

LUÍS FERNANDO MERCIER FRANCO

**MODELAGEM TERMODINÂMICA DO EQUILÍBRIO  
SÓLIDO-LÍQUIDO EM SISTEMAS CONTENDO  
PROTEÍNAS: EFEITO DO pH**

Trabalho de Conclusão de Curso em Engenharia  
Química Apresentado ao Departamento de  
Engenharia Química da Escola Politécnica da  
Universidade de São Paulo

Professor Orientador: Doutor Pedro de Alcântara Pessoa Filho

São Paulo

2010

## **Agradecimentos**

Ao Professor Doutor Pedro de Alcântara Pessôa Filho, orientador deste trabalho, que, além de ter sido incansável no intuito de me levar a compreender os fundamentos básicos para o desenvolvimento deste projeto, tem me ensinado o que é ciência, e como se dá a sua aplicação em Engenharia Química.

Aos amigos e colegas do Laboratório de Engenharia Bioquímica da Escola Politécnica, pelo carinho e pela acolhida. Em especial, a doutoranda Kelly Cristina Nascimento Alves, por toda a gentileza e auxílio na obtenção dos dados experimentais.

E, a Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP – que me concedeu uma bolsa de iniciação científica sem a qual seria inviável levar a diante este trabalho.

## RESUMO

Neste trabalho é apresentado um modelo para a curva de solubilidade de proteínas em solução aquosa em função do pH da solução, baseado em ferramentas da termodinâmica clássica e considerando os diferentes estados de ionização que moléculas protéicas podem assumir. Também foi estudada a modelagem descrita por Grönwall em seu artigo publicado no início da década de 1940, e deste estudo, resultou uma extensão à sua teoria que se provou ser idêntica à modelagem proposta inicialmente. Para efeito de comparação, foram modeladas curvas de solubilidade utilizando o modelo de atividade estendido de Debye-Hückel para soluções eletrolíticas.

Foram obtidos dados experimentais da solubilidade da lisozima de clara de ovo de galinha para alguns valores de pH na região do ponto isoelétrico. Tanto para a insulina suína como para lisozima de clara de ovo de galinha, o modelo desenvolvido descreveu com grande precisão o fenômeno observado experimentalmente. A aplicação do modelo permite concluir que a utilização de cálculos de estados de ionização é uma ferramenta eficaz na modelagem termodinâmica de soluções protéicas.

## ABSTRACT

In this work, it is presented a model for the protein solubility curve in aqueous solutions as a function of the solution pH, based on classical thermodynamic tools and considering the different ionization states that protein molecules may assume. It was also studied the model described by Grönwall in his article published in the early 1940's, and from this study, it was developed an extension of his theory which, as proved in this work, is identical to the initially proposed modeling. For comparison, solubility curves were modeled using the Debye-Hückel extended activity model for electrolytic solutions.

It was obtained experimental hen egg white lysozyme solubility data for some pH values in the region of the isoelectric point. So much for porcine insulin as for the hen egg white lysozyme, the developed model describes with high accuracy the phenomenon experimentally observed. The model application allows us to conclude that the utilization of the ionization states calculus is an effective tool in the thermodynamic modeling of protein solutions.

## Sumário

1. Introdução .....	1
2. Materiais e Métodos .....	7
2.1. Métodos computacionais.....	7
2.2. Curva de solubilidade da lisozima da clara de ovo de galinha .....	9
2.2.1. Reagentes .....	9
2.2.2. Equipamentos.....	9
2.2.3. Preparação das soluções-tampão .....	9
2.2.4. Procedimento experimental .....	10
3. Modelo Estendido de Debye-Hückel .....	11
3.1. Introdução .....	11
3.2. Desenvolvimento teórico.....	12
4. Modelagem Termodinâmica Proposta.....	16
4.1. Introdução .....	16
4.2. Cálculo das frações de moléculas eletricamente neutras .....	18
5. Extensão da Equação de Linderstrøm-Lang e Grönwall .....	21
5.1. Introdução .....	21
5.2. Desenvolvimento teórico.....	21
5.3. Solução Parabólica da Curva de Solubilidade .....	26
6. Identidade entre o Modelo Proposto e a Equação Estendida de Linderstrøm-Lang e Grönwall .....	28
7. Aplicação dos Modelos Expostos.....	31
7.1. Solubilidade da Insulina Suína.....	31
7.2. Solubilidade da Lisozima de clara de ovo de galinha.....	36
8. Conclusões.....	40
9. Referências Bibliográficas.....	41
Apêndice A – Equações de Henderson - Hasselbach.....	44
Apêndice B – Código – fonte do programa de cálculo dos modelos expostos.....	47

# 1. Introdução

Dentre as etapas de um processo biotecnológico, aquela dedicada a purificação de bioprodutos tende a ser determinante, em geral, para o custo do processo, seguindo a tendência dos processos de separação da indústria química tradicional. A produção de proteínas tem lugar de destaque no rol de bioprodutos de interesse, pois muitos antibióticos, hormônios e enzimas pertencem a esta classe de biomoléculas e possuem notório interesse industrial por suas propriedades terapêuticas e aplicações na indústria alimentícia. No entanto, proteínas a serem comercializadas para estes fins exigem altos graus de pureza que somente são obtidos pelas etapas de purificação. Como afirmado por Bončina et al. (2008), estudos de separação de proteínas, especificamente, têm sido motivados por um aumento da demanda por proteínas puras nas indústrias farmacêutica e correlatas.

Muitas moléculas biológicas – ou comumente chamadas *biomoléculas* -, e particularmente proteínas, são destruídas pelo calor, não podem ser destiladas, são geradas a baixas concentrações e têm uma estrutura tridimensional que pode ser alterada durante a purificação resultando na perda da atividade biológica. Por isso, bioprodutos, tais como proteínas, devem ser recuperados, purificados e concentrados em operações com membranas, etapas de adsorção, cromatografia, cristalização e outras técnicas que não requerem calor ou pH extremo para alcançar o fracionamento (Ladisich, 2001).

Em geral, nos processos biotecnológicos, as correntes de saída dos biorreatores seguem para as operações de *downstream* (etapas de separação que visam concentrar e dispor o produto final). As primeiras operações de *downstream*, por exemplo: a clarificação e a centrifugação, têm por objetivo isolar a proteína de

## 1. Introdução

---

interesse das células que a produziram. Após estas primeiras etapas de purificação, seguem-se as operações de baixa resolução. A precipitação e a extração em sistemas de duas fases aquosas são exemplos destas operações que tendem a concentrar a proteína-alvo, preparando-a para os próximos estágios de separação. Por fim, a proteína, ora na concentração obtida pelas operações anteriores, é tratada por operações de alta resolução, caracterizadas principalmente pela técnica de cromatografia, a fim de concentrar a proteína à pureza especificada para seu uso. Vale ressaltar que esta última etapa é extremamente dispendiosa, e quanto mais eficientes forem as etapas precedentes no que tange a pureza da proteína-alvo, menos custosa será a etapa posterior.

O escopo deste trabalho está centrado na operação de precipitação de proteínas cujos fundamentos são as relações do equilíbrio sólido-líquido (ESL) destes sistemas. A solubilidade de uma proteína em solução é governada pelos mesmos princípios fundamentais, ou seja, condições de equilíbrio termodinâmico, que regem o comportamento de qualquer substância no ESL. Contudo, algumas diferenças devem ser levadas em conta na descrição deste fenômeno e no projeto destas operações.

Proteínas são formadas por cadeias de aminoácidos através de ligações peptídicas, cada aminoácido pode ser considerado um monômero que ao possuir um grupo ionizável não ligado, tal como ocorre em alguns aminoácidos naturais: tirosina (Tyr), lisina (Lys), cisteína (Cys), histidina (His), arginina (Arg), ácido aspártico (Asp) e ácido glutâmico (Glu) e também em resíduos terminais, confere a proteína uma característica de polieletrólito. É importante, então, considerar a definição de polieletrólitos tal qual definida por Pessoa Filho e Maurer (2008): polieletrólitos são

## 1. Introdução

---

polímeros com muitas unidades monoméricas eletrolíticas – neste caso, aminoácidos que possuem grupos ionizáveis não ligados.

Para muitos processos industriais e mesmo naturais, informações a respeito de equilíbrios de fases de soluções eletrolíticas são necessárias; e necessárias também são modificações nos modelos tradicionais de coeficiente de atividade quando se trata de descrever equilíbrios de fases em sistemas que contém eletrólitos (Prausnitz et al., 1999).

Sendo assim, tendo que a pressão do sistema não influencia de modo significativo o ESL, tem-se que três variáveis o influenciam diretamente: a temperatura, o pH e a força iônica; estas duas últimas tornam-se importantes devido à característica eletrolítica das proteínas.

Na verdade, a força iônica não é uma variável que represente de modo abrangente os possíveis efeitos de íons em soluções protéicas. Isto se deve ao fato de que a força iônica do modo como é definida não contempla uma diferenciação entre os equilíbrios que se estabelecem entre os íons e alguns resíduos ionizáveis da cadeia protéica. De fato, a variação da força iônica de uma solução protéica altera o valor da solubilidade da proteína em solução, mas não da mesma forma para qualquer espécie de íon. Portanto, seria mais prudente acrescentar ao conjunto de variáveis que influenciam o ESL a espécie química dos íons presentes em solução.

O objetivo deste trabalho é estabelecer um modelo capaz de prever a influência do pH no ESL de proteínas mantendo-se temperatura e força iônica constantes.

Sabe-se, por observações experimentais que a variação do pH de um sistema altera o valor da solubilidade de proteínas em solução, e a solubilidade de proteínas por sua vez atinge seu valor mínimo em um pH denominado ponto isoelétrico (pI), no qual a carga líquida da proteína é nula.

## 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 eqüina 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 pelo tempo, 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. Quando da exposição da modelagem aqui tratada, retornar-se-á a este trabalho que muito contribuiu ao entendimento do fenômeno estudado.

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 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 obter parâmetros a partir de um ajuste de

## 1. Introdução

---

um conjunto de dados experimentais. Vale observar que resultados expressivos foram alcançados neste trabalho, de modo particular, o fato de que a presença do ânion tiocianato não altera a forma da curva de solubilidade da insulina, mas a desloca em direção a menores valores de pH e de solubilidade.

Contudo, os trabalhos mais recentes se concentraram 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.

Destaca-se neste grupo o trabalho de Pinho et al. (1994), que aplicaram o modelo UNIFAC associado ao termo de Debye-Hückel para aminoácidos como glicina, D-alanina, L-serina, e DL-valina. Também o trabalho de Khoshkbarchi e Vera (1996), no qual foi desenvolvido um modelo simplificado de esfera rígida perturbada para calcular o coeficiente de atividade de aminoácidos e de peptídeos, permite prever a curva de solubilidade destes em função do pH. Já o modelo NRTL foi aplicado tanto por Pradhan e Vera (1998) quanto Tseng et al. (2009) para aminoácidos. Já Seyfi et al. (2009) aplicaram a equação de estado SAFT para a predição da curva de solubilidade da DL-metionina.

Entretanto, o primeiro trabalho dedicado exclusivamente à modelagem e predição da curva de solubilidade de proteína 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

## *1. Introdução*

---

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 do virial.

No presente trabalho, propomos um novo modelo baseado no cálculo de estados de ionização e em ferramentas de termodinâmica clássica, bem como uma extensão à teoria de Linderstrøm-Lang e Grönwall. Vale ressaltar de antemão que tanto este novo modelo proposto quanto a extensão desenvolvida resultam na mesma equação para o fenômeno apesar das distintas abordagens.

## 2. Materiais e Métodos

### 2.1. Métodos computacionais

O código-fonte do programa em linguagem C está apresentado no Apêndice B. Os métodos numéricos utilizados se resumem a interpolação linear, o método de Newton-Raphson cujo uso será apresentado na próxima seção quando se deduzirá o algoritmo do modelo de Debye-Hückel e o método de busca unidirecional de mínimos de função pela Seção Áurea.

As Tabelas 2.1.1. e 2.1.2. apresentam os resíduos considerados para cada proteína considerada bem como os seus valores de  $pK_A$  que foram obtidos a partir da literatura e são indispensáveis para execução dos cálculos mais rigorosos.

**Tabela 2.1.1. – Resíduos ionizáveis considerados e valores de  $pK_A$  da insulina suína. Valores extraídos de Kaarsholm et al. (1990).**

Aminoácido	Grupo Ionizável	$pK_A$
A1 – Gly	$\alpha$ -NH <sub>2</sub>	7.38
A4 – Glu	$\gamma$ -COOH	4.37
A14 – Tyr	Fenol-OH	10.20
A17 – Glu	$\gamma$ -COOH	4.37
A19 – Tyr	Fenol-OH	10.20
A21 – Asn	$\alpha$ -COOH	3.45
B1 – Phe	$\alpha$ -NH <sub>2</sub>	7.38
B5 – His	Imidazol-NH	5.98
B10 – His	Imidazol-NH	5.98
B 13 – Glu	$\gamma$ -COOH	4.37
B16 – Tyr	Fenol-OH	10.20
B21 – Glu	$\gamma$ -COOH	4.37
B22 – Arg	Guanídio-NH <sub>2</sub>	11.10
B26 – Tyr	Fenol-OH	10.20
B29 – Lys	$\epsilon$ -NH <sub>2</sub>	8.52
B30 –Ala	$\alpha$ -COOH	3.45

## 2. Materiais e Métodos

**Tabela 2.1.2. – Resíduos ionizáveis considerados e valores de  $pK_A$  a 25 °C da lisozima de clara de ovo de galinha. Valores extraídos de Kuramitsu e Hamaguchi (1980).**

Aminoácido	Grupo Ionizável	$pK_A$
$\alpha$ -NH <sub>2</sub>	$\alpha$ -NH <sub>2</sub>	7.9
Lys1	$\epsilon$ -NH <sub>2</sub>	10.8
Arg5	Guanídio-NH <sub>2</sub>	12.0
Glu7	$\gamma$ -COOH	2.6
Lys13	$\epsilon$ -NH <sub>2</sub>	10.5
Arg14	Guanídio-NH <sub>2</sub>	12.0
His15	Imidazol-NH	5.8
Asp18	$\beta$ -COOH	2.0
Tyr20	Fenol-OH	10.3
Arg21	Guanídio-NH <sub>2</sub>	12.0
Tyr23	Fenol-OH	9.8
Lys33	$\epsilon$ -NH <sub>2</sub>	10.6
Glu35	$\gamma$ -COOH	6.1
Arg45	Guanídio-NH <sub>2</sub>	12.0
Asp48	$\beta$ -COOH	4.3
Asp52	$\beta$ -COOH	3.4
Tyr53	Fenol-OH	12.1
Arg61	Guanídio-NH <sub>2</sub>	12.0
Asp66	$\beta$ -COOH	1.6
Arg68	Guanídio-NH <sub>2</sub>	12.0
Arg73	Guanídio-NH <sub>2</sub>	12.0
Asp87	$\beta$ -COOH	2.1
Lys96	$\epsilon$ -NH <sub>2</sub>	10.8
Lys97	$\epsilon$ -NH <sub>2</sub>	10.3
Asp101	$\beta$ -COOH	4.5
Arg112	Guanídio-NH <sub>2</sub>	12.0
Arg114	Guanídio-NH <sub>2</sub>	12.0
Lys116	$\epsilon$ -NH <sub>2</sub>	10.4
Asp119	$\beta$ -COOH	2.5
Arg125	Guanídio-NH <sub>2</sub>	12.0
Arg128	Guanídio-NH <sub>2</sub>	12.0
$\alpha$ -COOH	$\alpha$ -COOH	3.1

## **2.2. Curva de solubilidade da lisozima da clara de ovo de galinha**

### **2.2.1. Reagentes**

Neste experimento de determinação da curva de solubilidade da lisozima da clara de ovo de galinha os reagentes utilizados foram: cloreto de sódio (Reatec, P. A.), bicarbonato de sódio (Research Becto Chemical, P. A.),  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$  (Reagen, P. A.), água destilada deionizada Milli-Q (Millipore) e lisozima de clara de ovo de galinha (Sigma, min 99%).

### **2.2.2. Equipamentos**

Neste experimento os equipamentos utilizados foram: espectrofotômetro UV-Visível (Beckman DU-530), centrífuga (Eppendorf), banho-térmico (Láctea) com controlador de temperatura (Julabo), pHmetro (Micronal) e balança (Mettler Toledo, AB204S).

### **2.2.3. Preparação das soluções-tampão**

Foram preparadas soluções-tampão com valores de pH=9,6, pH = 9,8, pH = 10,0, pH = 10,2, pH = 10,4, pH = 10,6, pH = 10,8, pH = 11,0, pH = 11,2, pH = 11,4, pH = 11,6 e pH = 11,8 a partir de quatro soluções: uma de hidróxido de sódio 0,1 M, uma de cloreto de sódio 1 M (para equalizar a força iônica), uma de bicarbonato de sódio 0,05 M e outra de  $\text{Na}_2\text{HPO}_4$  0,05 M.

A Tabela 2.2.3.1. apresenta os valores de volume usados no preparo de cada uma das soluções-tampão.

## 2. Materiais e Métodos

**Tabela 2.2.3.1. – Volumes de hidróxido de sódio a 0,1 M adicionados no preparo das soluções-tampão.**

Tampão	pH	V NaOH / mL
NaHCO <sub>3</sub>	9.6	10.0
0.05 M	9.8	15.2
(para cada solução-tampão 100 mL desta solução foram adicionados)	10.0	21.4
	10.2	27.6
	10.4	33.0
	10.6	38.2
	10.8	42.4
	11.0	45.4
	11.2	53.0
	11.4	58.3
Na <sub>2</sub> HPO <sub>4</sub>	11.0	8.2
0.05 M	11.2	12.6
(para cada solução-tampão 100 mL desta solução foram adicionados)	11.4	18.2
	11.6	27.0
	11.8	38.8

### 2.2.4. Procedimento experimental

Foram adicionadas quantidades não definidas de lisozima de clara de ovo de galinha a 1,3 mL de cada solução-tampão em tubos Eppendorf de 2 mL. Colocaram-se, então, as amostras após agitação vigorosa em banho térmico a  $25,0 \pm 0,1$  °C por 24 horas para que os sistemas atingissem o equilíbrio termodinâmico. Caso não houvesse a formação de precipitado, lisozima era adicionada e o procedimento, repetido. Ao se observar a formação de precipitado, o sistema era centrifugado a 13000 rpm por 15 minutos. Uma alíquota do sobrenadante era retirada, e sua concentração protéica determinada por medida de absorbância a 280 nm. Cada sistema foi feito ao menos em duplicata.

### 3. Modelo Estendido de Debye-Hückel

#### 3.1. Introdução

De fato, como descrito por Pessoa Filho (2002), a equação de Debye-Hückel não deve ser aplicada à modelagem de polieletrólitos tais como proteínas, porque nestas moléculas as cargas são distribuídas ao longo da cadeia e, portanto, não são cargas pontuais como suposto por Debye-Hückel.

Contudo, nesta modelagem considera-se uma região de  $pH$  em torno do ponto isoelétrico ( $pI$ ) na qual são inerentes duas condições favoráveis a aplicação da equação de Debye-Hückel: carga líquida próxima a zero, pois  $Z(pI) = 0$ ; e, região de mínima solubilidade que significa baixas concentrações de proteínas em solução. Isto significa em primeiro lugar que os efeitos da carga, tais como “desnovelamento” da cadeia protéica com exposição de resíduos outrora blindados, podem ser desprezados em uma região próxima ao ponto isoelétrico; e em segundo lugar, baixas concentrações de proteína em solução significam que os termos de curto alcance podem ser desprezados e, portanto, há prevalência do termo de longo alcance, que no caso de soluções eletrolíticas pode ser descrito por uma equação estendida de Debye-Hückel.

Assim, o modelo de Debye-Hückel foi utilizado não para verificar sua aplicabilidade à modelagem de soluções protéicas, mas sim para confrontar o modelo proposto com um modelo usual em modelagem de soluções eletrolíticas e será referido no texto sempre como abordagem clássica.

### 3.2. Desenvolvimento teórico

Pode-se escrever a atividade da proteína como o produto entre a solubilidade e o coeficiente de atividade da mesma como mostrado na equação (3.2. - 1).

$$a_p = \gamma S_p \quad (3.2. - 1)$$

Na qual, pode-se calcular o coeficiente de atividade através da equação estendida de Debye-Hückel (Pitzer, 1973):

$$\ln \gamma = -A_\phi Z_{pH}^2 \left[ \frac{\sqrt{I}}{1 + b\sqrt{I}} + \frac{2}{b} \ln(1 + b\sqrt{I}) \right] \quad (3.2. - 2)$$

Sendo que a constante universal empírica  $b = 1,2 \text{ kg}^{1/2} \text{ mol}^{-1/2}$  é usada para todos os eletrólitos. Já  $A_\phi$  é o parâmetro de Debye-Hückel para o coeficiente osmótico e é dado por:

$$A_\phi = \frac{1}{3} (2\pi N_A \rho_s)^{1/2} \left( \frac{e^2}{4\pi \epsilon_0 \epsilon_r kT} \right)^{3/2} \quad (3.2. - 3)$$

Em que  $N_A$  é a constante de Avogadro,  $\rho_s$  é a densidade do solvente em  $\text{g cm}^{-3}$ ,  $e$  é a carga eletrônica ( $e = 1.60218 \times 10^{-19} \text{ C}$ ),  $\epsilon_0$  é a permissividade do espaço livre ( $\epsilon_0 = 8.85419 \times 10^{-12} \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2}$ ),  $\epsilon_r$  é a constante dielétrica do solvente,  $k$  é a constante de Boltzmann e  $T$  é a temperatura absoluta. Já a carga líquida em função do pH pode ser calculada como:

$$Z_{pH} = \sum_{j+} \alpha_+ - \sum_{j-} \alpha_- \quad (3.2. - 4)$$

Sendo que  $\alpha_+$  e  $\alpha_-$  são dados pelas equações (A - 10) e (A - 16) do Apêndice A.

### 3. Modelo Estendido de Debye-Hückel

---

E, a força iônica é dada pela equação (3.2. - 5) descrita abaixo:

$$I = \frac{1}{2} \sum_{i=1}^n m_i Z_i^2 \quad (3.2. - 5)$$

Contudo, no ponto isoelétrico a carga média da proteína é nula, portanto, o termo de Debye-Hückel se anula e o coeficiente de atividade se torna unitário.

$$\ln a_p(pI) = \ln s_p(pI) \quad (3.2. - 6)$$

Considerando que a atividade da proteína seja constante com a variação de pH, isto implica que esta é sempre igual à solubilidade da proteína no ponto isoelétrico, e, portanto, vale que:

$$\ln a_p(pH) = \ln s_p(pH) + \ln \gamma(pH) = \ln s_p(pI) \quad (3.2. - 7)$$

Assim:

$$\ln s_p(pH) = \ln s_p(pI) + A_\phi Z_{pH}^2 \left[ \frac{\sqrt{I}}{1 + b\sqrt{I}} + \frac{2}{b} \ln(1 + b\sqrt{I}) \right] \quad (3.2. - 8)$$

Esta equação prevê a curva de solubilidade em função do pH através do cálculo da atividade da proteína na fase líquida pela equação estendida de Debye-Hückel. Para resolvê-la é necessário se utilizar de um método numérico, pois a força iônica depende da solubilidade naquele pH. Optou-se, neste trabalho, pelo Método de Newton-Raphson para executar o método iterativo que é exigido pela equação (3.2. - 8). Para aplicar este método arbitrou-se uma função  $f(s_p)$  definida aqui como:

$$f(s_p) = -\ln s_p(pH) + \ln s_p(pI) + A_\phi Z_{pH}^2 \left[ \frac{\sqrt{I}}{1 + b\sqrt{I}} + \frac{2}{b} \ln(1 + b\sqrt{I}) \right] \quad (3.2. - 9)$$

### 3. Modelo Estendido de Debye-Hückel

---

Assim, ao calcular o  $s_p$  que zera esta função  $f(s_p)$ , saber-se-á que é a própria solubilidade que se deseja calcular. Para aplicar o Método de Newton-Raphson, basta derivar a função  $f(s_p)$  e aplicar o seguinte algoritmo de resolução:

$$s_{i+1} = s_i - \frac{f(s_i)}{f'(s_i)} \quad (3.2. - 10)$$

Para calcular  $f'(s)$ , deve-se considerar que a força iônica depende da solubilidade. Para isto, observando a equação (3.2. - 5), vê-se que a força iônica é dada pelo somatório das contribuições de todos os íons presentes na solução, por isso aqui cabe um artifício de dividir estas contribuições em duas macro-contribuições: aquela devida a todos os íons presentes na solução, exclusive os grupos ionizáveis presentes nas moléculas da proteína em questão ( $I_0$ ), e que se considera constante com o pH; e aquela que engloba apenas os grupos ionizáveis presentes nas moléculas da proteína ( $I_p$ ).

$$I = I_0 + I_p = I_0 + \frac{1}{2} Z_{pH}^2 S_P \quad (3.2. - 11)$$

Considerando a função  $w$  tal que:

$$w = \sqrt{I} = \sqrt{I_0 + \frac{1}{2} Z_{pH}^2 S_P} \quad (3.2. - 12)$$

Sendo:

$$w' = \frac{1}{4} Z_{pH}^2 I^{-\frac{1}{2}} \quad (3.2. - 13)$$

Então, aplicando-se as devidas regras de derivação, tem-se:

### 3. Modelo Estendido de Debye-Hückel

---

$$f'(s_p) = -\frac{1}{s_p} + A_\phi Z_{pH}^2 \left[ \frac{3w' + 2w'bw}{(1 + bw)^2} \right] \quad (3.2. - 14)$$

Substituindo as equações (3.2. - 12) e (3.2. - 13) na equação (3.2. - 14), tem-se que:

$$f'(s_p) = -\frac{1}{s_p} + A_\phi Z_{pH}^4 \left[ \frac{1.5 + b\sqrt{I}}{2\sqrt{I}(1 + b\sqrt{I})^2} \right] \quad (3.2. - 15)$$

Então o algoritmo deve ser:

$$s_{i+1} = s_i - \left( \frac{\ln \frac{s_p(pI)}{s_i} + A_\phi Z_{pH}^2 \left[ \frac{\sqrt{I}}{1 + b\sqrt{I}} + \frac{2}{b} \ln(1 + b\sqrt{I}) \right]}{-\frac{1}{s_i} + A_\phi Z_{pH}^4 \left[ \frac{1.5 + b\sqrt{I}}{2\sqrt{I}(1 + b\sqrt{I})^2} \right]} \right) \quad (3.2. - 16)$$

## 4. Modelagem Termodinâmica Proposta

### 4.1. Introdução

A condição de equilíbrio entre duas fases ( $\alpha$  e  $\beta$ ) com  $m$  componentes é dada por

(Prausnitz et al., 1999):

$$p^\alpha = p^\beta \tag{4.1. - 1}$$

$$T^\alpha = T^\beta \tag{4.1. - 2}$$

$$\begin{cases} \mu_1^\alpha = \mu_1^\beta \\ \vdots \\ \mu_m^\alpha = \mu_m^\beta \end{cases} \tag{4.1. - 3}$$

Considerando o equilíbrio sólido-líquido (ESL) de sistemas contendo proteínas, sendo que os efeitos da pressão podem ser desprezados, que as temperaturas nas duas fases são iguais, e que na fase sólida só há proteína, tem-se que para a proteína a relação de equilíbrio pode ser descrita por:

$$\mu_p^L(T, x) = \mu_p^S(T) \tag{4.1. - 4}$$

Contudo, adotando-se a lei de Henry, que é válida para baixas concentrações de soluto, hipótese que não é distante dos valores observados da solubilidade de proteínas, pode se escrever o potencial químico na fase líquida por:

$$\mu_p^L(T, x) = \mu_p^{L^{ref}}(T, x^{ref}) + RT \ln x_p \tag{4.1. - 5}$$

#### 4. Modelagem Termodinâmica Proposta

---

Entretanto, variações de pH não alteram o potencial químico da fase sólida, o que resulta pela relação (4.1. – 4) que o mesmo não altera também o potencial químico da fase líquida. Com isso, pela relação (4.1. – 5), sendo o potencial químico de referência constante com o pH, uma contradição emerge: a variação do pH não poderia alterar o valor da solubilidade, porém a observação experimental mostra o contrário. Assim, deve-se lembrar que pelo princípio da eletroneutralidade, a fase sólida só possui moléculas protéicas eletricamente neutras, e, portanto, o equilíbrio de fases deve se estabelecer apenas entre as moléculas presentes na fase sólida e as moléculas eletricamente neutras em solução. Ou seja, como só moléculas eletricamente neutras podem estar na fase sólida, o equilíbrio de fases depende exclusivamente da fração de moléculas eletricamente neutras em solução, pois só estas podem precipitar. Assim, tanto a equação (4.1. – 4) quanto a equação (4.1. – 5) devem ser aplicadas não ao conjunto de todas as moléculas de proteína em solução, mas somente as moléculas protéicas eletricamente neutras. Portanto, a concentração das moléculas de proteína eletricamente neutras em solução é constante com o pH e pode ser interpretada como a multiplicação da fração de moléculas eletricamente neutras pela concentração total de proteínas que é a própria solubilidade. Assim, uma relação de inversa proporcionalidade é definida: quanto maior a fração de moléculas neutras, menor será a solubilidade de uma proteína em solução. Portanto, no ponto isoelétrico, no qual há o mínimo de solubilidade, tem-se o ponto máximo de fração de moléculas eletricamente neutras. A partir disto é possível escrever a equação que rege o comportamento da solubilidade de uma proteína em função do pH da solução.

$$\phi(pH)S_p(pH) = \lambda = \text{constante}$$

(4.1. – 6)

#### 4. Modelagem Termodinâmica Proposta

---

Na equação (4.1. – 6), o parâmetro  $\lambda$  pode ser um parâmetro de ajuste do modelo a um conjunto de valores experimentais tendo se estabelecido uma função objetivo, ou ainda dado pelo produto entre a solubilidade no ponto isoelétrico e a fração de moléculas eletricamente neutras em solução também no  $pI$ , neste caso vale a seguinte equação:

$$s_p(pH) = \frac{\phi(pI)s_p(pI)}{\phi(pH)} \quad (4.1. – 7)$$

Sendo a equação do modelo – equação (4.1. – 7) – extremamente simples, a única dificuldade reside no cálculo das frações de moléculas eletricamente neutras ( $\phi(pH)$ ).

#### 4.2. Cálculo das frações de moléculas eletricamente neutras

Dentre os aminoácidos que formam através de ligações peptídicas as proteínas, há aqueles que possuem resíduos com grupos ionizáveis além dos grupos amina e ácido carboxílico terminais na cadeia. Todos estes participam de equilíbrios químicos com íons  $H^+$  presentes na solução; então, ao se alterar o pH da solução, estes equilíbrios são alterados.

Assim, há várias possibilidades de combinações de estados ionizados destes grupos ionizáveis que resultam em uma molécula de proteína eletricamente neutra. É importante notar que, dada a fração de uma configuração de uma molécula neutra, todas as outras frações de configurações possíveis que também resultem em moléculas neutras são proporcionais à fração da configuração dada, pois todas as configurações participam de equilíbrios entre si.

#### 4. Modelagem Termodinâmica Proposta

---

Desta observação, implica-se que basta identificar uma configuração protéica cuja soma dos estados ionizados resulte em carga nula para se descrever o equilíbrio de fases em relação ao pH na região do ponto isoelétrico onde esta consideração é predominante em relação a outros possíveis fenômenos.

Contudo, é necessário contabilizar a contribuição de cada grupo ionizável via equação de Henderson-Hasselbach, equação (4.2. – 1):

$$\alpha_j(pH) = \begin{cases} \frac{10^{pH-pK_{A_j}}}{1 + 10^{pH-pK_{A_j}}}, & \text{se } v_j = -1 \\ \frac{1}{1 + 10^{pH-pK_{A_j}}}, & \text{se } v_j = +1 \end{cases} \quad (4.2. - 1)$$

Na qual  $pK_{A_j}$  é o co-logaritmo da constante de equilíbrio do grupo ionizável do resíduo  $j$ ,  $v_j$  é estado carregado potencial do resíduo  $j$  (-1 para grupos ácidos e +1 para grupos básicos) e  $\alpha_j$  é a fração ionizada para o pH dado. Para grupos cujo estado de ionização  $\varepsilon$  na dada configuração seja neutro, a sua fração é calculada tal que a soma de sua fração e da fração ionizada tenha como resultado o valor unitário.

Assim, a função  $\Phi_{P_n}$  pode ser escrita como o produto de cada fração de cada grupo ionizável  $j$  naquele estado  $\varepsilon$  (-1, 0 ou +1) na configuração neutra  $k$  em uma proteína com  $n$  grupos ionizáveis. Uma representação matemática possível para esta função pode ser escrita como:

$$\phi_k(pH) = \prod_{j=1}^n \alpha'_{k_j}(pH) \quad (4.2. - 2)$$

Na qual, a fração  $\alpha'_j(pH)$  é dada por:

#### 4. Modelagem Termodinâmica Proposta

---

$$\alpha'_j(pH) = \begin{cases} \alpha_j(pH), & \text{se } |\varepsilon_{jk}| = 1 \\ 1 - \alpha_j(pH), & \text{se } |\varepsilon_{jk}| = 0 \end{cases} \quad (4.2. - 3)$$

Substituindo a equação (4.2. - 1) na equação (4.2. - 3) e a resultante na equação (4.2. - 2), pode-se escrever que:

$$\phi_k(pH) = \prod_{j=1}^n \left\{ \left( \frac{1}{1 + 10^{pH-pK_{A_j}}} \right) \left[ \left( \frac{1 + \nu_j}{2} - \nu_j |\varepsilon_{jk}| \right) 10^{pH-pK_{A_j}} - \left( \frac{\nu_j - 1}{2} - \nu_j |\varepsilon_{jk}| \right) \right] \right\}, \text{ com } \sum_{j=1}^n \varepsilon_{jk} = 0 \quad (4.2. - 4)$$

A equação (4.2. - 4) junto à equação (4.1. - 7) constituem o modelo termodinâmico capaz de prever a razão entre as frações de moléculas neutras de uma proteína em função do pH na região do ponto isoelétrico mantendo fixas temperatura e força iônica. Substituindo a equação (4.2. - 4) na equação (4.1. - 7), obtém-se para a curva de solubilidade:

$$\frac{s_p(pH)}{s_p(pI)} = \prod_{j=1}^n \left\{ \left( \frac{1 + 10^{pH-pK_{A_j}}}{1 + 10^{pI-pK_{A_j}}} \right) \left[ \frac{\left( \frac{1 + \nu_j}{2} - \nu_j |\varepsilon_{jk}| \right) 10^{pI-pK_{A_j}} - \left( \frac{\nu_j - 1}{2} - \nu_j |\varepsilon_{jk}| \right)}{\left( \frac{1 + \nu_j}{2} - \nu_j |\varepsilon_{jk}| \right) 10^{pH-pK_{A_j}} - \left( \frac{\nu_j - 1}{2} - \nu_j |\varepsilon_{jk}| \right)} \right] \right\},$$

$$\text{com } \sum_{j=1}^n \varepsilon_{jk} = 0 \quad (4.2. - 8)$$

## 5. Extensão da Equação de Linderstrøm-Lang e Grönwall

### 5.1. Introdução

Segundo Grönwall (1941), na fase líquida há moléculas de proteínas com muitas configurações  $k$  distintas que possuem carga líquida  $i$ , e também há certas configurações que tem carga líquida nula e são chamadas configurações neutras. Além disso, estas últimas são aquelas que determinam o equilíbrio de fases já que elas estão em equilíbrio com a fase sólida na qual só há moléculas neutras pelo princípio da eletroneutralidade. É também afirmado que dado este equilíbrio a atividade das moléculas neutras não se altera com a variação do pH e é proporcional às atividades de qualquer tipo de configuração neutra.

Neste mesmo trabalho, é atribuída a Linderstrøm-Lang o desenvolvimento de uma teoria de predição da solubilidade de proteínas com a variação do pH na região do ponto isoelétrico. Este desenvolvimento apresenta um balanço de cargas de um modo implícito; contudo, como atestado por Grönwall, o escasso conhecimento sobre as estruturas protéicas impediram, àquele tempo, uma formulação mais pormenorizada do problema.

### 5.2. Desenvolvimento teórico

A base desta abordagem descrita por Grönwall e atribuída a Linderstrøm-Lang reside na hipótese de que a molécula de proteína de configuração  $k$  e carga nula participa do seguinte equilíbrio com os íons  $H^+$  e com sua forma de carga  $i$ :

### 5. Extensão da Equação de Linderstrøm-Lang e Grönwall

---



Note-se que o valor de  $i$  pode ser negativo. A constante do equilíbrio químico descrita pela equação (5.2. - 1) pode ser expressa como:

$$K_{ik} = \frac{a_{P_{k^*}^i}}{a_{P_k} a_{H^+}^i} \quad (5.2. - 2)$$

Sabendo que a atividade da molécula de proteína de configuração  $k$  e carga  $i$  é dada pelo produto entre sua concentração e seu coeficiente de atividade, tem-se que:

$$x_{P_{k^*}^i} = a_{P_k} a_{H^+}^i \frac{K_{ik}}{\gamma_{P_{k^*}^i}} \quad (5.2. - 3)$$

Entretanto, a solubilidade da proteína é soma das concentrações de todas as moléculas de carga  $i$ . Porém, a concentração das moléculas de carga  $i$  é a soma das concentrações das moléculas de configuração  $k$ . Assim, a solubilidade pode ser escrita como:

$$s_p(pH) = \sum_i \sum_k \left( a_{P_k} a_{H^+}^i \frac{K_{ik}}{\gamma_{P_{k^*}^i}} \right) \quad (5.2. - 4)$$

Derivando a equação (5.2. - 4) com respeito à atividade do íon  $H^+$ , considerando que atividade da molécula neutra é constante para qualquer configuração  $k$ , e ainda que o coeficiente de atividade da proteína de carga  $i$  independe da atividade do íon  $H^+$ ; tem-se que:

$$\frac{ds_p}{da_{H^+}} = \sum_i \left[ i \cdot a_{H^+}^{i-1} \sum_k \left( a_{P_k} \frac{K_{ik}}{\gamma_{P_{k^*}^i}} \right) \right] \quad (5.2. - 5)$$

### 5. Extensão da Equação de Linderstrøm-Lang e Grönwall

---

Sendo:

$$\frac{d \log s_p(pH)}{d(pH)} = - \frac{a_{H^+}}{s_p(pH)} \frac{ds_p}{da_{H^+}} \quad (5.2. - 6)$$

Substituindo as equações (5.2. - 4) e (5.2. - 5) na equação (5.2. - 6) e reescrevendo em termos de pH e das concentrações das proteínas de carga  $i$ , tem-se que:

$$\frac{d \log s_p}{d(pH)} = - \frac{\sum_i (i \cdot x_i)}{\sum_i x_i} \quad (5.2. - 7)$$

Contudo, o termo do lado direito da equação (5.2. - 7) está associado ao conceito de valência média estabelecido por Linderstrøm-Lang (1924):

$$\bar{i} = \frac{\sum_i (i \cdot x_i)}{\sum_i x_i} \quad (5.2. - 8)$$

Os autores não possuíam um modo mais adequado de construir a curva de solubilidade dada pela equação (5.2. - 7) do que aproximar por uma função linear o perfil de carga líquida da proteína obtido experimentalmente e integrá-lo, valendo-se da analogia entre o conceito de carga líquida ( $Z$ ) e de valência média ( $\bar{i}$ ). Entretanto, hoje já se tem uma descrição analítica do perfil de carga líquida de proteínas, que pode ser escrita como:

$$Z(pH) = \sum_{j=1}^n (v_j \alpha_j) \quad (5.2. - 9)$$

Substituindo a equação (4.2. - 1) na equação (5.2. - 9) e a resultante na equação (5.2. - 7), tem-se que:

### 5. Extensão da Equação de Linderstrøm-Lang e Grönwall

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

(5.2. – 10)

Integrando a equação (5.2. – 10) de um ponto de referência, por exemplo, do ponto isométrico (pl) até um pH qualquer tem-se a seguinte relação:

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

(5.2. – 11)

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

(5.2. – 12)

Fazendo a seguinte substituição:

$$u = 10^{pH-pK_{A_j}} \quad \text{tal que} \quad \frac{du}{u} = \ln 10 d(pH)$$

(5.2. – 13)

Tem-se que:

$$\log \frac{s_P(pH)}{s_P(pI)} = -\frac{1}{2} \sum_{j=1}^n \left[ \frac{(1 + v_j)}{\ln 10} \int_{uI}^u \left[ \frac{1}{u(1 + u)} \right] du + \frac{v_j(1 - v_j)}{\ln 10} \int_{uI}^u \left[ \frac{1}{(1 + u)} \right] du \right]$$

(5.2. – 14)

$$\log \frac{s_P(pH)}{s_P(pI)} = -\frac{1}{2} \sum_{j=1}^n \left\{ \frac{(1 + v_j)}{\ln 10} \left[ \int_{uI}^u \left( \frac{1}{u} \right) du - \int_{uI}^u \left( \frac{1}{1 + u} \right) du \right] + \frac{v_j(1 - v_j)}{\ln 10} \int_{uI}^u \left[ \frac{1}{(1 + u)} \right] du \right\}$$

(5.2 – 15)

### 5. Extensão da Equação de Linderstrøm-Lang e Grönwall

$$\log \frac{s_p(pH)}{s_p(pI)} = -\frac{1}{2} \sum_{j=1}^n \left[ (1 + v_j) \log \left( \frac{10^{pH-pK_{A_j}}}{10^{pI-pK_{A_j}}} \right) - [v_j(1 - v_j) - (1 + v_j)] \log \left( \frac{1 + 10^{pH-pK_{A_j}}}{1 + 10^{pI-pK_{A_j}}} \right) \right] \quad (5.2 - 16)$$

Sendo:

$$v_j(1 - v_j) - (1 + v_j) = v_j - v_j^2 - 1 - v_j = -(1 + v_j^2) = -2 \quad (5.2. - 17)$$

Resultando em:

$$\log \frac{s_p(pH)}{s_p(pI)} = \sum_{j=1}^n \left[ \log \left( \frac{1 + 10^{pH-pK_{A_j}}}{1 + 10^{pI-pK_{A_j}}} \right) - \frac{(1 + v_j)}{2} (pH - pI) \right] \quad (5.2. - 18)$$

Ainda é possível reescrever a equação (5.2. - 18) como:

$$\log \frac{s_p(pH)}{s_p(pI)} = n_+(pI - pH) + \sum_{j=1}^n \left[ \log \left( \frac{1 + 10^{pH-pK_{A_j}}}{1 + 10^{pI-pK_{A_j}}} \right) \right] \quad (5.2. - 19)$$

Na qual  $n_+$  representa o número de resíduos ionizáveis que na sua forma carregada apresentam carga positiva. A equação (5.2. - 19) é uma expressão analítica que a partir de dados da estrutura protéica e de um dado de solubilidade permite prever a curva de solubilidade de uma dada proteína em função do pH na região do ponto isoelétrico mantendo fixas a temperatura e a força iônica do sistema.

É possível provar que as equações (4.2. - 8) e (5.2. - 19) são idênticas - a prova disso é descrita na seção 6 - e, assim, tem-se que apesar das distintas abordagens a descrição do fenômeno é a mesma. Note-se que um passo fundamental no desenvolvimento da Equação (5.2. - 5) é admitir que, por hipótese, a atividade da

### 5. Extensão da Equação de Linderstrøm-Lang e Grönwall

---

forma neutra *não* depende do pH; essa hipótese também é a base da equação (4.1. – 7). Deve-se registrar o fato de que, pelo melhor do nosso conhecimento, esta é a primeira vez que uma expressão analítica para o cálculo da solubilidade de proteínas em função do pH é descrita, mesmo no limite em que se considera válida a lei de Henry.

### 5.3. Solução Parabólica da Curva de Solubilidade

Tanto Grönwall (1941) quanto Fredericq e Neurath (1950) adotaram a hipótese de que os perfis de carga líquida da  $\beta$ -lactoglobulina para um e da insulina para os outros dois eram lineares na região do ponto isoelétrico. Esta hipótese é bastante plausível e simplificadora, principalmente a época na qual estes trabalhos foram desenvolvidos em que a estrutura protéica bem como as constantes de ionização dos resíduos com grupos ionizáveis era pouco conhecida. De fato, ainda hoje a falta de dados das constantes de ionização na cadeia protéica é o maior desafio no uso do modelo rigoroso – equação (5.2. – 18). Por isso, o valor desta hipótese de perfil de carga líquida linear é notável.

Supondo, então, que o perfil de carga seja linear, é possível escrevê-lo como:

$$Z(pH) = -A(pH - pI), \quad \text{com } A > 0 \quad (5.3. - 1)$$

Substituindo (5.3. – 1) em (5.2. – 6):

$$\frac{d \log s_p(pH)}{d(pH)} = A(pH - pI) \quad (5.3. - 2)$$

Integrando (5.3. – 2) de um  $pH_0$  de referência até pH, tem-se que:

### 5. Extensão da Equação de Linderstrøm-Lang e Grönwall

---

$$\log \frac{s_p(pH)}{s_p(pH_0)} = K[pH^2 - 2pI(pH - pH_0) - pH_0^2], \quad \text{com } K = \frac{A}{2}$$

(5.3. – 3)

Ao se adotar como ponto de referência o ponto isoelétrico, a equação (5.3. – 3) reduz-se à:

$$\log \frac{s_p(pH)}{s_p(pI)} = K(pH - pI)^2$$

(5.3. – 4)

Contudo, apesar da solução quadrática, ou parabólica, para a equação estendida de Linderstrøm-lang e Grönwall expressa na equação (5.3. – 3) não requerer os dados das constantes de ionização dos resíduos com grupos ionizáveis, ela exige a definição do parâmetro  $A$  que fisicamente representa a inclinação da reta de aproximação da curva de carga líquida e que pode ser calculado por correlação a partir de um conjunto de dados experimentais de solubilidade ou de carga líquida.

## 6. Identidade entre o Modelo Proposto e a Equação Estendida de Linderstrøm-Lang e Grönwall

A equação final da modelagem termodinâmica proposta apresentada na seção 4 é a equação (4.2. – 8) dada por:

$$\frac{s_p(pH)}{s_p(pI)} = \prod_{j=1}^n \left\{ \left( \frac{1 + 10^{pH-pK_{A_j}}}{1 + 10^{pI-pK_{A_j}}} \right) \left[ \frac{\left( \frac{1 + v_j}{2} - v_j |\varepsilon_{jk}| \right) 10^{pI-pK_{A_j}} - \left( \frac{v_j - 1}{2} - v_j |\varepsilon_{jk}| \right)}{\left( \frac{1 + v_j}{2} - v_j |\varepsilon_{jk}| \right) 10^{pH-pK_{A_j}} - \left( \frac{v_j - 1}{2} - v_j |\varepsilon_{jk}| \right)} \right] \right\} \quad (6 - 1)$$

Aplicando-se o logaritmo decimal nos dois termos da equação (6 – 1), tem-se que:

$$\log \frac{s_p(pH)}{s_p(pI)} = \sum_{j=1}^n \left\{ \log \left( \frac{1 + 10^{pH-pK_{A_j}}}{1 + 10^{pI-pK_{A_j}}} \right) + \log \left[ \frac{\left( \frac{1 + v_j}{2} - v_j |\varepsilon_{jk}| \right) 10^{pI-pK_{A_j}} - \left( \frac{v_j - 1}{2} - v_j |\varepsilon_{jk}| \right)}{\left( \frac{1 + v_j}{2} - v_j |\varepsilon_{jk}| \right) 10^{pH-pK_{A_j}} - \left( \frac{v_j - 1}{2} - v_j |\varepsilon_{jk}| \right)} \right] \right\} \quad (6 - 2)$$

Pelo bem da simplificação desta dedução, adotar-se-ão as seguintes funções:

$$\eta_j(pH) = \log \left( \frac{1 + 10^{pH-pK_{A_j}}}{1 + 10^{pI-pK_{A_j}}} \right) \quad (6 - 3)$$

$$\theta_{jk}(pH) = \log \left[ \frac{\left( \frac{1 + v_j}{2} - v_j |\varepsilon_{jk}| \right) 10^{pI-pK_{A_j}} - \left( \frac{v_j - 1}{2} - v_j |\varepsilon_{jk}| \right)}{\left( \frac{1 + v_j}{2} - v_j |\varepsilon_{jk}| \right) 10^{pH-pK_{A_j}} - \left( \frac{v_j - 1}{2} - v_j |\varepsilon_{jk}| \right)} \right] \quad (6 - 4)$$

Assim, pode-se reescrever a equação (6 – 2) por:

6. Identidade entre o Modelo Proposto e a Equação Estendida de Linderstrøm-Lang e Grönwall

$$\log \frac{s_p(pH)}{s_p(pI)} = \sum_{j=1}^n [\eta_j(pH) + \theta_{jk}(pH)] \quad (6 - 5)$$

Observando-se a equação (6 – 5), tem-se que a função  $\eta_j$  independe do estado de ionização do resíduo  $j$ , porém a função  $\theta_{jk}$  depende. Por isso, deve-se analisar esta função para cada estado de ionização possível como mostrado na Tabela 6.1.

**Tabela 6. 1. – Análise da função  $\theta_{jk}$  para cada estado de ionização possível.**

$v_j$	$\varepsilon_{jk}$	$\theta_{jk}$	Quantidade de resíduos neste estado
-1	-1	$pI - pH$	$n_a$
	0	0	$n_b$
+1	+1	0	$n_c$
	0	$pI - pH$	$n_d$

Pela análise da Tabela 6.1. e considerando uma proteína com  $n$  resíduos com grupos ionizáveis, infere-se que a equação (6 – 5) pode ser reescrita como:

$$\log \frac{s_p(pH)}{s_p(pI)} = (n_a + n_d)(pI - pH) + \sum_{j=1}^n [\eta_j(pH)] \quad (6 - 6)$$

Contudo, as equações acima descritas são válidas apenas para configurações neutras, pois apenas as configurações neutras da fase líquida estão em equilíbrio com as moléculas de proteína da fase sólida. Para que uma configuração seja neutra vale a seguinte restrição:

$$\sum_{j=1}^n \varepsilon_{jk} = 0 \Leftrightarrow n_a = n_c \quad (6 - 7)$$

## 6. Identidade entre o Modelo Proposto e a Equação Estendida de Linderstrøm-Lang e Grönwall

---

Substituindo a restrição dada pela equação (6 – 7) na equação (6 – 6), tem-se que:

$$\log \frac{s_p(pH)}{s_p(pI)} = (n_c + n_d)(pI - pH) + \sum_{j=1}^n [\eta_j(pH)] \quad (6 - 8)$$

Ora, mas a soma de  $n_c$  e de  $n_d$ , não resulta em outra coisa senão no número de resíduos com grupos ionizáveis da estrutura protéica cujo  $\nu_j$  seja +1. Assim, substituindo  $n_c$  mais  $n_d$  por  $n_+$  e substituindo a equação (6 – 3) na equação (6 – 8), tem-se que:

$$\log \frac{s_p(pH)}{s_p(pI)} = n_+(pI - pH) + \sum_{j=1}^n \left[ \log \left( \frac{1 + 10^{pH - pK_{A_j}}}{1 + 10^{pI - pK_{A_j}}} \right) \right] \quad (6 - 9)$$

Ao se analisar a equação (6 – 9), observa-se que ela é idêntica a equação estendida de Linderstrøm-Lang e Grönwall apresentada pela equação (5.2. – 19). Com isso, tem-se que apesar das distintas formulações tanto a modelagem termodinâmica aqui proposta quanto a extensão a teoria de Linderstrøm-Lang e Grönwall resultam na mesma descrição do fenômeno. Além disso, a equação final dos modelos é extremamente simples e depende apenas de um ponto de referência e dos dados da estrutura da molécula de proteína cuja solubilidade se deseja estudar.

## 7. Aplicação dos Modelos Expostos

Para validar o modelo comparando-o com o cálculo feito a partir da teoria de Debye-Hückel, o critério que relaciona os pontos calculados pelo modelo com os pontos experimentais foi estabelecido a partir da seguinte função objetivo ( $FO$ ):

$$FO = \sqrt{\frac{1}{N} \sum_{k=1}^N \left( \frac{S_{p_k}^{calc} - S_{p_k}^{exp}}{S_{p_k}^{exp}} \right)^2}$$

(7 – 1)

Contudo, a equação (7 – 1) não é a melhor definição de função objetivo neste caso, pois a derivada da curva de solubilidade é razoavelmente grande nos dois lados do ponto isoelétrico, e isso permite que pequenos desvios na direção horizontal resultem em um grande aumento no valor da função objetivo. Uma solução possível seria substituir essa função objetivo por uma função definida pela soma do quadrado da distância do ponto experimental à curva do modelo, principalmente se fosse necessário calcular algum parâmetro do modelo via otimização da função objetivo estipulada. Não sendo este o escopo deste trabalho, optou-se pelo uso da  $FO$  definida na equação (7 – 1).

### 7.1. Solubilidade da Insulina Suína

As Figuras 7.1.1., 7.1.2. e 7.1.3. apresentam os gráficos da curva de solubilidade dada pelos pontos experimentais, pelo modelo proposto, pela abordagem clássica e também a curva obtida pela solução parabólica da Equação de Linderstrøm-Lang e Grönwall em três temperaturas distintas.

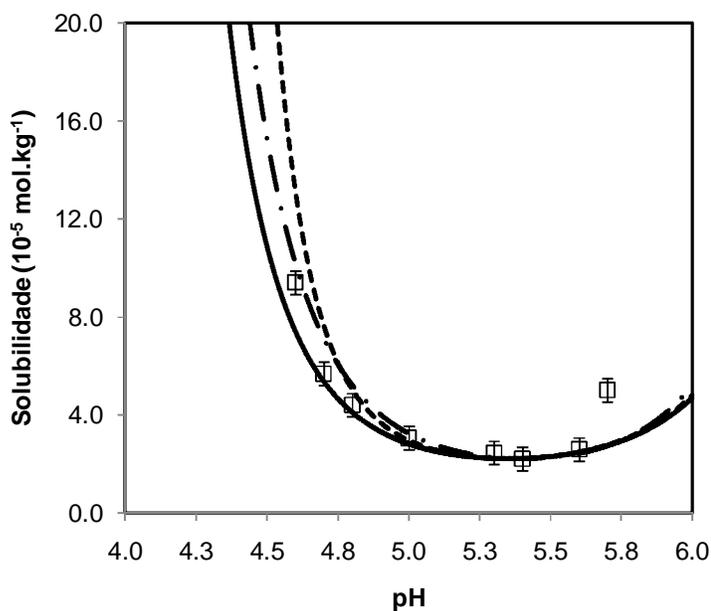


Figura 7.1.1. - Solubilidade da insulina suína em meio aquoso em função do pH a 25 °C. Dados experimentais ( $\square$ ), Tashima et al., 2009; modelo proposto (linha contínua), abordagem clássica (linha descontinua) e solução parabólica (linha semi-contínua).

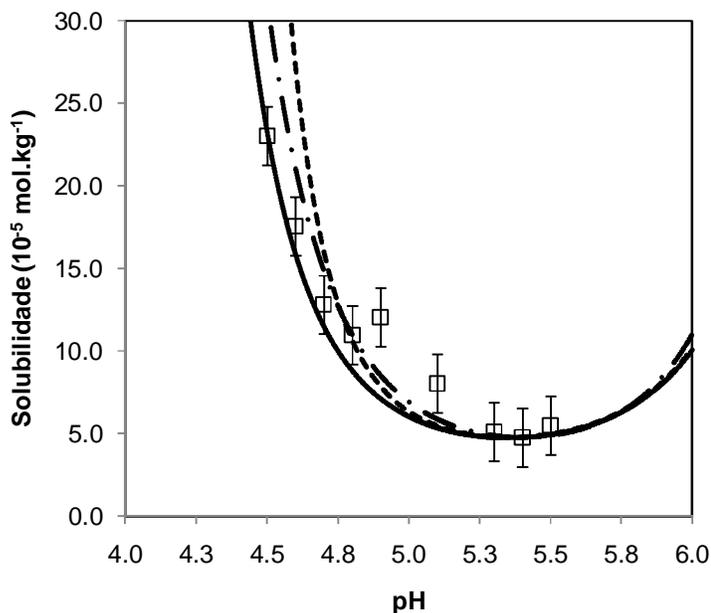
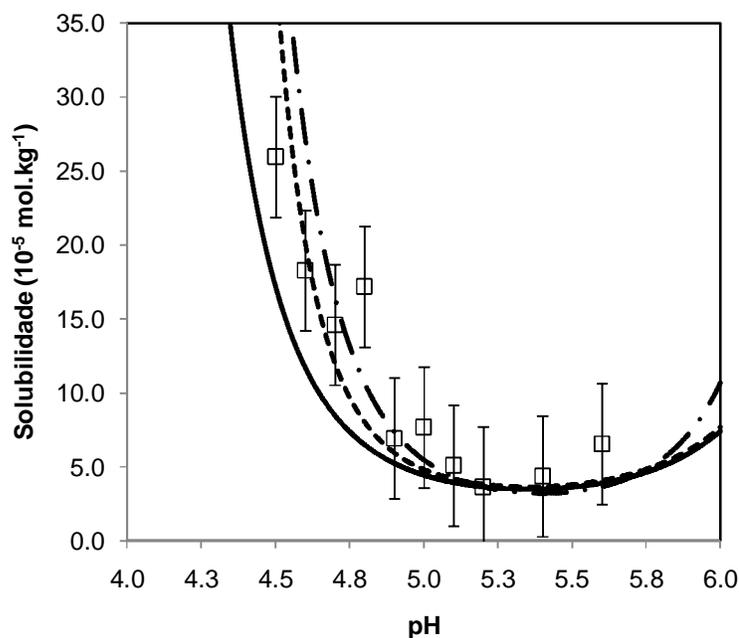


Figura 7.1.2. - Solubilidade da insulina suína em meio aquoso em função do pH a 15 °C. Dados experimentais ( $\square$ ), Tashima et al., 2009; modelo proposto (linha contínua), abordagem clássica (linha descontinua) e solução parabólica (linha semi-contínua).



**Figura 7.1.3. - Solubilidade da insulina suína em meio aquoso em função do pH a 5 °C. Dados experimentais (□), Tashima et al., 2009; modelo proposto (linha contínua), abordagem clássica (linha descontinua) e solução parabólica (linha semi-contínua).**

Os desvios da curva obtida pelo modelo proposto para valores de pH distantes do pI podem ser explicados por ao menos três importantes motivos: dois devidos a hipóteses do modelo e um devido a fenômenos intrínsecos a proteína. Sabendo que no pI a solubilidade é mínima, para valores de pH distantes do pI pode-se ter solubilidades suficientemente grandes para que a validade da lei de Henry – adotada na formulação do modelo - seja questionada, pois quanto mais distante do pI, maior a solubilidade e menor a validade da lei de Henry, já que há um aumento da concentração do soluto (proteína) na solução; e, também ao fato do modelo prever solubilidade infinita quando a fração de moléculas eletricamente neutras tende a zero.

$$\lim_{\Phi \rightarrow 0} s_p = \lim_{\Phi \rightarrow 0} \frac{\Phi(pI) s_p(pI)}{\Phi(pH)} = +\infty$$

## 7. Aplicação dos Modelos Expostos

Todavia, ainda há um fator de ordem natural que justifica o desvio apresentado para valores de pH distantes do  $pI$ : o desnovelamento da cadeia protéica. Este fenômeno pode ocorrer de maneira a expor grupos hidrófobos, resíduos antes não expostos, acrescentando relações de equilíbrio químico não existentes na região do  $pI$ , alterando o equilíbrio de fases, e, portanto, alterando o valor da solubilidade.

A Tabela 7.1.1. apresenta os valores obtidos para a função objetivo definida em (7 – 1) tanto para o modelo proposto quanto para a abordagem clássica para três diferentes temperaturas.

**Tabela 7.1.1. - Comparação entre os valores da Função Objetivo (FO) para cada abordagem (Proposta - P ou Clássica – C) em diferentes temperaturas para insulina suína.**

$T / ^\circ\text{C}$	Abordagem	FO
25	P	0.1857
	C	0.2535
15	P	0.1981
	C	0.5056
5	P	0.3595
	C	0.3064

Observando-se a Tabela 7.1.1., vê-se que o modelo apresentou melhores resultados do que o cálculo efetuado através da abordagem clássica no que tange a minimização da função objetivo, exceto para os dados a  $5\text{ }^\circ\text{C}$ .

Também é possível observar que a função objetivo sofreu aumento com a diminuição da temperatura. Isto ocorreu, principalmente, devido à dificuldade de obter dados experimentais de solubilidade a temperaturas baixas o que aumenta o desvio experimental fazendo com que se tenha um aumento do valor da função objetivo como a definida em (5 – 1). Um efeito secundário que também pressiona a

## 7. Aplicação dos Modelos Expostos

função objetivo é o fato dos dados de  $pK_A$  terem sido obtidos a uma temperatura específica o que tende a acrescentar erro no cálculo de solubilidades a temperaturas diferentes, ainda que o  $pK_A$  não varie de forma abrupta com a temperatura.

Outra observação que deve ser registrada é o fato da função objetivo para a abordagem clássica na temperatura de 15 °C divergir a um valor tão elevado em relação aos demais. Isto se deve a um ponto experimental específico: a um pH de 4.5 a solubilidade da insulina suína obtida experimentalmente é  $23.0 \times 10^{-5} \text{ mol kg}^{-1}$ , enquanto a calculada pelo modelo proposto é  $23.0 \times 10^{-5} \text{ mol kg}^{-1}$ , e a calculada pela abordagem clássica é  $52.0 \times 10^{-5} \text{ mol kg}^{-1}$ . Por conta do desvio deste ponto a função objetivo para a abordagem clássica nesta temperatura tem um aumento significativo.

A Tabela 7.1.2. apresenta os valores do parâmetro  $K$  da solução parabólica obtidos neste trabalho para a insulina suína comparados ao encontrado na literatura.

**Tabela 7.1.2. – Valores do parâmetro  $K$  da solução parabólica para a insulina suína.**

Referências	$T / ^\circ\text{C}$	$K$
Fredericq e Neurath (1950)	-	0.81
Neste trabalho	25	1.04
Neste trabalho	15	1.01
Neste trabalho	5	1.45

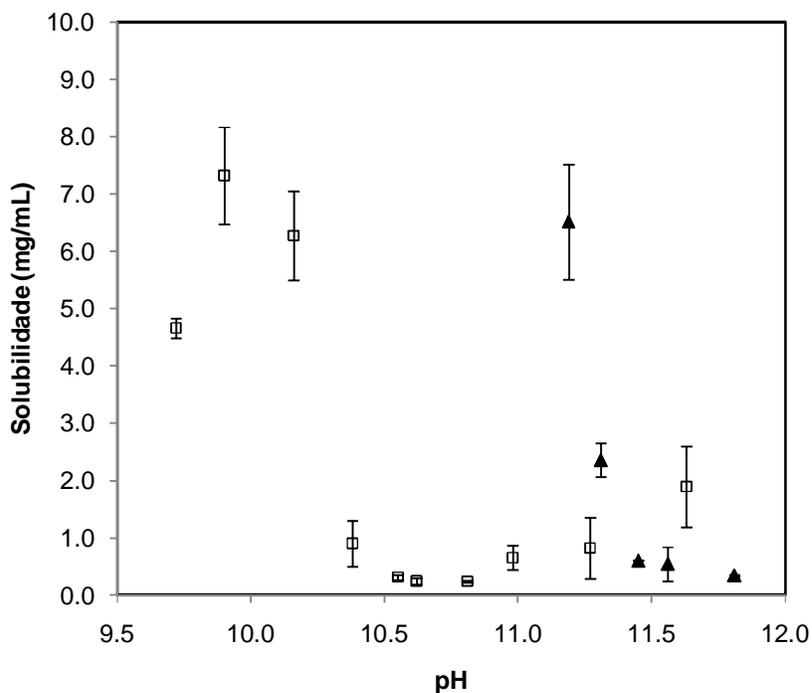
Pela Tabela 7.1.2., pode-se observar que há uma diferença pequena entre os valores do parâmetro  $K$  obtidos neste trabalho e por Fredericq e Neurath (1950). Por isso não se pode afirmar a que se deve esta diferença se é devido a diferenças entre procedimentos experimentais ou se é devido à diferença de método de cálculo do parâmetro. Apesar de no experimento realizado por Fredericq e Neurath (1950) haver a presença do ânion tiocianato, e este ânion interagir com grupos ionizáveis da cadeia protéica como observado pelos autores do trabalho; também estes

próprios autores afirmam que a presença deste ânion não altera a forma da curva que é basicamente determinada pelo parâmetro  $K$ . De qualquer modo é interessante observar a proximidade dos valores encontrados neste trabalho com o obtido por Fredericq e Neurath (1950), pelo menos no que diz respeito à ordem de grandeza do parâmetro em questão.

## **7.2. Solubilidade da Lisozima de clara de ovo de galinha**

Um modelo para ser validado necessita correlacionar dados experimentais de modo satisfatório, ou seja, um modelo deve buscar fidelidade ao fenômeno físico observado. Assim, na procura de dados de solubilidade de proteínas, poucos dados na região do  $pI$  foram encontrados. Este é o caso da lisozima de clara de ovo de galinha, que apesar de ter muitos estudos de solubilidade, nenhum deles se encontra na região do  $pI$ . Surgiu assim a necessidade de se obter estes dados experimentais para poder validar o modelo proposto. A escolha da lisozima de clara de ovo de galinha se deu em parte pela sua importância, mas também em parte por ter os valores de  $pK_A$  conhecidos.

A curva de solubilidade da lisozima de clara de ovo de galinha levantada neste trabalho está apresentada na Figura 7.2.1.



**Figura 7.2.1. – Solubilidade da lisozima da clara de ovo de galinha a 298.15 K e em solução 250 mmolar. Dados extraídos em solução-tampão a base de  $\text{Na}_2\text{HCO}_3$  (□), e dados obtidos a partir de soluções-tampão a base de  $\text{Na}_2\text{HPO}_4$  (▲).**

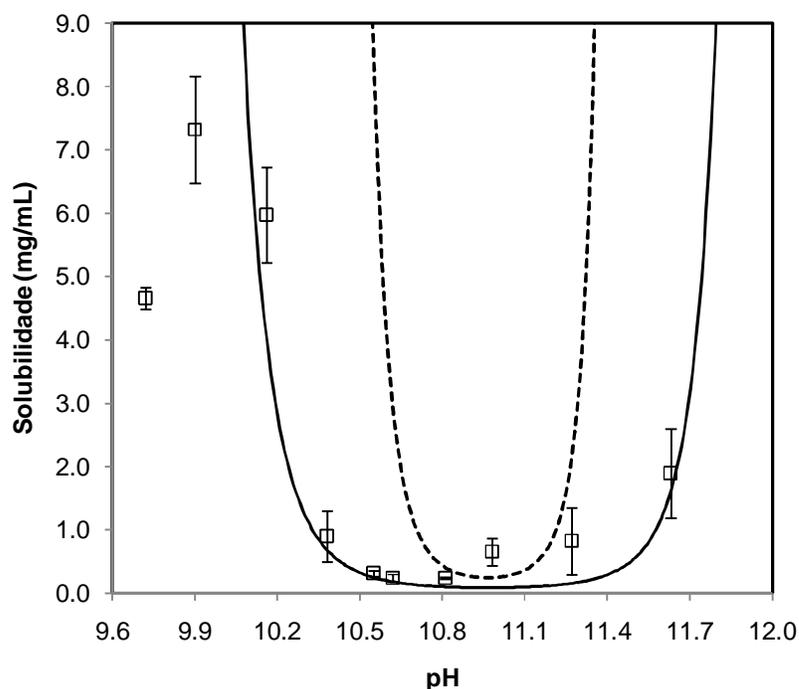
Pela Figura 7.2.1. é possível observar que a lisozima segue o comportamento esperado para proteínas atingindo o mínimo de solubilidade e um pH próximo ao pH isoelétrico calculado ( $pI = 10.97$ ) para soluções-tampão a base de bicarbonato de sódio. No entanto, para soluções-tampão a base de  $\text{Na}_2\text{HPO}_4$ , a lisozima se comporta de forma anômala, não apresentando mínimo de solubilidade no ponto isoelétrico. Dada a única diferença entre os experimentos ser o ânion do sal da solução tamponante, infere-se que o íon fosfato deve interagir com resíduos ionizáveis da lisozima alterando sua população de configurações eletricamente neutras e, por consequência, as suas propriedades físico-químicas como a solubilidade em soluções aquosas. Este fenômeno de interação entre o íon fosfato e

## 7. Aplicação dos Modelos Expostos

alguns resíduos de aminoácidos dispostos em cadeias protéicas tem sido observado em algumas proteínas e é denominado fosforização.

Esta interação por alterar a configuração protéica não pode ser prevista pelo modelo a menos que se conheça a constante de interação fosfato-resíduo.

Por conta de falta de dados a este respeito, modelou-se apenas a curva de solubilidade obtida em soluções tamponantes de bicarbonato de sódio. A Figura 7.2.2. apresenta a predição da curva de solubilidade da lisozima da clara de ovo de galinha em função do pH da solução.



**Figura 7.2.2.** – Curva de solubilidade de lisozima de clara de ovo de galinha em soluções aquosas com bicarbonato de sódio como sal tamponante a 298.15 K e 250 mmolar. Dados experimentais ( $\square$ ), modelo proposto (linha contínua) e abordagem clássica (linha descontinua).

O valor da função objetivo para cada modelo está apresentado na Tabela 7.2.1.

**Tabela 7.2.1. - Comparação entre os valores da Função Objetivo (*FO*) para cada abordagem (Proposta - P ou Clássica – C) em diferentes temperaturas para lisozima de clara de ovo de galinha.**

<i>T</i> / °C	Abordagem	<i>FO</i>
25	P	46.50
	C	>79.19 <sup>a</sup>

<sup>a</sup> Não foi possível calcular o valor da *FO* para todos os pontos experimentais, pois o método numérico aplicado divergiu em alguns pontos. Assim, sabe-se apenas um valor mínimo possível de ser calculado.

Pela Figura 7.2.2. e pela Tabela 7.2.1., observa-se que a modelagem proposta aproxima de forma muito mais satisfatória os dados experimentais do que a abordagem clássica de Debye-Hückel. Ainda assim, a modelagem é restrita a cerca de uma unidade de pH em torno do ponto isoelétrico pelos motivos expressos anteriormente. Já o insucesso da abordagem clássica se deve, principalmente, ao efeito da carga. Diferente de sais inorgânicos ou mesmo aminoácidos, a lisozima de clara de ovo de galinha tem uma amplitude muito maior de variação de carga com a variação de pH devido ao número de grupos ionizáveis em sua cadeia. A proximidade entre estes grupos é uma realidade que difere muito da hipótese de cargas pontuais de Debye-Hückel.

## 8. Conclusões

A modelagem termodinâmica proposta neste trabalho apresentou expressivos resultados quando comparada a abordagem clássica de Debye-Hückel.

Tanto para a insulina suína quanto para a lisozima de clara de ovo de galinha, o modelo aproximou melhor os pontos experimentais do que a referida abordagem comparativa. Verificou-se que o modelo tem uma validade restrita a região próxima ao ponto isoelétrico.

Entretanto, este modelo tem uma importância ímpar na modelagem termodinâmica para equilíbrios de fases de soluções protéicas devido, em primeiro lugar, a sua construção baseada em fundamentos da termodinâmica clássica e, em segundo lugar, por sua aplicabilidade.

Contudo, restam questões a respeito do fenômeno da interação de grupos ionizáveis da cadeia protéica e íons em solução; e, também a respeito da modelagem da curva em regiões distantes do ponto isoelétrico.

## 9. Referências Bibliográficas

BONČINA, M.; REŠČIČ, J.; VLACHY, V., Solubility of Lysozyme in Polyethylene Glycol-Electrolyte Mixtures: the depletion interaction and ion-specific effects., *Biophys. J.*, 95, 1285-1294, 2008

FREDERICQ, E.; NEURATH, H., The Interaction of Insulin with Thiocyanate and other Anions. The Minimum Molecular Weight of Insulin., *J. Am. Chem. Soc.*, 72, 2684 - 2691, 1950

GREEN, A. A. Studies in the physical chemistry of the proteins (IX). *J. Biol. Chem.*, v. 93, p. 517-542, 1931

GRÖNWALL, A. Studies on the solubility of lactoglobulin. *Compt. rend. trav. Lab. Carlsberg*, v. 24, p. 185-200, 1941

KAARSHOLM, N. C.; HAVELUNG, S.; HOUGAARD, 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. 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, 1996

## 9. Referências Bibliográficas

---

KURAMITSU, S.; HAMAGUCHI, K. Analysis of the acid-base titration curve of hen lysozyme. *J. Biochem.*, v. 87, p. 1215-1219, 1980

LADISCH, M. R. **Bioseparations Engineering: principles, practice, and economics.** John Wiley & Sons, 2001. pp.1, 116 – 117, 133

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

PESSÔA FILHO, P. A., **Contribuições à modelagem do equilíbrio de fases e da partição de moléculas biológicas em sistemas aquosos bifásicos.**, Campinas, 2002, Tese de Doutorado, Faculdade de Engenharia Química, UNICAMP.

PESSÔA FILHO, P. A.; MAURER, G. An extension of the Pitzer equation for the excess Gibbs energy of aqueous electrolyte systems to aqueous polyelectrolyte solutions. *Fluid Phase Equilib.*, 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

PRADHAN, A. A.; VERA, J. H. Effects of acids and bases on the solubility of amino acids. *Fluid Phase Equilib.*, v. 152, p. 121-132, 1998

## 9. Referências Bibliográficas

---

PRAUSNITZ, J. M.; LICHTENTHALER, R. N.; AZEVEDO, E. G. De. **Molecular Thermodynamics of Fluid-Phase Equilibria**. 3 ed. Prentice Hall PTR, 1999.

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 Equilib.*, v. 287, p. 15-22, 2009

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

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

## Apêndice A – Equações de Henderson - Hasselbach

Considere-se que haja grupos ionizáveis ácidos e grupos ionizáveis básicos na cadeia protéica.

Então, pode-se descrever a dissociação de cada grupo ácido pela seguinte reação de ionização:



Assim, considerando que a constante de equilíbrio:

$$K = \prod_{i=1}^n a_i^{v_i} \tag{A - 2}$$

Para a reação (A - 1), considerando soluções ideais, tem-se que:

$$K = \frac{[A^-][H^+]}{[HA]} \tag{A - 3}$$

Aplicando o logaritmo decimal a equação (A - 3), pode-se escrever que:

$$\log K = -\log \frac{[HA]}{[A^-]} + \log[H^+] \tag{A - 4}$$

Considerando uma grandeza  $\varphi$  tem-se que o seu co-logaritmo pode ser denotado por:

$$-\log[\varphi] = p\varphi \tag{A - 5}$$

Assim, substituindo esta notação em (A - 4), tem-se que:

$$pK = \log \frac{[HA]}{[A^-]} + pH \quad (A - 6)$$

Contudo, a fração de resíduos carregados  $\alpha$  é dada por:

$$\alpha = \frac{[A^-]}{[HA] + [A^-]} \quad (A - 7)$$

Reescrevendo (A – 7), tem-se que:

$$\alpha = \frac{[A^-]/[HA]}{1 + [A^-]/[HA]} \Rightarrow \frac{[A^-]}{[HA]} = \frac{\alpha}{1 - \alpha} \quad (A - 8)$$

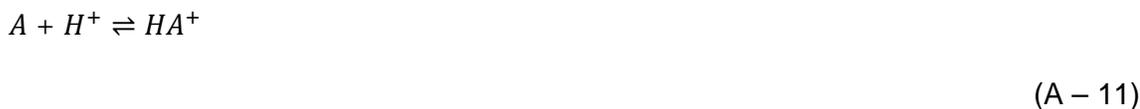
Substituindo (A – 8) em (A – 6), tem-se que:

$$pK = \log \left( \frac{\alpha}{1 - \alpha} \right) + pH \Rightarrow \frac{1 - \alpha}{\alpha} = 10^{pH - pK} \quad (A - 9)$$

Rearranjando (A – 9), tem-se para resíduos ácidos:

$$\alpha = \frac{1}{1 + 10^{pH - pK}} \quad (A - 10)$$

Analogamente, para resíduos básicos pode se escrever uma reação de equilíbrio:



Pode se escrever a constante de equilíbrio como:

$$K = \frac{[HA^+]}{[A][H^+]} \quad (A - 12)$$

Aplicando o logaritmo decimal a equação (A – 12), tem-se que:

$$\log K = -\log \frac{[A]}{[HA^+]} - \log[H^+]$$

(A -13)

Por analogia a fração de resíduos carregados é dada por:

$$\alpha = \frac{[HA^+]}{[A] + [HA^+]}$$

(A – 14)

Assim, a equação (A – 13) fica:

$$\frac{\alpha}{1 - \alpha} = 10^{pH - pK}$$

(A – 15)

Por fim, pode se escrever para resíduos básicos a seguinte equação:

$$\alpha = \frac{10^{pH - pK}}{1 + 10^{pH - pK}}$$

(A – 16)

## Apêndice B – Código – fonte do programa de cálculo dos modelos expostos

```
/* **** */
/* Polytechnic School of the University of São Paulo */
/* Biochemical Engineering Laboratory - Department of Chemical Engineering */
/* "Thermodynamic modeling of solid-liquid equilibrium in protein */
/* aqueous systems: effect of pH" */
/* Calculus of the protein solubility as a function of pH */
/* Developed by Luís Fernando Mercier Franco */
/* Advisor: Professor Pedro de Alcântara Pessoa Filho */
/* **** */

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

#define MAX 10000
#define MIN 150
#define NITER 1001

/* Prototypes of the auxiliaries functions */

/* Function that minimizes the objective function by Golden Section Method*/
double FMIN(double *XMIN, double pexp[MAX], double sexp[MAX], double solpH0,
            double pH0, double pI, int n);

/* Function that calculates the objective function */
double of(double A, double spH0, double pH0, double pI, double pH[MAX],
          double sexp[MAX], int n);

int main(){

    /* Variables' description */
    int i; /* counter of pH */
    int j; /* counter of g.i. */
    int k; /* counter of configurations */
    int n; /* number of g.i. */
    int npexp; /* number of experimental points */
    int nexpt;
    int gi[MAX]; /* vector of g.i. */
    int gil[MAX]; /* vector of g.i. that have variable alphas in the interval */
    int soma; /* sum of the charges of g.i. that have imutable alphas */
    int sum;
    int min; /* minimum */
    int med; /* medium */
    int iter; /* number of iterations */
    int npos;
    double pH[MAX]; /* vector of pH */
    double pKA[MIN]; /* vector of pKA */
    double ex;
    double toler;
    double termol;
    double pI; /* pI: isoelectric point */
    double mod; /* module of a number */
    double epsilon;
    double solexp[MAX]; /* experimental solubility */
    double pHexp[MAX];
    double solexp1[MAX];
```

## Apêndice B – Código – fonte do programa de cálculo dos modelos expostos

---

```

double pHexpl[MAX];
double solcalc[MAX];
double solpI; /* pI solubility */
double pitch; /* pitch between two pH */
double soli;
double solant;
double tol;
double aux;
double ionforce; /* ionic force */
double PI; /* the number pi = 3.1415... */
double KB;
double NAV; /* Avogadro's constant */
double PER0;
double EE; /* eletric charge of an eletron */
double TC;
double DC;
double TDL;
double T; /* temperature */
double DDL;
double DM;
double CPER;
double BPER;
double PER;
double D;
double BB;
double Z2; /* the square of the average charge */
double Z[MAX]; /* the average charge */
double soldh[MAX]; /* solubility calculated by the classical approach */
double e; /* the number e = 2.71... */
double CC;
double ifr;
double Z4;
double CC2;
double ifol;
double solcalcdh[MAX];
double soldhit[NITER];
double ionforce0;
double fun;
double fun1[MAX];
double derivfun;
double lnCC;
double lns;
double del;
double range;
double OF;
double AY;
double BY;
double CY;
double ymin;
double OFDH;
double tacid;
double tbasic;
double tlog;
double termof;
double OFML;
double somatorio;
double prod;
double s[MAX];
double s0;
double pH0;
double intcarga[MAX];
double termosum;
double ex0;
double scalc[MAX];
double OFLLG;
double ZLLG[MAX];
double OFMDH;
double slin[MAX];

```

## Apêndice B – Código – fonte do programa de cálculo dos modelos expostos

```

double A;
double XMIN;
double OFLIN;
double termo2;
double termo22;
double termo23;
double termo3;
double termo12;
double AF;
double OFMLIN;
double solcalclin[MAX];
char name[40];
char aminoacid[40];
char nfile[40];
char get[40];
time_t t1, t2;

FILE *ARQUI, *RESULTS, *RES, *HELPPFILE;

(void) time(&t1);

printf("-----");
printf("\n");
printf("\n\t POLYTECHNIC SCHOOL OF THE UNIVERSITY OF SAO PAULO \n");
printf("\n\t DEPARTMENT OF CHEMICAL ENGINEERING \n");
printf("\n\t BIOCHEMICAL ENGINEERING LABORATORY \n");
printf("\n\t .....");
printf("\n\t : PROTEIN SOLID-LIQUID EQUILIBRIUM : ");
printf("\n\t : .....: \n");
printf(" \n\t Developed by Luis Fernando Mercier Franco");
printf(" \n\t Advisor: Professor Pedro de Alcantara Pessoa Filho");
printf(" \n\t FAPESP (IC 2009/06575-0)\n");
printf(" \n\t Calculus of the protein solubility curve as a function");
printf(" \n\t of pH.\n\n");

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

med = 10000;
min = 150;

/* Reading the file (.txt) that contains the input values */
ARQUI = fopen(nfile, "r");
fscanf(ARQUI, "%S", get);
fscanf(ARQUI, "%s", &name);
fscanf(ARQUI, "%S", get);
fscanf(ARQUI, "%lf", &T);
fscanf(ARQUI, "%S", get);
fscanf(ARQUI, "%lf", &ionforce0);
fscanf(ARQUI, "%S", get);
fscanf(ARQUI, "%d", &n);
fscanf(ARQUI, "%S", get);
for (j = 0; j < n; j++){
    fscanf(ARQUI, "%s", &aminoacid[j]);
    fscanf(ARQUI, "%d", &gi[j]);
    fscanf(ARQUI, "%lf", &pKA[j]);
}
fscanf(ARQUI, "%S", get);
fscanf(ARQUI, "%lf", &pH[0]);
fscanf(ARQUI, "%lf", &pH[MAX-1]);
fscanf(ARQUI, "%S", get);
fscanf(ARQUI, "%lf", &pI);
fscanf(ARQUI, "%lf", &solpI);
fscanf(ARQUI, "%S", get);
fscanf(ARQUI, "%lf", &pH0);
fscanf(ARQUI, "%lf", &s0);
fscanf(ARQUI, "%S", get);

```

## Apêndice B – Código – fonte do programa de cálculo dos modelos expostos

```

fscanf(ARQUI, "%d", &npexp);
fscanf(ARQUI, "%S", get);
for (k = 0; k < npexp; k++){
    fscanf(ARQUI, "%lf", &pHexpl[k]);
    fscanf(ARQUI, "%lf", &solexpl[k]);
}
fclose(ARQUI);

/* Vector pH (1 - 14) */
pitch = (pH[med-1] - pH[0]) / (med - 1);
epsilon = 0.5 * pitch;
for (i = 0; i < med - 1; i++){
    pH[i+1] = pH[i] + pitch;
}

/* Calculating the net charge profile */
for (i = 0; i < med; i++){
    somatorio = 0;
    for (j = 0; j < n; j++){
        ex = pow(10, pH[i] - pKA[j]);
        tacid = (1.0 * gi[j]) * (1 - (1.0 * gi[j])) * ex;
        tbasic = (1 + (1.0 * gi[j]));
        termosum = (1 / (1 + ex)) * (tbasic + tacid);
        somatorio = somatorio + termosum;
    }
    Z[i] = 0.5 * somatorio;
}

/* Calculating solubility by LLG theory */
for (i = 0; i < med; i++){
    prod = 1;
    npos = 0;
    for (j = 0; j < n; j++){
        ex = 1 + pow(10, pH[i] - pKA[j]);
        ex0 = 1 + pow(10, pH0 - pKA[j]);
        prod = prod * (ex / ex0);
        npos = npos + (1 + gi[j]) / 2;
    }
    s[i] = s0 * pow(10, 1.0 * npos * (pH0 - pH[i])) * prod;
}
termof = 0;
for (k = 0; k < npexp; k++){
    for (i = 0; i < med; i++){
        funl[i] = pHexpl[k] - pH[i];
        if (i > 0 && funl[i] * funl[i-1] < 0){
            del = funl[i-1] / (funl[i-1] - funl[i]);
            solant = s[i-1];
            soli = s[i];
            scalc[k] = solant + del * (soli - solant);
            i = med;
        }
    }
    termof = termof + pow((scalc[k] - solexpl[k])/solexpl[k], 2);
}
OFML = pow(termof / npexp, 0.5);

/* Quadratic solution of Linderstrom-Lang and Grönwall equation */
OFLIN = FMIN(&XMIN, pHexpl, solexpl, s0, pH0, pI, npexp);
A = XMIN;
for (i = 0; i < med; i++){
    slin[i] = s0*pow(10, A/2*(pH[i]*pH[i]-2*pI*(pH[i]-pH0)-pH0*pH0));
}
termof = 0;
for (k = 0; k < npexp; k++){
    for (i = 0; i < med; i++){
        funl[i] = pHexpl[k] - pH[i];
        if (i > 0 && funl[i] * funl[i-1] < 0){
            del = funl[i-1] / (funl[i-1] - funl[i]);

```

## Apêndice B – Código – fonte do programa de cálculo dos modelos expostos

```

        solant = slin[i-1];
        soli = slin[i];
        solcalclin[k] = solant + del * (soli - solant);
        i = med;
    }
}
termof = termof + pow((solcalclin[k] - solexpl[k])/solexpl[k], 2);
}
OFMLIN = pow(termof / npexp, 0.5);

/* Classical approach with Debye-Hückel theory */
PI = 3.14159265;
KB = 1.38066 * pow(10, -23);
NAV = 6.02205 * pow(10, 23);
PER0 = 8.85419 * pow(10, -12);
EE = 1.60218 * pow(10, -19);
TC = 647.14;
DC = 322.0;
e = exp(1);
toler = solpI * pow(10, -3);
TDL = 1 - T / TC;
DDL = 1+1.99206*pow(TDL,1/3)+1.10123*pow(TDL,2/3)-0.512506*pow(TDL,5/3);
DDL = DDL-1.75263*pow(TDL,16/3)-45.4485*pow(TDL,43/3);
DDL = DDL-6.75615*pow(10,5)*pow(TDL,110/3);
DM = DDL * DC;
CPER = -2.0525+3115.9/(-182.89+T);
BPER = -8032.5+4.2142*pow(10,6)/T+2.1417*T;
PER = 342.79*exp(-5.0866*pow(10,-3)*T+9.469*pow(10,-7)*pow(T,2));
PER = PER + CPER * log((BPER + 1) / (BPER + 1000));
D = 4 * PI * PER * PER0;
termol = 1.0 / 3.0;
termol2 = 2 * PI * NAV * DM;
termo2 = pow(termol2, 0.5);
termo22 = pow(EE, 2)/D/KB/T;
termo23 = 3.0 / 2.0;
termo3 = pow(termo22, termo23);
AF = termol * termo2 * termo3;
BB = 1.2;
/* Newton-Raphson Method */
for (i = 0; i < med; i++){
    soldhit[0] = solpI;
    Z2 = pow(Z[i], 2);
    Z4 = pow(Z[i], 4);
    for (iter = 0; iter < NITER - 1; iter++){
        ifol = 0.5 * soldhit[iter] * Z2;
        ionforce = ionforce0 + ifol;
        ifr = pow(ionforce, 1/2);
        CC = 1 + BB * ifr;
        lnCC = log(CC)/log(e);
        CC2 = pow(CC, 2);
        lns = log(solpI / soldhit[iter]) / log(e);
        fun = lns + AF * Z2 * ((ifr / CC) + (2 / BB * lnCC));
        derivfun = -1/soldhit[iter]+AF*Z4*(1.5+BB*ifr)/(2*ifr*CC2);
        soldhit[iter+1] = soldhit[iter] - fun / derivfun;
        mod = fabs(soldhit[iter+1] - soldhit[iter]);
        if (mod <= toler){
            soldh[i] = soldhit[iter+1];
            iter = NITER;
        }
    }
    if (iter == NITER - 1){
        soldh[i] = soldhit[iter+1];
    }
}
}
termof = 0;
for (k = 0; k < npexp; k++){
    for (i = 0; i < med; i++){
        funl[i] = pHexpl[k] - pH[i];
    }
}

```

## Apêndice B – Código – fonte do programa de cálculo dos modelos expostos

```

        if (i > 0 && funl[i] * funl[i-1] < 0){
            del = funl[i-1] / (funl[i-1] - funl[i]);
            solant = soldh[i-1];
            soli = soldh[i];
            solcalcdh[k] = solant + del * (soli - solant);
            i = med;
        }
    }
    if (solcalcdh[k] > 0.5 * solpI){
        termof = termof + pow((solcalcdh[k] - solexpl[k])/solexpl[k], 2);
    }
}
OFMDH = pow(termof / npexp, 0.5);

/* Printing the results in a file which extension is .csv */
RESULTS = fopen("RES.csv", "w");
fprintf(RESULTS, "pH;Z;sol(DH);sol(LLG);sollin;;pHexp;solexp\n");
for (i = 0; i < med; i++){
    if (i < npexp){
        fprintf(RESULTS, "%e;%e;%e;%e;%e;%e\n",
            pH[i] * pow(10, -6), Z[i] * pow(10, -6), soldh[i] * pow(10, -6),
            s[i] * pow(10, -6), slin[i] * pow(10, -6), pHexpl[i] * pow(10, -6),
            solexpl[i] * pow(10, -6));
    }
    else{
        fprintf(RESULTS, "%e;%e;%e;%e;%e\n", pH[i] * pow(10, -6),
            Z[i] * pow(10, -6), soldh[i] * pow(10, -6),
            s[i] * pow(10, -6), slin[i] * pow(10, -6));
    }
}
fclose(RESULTS);

/* Printing the results in a file which extension is .txt */
RES = fopen("RES.txt", "w");
fprintf(RES, "\n Protein: %s\n\n", name);
fprintf(RES, "\n Temperature: %.2lfK\n", T);
fprintf(RES, "\n I: %.1lf mM\n", ionforce0 * 1000);
fprintf(RES, "\n Number of I. G.: %d\n", n);
fprintf(RES, "\n aminoacids | I. G. | pKA\n");
fprintf(RES, " -----\n");
for (j = 0; j < n; j++){
    fprintf(RES, " %c |", aminoacid[j]);
    if (gi[j] == 1){
        fprintf(RES, " %d |", gi[j]);
    }
    if (gi[j] == -1){
        fprintf(RES, " %d |", gi[j]);
    }
    if (pKA[j] / 10.0 < 1){
        fprintf(RES, " %.2lf\n", pKA[j]);
    }
    if (pKA[j] / 10.0 >= 1){
        fprintf(RES, " %.2lf\n", pKA[j]);
    }
}
fprintf(RES, "\n\n Values of solubility:\n\n");
fprintf(RES, " pH | sol exp | sol calc | sol (D-H)\n");
fprintf(RES, " -----\n");
for (k = 0; k < npexp; k++){
    if (pHexpl[k] / 10.0 < 1){
        fprintf(RES, " %.2lf | ", pHexpl[k]);
    }
    if (pHexpl[k] / 10.0 >= 1){
        fprintf(RES, " %.2lf | ", pHexpl[k]);
    }
    if (solcalcdh[k] < solpI){
        fprintf(RES, "%.2e | %.2e | NC\n", solexpl[k], scalc[k]);
    }
}

```

## Apêndice B – Código – fonte do programa de cálculo dos modelos expostos

```

        if (solcalcdh[k] >= solpI){
            fprintf(RES, "%.2e | %.2e | %.2e\n", solexpl[k],
                scalc[k], solcalcdh[k]);
        }
    }
    fprintf(RES, "\n OF values:\n");
    fprintf(RES, "\n For the Debye-Hückel theory: %.3e\n", OFMDH);
    fprintf(RES, "\n For the Linderstrom-Lang theory: %.3e\n", OFML);
    (void) time(&t2);
    fprintf(RES, "\n\n Time to do this = %d seconds\n", (int) t2-t1);
    fclose(RES);

    /* Printing the results in the screen */
    printf("\n Result summary:\n\n");
    printf("\n Protein name: %s\n\n", name);
    printf("\n Temperature: %.2lfK\n", T);
    printf("\n I: %.1lf mM\n", ionforce0 * 1000);
    printf("\n Number of ionizable groups (G. I.): %d\n", n);
    printf("\n Aminoacids | G. I. | pKA\n");
    printf("-----\n");
    for (j = 0; j < n; j++){
        printf(" %c |", aminoacid[j]);
        if (gi[j] == 1){
            printf(" %d |", gi[j]);
        }
        if (gi[j] == -1){
            printf(" %d |", gi[j]);
        }
        if (pKA[j] / 10.0 < 1){
            printf(" %.2lf\n", pKA[j]);
        }
        if (pKA[j] / 10.0 >= 1){
            printf(" %.2lf\n", pKA[j]);
        }
    }
    printf("\n\n Solubility values:\n\n");
    printf(" pH | sol exp | sol calc | sol (D-H) | sol Q\n");
    printf("-----\n");
    for (k = 0; k < npexp; k++){
        if (pHexpl[k] / 10.0 < 1){
            printf(" %.2lf | ", pHexpl[k]);
        }
        if (pHexpl[k] / 10.0 >= 1){
            printf(" %.2lf | ", pHexpl[k]);
        }
        if (solcalcdh[k] < solpI){
            printf("%.2e | %.2e | NC | %.2e\n", solexpl[k],
                scalc[k], solcalclin[k]);
        }
        if (solcalcdh[k] >= solpI){
            printf("%.2e | %.2e | %.2e | %.2e\n", solexpl[k], scalc[k],
                solcalcdh[k], solcalclin[k]);
        }
    }
    printf("\n OF values:\n");
    printf("\n For Debye-Huckel theory: %.3e\n", OFMDH);
    printf("\n For Linderstrom-Lang and Gronwall Extended Equation: %.3e\n", OFML);
    printf("\n K = %lf\n", A / 2);
    (void) time(&t2);
    printf("\n\n Execution time = %d seconds\n", (int) t2-t1);
    printf("\n The results are stored in a file called RES.txt...\n\n");
    printf(" Press 'Enter' to close this window.\n\n");
    printf("-----");
    getch();
    return (0);
}

```

## Apêndice B – Código – fonte do programa de cálculo dos modelos expostos

```

double FMIN(double *XMIN, double pexp[MAX], double sexp[MAX], double spH0,
  double pH0, double pI, int n){
  int max;
  double AX;
  double BX;
  double CX;
  double tol;
  double R;
  double C;
  double X0;
  double X1;
  double X2;
  double X3;
  double F1;
  double F2;
  double FMIN;
  max = 280;
  AX = 0;
  CX = 100;
  BX = (CX - AX) / 3;
  tol = pow(10, -10);
  R = 0.61803399;
  C = 1 - R;
  X0 = AX;
  X3 = CX;
  if (fabs(CX - BX) < fabs(BX - AX)){
    X1 = BX;
    X2 = BX + C * (CX - BX);
  }
  else{
    X2 = BX;
    X1 = BX - C * (CX - BX);
  }
  F1 = of(X1, spH0, pH0, pI, pexp, sexp, n);
  F2 = of(X2, spH0, pH0, pI, pexp, sexp, n);
  while (fabs(X3 - X0) > tol * (X1 + X2)){
    if (F2 < F1){
      X0 = X1;
      X1 = X2;
      X2 = R * X1 + C * X3;
      F1 = F2;
      F2 = of(X2, spH0, pH0, pI, pexp, sexp, n);
    }
    else{
      X3 = X2;
      X2 = X1;
      X1 = R * X2 + C * X0;
      F2 = F1;
      F1 = of(X1, spH0, pH0, pI, pexp, sexp, n);
    }
  }
  if (F1 < F2){
    FMIN = F1;
    *XMIN = X1;
  }
  else{
    FMIN = F2;
    *XMIN = X2;
  }
  return FMIN;
}

double of(double A, double spH0, double pH0, double pI, double pH[MAX],
  double sexp[MAX], int n){
  int k;
  double FO;
  double sum;
  double pH2;

```

*Apêndice B – Código – fonte do programa de cálculo dos modelos expostos*

---

```
double m;  
double pH02;  
sum = 0;  
for (k = 0; k < n; k++){  
    pH2 = pH[k] * pH[k];  
    m = 2 * pI * (pH[k] - pH0);  
    pH02 = pH0 * pH0;  
    sum = sum + (sexp[k] - spH0*pow(10, A/2*(pH2 - m - pH02)))/sexp[k];  
}  
FO = pow(sum / (1.0 * n), 0.5);  
return FO;  
}
```