilidade da DL-Metionina em função do pH e da temperatura com grande acurácia. Tseng et al. (2009) investigaram a solubilidade da DL-Alanina, L-Leucina, L-Isoleucina, L-Serina e DL-Fenilalanina em função do pH em solução aquosa. Considerando a dissociação de aminoácidos em solução aquosa e usando o modelo NRTL, eles obtiveram resultados satisfatórios com dois parâmetros ajustados. Em suma, a revisão da literatura mostra que enquanto os modelos diferem consideravelmente em suas formulações do modelo do coeficiente de atividade, o cálculo do equilíbrio ácido-base permanece uma etapa crucial na descrição da curva de solubilidade. A Tabela 4.1 apresenta um quadro-resumo no intuito de mapear as diferentes estratégias para o cálculo da não-idealidade de soluções contendo aminoácidos.

Tabela 4.1 - Quadro-resumo de trabalhos sobre a não-idealidade de soluções de aminoácidos.

Equação de Poisson-Boltzmann	Energia Livre de Gibbs Excedente	Contribuição de Grupos	Termodinâmica Estatística
Kirkwood (1934)	Nass (1988)	Gupta e Heidemann (1990)	Khoshkbarchi e Vera (1996b)
Kirkwood (1939)	Chen; Zhu e Evans (1989)	Pinho; Silva e Macedo (1994)	Park; Yoo e Lee (2003)
	Rodríguez-Raposo; Fernández-Mérida e Esteso (1994)		Fuchs et al. (2006)
	Khoshkbarchi e Vera (1996a)		Seyfi et al. (2009)
	Pradhan e Vera (1998)		
	Xu; Pinho e Macedo (2004)		
	Tseng et al. (2009)		

4.2 Desenvolvimento teórico

Do mesmo modo que foi feito para proteínas no capítulo anterior, para aminoácidos ao longo da curva de solubilidade vale a afirmação de que, com a variação do pH da solução, não se altera o potencial químico da fase sólida e, pelo critério de equilíbrio, permanece inalterado o potencial químico das moléculas neutras de aminoácido na fase líquida. Supondo que a variação do potencial químico de referência desta espécie na fase líquida seja desprezível, a atividade das moléculas eletricamente neutras de aminoácido deve ser constante com o pH:

$$a_{A^0}(pH) = \text{constante} \quad (4 - 1)$$

Assim, pode-se escrever que:

$$\ln \frac{S_A(pH)}{S_A(pI)} = \ln \frac{\phi_{A^0}(pI)}{\phi_{A^0}(pH)} + \ln \gamma_{A^0}(pI) - \ln \gamma_{A^0}(pH) \quad (4 - 2)$$

em que S_A é a solubilidade do aminoácido, ϕ_A é a fração de aminoácido eletricamente neutro e γ_A é o coeficiente de atividade do aminoácido eletricamente neutro.

O cálculo de ϕ_A é exatamente análogo ao cálculo da fração de moléculas de proteína eletricamente neutras exposto na subseção 3.2.1.

4.2.1 Cálculo das frações eletricamente neutras e carregadas

Considerando-se os seguintes equilíbrios químicos de ionização de um aminoácido cujo grupo R seja apolar (na seção de resultados e discussões, seção 4.3, apresentar-se-á uma formulação para aminoácidos cujo grupo R seja ionizável) em solução aquosa:

e, suas respectivas constantes de equilíbrio:

$$K_{A1} = \frac{[NH_2RCOO^-][H^+]}{[NH_3^+RCOO^-]} \quad (4-5)$$

$$K_{A2} = \frac{[NH_3^+RCOO^-][H^+]}{[NH_3^+RCOOH]} \quad (4-6)$$

Sendo que as frações das formas ionizadas dos aminoácidos são dadas pela equação de Henderson-Hasselbach:

$$\alpha_{-1}(pH) = \frac{10^{pH-pK_{A1}}}{1 + 10^{pH-pK_{A1}}} \quad (4-7)$$

$$\alpha_{+1}(pH) = \frac{1}{1 + 10^{pH-pK_{A2}}} \quad (4-8)$$

Assim, tem-se que as frações de cada uma das espécies de aminoácido em solução bem como suas respectivas cargas líquidas c são dadas por:

$$\phi_{-1}(pH) = \alpha_{-1}(pH)[1 - \alpha_{+1}(pH)], \quad c = -1 \quad (4-9)$$

$$\phi_{+1}(pH) = [1 - \alpha_{-1}(pH)]\alpha_{+1}(pH), \quad c = +1 \quad (4-10)$$

$$\phi_0(pH) = [1 - \alpha_{-1}(pH)][1 - \alpha_{+1}(pH)], \quad c = 0 \quad (4-11)$$

$$\phi_{zwitt}(pH) = \alpha_{-1}(pH)\alpha_{+1}(pH), \quad c = 0 \quad (4-12)$$

Uma observação importante é que, assim como para proteínas, a razão entre duas frações de configurações de moléculas eletricamente neutras distintas é invariável com pH, logo é indiferente qual configuração de moléculas eletricamente neutras se utiliza na Equação (4-2). Obviamente, a forma zwitteriônica é muito mais comum, e sua utilização evita problemas numéricos. Para ilustrar isto, considere-se um aminoácido hipotético para o qual os valores de pK_{A1} , pK_{A2} e pI sejam respectivamente: 2,35; 9,87 e 6,11. Para este aminoácido, no ponto isoelétrico, a fração de moléculas na forma zwitteriônica será de 0,9997 enquanto a fração de moléculas neutras na forma não-zwitteriônica será de $3,019 \times 10^{-8}$. Ou seja, ao se executar o cálculo da solubilidade com a forma neutra não-zwitteriônica, trabalhar-se-á com números em ordens de grandeza inferiores a 10^{-8} . Além disso, este cálculo feito ilustra também porque a forma neutra não-zwitteriônica é virtualmente inexistente em solução.

4.2.2 Modelo de Pitzer

Usualmente o modelo de Pitzer para soluções eletrolíticas é empregado para soluções salinas nas quais a não-idealidade depende das concentrações e das cargas dos íons dissolvidos. No entanto, do modo como, tradicionalmente, este modelo de energia de livre de Gibbs proposto por Pitzer é formulado (vide seção 2.6), não é possível aplicá-lo a sistemas de polieletrólitos. Para contemplar estes sistemas, uma nova formulação do modelo foi desenvolvida por Pessoa Filho e Maurer (2008). Esta nova formulação é mais adequada para aplicação a aminoácidos cuja forma eletricamente neutra predominante é a forma zwitteriônica. Assim, tem-se que para um soluto i em solução aquosa eletrolítica, o logaritmo natural do coeficiente de atividade é dado pela soma de um termo para interações de longo alcance (LR, do inglês *long range*) e um termo para interações de curto alcance (SR, do inglês *short range*):

$$\ln \gamma_i = \ln \gamma_i^{LR} + \ln \gamma_i^{SR} \quad (4 - 13)$$

Com o termo de longo alcance dado por uma extensão ao modelo de Debye-Hückel apresentado no capítulo segundo:

$$\ln \gamma_i^{LR} = -A_\phi z_i^2 \left[\frac{2}{b} \ln(1 + b\sqrt{I}) + \frac{\sqrt{I}}{1 + b\sqrt{I}} \right] \quad (4 - 14)$$

Sendo:

$$A_\phi = \frac{1}{3} \left(\frac{e^2}{\varepsilon_0 \varepsilon_r k_B T} \right)^{3/2} \frac{(2d_s N_A)^{1/2}}{8\pi} \quad (4 - 15)$$

Na qual ε_0 é a permissividade do vácuo ($8,8541878 \times 10^{-12} \text{ A}^2 \text{ s}^4 \text{ kg}^{-1} \text{ m}^{-3}$), ε_r é a constante dielétrica, d_s é a densidade do solvente, N_A é o número de Avogadro ($6,0221415 \times 10^{23}$), e é a carga eletrônica ($1,60217646 \times 10^{-19} \text{ A s}$), k_B é a constante de Boltzmann ($1,3806503 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1}$), z_i é a carga da espécie i e b é um parâmetro universal igual a $1,2 \text{ kg}^{1/2} \text{ mol}^{-1/2}$ (PRAUSNITZ; LICHTENTHALER; AZEVEDO, 1999).

E o termo de curto alcance escrito em sua formulação para polieletrólitos (PESSÔA FILHO; MAURER, 2008) dado por:

$$\ln \gamma_i^{SR} = 2 \sum_{j \neq s} \lambda_{ij}(I) m_j + 3 \sum_{j \neq s} \sum_{k \neq s} \Lambda_{ijk} m_j m_k + \dots$$

$$\dots - z_i^2 M_s \sum_{j \neq s} \sum_{k \neq s} \lambda_{ij}^1 \frac{1}{a^2 I^2} \left[1 - \left(1 + a \sqrt{I} + \frac{a^2 I}{2} \right) \exp(-a \sqrt{I}) \right] m_j m_k \quad (4 - 16)$$

No qual M_s é a massa molar do solvente em kg.mol^{-1} , a é outro parâmetro universal igual a $2,0 \text{ kg}^{1/2}\text{mol}^{-1/2}$, Λ_{ijk} é o parâmetro devido às interações três-a-três, e λ_{ij} é o parâmetro devido às interações por pares que depende da força iônica:

$$\lambda_{ij}(I) = \lambda_{ij}^{(0)} + \lambda_{ij}^{(1)} \frac{2}{a^2 I^2} \left[1 - \left(1 + a \sqrt{I} \right) \exp(-a \sqrt{I}) \right] \quad (4 - 17)$$

Com a força iônica definida como:

$$I = \frac{1}{2} \sum_i m_i z_i^2 \quad (4 - 18)$$

No caso de aminoácidos, quatro espécies podem ser encontradas na solução, duas formas neutras sendo uma forma zwitteriônica; e duas formas não neutras, uma carregada positivamente e outra negativamente. Para a descrição do equilíbrio de fases, apenas uma das formas neutras é necessária para o cálculo. Assim, arbitrar-se-á a forma zwitteriônica como configuração neutra para o cálculo do equilíbrio de fases. Tendo a molécula zwitteriônica carga líquida nula, o termo de longo alcance é nulo e apenas o termo de curto alcance entra no cômputo da não-idealidade. De fato, como observado por Scatchard e Kirkwood (1932), para aminoácidos alifáticos não foi encontrada nenhuma dependência entre o logaritmo natural do coeficiente de atividade e a raiz quadrada da força iônica. O próprio modelo de Kirkwood (1934), que propõe uma solução analítica para a equação linearizada de Poisson-Boltzmann em coordenadas esféricas, tem seu primeiro termo para o potencial de interação da série obtida idêntico ao modelo Debye-Hückel e nulo como consequência da carga líquida nula do zwitterion. Isto corrobora com o fato de aqui o termo de longo alcance do modelo de Pitzer, que nada mais é do que uma extensão ao modelo de Debye-Hückel (PITZER, 1973), ser nulo.

Além disso, na tentativa de diminuir os parâmetros ajustáveis, considerar-se-ão apenas as interações binárias e não se considerará a dependência do parâmetro λ_{ij} com a força iônica, mas apenas sua variação com a temperatura. Adotar-se-á a equação proposta por Pitzer e Peiper (1984) com o menor número de parâmetros

possível (na seção seguinte, dar-se-á uma justificativa teórica para a forma desta equação através de termodinâmica estatística):

$$\lambda_{ij}(T) = C1 + \frac{C2}{T} \quad (4 - 19)$$

Assim, para aminoácidos, pode-se reescrever a Equação (4 – 2) utilizando o modelo de Pitzer para calcular a não-idealidade devido a presença de cargas elétricas em solução como:

$$\begin{aligned} \ln \frac{S_A(pH)}{S_A(pI)} = & \ln \frac{\phi_{zwitt}(pI)}{\phi_{zwitt}(pH)} + 2\lambda_{-1}[\phi_{-1}(pI)S_A(pI) - \phi_{-1}(pI)S_A(pH)] + \dots \\ & \dots + 2\lambda_{+1}[\phi_{+1}(pI)S_A(pI) - \phi_{+1}(pI)S_A(pH)] + 2\lambda_0[\phi_{(pI)}S_A(pI) - \phi_0(pI)S_A(pH)] + \dots \\ & \dots + 2\lambda_{zwitt}[\phi_{zwitt}(pI)S_A(pI) - \phi_{zwitt}(pI)S_A(pH)] \end{aligned} \quad (4 - 20)$$

Sendo a curva experimental de solubilidade de um aminoácido em função do pH simétrica em relação ao ponto isoelétrico é possível inferir que as interações entre as moléculas neutras deste aminoácido e as moléculas carregadas são indistinguíveis, ou seja, a molécula eletricamente neutra de aminoácido interage da mesma forma com a molécula carregada positiva ou negativamente. Esta hipótese implica na seguinte igualdade *a priori*:

$$\lambda_{-1} = \lambda_{+1} = \lambda_0 = \lambda_{zwitt} = \lambda \quad (4 - 21)$$

É evidente que se tal simetria não for observada na curva experimental esta hipótese perde sua validade. Assim, a equação do modelo pode ser escrita como:

$$\ln \frac{S_A(pH)}{S_A(pI)} = \ln \frac{\phi_{zwitt}(pI)}{\phi_{zwitt}(pH)} + 2\lambda[S_A(pI) - S_A(pH)] \quad (4 - 22)$$

Uma deficiência deste modelo poderia ser encontrada no seguinte fato: a Equação (4 – 22) pode ser obtida da mesma forma se a configuração neutra escolhida não fosse a zwitteriônica, isso significa que a interação entre as moléculas carregadas e as moléculas neutras são iguais ainda que a molécula na forma zwitteriônica se comporte como uma molécula altamente polar ao contrário da outra configuração neutra. Mas apesar desta aparente inconsistência, o fato é que para qualquer efeito prático a fração de moléculas neutras na configuração não-zwitteriônica é nula.

Pela direta substituição das Equações (4 – 7) a (4 – 10) e (4 – 12) na Equação (4 – 22), tem-se uma relação explícita entre a solubilidade de aminoácidos e o pH

apenas sendo necessárias informações da estrutura primária, a solubilidade no ponto isoelétrico e o parâmetro de interação:

$$\log \frac{S_A(pH)}{S_A(pI)} = pI - pH + \log \left[\frac{(1 + 10^{pH - pK_{A1}})}{(1 + 10^{pI - pK_{A1}})} \right] + \log \left[\frac{(1 + 10^{pH - pK_{A2}})}{(1 + 10^{pI - pK_{A2}})} \right] + \dots$$

$$\dots + \frac{2}{\ln 10} \lambda [S_A(pI) - S_A(pH)] \quad (4 - 23)$$

A exceção do termo de não-idealidade a Equação (4 – 23) é um caso particular da Equação (3 – 32) em que o número de resíduos ionizáveis é dois. Poder-se-ia também incluir este mesmo termo de não-idealidade a Equação (3 – 32) assumindo as mesmas simplificações, apesar disso não parecer necessário ao menos nos sistemas estudados no capítulo anterior.

4.2.3 Uma abordagem por termodinâmica estatística da não-idealidade

A epígrafe que está no *caput* deste capítulo não faz sentido senão com esta seção que é intencionada a aprofundar o estudo das interações que até agora foram resumidas em um único parâmetro de ajuste sem uma interpretação física mais profunda. A frase de Brecht (a grosso modo traduzida como: “primeiro vem a comida, depois a moral”) resume o espírito desta seção, findado o trabalho propriamente de engenharia química, o trabalho prático de desenvolver um método de cálculo de uma propriedade de interesse que está dado pela Equação (4 – 22) - ou em sua outra representação (4 – 23) - pode-se agora dar-se ao trabalho de compreender melhor as interações responsáveis pela não-idealidade de aminoácidos em solução. Far-se-á isso empregando uma abordagem que até o presente momento foi ignorada neste texto: a termodinâmica estatística.

Pela teoria de gases imperfeitos, é possível escrever o logaritmo do coeficiente de atividade como uma expansão virial desde que o potencial de interação possa ser dividido em contribuições de interações dois-a-dois, três-a-três e assim por diante. Do mesmo modo, pela teoria de McMillan-Mayer é possível emular do ponto de vista da modelagem termodinâmica a situação dada pela não-idealidade de gases para a situação de um soluto em um solvente líquido (HILL, 1986). No Apêndice C estão

apresentados argumentos de um modo mais pormenorizado e rigoroso desta teoria. Aqui, assumir-se-á que isso é possível de tal modo que o logaritmo natural do coeficiente de atividade da molécula zwitteriônica seja escrito como:

$$\ln \gamma_{zwitt} = 2 \sum_j \rho_j B_{j,zwitt} \quad (4 - 24)$$

em que $B_{j,zwitt}$ é o segundo coeficiente virial osmótico e ρ_j é a número de moléculas do componente j por unidade de volume.

Sabendo que o segundo coeficiente virial osmótico pode ser relacionado ao potencial da força média de interação entre os solutos no dado solvente no limite de baixas concentrações de soluto, tem-se que (NEAL et al., 1999):

$$B_{ij} = -\frac{1}{16\pi^2} \int_0^{2\pi} \int_0^\pi \int_0^{2\pi} \int_0^{2\pi} \int_0^\pi \int_0^{+\infty} \left[e^{\left(\frac{-W_{ij}(r)}{k_B T} \right)} - 1 \right] r^2 \sin \varphi \sin \beta \, dr \, d\varphi \, d\theta \, d\alpha \, d\beta \, d\gamma \quad (4 - 25)$$

Considerando que o potencial de força média seja dado pela contribuição de três potenciais: o de esferas rígidas (W^{HS}), o potencial atrativo devido às forças de van der Waals (W^{vdW}) e o potencial devido às interações eletrostáticas (W^{elet}) tal como a teoria DLVO (Derjagun, Landau, Verwey e Overbeeck) sugere:

$$W_{ij}(r) = W^{HS}(r) + W^{vdW}(r) + W^{elet}(r) \quad (4 - 26)$$

Pode-se então escrever este potencial do modo mais simples como:

$$W_{ij}(r) = \begin{cases} +\infty, & r \leq \sigma \\ \frac{z_i z_j e^2 e^{-\kappa(r-\sigma)}}{\epsilon r (1 + \kappa \sigma)} - \epsilon \left(\frac{\sigma}{r} \right)^6, & r \geq \sigma \end{cases} \quad (4 - 27)$$

em que z_i é a carga da molécula i , z_j é a carga da molécula j , e é a carga eletrônica, σ é o raio de van der Waals médio das moléculas i e j , ϵ é o parâmetro de interação entre as moléculas i e j .

Considerando aqui que qualquer configuração de uma molécula de aminoácido possui o mesmo raio de van der Waals, σ , hipótese razoável já que a máxima diferença de estrutura entre elas é de um próton. Assumir-se-á também que o valor do parâmetro de interação de van der Waals entre a forma zwitteriônica e as demais formas é o mesmo, o que é razoável para moléculas de estrutura química tão semelhante como as formas ionizadas de um aminoácido.

No caso do soluto i ser o zwitterion, o termo eletrostático é nulo e o segundo coeficiente virial osmótico para um sistema de interações isotrópicas fica sendo

apenas:

$$B_{ij} = -2\pi \left\{ \int_0^\sigma [e^{-\infty} - 1] r^2 dr + \int_\sigma^{+\infty} \left[e^{\frac{\varepsilon}{k_B T} \left(\frac{\sigma}{r} \right)^6} - 1 \right] r^2 dr \right\} \quad (4 - 28)$$

Considerando a expansão em série de Taylor truncada para o segundo termo exponencial:

$$e^{\frac{\varepsilon}{k_B T} \left(\frac{\sigma}{r} \right)^6} \approx 1 + \frac{\varepsilon}{k_B T} \left(\frac{\sigma}{r} \right)^6 \quad (4 - 29)$$

Tem-se, então, que o segundo coeficiente virial osmótico é dado por:

$$B_{ij} = 2\pi \left\{ \int_0^\sigma r^2 dr - \frac{\varepsilon \sigma^6}{k_B T} \int_\sigma^{+\infty} \frac{1}{r^4} dr \right\} = \frac{2\pi \sigma^3}{3} \left(1 - \frac{\varepsilon}{k_B T} \right) \quad (4 - 30)$$

Assim, o logaritmo do coeficiente de atividade do zwitterion pode ser escrito como:

$$\ln \gamma_{zwitt} = \frac{4\pi \sigma^3}{3} \left(1 - \frac{\varepsilon}{k_B T} \right) \sum_j \rho_j \quad (4 - 31)$$

Substituindo a Equação (4 - 31) na Equação (4 - 2), para solubilidade dada em unidades molais:

$$\ln \frac{S_A(pH)}{S_A(pI)} = \ln \frac{\phi_{zwitt}(pI)}{\phi_{zwitt}(pH)} + \frac{4\pi \sigma^3 N_A \rho}{3} \left(1 - \frac{\varepsilon}{k_B T} \right) [S_A(pI) - S_A(pH)] \quad (4 - 32)$$

A semelhança entre as Equações (4 - 32) e (4 - 22) é notável e torna a aplicação do modelo mais consistente. A partir desta comparação, as constantes de ajuste do parâmetro de interação do modelo de Pitzer apresentadas na Equação (4 - 19) ganham um significado físico mais forte:

$$CI = \frac{2\pi \sigma^3 N_A \rho}{3} \quad (4 - 33)$$

$$C2 = -CI \frac{\varepsilon}{k_B T} \quad (4 - 34)$$

4.3 Resultados e Discussões

4.3.1 Solubilidade de aminoácidos em função do pH

A Equação (4 – 23) por ser uma equação transcendental exige na sua aplicação uma solução numérica. Como a derivada analítica desta equação é de simples obtenção, foi utilizado o método de Newton-Raphson para o cálculo da solubilidade em um dado pH pela Equação (4 – 23). Além disso, o parâmetro de interação λ deve ser ajustado ao conjunto de dados experimentais. Para tanto utilizou-se o método de Nelder e Mead (1965) que apesar de ser um método para otimização de funções multivariáveis reduz-se a um simples método de busca unidirecional no caso de uma função monovariável. A função objetivo considerada foi o próprio critério quantitativo expresso na Equação (3 – 40). A Tabela 4.2 apresenta o resultado do ajuste do parâmetro de interação para os sistemas de aminoácidos estudados. Os dados de pK_A dos aminoácidos estudados neste capítulo estão tabelados no Anexo A.

Tabela 4.2 - Valores do parâmetro de interação λ e do CQ [Equação (3 – 40)] para diversos conjuntos de dados experimentais de solubilidade de aminoácidos em função do pH.

Aminoácidos	$\lambda / \text{kg mol}^{-1}$	CQ	Ref. Dados Exp.
DL-Alanina	0,0267	2,0 %	Tseng et al. (2009)
DL-Metionina	0,0935	6,8 %	Fuchs et al. (2006)
L-Isoleucina	0,234	5,2 %	Tseng et al. (2009)
L-Leucina	0,273	9,4 %	Tseng et al. (2009)
DL-Fenilalanina	3,95	2,9 %	Tseng et al. (2009)
L-Serina	0,0165	4,6 %	Tseng et al. (2009)

As Figuras 4.1, 4.2, 4.3, 4.4, 4.5 e 4.6 apresentam os resultados da aplicação de tais modelos aos sistemas de DL-Alanina, DL-Metionina, L-Isoleucina, L-Leucina, DL-Fenilalanina e L-Serina, respectivamente.

Figura 4.1 - Valores de solubilidade da DL-Alanina a 298,15 K em solução aquosa. Dados experimentais (o), Tseng et al. (2009); solução com NaOH ou HCl (●), Pradhan e Vera (1998); solução com KOH ou HNO₃ (□), Pradhan e Vera (1998). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Figura 4.2 - Valores de solubilidade de DL-Metionina a 303,0 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Fuchs et al. (2006). Modelagem: solução ideal, linha pontilhada; modelo proposto, linha contínua.

Figura 4.3 - Valores de solubilidade da L-Isoleucina a 298,15 K em solução aquosa (com HCl ou NaOH). Dados experimentais (o), Tseng et al. (2009). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Figura 4.4 - Valores de solubilidade da L-Leucina a 298,15 K em solução aquosa (com HCl ou NaOH). Dados experimentais (o), Tseng et al. (2009). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Figura 4.5 - Valores de solubilidade da DL-Fenilalanina a 298,15 K em solução aquosa (com HCl ou NaOH). Dados experimentais (o), Tseng et al. (2009). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Figura 4.6 - Valores de solubilidade da L-Serina 298,15 K em solução aquosa (com HCl ou NaOH). Dados experimentais (o), Tseng et al. (2009). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Pelas Figuras 4.1, 4.2, 4.3, 4.4, 4.5 e 4.6 é possível afirmar que a aplicação do modelo proposto é uma abordagem profícua no tratamento termodinâmico do equilíbrio sólido-líquido destes sistemas dado o sucesso na correlação de dados experimentais de diversos aminoácidos. Também é possível observar que, apesar da vasta literatura dedicada à modelagem da não-idealidade de soluções contendo aminoácidos, por vezes a consideração de solução ideal é suficiente para uma descrição razoavelmente precisa da curva de solubilidade de aminoácidos.

Para uma comparação quantitativa da aplicação dos modelos, a Equação (3 – 40) foi usada como um critério de validação. Além do modelo proposto e do modelo considerando solução ideal, também foram comparados os resultados de Tseng et al. (2009) que aplicaram o modelo NRTL (vide seção 2.5) para calcular a não-idealidade considerando a interação entre as moléculas de aminoácidos – DL-Alanina, L-Isoleucina, L-Leucina, DL-Fenilalanina e L-Serina – e as moléculas do solvente, no caso água.

A Tabela 4.3 apresenta os valores do critério quantitativo (CQ) dos três modelos avaliados.

Tabela 4.3 - Valores de CQ [Equação (3 – 40)] em porcentagem para os diferentes modelos aplicados aos conjuntos de dados de solubilidade de DL-Alanina, L-Isoleucina, L-Leucina, DL-Fenilalanina e L-Serina a 298,15 K. Dados experimentais de Tseng et al. (2009).

Aminoácido	Modelo proposto considerando solução ideal ($\lambda = 0$)	Modelo proposto com valores de λ da Tabela 4.2	Modelo NRTL (TSENG et al., 2009)
DL-Alanina	2,4 %	2,0 %	2,2 %
L-Isoleucina	24,9 %	5,2 %	6,3 %
L-Leucina	32,3 %	9,4 %	13,3 %
DL-Fenilalanina	62,5 %	2,9 %	3,0 %
L-Serina	6,0 %	4,6 %	5,1 %

A Tabela 4.3 mostra claramente que o modelo proposto na Equação (4 – 23) na sua forma completa, isto é, incluindo o termo de não-idealidade conseguiu correlacionar os dados experimentais de solubilidade destes aminoácidos com grande precisão. Uma ressalva que deve ser feita é que apesar dos desvios do modelo NRTL utilizado por Tseng et al. (2009) serem apenas ligeiramente superiores, este modelo exige o ajuste de no mínimo dois parâmetros (considerando fixo o fator de não-aleatoriedade na mistura, $\alpha_{ij} = 0,3$).

É importante notar também que tanto para a DL-Alanina quanto para a L-Serina a consideração de solução ideal é uma hipótese razoável e que traz a vantagem de não necessitar de nenhum ajuste de parâmetros.

Interessante notar que o desvio em relação a consideração de solução ideal aumenta com o aumento da massa molecular destes aminoácidos. Tendo por base a interpretação dada ao parâmetro de interação de Pitzer na seção 4.2.3, pode-se inferir que o aumento da não-idealidade está diretamente ligado ao aumento do efeito de “volume excluído”. No *ensemble* de McMillan-Mayer, o termo referente a idealidade, assim como no gás ideal, considera partículas pontuais que não interagem. O modelo de interação de van der Waals inclui um termo que considera um volume molecular (no fundo, além do termo atrativo de van der Waals, há um termo de esferas rígidas que é responsável pelo aparecimento de um termo de volume excluído) e, aqui, parece fazer sentido que quanto mais pesada a molécula, maior seu volume; e, por conseqüência, maior o termo de não-idealidade devido ao volume excluído. Essa observação contribui para a consistência da modelagem aqui desenvolvida. A abordagem mais precisa seria considerar a não-idealidade nos equilíbrios químicos de ionização, já que a constante de equilíbrio é o produto das atividades dos componentes elevadas aos correspondentes coeficientes estequiométricos. Assim, uma sutileza até agora não mencionada é que ao se adotarem as equações de Henderson-Hasselbach se está considerando solução ideal. Logo, a inclusão da não-idealidade *a posteriori* é um modo menos rigoroso do ponto de vista do formalismo termodinâmico, mas ainda assim válido. Essa estratégia tem sido usada amplamente em modelagem de equilíbrio de fases nas quais há que se considerar uma série de equilíbrios químicos. Mas isso coloca uma questão importante, parte da informação sobre a não-idealidade do sistema já está nos valores de pK_A . Isso ajuda a compreender por que alguns sistemas quiçá necessitam de um modelo de não-idealidade.

Há ainda que se estender este modelo a aminoácidos que possuam um grupo *R* ionizável (por exemplo: a tirosina, a cisteína, a histidina, o ácido glutâmico, o ácido aspártico, a lisina e a asparagina).

De fato, havendo mais um grupo ionizável, no total de três, segue-se pela Equação (3 – 9) que haverá oito formas distintas, em termos de distribuição de cargas, deste aminoácido em solução. Portanto, a Equação (4 – 20) teria ao invés de quatro parâmetros de interação, oito. Um parâmetro de interação entre cada forma de

aminoácido e a forma zwitteriônica. Se se adotar por hipótese que estes parâmetros são iguais, é de fácil demonstração que a forma final da equação do modelo seria:

$$\log \frac{S_A(pH)}{S_A(pI)} = pI - pH + \log \left[\frac{(1 + 10^{pH - pK_{A1}})}{(1 + 10^{pI - pK_{A1}})} \right] + \log \left[\frac{(1 + 10^{pH - pK_{A2}})}{(1 + 10^{pI - pK_{A2}})} \right] + \dots$$

$$\dots + \log \left[\frac{(1 + 10^{pH - pK_{A3}})}{(1 + 10^{pI - pK_{A3}})} \right] + \frac{2}{\ln 10} \lambda [S_A(pI) - S_A(pH)] \quad (4 - 35)$$

Os dados de solubilidade de aminoácidos com três resíduos ionizáveis são poucos; aqui, usar-se-ão dados de solubilidade de tirosina publicados a quase um século atrás em um artigo (HITCHCOCK, 1924) que também apresentava, possivelmente, a primeira modelagem para a curva de solubilidade de aminoácidos em função do pH considerando solução ideal. A tirosina apesar de apresentar um terceiro resíduo ionizável aparece geralmente na literatura especializada como um aminoácido apolar. Sendo assim, é interessante comparar qual o efeito de considerar ou não a ionização do grupo *R* da tirosina na modelagem da curva de solubilidade. Neste caso, considerar-se-á apenas a situação de solução ideal, ou seja, $\lambda = 0$. A Figura 4.7 apresenta o resultado desta modelagem.

Figura 4.7 - Valores de solubilidade da tirosina 298,15 K em solução aquosa. Dados experimentais (o), Hitchcock (1924). Modelo considerando o grupo *R* ionizável (linha contínua). Modelo considerando o grupo *R* não-ionizável (linha pontilhada).

Pela Figura 4.7 é possível afirmar que a consideração do equilíbrio químico de ionização do grupo *R* da tirosina melhora a descrição do equilíbrio sólido-líquido. Também observa-se que a hipótese de $\lambda = 0$ é razoável a este sistema.

4.3.2 Solubilidade da glicina e de seus oligopeptídeos em função do pH

Também foram investigados alguns sistemas com peptídeos tendo apenas os resíduos terminais com grupos ionizáveis. A Tabela 4.4 apresenta os valores do parâmetro de interação λ e o valor de CQ – definido na Equação (3 – 40) – para a glicina, diglicina, triglicina, tetraglicina, pentaglicina e hexaglicina. Apesar da glicina ser um aminoácido, para efeito de comparação da sua curva de solubilidade com a de seus oligômeros, apresentam-se seus resultados nesta seção. Os dados de pK_A da glicina e de seus oligômeros estão também relacionados no Anexo A.

Tabela 4.4 - Valores do parâmetro de interação λ e do CQ [Equação (3 - 40)] para diversos conjuntos de dados experimentais de solubilidade de glicina e seus oligopeptídeos em função do pH. Dados experimentais de Lu et al. (2006).

Molécula	$\lambda / \text{kg mol}^{-1}$	CQ
Glicina	0,200	11,1 %
Diglicina	3,26	6,9 %
Triglicina	13,8	9,8 %
Tetraglicina	93,6	1,7 %
Pentaglicina	238,8	4,7 %
Hexaglicina	258,2	7,9 %

A Figura 4.8 traz as curvas de solubilidade calculadas pelo modelo – Equação (4 – 23) – e os dados experimentais para a glicina, a diglicina e a triglicina, e a Figura 4.9 traz os mesmos resultados para os sistemas de tetraglicina, de pentaglicina e de hexaglicina.

Figura 4.8 - Curva de solubilidade em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais de glicina (○), de diglicina (■) e de triglicina (Δ), Lu et al. (2006). Modelagem - Equação (4 - 23) - com valores de λ da Tabela 4.4: glicina, linha contínua; diglicina, linha semi-contínua; triglicina, linha pontilhada.

Figura 4.9 - Curva de solubilidade em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais de tetraglicina (○), de pentaglicina (■) e de hexaglicina (Δ), Lu et al. (2006). Modelagem - Equação (4 - 23) - com valores de λ da Tabela 4.4: tetraglicina, linha contínua; pentaglicina, linha semi-contínua; hexaglicina, linha pontilhada.

Tanto pela observação dos valores de CQ da Tabela 4.4, quanto pela observação das curvas do modelo contrastadas aos dados experimentais de solubilidade apresentadas nas Figuras 4.8 e 4.9, é possível afirmar que o modelo foi exitoso na reprodução da curva de solubilidade em função do pH da glicina e de seus oligopeptídeos.

Interessante notar que a magnitude do parâmetro de interação λ aumenta consistentemente com o aumento da cadeia do peptídeo. Retomando a interpretação via termodinâmica estatística dada ao parâmetro de interação λ , tem-se que:

$$\lambda = \frac{2\pi\sigma^3}{3} N_A \rho \left(1 - \frac{\varepsilon}{k_B T} \right) \quad (4 - 36)$$

De certo modo, o parâmetro ε carrega parte da não-idealidade entre o soluto e o solvente, já que se tratando de um potencial da força média, as interações solventes-solutos estão sob a forma de média no cômputo das interações soluto-soluto. Assim, é de se supor que o aumento da cadeia apolar em um soluto em água cause um aumento do valor de ε , todavia o que se observa é que o valor de λ fica cada vez mais positivo a medida que se aumenta a cadeia apolar do soluto. Assim, provavelmente esse aumento de λ deve estar relacionado ao aumento de σ , ou seja, do efeito de “volume excluído”.

4.3.3 Solubilidade de compostos β -lactâmicos em função do pH, da temperatura e do solvente

Os compostos que contêm grupos β -lactâmicos apresentam uma característica comum aos aminoácidos: um grupo amina e um grupo ácido carboxílico participantes de equilíbrios de protonação. Essa característica comum faz com que a modelagem do equilíbrio sólido-líquido desenvolvida para aminoácidos possa ser aplicada a moléculas que possuem grupos β -lactâmicos.

A Figura 4.10 apresenta a estrutura da ampicilina e a Figura 4.11 a estrutura do ácido 6-aminopenicilânico.

Figura 4.10 - Estrutura da ampicilina.

Figura 4.11 - Estrutura do ácido 6-aminopenicilânico.

Há especial interesse nestas moléculas, pois algumas delas são antibióticos, como a ampicilina, assim há um claro objetivo industrial na produção de tais moléculas. Em particular a ampicilina, que hoje é produzida em um processo muito agressivo ao meio ambiente, ganhou nos últimos anos atenção no que tange a sua produção por uma rota biossintética. Esta rota exige modelos termodinâmicos para o equilíbrio sólido-líquido confiáveis para a otimização do processo reacional no qual o antibiótico deve ser precipitado e retirado do reator (SANTANA et al, 2010).

Aqui serão investigadas algumas moléculas importantes neste processo: a ampicilina, obviamente; e, o ácido 6-aminopenicilânico envolvido na produção de ampicilina. Variáveis importantes para esse processo são: pH, temperatura e concentração de metanol do meio. Apesar de, recentemente, dados de solubilidade nas diversas condições terem sido publicados (SANTANA et al., 2010), a modelagem ainda é algo que precisa ser melhorado.

O modelo usado é o dado pela Equação (4 – 23); entretanto, como há disponibilidade de dados de solubilidade em função do pH e da temperatura, foi possível ajustar dois parâmetros $C1$ e $C2$ que estão relacionados aos parâmetros σ e ε . O ajuste foi realizado através da minimização da função definida pela Equação (3 – 41) pelo método de Nelder e Mead (1965) com duas variáveis a serem

ajustadas. Os dados de pK_A usados no cálculo estão apresentados no Anexo A. O código-fonte do programa de cálculo no ajuste de parâmetros está apresentado no Apêndice D.

A Tabela 4.5 apresenta os valores de σ e ε ajustados bem como os valores de CQG – definido na Equação (3 – 41) – tanto para a ampicilina quanto para o ácido 6-aminopenicilânico.

Tabela 4.5 - Valores dos parâmetros σ , ε e CQG para os sistemas contendo ampicilina e ácido 6-aminopenicilânico. Dados experimentais de Santana et al. (2010).

Molécula	$\sigma / \text{\AA}$	$\varepsilon/k_B / \text{K}$	CQG
Ampicilina	5,25	299,8	8,2 %
Ácido 6-aminopenicilânico	2,68	275,6	12,0 %

As Figuras 4.12, 4.13, 4.14 e 4.15 apresentam as curvas de solubilidade da ampicilina em função do pH às temperaturas de 283,06 K; 288,01 K; 292,95 K e 298,03 K respectivamente. Já as Figuras 4.16, 4.17, 4.18 e 4.19 as curvas de solubilidade do ácido 6-aminopenicilânico às temperaturas de 283,06 K; 288,01 K; 292,95 K e 298,03 K respectivamente.

Figura 4.12 - Valores de solubilidade da ampicilina 283,06 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Figura 4.13 - Valores de solubilidade da ampicilina 288,01 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Figura 4.14 - Valores de solubilidade da ampicilina 292,95 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Figura 4.15 - Valores de solubilidade da ampicilina 298,03 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Figura 4.16 - Valores de solubilidade do ácido 6-aminopenicilânico 283,06 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Figura 4.17 - Valores de solubilidade do ácido 6-aminopenicilânico 288,01 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Figura 4.18 - Valores de solubilidade do ácido 6-aminopenicilânico 292,95 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Figura 4.19 - Valores de solubilidade do ácido 6-aminopenicilânico 298,03 K em solução aquosa (com HCl ou NaOH) em função do pH. Dados experimentais (o), Santana et al. (2010). Modelo Proposto (linha contínua). Modelo Ideal (linha pontilhada).

Diferentemente do que foi observado para os aminoácidos de modo geral, há um grande desvio em relação a idealidade, isto é, a consideração de solução ideal não é apropriada para estes sistemas da ampicilina e do ácido 6-aminopenicilânico principalmente para valores de pH distantes do ponto isoelétrico nos quais a concentração de moléculas de soluto é maior do que no ponto isoelétrico e, portanto, a interação entre elas não pode ser desprezada. De fato, o modelo proposto com a consideração de não-idealidade do modo como exposto na Equação (4 – 23) consegue representar com bastante precisão a curva obtida experimentalmente. No artigo no qual estão publicados estes dados experimentais (SANTANA et al., 2010), a modelagem que foi objeto de aplicação – modelo de Khoshkbarchi e Vera (1996c) – não conseguiu representar os dados senão de modo qualitativo, inclusive no caso dos sistemas contendo ácido 6-aminopenicilânico este modelo tentado violou uma restrição destas curvas: a de que o ponto de mínima solubilidade ocorre no ponto isoelétrico.

Para evitar ajuste de mais parâmetros a variação da concentração de metanol no sistema foi negligenciada no termo de não-idealidade bem como nos valores de pK_A .

Mesmo assim, considerando apenas a variação da solubilidade do ponto isoelétrico com a variação da concentração de metanol no sistema foi possível correlacionar os dados experimentais de solubilidade de ampicilina e do ácido 6-aminopenicilânico com grande acurácia como pode ser visto nas Figuras 4.20 e 4.21.

As Figuras 4.20 e 4.21 apresentam os resultados da modelagem para os sistemas contendo ampicilina e ácido 6-aminopenicilânico, respectivamente, em diferentes concentrações de metanol.

É digno de atenção o fato de que o efeito da presença de metanol no sistema tão logo diminua a constante dielétrica do solvente alterando a magnitude da concentração de soluto em equilíbrio na fase líquida não altera o formato da curva de solubilidade o que coaduna com a hipótese de o termo de não-idealidade e os valores de pK_A permanecerem inalterados.

No entanto, ao assumir tal hipótese deixa-se de se discutir uma questão em aberto na Termodinâmica no *ensemble* de McMillan-Mayer: como devem ser tratadas as misturas de solventes e até que ponto um componente deve ser tratado como co-soluto ou como co-solvente. Não fazendo parte do escopo deste trabalho tratar deste tema, ele é apenas mencionado.

Figura 4.20 - Valores de solubilidade da ampicilina 298,15 K em solução aquosa em função do pH. Dados experimentais a diferentes frações mássicas de metanol: (○) 0,0 %, (■) 15,0 %, (△) 30,0 % e (◆) 50,0 %. Santana; Mattedi e Giordano (2010). Modelo proposto: 0,0 % (linha contínua), 15,0 % (linha com travessões), 30,0 % (linha semi-contínua) e 50,0 % (linha pontilhada).

Figura 4.21 - Valores de solubilidade do ácido 6-aminopenicilânico 298,15 K em solução aquosa em função do pH. Dados experimentais a diferentes frações mássicas de metanol: (o) 0,0 %, (■) 15,0 %, (Δ) 30,0 % e (◆) 50,0 %. Santana; Mattedi e Giordano (2010). Modelo proposto: 0,0 % (linha contínua), 15,0 % (linha com travessões), 30,0 % (linha semi-contínua) e 50,0 % (linha pontilhada).

5 Segundo coeficiente virial osmótico de proteínas e suas relações

"The sciences, not yet split by nineteenth-century academicism into a superior 'pure' and an inferior 'applied' branch, devoted themselves to the solution of productive problems: the most striking advances of the 1780s were those of chemistry, which was by tradition most closely linked to workshop practice and the needs of industry."
(Eric Hobsbawm, in *The Age of Revolution*)

5.1 Introdução

O célebre trabalho de George et al. (1997) revelou uma importante interface entre as teorias termodinâmicas a respeito das forças de interação entre proteínas e os processos de cristalização das mesmas, interface esta embasada na relação entre os valores do segundo coeficiente virial osmótico e as condições de nucleação de diversas proteínas. Esta notável observação da janela de cristalização motivou uma série de pesquisas tanto de cunho experimental na determinação de valores de segundo coeficiente virial osmótico quanto trabalhos de viés fundamentalmente teórico na proposição de modelos de potencial da força média, como Lima et al. (2009), e de equações de estado (MOLLERUP; BREIL, 2009ab) no intuito de calcular o valor deste parâmetro e comparar com os dados obtidos experimentalmente.

Segundo Prausnitz (2003) há pelo menos quatro métodos bem estabelecidos para a determinação experimental do segundo coeficiente virial osmótico: osmometria, espalhamento de luz, medida do ponto de nuvem e fluorescência anisotrópica.

Além destes, Tessier; Lenhoff e Sandler (2002) empregaram a cromatografia de auto-interação para determinar o valor do segundo coeficiente virial osmótico da lisozima e do quimiotripsinogênio; é possível relacionar a retenção relativa de proteínas na coluna cromatográfica com o segundo coeficiente virial osmótico.

Mesmo assim, há poucos dados disponíveis na literatura comparado ao número de proteínas de interesse industrial e a infinitude de condições a que estas proteínas podem ser submetidas. Por essa razão, a relação entre o segundo coeficiente virial

osmótico e a solubilidade de proteína já apontada no trabalho de George et al. (1997) tem atraído especial atenção na proposição de modelagens termodinâmicas que fundamentam esta observação até então empírica.

Há pelo menos três trabalhos publicados que apresentam relações teóricas com distintas formulações: Guo et al. (1999), Haas; Drenth e Wilson (1999) e Ruppert; Sandler e Lenhoff (2001). O primeiro aplica a teoria de McMillan-Mayer e a equação de Gibbs-Duhem a fim de obter uma equação para o potencial químico da proteína em solução e assim relacionar o segundo coeficiente virial osmótico com a solubilidade, ajustando a diferença entre o potencial químico de referência da proteína em solução e o potencial químico da proteína no cristal (GUO et al., 1999). Apesar da relação teórica proposta, nenhuma aplicação concreta foi apresentada. No trabalho de Haas; Drenth e Wilson (1999) foi desenvolvida uma relação entre o segundo coeficiente virial e a solubilidade de proteína partindo da definição do segundo coeficiente virial dada pela termodinâmica estatística. Haas; Drenth e Wilson (1999) fizeram também considerações sobre a anisotropia das interações entre moléculas não-esféricas e aplicaram o potencial de poço quadrático e o potencial de Yukawa para sistemas hipotéticos. Por fim, o trabalho de Ruppert; Sandler e Lenhoff (2001), considerando a lei de Henry e uma proposta de conversão do segundo coeficiente virial osmótico experimental para o segundo coeficiente virial osmótico limite de McMillan-Mayer, chegaram a uma relação entre o segundo coeficiente virial osmótico determinado experimentalmente e a solubilidade de proteína, porém com a necessidade de parâmetros como o incremento do índice de refração. Ruppert; Sandler e Lenhoff (2001) aplicaram esta relação a alguns sistemas de lisozima e de ovalbumina.

Há que se fazer menção também ao trabalho de Allahyarov et al. (2002) que calcularam valores do segundo coeficiente virial em função da concentração de sal para sistemas hipotéticos de proteínas através de dinâmica molecular e concluíram que a dependência entre estes parâmetros não é monotônica como aquela prevista pela teoria de Derjaguin-Landau-Verwey-Overbeek.

No trabalho aqui apresentado, uma nova relação baseada exclusivamente em termodinâmica clássica é proposta. Contudo, o desenvolvimento teórico percorrido permite em primeiro plano uma relação entre o segundo coeficiente virial osmótico e a concentração do agente precipitante, neste caso, sais; a partir da equação de Cohn (COHN, 1925), obtém-se uma relação entre o segundo coeficiente virial

osmótico e a solubilidade de proteína na região de *salting-out* para variações na concentração de sal.

5.2 Desenvolvimento teórico

De fato, como observado por Watanabe (2007), a fase sólida pode conter sais e água além de moléculas de proteína; no entanto, os íons salinos estão de tal forma ligados a resíduos da cadeia protéica que tal complexo íon-água-proteína se comporta como um único componente eletricamente neutro.

Considerando, ainda, que o potencial químico da molécula eletricamente neutra de proteína da fase líquida possa ser escrito, para um processo isotérmico no qual o efeito da pressão sobre o potencial químico seja desprezível, como uma expansão virial truncada após o segundo coeficiente virial osmótico³, tem-se que:

$$\mu_0^L(\mathbf{m}) = \mu_0^*(\mathbf{m}^*) + RT \ln \frac{m_0}{m_0^*} + 2RTB \sum_i m_i = \mu_0^*(\mathbf{m}^*) + RT \ln \frac{m_0}{m_0^*} + 2RTBS \quad (5-1)$$

Derivando a expressão acima em relação a molalidade de sal na fase líquida:

$$\frac{\partial \mu_0^L(\mathbf{m})}{\partial m_{sal}} = \frac{\partial \mu_0^*(\mathbf{m}^*)}{\partial m_{sal}} + RT \frac{\partial \ln m_0}{\partial m_{sal}} + 2RTB \frac{\partial S}{\partial m_{sal}} + 2RTS \frac{\partial B}{\partial m_{sal}} \quad (5-2)$$

em que m_0 é a molalidade de moléculas neutras de proteína na fase líquida e pode ser calculada pelo produto entre a fração de moléculas de proteína eletricamente neutras na fase líquida e a própria solubilidade (S) da proteína conforme a Equação (3 – 6). Substituindo a Equação (3 – 6) na Equação (5 – 2) e com alguma manipulação algébrica, pode-se mostrar que:

$$\frac{1}{SB} \frac{\partial \mu_0^*(\mathbf{m}^*)}{\partial m_{sal}} + \frac{RT}{SB} \left(\frac{\partial \ln \phi_0}{\partial m_{sal}} + \frac{\partial \ln S}{\partial m_{sal}} \right) + 2RT \left(\frac{\partial \ln B}{\partial m_{sal}} + \frac{\partial \ln S}{\partial m_{sal}} \right) = \frac{1}{SB} \frac{\partial \mu_0^L(\mathbf{m})}{\partial m_{sal}} \quad (5-3)$$

Pode-se considerar ainda que o sistema esteja em uma região do diagrama de fases onde não há variação do potencial químico da fase sólida; esta hipótese é coerente

³ Outro modo de pensar a Equação (5 – 1) seria considerar a Equação (C – 14), desenvolvida no Apêndice C, para o cálculo do logaritmo natural do coeficiente de atividade de uma dada molécula neutra de proteína e assumindo por hipótese que os segundos coeficientes viriais entre as diversas configurações protéicas são iguais.

com o trabalho publicado por Watanabe (2007) no qual fica evidente que o diagrama de fases da lisozima da clara de ovo de galinha apresenta no máximo dois precipitados verdadeiros em regiões distintas do diagrama de fases. Pelo mesmo critério de equilíbrio estabelecido na Equação (3 – 1), tem-se que o potencial químico da proteína eletricamente neutra da fase líquida é invariante e, por consequência, a Equação (5 – 3) pode ser reescrita como:

$$\frac{1}{SB} \frac{\partial \mu_0^*(m^*)}{\partial m_{sal}} + \frac{RT}{SB} \left(\frac{\partial \ln \phi_0}{\partial m_{sal}} + \frac{\partial \ln S}{\partial m_{sal}} \right) + 2RT \left(\frac{\partial \ln B}{\partial m_{sal}} + \frac{\partial \ln S}{\partial m_{sal}} \right) = 0 \quad (5 - 4)$$

A Equação (5 – 4) é uma relação teórica entre o segundo coeficiente virial, a solubilidade da proteína na fase líquida e a molalidade de sal na mesma fase desde que se estabeleça como o potencial químico de referência deve variar com a molalidade de sal e como deve se calcular a fração de moléculas de proteína eletricamente neutras. As próximas subseções serão dedicadas a estas questões.

5.2.1 Potencial químico de referência

De acordo com Sinanoğlu e Abdunur (1965), e com Melander e Horváth (1977), a variação do potencial químico de uma molécula de proteína no estado de um gás hipotético e o potencial químico de uma molécula de proteína no interior de uma solução, ou seja, a energia livre de transferência entre estados de referência, pode ser escrita por:

$$\Delta \mu^* = \Delta \mu_{cav} + \Delta \mu_{elet} + \Delta \mu_{vdW} + RT \ln \left(\frac{RT}{pV} \right) \quad (5 - 5)$$

em que $\Delta \mu^*$ é o potencial químico de transferência de uma molécula de proteína do estado de um gás hipotético para o interior de uma solução, $\Delta \mu_{cav}$ é a variação do potencial químico devida à energia envolvida na formação de uma cavidade para alojar a molécula na solução, $\Delta \mu_{elet}$ é a contribuição eletrostática à variação do potencial químico, $\Delta \mu_{vdW}$ expressa a variação sofrida pelo potencial químico devido as interações atrativas de van der Waals entre o solvente e o soluto. O último termo da Equação (5 – 5) mede a mudança no “volume livre”.

Outra importante consideração é: quando somente a concentração de sal no

solvente aquoso é alterada e não há alterações significativas nas interações entre a proteína e o sal, assume-se que o processo de transferência de energia é afetado somente por mudanças em $\Delta\mu_{cav}$ e $\Delta\mu_{elet}$. Já que os outros termos permanecem inalterados com a variação da concentração de sal, é suficiente avaliar apenas estes dois termos que sofrem alterações. Sendo que, assim como proposto por Melander e Horváth (1977):

$$\Delta\mu_{cav} = [N_A A + 4.8 N_A^{1/3} (\kappa^e - 1) V^{2/3}] \gamma \quad (5 - 6)$$

em que N_A é o número de Avogadro, A é a área superficial molar do soluto, κ^e é um termo de correção, V é o volume molar e γ é a tensão superficial do solvente cuja expressão como função da molalidade de sal é, em geral, aproximada por um relação linear:

$$\gamma = \gamma^0 + \sigma m_{sal} \quad (5 - 7)$$

em que γ^0 é a tensão superficial da água pura e σ é uma constante chamada de incremento molal da tensão superficial.

Já para o cálculo da componente eletrostática, pode-se combinar a teoria de Debye-Hückel, que é uma lei limite para baixos valores de força iônica, com o modelo de Kirkwood para o dipolo da proteína, que é válido para valores mais elevados de força iônica. Assim, tem-se que:

$$\Delta\mu_{elet} = A - \frac{B\sqrt{m_{sal}}}{1 + C\sqrt{m_{sal}}} - D\mu_d m_{sal} \quad (5 - 8)$$

Considerando, por outro lado, um estado de referência à infinita diluição para uma molalidade de sal de equilíbrio m_1 , e outro estado de referência à infinita diluição para uma molalidade de sal de equilíbrio m_2 , tem-se que a diferença entre os potenciais químicos destes estados é igual a diferença de entre as energias livres molares de transferência entre estes estados e o estado de uma gás hipotético, tal que:

$$\mu^{2*} - \mu^{1*} = \Delta\mu^{2*} - \Delta\mu^{1*} = (\Delta\mu_{cav} + \Delta\mu_{elet})_2 - (\Delta\mu_{cav} + \Delta\mu_{elet})_1 \quad (5 - 9)$$

Dividindo a Equação (5 - 9) pela diferença entre as molalidades de sal entre os dois estados e aplicando o limite no qual tal diferença tende a zero, tem-se que:

$$\lim_{m_2 - m_1 \rightarrow 0} \frac{\mu^{2*} - \mu^{1*}}{m_2 - m_1} = \lim_{m_2 - m_1 \rightarrow 0} \frac{(\Delta\mu_{cav} + \Delta\mu_{elet})_2 - (\Delta\mu_{cav} + \Delta\mu_{elet})_1}{m_2 - m_1} \quad (5 - 10)$$

Portanto, a Equação (5 – 10) pode, evidentemente, ser reescrita em termos diferenciais como:

$$\frac{\partial \mu^*}{\partial m_{sal}} = \frac{\partial \Delta \mu_{cav}}{\partial m_{sal}} + \frac{\partial \Delta \mu_{elet}}{\partial m_{sal}} \quad (5 - 11)$$

Na região de *salting-out*, região na qual a adição de um co-solvente diminui a solubilidade da proteína, as molalidades salinas são suficientemente elevadas para que a seguinte aproximação possa ser válida no cálculo da derivada de $\Delta \mu_{elet}$ em relação à molalidade do sal através da Equação (5 – 8):

$$C\sqrt{m_{sal}} \gg 1 \Rightarrow \frac{\partial \Delta \mu_{elet}}{\partial m_{sal}} \approx -D \mu_d = -RT \Lambda \quad (5 - 12)$$

em que Λ é um coeficiente de *salting-in* relacionado a interações eletrostáticas.

Por outro lado, diferenciando-se a Equação (5 – 6) com respeito a molalidade salina, tem-se que:

$$\frac{\partial \Delta \mu_{cav}}{\partial m_{sal}} = [N_A A + 4.8 N_A^{1/3} (\kappa^e - 1) V^{2/3}] \sigma = RT \Omega \sigma \quad (5 - 13)$$

em que $\Omega \sigma$ é o coeficiente de *salting-out* relacionado a interações hidrofóbicas.

Assim, tem-se que a variação do potencial químico de referência da molécula de proteína na fase líquida em relação à molalidade de sal é dada por:

$$\frac{\partial \mu^*}{\partial m_{sal}} = RT K_s \quad (5 - 14)$$

Sendo K_s a constante de *salting-out*, a mesma que aparece na equação de Cohn que será tratada na próxima subseção, descrita como:

$$K_s = \Omega \sigma - \Lambda \quad (5 - 15)$$

5.2.2 Equação de Cohn

A famosa equação de Cohn (COHN, 1925) é uma equação fundamentalmente empírica que correlaciona linearmente o logaritmo da solubilidade de proteínas e a molalidade de sal na região de *salting-out*. O fato de Melander e Horváth (1977)

terem proposto um embasamento teórico para justificar a forma da equação de Cohn não a torna uma equação teórica, apenas apresenta uma formulação interpretativa dos parâmetros desta equação.

A equação de Cohn, por sua grande capacidade de descrever diversos sistemas com tão poucos parâmetros, ganhou notoriedade e é ainda hoje bastante utilizada no estudo de operações unitárias fundadas na precipitação de proteínas. Aqui, usar-se-á a equação de Cohn para facilitar a compreensão do que representa a Equação (5 – 4).

Esta equação pode ser escrita como (COHN, 1925):

$$\ln S = \beta - K_s m_{sal} \quad (5 - 16)$$

Portanto, a derivada do logaritmo natural da solubilidade de proteína em relação à molalidade de sal é simplesmente:

$$\frac{\partial \ln S}{\partial m_{sal}} = -K_s \quad (5 - 17)$$

Deve-se observar que teoricamente K_s não depende do pH da solução.

5.2.3 Modelo proposto

Substituindo as Equações (5 – 14) e (5 – 17) na Equação (5 – 4), tem-se que:

$$\frac{1}{SB} \frac{\partial \ln \phi_0}{\partial m_{sal}} + 2 \left(\frac{\partial \ln B}{\partial m_{sal}} + \frac{\partial \ln S}{\partial m_{sal}} \right) = 0 \quad (5 - 18)$$

A Equação (5 – 18) acima escrita é uma equação geral que relaciona o segundo coeficiente virial, a solubilidade da proteína e a molalidade do sal. A fração de moléculas neutras de proteínas é função da molalidade de sal, do pH, da temperatura, dos valores de pK_A e das constantes de equilíbrio químico que envolvem os íons salinos e os resíduos polares da cadeia protéica. Esta descrição não é trivial e nem geral, pois depende do tipo de sal usado. Para tanto aqui se assumirá a hipótese de que na região do *salting-out* pode-se admitir que:

$$\frac{\partial \ln \phi_0}{\partial m_{sal}} \approx 0 \quad (5 - 19)$$

Assim, a Equação (5 – 18) reduz-se a:

$$\frac{\partial \ln B}{\partial m_{sal}} = K_s \quad (5 - 20)$$

Integrando-se a Equação (5 – 20), tem-se que:

$$\ln \left(\frac{B}{B^*} \right) = K_s (m_{sal} - m_{sal}^*) \quad (5 - 21)$$

A Equação (5 – 21) relaciona o segundo coeficiente virial com a molalidade de sal, tendo apenas um ponto da curva e a constante de *salting-out* como parâmetros.

Do mesmo modo, pode-se observar que pela equação de Cohn e pela Equação (5 – 20) é possível escrever que:

$$\frac{\partial \ln B}{\partial m_{sal}} + \frac{\partial \ln S}{\partial m_{sal}} = 0 \quad (5 - 22)$$

E, finalmente, integrando a Equação (5 – 22) encontra-se uma relação entre o segundo coeficiente virial e a solubilidade de proteína na região do *salting-out* para uma variação da molalidade de sal:

$$S = \frac{B^* S^*}{B} \quad (5 - 23)$$

É de importante menção que a Equação (5 – 23) tem sua validade restrita a valores de solubilidade de proteína em função do segundo coeficiente virial para diferentes valores de molalidade de sal, ou seja, mudanças em pH e em temperatura *a priori* não estão contempladas neste modelo.

5.3 Resultados e Discussões

5.3.1 Resultados para a lisozima

Um primeiro resultado que servirá, na verdade, para comparação e validação das Equações (5 – 21) e (5 – 23) é a curva de solubilidade da lisozima em função da concentração de cloreto de sódio na região de *salting-out*. Essa curva nestas

condições deve seguir a equação de Cohn – Equação (5 – 16) – e dela pode-se obter o valor da constante de *salting-out*, K_s , por simples regressão linear. A Figura 5.1 apresenta esses resultados a partir de dados experimentais obtidos por Watababe (2007).

Figura 5.1 - Solubilidade da lisozima da clara de ovo de galinha em função da molalidade de cloreto de sódio. Dados experimentais (○), Watanabe (2007). Equação de Cohn (linha contínua).

O valor de K_s para este sistema foi calculado como $0,93 \text{ kg mol}^{-1}$ e servirá de base para comparação com o valores de K_s ajustados pela Equação (5 – 21).

A Equação (5 – 21) foi aplicada a um conjunto de dados experimentais publicados por Curtis et al. (2002). Nesse trabalho foram apresentados os valores do segundo coeficiente virial osmótico em função da molalidade de cloreto de sódio para a lisozima nativa cujo resíduo de número 101 é um ácido aspártico e para uma variante da lisozima cujo resíduo 101 é uma fenilalanina. Esta variante da lisozima será referenciada no texto como lisozima D101F. Os valores do segundo coeficiente virial osmótico foram obtidos por espalhamento de luz.

A Figura 5.2 apresenta os resultados da aplicação da Equação (5 – 21) aos dados de Curtis et al. (2002).

Figura 5.2 - Valores do segundo coeficiente virial osmótico em função da molalidade de cloreto de sódio em pH = 4,5. Dados experimentais obtidos em Curtis et al. (2002). Lisozima nativa, modelo (linha contínua), dados experimentais (□). Lisozima D101F, modelo (linha pontilhada), dados experimentais (●).

Os valores das constantes de *salting-out* para a lisozima nativa e para lisozima mutante D101F obtidos pelo ajuste da Equação (5 – 21) foram 0,70 e 1,93 kg mol⁻¹ respectivamente. Pode-se observar assim que a abordagem aqui pretendida reproduz resultados coerentes com aquilo que tem sido reportado na literatura.

O fato de a alteração de apenas um resíduo na cadeia protéica ser responsável por uma elevação tão grande no valor de K_s mostra a importância que a estrutura primária e, por conseqüência, a estrutura terciária tem para a modelagem termodinâmica de sistemas contendo proteínas, não podendo, portanto, serem ignoradas. Há que se registrar também que o ponto de referência B^* foi escolhido arbitrariamente dentre os dados experimentais, mas poderia ter sido ajustado.

Já a Equação (5 – 23) foi também aplicada a sistemas contendo lisozima. A Figura 5.3 mostra o resultado desta aplicação e a comparação com outros modelos publicados por outros autores no intento de propor uma relação teórica entre a solubilidade da proteína e o segundo coeficiente virial osmótico. Intento esse motivado pela hipótese de George et al. (1997), que apresentam uma correlação entre os dados experimentais de ambas as propriedades.

Figura 5.3 - Valores da solubilidade da lisozima em função do segundo coeficiente virial osmótico. Dados experimentais obtidos em Guo et al. (1999). Lisozima a 25 °C e pH = 4,2 com variações na concentração de cloreto de sódio (○). Lisozima a 18°C e pH = 4,5 com variações na concentração de cloreto de amônio (□). Lisozima a 23 °C e pH = 7,8 com variações na concentração de brometo de magnésio (♦). Modelo de Haas; Drenth e Wilson (1999) com $z = 3$ e $A = 0.001$ (linha pontilhada). Modelo de Guo et al. (1999) com $\Delta\mu = 250 \text{ kJ mol}^{-1}$ (linha semi-contínua). Modelo de Ruppert; Sandler e Lenhoff (2001), (linha com travessões). Equação (5 – 23) (linha contínua).

A Tabela 5.1 apresenta a comparação entre os modelos considerando o valor de CQ definido pela Equação (3 – 40).

Tabela 5.1 - Comparação dos valores de CQ definido pela Equação (3 – 40) para o sistema contendo lisozima.

Modelos	CQ
Modelo Proposto – Equação (5 - 23)	39,6 %
Modelo de Guo et al. (1999)	58,4 %
Modelo de Haas; Drenth e Wilson (1999)	44,2 %
Modelo de Ruppert; Sandler e Lenhoff (2001)	27,7 %

Observa-se pela Tabela 5.1 que todos os modelos são pouco eficientes para uma avaliação quantitativa, no entanto pela Figura 5.3 observa-se que os modelos

conseguem ao menos representar qualitativamente o fenômeno estudado. Além disso, a modelagem proposta se equipara aos modelos já publicados na literatura no que tange a eficácia de representação dos dados experimentais; sendo, todavia, que esta nova modelagem resulta em uma equação muitíssimo mais simples e que exige muito menos parâmetros.

Há uma ressalva, no entanto, ao uso das Equações (5 – 21) e (5 – 23). Sendo a região do diagrama de fases aquela de *salting-out*, as equações são válidas apenas para valores negativos do segundo coeficiente virial osmótico, ou seja, para forças intermoleculares atrativas entre as moléculas de proteína (PRAUSNITZ, 2003), que é fisicamente o que acontece na cristalização e na precipitação de proteínas.

5.3.2 Resultados para a ovalbumina

A Figura 5.4 apresenta a aplicação da equação de Cohn aos dados experimentais de solubilidade de ovalbumina em soluções de sulfato de amônio. O valor de K_s ajustado por regressão linear é $2,45 \text{ kg mol}^{-1}$.

Figura 5.4 - Solubilidade da ovalbumina em soluções de sulfato de amônio. Dados experimentais ($\square$), Judge; Johns e White (1996). Equação de Cohn (linha contínua).

Usando a equação de Cohn para ajustar o valor de K_s com os dados de solubilidade de ovalbumina em soluções de sulfato de amônio, foi possível prever sem nenhum ajuste de parâmetro adicional a relação entre o segundo coeficiente virial osmótico e a força iônica (no fundo, a concentração de sulfato de amônio) pela Equação (5 – 21).

A Figura 5.5 traz os resultados desta modelagem para os dados do segundo coeficiente virial osmótico da ovalbumina.

Figura 5.5 - Segundo coeficiente virial osmótico da ovalbumina em soluções de sulfato de amônio a pH = 4,0. Dados experimentais (○), Mehta; White e Litster (2011). Equação (5 – 21) com $K_s = 2,45 \text{ kg mol}^{-1}$, (linha contínua).

De fato, apesar do valor de K_s ter sido ajustado a partir de um conjunto de dados de solubilidade obtidos em um pH ligeiramente distinto daquele nos quais foram obtidos os valores do segundo coeficiente virial osmótico, observa-se que ele foi capaz de prever juntamente a Equação (5 – 21) com grande precisão os valores do segundo coeficiente virial osmótico.

O fato de ter se trocado na equação de Cohn a molalidade do agente precipitante pela força iônica não traz nenhuma alteração às relações propostas desde que se observe certo zelo em qual formulação da equação de Cohn foi calculada a constante de *salting-out*, K_s .

5.3.3 Resultados para a imunoglobulina humana

Por fim, verificou-se, também, a validade da relação expressa na Equação (5 – 21) aplicando-a a sistemas contendo imunoglobulina humana. Esta por sua vez é constituída de uma população de proteínas e apresenta em sua modelagem situações ainda de difícil tratamento tal como o cálculo do ponto isoelétrico. De qualquer modo, as relações aqui propostas levam em conta a identidade da proteína apenas no parâmetro K_s , assumindo que tal população de proteínas obedeça a equação de Cohn, pode-se supor que as relações aqui apresentadas devem ser obedecidas do mesmo modo. E isso pode ser verificado na Figura 5.6 para dados do segundo coeficiente virial osmótico para a imunoglobulina humana em função da fração mássica de cloreto de sódio. Os dados experimentais deste sistema ainda não foram publicados, mas foram gentilmente cedidos por Kelly Cristina do Nascimento Alves Cruvinel (*comunicação pessoal*).

Figura 5.6 - Segundo coeficiente virial osmótico da imunoglobulina humana (IgG) em soluções de cloreto de sódio. Dados experimentais ($\circ$), estes dados ainda não foram publicados, mas foram cedidos por Kelly Cristina do Nascimento Alves Cruvinel (*comunicação pessoal*). Equação (5 – 21), (linha contínua).

6 Conclusões

Terminada a exposição dos principais desenvolvimentos, cabe agora reunir aqui os pontos fundamentais tratados neste trabalho. Talvez a maior contribuição desta pesquisa seja a consideração de que uma dada proteína em solução não se encontrando em uma forma única, mas como uma população de moléculas com distintas configurações de carga elétrica ao longo de suas cadeias, deva ter um tratamento termodinâmico que leve em conta este fato. E que a analogia para moléculas menores tais como aminoácidos e compostos β -lactâmicos é total.

Com certeza um dos maiores êxitos deste projeto tenha sido o desenvolvimento de uma relação explícita e unívoca entre a solubilidade de proteínas e o pH da solução. Esta relação foi aplicada a sistemas de insulina suína e algumas mutantes de insulina bem como a sistemas de β -lactoglobulina com um resultado bastante expressivo no que tange a precisão na representação dos dados experimentais.

Pela natureza própria da equação proposta, a necessidade de dados de pK_A é uma realidade desafiadora, pois a escassez de informações a respeito destes dados na literatura é ainda um impedimento a uma aplicação mais ampla do modelo.

Ainda assim, tanto quanto foi possível verificar na literatura aberta, é a primeira vez que uma tal relação, fundada em argumentos teóricos de físico-química básica, entre a solubilidade de proteínas e o pH é apresentada na literatura aberta.

A extensão deste modelo ao caso de sistemas contendo aminoácidos e grupos β -lactâmicos abriu uma possibilidade de aprofundar a discussão da não-idealidade destes sistemas. A possibilidade, no entanto, de aplicação deste modelo somado as considerações de não-idealidade a sistemas de aminoácidos é imensa já que os dados de pK_A destas moléculas estão disponíveis na literatura para uma vasta gama de moléculas.

Uma interpretação via termodinâmica estatística para o parâmetro de interação de Pitzer permitiu, também, uma discussão mais minuciosa desta questão. A relação entre o tamanho da molécula e o aumento do desvio da idealidade para os sistemas analisados, inclusive para a glicina e seus oligopeptídeos nos quais isto fica evidente, concorre a interpretação dada ao parâmetro de interação de Pitzer e tem profunda relação com o potencial de esferas rígidas que é associado ao efeito de

“volume excluído”.

Há no capítulo dedicado a este tema uma seção que contempla uma aplicação tecnológica bastante interessante do ponto de vista de engenharia: a modelagem do equilíbrio sólido-líquido de sistemas que contém moléculas com grupos β -lactâmicos, por exemplo a ampicilina, cujo interesse industrial é notório dada a sua função antibiótica.

Em termos de processo a dependência da solubilidade com o pH é crucial em operações como a precipitação isoelétrica, ainda assim parâmetros como temperatura e composição do solvente foram avaliados no caso da ampicilina e do ácido 6-aminopenicilânico com grande sucesso na correlação dos dados experimentais.

Além disso, foram desenvolvidas relações entre o segundo coeficiente virial osmótico, a concentração de agente precipitante e a solubilidade de proteínas. Neste campo, alguns modelos foram desenvolvidos, mas a necessidade de ajuste de parâmetros e de informações que estes modelos requerem evocam a necessidade de relações tão precisas quanto, mas de mais fácil implementação.

Pôde-se observar que os modelos até hoje desenvolvidos incluindo o que aqui se apresenta conseguem descrever a relação entre o segundo coeficiente virial osmótico e a solubilidade da proteína ao menos em termos qualitativos. Contudo, os próprios dados experimentais, por serem raros, limitam uma maior exploração da validade dos modelos.

Ficou também evidente a característica que o segundo coeficiente virial osmótico tem de representar a magnitude das forças de interação entre as moléculas de proteína, ou seja, quanto mais negativo ele o for, mais atrativas são as forças entre as moléculas de proteína facilitando a agregação e diminuindo a solubilidade em solução. Este fato pode ser visualizado em todos os modelos.

Como sugestões a trabalhos futuros é possível elencar algumas questões que permanecem em aberto tanto na área experimental quanto em termos de modelagem termodinâmica: i) determinação experimental dos valores de pK_a na cadeia protéica para diferentes proteínas e em diferentes condições tais como temperatura e composição do solvente; ii) determinação de curvas de solubilidade de proteínas em função do pH em diversas condições a fim de consolidar a validação do modelo proposto; iii) determinação experimental de curvas de solubilidade de aminoácidos com três resíduos ionizáveis e estudo de como

contemplar a questão da não-idealidade nestes sistemas; iv) estudo de como levar em conta o fato de uma dada proteína em um dado pH ser de fato uma população de moléculas com distintos estados de ionização no cômputo do potencial da força média de proteínas a fim de prever o diagrama de fases com uma abordagem mais realística; v) estudo da possível relação entre uma dada configuração protéica em um estado de ionização e a atividade biológica desta proteína; é possível supor que sabendo qual é esta configuração haja uma relação entre a curva da fração desta configuração e a curva de atividade biológica da proteína em função do pH.

7 Referências bibliográficas

- ALLAHYAROV, E.; LÖWEN, H.; LOUIS, A. A.; HANSEN, J. P. Discrete charges patterns, Coulomb correlations and interactions in protein solutions. *Europhys. Lett.*, v. 57, p. 731-737, 2002
- BROMLEY, L. A. Thermodynamic properties of strong electrolytes in aqueous solutions, *AIChE J.*, v. 19, p. 313-320, 1973
- CHEN, C. C.; ZHU, Y.; EVANS, L. B. Phase partitioning of biomolecules: solubilities of aminoacids, *Biotechnol. Prog.*, v. 5, p. 111-118, 1989
- COHN, E. J. The physical chemistry of the proteins. *Physiol. Rev.*, v. 5, p. 349-437, 1925
- COX, M. M.; NELSON, D. L. **Lehninger Principles of Biochemistry**, 3 ed., Worth Publishers, New York, 2003
- CURTIS, R. A.; STEINBRECHER, C.; HEINEMANN, M.; BLANCH, H. W.; PRAUSNITZ, J. M. Hydrofobic forces between protein molecules in aqueous solutions of concentrated electrolyte. *Biophys. Chem.*, v. 98, p. 249-265, 2002
- FRANCO, L. F. M.; PESSÔA FILHO, P. de A. On the solubility of proteins as a function of pH: mathematical development and application. *Fluid Phase Equilibr.*, v. 306, p. 242-250, 2011
- FREDERICQ, E; NEURATH, H. The interaction of insulin with thiocyanate and other anions. The minimum molecular weight of insulin. *J. Am. Chem. Soc.*, v. 72, p. 2684–2691, 1950
- FUCHS D.; FISCHER, J.; TUMAKAKA, F.; SADOWSKI, G. Solubility of Amino Acids: Influence of the pH value and the Addition of Alcoholic Cosolvents on Aqueous Solubility, *Ind. Eng. Chem. Res.*, v. 45, p. 6578-6584, 2006
- GEORGE, A.; CHIANG, Y.; GUO, B.; ARABSHAHI, A.; CAI, Z.; WILSON, W. W. Second virial coefficient as predictor in protein crystal growth. *Methods Enzymol.*, v. 276, p. 100-110, 1997

- GRENN, A. A. Studies in the physical chemistry of the proteins. *J. Biol. Chem.*, v. 95, p. 517-542, 1931
- GREENSTEIN, J. P.; WINITZ, M. **Chemistry of the Amino Acids**, v. 1, John Wiley & Sons, New York, 1961
- GRÖNWALL, A. Studies on the solubility of lactoglobulin. *Compt. rend. trav. Lab. Carlsberg*, v. 24, p. 185-200, 1941
- GUO, B.; KAO, S.; MCDONALD, H.; ASANOV, A.; COMBS, L. L.; WILSON, W. W. Correlation of second virial coefficients and solubilities useful in protein crystal growth. *J. Cryst. Growth*, v. 196, p. 424-433, 1999
- GUPTA, R. B.; HEIDEMANN, R. A. Solubility models for amino acids and antibiotics, *AIChE J.*, v. 36, p. 333-341, 1990
- HAAS, C.; DRENTH, J.; WILSON, W. W. Relation between the solubility of proteins in aqueous solutions and the second virial coefficient of the solution. *J. Phys. Chem. B*, v. 103, p. 2808-2811, 1999
- HILL, T. L. **An Introduction to Statistical Thermodynamics**, Dover, 1986
- HITCHCOCK, D. The solubility of tyrosine in acid and in alkali, *J. Gen. Physiol.*, v. 6, p. 747-757, 1924
- JUDGE, R. A.; JOHNS, M. R.; WHITE, E. T. Solubility of ovalbumin in ammonium sulfate solutions, *J. Chem. Eng. Data*, v. 41, p. 422-424, 1996
- KAARSHOLM, N. C.; HAVELUNG, S.; HOUGARD, P. Ionization behavior of native and mutant insulins: pK perturbation of B13-Glu in aggregated species. *Arch. Biochem. Biophys.*, v. 283, p. 496-502, 1990
- KHOSHBARCHI, M. K.; VERA, J. H. Measurement and correlation of ion activity in aqueous single electrolyte solutions. *AIChE J.*, v. 42, p. 249-258, 1996a
- _____. Measurement and modeling of activities of amino acids in aqueous salt systems. *AIChE J.*, v. 42, p. 2354-2364, 1996b

KHOSHKBARCHI, M. K.; VERA, J. H. A simplified perturbed hard-sphere model for the activity coefficients of amino acids and peptides in aqueous solutions. *Ind. Eng. Chem. Res.*, v. 35, p. 4319-4327, 1996c

KIRKWOOD, J. G. Theory of Solutions of Molecules Containing Widely Separated Charges with Special Application to Zwitterions. *J. Chem. Phys.*, v. 2, p. 351-361, 1934

_____. Theoretical studies upon dipolar ions. *Chem. Rev.*, v. 24, p. 233-251, 1939

LARSEN, B. L.; RASMUSSEN, P.; FREDENSLUND, A. A modified UNIFAC group-contribution model for prediction of phase equilibria and heats of mixing, *Ind. Eng. Chem. Res.*, v. 26, p. 2274-2286, 1987

LIMA, E. R. A.; BISCAIA JR.; E. C.; BOSTRÖM, M.; TAVARES, F. W.; PRAUSNITZ, J. M. Osmotic second virial coefficients and phase diagrams for aqueous proteins from a much-improved Poisson-Boltzmann equation. *J. Phys. Chem. C*, v. 111, p. 16055-16059, 2009

LINDERSTRØM-LANG, K. On the ionisation of proteins. *Compt. rend. trav. Lab. Carlsberg*, v. 15, p. 1-29, 1924

LU, J.; WANG, X. J.; YANG, X.; CHING, C. B. Solubilities of Glycine and Its Oligopeptides in Aqueous Solutions, *J. Chem. Eng. Data*, v. 51, p. 1593-1596, 2006

McMILLAN, W. G.; MAYER, J. E. The statistical thermodynamics of multicomponents systems. *J. Chem. Phys.*, v. 13, p. 276-305, 1945

MEHTA, C. H.; WHITE, E. T.; LITSTER, J. D. Correlation of Second Virial Coefficient with solubility for proteins in salt solutions, *Biotechnol. Prog.*, 2011

MELANDER, W.; HORVÁTH, C. Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: an interpretation of the lyotropic series. *Arch. Biochem. Biophys.*, v. 183, p. 200-215, 1977

MOLLERUP, J. M.; BREIL, M. P. On the thermodynamics of the McMillan-Mayer state function, *Fluid Phase Equilib.*, v. 276, p. 18-23, 2009a

- MOLLERUP, J. M.; BREIL, M. P. The osmotic second virial coefficient and the Gibbs-McMillan-Mayer framework, *Fluid Phase Equilibr.*, v. 286, p. 88-94, 2009b.
- NASS, K. K. Representation of the solubility behavior of amino acids in water, *AIChE J.*, v. 34, p. 1257-1266, 1988
- NEAL, B. L.; ASTHAGIRI, D.; VELEV, O. D.; LENHOFF, A. M.; KALER, E. W. Why is the osmotic second virial coefficient related to protein crystallization? *J. Cryst. Growth*, v. 196, p. 377-387, 1999
- NELDER, J. A.; MEAD, R. A simplex method for function minimization, *Comp. J.*, v. 7, p. 308-313, 1965
- PARK, B. H.; YOO, K. P.; LEE, C. S. Phase equilibria and properties of amino acids + water mixtures by hydrogen-bonding lattice fluid equation of state, *Fluid Phase Equilibr.*, v. 212, p. 175-182, 2003
- PESSÔA FILHO, P. de A.; MAURER, G. An extension of the Pitzer equation for the excess Gibbs energy of aqueous electrolyte systems to aqueous polyelectrolyte solutions, *Fluid Phase Equilibr.*, v. 269, p. 25-35, 2008
- PINHO, S. P.; SILVA, C. M.; MACEDO, E. A. Solubility of amino acids: a group-contribution model involving phase and chemical equilibrium. *Ind. Eng. Chem. Res.*, v. 33, p. 1341-1347, 1994
- PITZER, K. S. Thermodynamics of electrolytes, *J. Phys. Chem.*, v. 77, p. 268-277, 1973
- PITZER, K. S.; PEIPER, J. C. Thermodynamic Properties of Aqueous Sodium Chloride Solutions, *J. Phys. Chem. Ref. Data*, v. 13, p. 1-101, 1984
- PRADHAN, A. A.; VERA, J. H. Effects of acids and bases on the solubility of amino acids. *Fluid Phase Equilibr.*, v. 152, p. 121-132, 1998
- PRAUSNITZ, J. M.; LICHTENTHALER, R. N.; AZEVEDO, E. G. de, **Molecular Thermodynamics of Fluid-Phase Equilibria**, 3 ed, Prentice Hall, 1999
- PRAUSNITZ, J. M. Molecular thermodynamics for some applications in biotechnology. *J. Chem. Thermodynamics*, v. 35, p. 21-39, 2003

- RENON, H.; PRAUSNITZ, J. M. Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures, *AIChE J.*, v. 14, p. 135-144, 1968
- RODRÍGUEZ-RAPOSO, R.; FERNÁNDEZ-MÉRIDA, L.; ESTESO, M. A. Activity coefficients in (electrolyte + amino acid)(aq), the dependence of the of the ion-zwitterion interactions on the ionic strength and on the molality of the amino acid analysed in terms of the Pitzer's equation, *J. Chem. Thermodynamics*, v. 26, 1121-1128, 1994
- RUPPERT, S.; SANDLER, S. I.; LENHOFF, A. M. Correlation between the osmotic second virial coefficient and the solubility of proteins. *Biotechnol. Prog.*, v. 17, 182-187, 2001
- SANTANA, M. C.; RIBEIRO, M.P.A.; LEITE, G.A.; GIORDANO, R.L.C.; GIORDANO, R.C.; MATTEDI, S. Solid-Liquid Equilibrium of Substrates and Products of the Enzymatic Synthesis of Ampicillin, *AIChE J.*, v. 56, p. 1578-1583, 2010
- SANTANA, M. C.; MATTEDI, S.; GIORDANO, R. C. Efeito do metanol na precipitação dos compostos presentes na reação enzimática de ampicilina. In: Congresso Brasileiro de Termodinâmica Aplicada, 5, 2010, Foz do Iguaçu. **Anais.** p. 4863-4868
- SCATCHARD, G.; KIRKWOOD, J. G. *Phys. Zeits.*, v. 33, 1932
apud KIRKWOOD, J. G. Theory of Solutions of molecules containing widely separated charges with special application to zwitterions, *J. Chem. Phys.*, v. 2, p. 351-361, 1934
- SEYFI, S.; PAZUKI, G.; AGHAMIRI, S. F., BEHESHTI, M. On the prediction of equilibrium phase behavior of amino acids in aqueous and aqueous-electrolyte solutions using SAFT equation of state. *Fluid Phase Equilibr.*, v. 287, p. 15-22, 2009
- SINANOĞLU. O.; ABDULNUR, S. Effect of water and other solvents on the structure of biopolymers. *Fed. Proc.*, v. 24, p. 12-23, 1965
- STRYER, L. **Biochemistry**, 4 ed., W.H. Freeman and Company, New York, 1995
- TASHIMA, A. K.; OTTENS, M.; VAN DER WIELEN, L. A.; CINTRA, D. E.; PAULI, J. R.; PESSÔA FILHO, P. A.; MIRANDA, E. A. Precipitation of porcine insulin with carbon dioxide. *Biotech. Bioeng.*, v.103, p. 909-919, 2009

- TESSIER, P. M.; LENHOFF, A. M.; SANDLER, S. I. Rapid measurement of protein osmotic second virial coefficients by self-interaction chromatography. *Biophys. J.*, v. 82, p.1620-1631, 2002
- TJONG, H.; ZHOU, H. X. Prediction of protein solubility from calculation of transfer free energy. *Biophys. J.*, v. 95, p. 2601-2609, 2008
- TSENG, H. C.; LEE, C. Y.; WENG, W. L.; SHIAH, I. M. Solubilities of amino acids in water at various pH values under 298.15 K. *Fluid Phase Equilib.*, v. 285, p. 90-95, 2009
- WATANABE, E. O. *Equilíbrio de fases na precipitação de lisozima e albumina de soro bovino com o uso de sais*; Faculdade de Engenharia Química, Universidade de Campinas, Campinas, 2007
- WILSON, G. M. Vapor-Liquid Equilibrium XI. A new expression for the excess free energy of mixing, *J. Am. Chem. Soc.*, v. 86, p. 127-130, 1964
- XU, X.; PINHO, S. P.; MACEDO, E. A. Activity coefficient and solubility of amino acids in water by the modified Wilson model, *Ind. Eng. Chem. Res.*, v. 43, p. 3200-3204, 2004

Apêndice A – Integração do perfil de carga em função do pH

Para se chegar à Equação (3 – 31) a partir da Equação (3 – 29) é preciso integrar o perfil de carga líquida da proteína em função do pH. Assim, tem que se calcular a seguinte integral:

$$\int_{pI}^{pH} \bar{c}(pH) d pH = \frac{1}{2} \int_{pI}^{pH} \left\{ \sum_{j=1}^n \left(\frac{1}{1 + 10^{pH - pK_{Aj}}} \right) [1 + v_j - (1 - v_j) 10^{pH - pK_{Aj}}] \right\} d pH \quad (C - 1)$$

Para isso, deve-se fazer a seguinte substituição de variáveis a fim de tornar o desenvolvimento mais simples de se compreender:

$$y = 10^{pH - pK_{Aj}} \quad (C - 2)$$

Assim, ter-se-á que:

$$\log y = pH - pK_{Aj} \quad (C - 3)$$

E, que:

$$d pH = \frac{1}{\ln 10} \frac{1}{y} dy \quad (C - 4)$$

Portanto, substituindo as Equações (C – 3) e (C – 4) na Equação (C – 1), tem-se que:

$$\begin{aligned} \int_{pI}^{pH} \bar{c}(pH) d pH &= \frac{1}{2} \int_{y^0}^y \left\{ \sum_{j=1}^n \left(\frac{1}{1 + y} \right) [1 + v_j - (1 - v_j) y] \frac{1}{\ln 10} \frac{1}{y} \right\} dy = \dots \\ &= \frac{1}{2 \ln 10} \sum_{j=1}^n \left[(1 + v_j) \int_{y^0}^y \left(\frac{1}{1 + y} \right) \frac{1}{y} dy - (1 - v_j) \int_{y^0}^y \left(\frac{y}{1 + y} \right) \frac{1}{y} dy \right] = \dots \\ &= \frac{1}{2 \ln 10} \sum_{j=1}^n \left\{ (1 + v_j) \left[\int_{y^0}^y \left(\frac{1}{y} \right) dy - \int_{y^0}^y \left(\frac{1}{1 + y} \right) dy \right] - (1 - v_j) \int_{y^0}^y \left(\frac{1}{1 + y} \right) dy \right\} = \dots \\ &= \frac{1}{2 \ln 10} \sum_{j=1}^n \left\{ (1 + v_j) \left[\ln \frac{y}{y^0} - \ln \frac{(1 + y)}{(1 + y^0)} \right] - (1 - v_j) \ln \frac{(1 + y)}{(1 + y^0)} \right\} \quad (C - 5) \end{aligned}$$

Pode-se reescrever a Equação (C – 5) em termos do logaritmo decimal tal que:

$$\begin{aligned}
\int_{pI}^{pH} \bar{c}(pH) d pH &= \frac{1}{2} \sum_{j=1}^n \left[(1+v_j) \log \frac{y}{y^0} - (1+v_j) \log \frac{(1+y)}{(1+y^0)} - (1-v_j) \log \frac{(1+y)}{(1+y^0)} \right] = \dots \\
&= \frac{1}{2} \sum_{j=1}^n \left\{ (1+v_j) \log \frac{y}{y^0} - [(1+v_j) + (1-v_j)] \log \frac{(1+y)}{(1+y^0)} \right\} = \dots \\
&= \frac{1}{2} \sum_{j=1}^n \left[(1+v_j) \log \frac{y}{y^0} - 2 \log \frac{(1+y)}{(1+y^0)} \right] \quad (C-6)
\end{aligned}$$

Substituindo as variáveis originais a Equação (C – 6):

$$\begin{aligned}
\int_{pI}^{pH} \bar{c}(pH) d pH &= \sum_{j=1}^n \left[\frac{(1+v_j)}{2} \log \left(\frac{10^{pH-pK_{A_j}}}{10^{pI-pK_{A_j}}} \right) - \log \frac{(1+10^{pH-pK_{A_j}})}{(1+10^{pI-pK_{A_j}})} \right] = \dots \\
&= \sum_{j=1}^n \left[\frac{(1+v_j)}{2} (pH-pI) - \log \frac{(1+10^{pH-pK_{A_j}})}{(1+10^{pI-pK_{A_j}})} \right] = \dots \\
&= (pH-pI) \sum_{j=1}^n \left[\frac{(1+v_j)}{2} \right] - \sum_{j=1}^n \left[\log \frac{(1+10^{pH-pK_{A_j}})}{(1+10^{pI-pK_{A_j}})} \right] \quad (C-7)
\end{aligned}$$

Entretanto:

$$\sum_{j=1}^n \left[\frac{(1+v_j)}{2} \right] = n_+ \quad (C-8)$$

Por fim, substituindo a Equação (C – 8) na Equação (C – 7):

$$\int_{pI}^{pH} \bar{c}(pH) d pH = n_+ (pH-pI) - \sum_{j=1}^n \left[\log \frac{(1+10^{pH-pK_{A_j}})}{(1+10^{pI-pK_{A_j}})} \right] \quad (C-9)$$

Apêndice B – Código-fonte do programa de cálculo de solubilidade de proteínas

```

/*****/
/* Escola Politécnica da Universidade de São Paulo */
/* Departamento de Engenharia Química */
/* Grupo de Engenharia de Bioprocessos (GenBio) */
/* Programa de cálculo da solubilidade de proteínas em função do pH */
/* Luís Fernando Mercier Franco & Pedro de Alcântara Pessoa Filho */
/*****/

```

```

/*****/

```

1) Para usar este programa um arquivo de entrada é requerido. A extensão deste arquivo precisa ser ".txt" e este arquivo deve ser salvo nos mesmos diretório e pasta do executável. Abaixo um exemplo para a insulina suína a 298.2 K, dados de Tashima et al. (2009).

```

Nome_da_proteína:
Insulina_suina
Número_de_resíduos:
16
(Aminoácidos,nu[j],pKA):
G 1 7.38
E -1 4.37
Y -1 10.20
E -1 4.37
Y -1 10.20
N -1 3.45
F 1 7.38
H 1 5.98
H 1 5.98
E -1 4.37
Y -1 10.20
E -1 4.37
R 1 11.10
Y -1 10.20
K 1 8.52
A -1 3.45
(pI,pIsol):
5.4 0.0000222
Número_de_dados_de_solubilidade:
8
Dados_de_solubilidade(pH,sol):
4.6 0.0000941
4.7 0.0000569
4.8 0.0000441
5.0 0.0000307
5.3 0.0000246
5.4 0.0000222
5.6 0.0000259
5.7 0.0000502

```

2) Um arquivo chamado "_helpfile.txt" é criado durante a execução do programa. Este arquivo contém os valores dos parâmetros discretos $\epsilon[j,k]$ para todos os resíduos e para todas as configurações neutras possíveis.

3) Os autores não têm qualquer responsabilidade sobre o uso deste programa.

```

/*****/

```



```

#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <string.h>
#include <time.h>

#define MAX 100
#define MIN 50
#define NUM 40

int main(){

    /* Descrição das variáveis */
    int i;                /* contador do vetor pH */
    int j;                /* contador de grupos ionizáveis */
    int k;                /* contador de configurações */
    int n;                /* número de grupos ionizáveis */
    int nexp;             /* número de pontos experimentais */
    int nu[MIN];          /* vetor de grupos ionizáveis */
    int M;                /* número total de configurações */
    int npos;             /* número de grupos que têm nu[j] = +1 */
    int count;
    int epsm[MIN];        /* carga de cada grupo na configuração k */
    int aux;
    int rem;
    int quot;
    int eps;
    int c;
    int AS;
    int BS;
    int pmc;              /* posição da configuração mais comum */
    int epsmc[MIN];       /* carga de cada grupo na configuração mais comum */
    double pH[MAX];       /* vetor de pH */
    double pKA[MIN];      /* vetor de pKA */
    double ex;
    double ex0;
    double pI;            /* pI: ponto isoelétrico */
    double sexp[MIN];     /* solubilidade experimental */
    double pIsol;         /* solubilidade no pI */
    double Z[MAX];        /* carga líquida da proteína em um dado pH */
    double s18[MAX];      /* solubilidade calculada pela Equação(3 - 18) */
    double s32[MAX];      /* solubilidade calculada pela Equação(3 - 32) */
    double mc;
    double frac;
    double auxiliar;
    double phimc[MAX];    /* fração da configuração mais comum */
    double phitot[MAX];   /* fração total de moléculas neutras */
    double prod;

    char name[NUM];
    char res[NUM];
    char nfile[NUM];
    char get[NUM];
    time_t t1, t2;

    FILE *ARQUI, *HELPPFILE;

    (void) time(&t1);

    printf("-----");
    printf("\n");
    printf("\n\t ESCOLA POLITECNICA DA UNIVERSIDADE DE SAO PAULO\n");
    printf("\n\t DEPARTAMENTO DE ENGENHARIA QUIMICA\n");
    printf("\n\t GRUPO DE ENGENHARIA DE BIOPROCESSOS\n");
    printf("\n\t ..... ");
    printf("\n\t : SOLUBILIDADE DE PROTEÍNAS EM FUNÇÃO DO pH : ");
    printf("\n\t : .....:");
    printf("\n\t \n\t Desenvolvido por Luis Fernando Mercier Franco");

```

```

printf(" \n\t & Pedro de Alcantara Pessoa Filho");

printf("\n Data input:\n");
printf(" Data file name: ");
scanf("%s", &nfile);

/*****
/* Lendo o arquivo (.txt) que contém os valores de entrada */
*****/

ARQUI = fopen(nfile, "r");
fscanf(ARQUI, "%S", get);
fscanf(ARQUI, "%s", &name);
fscanf(ARQUI, "%S", get);
fscanf(ARQUI, "%d", &n);
fscanf(ARQUI, "%S", get);
for (j = 0; j < n; j++){
    fscanf(ARQUI, "%s", &res[j]);
    fscanf(ARQUI, "%d", &nu[j]);
    fscanf(ARQUI, "%lf", &pKA[j]);
}
fscanf(ARQUI, "%S", get);
fscanf(ARQUI, "%lf", &pI);
fscanf(ARQUI, "%lf", &pIsol);
fscanf(ARQUI, "%S", get);
fscanf(ARQUI, "%d", &nexp);
fscanf(ARQUI, "%S", get);
for (k = 0; k < nexp; k++){
    fscanf(ARQUI, "%lf", &pH[k]);
    fscanf(ARQUI, "%lf", &sexp[k]);
}
fclose(ARQUI);

/*****
*****/

/*****
/* Cálculo do perfil de carga pela Equação (3 - 28) */
*****/

for (i = 0; i < nexp; i++){
    Z[i] = 0.0;
    for (j = 0; j < n; j++){
        ex = pow(10.0, pH[i] - pKA[j]);
        AS = 1 + nu[j];
        BS = 1 - nu[j];
        Z[i] = Z[i] + 0.5 * (1.0 * AS - 1.0 * BS * ex) / ex;
    }
}

/*****
/* Cálculo da solubilidade da proteína pela Equação (3 - 18) */
*****/

/* Cálculo de configurações neutras */

HELPPFILE = fopen("_helpfile.txt", "w");
count = 0;
M = pow(2, n);
for (k = 0; k < M; k++){
    aux = k;
    c = 0;
    for (j = 0; j < n; j++){
        rem = aux % 2;
        quot = aux / 2;
        epsm[j] = rem;
        aux = quot;
    }
}

```

```

    c = c + epsm[j] * nu[j];
}
if (c == 0){
    for (j = 0; j < n; j++){
        eps = epsm[j] * nu[j];
        fprintf(HELPPFILE, "%d ", eps);
    }
    count = count + 1;
}
}
fclose(HELPPFILE);

pH[nexp] = pI;

/* Cálculo da fração total e da mais comum */

for (i = 0; i <= nexp; i++){
    frac = 0;
    mc = 0;
    HELPPFILE = fopen("_helpfile.txt", "r");
    for (k = 0; k < count; k++){
        prod = 1.0;
        for (j = 0; j < n; j++){
            ex = pow(10.0, pH[i] - pKA[j]);
            fscanf(HELPPFILE, "%d", &eps);
            AS = (1 + nu[j]) / 2 - eps;
            BS = (nu[j] - 1) / 2 - eps;
            prod = prod * (1.0 / (1.0 + ex)) * (1.0 * AS * ex - 1.0 * BS);
        }
        auxiliar = prod;
        if (k > 0 && auxiliar > mc){
            mc = auxiliar;
            pmc = k;
        }
        frac = frac + prod;
    }
    fclose(HELPPFILE);
    phimc[i] = mc;
    phitot[i] = frac;
}

HELPPFILE = fopen("_helpfile.txt", "r");
for (k = 0; k < count; k++){
    for (j = 0; j < n; j++){
        fscanf(HELPPFILE, "%d", &eps);
    }
    if(k == pmc - 1)
        for (j = 0; j < n; j++){
            fscanf(HELPPFILE, "%d", &epsmc[j]);
        }
}
fclose(HELPPFILE);

/* Cálculo da solubilidade */

for (i = 0; i < nexp; i++)
    sl8[i] = pIsol * phitot[nexp] / phitot[i];

/*****
/*****

/*****
/* Cálculo da solubilidade da proteína pela Equação (3 - 32) */
/*****

for (i = 0; i < MAX; i++){
    /* Cálculo da função eta[j] - Equação (3 - 34) */
    prod = 1;
    npos = 0;

```

```

for (j = 0; j < n; j++){
    ex = 1.0 + pow(10.0, pH[i] - pKA[j]);
    ex0 = 1.0 + pow(10.0, pI - pKA[j]);
    prod = prod * (ex / ex0);
    npos = npos + (1 + nu[j]) / 2;
}
/* Cálculo da solubilidade */
s32[i] = pIsol * pow(10.0, (1.0 * npos) * (pI - pH[i])) * prod;
}
/*****
/*****

/*****
/* Impressão dos resultados na tela */
/*****

printf("\n\n RESULTADOS:\n\n");
printf(" Proteína: %s\n\n", name);
printf(" Tabela 1 - Parametros de calculo\n\n");
printf(" Residue | nu[j] | eps[j,k*] | pKA\n");
printf(" -----\n");
for (j = 0; j < n; j++){
    if (nu[j] > 0)
        printf(" %c | %d |", res[j], nu[j]);
    else
        printf(" %c | %d |", res[j], nu[j]);
    if (epsmc[j] < 0)
        printf(" %d |", epsmc[j]);
    else
        printf(" %d |", epsmc[j]);
    if (pKA[j] < 10.0)
        printf(" %.2lf\n", pKA[j]);
    else
        printf(" %.2lf\n", pKA[j]);
}
printf(" *Refere-se a configuracao mais comum.\n\n");
printf("\n\n Tabela 2 - Carga e fracao de moleculas neutras\n");
printf("\n pH | Z | phimc | phitot\n");
printf(" -----\n");
for (i = 0; i < nex; i++)
    printf(" %.2lf | %.2e | %.2e | %.2e\n", pH[i], Z[i], phimc[i],
        phitot[i]);
printf("\n\n Tabela 3 - Valores de solubilidade\n");
printf("\n pH | s18 | s32 | sexp\n");
printf(" -----\n");
for (i = 0; i < nex; i++)
    printf(" %.2lf | %.2e | %.2e | %.2e\n", pH[i], s18[i], s32[i],
        sexp[i]);

(void) time(&t2);
printf("\n\n Tempo de execução = %d segundos\n", (int) t2-t1);
printf("-----\n");
/*****
/*****

return 0;
}

```

Apêndice C – Expressão para o coeficiente de atividade de um soluto no *ensemble* de McMillan-Mayer

De fato, o *ensemble* de McMillan-Mayer parece ser o mais apropriado para aplicações como a que se pretende neste trabalho: sistemas solventes-solutos. Do ponto de vista formal da Termodinâmica este *ensemble* poderia ser pensado simplesmente como uma transformada de Legendre da equação fundamental – Equação (2 – 7) – muito particular na qual são permutadas: as seguintes coordenadas conjugadas: entropia e temperatura; e, quantidade de matéria do solvente e potencial químico de solvente de modo a definir uma nova função de estado característica deste *ensemble* cujas variáveis seriam: temperatura, volume, potencial químico do solvente e quantidade de matéria do soluto (MOLLERUP; BREIL, 2009a). Há claro outras formas de se estabelecer formalmente e de forma consistente a existência deste *ensemble*, mas o que o torna importante é o fato de se ter uma separação no sistema do que é solvente e o que é soluto. A característica principal deste *ensemble* é ter como variáveis fixas do sistema macroscópico o potencial químico do solvente e a quantidade de matéria do soluto.

Deste modo, sendo possível imaginar tal *ensemble* e estabelecê-lo de tal modo consistente com os outros *ensembles*: microcanônico, canônico, grande canônico, etc.; a validade desta teoria não a restringe a um limite da concentração de solutos, outrossim as aproximações que se seguirão no tratamento do problema acabarão por limitar a aplicação a soluções diluídas.

Uma aproximação comumente empregada – e, tão comumente empregada que por vezes sua identidade de aproximação é negligenciada – é a hipótese de que pode-se representar os problemas de interação entre muitos corpos de modo a somar contribuições devidas às interações de dois corpos, três corpos e assim em diante (HILL, 1986).

No caso de sistemas solventes-solutos, a hipótese da teoria de McMillan-Mayer (McMILLAN; MAYER, 1945) é que as interações entre as moléculas de soluto no interior de um solvente podem ser tratadas de modo análogo ao clássico problema de gases imperfeitos, por exemplo, usando uma expansão virial da pressão osmótica o que introduz as limitações próprias de uma expansão virial.

Partindo, então, da teoria de McMillan-Mayer e truncando a expansão virial da pressão osmótica após o segundo coeficiente, tem-se que (PITZER, 1973):

$$\Pi = \rho k_B T - \frac{1}{6} \sum_i \sum_j \rho_i \rho_j \int_0^\infty \left(\frac{\partial u_{ij}}{\partial r} \right) g_{ij}(r) 4\pi r^3 dr \quad (\text{C} - 1)$$

em que Π é a pressão osmótica, ρ é a densidade, ρ_i é a concentração do soluto i , ρ_j é a concentração do soluto j , u_{ij} é o potencial de interação entre os solutos i e j e $g_{ij}(r)$ é a função de distribuição radial que está diretamente relacionada ao potencial da força média. Ou ainda em termos do segundo coeficiente virial osmótico entre os solutos i e j , B_{ij} :

$$\frac{\Pi}{k_B T} = \rho + \sum_i \sum_j \rho_i \rho_j B_{ij} \quad (\text{C} - 2)$$

Ou ainda em termos de número de moléculas:

$$\frac{\Pi}{k_B T} = \frac{N}{V} + \frac{1}{V^2} \sum_i \sum_j N_i N_j B_{ij}, \quad \text{sendo } N = \sum_{i=1}^m N_i \quad (\text{C} - 3)$$

Por simples cálculo de propriedades de afastamento, é possível calcular o coeficiente de fugacidade, ϕ_k , de um dado componente k através de uma equação de estado, em termos osmóticos:

$$\ln \bar{\phi}_k = \int_\infty^v \left[\frac{1}{v} - \frac{N}{k_B T} \left(\frac{\partial \Pi}{\partial N_k} \right)_{T, \rho, N_{i \neq k}} \right] dv - \ln \left(\frac{\Pi}{\rho k_B T} \right) \quad (\text{C} - 4)$$

Para calcular este coeficiente de fugacidade, deve-se primeiro avaliar a derivada parcial interna a integração:

$$\frac{N}{k_B T} \left(\frac{\partial \Pi}{\partial N_k} \right) = \frac{1}{v} + \frac{1}{N v^2} \sum_i \sum_j \left[\frac{\partial (N_i N_j)}{\partial N_k} \right] B_{ij} \quad (\text{C} - 5)$$

Logo:

$$\begin{aligned} \ln \bar{\phi}_k &= \int_\infty^v \left\{ \frac{1}{v} - \frac{1}{v} - \frac{1}{N v^2} \sum_i \sum_j \left[\frac{\partial (N_i N_j)}{\partial N_k} \right] B_{ij} \right\} dv - \ln \left(\frac{\Pi}{\rho k_B T} \right) = \dots \\ &\dots = \frac{1}{V} \sum_i \sum_j \left[\frac{\partial (N_i N_j)}{\partial N_k} \right] B_{ij} - \ln \left(\frac{\Pi}{\rho k_B T} \right) \end{aligned} \quad (\text{C} - 6)$$

Resta ainda avaliar a derivada parcial do duplo somatório:

$$\begin{aligned}
\sum_i \sum_j \left[\frac{\partial(N_i N_j)}{\partial N_k} \right] B_{ij} &= \left[\frac{\partial(N_1 N_1)}{\partial N_k} \right] B_{11} + \dots + \left[\frac{\partial(N_1 N_k)}{\partial N_k} \right] B_{1k} + \dots + \left[\frac{\partial(N_1 N_m)}{\partial N_k} \right] B_{1m} + \dots \\
&\dots + \left[\frac{\partial(N_k N_1)}{\partial N_k} \right] B_{k1} + \dots + \left[\frac{\partial(N_k N_k)}{\partial N_k} \right] B_{kk} + \dots + \left[\frac{\partial(N_k N_m)}{\partial N_k} \right] B_{km} + \dots \\
&\dots + \left[\frac{\partial(N_m N_1)}{\partial N_k} \right] B_{m1} + \dots + \left[\frac{\partial(N_m N_k)}{\partial N_k} \right] B_{mk} + \dots + \left[\frac{\partial(N_m N_m)}{\partial N_k} \right] B_{mm} \quad (C-7)
\end{aligned}$$

Contudo:

$$\frac{\partial(N_i N_j)}{\partial N_k} = \begin{cases} 0, & \text{se } i \neq k, j \neq k \\ N_j, & \text{se } i = k, j \neq k \\ N_i, & \text{se } i \neq k, j = k \\ 2N_i, & \text{se } i = k, j = k \end{cases} \quad (C-8)$$

Portanto:

$$\begin{aligned}
\sum_i \sum_j \left[\frac{\partial(N_i N_j)}{\partial N_k} \right] B_{ij} &= 0 + \dots + N_1 B_{1k} + \dots + 0 + \dots + N_1 B_{k1} + \dots + 2N_k B_{kk} + \dots \\
&\dots + N_m B_{km} + \dots + 0 + \dots + N_m B_{mk} + \dots + 0 \quad (C-9)
\end{aligned}$$

Considerando ainda que $B_{ij} = B_{ji}$, ou seja, a matriz de interações é simétrica já que por uma hipótese bastante razoável o modo e a intensidade da interação de i com j deve ser o mesmo que de j com i , tem-se que:

$$\sum_i \sum_j \left[\frac{\partial(N_i N_j)}{\partial N_k} \right] B_{ij} = 2 \sum_j N_j B_{jk} \quad (C-10)$$

Assim, tem-se que o logaritmo natural do coeficiente de fugacidade de um soluto k é:

$$\ln \bar{\phi}_k = \frac{2}{V} \sum_j N_j B_{jk} - \ln \left(\frac{\Pi}{\rho k_B T} \right) \quad (C-11)$$

Isto implica que:

$$\ln \left(\frac{\bar{\phi}_k \Pi}{\rho k_B T} \right) = 2 \sum_j \rho_j B_{jk} \quad (C-12)$$

Ou ainda ,em termos de fugacidade:

$$\ln\left(\frac{\bar{f}_k}{\rho_k k_B T}\right) = 2 \sum_j \rho_j B_{jk} \quad (\text{C} - 13)$$

E, portanto, o coeficiente de atividade do soluto k pode ser escrito como:

$$\ln \gamma_k = 2 \sum_j \rho_j B_{jk} \quad (\text{C} - 14)$$

Apêndice D – Código-fonte do programa de ajuste de parâmetros para o cálculo da solubilidade de aminoácidos e compostos β -lactâmicos

```

/*****
/* ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO
/* DEPARTAMENTO DE ENGENHARIA QUÍMICA
/* GRUPO DE ENGENHARIA DE BIOPROCESSOS
/* SOLUBILIDADE DE AMINOÁCIDOS
*****/

#include <stdlib.h>
#include <stdio.h>
#include <math.h>
#include <string.h>
#include <dirent.h>

#define NPM 100
#define MAX 100
#define NITERMAX 1000
#define VEC 40

double FO(double par[MAX], int n, char dfn[VEC]);

int main(){
    int i, N, LS;
    double X1[NPM], X2[NPM], X3[NPM], XK[NPM], XMIN[NPM], XMAX[NPM], EMAX[NPM];
    double EFAC, Y, Y1, YSTAR, YTEST, YBASE;
    char nfile[VEC], dfn[VEC], get[VEC];
    FILE *ENT;

/*****
1) Para usar este programa dois arquivos de entrada são requeridos. A extensão
destes arquivos precisa ser ".txt" e estes arquivos devem ser salvos nos
mesmos diretório e pasta do executável.

2) O primeiro arquivo deve conter as estimativas iniciais dos parâmetros a serem
ajustados bem como seus limites inferior e superior:

Numero_de_parametros:
2
Estimativa_inicial:
0.032
-16.0
Valores_mínimo_e_máximo:
0.0      100.0
-100.0   0.0

3) O segundo arquivo deve conter os valores de temperatura, de pKA, do pI e da
solubilidade no pI bem como os dados experimentais de solubilidade com os
quais se pretende ajustar os parâmetros do modelo. Abaixo um exemplo para a
ampicilina, dados de Santana et al. (2010):

ndata:
4

AMP1
Temperature(K):
283.06
pKA1_pKA2:
2.27      7.66

```

```

pI_solpI:
5.5      13.45
nexp:
5
pH_sexp:
5.5      13.45
6.0      14.07
6.5      15.96
7.0      18.73
7.5      39.59

```

```

AMP2
Temperature(K):
288.01
pKA1_pKA2:
2.24      7.61
pI_solpI:
5.5      14.75
nexp:
5
pH_sexp:
5.5      14.75
6.0      16.35
6.5      17.19
7.0      19.29
7.5      47.00

```

```

AMP3
Temperature(K):
292.95
pKA1_pKA2:
2.14      7.45
pI_solpI:
5.5      16.88
nexp:
5
pH_sexp:
5.5      16.88
6.0      17.17
6.5      18.50
7.0      27.48
7.5      49.80

```

```

AMP4
Temperature(K):
298.03
pKA1_pKA2:
2.14      7.31
pI_solpI:
5.5      20.89
nexp:
5
pH_sexp:
5.5      20.89
6.0      21.61
6.5      22.27
7.0      28.39
7.5      51.37

```

4) Os autores não têm qualquer responsabilidade sobre o uso deste programa.

```

/*****

```

```

printf("-----");
printf("\n");
printf("\n\t ESCOLA POLITECNICA DA UNIVERSIDADE DE SAO PAULO \n");
printf("\n\t DEPARTAMENTO DE ENGENHARIA QUIMICA \n");

```



```

P1011:
    X2[i] = X1[i];
    XK[i] = 0.5 * XK[i];
    if (XK[i] < 0.0) goto P1015;
    if (XK[i] < 0.5 * EFAC) XK[i] = 0.5 * EFAC;
    goto P1018;
P1015:
    if (fabs(XK[i]) < 0.5 * EFAC) XK[i] = -0.5 * EFAC;
P1018:
    Y1 = YTEST;
P103:
    YTEST = Y1;
}
if (LS == 1) goto P350;
YBASE = YTEST;
if (YBASE < YSTAR) goto P300;
for (i = 0; i < N; i++)
    if (fabs(XK[i]) > 0.5 * EFAC) goto P100;
goto P600;
P300:
    for (i = 0; i < N; i++){
        X1[i] = X3[i] + 2.0 * (X2[i] - X3[i]);
        if (X1[i] < XMIN[i]) X1[i] = XMIN[i];
        if (X1[i] > XMAX[i]) X1[i] = XMAX[i];
        X3[i] = X2[i];
        X2[i] = X1[i];
    }
    YSTAR = FO(X1, N, dfn);
    LS = 1;
    goto P100;
P350:
    LS = 0;
    if (YTEST >= YBASE) goto P500;
    YBASE = YTEST;
    goto P300;
P500:
    for (i = 0; i < N; i++){
        X1[i] = X3[i];
        X2[i] = X1[i];
    }
    YSTAR = YBASE;
    goto P100;
P600:
    Y = FO(X1, N, dfn);

    /*****
    /* Impressão dos resultados
    /*****
    printf("\n\n Valor otimo da funcao objetivo: %.3lf\n\n", Y);
    printf("\t\t Variaveis | Valores de otimo\n");
    printf("\t\t ----- \n");
    for (i = 0; i < N; i++)
        printf("\t\t XOTM[%d] | %e\n", i + 1, X1[i]);
    printf("-----\n");

    return 0;
}

double FO(double par[MAX], int n, char dfn[VEC]){
    int i, j, k, ndata, nexp;
    int cont[MAX];
    double pI, spI, pKA1, pKA2, exp, aminus, aplus, tau;
    double fipI, lnf, dtau, FUN, tol, lam, T, penalty;
    double fi[MAX], sol[NITERMAX], sexp[MAX], pH[MAX], s[MAX], F[MAX];
    char get[MAX];
    FILE *ENT;

    tol = pow(10.0, -6.0);

```

```

penalty = 1000000.0;

FUN = 0.0;
ENT = fopen(dfn, "r");
fscanf(ENT, "%s", get);
fscanf(ENT, "%d", &ndata);
for (k = 0; k < ndata; k++){
    fscanf(ENT, "%s", get);
    fscanf(ENT, "%s", get);
    fscanf(ENT, "%lf", &T);
    fscanf(ENT, "%s", get);
    fscanf(ENT, "%lf", &pKA1);
    fscanf(ENT, "%lf", &pKA2);
    fscanf(ENT, "%s", get);
    fscanf(ENT, "%lf", &pI);
    fscanf(ENT, "%lf", &spI);
    fscanf(ENT, "%s", get);
    fscanf(ENT, "%d", &nexp);
    fscanf(ENT, "%s", get);
    for (i = 0; i < nexp; i++){
        fscanf(ENT, "%lf", &pH[i]);
        fscanf(ENT, "%lf", &sexp[i]);
    }

    lam = par[0] + par[1] / T;
    exp = pow(10.0, pI - pKA1);
    aminus = exp / (1.0 + exp);
    exp = pow(10.0, pI - pKA2);
    aplus = 1.0 / (1.0 + exp);
    fipI = aminus * aplus;

    F[k] = 0.0;
    cont[k] = 0;
    for (i = 0; i < nexp; i++){
        exp = pow(10.0, pH[i] - pKA1);
        aminus = exp / (1.0 + exp);
        exp = pow(10.0, pH[i] - pKA2);
        aplus = 1.0 / (1.0 + exp);
        fi[i] = aminus * aplus;
        lnfi = log(fipI / fi[i]);
        sol[0] = spI;
        /* Método de Newton-Raphson para cálculo da Equação (4 - 23) */
        for (j = 0; j < NITERMAX; j++){
            tau = log(sol[j]/spI)-lnfi+lam*(sol[j]-spI);
            dtau = 1.0 / sol[j] + lam;
            sol[j+1] = sol[j] - tau / dtau;
            if (fabs(sol[j+1] - sol[j]) < tol){
                s[i] = sol[j+1];
                cont[k] = cont[k] + 1;
                j = NITERMAX + 1;
            }
        }
        F[k] = F[k] + pow(((sexp[i] - s[i]) / sexp[i]), 2.0);
    }
    F[k] = F[k] + penalty * (nexp - cont[k]);
    FUN = FUN + F[k];
}
fclose(ENT);
FUN = sqrt(FUN / (1.0 * ndata * nexp));

return FUN;
}

```

Anexo A – Valores de pK_A usados nos cálculos do quarto capítulo

No capítulo quatro foram calculadas curvas de solubilidade para diversos aminoácidos, alguns peptídeos e duas moléculas com grupos β -lactâmicos. Para tanto foram utilizados valores de pK_A destas moléculas que estão publicados na literatura. A Tabela A.1 apresenta estes valores e suas respectivas referências.

Tabela A.1 - Valores de pK_A para aminoácidos, peptídeos e moléculas com grupos β -lactâmicos estudados no capítulo quatro.

Molécula	T / K	pK_{A1}	pK_{A2}	Referências
DL-Alanina	298,15	2,35	9,87	Tseng et al. (2009)
DL-Metionina	298,15	2,12	9,28	Fuchs et al. (2006)
L-Isoleucina	298,15	2,32	9,76	Tseng et al. (2009)
L-Leucina	298,15	2,33	9,74	Tseng et al. (2009)
DL-Fenilalanina	298,15	2,58	9,24	Tseng et al. (2009)
L-Serina	298,15	2,21	9,15	Tseng et al. (2009)
Tirosina ^a	298,15	2,20	9,11	Greenstein e Winitz (1961)
Glicina	298,15	2,34	9,60	Greenstein e Winitz (1961)
Diglicina	298,15	3,12	8,17	Greenstein e Winitz (1961)
Triglicina	298,15	3,26	7,91	Greenstein e Winitz (1961)
Tetraglicina	298,15	3,05	7,75	Greenstein e Winitz (1961)
Pentaglicina	298,15	3,05	7,70	Greenstein e Winitz (1961)
Hexaglicina	298,15	3,05	7,60	Greenstein e Winitz (1961)
Ampicilina	283,1	2,27	7,66	Santana et al. (2010)
	288,1	2,24	7,61	Santana et al. (2010)
	293,1	2,14	7,45	Santana et al. (2010)
	298,1	2,14	7,31	Santana et al. (2010)
Ácido 6-aminopenicilânico	283,1	1,98	5,34	Santana et al. (2010)
	288,1	2,04	5,25	Santana et al. (2010)
	293,1	2,10	5,18	Santana et al. (2010)
	298,1	2,20	4,83	Santana et al. (2010)

^a O valor do pK_{A3} da tirosina utilizado foi 10,07 (GREENSTEIN; WINITZ, 1961).