



**Universidade Federal de São Carlos**  
**Programa de Pós-Graduação em Engenharia Química**

**Modelagem e otimização da cromatografia de troca aniônica no processo de produção de  
proteína de pneumococo**

Aluno: Leandro Junqueira Benedini  
Orientadora: Profª. Dra. Teresa Cristina Zangirolami  
Coorientador: Prof. Dr. Felipe Fernando Furlan

**São Carlos**

**2019**



**Universidade Federal de São Carlos**  
**Programa de Pós-Graduação em Engenharia Química**

**Modelagem e otimização da cromatografia de troca aniônica no processo de produção de proteína de pneumococo**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Doutor em Engenharia Química.

Aluno: Leandro Junqueira Benedini  
Orientadora: Profa. Dra. Teresa Cristina Zangirolami  
Coorientador: Prof. Dr. Felipe Fernando Furlan

**São Carlos**

**2019**

Junqueira Benedini, Leandro

Modelagem e otimização da cromatografia de troca aniônica no processo de produção de proteína de pneumococo / Leandro Junqueira Benedini. -- 2019.

193 f. : 30 cm.

Tese (doutorado)-Universidade Federal de São Carlos, campus São Carlos, São Carlos

Orientador: Profa. Dra. Teresa Cristina Zangirolami

Banca examinadora: Profa. Dra. Teresa Cristina Zangirolami, Dra. Viviane Maimoni Gonçalves, Prof. Dr. Adriano Rodrigues Azzoni, Prof. Dr. Felipe Fernando Furlan, Prof. Dr. Marcelo Perencin de Arruda Ribeiro

Bibliografia

1. Proteínas recombinantes. 2. Cromatografia de troca aniônica. 3. Modelagem matemática. I. Orientador. II. Universidade Federal de São Carlos. III. Título.

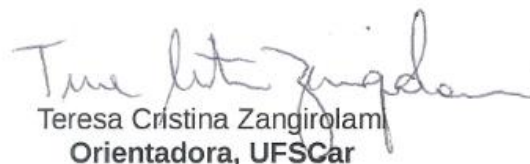
Ficha catalográfica elaborada pelo Programa de Geração Automática da Secretaria Geral de Informática (SIn).

DADOS FORNECIDOS PELO(A) AUTOR(A)

Bibliotecário(a) Responsável: Ronildo Santos Prado – CRB/8 7325

MEMBROS DA BANCA EXAMINADORA DA DEFESA DE TESE DE LEANDRO JUNQUEIRA BENEDINI APRESENTADA AO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA DA UNIVERSIDADE FEDERAL DE SÃO CARLOS, EM 12 DE DEZEMBRO DE 2019.

BANCA EXAMINADORA:

  
Teresa Cristina Zangirolami  
Orientadora, UFSCar

  
Viviane Maimoni Gonçalves  
Instituto Butantan

  
Adriano Rodrigues Azzoni  
USP

  
Felipe Fernando Furlan  
UFSCar

  
Marcelo Perencin de Arruda Ribeiro  
UFSCar

## AGRADECIMENTOS

Agradeço primeiramente à minha esposa Larissa, que sempre me apoiou em todas as minhas decisões e esteve ao meu lado em todos momentos. Não posso deixar de agradecer também aos meus pais que sempre me incentivaram a estudar, buscar conhecimento e melhorar cada dia mais.

Agradeço aos professores Drs. Teresa Zangirolami e Felipe Furlan pelo acolhimento, apoio e todos os conhecimentos compartilhados e construídos ao longo desses anos, as conversas e conselhos. A orientação de vocês foi indispensável para o desenvolvimento desse trabalho.

Agradeço especialmente ao Instituto Federal de Educação Ciência e Tecnologia de São Paulo – IFSP pelo incentivo à pesquisa através do afastamento para qualificação docente, o que me permitiu dedicar-me integralmente aos estudos.

Agradeço também aos professores que muito gentilmente contribuíram com a minha pesquisa através de suas sugestões na banca da qualificação: Drs. Marcelo Ribeiro e Jorge Pereira. Gostaria de destacar que a arguição de vocês foi fundamental para o desenvolvimento da pesquisa.

Agradeço a todos os colegas do Instituto Butantan, em especial à Dra. Viviane Gonçalves pelo comprometimento e disponibilidade, pois sem você a execução desse trabalho não teria sido possível.

Finalmente, não posso deixar de agradecer a todos os professores e colegas de profissão que atualmente lutam pelo desenvolvimento de pesquisa séria e de qualidade em nosso país, acreditando no poder da educação como ferramenta para a construção de uma sociedade melhor, com menos ódio, intolerância, preconceito e discriminação.

## RESUMO

A proteína A de superfície de pneumococo (PspA) é uma proteína ligante de lactoferrina que pode ser encontrada na superfície de *Streptococcus pneumoniae*. O seu uso como agente carreador na formulação de vacinas conjugadas contribuiu para a proteção contra infecções como pneumonia, sinusite e meningite. Recentemente foi desenvolvido pelo Instituto Butantan um processo de purificação de PspA4Pro produzida em cultivos de *Escherichia coli* BL21(DE3) pET37b+PspA4Pro. Apesar dos resultados alcançados em termos de pureza serem satisfatórios verificou-se perda significativa do produto de interesse ao longo da purificação, principalmente na cromatografia de troca aniônica, a etapa mais cara do processo *downstream*. Neste contexto, o presente trabalho tem como objetivo modelar e simular esta operação, desenvolvendo uma metodologia aplicável a misturas complexas. Na primeira etapa deste trabalho extratos celulares obtidos de biomassas de *E. coli* produzidas em diferentes condições de cultivo foram purificados. A cromatografia de troca aniônica foi executada no sistema Akta Avant 150 e as amostras foram analisadas por metodologia Bradford e densitometria de bandas. O processo foi modelado no software EMSO (*Environment for modeling, simulation and optimization*) e as isotermas de ação em massa estérica e de Langmuir modificada foram testadas com a equação do equilíbrio dispersivo para estimar parâmetros e validar modelos, mostrando-se adequadas para simular o perfil de eluições. A isoterma de ação em massa estérica foi usada para simular o processamento de uma amostra virtual, com parâmetros de PspA4Pro e impurezas proteicas ajustados separadamente. Os resultados da simulação foram comparados com os resultados de uma eluição experimental. Em ambos os casos foi observado que PspA4Pro eluiu no início do pico, mostrando semelhança entre os perfis da simulação e do experimento e permitindo estabelecer um esquema de purificação alternativo que forneceu PspA4Pro a 82,8% de pureza, representando um aumento de 34% em relação ao processo de purificação original. Em uma segunda etapa, diferentes experimentos envolvendo modificações nas forças iônicas de eluição e aumento de massa proteica processada foram testados na coluna e uma metodologia validações cruzadas foi aplicada. Simulações demonstraram que a coluna cromatográfica poderia processar o dobro de proteína sem perda de qualidade da PspA4Pro obtida. Os resultados foram confirmados experimentalmente, obtendo-se um rendimento específico de  $22 \times 10^{-3} \text{ g PspA4Pro.mL}_{\text{resina}}^{-1}$  com 81,2% de pureza, superior à rendimento específico de  $10 \times 10^{-3} \text{ g PspA4Pro.mL}_{\text{resina}}^{-1}$  com 73,9% de pureza do processo original. Em uma última etapa, a metodologia de validação cruzada foi utilizada para estimar separadamente os parâmetros de PspA4Pro e impurezas proteicas. Os resultados das simulações, mais uma vez mostraram a tendência de PspA4Pro eluir no início do pico, mas em perfis variáveis de acordo com a força iônica aplicada, possibilitando a identificação de estratégias alternativas de eluição. Os melhores resultados foram obtidos aplicando menores forças iônicas, levando a um expressivo aumento de pureza (91,0%) sem redução de rendimento específico ( $9,7 \times 10^{-3} \text{ g}_{\text{PspA4Pro.mL}_{\text{resina}}^{-1}}$ ). As conclusões deste trabalho mostram que é possível utilizar modelos matemáticos para descrever cromatografias e aumentar eficiência, mesmo para misturas complexas, como é o caso dos extratos celulares com proteínas recombinantes. Esta abordagem facilita a análise do processo de forma sistemática e permite a melhor compreensão de operações reais de purificação.

Palavras-chave: *Escherichia coli*. Proteínas recombinantes. Cromatografia de troca aniônica. Modelagem matemática. Modelo do Equilíbrio Dispersivo.

## ABSTRACT

Pneumococcus surface protein A (PspA) is a lactoferrin-binding protein that can be found on the surface of *Streptococcus pneumoniae*. The use of this protein as a carrier in the formulation of conjugate vaccines contributes to the protection against infections such as pneumonia, sinusitis and meningitis. Recently, the Butantan Institute developed a purification process for PspA4Pro produced by cultures of *Escherichia coli* BL21 (DE3) pET37b + PspA4Pro. Although the results achieved in terms of purity were satisfactory, it was observed a significant loss of the product during the purification, especially in anion exchange chromatography, the most expensive step in the downstream process. In this context, the present work aims to model and simulate this operation, developing an applicable methodology to model complex mixtures. In the first stage of this work, clarified cell extracts obtained from *E. coli* biomass produced under different cultivation conditions were purified. Anion exchange chromatography was performed on the Akta Avant 150 system and the samples were analyzed by Bradford methodology and band densitometry. The process was modeled by the EMSO software (Environment for modeling, simulation and optimization) and the modified Langmuir and the steric mass action isotherms were separately tested along with the equilibrium dispersive model to estimate the parameters and validate the models, proving to be adequate to simulate the process. The steric mass action isotherm was also used to simulate the chromatography of a virtual sample, with the parameters of PspA4Pro and the protein impurities separately adjusted. The simulation results were compared with the results of an experimental elution. In both cases it was observed that PspA4Pro eluted at the beginning of the peak, showing similarity between the simulation and experiment profiles and allowing to establish an alternative purification scheme that provided PspA4Pro at a purity of 82.8%, representing an increase of 34% when compared to the original purification process. In the second stage of this work, different experiments involving modifications in ionic elution forces and in the processed protein mass were tested on the column and a cross-validation methodology was applied. Simulations demonstrated that the chromatographic column could process twice the protein without loss of quality of the obtained PspA4Pro. These results were experimentally confirmed, providing a specific yield of  $22 \times 10^{-3} \text{ g}_{\text{PspA4Pro}} \cdot \text{mL}_{\text{resin}}^{-1}$  with purity of 81.2%, which was higher than the specific yield of  $10 \times 10^{-3} \text{ g}_{\text{PspA4Pro}} \cdot \text{mL}_{\text{resin}}^{-1}$  with purity of 73.9% obtained at the original process. In the last stage of this work, the cross validation methodology was once again used to separately estimate the PspA4Pro and the protein impurities parameters. The results of the simulations once again showed the tendency of PspA4Pro to elute at the beginning of the peak, but in variable profiles according to the applied ionic force, enabling the identification of alternative elution strategies. The best results were obtained by applying lower ionic forces, leading to a significant increase of purity (91.0%) without reduction of specific yield ( $9.7 \times 10^{-3} \text{ g}_{\text{PspA4Pro}} \cdot \text{mL}_{\text{resin}}^{-1}$ ). The conclusions of this work show that it is possible to use mathematical models to describe chromatographic processes and to increase efficiency, even when complex mixtures are being processed, such as cell extracts with recombinant proteins. This approach facilitates systematic process analysis and allows a better understanding of actual purification operations.

Keywords: *Escherichia coli*. Recombinant proteins. Anion exchange chromatography. Mathematical modeling. Equilibrium Dispersive Model.

## LISTA DE FIGURAS

- Figura 2.1. Mortalidade estimada em crianças com menos de cinco anos de idade expostas a doenças causadas por pneumococos. 31
- Figura 2.2. Etapas da transferência de oxigênio desde a fase gasosa até o interior da célula (Fonte: CAMPANI, 2014). 34
- Figura 2.3. Sequência de operações unitárias propostas para purificação de PspA4Pro em células de *Escherichia coli* recombinante. 37
- Figura 2.4. Representação de um homogeneizador de alta pressão (Adaptado de SUBSTECH, 2013). 39
- Figura 2.5. Representação da curva de ruptura em sistema de adsorção com leito fixo. Adaptado de TREYBAL (1981). ZTM: Zona de transferência de massa. 42
- Figura 2.6. Esquema do mecanismo de troca iônica (Fonte: COLLINS, 2006). 44
- Figura 2.7. Perfil de eluição hipotético gerado pelo processo cromatográfico descrito na figura anterior (Fonte: COLLINS, 2006). 44
- Figura 2.8. Representação da transferência de massa de uma proteína em uma partícula esférica de resina. Fonte: Autoria própria. 46
- Figura 2.9. Representação dos aspectos de transferência de massa observados no Modelo do equilíbrio dispersivo. Fonte: Autoria própria. 48
- Figura 2.10. Representação dos aspectos de transferência de massa observados no modelo do transporte dispersivo. Fonte: Autoria própria. 49
- Figura 2.11. Representação dos aspectos de transferência de massa observados no modelo geral. Fonte: Autoria própria. 52
- Figura 2.12. Representação gráfica da isoterma de adsorção linear. Fonte: Autoria própria. 53
- Figura 2.13. Representação gráfica da isoterma de adsorção de Langmuir. Fonte: Autoria própria. 55
- Figura 2.14. Representação gráfica da isoterma de adsorção de Langmuir com competição entre as espécies químicas A e B. Fonte: Autoria própria. 58
- Figura 2.15. Representação tridimensional da isoterma de adsorção de Langmuir com competição entre as espécies químicas A e B. Adaptado de GUIOCHON, 2006. 58
- Figura 2.16. Representação gráfica da isoterma de adsorção de Langmuir modificada a partir de diferentes concentrações de sal. Fonte: Autoria própria. 60
- Figura 2.17. Esquema representativo da ligação de proteínas em uma superfície de troca catiônica. Adaptado de: GALLANT, 1995. 63

Figura 3.1. Descrição da estratégia de modelagem e simulação do processo utilizando o modelo dispersivo de equilíbrio (MED) em conjunto com as isotermas de ação em efeito estérico (SMA) ou Langmuir modificada (LM). Os parâmetros foram estimados a partir dos dados da CTA-I e a validação do modelo foi realizada com os dados da CTA-II. 77

Figura 3.2. Descrição da estratégia de simulação do extrato virtual de células clarificadas (ECC) utilizando o modelo do equilíbrio dispersivo (MED) em conjunto com a isoterma de ação em efeito estérico (SMA). Os parâmetros do pool de proteínas e PspA4Pro foram estimados a partir dos dados CTA-III e CTA-IV, respectivamente. 80

Figura 3.3. SDS-PAGE dos extratos celulares clarificados (ECC) com PspA4Pro recombinante e outras proteínas. 1) Marcador molecular contendo fosforilase (97 kDa), albumina (66 kDa), ovalbumina (45 kDa), anidro (30 kDa), tripsina (20 kDa) e lactalbumina (14,4 kDa). 2) ECC<sub>A</sub> produzido por células cultivadas com indução de lactose a 27 °C. 3) ECC<sub>B</sub> produzido por células cultivadas com indução de IPTG a 37 °C 4) ECC<sub>C</sub> produzido por células portadoras do vetor pET37b sem o inserto de PspA4Pro. 5) Amostra pura de PspA4Pro (SPP). 82

Figura 3.4. SDS-PAGE das frações de eluições Q300 produzidas pela cromatografia. 1) Marcador molecular contendo fosforilase (97 KDa), albumina (66 KDa), ovalbumina (45 KDa), anidro (30 KDa) e tripsina (20 KDa). 2) Fração Q300 produzida por células cultivadas com indução de lactose a 27 °C 3) Fração Q300 produzida por células cultivadas com indução de IPTG a 37 °C 4) Fração Q300 produzida por células portadoras do vetor pET37b sem o inserto de PspA4Pro. 5) Fração Q300 produzida por amostra pura de PspA4Pro (SPP). 83

Figura 3.5. Comparação de cromatogramas experimentais: CTA-I realizada com extrato celular clarificado (ECC) do cultivo A (células induzidas por lactose a 27 °C), CTA-II realizada com ECC do cultivo B (células induzidas por IPTG a 37 °C), CTA-III com ECC do cultivo C (células portadoras do vetor pET37b sem inserto de PspA4Pro) e CTA-IV (PspA4Pro pura). 84

Figura 3.6. Perfis de eluição para as proteína totais dos extratos celulares ECC<sub>A</sub> na CTA-I (estimativa dos parâmetros do modelo): a) Dados experimentais (linha preta sólida), perfis simulados usando a isoterma de ação em efeito estérico (SMA) em conjunto com o modelo do equilíbrio dispersivo (MED) (linha preta tracejada) e perfis de eluição obtidos a partir de aplicação de NaCl a 150 mM, 300 mM e 1000 mM (linha cinza tracejada); b) Dados experimentais (linha preta sólida), perfis simulados usando a isoterma de Langmuir modificada (LM) em conjunto com o modelo do equilíbrio dispersivo (MED), e perfis de eluição obtidos a partir de aplicação de NaCl a 150 mM, 300 mM e 1000 mM (linha cinza tracejada). 88

Figura 3.7. Perfis de eluição para as proteína totais dos extratos celulares ECC<sub>B</sub> na CTA-II (validação do modelo): a) Dados experimentais (linha preta sólida), perfis simulados usando a isoterma de ação em efeito estérico (SMA) em conjunto com o modelo do equilíbrio dispersivo (MED) (linha preta tracejada) e perfis de eluição obtidos a partir de aplicação de NaCl a 150 mM, 300 mM e 1000 mM (linha cinza tracejada); b) Dados experimentais (linha preta sólida), perfis simulados usando a isoterma de Langmuir modificada (LM) em conjunto com o modelo do equilíbrio dispersivo (MED), e perfis de eluição obtidos a partir de aplicação de NaCl a 150 mM, 300 mM e 1000 mM (linha cinza tracejada). 89

Figura 3.8. Modelo da isoterma de ação em efeito estérico (SMA) (linhas tracejadas) ajustado aos perfis de eluição experimental Q300 (linhas sólidas). a) cromatografia de troca aniônica (CTA-III) aplicada em extrato celular clarificado do cultivo C (células portadoras do vetor pET37b sem o gene de PspA4Pro); b) cromatografia de troca aniônica (CTA-IV) aplicada em solução de proteína pura (SPP). 91

Figura 3.9. Perfis de eluição simulados para PspA4Pro, para o conjunto de impurezas proteicas e para a mistura de proteínas totais da cromatografia de troca aniônica (perfil de eluição obtido com aplicação de 300 mM de NaCl). As simulações foram realizadas utilizando a isoterma de ação em efeito estérico (SMA) em conjunto com o modelo do equilíbrio dispersivo (MED). Os parâmetros estimados estão expostos na Tabela 3.3. 92

Figura 4.1. Descrição da cromatografia de troca aniônica. A amostra QNAds contém a fração não adsorvida. As amostras Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> contêm as eluições fraca, média e forte, respectivamente. 100

Figura 4.2. Procedimento numérico utilizado para estimar os parâmetros e simular o sistema descrito pela Figura 4.1. Foram utilizados dados dos experimentos A, B, C e D. 105

Figura 4.3. Curvas das eluições Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> simuladas (vermelhas) e ajustadas às curvas experimentais (azuis) dos experimentos B, C e D. A performance da simulação foi validada pelo experimento A. 111

Figura 4.4. Curvas das eluições Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> simuladas (vermelhas) e ajustadas às curvas experimentais (azuis) dos experimentos A, C e D. A performance da simulação foi validada pelo experimento B. 112

Figura 4.5. Curvas das eluições Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> simuladas (vermelhas) e ajustadas às curvas experimentais (azuis) dos experimentos A, B e D. A performance da simulação foi validada pelo experimento C. 112

Figura 4.6. Curvas das eluições Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> simuladas (vermelhas) e ajustadas às curvas experimentais (azuis) dos experimentos A, B e C. A performance da simulação foi validada pelo experimento D. 113

Figura 4.7. Perfis de eluição simulados ( $q_i$ ) dos pools de proteínas Q<sub>1</sub>-I (linhas azuis), Q<sub>1</sub>-II (linhas laranjas), Q<sub>2</sub> (linha cinza) e Q<sub>3</sub> (linha amarela) nos N pratos teóricos da coluna cromatográfica a partir do processamento da massa de proteínas adsorvidas no experimento A, chamada de M<sub>A</sub> (linha A), 2 x M<sub>A</sub> (linha B) e 3 x M<sub>A</sub> (linha C). As condições simuladas foram apresentadas imediatamente após aplicações do extrato celular clarificado (ECC) e eluições Q<sub>1</sub> e Q<sub>2</sub>. 115

Figura 5.1. Representação do perfil de eluição Q<sub>2</sub> para os experimentos realizados com ECCB. O volume total eluído foi dividido em dez subfrações, cada uma com 0,5 volume de coluna. 126

Figura 5.2. Estratégia experimental utilizada para identificar os parâmetros do modelo e simular a cromatografia de troca aniônica usando dados das corridas I a VII. 131

Figura 5.3. Estratégia experimental utilizada para simular o sistema a partir de variações na concentração de NaCl em Q<sub>2</sub> (250 mM a 350 mM). Os dados utilizados pertenceram às cromatografias (CTAs) VII a IX. 132

Figura 5.4. Comparações entre as cromatografias de troca aniônica CTA-I, CTA-II e CTA-III experimentais (linhas sólidas) e simuladas (linhas pontilhadas) representadas nos quadros “a”, “b” e “c”, respectivamente: Os experimentos foram realizados com extrato celular clarificado (ECC) do cultivo A (células não produtoras de PspA4Pro cultivadas a 37 °C, com indução por lactose). Os primeiros e últimos picos de todas as corridas foram obtidos pela aplicação de 5 VC de eluições com 150 e 1000 mM de NaCl, respectivamente. Os segundos picos foram obtidos pela aplicação de 5 VC de eluições com 300, 250 e 350 mM de NaCl, respectivamente. 138

Figura 5.5. Comparação entre os cromatogramas experimentais (linhas sólidas) e simulados (linhas pontilhadas) das cromatografias de troca aniônica CTA-IV, CTA-V e CTA-VI: Os experimentos foram realizados com solução de PspA4Pro pura (SPP). As primeiras e últimas eluições não foram detectadas. As segundas eluições foram obtidas pela aplicação de 5 VC de solução com 300, 250 e 350 mM de NaCl, respectivamente. 139

Figura 5.6. Comparação entre os cromatogramas experimentais e simulados de CTA-VII. Os dados experimentais foram obtidos com extrato celular clarificado (ECC) do cultivo B (células produtoras de PspA4Pro induzidas por lactose a 37 °C). A simulação foi realizada considerando as MTPs e purezas informadas na Tabela 5.2 (CTA-VII), as condições iniciais e parâmetros fixos da Tabela 5.3 e os parâmetros estimados da Tabela 5.4 para todos os pools de proteínas e PspA4Pro. 140

Figura 5.7: Géis SDS-PAGE do extrato celular clarificado (ECC<sub>B</sub>) que apresentava PspA4Pro recombinante junto com outras proteínas e dos subfragmentos 2 a 6 da eluição Q<sub>2</sub> de CTA-VII. MM) Marcador molecular contendo fosforilase (97 KDa), albumina (66 KDa), ovalbumina (45 KDa), anidro (30 KDa) e tripsina (20 KDa). 1) ECC<sub>B</sub> obtido por células produtoras de PspA4Pro cultivadas a 37 °C com indução por lactose. 2) subfragmento 2 de Q<sub>2</sub>. 3) subfragmento 3 de Q<sub>2</sub>. 4) subfragmento 4 de Q<sub>2</sub>. 5) subfragmento 5 de Q<sub>2</sub>. 6) subfragmento 6 de Q<sub>2</sub>. 143

Figura 5.8. Valores simulados das recuperações de PspA4Pro (linhas amarelas) e purezas (linhas azuis) para os perfis alternativos de eluições baseados em variações da força iônica da eluição central (250 mM a 350 mM) . 145

Figura 5.9. Valores simulados das recuperações de PspA4Pro REC(t) para os perfis alternativos de eluições baseados em variações da força iônica da eluição central (250 mM a 350 mM). 146

Figura 5.10. Valores simulados das purezas P(t) para os perfis alternativos de eluições baseados em variações da força iônica da eluição central (250 mM a 350 mM). 146

Figura 5.11. Comparação dos cromatogramas simulados de CTA-VII quando concentrações de 250 mM e 350 mM são aplicadas em Q<sub>2</sub>. Para cada caso, picos de PspA4Pro e do pool de proteínas eluídas no mesmo momento foram descritos separadamente . 147

Figura 5.12. Comparação entre cromatogramas experimentais: CTA-VIII e CTA-IX realizados com extrato celular clarificado (ECC) do cultivo B (células produtoras de PspA4Pro cultivadas a 27 °C com indução por lactose). As eluições foram realizadas com 150 mM, 250 mM e 1000 mM de NaCl e 150 mM, 350 mM e 1000 mM de NaCl, respectivamente. 148

Figura 5.13: Géis SDS-PAGE do extrato celular clarificado (ECC<sub>B</sub>) que apresentava PspA4Pro recombinante junto com outras proteínas e dos subfragmentos 2 a 6 da eluição Q<sub>2</sub> de CTA-VIII. MM) Marcador molecular contendo fosforilase (97 KDa), albumina (66 KDa), ovalbumina (45 KDa), anidro (30 KDa) e tripsina (20 KDa). 1) ECC<sub>B</sub> obtido por células produtoras de PspA4Pro cultivadas a 37 °C com indução por lactose. 2) subfragmento 2 de Q<sub>2</sub>. 3) subfragmento 3 de Q<sub>2</sub>. 4) subfragmento 4 de Q<sub>2</sub>. 5) subfragmento 5 de Q<sub>2</sub>. 6) subfragmento 6 de Q<sub>2</sub>. 149

Figura 5.14: Géis SDS-PAGE do extrato celular clarificado (ECC<sub>B</sub>) que apresentava PspA4Pro recombinante junto com outras proteínas e dos subfragmentos 2 a 6 da eluição Q<sub>2</sub> de CTA-IX. MM) Marcador molecular contendo fosforilase (97 KDa), albumina (66 KDa), ovalbumina (45 KDa), anidro (30 KDa) e tripsina (20 KDa). 1) ECC<sub>B</sub> obtido por células produtoras de PspA4Pro cultivadas a 37 °C com indução por lactose. 2) subfragmento 2 de Q<sub>2</sub>. 3) subfragmento 3 de Q<sub>2</sub>. 4) subfragmento 4 de Q<sub>2</sub>. 5) subfragmento 5 de Q<sub>2</sub>. 6) subfragmento 6 de Q<sub>2</sub>. 149

Figura A.1. Esquema de reatores de mistura perfeita para a conversão de A em B através de reação de primeira ordem . 164

Figura A.2. Definição dos modelos envolvidos em um processo que apresenta três reatores em série. 165

Figura A.3. Desenvolvimento do fluxograma que representa o processo de três reatores em série. 166

Figura A.4. Resultados do processo simulado composto por três reatores em série. 166

Figura B.1. Variação do perfil de absorvância (Abs) e porcentagem de condutividade (%C) durante o tempo no processo de cromatografia aniônica. Eluições: Q150 (0 a 20 minutos), Q300 (20 a 40 minutos) e Q1000 (40 a 60 minutos). 167

Figura B.2. Variação de concentração de proteínas (P) e NaCl (C<sub>NaCl</sub>) durante o tempo no processo de cromatografia aniônica. 169

Figura D.1. Perfis de eluição de proteínas da mistura CCE<sub>A</sub> na AEC-I (estimativa de parâmetros): Dados experimentais (linha sólida), perfis simulados usando a isoterma de Langmuir competitiva (linha pontilhada escura) e perfis de eluição usando 150 mM, 300 mM e 1000 mM de NaCl (linha pontilhada clara). 181

Figura D.2. Perfis de eluição de proteínas da mistura CCE<sub>B</sub> na AEC-II (validação do modelo): Dados experimentais (linha sólida), perfis simulados usando a isoterma de Langmuir competitiva (linha pontilhada escura), e perfis de eluição usando 150 mM, 300 mM e 1000 mM de NaCl (linha pontilhada clara). 181

## LISTA DE TABELAS

- Tabela 3.1. Características das condições de entrada, frações não adsorvidas (QNAds) e frações de eluição (Q150, Q300 e Q1000) para as cromatografias de troca aniônica (CTA) I a IV executadas em uma coluna XK 50/30. 81
- Tabela 3.2. Condições iniciais e parâmetros fixos usados para a simulação da cromatografia de troca aniônica. 85
- Tabela 3.3. Parâmetros estimados das isotermas de ação em efeito estérico (SMA) e Langmuir modificada (ML) durante as corridas cromatográficas CTA-I, CTA-III e CTA-IV. 86
- Tabela 3.4. Valores simulados e experimentais da massa total de proteínas (MTP) e pureza para o perfil de eluição alternativo obtido com base no subfracionamento da eluição Q300 (CTA-V, realizado usando a coluna XK 26/20). 93
- Tabela 4.1. Características do extrato celular clarificado e perfis de eluição dos experimentos realizados em uma coluna de 70 mL com resina Q-Sepharose. 101
- Tabela 4.2. Massa ( $M_i$ ), pureza ( $P_i$ ) e recuperações de PspA4Pro ( $R_i$ ) nos extratos celulares clarificados (ECC) e frações QNAds,  $Q_1$ ,  $Q_2$  e  $Q_3$ . As amostras foram obtidas a partir da realização dos experimentos A a E para purificar PspA4Pro com cromatografia de troca aniônica. 108
- Tabela 4.3. Condições iniciais e parâmetros fixos utilizados nas simulações da cromatografia de troca aniônica. 109
- Tabela 4.4. Parâmetros estimados das simulações I a V para os pools de proteínas ( $Q_{1-I}$ ,  $Q_{1-II}$ ,  $Q_2$  e  $Q_3$ ) observados durante a cromatografia de troca aniônica. 110
- Tabela 5.1. Descrição das amostras de entrada e definição dos perfis de eluição  $Q_1$ ,  $Q_2$  e  $Q_3$  para as cromatografias de troca aniônica (CTA) I a IX realizadas em coluna Q-Sepharose de 70 mL. 125
- Tabela 5.2. Características dos materiais de partida e dos perfis de eluição ( $Q_1$ ,  $Q_2$ , e  $Q_3$ ) para as corridas cromatográficas (CTA) I a IX. 133
- Tabela 5.3. Condições iniciais e parâmetros fixos utilizados para simular o processo de cromatografia de troca aniônica. 136
- Tabela 5.4. Parâmetros estimados da isoterma de ação em efeito estérico (SMA) para os pools de proteínas (1-I, 1-II, 2 e 3) das cromatografias de troca aniônica CTA-I, CTA-II e CTA-III e para PspA4Pro eluído nas corridas CTA-IV, CTA-V e CTA-VI. 136
- Tabela 5.5. Comparação dos tempos de residências e concentração máxima de proteínas ( $C_{P,max}$ ) para os picos 1-I, 1-II, 2/PspA4Pro e 3 dos cromatogramas experimentais e simulados de CTA-VII. 141
- Tabela 5.6. Valores experimentais e simulados de recuperações ( $REC_{1:N}$ ) e purezas ( $P_{1:N}$ ) de CTA-VII seguidos pelos valores experimentais de recuperações ( $REC_{0:\Delta t}$ ) e purezas ( $P_{0:\Delta t}$ ) de

CTA-VIII e CTA-IX para os perfis de eluições alternativos compostos por N subfrações. 141

Tabela D.1. Parâmetros ajustados durante a simulação usando o modelo de Langmuir competitivo. 180

## LISTA DE SÍMBOLOS

$\epsilon_B$ : Porosidade do leito

$\Delta$ : Largura do pico gerado por uma molécula sem afinidade pela resina

$\Psi$  ou  $\Psi_i$ : Parâmetro empírico da isoterma de Langmuir modificada

$\Pi$  ou  $\Pi_i$ : Parâmetro empírico da isoterma de Langmuir modificada

$\epsilon_P$ : Porosidade da partícula

$\epsilon_T$ : Porosidade total da coluna de adsorção

$\hat{y}_i$ : Valores simulados na saída do processo

$\alpha$  ou  $\alpha_i$ : Fator estérico de uma proteína ou adsorbato

$\Delta t$ : Tempo de eluição total de um perfil de eluição alternativo

$\Delta Z$ : Comprimento de cada prato teórico da coluna

$\theta$ : Fração de sítios ativos ocupados em uma resina

$\theta$ : Número de elementos de um perfil de eluição alternativo quando  $t_i - t_{i-1} \simeq 0$

$\theta_A$ : Fração de sítios ativos ocupados por A em uma resina

$\theta_B$ : Fração de sítios ativos ocupados por B em uma resina

$\theta_V$ : Fração de sítios ativos vazios em uma resina

$\Lambda$ : Número de sítios ativos ligantes disponíveis por volume de resina (mols/mL<sub>resina</sub>)

$\sigma$ : Número de parâmetros estimados

## LISTA DE ABREVIACÕES

%C: Porcentagem de condutividade

1-I: Primeiro pool de proteínas eluídas em  $Q_1$

1-II: Segundo pool de proteínas eluídas em  $Q_1$

Abs: Absorbância

Abs<sub>i</sub>: Absorbância de cada pico “i” do cromatograma

A<sub>Q150-I</sub> e A<sub>Q150-II</sub>: Áreas observadas dos picos Q150-I e Q150-II no cromatograma

BSA: *Bovine serum albumin*

C<sub>0</sub>, C<sub>A</sub>, C<sub>B</sub> e C<sub>C</sub>: Concentrações de adsorbato na saída da coluna de adsorção

C<sub>1</sub>: Clarificado obtido por aplicação de CTAB

C<sub>2</sub>: Clarificado obtido por crioprecipitação

c<sub>a,i</sub>: Concentração do elemento A na saída do reator “i” (mol/L)

c<sub>a,in</sub>: Concentração do elemento A na entrada do sistema de reatores (mol/L)

c<sub>a,out</sub>: Concentração do elemento A na saída do sistema de reatores (mol/L)

CA: Concentração de PspA4Pro na fase líquida (Apêndice E)

C<sub>AS</sub>, C<sub>BS</sub>, C<sub>A</sub>, C<sub>B</sub> e C<sub>S</sub>: Concentrações dos elementos teóricos AS, BS, A, B e S na fase líquida (g/mL)

CB: Concentração de impurezas proteicas na fase líquida (Apêndice E)

c<sub>ext</sub>: Concentração de adsorbato na fase líquida (g/mL)

c<sub>i(1-Nt)</sub>: Concentração inicial de proteínas não adsorvidas na resina no interior da coluna (g/mL<sub>resina</sub>)

c<sub>i,e</sub>: Concentração dos “i” grupos proteicos no extrato celular clarificado

c<sub>i</sub>: Concentração de adsorbato ou proteína na fase móvel (g/mL),

c<sub>N</sub>: Concentração na fase líquida do elemento teórico N (g/mL)

c<sub>NaCl</sub>: Concentração de NaCl na fase líquida (g/mL)

C<sub>NaCl</sub>: Concentração de NaCl no cromatograma (Apêndice B)

c<sub>p(r)</sub>: Concentração de proteína ou adsorbato na fase líquida no interior da partícula esférica de resina (g/mL)

C<sub>P,max</sub>: Concentração máxima de proteínas observada em um pico

c<sub>p</sub>: Concentração de proteína ou adsorbato na fase líquida quando  $r=R_p$  (g/mL)

CS: Concentração de sal na fase líquida (Apêndice E)

$c_s$ : Concentração do soluto na fase estacionária que não está em equilíbrio com a fase líquida (g/mL)

$c_{sal}$  ou  $c_{salt}$ : Concentração de sal na fase líquida (g/mL)

CTA: Cromatografia de troca aniônica

CTAB: Brometo de cetiltrimetilamônio

D: Distância que a luz atravessa pelo corpo de medição do cromatógrafo

$D_a$ : Coeficiente de dispersão axial ( $\text{cm}^2/\text{s}$ )

$D_p$ : Coeficiente de difusividade aparente no interior de uma partícula esférica de resina ( $\text{cm}^2/\text{s}$ ).

$d_p$ : Diâmetro da partícula esférica de resina

dt: Elemento diferencial do tempo

dz: Elemento diferencial da distância percorrida na coluna

ECC: Extrato celular clarificado

EMSO: *Environment for modeling, simulation and optimization*

F: Fluxo da fase móvel

FDA: *Food and Drug Administration*

$F_{in}$ : Fluxo na entrada no reator (L/s)

$F_{out}$ : Fluxo na saída no reator (L/s)

G-REC: Recuperação global da cromatografia de troca aniônica

HCAP: Homogeneizador contínuo de alta pressão

ILM: Isoterma de Langmuir modificada

IPTG: Isopropil  $\beta$ -D-1-tiogalactopiranosídeo

$K_A$ : Constante de adsorção do elemento teórico A (mL/g)

$K_{Ads}$  ou  $K_{Ads,i}$ : Constante de adsorção

$K_{Ads}^*$  ou  $K_i'$ : Constante de adsorção corrigida ( $\text{g}/\text{mL}_{resina}$ )

$k_{a,i}$ : Constante de adsorção da proteína “i” (Apêndice E)

$K_B$ : Constante de adsorção do elemento teórico B (mL/g)

kDa: Kilodalton

$K_{Des}$  ou  $K_{Des,i}$ : Constante de dessorção

$K_{d_i}$ : Constante de dessorção da proteína “i” (Apêndice E)

$K_{Eq}$  ou  $K_{Eq,i}$  : Constante de equilíbrio da adsorção

$K_{Eq,i}$ : Constante de equilíbrio de um grupo de proteínas “i”

$K_i$ : Constante de absorvidade molar de cada pico “i” do cromatograma (Apêndice B)

$K_i$ : Constante de adsorção do elemento “i” (mL/g)

$k_m$ : Coeficiente aparente de transferência de massa na parede da partícula de resina ( $s^{-1}$ )

$K_N$ : Constante de adsorção do elemento teórico N (mL/g)

$K'_i$ : Constante de absorvidade molar de cada pico “i” do cromatograma multiplicada pela distância do corpo de medição (Apêndice B)

L: Comprimento da coluna (cm)

LPS: Lipopolissacarídeos

$M_A$ : Massa total de proteínas eluídas no experimento A

MED: Modelo do Equilíbrio Dispersivo

$M_r$ : Fluxo molar de soluto observado entre a solução e a superfície externa da partícula esférica de resina (mol/s.L).

MG: Modelo geral

$M_i$ : Massa de proteínas na fração “i” (g)

MI: Modelo ideal

$M_{P,i}$ : Massa de proteínas totais em cada pico “i” do cromatograma (Apêndice B)

$M_{P,Q150-I}$  e  $M_{P,Q150-II}$ : Massas de proteínas totais dos picos Q150-I e Q150-II no cromatograma

$M_{PspA4Pro,i}$ : Massa de PspA4Pro na fração “i” (g)

$M_{Q1-I}$ : Massa de proteínas eluídas no grupo proteico Q1-I

$M_{Q1-II}$ : Massa de proteínas eluídas no grupo proteico Q1-II

$M_{Q2}$ : Massa de proteínas eluídas no grupo proteico Q2

$M_{Q3}$ : Massa de proteínas eluídas no grupo proteico Q3

MTD: Modelo do transporte dispersivo

MTP: Massa total de proteínas

MTP(t): Perfil de massa total de proteínas

$MTP_{ECCB}$ : Massa total de proteínas na fração  $ECC_B$  (g)

$MTP_i$ : Massa total de proteínas na subfração “i” de  $Q_2$  (g)

N: Número de elementos do perfil de eluição alternativo (Capítulo 5)

N: Número de pontos experimentais

$N_t$ : Número total de pratos teóricos da coluna

P(t): Perfil de pureza de PspA4Pro

P: Concentração de proteínas no cromatograma (Apêndice B)

$P_{0-\Delta t}$ : Pureza dos pools alternativos eluídos em um  $\Delta t$  específico (%)

$P_{1-N}$ : Pureza dos pools alternativos compostos por N subfrações (%)

$P_{ECCB}$ : Pureza da fração  $ECC_B$  (%)

pH: Potencial hidrogeniônico

$P_{i,n}$ : Concentração de proteínas em cada ponto amostrado de cada pico “i” (Apêndice B)

$P_i$ : Concentração de proteínas de cada pico “i” do cromatograma (Apêndice B)

pI: Ponto isoelétrico

$P_i$ : Pureza de PspA4Pro na fração “i” (%)

$P_i$ : Pureza da subfração “i” de  $Q_2$  (%)

PspA: Proteína A de Superfície de pneumococo

PspA4Pro: Proteína recombinante produzida em *Escherichia coli* BL21 (DE3) com o plasmídeo pET37b + PspA4Pro, expressando o gene de um fragmento de PspA.

$P_{T,i}$ : Concentração de proteínas totais em cada pico “i” do cromatograma

Q: Concentração de sal nas fases sólidas (mol/mL<sub>resina</sub>) (Capítulo 2)

Q: Número total de sítios ativos disponíveis na resina (mol/mL<sub>resina</sub>)

$Q_1^*$ : Eluição aplicada na coluna de troca aniônica com baixa força iônica

$Q_1$ : Fração obtida pela eluição  $Q_1^*$

Q1000: Grupo proteico observado na terceira eluição

Q150-I: Grupo proteico observado na primeira eluição

Q150-II: Grupo proteico observado na primeira eluição

$Q_1$ -I e  $Q_1$ -II: Pools de proteínas I e II eluídas na fração  $Q_1$

$Q_2^*$ : Eluição aplicada coluna de troca aniônica com média força iônica

$Q_2$ : Fração obtida pela eluição  $Q_2^*$

$Q_3^*$ : Eluição aplicada coluna de troca aniônica com alta força iônica

$Q_3$ : Fração obtida pela eluição  $Q_3^*$

Q300: Grupo proteico observado na segunda eluição

qA: Concentração de PspA4Pro na fase sólida (Apêndice E)

qB: Concentração de impurezas proteicas na fase sólida (Apêndice E)

$q_{i(1-Nt)}$ : Concentração inicial de proteínas adsorvidas pela resina no interior da coluna ( $\text{g/mL}_{\text{resina}}$ )

$q_i$ : Quantidade de adsorbato ou proteína na estrutura sólida em equilíbrio com a fase líquida ( $\text{g/mL}_{\text{resina}}$  ou  $\text{g/g}_{\text{resina}}$ ),

$q_m$ : Concentração média de soluto adsorvido ( $\text{g/mL}_{\text{resina}}$ )

$q_{\text{max}}^*$  ou  $q_{\text{max},i}$ : Quantidade máxima corrigida de adsorção na fase sólida

$q_{\text{max},A}$ : Quantidade máxima de sítios ativos possíveis a serem ocupados por A ( $\text{g/mL}_{\text{resina}}$ )

$q_{\text{max},B}$ : Quantidade máxima de sítios ativos possíveis a serem ocupados por B ( $\text{g/mL}_{\text{resina}}$ )

$q_{\text{max},i}$ : Quantidade máxima de sítios ativos possíveis a serem ocupados pelo elemento “i” ( $\text{g/mL}_{\text{resina}}$ )

$q_{\text{max}}$ : Quantidade máxima de sítios ativos possíveis a serem ocupados pelo adsorbato ( $\text{g/mL}_{\text{resina}}$ )

QNAds: Fração não adsorvida pela coluna de troca aniônica

qS: Concentração de sal na fase sólida (Apêndice E)

$r_a$ : Velocidade de reação ( $\text{mol/L.s}$ )

REC(t): Perfil de recuperação de PspA4Pro

$\text{REC}_{0-\Delta t}$ : Recuperação de PspA4Pro em pools alternativos eluídos em um  $\Delta t$  específico (%)

$\text{REC}_{1-N}$ : Recuperação de PspA4Pro em pools alternativos compostos por N subfrações (%)

$\text{REC}_i$ : Recuperação de PspA4Pro na subfração “i” de  $Q_2$  (%)

$R_i$ : Recuperação de PspA4Pro na fração “i” (%)

RMSE: *Root mean square error*

$R_p$ : Raio de uma partícula esférica de resina (cm)

rpm: *Rotations per minute*

SMA: *Steric mass action isotherm* ou isoterma de ação em efeito estérico

SP<sub>1</sub>\*: Eluição aplicada na coluna de troca catiônica com baixa força iônica

SP<sub>1</sub>: Fração obtida pela eluição SP<sub>1</sub>\*

SP<sub>2</sub>\*: Eluição aplicada na coluna de troca catiônica com média força iônica

SP<sub>2</sub>: Fração obtida pela eluição SP<sub>2</sub>\*

SP<sub>3</sub>\*: Eluição aplicada coluna de troca catiônica com alta força iônica

SP<sub>3</sub>: Fração obtida pela eluição SP<sub>3</sub>\*

SPNAds: Fração não adsorvida pela coluna de troca catiônica

SPP: Solução de proteína pura ou solução de PspA4Pro pura

t<sub>i</sub>: Tempo de eluição da subfração “i”

t<sub>0</sub>: Tempo de retenção de uma molécula sem afinidade pela resina

u: Velocidade da fase móvel (cm/s)

UV: Ultravioleta

v ou v<sub>i</sub>: Número de sítios ligantes presentes na molécula do adsorbato ou proteína

v: Volume de cada ponto amostrado pelo cromatógrafo (Apêndice B)

V: Volume do reator (L)

V<sub>0</sub>: Volume para obtenção do pico de uma molécula que não apresenta afinidade com a resina

V<sub>1</sub> a V<sub>5</sub>: Válvulas 1 a 5 do sistema de cromatografia de troca aniônica

v<sub>1</sub> e v<sub>2</sub>: Coeficientes estequiométricos da proteína e do sal, respectivamente, em uma equação representando o equilíbrio.

VC: Volume de coluna

V<sub>C</sub>: Volume de coluna

V<sub>ECC</sub>: Volume do extrato celular clarificado (*clarified cell extract*)

v<sub>i</sub>: Número de sítios ligantes da proteína “i” (Apêndice E)

V<sub>T,i</sub>: Volume total da fração eluída para cada pico “i” do cromatograma

x<sub>i</sub>: Fator estérico da proteína “i” (Apêndice E)

y<sub>i</sub>: Valores experimentais na saída do processo

ZTM: Zona de transferência de massa

## SUMÁRIO

<b>1. INTRODUÇÃO</b>	<b>26</b>
1.1. Contextualização	26
1.2. Objetivos	28
1.3. Estrutura do texto	29
<b>2. REVISÃO BIBLIOGRÁFICA</b>	<b>30</b>
2.1. <i>Escherichia coli</i> como fábrica celular de proteínas para fins farmacêuticos	30
2.2. As vacinas antipneumocócicas	31
2.3. Processos <i>Upstream</i> para a produção de PspA4Pro	32
2.4. Processos <i>Downstream</i> para purificação de proteínas com aplicação farmacêutica	35
2.4.1. Metodologias de ruptura celular	38
2.4.2. Operações de precipitação	39
2.4.3. Processo de adsorção	40
2.4.3.A. Princípios de adsorção	40
2.4.3.B. Adsorção em leito fixo	41
2.4.4. Processos cromatográficos para purificação de proteínas	43
2.5. Modelagem matemática de processos cromatográficos	45
2.5.A. O Modelo ideal (MI)	46
2.5.B. Modelo do equilíbrio dispersivo (MED)	47
2.5.C. Modelo do transporte dispersivo (MTD)	48
2.5.D. O Modelo geral (MG)	50
2.6. Isotermas de adsorção	52
2.6.A. A isoterma linear	52
2.6.B. A isoterma de Langmuir	53
2.6.C. A isoterma de Langmuir competitiva	55
2.6.D. A isoterma de Langmuir modificada	59
2.6.E. A isoterma de ação em efeito estérico	60
2.7. O software EMSO como simulador orientado a equações	64

<b>3. MODELAGEM E SIMULAÇÃO DA CROMATOGRAFIA DE TROCA ANIÔNICA PARA PURIFICAÇÃO DE PROTEÍNAS EM MISTURAS COMPLEXAS</b>	<b>66</b>
3.1. Introdução	68
3.2. Teoria	70
3.2.1. Modelo do equilíbrio dispersivo (MED)	70
3.2.2. Isoterma de Langmuir modificada	71
3.2.3. Isoterma de ação em efeito estérico ( ou “ <i>steric mass action isotherm</i> ”)	73
3.3. Materiais e métodos	74
3.3.1. Produção de PspA4Pro	74
3.3.2. Cromatografia de troca aniônica	75
3.3.3. Quantificação de proteínas e determinação de pureza da PspA4Pro	75
3.3.4. Modelagem e simulação	76
3.3.4.A. Características de fluxo da coluna	76
3.3.4.B. Estimativas de parâmetros, simulações e validações experimentais	76
3.3.5. Estimativa dos parâmetros e validações experimentais do extrato celular clarificado virtual	79
3.4. Resultados e Discussão	81
3.4.1. Visão geral das cromatografias de troca aniônica	81
3.4.2. Estimativa de parâmetros, simulações e validações experimentais dos modelos de adsorção	85
3.4.3. Estimativa de parâmetros de extratos celulares virtuais, simulações e validações experimentais	90
3.5. Conclusões	92
<b>4. OTIMIZAÇÃO DA CAPACIDADE DE PROCESSAMENTO DA CROMATOGRAFIA DE TROCA ANIÔNICA NA PURIFICAÇÃO DE PROTEÍNAS DE SUPERFÍCIE DE PNEUMOCOCO</b>	<b>95</b>
4.1. Introdução	97
4.2. Materiais e métodos	98
4.2.1. Produção de PspA4Pro	98
4.2.2. Cromatografia de troca aniônica	99

4.2.3. Quantificação de proteínas, pureza de PspA4Pro e determinação da recuperação	101
4.2.4. Determinação de parâmetros experimentais	102
4.2.5. Estratégias de simulação do processo	102
4.2.6. Estimativas de parâmetros e validações experimentais	104
4.2.7. Determinação da capacidade de processamento da resina	106
4.3. Resultados e discussão	106
4.3.1. Visão geral das cromatografias de troca aniônica	106
4.3.2. Estimativas de parâmetros, simulações e validações experimentais das cromatografias	109
4.3.3. Simulações dos ECC virtuais e determinação da capacidade de processamento da resina	114
4.4. Conclusões	116
<b>5. UMA ABORDAGEM ROBUSTA PARA MODELAR E SIMULAR PROCESSOS CROMATOGRÁFICOS NA PURIFICAÇÃO DE MISTURAS COMPLEXAS</b>	<b>117</b>
5.1. Introdução	119
5.2. Teoria	121
5.2.1. Modelo do equilíbrio dispersivo	121
5.2.2. A isoterma de ação em efeito estérico	122
5.3. Materiais e métodos	123
5.3.1. Produção de PspA4Pro	123
5.3.2. Cromatografia de troca aniônica	124
5.3.3. Quantificação de proteínas e determinação de pureza de PspA4Pro	126
5.3.4. Modelagem e simulação	128
5.3.4.A. Estimativas de parâmetros de PspA4Pro e das impurezas proteicas	129
5.3.4.B. Simulações e validações experimentais	130
5.3.4.C. Simulações dos perfis de eluições modificados e validações experimentais	132
5.4. Resultados e discussão	133
5.4.1. Visão geral das cromatografias de troca aniônica	133
5.4.2. Estimativa de parâmetros e simulação da cromatografia de troca aniônica	135
5.4.3. Simulações e validações experimentais dos perfis de eluição modificados	143
5.5. Conclusões	150

<b>6. CONSIDERAÇÕES FINAIS</b>	<b>152</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>154</b>
<b>APÊNDICES</b>	<b>164</b>
APÊNDICE A: Exemplo de aplicação do EMSO	164
APÊNDICE B: Disponibilização dos dados de condutividade e absorbância em forma de concentrações	167
APÊNDICE C: Roteiro de modelagem da cromatografia de troca aniônica	171
APÊNDICE D: Simulação da cromatografia de troca aniônica com o modelo de Langmuir competitivo	179
APÊNDICE E: Material suplementar - Código da modelagem e simulação da cromatografia de troca aniônica	182

## 1. INTRODUÇÃO

### 1.1. Contextualização

O desenvolvimento da tecnologia de DNA recombinante a partir de 1970 permitiu a expressão de genes heterólogos em hospedeiros eucarióticos ou procarióticos, com grande impacto na produção de proteínas com aplicações farmacêuticas. No ano de 2017 este mercado foi avaliado em US \$91,2 bilhões com crescimento anual projetado de 12,2% até 2026 (INSIGHTS, 2018). Nos últimos trinta anos, cerca de 400 produtos baseados em proteínas recombinantes provenientes de organismos vivos foram utilizados como biofármacos. Outros 1300 ainda estão em desenvolvimento, dos quais 33% se encontram em fase de ensaios clínicos (GLOBAL DATA, 2015).

Um terço dessas proteínas e metade dos produtos aprovados por agências reguladoras internacionais são produzidos tendo *Escherichia coli* geneticamente modificada como plataforma de expressão (MAMAT et al., 2015). A escolha desta bactéria como fábrica microbiana se deve ao vasto conhecimento acumulado sobre sua genética e metabolismo, à disponibilidade de métodos precisos e rápidos para a modificação de seu genoma (SHILOACH et al., 2005; SHOJA et al., 2008) e à facilidade de obtenção de elevadas quantidades de biomassa utilizando protocolos de cultivo bem estabelecidos (MAMAT et al., 2015). Vetores de expressão muito difundidos para a produção de proteínas recombinantes em *E. coli* são baseados no operon *lac*, onde lactose e seu análogo, o isopropil  $\beta$ -D-1-tiogalactopiranosídeo (IPTG), são utilizados para a indução da produção da proteína de interesse, que geralmente se acumula no interior da célula (SORENSEN et al., 2005).

A recuperação do bioproduto ocorre de acordo com o procedimento convencional, envolvendo etapas de clarificação, rompimento, isolamento e purificação (PESSOA JR. et al., 2005). O principal desafio reside na obtenção do produto com alta pureza, contendo teores de impurezas proteicas, ácidos nucleicos e pirógenos abaixo dos limites estabelecidos pelas agências reguladoras (MAMAT et al., 2015), o que demanda que uma adequada estratégia de purificação seja desenvolvida com base nas diferenças entre as características físico-químicas da molécula de interesse e dos contaminantes, portanto, os protocolos podem ser extremamente variados (SCOPES et al., 2013). As etapas de purificação de produtos biotecnológicos contribuem com 20 a 80% do custo total do processo, dependendo das características do material de partida, tais como concentração do produto de interesse, concentração de contaminantes, tipo de contaminantes e grau de pureza exigido (SOFER et al., 1997).

Tendo em vista estes aspectos, o projeto de pesquisa “Intensificação e integração de processos de produção e purificação de proteína A de Superfície de pneumococo” (FAPESP 2015/10.291-8), no qual se insere o presente tema de Doutorado, tem se dedicado a estudar a purificação da proteína PspA4Pro presente na biomassa obtida a partir dos cultivos da linhagem de *Escherichia coli* BL21(DE3) pET37b+PspA4Pro, que expressa o fragmento da proteína PspA da família 2 do clado 4 (MIYAJI et al., 2013). O uso de PspA como proteína carreadora na formulação de vacinas conjugadas contribui para aumentar a proteção contra infecções de *Streptococcus pneumoniae*, bactéria causadora de doenças como pneumonia, infecções do ouvido, sinusite e meningite (ALONSO DE VELASCO et al., 1995). Além disso, a PspA também pode ser empregada como vacina de subunidade antigênica, o que tem sido bastante explorado por diversos grupos de pesquisa em todo mundo (BOGAERT et al., 2003; BRILES et al., 2000) e também pelo Centro de Biotecnologia do Instituto Butantan (MIYAJI et al., 2013; VADESILHO et al., 2014; PERCIANI et al., 2013; MORENO et al., 2013).

Sob a coordenação da Dra. Viviane M. Gonçalves foi desenvolvido um processo de purificação da PspA4Pro envolvendo a seguinte sequência de etapas: i) ressuspensão e ruptura da biomassa celular em homogeneizador contínuo a alta pressão, ii) precipitação de debris celulares com o detergente catiônico brometo de cetiltrimetilamônio (CTAB), iii) realização de cromatografia de troca aniônica, iv) congelamento, descongelamento e centrifugação da amostra e v) realização de cromatografia de troca catiônica. Apesar dos resultados alcançados em termos de pureza atenderem aos requisitos de qualidade estabelecidos pelos órgãos reguladores da indústria farmacêutica, verificou-se perda significativa do produto de interesse ao longo da purificação, principalmente na cromatografia de troca aniônica, a etapa mais cara do processo *downstream* devido ao alto valor de mercado da resina cromatográfica (FIGUEIREDO, 2014). Nota-se portanto a necessidade de um estudo direcionado à modelagem desta etapa, podendo ser uma estratégia interessante para identificar pontos de otimização de pureza e rendimento, viabilizando a descrição do processo com modelos matemáticos que visem representar os comportamentos adsorptivos das proteínas e os fenômenos de transferência de massa. Recentemente, os avanços computacionais permitiram o desenvolvimento de modelos de simulação capazes de descrever aspectos da cromatografia relacionados à cinética e difusividade na resina (ORELLANA et al., 2009; LINQUEO et al., 2012). No entanto, na maioria dos casos, misturas sintéticas formuladas com um número limitado de proteínas puras são usadas para extrair dados para a simulação, o que é adequado

para o planejamento experimental, mas reduz o poder descritivo das simulações, já que extratos celulares reais são compostos pela proteína de interesse e uma mistura de proteínas desconhecidas com diferentes propriedades físico-químicas.

Neste contexto, o presente trabalho pretende realizar a modelagem do processo de purificação de PspA4Pro, além de contribuir para o desenvolvimento de estratégias robustas para modelar processos cromatográficos aplicados a misturas complexas, contendo impurezas proteicas, representativas das situações reais de purificação encontradas em processos industriais. A abordagem para desenvolvimento do trabalho envolveu inicialmente realizar um levantamento das isotermas de adsorção mais viáveis para descrever matematicamente a cromatografia de troca aniônica durante a purificação de PspA4Pro. Em um segundo momento o processo foi modelado utilizando a plataforma EMSO (*Environment for modeling, simulation and optimization*), um simulador orientado a equações que permite descrever sistemas dinâmicos e estacionários (SOARES e SECCHI, 2003; THOMBRE et al., 2015; IPSITABANERJEE et al., 2010; FURLAN, 2016). Para isso, balanços materiais da proteína de interesse e dos diferentes contaminantes foram aliados às isotermas de adsorção e usados como base para a modelagem. Por fim, uma vez que a simulação foi capaz de descrever os dados experimentais de forma eficiente, uma metodologia de validação cruzada foi desenvolvida e estratégias alternativas de purificação foram testadas, visando o aumento de rendimento e pureza da PspA4Pro produzida e diminuindo a complexidade do processo e os custos de produção.

## 1.2. Objetivos

O objetivo geral deste trabalho é modelar e otimizar a etapa de cromatografia de troca aniônica na purificação da proteína PspA4Pro, visando compreender os fenômenos de adsorção e transferência de massa envolvidos e identificar condições operacionais que aumentem a eficiência do processo. Como objetivos específicos, destacam-se:

- Avaliar a capacidade descritiva de diferentes isotermas de adsorção na modelagem da cromatografia de troca aniônica.
- Desenvolver uma metodologia adequada para descrever modelos matemáticos que possam ser aplicados a processos cromatográficos de misturas complexas.
- Utilizar os modelos desenvolvidos na simulação e variar características operacionais da cromatografia de troca aniônica, visando encontrar alternativas de aumento de pureza e rendimento da PspA4Pro obtida.

### **1.3. Estrutura do texto**

O texto desta tese foi estruturado conforme detalhado abaixo. Após a Revisão Bibliográfica, os estudos realizados (partes B a D) foram disponibilizados em forma de artigos científicos. No final do texto, encontra-se o tópico de conclusões e sugestões para trabalhos futuros e os Apêndices.

**A) Revisão Bibliográfica**

**B) Artigo 1: “Modelagem e simulação da cromatografia de troca aniônica para a purificação de proteínas em misturas complexas”**

**C) Artigo 2: “Otimização da capacidade de processamento da cromatografia de troca aniônica na purificação de proteínas de superfície de pneumococo”**

**D) Artigo 3: “Uma abordagem sólida para modelar e simular processos cromatográficos na purificação de misturas complexas”**

**E) Considerações finais e sugestões para trabalhos futuros**

**F) Apêndices**

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1. *Escherichia coli* como fábrica celular de proteínas para fins farmacêuticos

O desenvolvimento da tecnologia de DNA recombinante a partir de 1970 permitiu a expressão de genes heterólogos em hospedeiros eucarióticos ou procarióticos, com grande impacto na produção microbiológica de proteínas com aplicações farmacêuticas (MAMAT et al., 2015; KAMIONKA et al., 2011). Em 1982, o primeiro produto desta tecnologia foi aprovado: a insulina humana para tratar diabetes (GOEDDEL et al., 1979). Desde então, aproximadamente 400 produtos baseados em proteínas recombinantes provenientes de organismos vivos foram desenvolvidos com sucesso e utilizados como biofármacos. Outros 1300 ainda estão em desenvolvimento, dos quais 33% se encontram em fase de ensaios clínicos (Global Data, 2015). A possibilidade da produção escalonada dessas proteínas gerou novas oportunidades às empresas farmacêuticas, pois apresentam alto valor agregado e oferecem alternativas tecnológicas para a produção através de procedimentos relativamente baratos (FERRER-MIRALLES et al., 2009).

A *Escherichia coli* é um dos microrganismos que mais se destaca neste mercado, já que suas características de clonagem e modificação genética são amplamente conhecidas. O vasto conhecimento acumulado de seu metabolismo facilita a obtenção de elevadas quantidades de biomassa através de otimização da composição do meio de cultura e utilização de protocolos de cultivo bem estabelecidos (MAMAT et al., 2015; ASSENBERG et al., 2013). Apesar disso, os principais problemas observados no cultivo de *Escherichia coli* estão relacionados à alta produção de CO<sub>2</sub>, redução de pH do meio de cultivo devido à presença de acetato e limitações na transferência de oxigênio para a célula (ASSENBERG et al., 2013). Dentre os produtos aprovados provenientes de *Escherichia coli* se incluem vários hormônios importantes para o mercado farmacêutico, tais como insulina, calcitonina, hormônio de crescimento e somatropina (FERRER-MIRALLES et al., 2009). Dentre os 151 fármacos baseados em proteínas recombinantes licenciados até Janeiro de 2009 pela FDA, 29% são obtidos por *Escherichia coli* (REDWAN, 2007).

## 2.2. As vacinas antipneumocócicas

*Streptococcus pneumoniae*, agente causador de patologias como pneumonia, sinusite e meningite, é um microrganismo Gram-positivo que apresenta a capacidade de se ligar à lactoferrina, glicoproteína que faz parte do sistema imune inato, o que impede a ação bactericida sobre o mesmo (SENKOVICH et al., 2007). Sua membrana celular é recoberta por uma cápsula de polissacarídeo, cuja estrutura é responsável pela definição dos diferentes sorotipos existentes. Além disso, a cápsula de polissacarídeo é o principal fator de virulência (MADHI et al., 2008). A incidência de doenças causadas por pneumococos é mais elevada em países subdesenvolvidos (Figura 2.1), onde o acesso a tratamentos médico hospitalares é menor (MADHI et al., 2008). A taxa de mortalidade associada à infecção chega a ser de 20% para crianças com menos de dois anos e 60% para indivíduos com mais de 65 anos (CILLÓNIZ, 2012). No Brasil, encontramos doze sorotipos que estão diretamente relacionados a 75% das infecções, das quais 45% provém dos sorotipos 14, 6B e 1 (WAEGEMAN et al., 2011). O aumento da resistência do pneumococo à penicilina e a outros antibióticos contribui para agravar ainda mais esse quadro (MANGTANI et al., 2003). A imunização por meio de vacinas antipneumocócicas é a forma mais eficaz de prevenção das doenças causadas por *S. pneumoniae*. A maioria das vacinas existentes no mercado são formuladas com a combinação de polissacarídeos de diferentes sorotipos, conjugados ou não a uma proteína, e apresentam cobertura limitada.

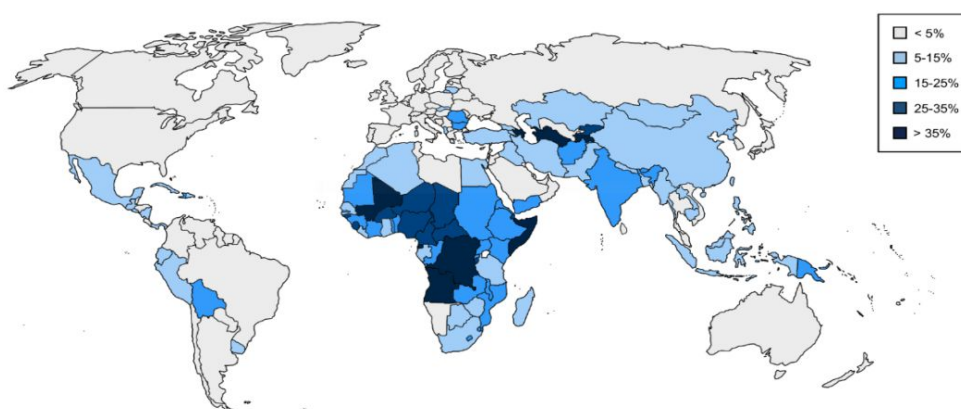


Figura 2.1. Mortalidade estimada em crianças com menos de cinco anos de idade expostas a doenças causadas por pneumococos (Fonte: Abcombi Biosciences, Data de acesso: 05/2018).

Para que a cobertura de imunidade seja ampliada, uma nova vacina está sendo desenvolvida para o mercado nacional, com a finalidade de conjugar os polissacarídeos dos três sorotipos prevalentes no Brasil (14, 6B e 1) com PspA, uma proteína da parede celular do pneumococo exposta em sua cápsula polissacarídica (BARAZZONE et al., 2011). A produção de antígenos vacinais em grande quantidade é viabilizada pela clonagem destas proteínas em organismos de fácil cultivo, como *E. coli*. No presente trabalho, utilizou-se *Escherichia coli* recombinante para a produção de PspA. Para isso, células de *E. coli* foram transformadas com o fragmento do clado 4 que contém a porção N-terminal, a região definidora de clado e o primeiro trecho da região rica em prolinas. O fragmento foi denominado PspA4Pro, com massa molecular aproximada de 40 kDa, ponto isoelétrico de 4.8 e sem cauda de histidina. A célula *E. coli* BL21(DE3) pET37b+PspA4Pro foi obtida pela Dra. Eliane Miyaji do Centro de Biotecnologia do Instituto Butantan. Nesta construção, a proteína PspA4Pro se acumula no interior da célula, na forma solúvel. A produção da proteína utiliza vetores de expressão baseados no operon *lac*, onde lactose e/ou seu análogo, o isopropil  $\beta$ -D-1-tiogalactopiranosídeo (IPTG), são os indutores (SORENSEN et al., 2005).

Neste contexto, a obtenção de PspA4Pro pura e dentro dos critérios de qualidade exigidos pelo mercado farmacêutico depende das estratégias utilizadas durante o cultivo e a purificação (etapas *upstream* e *downstream*, respectivamente). Nas etapas *upstream*, alguns fatores são determinantes para a produção de menor quantidade de proteínas contaminantes, como o tipo de meio utilizado, a temperatura de cultivo e técnicas de aeração e indução (SCOPEL et al., 2013). As características físico-químicas desses contaminantes também devem ser consideradas no desenvolvimento de estratégias para o processo *downstream*. Por isso, é fundamental identificar as respostas fisiológicas das células à indução e, principalmente incorporar essa questão no desenvolvimento de estratégias de cultivo.

### **2.3. Processos *Upstream* para a produção de PspA4Pro**

Durante o cultivo de *Escherichia coli* são definidas estratégias relacionadas às características do meio, o tipo de biorreator a ser utilizado, as formas de controle de oxigênio, o pH, a temperatura de operação e o uso de indutores. Todas estas decisões resultam em um comportamento metabólico específico. Como se trata de um processo que envolve microrganismo recombinantes, a síntese da proteína de interesse não é um fenômeno comum

à célula. Por isso, o esforço para sua produção pode gerar condições de estresse (KILIKIAN et al., 2000). Cultivos em alta densidade celular têm sido considerados uma estratégia interessante para reduzir custos de produção e tornar o processo mais sustentável, pois apresentam grande acúmulo de biomassa, chegando a atingir concentrações maiores que 100 g/L, o que aumenta também a produtividade da proteína de interesse (RIESENBERG e GUTHKE, 1999; HORTA, 2011). Essa estratégia pode parecer uma opção economicamente viável, já que no processo *downstream* cerca de 70% da proteína é perdida, mas em muitos casos não se observa produção expressiva de PspA4Pro após algumas horas de indução para concentrações de biomassa muito altas. Além disso, o aumento da concentração de biomassa desencadeia várias respostas fisiológicas e reológicas no caldo (HORTA, 2011) causando diversas dificuldades, tais como inibição da célula pelo substrato ou por outros ácidos orgânicos e dificuldades na transferência de oxigênio, já que caldos mais viscosos são mais resistentes à agitação (LEE, 1991).

Para compreender melhor esses fenômenos, SILVA (2015) avaliou as mudanças morfológicas e reológicas do caldo de cultivo de *Escherichia coli*. Os cultivos foram conduzidos em batelada, utilizando diferentes temperaturas e indutores. Foi observado que o uso de lactose como indutor gerou acúmulo de galactose no interior da célula, já que esse açúcar não é assimilado pela linhagem BL21(DE3). Foram observadas mudanças na reologia ligadas à concentração celular e ao estresse causado pela síntese de proteína, respectivamente, e a intensidade da mudança se mostrou dependente do indutor usado. A ocorrência de permeabilização celular, com liberação de conteúdo intracelular para o meio e o aumento na concentração de polissacarídeos também podem ter contribuído para alterações na reologia do caldo (SILVA, 2015).

É muito importante que condições de agitação e aeração do cultivo também sejam otimizadas. Como se trata de um processo aeróbico, a deficiência em oxigênio pode levar a drásticas reduções de produtividade e desempenho. A reduzida transferência de oxigênio em biorreatores ocorre principalmente devido à baixa solubilidade desta molécula em água (7 mg/L em ar atmosférico a 1 atm e 35 °C) e à heterogeneidade do sistema, que é formado por gás, líquido e sólidos (CAMPANI, 2015).

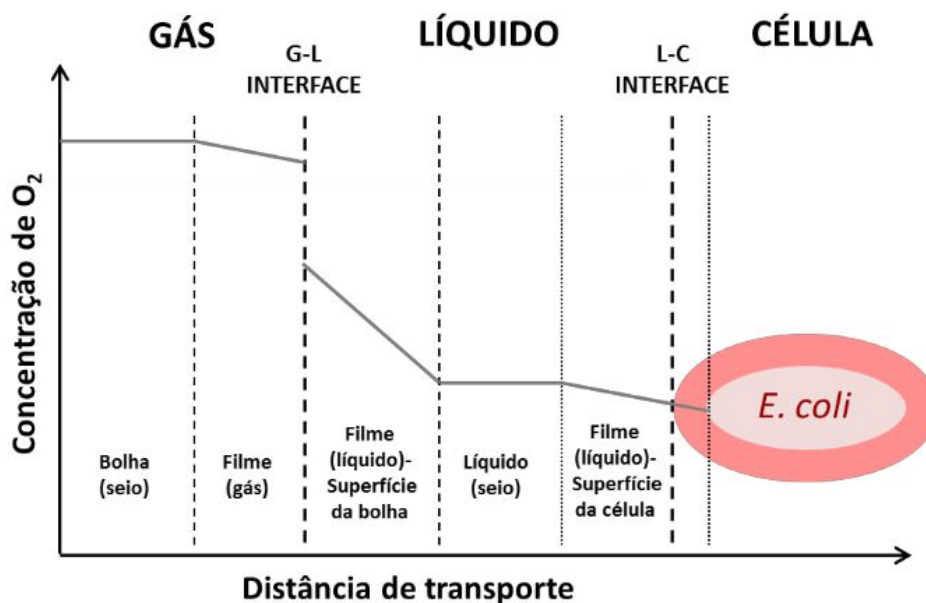


Figura 2.2. Etapas da transferência de oxigênio desde a fase gasosa até o interior da célula (Fonte: CAMPANI, 2014).

A Figura 2.2 mostra o trajeto percorrido pelo  $O_2$  até o interior da célula. A etapa limitante na transferência de massa ocorre no filme líquido que fica estagnado ao redor das bolhas de ar. Pode-se esperar, portanto, melhores produtividades em condições mais intensas de agitação. No entanto, alguns autores mostraram exatamente o contrário: diminuição no rendimento específico de proteínas nessas situações. As justificativas estão relacionadas ao estresse oxidativo e cisalhamento (CASTRO-MARTÍNEZ et al., 2012; WAEGEMAN et al., 2011). Neste contexto, uma estratégia abordada recentemente foi a substituição do reator em batelada convencional tipo tanque agitado pelo reator *airlift*, que apresenta cisalhamento menos intenso, simplicidade de construção, menor risco de contaminação, eficiente dispersão gás-líquido e baixo consumo de energia. O ar é inserido através de injeção em um aspersor localizado na parte inferior, no interior de um tubo concêntrico. A menor densidade da mistura gás-líquido no interior do tubo faz com que haja circulação da suspensão. Desta forma, a fase gasosa percorre todo o reator várias vezes e a transferência de massa é melhorada (CAMPANI, 2014). Em trabalhos recentes foi observada melhor transferência de  $O_2$  em situações em que a pressão interna do reator *airlift* foi aumentada, o que minimizou o uso de oxigênio puro para suprir o cultivo, reduzindo os custos da produção (CAMPANI, 2015; CAMPANI, 2016).

#### **2.4. Processos *Downstream* para purificação de proteínas com aplicação farmacêutica**

No caso de proteínas terapêuticas, o principal desafio da etapa *downstream* reside na obtenção do produto com alta pureza, contendo impurezas como ácidos nucleicos, pirógenos e enzimas abaixo dos limites estabelecidos pelas agências reguladoras (MAMAT et al., 2015; FIGUEIREDO et al., 2014). As etapas de purificação de produtos biotecnológicos contribuem com 20 a 80 % do custo total do processo, dependendo das características do material de partida, como concentração do produto de interesse, concentração de contaminantes, tipo de contaminantes e grau de pureza exigido (SOFER et al., 1997). Assim, é fundamental o desenvolvimento de processos eficientes e de baixo custo. Atualmente tecnologias descartáveis como colunas cromatográficas de uso único têm sido utilizadas, já que não demandam gastos com o tempo de limpeza e reequilíbrio. Além disso, a ausência de processos cíclicos de regeneração faz com que as resinas apresentem maior estabilidade, o que facilita na validação de processos. No entanto, apesar desses benefícios, muitos experimentadores ainda utilizam as resinas clássicas que demandam procedimentos limpeza e sanitização (FIGUEIREDO, 2014).

A precipitação e a cromatografia líquida estão entre as técnicas mais comumente utilizadas na purificação industrial de proteínas. No primeiro caso, o princípio de separação muitas vezes envolve centrifugação: enquanto as proteínas contaminantes ficam predominantemente no precipitado, a proteína de interesse permanece solúvel na fase líquida clarificada. No segundo caso, a amostra a ser purificada é submetida a contato físico-químico com uma resina sólida. Proteínas são então adsorvidas e posteriormente dessorvidas por eluentes que possuem maior interação com a matriz sólida. Muitas vezes os processos de precipitação antecedem a cromatografia para reduzir a quantidade de impurezas totais no material que será transferido para a coluna, sobrecarregando menos a resina cromatográfica que é um insumo de alto custo (BRACEWELL, 2008).

Embora existam muitos estudos na literatura sobre a purificação de proteínas recombinantes, a grande maioria aborda processos de recuperação de proteínas ligadas a elementos como cauda de histidinas (His-tag), glutatona S-transferase ou proteína ligante maltose. Embora esses elementos de fusão facilitem a purificação, os mesmos não podem estar presentes em proteínas com aplicação terapêutica, de acordo com a legislação vigente determinada pelas agências reguladoras. A remoção destes elementos em momento posterior à

síntese da proteína envolve processos enzimáticos, o que impactaria no custo do biofármaco devido ao elevado custo das enzimas e tempo demandado nas etapas exigidas para a recuperação de proteína isenta desses elementos (WAUGH et al., 2011). Uma alternativa em relação a este problema seria produzir as proteínas de interesse sem a fusão. Porém, neste caso, um processo específico de purificação precisaria ser desenvolvido, levando em consideração as características físico-químicas da molécula de interesse e de contaminantes como outras proteínas, DNA, lipopolissacarídeos e endotoxinas.

Um exemplo deste tipo de aplicação foi abordado recentemente por KRISHNAN et al. (2016), através da proposta de um processo de purificação de hipocalcina produzida via intracelular por *Escherichia coli* recombinante. Após lise celular e remoção dos corpos de inclusão a purificação foi realizada em uma única etapa envolvendo cromatografia hidrofóbica, possibilitando obtenção da proteína de interesse sem impurezas proteicas e sem compostos ligantes. CHEN et al. (2017) também apresentou uma eficiente metodologia de purificação de cistatina C produzida por *Escherichia coli* sem a utilização de tags de afinidade. O autor optou por utilizar cromatografia de afinidade com metais imobilizados e cromatografia por exclusão de tamanho, respectivamente, possibilitando a obtenção de proteína com aproximadamente 95% de pureza. FIGUEIREDO (2014) também desenvolveu uma metodologia de purificação de proteína PspA4Pro sem elementos de fusão. O processo desenvolvido é tema central desta tese, e cada uma de suas etapas podem ser verificadas na figura 2.3, conforme descrito a seguir: i) ressuspensão e ruptura da biomassa celular em homogeneizador contínuo a alta pressão, obtendo-se uma fração inicial com baixo teor de pureza chamada homogenato (H); ii) precipitação de debris celulares e outras impurezas com o detergente catiônico brometo de cetiltrimetilamônio (CTAB) e remoção do precipitado por centrifugação, obtendo-se a fração clarificada ( $C_1$ ) separada das impurezas sólidas no corpo de fundo; iii) carregamento da amostra  $C_1$  em coluna de troca aniônica (resina Q-Sepharose) gerando separação das proteínas com afinidade pela resina das não adsorvidas (QNAds); iv) aplicação de eluições  $Q_1^*$ ,  $Q_2^*$  e  $Q_3^*$  com baixa, média e alta forças iônicas, respectivamente, na coluna de troca aniônica, gerando as respectivas frações desorvidas onde  $Q_2$  apresentou PspA4Pro eluída com maior pureza; v) congelamento, descongelamento e centrifugação da fração  $Q_2$  em um processo denominado crioprecipitação, produzindo uma fração clarificada ( $C_2$ ) e impurezas sólidas no corpo de fundo; vi) carregamento da amostra  $C_2$  em coluna de troca catiônica (resina SP-Sepharose), gerando separação das proteínas com afinidade pela resina das não adsorvidas (SPNAds); vii) aplicação de eluições  $SP_1^*$ ,  $SP_2^*$  e  $SP_3^*$  com baixa,

média e alta forças iônicas, respectivamente, em coluna de troca catiônica, gerando as respectivas frações dessorvidas onde SP<sub>2</sub> apresentou PspA4Pro eluída com aproximadamente 95% de pureza e 30% de recuperação mássica global (FIGUEIREDO et al., 2017).

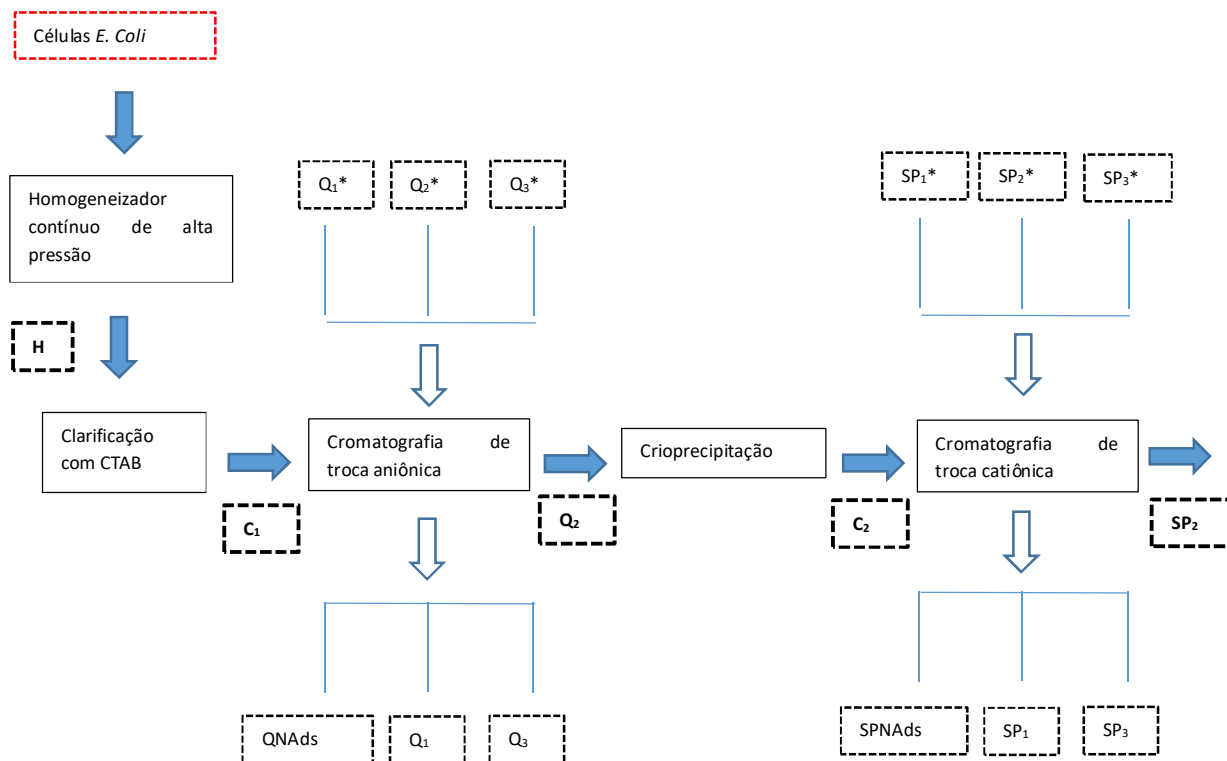


Figura 2.3. Sequência de operações unitárias propostas para purificação de PspA4Pro em células de *Escherichia coli* recombinante.

Apesar dos resultados alcançados em termos de pureza (>95%) atenderem aos requisitos de qualidade estabelecidos pelos órgãos reguladores da indústria farmacêutica, verificou-se perda significativa de PspA4Pro ao longo da purificação, com recuperações entre 14 e 33 %, e desempenho dependente da qualidade do material de partida (FIGUEIREDO, 2015), o que mostra que melhorias ainda podem ser implementadas para reduzir custos operacionais e aumentar a eficiência de recuperação proteica.

Neste contexto, é possível observar que apesar das dificuldades já citadas, alguns autores obtiveram sucesso no desenvolvimento de processos de purificação de proteínas recombinantes sem fusão com elementos ligantes. Sendo assim, nos tópicos a seguir pretendemos apresentar alguns conceitos teóricos relacionados às principais etapas de purificação de proteínas intracelulares.

### 2.4.1. Metodologias de ruptura celular

Muitas vezes a proteína de interesse se encontra no interior da célula, como é o caso das proteínas produzidas por *E. coli*. Quando isso acontece, a primeira etapa do processo de purificação deve envolver a ruptura celular para disponibilizar o biocomposto para as próximas etapas. Os métodos utilizados se dividem em mecânicos, a exemplo do homogeneizador contínuo de alta pressão (HCAP), prensa de French, moinho de bolas e ultrassom; não mecânicos, como o choque osmótico e de temperatura; e químicos e enzimáticos, quando um reagente químico ou enzima catalisa uma reação que gera fragmentação da parede celular (GECIOVA et al, 2002).

A ruptura celular realizada por HCAP é um método acessível com alto potencial para escalonamento industrial e ampla variedade de dados na literatura que comprovam a sua eficácia, principalmente quando aplicado em microrganismos como *E. coli*, *B. subtilis* e *S. cerevisiae* (GECIOVA et al, 2002). Em 2013, SAFI et al. testaram diversos métodos para recuperação de proteínas produzidas por microalgas, dentre eles o HCAP, tratamento químico e ultrassom. Foi observada maior recuperação e estabilidade proteica no primeiro caso. O mesmo foi observado por GRIMI et al. (2014), que apesar de listar os benefícios da técnica pontuou o grande gasto de energia gerado pelo cisalhamento e manutenção da alta pressão.

A Figura 2.4 descreve o processo de operação de um HCAP, onde a suspensão celular é comprimida pelo pistão de uma bomba e submetida a alta pressão para passar por um orifício estreito, colidindo em seguida contra um anel de impacto. A tensão de cisalhamento gerada é responsável por romper a parede celular, o que resulta em um produto homogeneizado que será recirculado para nova passagem no sistema. Um trocador de calor também é conectado ao sistema de recirculação para que não haja aquecimento da suspensão em processamento (HARRISON, 1991). Após a completa ruptura celular é obtida uma mistura que denominamos homogenato, contendo todas as proteínas solúveis e resíduos celulares. Para que os fragmentos sejam removidos, o homogenato segue para etapas de precipitação e clarificação, de forma que as proteínas solúveis se tornem adequadamente disponíveis para as etapas cromatográficas.

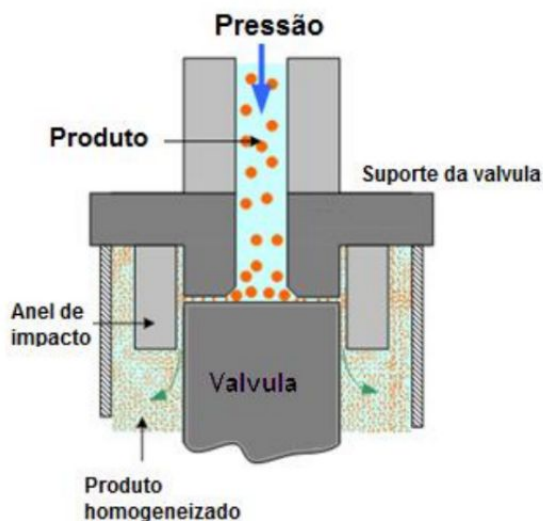


Figura 2.4. Representação de um homogeneizador de alta pressão (Adaptado de SUBSTECH, 2013).

#### 2.4.2. Operações de precipitação

A finalidade da precipitação é transformar moléculas solúveis em insolúveis através da modificação de condições da solução em que se encontram. Com isso é possível removê-las da mistura por processos físicos de separação como filtração, centrifugação ou decantação. O precipitado pode ser a molécula de interesse ou não. Caso não seja, um sobrenadante mais límpido é formado no qual produto de interesse estará diluído. Por se tratar de um processo simples, a precipitação é muitas vezes utilizada como etapa anterior a processos de separação mais complexos (HILBRIG, 2003).

Muitas vezes a clarificação pode ser um desafio, principalmente se a solução possui uma grande quantidade de partículas dispersas. Nesse caso, agentes flocculantes são utilizados para gerar agregação e aumentar o tamanho dos sólidos insolúveis, o que facilita e reduz o custo da separação (BUYEL et al., 2014). Um exemplo dos benefícios da agregação de partículas foi observado por BRACEWELL et al. (2009), que estudou os efeitos da filtração e precipitação na recuperação de uma proteína intracelular produzida por *Saccharomyces cerevisiae*. O autor notou que ambos os processos eram eficientes para clarificar e remover impurezas grosseiras do homogenato. No entanto, quando submetidos à cromatografia hidrofóbica, amostras clarificadas por precipitação sobrecarregaram menos a coluna cromatográfica, pois houve agregação e precipitação de lipídios. Já a filtração não apresentou

a mesma eficiência para a remoção dessas moléculas, reduzindo a eficiência da cromatografia.

Muitos fatores podem afetar as características dos floculantes, como por exemplo o pH, a condutividade do meio, a carga e tamanho das partículas e a densidade de carga (BUYEL et al., 2014). Portanto, métodos muito empregados para a redução de solubilidade de proteínas envolvem precipitação por adição de sais iônicos, mudança de pH e temperatura, ou precipitação a partir de adição de solventes orgânicos (etanol e acetona), íons metálicos ( $\text{Cu}^{2+}$ ,  $\text{Zn}^{2+}$ ) ou polímeros não iônicos (HARRISON, 1991). Outros agentes floculantes também muito comuns para purificação de proteínas são os detergentes (MICOLI et al., 2012). Por exemplo, o brometo de cetiltrimetilamonio (CTAB) é um detergente catiônico que possui afinidade por moléculas com alto peso molecular e cargas negativas, principalmente lipopolissacarídeos (LPS) e ácidos nucleicos, e pode ser utilizado como agente precipitante em etapas anteriores às cromatografias (LANDER et al., 2002).

### **2.4.3. Processo de adsorção**

#### **A) Princípios de adsorção**

O processo de adsorção é aplicado de forma ampla e variada na indústria farmacêutica. Sua relevância pode ser notada na pesquisa e desenvolvimento de materiais biocompatíveis, no melhoramento de dispositivos analíticos, na indústria de alimentos e em sistemas de tratamento de efluentes industriais (RABE et al., 2011). Trata-se de um processo em que uma molécula adere a uma superfície sólida. O soluto adsorvido, que pode estar na fase gasosa ou líquida, é chamado de adsorbato, e a matriz sólida que retém a substância é chamada de adsorvente. Geralmente moléculas submetidas a processos de adsorção estão presentes em misturas multi-componentes, e um ou mais de seus compostos podem possuir afinidade com a matriz sólida. A migração desses componentes de uma fase para outra ocorre devido à existência da força motriz que é gerada pela diferença de concentrações entre as duas fases do sistema (YUSUFF et al., 2013).

Processos de adsorção ocorrem a partir de interações química ou física. Na adsorção química há a formação de ligações covalentes entre os adsorventes e adsorbato. O compartilhamento de elétrons ocorre com a formação de uma única camada que envolve a superfície sólida, com liberação de considerável quantidade de energia. Já a adsorção física

ocorre através da existência de forças de Van der Waals, quando não há compartilhamento de elétrons, mas somente interações eletrostáticas geradas pelas diferenças de carga. Trata-se de um fenômeno reversível que envolve menor liberação de energia e formação de mais de uma camada de adsorbato sobre a superfície da matriz sólida (MEZZARI, 2002). Outros fatores relacionados às características do sistema também podem influenciar a adsorção, a exemplo de porosidade, área superficial e aspectos granulométricos do adsorvente, tempo de contato entre os fluidos envolvidos, assim como a temperatura e o pH do sistema (BORBA et al., 2006). A temperatura afeta diretamente o estado de equilíbrio e a cinética de interação, uma vez que adsorção é um processo exotérmico (NORDE et al., 1996). O pH determina o estado eletrostático dos adsorbatos. Quando se aproxima do ponto isoelétrico de uma proteína, por exemplo, há mesma quantidade de cargas negativas e positivas, o que resulta em uma molécula neutra. Em condições de baixo pH as proteínas se tornam positivamente carregadas, e o oposto ocorre em condições de alto pH. Desta forma, o controle do pH em um processo de adsorção pode fazer com que o experimentador manipule a carga do adsorbato para que se torne oposta à do adsorvente, favorecendo a interação física (BREMER et al., 2004; DEMANECHÉ et al., 2009).

### **B) Adsorção em leito fixo**

Industrialmente, é muito importante que o processo de adsorção seja desenvolvido de forma eficiente e contínua para que os fenômenos de interação moleculares sejam otimizados (RABE et al., 2011). Por isso, a configuração industrial mais comum é a coluna de leito fixo, com a matriz sólida imóvel interagindo com uma fase móvel rica em adsorbato. Com essa configuração, é possível operar com colunas em paralelo para que o processo seja cíclico e contínuo, já que as etapas de adsorção e dessorção são realizadas de forma intercalada devido à necessidade de regeneração do adsorvente, o que reduz os custos operacionais e aumenta a eficiência da operação (SIAHPOOSH, 2009). Em uma coluna de leito fixo a concentração de adsorbato varia com o tempo e com a posição no leito. Uma ferramenta apropriada para analisar o comportamento deste sistema é a curva de ruptura (McCABE, 1993) apresentada pela Figura 2.5. No instante inicial, verifica-se alta adsorção do soluto pela resina, já que possui todos os sítios ativos livres e concentração de adsorbato ( $C_A$ ) nula. Por isso, a zona de transferência de massa (ZTM) é identificada no topo da coluna. Com o passar do tempo a ZTM se desloca para camadas inferiores. Na figura “b” verifica-se uma

situação em que quase metade da coluna se encontra saturada, porém a concentração de adsorbato na saída ( $C_B$ ) ainda é nula. O primeiro momento em que haverá indícios de soluto na saída é quando a ZTM atinge o final da coluna (como mostrado em “c”, quando  $C_C$  é aproximadamente 5% da concentração de entrada). Posteriormente são atingidas condições de saturação, e o valor de  $C_D$  se torna muito próximo ao da concentração inserida na coluna (McCABE, 1993).

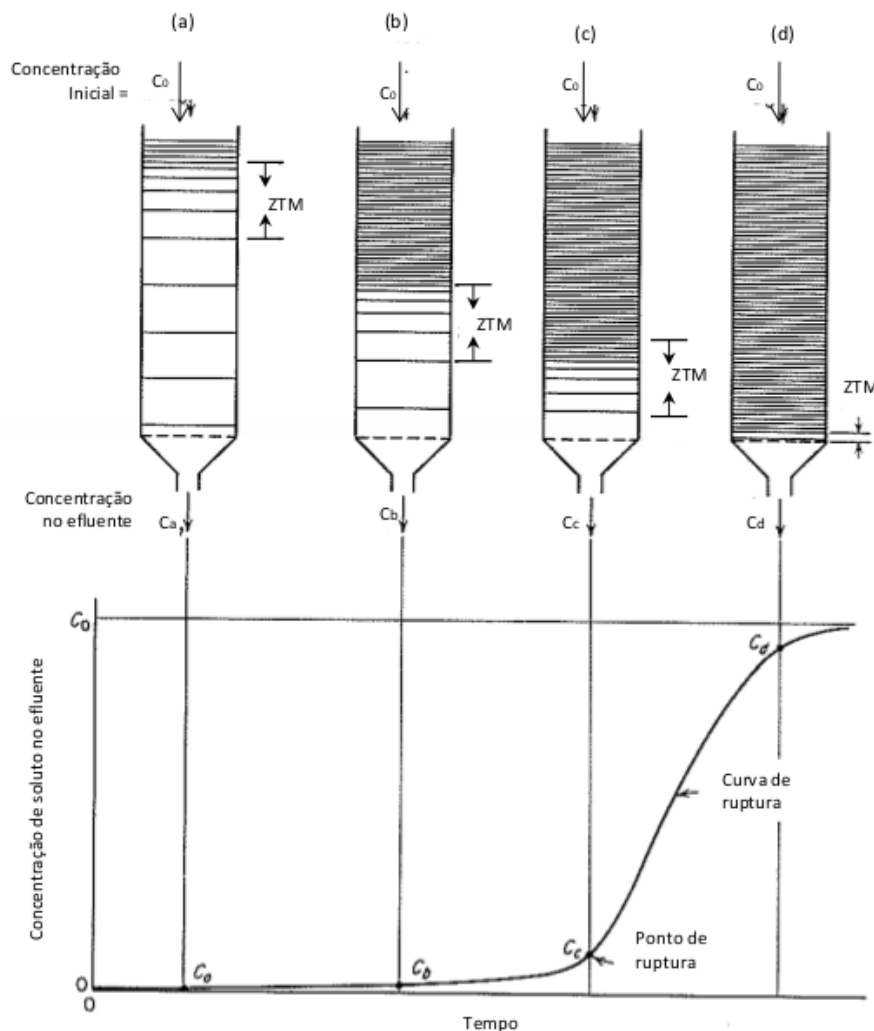


Figura 2.5. Representação da curva de ruptura em sistema de adsorção com leito fixo. Adaptado de TREYBAL (1980). ZTM: Zona de transferência de massa.

#### 2.4.4. Processos cromatográficos para purificação de proteínas

A cromatografia é um método físico-químico de separação no qual os princípios de adsorção em leito fixo são aplicados. É utilizada para separar moléculas que apresentam diferentes afinidades com a resina. Durante a passagem da fase móvel pela estacionária os componentes ficam retidos de forma seletiva, possibilitando migrações diferenciais para cada um (COLLINS, 2006). Considerando o estado físico da fase móvel, é possível dividir a cromatografia em supercrítica, gasosa ou líquida. No primeiro caso, a fase móvel é representada por um fluido supercrítico. No segundo caso é um gás a baixa pressão. No último caso, que será o objeto explorado neste trabalho, é um líquido que interage com os solutos. A cromatografia líquida pode ser executada de duas formas: A clássica, feita em colunas de vidro sob pressão atmosférica, quando há escoamento por gravidade; e a de pressões controladas, que utiliza colunas conectadas a bombas. Este último caso é a configuração mais comum (COLLINS, 2006). Em relação à fase estacionária a matriz ideal na cromatografia líquida de proteínas deve ser insolúvel e conter grupos que se liguem às proteínas de interesse. Deve também possuir estabilidade química, física, e rigidez necessária para permitir altos fluxos de fase móvel em colunas empacotadas por partículas pequenas. Dentre os tipos de matrizes mais comuns utilizadas para a separação de proteínas estão as inorgânicas, poliméricas e polissacarídeas (JANSON, 2012). No caso particular da cromatografia de troca iônica, as proteínas são separadas com base nas cargas iônicas que possuem e são seletivamente retidas pela matriz sólida, carregada com carga oposta. A distribuição das cargas nas superfícies das proteínas é determinada pelo seu ponto isoelétrico (pI) e pelo pH do sistema. Em situações com pH acima do pI as proteínas se tornam carregadas negativamente. O oposto ocorre em situações cujo pH do sistema se encontra abaixo do pI. Quanto mais distante o pH estiver do pI, mais forte é a interação da proteína com a resina (JANSON, 2012). Os solutos adsorvidos são posteriormente eluídos através de deslocamento gerado por outros íons com o mesmo tipo de carga (COLLINS, 2006). Isso acontece quando se aumenta a força iônica do meio utilizado na fase móvel. Proteínas com menor densidade de carga deixam primeiro o sistema (SCOPES et al., 2013). A Figura 2.6 explica de forma didática todas as etapas de operação de uma coluna de troca catiônica, quando a resina negativamente carregada retém dois tipos diferentes de íons positivos.

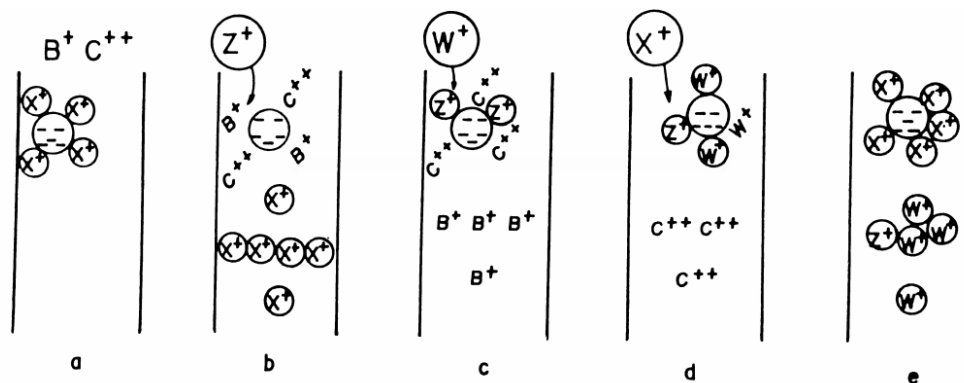


Figura 2.6. Esquema do mecanismo de troca iônica (Fonte: COLLINS, 2006).

Os cátions  $B^+$  e  $C^{++}$  são materiais a serem separados em uma amostra hipotética. O segundo cátion, por possuir maior carga, apresenta maior afinidade com a resina. Os cátions  $Z^+$ ,  $W^+$  e  $X^+$  são íons presentes nos eluentes 1, 2 e 3, respectivamente. A Figura 2.6-a mostra a etapa da adsorção, quando os átomos carregados da amostra entram em contato com a resina e substituem os cátions  $X^+$ , responsáveis pelo equilíbrio da fase sólida. Na etapa 5-b o eluente 1 é alimentado ao sistema. Por possuir baixa afinidade com a resina, acaba competindo somente com os cátions  $B^+$ , que são eluídos na primeira etapa. Na etapa 5-c o eluente 2 é alimentado ao sistema. Por possuir cátions  $W^+$ , com maior afinidade pela resina, leva à eluição dos cátions  $C^{++}$ . A etapa 5-d mostra a aplicação de uma última eluição com alta concentração de íons  $X^+$ , e a resina é regenerada à sua condição inicial, representada em 5-e (COLLINS, 2006). O cromatograma apresentado na Figura 2.7 é obtido por um sensor localizado no final da coluna cromatográfica.

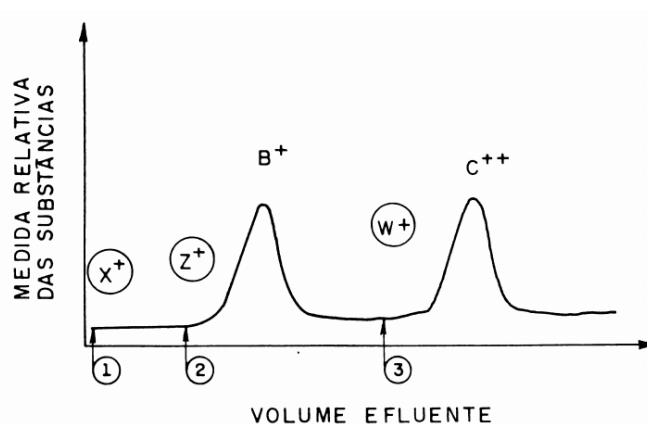


Figura 2.7. Perfil de eluição hipotético gerado pelo processo cromatográfico descrito na figura anterior (Fonte: COLLINS, 2006).

Em um sistema real onde proteínas serão separadas, os eluentes  $Z^+$  e  $W^+$  podem ser representados por um sal diluído na fase móvel em diferentes concentrações. Naturalmente, eluentes com maiores concentrações de sal apresentam maior força iônica e deslocam o equilíbrio no sentido da adsorção do sal, competindo com adsorbatos mais fortemente ligados e causando a dessorção desses. A maioria das proteínas são eluídas em concentrações de sal menores que 1 M. O tipo de sal utilizado pode ser um fator determinante no perfil de eluição. Em cromatografias de troca aniônica, por exemplo, ânions polivalentes são mais eficientes para deslocar moléculas pequenas (SCOPES et al., 2013).

## 2.5. Modelagem matemática de processos cromatográficos

Assim como a maioria dos processos químicos e operações unitárias, os processos cromatográficos também podem ser representados por modelos que possuem finalidade de prever o comportamento de um cromatograma sem a realização de experimentos. Essa representação matemática pode ser muito útil ao experimentador, pois possibilita estudo mais detalhado do processo em menos tempo e com menor esforço experimental, além de viabilizar otimizações e escalonamentos. Para que os modelos cromatográficos sejam melhor explorados devemos compreender de que forma ocorrem os fenômenos de transferência de massa, já que para que a interação exista é necessário que o adsorbato esteja relativamente próximo ao adsorvente. A Figura 2.8 descreve de forma detalhada o caminho que uma partícula percorre para que interaja com a matriz sólida. A primeira barreira é representada pela difusividade na fase líquida até as proximidades da partícula de adsorvente (A), seguido pela resistência à transferência de massa imposta pela película de líquido estagnado em volta da da parede da matriz sólida (B) e pela difusividade no interior do sólido (C), para que finalmente ocorra a interação entre o adsorvente e o adsorbato (D) (FINETTE, 1997; GUIOCHON, 2006). Nos últimos anos, os avanços computacionais possibilitaram o desenvolvimento de diferentes modelos matemáticos que descrevem todos estes aspectos de transferência de massa. A seguir, apresentaremos os mais comuns em ordem crescente de complexidade.

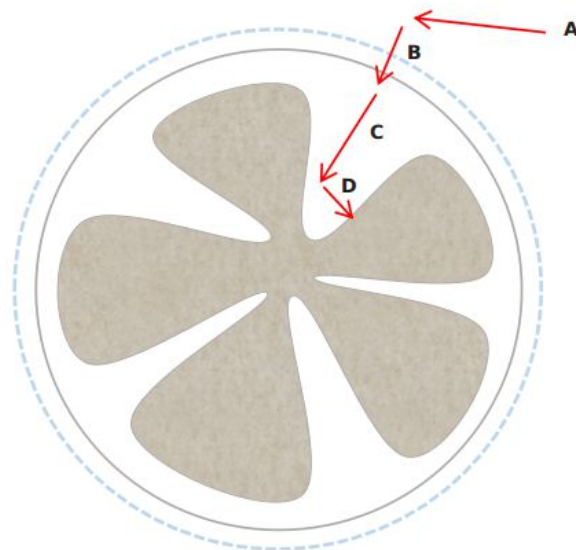


Figura 2.8. Representação da transferência de massa de uma proteína em uma partícula esférica de resina. Fonte: Autoria própria.

#### A) O Modelo ideal (MI)

Trata-se do modelo mais simples utilizado para descrever a interação entre adsorvente e adsorbato, pois desconsidera qualquer tipo de resistência à transferência de massa, ou seja, o adsorbato acessa qualquer região da matriz sólida sem nenhum tipo de impedimento. Os efeitos de dispersão axial na fase líquida são desconsiderados, o que significa que a difusividade é muito baixa ou desprezível (KACZMARSKI et al., 2001). O MI pode ser descrito pela equação 2.1 que representa o balanço de massa em um prato teórico da coluna cromatográfica, onde  $c$  é a concentração de adsorbato na fase móvel (g/mL),  $q$  é a quantidade de adsorbato retido na estrutura sólida em equilíbrio com a fase líquida (g/mL<sub>resina</sub>),  $u$  é a velocidade da fase móvel (cm/s) e  $\epsilon_T$  é a porosidade total do sistema (GRITTI et al., 2013; GUIOCHON, 2006).

$$\frac{\partial c}{\partial t} + \frac{(1 - \epsilon_T)}{\epsilon_T} \frac{\partial q}{\partial t} + u \frac{\partial c}{\partial z} = 0 \quad (2.1)$$

### B) Modelo do equilíbrio dispersivo (MED)

O MED é uma versão um pouco mais complexa do MI e deve ser aplicado em situações em que a dispersão axial tem influência relevante na transferência de massa. Portanto, o coeficiente de dispersão axial ( $D_a$ ) deve ser incluído na equação do modelo (2.2) do prato teórico (KACZMARSKI et al., 2001; GUIOCHON, 2006). Assim como o MI, o MED também desconsidera qualquer outro tipo de resistência à transferência de massa, o que facilita o acesso do adsorbato na região interna da matriz sólida. Se observarmos a figura 2.8, podemos dizer que o único efeito considerado pelo MED é a difusão mássica no líquido (A), aspecto diretamente relacionado à dispersão axial. O MED também propõem que  $D_a$  é constante em todo o comprimento da coluna, desde que a velocidade de escoamento da fase móvel se mantenha (DONAT et al., 2018; GUIOCHON, 2006). Portanto, este parâmetro pode ser calculado pela equação 2.3, onde  $L$  é o comprimento da coluna e  $N_t$  é o número total de pratos teóricos, que também pode ser estimado pela equação 2.4, desde que o perfil cromatográfico de uma molécula sem afinidade pela resina seja conhecido, onde  $t_o$  é o tempo de retenção dessa molécula no interior da coluna e  $\Delta$  é a largura do pico gerado (CLOSE et al., 2014).

$$\frac{\partial c}{\partial t} + \frac{(1 - \epsilon_T)}{\epsilon_T} \frac{\partial q}{\partial t} + u \frac{\partial c}{\partial z} = D_a \frac{\partial^2 c}{\partial z^2} \quad (2.2)$$

$$D_a = \frac{Lu}{2N_t} \quad (2.3)$$

$$N_t = 5.54 \left( \frac{t_o}{\Delta} \right)^2 \quad (2.4)$$

A figura 2.9 representa de forma didática quais aspectos fenomenológicos são levados em consideração pelo MED. É possível verificar que a única resistência imposta pelo sistema para que o adsorbato alcance a matriz sólida é gerada pela dispersão axial. Uma vez próximo à parede, o adsorbato atinge equilíbrio com a fase sólida, sem encontrar resistências da parede externa e acessando livremente o volume interno da matriz sólida.

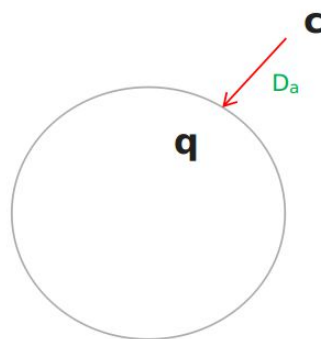


Figura 2.9. Representação dos aspectos de transferência de massa observados no Modelo do equilíbrio dispersivo. Fonte: Autoria própria.

Apesar de simples, o modelo do equilíbrio dispersivo tem sido utilizado com sucesso para modelar colunas cromatográficas. CLOSE et al. (2014) utilizaram o MED para prever o perfil cromatográfico de uma proteína separada por cromatografia hidrofóbica. As previsões apresentaram menos de 3% de desvio dos dados experimentais. TEOH et al. (2001) também utilizaram esse modelo para descrever o perfil cromatográfico da separação de uma biomolécula em um HPLC e conseguiram ajustar bem o modelo proposto aos cromatogramas experimentais.

### C) Modelo do transporte dispersivo (MTD)

O MTD é uma versão mais complexa do MED e pode ser representado pelas equações 2.5 e 2.6. No MTD, além dos efeitos da dispersão axial também são consideradas as resistências à transferência de massa impostas pela película de líquido estagnado em volta da parede da resina. Essas resistências são representadas pela difusão na parede (B) da figura 2.8. Neste caso, é introduzida a variável  $c_s$  que representa a concentração do soluto na fase estacionária que não está em equilíbrio com a fase líquida (g/mL), já que existe uma barreira de transferência de massa entre  $c_s$  e  $q$  imposta por  $k_m$  ( $s^{-1}$ ), o coeficiente aparente de transferência de massa (CANDY, 2012; FARAJI, 2015; GUIOCHON, 2006).

$$\frac{\partial c}{\partial t} + \frac{(1-\epsilon_T)}{\epsilon_T} \frac{\partial c_s}{\partial t} + u \frac{\partial c}{\partial z} = D_a \frac{\partial^2 c}{\partial z^2} \quad (2.5)$$

$$\frac{\partial c_s}{\partial t} = k_m (q - c_s) \quad (2.6)$$

A figura 2.10 representa os aspectos de transferência de massa considerados para desenvolver o MTD. Além da dispersão axial, observa-se também a incorporação da resistência à transferência de massa do adsorbato para o interior da matriz sólida, representada pelo coeficiente  $k_m$ . No entanto, uma vez que atravessa a parede o adsorbato a molécula adsorvida acessa livremente o volume interno da partícula sem outros tipos de resistência, já em equilíbrio com a fase sólida.

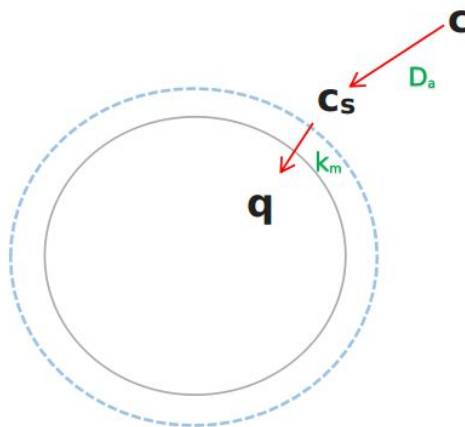


Figura 2.10. Representação dos aspectos de transferência de massa observados no modelo do transporte dispersivo. Fonte: Autoria própria.

CANDY et al. (2012) utilizaram o MTD para estudar o comportamento de proteínas em colunas cromatográficas. Após estimativa de parâmetros, otimização da produção e validação experimental atingiram uma produtividade de  $2,9 \text{ kg.L}^{-1}.\text{dia}^{-1}$ . FARAJI et al. (2015) utilizaram o MTD para estudar a separação de lactoperoxidase e lactoferrina em uma coluna de troca catiônica. Inicialmente, os autores analisaram as capacidades de equilíbrio de ambas as moléculas através de ensaios em batelada. Depois, realizaram experimentos na coluna para permitir a identificação dos parâmetros e validação do modelo. Os resultados gerados pela simulação apresentaram erro inferior a 10% em relação aos valores experimentais. Com isso, foi possível investigar os efeitos da variação do perfil de eluição e concentrações iniciais de proteína na eficiência de separação e qualidade do produto.

#### D) O Modelo geral (MG)

O MG é um dos modelos mais completos utilizados para descrição de processos cromatográficos, pois leva em consideração todos os aspectos do sistema de adsorção, tais como a influência da dispersão axial (figura 2.8.A), a resistência à transferência de massa gerada pelo filme líquido estagnado na parede da matriz sólida (figura 2.8.B) e o transporte difusivo através dos poros (figura 2.8.C). Finalmente, após romper todas essas barreiras o adsorbato se coloca em condições de equilíbrio para ser adsorvido pela matriz sólida (figura 2.8.D) (KACZMARSKI et al., 2001; ORELLANA et al., 2009; LIENQUEO et al., 2012; GUIOCHON, 2006). A configuração do modelo é feita por equações diferenciais que descrevem o movimento do adsorbato pelo leito da coluna e sua transferência de massa no interior da resina (LIENQUEO et al., 2012; ORELLANA et al., 2009). O alto nível de detalhamento fenomenológico apresentado pelo MG já foi comprovado diversas vezes nos últimos anos. KACZMARSKI et al. (2001) recorreram ao MG para ajustar curvas cromatográficas de albumina. Foram testadas grandes variações de concentração proteica e em todos os casos os ajustes e validações foram eficientes. ORELLANA et al. (2009) partiram da mesma estratégia para estudar o comportamento de curvas cromatográficas de  $\alpha$ -lactalbumina, BSA e conalbumina em cromatografias de troca aniônica com resina Q-Sepharose. As curvas de eluição obtidas foram simuladas com sucesso pelo modelo proposto, com erro de aproximadamente 5%.

O MG é inicialmente descrito pela equação 2.7 que representa o balanço de massa para a fase móvel. Considera-se neste caso que há uma concentração média ( $q_m$ ) de soluto adsorvido (g/mL), a qual pode ser representada pela equação 2.8 no caso de uma partícula esférica, onde  $R_p$  é o raio da partícula (cm) e  $M_f$  é o fluxo molar de soluto observado entre a solução e a superfície externa da partícula (mol/s.L). Este fluxo pode ser descrito pela equação 2.9, que mais uma vez utiliza o coeficiente aparente de transferência de massa ( $k_m$ ) para incorporar a resistência à transferência de massa existente entre as fases líquidas (c) e sólidas na parede da partícula ( $c_p$  quando  $r = R_p$ ) (GUIOCHON, 2006).

$$\frac{\partial c}{\partial t} + \frac{(1 - \epsilon_T)}{\epsilon_T} \frac{\partial q_m}{\partial t} + u \frac{\partial c}{\partial z} = D_a \frac{\partial^2 c}{\partial z^2} \quad (2.7)$$

$$\frac{\partial q_m}{\partial t} = \frac{3}{R_p M_f} \quad (2.8)$$

$$M_f = k_m [c - c_p(r = R_p)] \quad (2.9)$$

Naturalmente, como se trata de um processo em que a difusão no interior do sólido é relevante, espera-se que o modelo leve em consideração um coeficiente de difusividade aparente  $D_p$  (cm<sup>2</sup>/s), assim como o gradiente de concentração de adsorbato no interior da partícula em função do raio,  $c_p(r)$ . Para isso, a equação 2.10 que descreve o balanço de massa no interior da partícula deve ser integrada às equações anteriores para que todos os efeitos do MG sejam completamente descritos. Vale observar que neste caso a equação leva em consideração a porosidade da partícula ( $\epsilon_p$ ), não a porosidade total (GUIOCHON, 2006).

$$\epsilon_p \frac{\partial c_p}{\partial t} + (1 - \epsilon_p) \frac{\partial q}{\partial t} = D_p \left( \frac{\partial^2 c_p}{\partial r^2} + \frac{2}{r} \frac{\partial c_p}{\partial r} \right) \quad (2.10)$$

Neste momento já se torna necessário destacar a diferença entre as porosidades do leito ( $\epsilon_B$ ), da partícula ( $\epsilon_p$ ) e total ( $\epsilon_T$ ). A primeira é definida por uma relação entre os volumes vazios exteriores às partículas (interstícios) e o volume da coluna. A segunda é definida por uma relação entre os volumes vazios interiores à partícula e o volume da coluna. Por sua vez, a porosidade total ( $\epsilon_T$ ) é definida por uma relação entre os volumes vazios totais no leito (interparticulares e intraparticulares) e o volume da coluna (GUIOCHON, 2006).

A figura 2.11 representa todas as variáveis e parâmetros envolvidos no desenvolvimento do MG. É possível verificar que além da dispersão axial ( $D_a$ ) e do coeficiente de transferência de massa na parede ( $k_m$ ) o adsorbato também sofre resistência no interior da matriz sólida enquanto percorre caminhos líquidos até os sítios ativos. Esta resistência é representada pela difusividade  $D_p$ , e uma vez que o adsorbato consiga rompê-la finalmente alcança os sítios ativos, atingindo equilíbrio com a fase sólida.

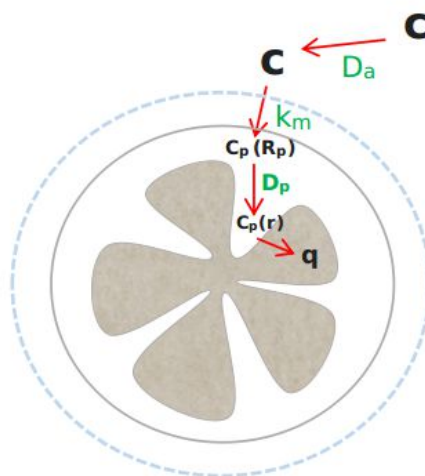


Figura 2.11. Representação dos aspectos de transferência de massa observados no modelo geral. Fonte: Autoria própria.

## 2.6. Isotermas de adsorção

É importante destacar que todos os modelos da seção anterior consideram que em algum momento do processo de transferência de massa o adsorvente entra em equilíbrio com a matriz sólida. Esta condição foi representada por “q”, uma variável que pode ser correlacionada com a concentração da fase líquida externa ( $c_{ext}$ ) na qual se encontra em equilíbrio através de uma isoterma de adsorção, conforme mostra de forma genérica a equação 2.11.

$$q = f(c_{ext}) \quad (2.11)$$

As isotermas de adsorção possuem função de complementar os modelos apresentados na seção anterior, fornecendo ao sistema condições de definir a quantidade de adsorbato que se encontra em fase líquida ou efetivamente adsorvido na resina. Nesta seção pretendemos apresentar conceitos teóricos das isotermas mais comuns.

### A) A isoterma linear

A forma mais simples de representar a função da equação 2.11 é a partir da equação 2.12, que transmite uma relação linear entre a quantidade de componentes adsorvidos e não adsorvidos a partir do uso da constante de adsorção  $K_{Ads}$ . Um dos principais problemas desta abordagem está relacionada à não limitação de uma quantidade máxima de sítios ativos

disponíveis para adsorção, já que esta relação mostra que quanto maior é a concentração de adsorbato ( $c_{ext}$ ) no sistema maior a quantidade de espécies químicas adsorvidas ( $q$ ), conforme representado pela figura 2.12. Portanto, o modelo linear deixa de ser válido em situações em que há grande quantidade de adsorbato no meio, já que a matriz sólida possui uma quantidade finita de sítios ativos para dessorção (GUIOCHON, 2006; LE HA et al., 1982; LIENQUEO et al., 2012; ORELLANA et al. 2009).

$$q = K_{Ads} c_{ext} \quad (2.12)$$

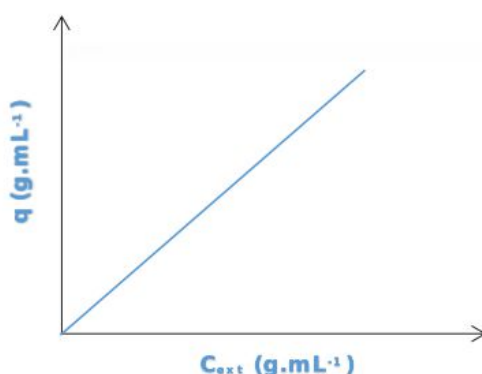
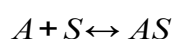


Figura 2.12. Representação gráfica da isoterma de adsorção linear. Fonte: Autoria própria.

### B) A isoterma de Langmuir

A isoterma de Langmuir é uma das mais utilizadas para descrever o fenômeno de adsorção sólido-líquido, possuindo capacidade de se adequar a grande parte de dados experimentais em uma ampla faixa de concentrações. Assim, este modelo aparece como a primeira escolha para ajustes de resultados experimentais, pois se baseia na hipótese que as espécies são adsorvidas em sítios particulares que acomodam apenas uma molécula. A velocidade de dessorção depende somente do material adsorvido na superfície e do perfil de colisão destas moléculas, já que todos os sítios ativos apresentam a mesma entalpia de adsorção (GUIOCHON, 2006). Este modelo parte do princípio que existe equilíbrio químico entre um adsorbato A e o sítio ativo S, formando o agregado AS que representa a adsorção, conforme representado pela reação genérica a seguir, cuja constante de adsorção é calculada pela equação 2.13, sendo que  $c_i$  representa a concentração da espécie química “i” adsorvida ou não (MAHMOUD et al., 2012; GUIOCHON, 2006).



$$K_{Ads} = \frac{c_{AS}}{c_A c_S} \quad (2.13)$$

Considera-se então que a matriz sólida apresenta um número finito de sítios ativos disponíveis para adsorção, e define-se  $\theta$  como a fração de sítios efetivamente ocupados por A. É possível portanto relacionar  $\theta$  às concentrações  $C_A$  e  $C_{AS}$ , conforme mostra a equação 2.14. A substituição da equação 2.14 na equação 2.13 gera a equação 2.15 que representa a constante de adsorção em função de  $C_A$  e  $\theta$ , e se isolarmos a variável  $\theta$  obteremos a equação 2.16.

$$\frac{\theta}{1-\theta} = \frac{c_{AS}}{c_S} \quad (2.14)$$

$$K_{Ads} = \frac{\theta}{c_A(1-\theta)} \quad (2.15)$$

$$\theta = \frac{K_{Ads} c_A}{1 + K_{Ads} c_A} \quad (2.16)$$

Finalmente, é possível ainda definir a variável  $\theta$  como uma relação entre a quantidade “ $q$ ” de sítios ativos ocupados e a quantidade máxima “ $q_{max}$ ” de sítios ativos possíveis de serem ocupados pelo adsorbato, conforme representa a equação 2.17, que substituída na equação 2.16 gera a isoterma de Langmuir (2.18) em sua forma mais comum, alguns casos também expressa no formato da equação 2.19, principalmente em situações em que os parâmetros  $q_{max}$  e  $K_{Ads}$  são ajustados a dados experimentais (MAHMOUD et al., 2012; GUIOCHON, 2006).

$$\theta = \frac{q}{q_{max}} \quad (2.17)$$

$$q = \frac{q_{max} K_{Ads} c_A}{1 + K_{Ads} c_A} \quad (2.18)$$

$$q = \frac{q_{max} c_A}{1 + K_{Ads} c_A} \quad (2.19)$$

A figura 2.13 mostra a representação gráfica da isoterma de Langmuir. Neste caso, diferente da isoterma linear, é importante observar que há uma limitação adsorptiva imposta por  $q_{\max}$ , interrompendo a linearidade do modelo e representando o número finito de sítios ligantes disponíveis nas matrizes sólidas reais (GUIOCHON, 2006).

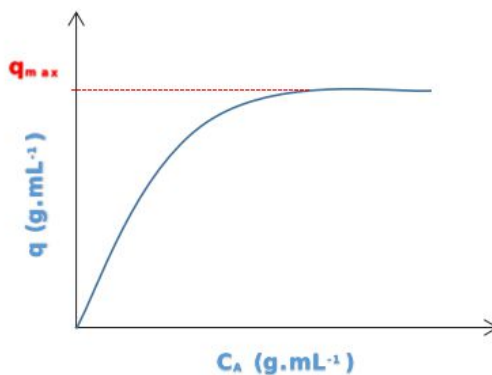
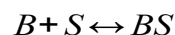
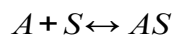


Figura 2.13. Representação gráfica da isoterma de adsorção de Langmuir. Fonte: Autoria própria.

### C) A isoterma de Langmuir competitiva

Em muitas situações é comum que o experimentador tenha de lidar com sistemas que apresentam mais de uma molécula competindo pelo mesmo sítio ativo. Trata-se de um desafio esperado principalmente quando se trata da adsorção de proteínas recombinantes, já que o metabolismo celular produz diversas impurezas proteicas além da proteína de interesse. A forma mais prática de contornar este desafio é adaptando a isoterma de Langmuir a situações em que mais de uma molécula compete pelos sítios ativos da matriz sólida. Por exemplo, podemos considerar que A e B são adsorbatos sem afinidade química entre si em um sistema de adsorção que apresente sítios ativos S (GUIOCHON, 2006; LIENQUEO et al., 2012;). Teremos portanto a seguinte situação:



$$K_A = \frac{c_{AS}}{c_A c_S} \tag{2.20}$$

$$K_B = \frac{c_{BS}}{c_B c_S} \quad (2.21)$$

Vale lembrar que  $K_i$  e  $c_i$  nas equações 2.20 e 2.21 representam as constantes de adsorção e as concentrações das espécies adsorvidas ou não, respectivamente. É importante destacar que neste caso teremos uma fração de sítios ocupados por A, chamada  $\theta_A$ , e outra ocupada por B, chamada  $\theta_B$ . Ainda assim, teremos uma fração de sítios ativos que permanecerá vazia no equilíbrio, chamada  $\theta_V$ . Podemos portanto considerar que há uma relação entre todas as frações citadas, conforme mostra a equação 2.22. Podemos também correlacionar estas variáveis às condições de equilíbrio de A, conforme mostra a equação 2.23, que ao ser substituída na equação 2.20 leva à equação 2.24 (GUIOCHON, 2006).

$$\theta_V + \theta_A + \theta_B = 1 \quad (2.22)$$

$$\frac{\theta_A}{1 - \theta_A - \theta_B} = \frac{c_{AS}}{c_S} \quad (2.23)$$

$$K_A c_A = \frac{\theta_A}{\theta_V} \quad (2.24)$$

Ainda assim, é pertinente destacar que a fração vazia nada mais é do que uma relação entre a concentração de sítios ativos vazios e a concentração de sítios ativos totais, vazios e ocupados, conforme mostra a equação 2.25, que substituída pelas equações 2.20 e 2.21 leva à equação 2.26.

$$\theta_V = \frac{c_S}{c_S + c_{AS} + c_{BS}} = \frac{1}{1 + \frac{c_{AS}}{c_S} + \frac{c_{BS}}{c_S}} \quad (2.25)$$

$$\theta_V = \frac{1}{1 + K_A c_A + K_B c_B} \quad (2.26)$$

A combinação entre as equações 2.24 e 2.26 levará à equação 2.27. Assim como foi feito na equação 2.17 podemos definir a variável  $\theta_A$  como uma relação entre a quantidade “ $q_A$ ” de sítios ativos ocupados e a quantidade máxima “ $q_{\max,A}$ ” de sítios ativos possíveis a

serem ocupados por A, conforme mostra a equação 2.28, que substituída na equação 2.27 leva à isoterma de Langmuir competitiva para A (equação 2.29). A isoterma competitiva de B pode ser obtida de forma análoga (equação 2.30).

$$\theta_A = K_A \frac{c_A}{1 + K_A c_A + K_B c_B} \quad (2.27)$$

$$\theta_A = \frac{q_A}{q_{max,A}} \quad (2.28)$$

$$q_A = \frac{q_{max,A} K_A c_A}{1 + K_A c_A + K_B c_B} \quad (2.29)$$

$$q_B = \frac{q_{max,B} K_B c_B}{1 + K_A c_A + K_B c_B} \quad (2.30)$$

Todo este desenvolvimento matemático também pode ser aplicado em um sistema com N componentes onde “i” representa cada espécie química. Neste caso, chegaremos à equação 2.31 que pode ser aplicada individualmente para cada componente.

$$q_i = \frac{q_{max,i} K_i c_i}{1 + K_A c_A + K_B c_B + \dots + K_N c_N} \quad (2.31)$$

A figura 2.14 descreve graficamente as relações entre A e B no sistema de adsorção em questão. Se considerarmos que  $C_A(3) > C_A(2) > C_A(1)$  e que  $C_B(3) > C_B(2) > C_B(1)$  é possível observar tanto para A quanto para B que a espécie competidora impõe uma limitação à capacidade de adsorção, o que pode ser representado pelas curvas de adsorção menos acentuadas nas situações em que a concentração do competidor no meio é maior (GUIOCHON, 2006). A relação abordada na figura 2.14 também pode ser representada tridimensionalmente, conforme verifica-se na figura 2.15

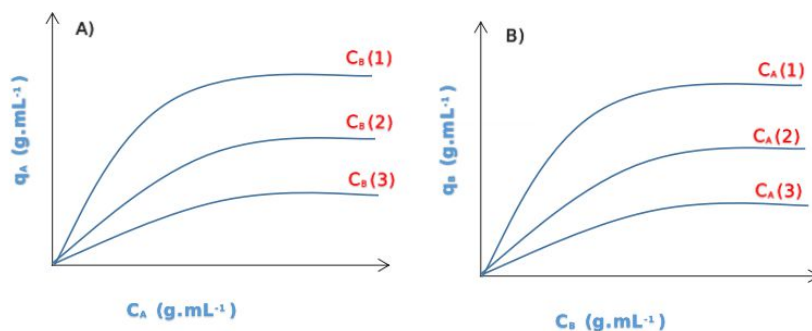


Figura 2.14. Representação gráfica da isoterma de adsorção de Langmuir com competição entre as espécies químicas A e B. Fonte: Autoria própria.

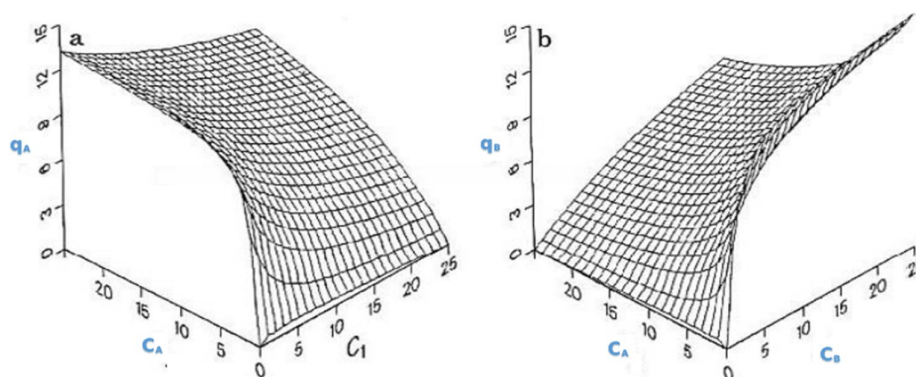


Figura 2.15. Representação tridimensional da isoterma de adsorção de Langmuir com competição entre as espécies químicas A e B. Adaptado de GUIOCHON, 2006.

ORELLANA et al. (2009) utilizaram a isoterma de Langmuir competitiva em complemento ao modelo geral (MG) para modelar curvas cromatográficas obtidas a partir de misturas sintéticas de BSA, lactalbumina e conalbumina em Q-Sepharose. Os autores utilizaram correlações empíricas para estimar as constantes de transferência de massa e testaram uma extensa variação de concentrações. Após ajuste dos parâmetros, foram simuladas condições que modificaram a vazão da fase móvel, o gradiente de sal nas eluições, a concentração das proteínas nas amostras, etc. Em todos os casos as condições foram validadas experimentalmente e apresentaram os comportamentos esperados.

#### D) A isoterma de Langmuir modificada

Quando lidamos com processos cromatográficos frequentemente nos deparamos com situações em que um sal é utilizado para aumentar a força iônica da fase móvel, competindo com as moléculas adsorvidas e gerando a dessorção responsável pela formação das curvas cromatográficas. Neste sentido, é importante dispormos de isotermas que expressem a concentração de sal como um agente competidor à ocupação dos sítios ativos. A isoterma de Langmuir modificada se insere exatamente neste contexto, pois através de equações empíricas propõe modificações nas constantes de Langmuir  $q_{\max}$  e  $K_{\text{Ads}}$ . A concentração de sal ( $c_{\text{sal}}$ ) expressa nas equações 2.32 e 2.33 é variável necessária para correção dos valores teóricos de  $q_{\max}$  e  $K_{\text{Ads}}$  a valores corrigidos e observáveis denominados  $q_{\max}^*$  e  $K_{\text{Ads}}^*$ , onde  $\Psi$  e  $\eta$  são parâmetros empíricos que apresentam unidades de concentração. As variáveis de adsorção corrigidas são então consideradas pela equação 2.34, que nada mais é que uma adaptação da isoterma de Langmuir original (LAN et al., 2000; WAKKEL et al., 2015).

$$q_{\max}^* = \frac{q_{\max}}{1 + \frac{c_{\text{sal}}}{\Psi}} \quad (2.32)$$

$$K_{\text{Ads}}^* = \frac{\eta K_{\text{Ads}}}{\eta - c_{\text{Sal}}} \quad (2.33)$$

$$q = \frac{q_{\max}^* c_i}{(c_i + K_{\text{Ads}}^*)} \quad (2.34)$$

É importante lembrar que esta isoterma leva em consideração que as espécies químicas competem somente com o sal, ou seja, não há competição entre diferentes adsorbatos de um mesmo meio, conforme foi considerado pela isoterma de Langmuir competitiva. Por isso, cabe ao pesquisador selecionar a isoterma que mais se adequa à sua situação. Apesar de recentemente ter sido pouco utilizada para fins de modelagem, a isoterma competitiva expressa bem a atuação do sal como agente competidor, conforme pode ser observado na figura 2.16 que demonstra o efeito de concentrações de sal em um meio adsortivo, tal que  $C_{\text{Sal}}(3) > C_{\text{Sal}}(2) > C_{\text{Sal}}(1)$ . É fácil perceber que sistemas com maior força iônica acarretam em curvas de adsorção mais discretas (LAN et al., 2000).

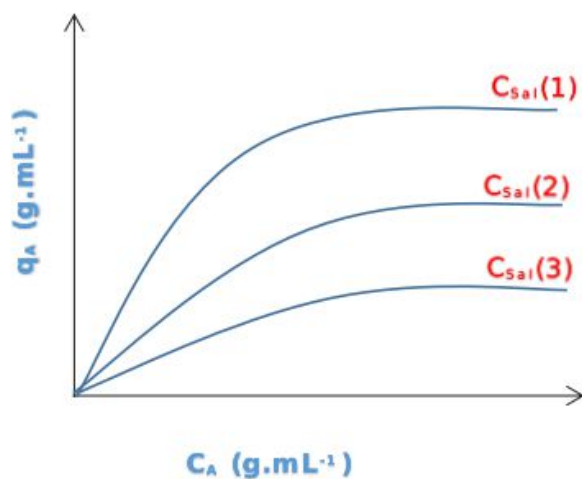


Figura 2.16. Representação gráfica da isoterma de adsorção de Langmuir modificada a partir de diferentes concentrações de sal. Fonte: Autoria própria.

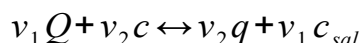
A isoterma de Langmuir modificada já foi utilizada para descrever situações reais em que a adsorção é limitada pela presença de sal na fase líquida. LAN et al. (2000) estudaram os efeitos de diferentes forças iônicas na capacidade de adsorção de BSA em uma resina de troca iônica. A determinação dos parâmetros foi realizada a partir de linearização das equações 2.32, 2.33 e 2.34. O comportamento experimental foi exatamente o representado na figura 2.16, ou seja, curvas de adsorção mais discretas foram observadas em sistemas que apresentavam maior força iônica. WAKKEL et al. (2015) também abordaram o assunto de forma semelhante, mas avançaram um pouco mais em alguns pontos, pois estudaram os efeitos da variação da força iônica na adsorção BSA e ferritina em cromatografia aniônica. Em ambos os casos o comportamento encontrado também refletiu o da figura 2.14.

#### E) A isoterma de ação em efeito estérico

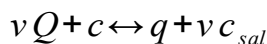
Muitas vezes é necessário que o processo de adsorção seja descrito de maneira mais detalhada para que a simulação se aproxime de condições reais. Em trabalhos recentes, autores tem utilizado com sucesso a isoterma de ação em efeito estérico (IAEE), ou *Steric Mass Action Isotherm* (SMA), para descrever a interação entre proteínas e resinas iônicas. Uma das grandes vantagens é que a abordagem teórica desta isoterma considera competição entre o sal e múltiplas proteínas presentes no meio de adsorção, uma situação bastante comum quando se trabalha com a separação de misturas proteicas reais (GUIOCHON et al., 2006;

GU, 2015; HUUK et al., 2014; KARKOV et al., 2013). O trabalho de HUUK et al (2014) é um bom exemplo da aplicação da IAEE em complemento ao modelo do transporte dispersivo para descrever o comportamento cromatográfico de citocromo, quimotripsina e ribonuclease nas resinas de troca iônica Q-Sepharose e Poros 50HS. Os autores conseguiram com sucesso validar cromatogramas obtidos a partir de modificações da concentração de sal nas eluições. KARKOV et al. (2013) também obtiveram bons resultados quando utilizaram a IAEE em complemento ao modelo do equilíbrio dispersivo para descrever cromatogramas de BSA e albumina humana em colunas de troca aniônica. Assim como no caso anterior, experimentos mostraram sucesso na previsibilidade do modelo mediante modificações na força iônica da fase móvel.

Neste contexto, observamos que a IAEE pode ser uma estratégia interessante a ser utilizada para complementar os modelos que serão desenvolvidos neste trabalho. Para compreendermos a base teórica relacionada ao seu desenvolvimento consideraremos inicialmente uma situação de equilíbrio sólido líquido entre uma proteína (1) e um determinado sal (2). A reação genérica representada abaixo descreve bem esta situação, onde  $v_1$  e  $v_2$  são os coeficientes estequiométricos da proteína e do sal, respectivamente;  $Q$  e  $c_{sal}$  são as concentrações de sal nas fases sólidas e líquida, respectivamente, e  $q$  e  $c$  são as concentrações de proteína nas fases sólidas e líquida. Todas as concentrações são expressas em mol/mL ou mol/mL<sub>resina</sub> (GALLANT, 1995; GUIOCHON et al., 2006; GU, 2015).



Se dividirmos todos os coeficientes estequiométricos por  $v_2$  obteremos a reação química representada a seguir. Denominaremos  $v$  como o número de sítios ligantes presentes na molécula do adsorbato. Sendo assim, a constante de equilíbrio poderá ser representada pela equação 2.35, cujo desenvolvimento leva à equação 2.36.



$$K_{eq} = \frac{k_{Ads}}{k_{Des}} = q \frac{c_{sal}^v}{c Q^v} \quad (2.35)$$

$$k_{Ads} c Q^v = k_{Des} q c_{sal}^v \quad (2.36)$$

No entanto, o desenvolvimento teórico da IAEE considera que as fases sólida e líquida não se encontram em equilíbrio. Portanto, a afirmação imposta pela equação 2.36 é irreal, já que em situações de não equilíbrio a diferença entre os termos adsorptivos e desorptivos não pode ser nula. Por isso, podemos corrigir esta equação à situação de não equilíbrio representada pela equação 2.37.

$$\frac{\partial q}{\partial t} = k_{Ads} c Q^v - k_{Des} q c_{sal}^v \quad (2.37)$$

É importante ainda destacar que o sistema estudado apresenta neutralidade eletrostática, o que significa que o número total de sítios ativos disponíveis na resina ( $\wedge$ ) é preenchido ou pela proteína em questão ou pelos íons provenientes do sal no meio, sejam eles positivos ou negativos. Sendo assim, podemos dizer que a equação 2.38 expressa esta condição de eletroneutralidade. Vale observar que a unidade de  $\wedge$  é mols/mL<sub>resina</sub>, ou seja, expressa o número de mols de sítios ativos disponíveis por volume de resina.

$$\wedge = Q + q v \quad (2.38)$$

No entanto, é importante destacar que muitas proteínas possuem estruturas muito mais complexas do que os íons presentes em solução. Por isso, quando suas regiões eletricamente carregadas se ligam aos sítios ativos é esperado que sua ampla estrutura tridimensional bloqueie e inviabilize a ocupação de outros sítios ativos disponíveis. A figura 2.17 ilustra bem esta situação para uma superfície de troca catiônica. Neste caso, uma proteína eletricamente carregada se aproxima e se conecta a quatro sítios ativos. No entanto, outros três permanecem bloqueados por sua estrutura, inviabilizados de se ligarem a outras proteínas. Esse fenômeno de bloqueio é teoricamente conhecido como fator estérico ( $\alpha$ ) e pode ser representado matematicamente através de uma correção na equação 2.38, conforme mostram as equações 2.39 e 2.40.

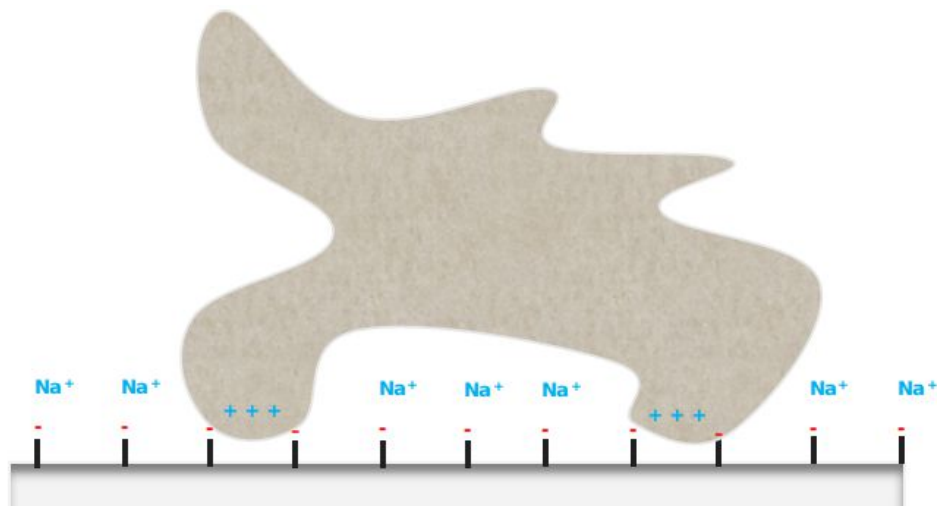


Figura 2.17. Esquema representativo da ligação de proteínas em uma superfície de troca catiônica. Adaptado de: GALLANT, 1995.

$$\Lambda = Q + q(v + \alpha) \quad (2.39)$$

$$Q = \Lambda - q(v + \alpha) \quad (2.40)$$

As equações 2.37 e 2.40 representam todos os fenômenos teóricos considerados pela IAEE. No entanto, elas devem ser adaptadas a um sistema que apresente “i” proteínas, conforme se pode verificar nas equações 2.41 e 2.42, que compõem a IAEE no formato em que será utilizada para o desenvolvimento deste trabalho (GALLANT et al., 1995; CHEN et al., 2017; GUIOCHON et al., 2006; KARKOV et al., 2013; HUUK et al., 2014; GU, 2015; CARVALHO et al., 2016).

$$\frac{\partial q_i}{\partial t} = k_{Ads} c_i Q^{v_i} - k_{Des} q_i c_{sal}^{v_i} \quad (2.41)$$

$$Q = \Lambda - \sum_{i=1}^N q_i (v_i + \alpha_i) \quad (2.42)$$

## 2.7. O software EMSO como simulador orientado a equações

A integração eficiente dos modelos apresentados na seção 2.5 com as isotermas de adsorção da seção 2.6 possibilitam ao pesquisador descrever matematicamente o processo cromatográfico. Sendo assim, uma vez que parâmetros fixos do sistema sejam estimados, é possível executar a simulação dispondo de um software apropriado. O EMSO (*Environment for Modeling, Simulation and Optimization*) é um simulador orientado a equações que possui uma extensa biblioteca com informações termodinâmicas, capacidade de criação de modelos estacionários ou dinâmicos, possibilidade de incluir rotinas externas de cálculo (plug-ins) e rotinas para resolução do sistema de equações algébrico-diferenciais (THOMBRE et al., 2015; IPSITABANERJEE et al., 2010; FURLAN et al., 2015; FURLAN et al., 2016). Trata-se do simulador escolhido para este trabalho, pois não emprega uma ordem específica de rotinas para a solução das equações do processo, diferente do que acontece com os simuladores sequenciais modulares, como é o exemplo do Aspen Plus (SOARES, 2003, THOMBRE et al., 2015; IPSITABANERJEE et al., 2010; FURLAN, 2016).

Para que a estrutura do EMSO seja compreendida, sua linguagem é dividida em três diferentes blocos: os modelos, os dispositivos e os fluxogramas. Um modelo consiste em uma descrição matemática e teórica de um processo físico ou químico que ocorre em um equipamento, seja ele um tanque, um reator, um decantador, uma coluna de destilação, um trocador de calor, etc. Enfim, independente da origem do processo há possibilidades de representação matemática uma vez que as equações fundamentais sejam conhecidas, assim como os parâmetros, variáveis, condições iniciais e condições de contorno envolvidos. Um dispositivo nada mais é do que a representação virtual de uma operação em particular do processo, agregando em sua estrutura os modelos. Um único modelo pode ser utilizado para representar diferentes dispositivos que apresentem estruturas teóricas semelhantes, mas que operem mediante diferentes condições. E por fim, um fluxograma é a conexão de todos os dispositivos descritos, o que leva à descrição de todos os processos envolvidos integrados uns aos outros (SOARES, 2003)<sup>1</sup>.

Muitos trabalhos têm recentemente utilizado o EMSO para descrever processos químicos, ajustar parâmetros e otimizar processo. ALBARELLI et al. (2017) realizaram uma otimização em uma biorrefinaria de açúcar produtora de etanol de primeira e segunda geração

---

<sup>1</sup> Exemplo simples de aplicação do EMSO disponível no Apêndice A.

e eletricidade. Através do modelo e de parâmetros econômicos conseguiram determinar a quantidade ótima de bagaço que deveria ser encaminhada para produção de energia e a melhor concentração final para vinhaça concentrada em um evaporador. DA ROCHA NOVAES et al. (2017) simularam e estimaram parâmetros cinéticos do hidrotreatamento de diesel. O modelo proposto foi capaz de reproduzir diferentes condições de operação com bom ajuste e precisão, apresentando desvios relativos abaixo de 4% para os dados experimentais, além de adequada previsibilidade. FONSECA et al. (2017) realizaram simulações para comparar as eficiências entre fermentações de etanol contínuas e em batelada alimentada. CARPIO et al. (2016) simularam e otimizaram a destilação extrativa de etanol com glicerol, o que possibilitou a determinação de vários pontos de otimização. ALMEIDA et al. (2017) simularam um processo em batelada para a destilação de óleos essenciais de eucalipto, e através de otimizações conseguiram encontrar condições processuais que apresentaram produto com aproximadamente 99% de pureza. Sendo assim, devido a todas as características apresentadas, o EMSO se mostra adequado para modelar e simular o processo dinâmico de adsorção/dessorção de proteínas em colunas de troca iônica, bem como propor condições de otimização para o processo.

### 3. MODELAGEM E SIMULAÇÃO DA CROMATOGRAFIA DE TROCA ANIÔNICA PARA PURIFICAÇÃO DE PROTEÍNAS EM MISTURAS COMPLEXAS

Assim como já foi afirmado, o processo de cromatografia de troca aniônica é o passo mais importante da purificação de PspA4Pro. Além disso, devido ao alto preço da resina Q-Sepharose, o custo desta etapa é bastante elevado. Neste contexto, no presente capítulo pretende-se modelar esse processo e identificar isotermas de adsorção que representam o seu comportamento. Utilizou-se o modelo do equilíbrio dispersivo para descrever o balanço de massa de proteínas em cada prato teórico da coluna. Em paralelo foram avaliadas duas isotermas para descrever a adsorção: a isoterma de Langmuir modificada e a isoterma de ação em efeito estérico. O ajuste de parâmetros e validação foram realizados com dados obtidos pelo processamento de biomassas cultivadas por estratégias distintas, levando à eluição de proteínas com características diferentes e desafiando a capacidade preditiva dos modelos. Apesar dessas dificuldades, ambas as isotermas avaliadas apresentaram bons resultados para a proposta de modelagem.

Em um segundo momento, os parâmetros da isoterma de ação em efeito estérico foram separadamente ajustados aos dados de outras duas cromatografias executadas com amostra sintética de PspA4Pro e amostra proveniente de células não produtoras de PspA4Pro. Desta forma, foi possível compreender separadamente o comportamento cromatográfico de PspA4Pro e impurezas proteicas eluídas no mesmo momento, permitindo a simulação de uma amostra virtual com composição semelhante aos clarificados reais e comparação do perfil simulado com picos experimentais, comprovando a eficiência da estratégia de modelagem escolhida e permitindo a detecção de regiões que apresentassem maior pureza.<sup>2</sup>

---

<sup>2</sup> O texto do presente capítulo aborda os resultados apresentados no artigo intitulado “*Modeling and simulation of anion exchange chromatography for purification of proteins in complex mixtures*”, publicado na revista “*Journal of Chromatography A*” (ISSN: 0021-9673) como requisito para obtenção do título de Doutor.

### ***Resumo***

A cromatografia de troca iônica é uma operação de separação amplamente utilizada na purificação de compostos biológicos. Atualmente, é possível encontrar na literatura modelos matemáticos confiáveis que descrevem essa técnica e que podem ser usados para melhorar o seu desempenho. No entanto, o uso de misturas sintéticas para o desenvolvimento desses modelos dificulta suas aplicações em extratos celulares reais processados por operações *downstream*. Neste contexto, este trabalho visa apresentar uma abordagem original para a modelagem de misturas reais aplicada na purificação de uma proteína recombinante de superfície de pneumococo (PspA4Pro). Em um primeiro momento, foram avaliadas as eficiências das isotermas de ação em efeito estérico (ou “*steric mass action isotherm*”, SMA) e de Langmuir modificada, as quais foram separadamente testadas com o modelo do equilíbrio dispersivo (MED). Os dados utilizados para estimar os parâmetros e validar o modelo foram obtidos a partir de cromatografia de troca aniônica (com Q-Sepharose FF) aplicada a extratos celulares reais produzidos por diferentes estratégias de cultivo. Simulações mostraram que os modelos foram capazes de descrever misturas complexas de proteínas desconhecidas. Em seguida, as equações MED e SMA foram usadas para descrever separadamente o perfil da PspA4Pro e do conjunto de impurezas proteicas eluídas no mesmo momento. As simulações mostraram que a PspA4Pro tendia a eluir no início do pico, permitindo estabelecer um esquema de eluição alternativo que proporcionou um aumento de 34% na pureza alcançada.

### 3.1. Introdução

A cromatografia de troca iônica (CTI) é um método físico-químico utilizado para separar moléculas que apresentam diferentes afinidades com uma resina sólida (GU, 2015). Por muitos anos, a CTI tem sido usada com sucesso para purificar compostos biológicos. No caso particular das proteínas, a separação é realizada com base nas cargas iônicas determinadas pelo ponto isoelétrico (pI) da proteína e pelo pH do sistema, proporcionando diferentes tempos de retenção na matriz sólida, o que aumenta a eficiência da separação (LIN et al., 2001; RABE et al., 2011; JANSON et al., 2012). Os produtos resultantes possuem alta pureza, apresentando contaminantes como ácidos nucleicos, pirogênicos e enzimas em concentrações abaixo dos limites estabelecidos pelas agências reguladoras (MAMAT et al., 2015; FIGUEIREDO et al., 2017).

Os altos custos dos materiais e as demandas por produtos de alta qualidade levaram ao desenvolvimento de técnicas de simulação de processos cromatográficos com o objetivo de otimização e aumento de eficiência. Nos últimos anos, os avanços computacionais permitiram o desenvolvimento de modelos de simulação com diferentes níveis de complexidade, e capazes de descrever aspectos relacionados à cinética de adsorção, transferência de massa, dispersão axial e difusividade do adsorbato na matriz sólida (ORELLANA et al. 2009; LINQUEO et al., 2012). Muitos trabalhos já utilizaram estes modelos para propor modificações no processo e aumentar a eficiência operacional (ORELLANA et al., 2009; LINQUEO et al., 2012; TEOH et al., 2001; KACZMARSKIET et al., 2001; CHAN et al., 2008; ORELLANA et al., 2009; CLOSE et al., 2014; FARAJI et al., 2015; LÓPES-UREÑA et al., 2018). No entanto, na maioria dos casos, misturas sintéticas formuladas com um número limitado de proteínas conhecidas e puras são usadas para extrair dados para caracterizar o processo. Trata-se de uma estratégia eficaz para o planejamento experimental, estimativa de parâmetros e simulação. No entanto, as misturas sintéticas apresentam muitas diferenças em relação aos extratos celulares clarificados processados em procedimentos *downstream*, uma vez que esses extratos contêm muitas proteínas desconhecidas com diferentes tamanhos e propriedades físico-químicas, juntamente com a proteína alvo a ser purificada. Além disso, as características das misturas reais estão diretamente relacionadas ao processo de cultivo, com a composição da mistura de proteínas desconhecidas variando muito de acordo com o material de partida.

Neste trabalho, foi realizada a modelagem da cromatografia de troca aniônica no processamento de misturas proteicas reais. A proteína alvo a ser purificada foi a PspA4Pro, uma

proteína recombinante produzida por *Escherichia coli* BL21 (DE3) apresentando o plasmídeo pET37b + PspA4Pro, que expressa o gene do fragmento N-terminal da proteína de superfície de pneumococo A (PspA) da família 2 do clado 4 (PERCIANI et al., 2013). O uso de PspA4Pro como proteína carreadora na formulação de vacinas conjugadas pode contribuir para aumentar a proteção contra *Streptococcus pneumoniae* (PERCIANI et al., 2013; DA SILVA et al., 2017), uma bactéria causadora de doenças como pneumonia, infecções de ouvido, sinusite, meningite e sepsis (HENRIQUES-NORMAK et al., 2013). Além disso, a PspA pode ser usada como uma vacina de subunidade antigênica, o que tem sido extensivamente estudado por vários grupos de pesquisa em todo o mundo (BRILES et al., 2000; LI et al., 2018; ROGRIGUES et al., 2018; TADA et al., 2018; KYE et al., 2019).

A abordagem apresentada neste trabalho é baseada em uma estratégia que não foi explorada anteriormente para a modelagem de misturas complexas. O procedimento envolve a modelagem de cromatogramas separados obtidos pelo processamento de PspA4Pro pura e de um conjunto de impurezas proteicas derivadas de células não produtoras de PspA4Pro, possibilitando avaliar as características físico-químicas específicas de PspA4Pro e das demais proteínas presentes na mesma eluição. Misturas não sintéticas também foram utilizadas para investigar a capacidade de duas isotermas de adsorção diferentes (Langmuir modificada e de ação em efeito estérico) para descrever o processo, complementando o modelo do equilíbrio dispersivo (LÓPEZ-UREÑA et al., 2018). As simulações foram realizadas no EMSO (*Environment for Modeling, Simulation and Optimization*), um software orientado a equações (SOARES et al., 2003; BANERJEE et al., 2010; THOMBRE et al., 2015; FURLAN et al., 2016) que já foi empregado com sucesso para simular biorrefinarias (ALBARELLI et al., 2017), processos de fermentação (FONSECA et al., 2017), produção de diesel (DA ROCHA NOVAES et al., 2017) e extração de óleos (ALMEIDA et al., 2018). Apesar de sua ampla aplicabilidade, até o presente momento o EMSO ainda não foi utilizado para descrever processos cromatográficos.

É esperado que a abordagem de modelagem proposta neste trabalho aproxime os dados simulados da realidade, melhorando o entendimento de processos reais e auxiliando no desenvolvimento de estratégias futuras para aumentar o desempenho de processos de purificação de extrato celular.

## 3.2. Teoria

### 3.2.1. Modelo do equilíbrio dispersivo (MED)

O modelo do equilíbrio dispersivo (MED) é um modelo matemático usado para simular o perfil de distribuição de proteínas no interior de uma coluna cromatográfica. Esse modelo assume que a cinética da transferência de massa entre as fases sólida e líquida é muito rápida. Desta forma, a concentração de proteínas na fase móvel é considerada similar à concentração média da região intra-particular. Esse modelo também assume que a coluna é radialmente homogênea, que o sistema é isotérmico e adiabático e que os parâmetros de transferência de massa são independentes da concentração de proteínas (CLOSE et al., 2014).

Dadas as considerações acima, o balanço de massa diferencial na fase móvel pode ser descrito pela Equação 3.1, onde  $c_i$  é a concentração de uma proteína "i" na fase móvel ( $\text{mol.mL}^{-1}$  ou  $\text{g.mL}^{-1}$ ),  $q_i$  é a quantidade de adsorbato retido na estrutura sólida ( $\text{mol.mL}_{\text{resina}}^{-1}$  ou  $\text{g.mL}_{\text{resina}}^{-1}$ ),  $u$  é a velocidade da fase móvel ( $\text{cm.s}^{-1}$ ) e  $D_a$  é a dispersão axial ( $\text{cm}^2.\text{s}^{-1}$ ) (TEOH et al., 2001; KACZMARSKI et al., 2001). A dispersão axial ( $D_a$ ) pode ser calculada pela Equação 3.2, onde  $L$  é o comprimento da coluna e  $N_t$  é o número total de pratos teóricos (CLOSE et al., 2014).

$$\frac{\partial c_i}{\partial t} + \frac{(1-\epsilon_T)}{\epsilon_T} \frac{\partial q_i}{\partial t} + u \frac{\partial c_i}{\partial z} = D_a \frac{\partial^2 c_i}{\partial z^2} \quad (3.1)$$

$$D_a = \frac{Lu}{2N_t} \quad (3.2)$$

O parâmetro  $N_t$  pode ser calculado pela Equação 3.3 para um pico cromatográfico obtido por uma molécula pequena que não apresente afinidade pela coluna, onde  $t_0$  é o tempo de retenção na coluna e  $\Delta$  é a largura do pico obtido na metade de sua altura (AMERSHAM-BIOSCIENCES, 2004). O parâmetro  $\epsilon_T$  pode ser definido como a razão entre o volume vazio ( $V_0$ ) e o volume total da coluna ( $V_C$ ), enquanto  $V_0$  pode ser calculado pelo tempo de retenção da molécula ( $t_0$ ) e pela vazão da fase móvel ( $F$ ), como mostram as Equações 3.4 e 3.5 (CLOSE et al., 2014).

$$N_t = 5.54 \left( \frac{t_0}{\Delta} \right)^2 \quad (3.3)$$

$$\epsilon_T = \frac{V_o}{V_c} \quad (3.4)$$

$$V_o = t_o \cdot F \quad (3.5)$$

Para uma descrição matemática completa do sistema cromatográfico, é necessário complementar o MED com uma isoterma de adsorção. Neste trabalho, foram testadas as isotermas de Langmuir modificada (LM) e de ação em efeito estérico (SMA).

### 3.2.2. Isoterma de Langmuir modificada

A isoterma de Langmuir modificada é utilizada para representar a adsorção competitiva de proteínas na presença de um sal (LAN et al., 2001). Sua teoria deriva da isoterma clássica de Langmuir, que considera o equilíbrio químico entre um adsorbato A e o sítio ativo S, formando o agregado químico AS. A constante de adsorção ( $K_{Ads}$ ) pode ser descrita pela Equação 3.6 (MAHMOUD et al., 2013; GHOSAL et al., 2017). A variável  $\theta$ , definida como a fração dos sítios ativos efetivamente ocupados, é representada pela equação 3.7, e a equação 3.8 pode então ser obtida pela combinação das Equações 3.6 e 3.7. A variável  $\theta$  é definida como a razão entre a quantidade "q" de sítios ativos ocupados e a quantidade máxima "q<sub>max</sub>" de possíveis sítios ativos a serem ocupados pelo adsorbato ( $\text{g.mL}^{-1}$ ), levando à isoterma de Langmuir (Equação 3.9) em sua forma mais usual, onde  $K_i$  é  $K_{Ads}^{-1}$ , a constante de equilíbrio de adsorção "i" representa uma proteína ou adsorbato qualquer.

$$K_{Ads} = \frac{C_{AS}}{C_A \cdot C_S} \quad (3.6)$$

$$\frac{\theta}{1-\theta} = \frac{C_{AS}}{C_S} \quad (3.7)$$

$$\theta = \frac{K_{Ads} \cdot C_A}{1 + K_{Ads} \cdot C_A} \quad (3.8)$$

$$q_i = \frac{q_{\max,i} \cdot c_i}{c_i + K_i} \quad (3.9)$$

A teoria clássica da isoterma de Langmuir considera que todos os sítios ativos são equivalentes e independentes, se ligando apenas a uma molécula de soluto, sem influenciar a adsorção das moléculas vizinhas. Entretanto, no caso da adsorção de proteínas, algumas dessas considerações deixam de ser viáveis, uma vez que nesses casos também devem ser contabilizados os efeitos da força iônica, fenômeno que domina a cromatografia de troca iônica e que pode ser caracterizado pela competição dos contra-íons de sal com as proteínas (LAN et al., 2001). Além disso, variações da força iônica exercem efeito no enovelamento das proteínas, modificando suas interações hidrofóbicas com a resina, além de causar um encolhimento da matriz sólida de resina, o que diminui a quantidade de sítios ativos disponíveis (STIGTER et al., 1990; KHARE et al., 1995). Neste contexto, a equação de Langmuir modificada objetiva levar em consideração esses efeitos, assumindo que  $K_i$  e  $q_{\max,i}$  da Equação 3.9 possam ser modificados pela concentração de sal,  $c_{\text{salt}}$  ( $\text{g.mL}^{-1}$ ), representando as alterações que a variação de força iônica gera no sistema. Essa interferência é expressa pelas constantes Langmuir aparentes  $K_i'$  ( $\text{mL.g}^{-1}$ ) e  $q_{\max,i}'$  ( $\text{g.mL}^{-1}$ ), representadas pelas Equações 3.11 e 3.12, respectivamente, que estão associadas às constantes empíricas  $\Psi$  e  $\eta$  ( $\text{g.mL}^{-1}$ ) (LAN et al., 2001).

$$q_i = \frac{q_{\max,i}' c_i}{(c_i + K_i')} \quad (3.10)$$

$$q_{\max,i}' = \frac{q_{\max,i}}{1 + \frac{c_{\text{salt}}}{\Psi}} \quad (3.11)$$

$$K_i' = \frac{\eta K_i}{\eta - c_{\text{salt}}} \quad (3.12)$$

### 3.2.3. Isoterma de ação em efeito estérico ( ou “*steric mass action isotherm*”)

A isoterma de ação em efeito estérico (SMA) tem sido utilizada como estratégia para representar a competição entre proteínas e soluções iônicas na cromatografia de troca iônica (GU, 2015). Sua abordagem teórica considera explicitamente a cinética de adsorção e dessorção, sem impor que haja equilíbrio nesse processo. O sal da mistura compete com todas as proteínas adsorvidas na fase sólida. O modelo também considera que as proteínas são capazes de se ligar a mais de um sítio ativo, devido ao seu tamanho, bem como bloquear diversos outros (GU, 2015; GUIOCHON et al., 2006; KARKOV et al., 2013 ; HUUK et al., 2014). A equação 3.13 representa bem a situação descrita acima, considerando que a quantidade de proteína “i” adsorvida ( $q_i$ ) depende das constantes de adsorção ( $k_{Ads}$ ) e dessorção ( $k_{Des}$ ) ( $\text{mL}\cdot\text{mol}^{-1}\cdot\text{s}^{-1}$ ), do número de sítios ligantes da proteína ( $v_i$ ), da concentração de sal na fase líquida ( $c_{\text{salt}}$ ) e da carga livre resultante ( $Q$ ) na resina ( $\text{mol}\cdot\text{mL}^{-1}$ ).

$$\frac{\partial q_i}{\partial t} = K_{Ads,i} c_i Q^{v_i} - K_{Des,i} q_i c_{\text{salt}}^{v_i} \quad (3.13)$$

O sistema descrito acima apresenta neutralidade eletrostática, o que significa que o número total de sítios ativos disponíveis na resina ( $\Lambda$ ,  $\text{mol}\cdot\text{mL}_{\text{resina}}^{-1}$ ) pode ser ocupado por íons de sal ou proteínas. A equação 3.14 expressa essa condição, mostrando que  $Q$  pode ser descrito pela capacidade iônica total da resina,  $\Lambda$ , que representa o número total de sítios ativos disponíveis, e pela soma dos sítios ativos ocupados, onde o fator estérico de cada proteína ( $\alpha_i$ ) representa o seu bloqueio molecular (KARKOV et al., 2013; HUUK et al., 2014; SHEN et al., 2005; CARVALHO et al., 2016). É essencial observar a importância teórica do parâmetro  $\alpha_i$ , uma vez que muitas proteínas têm estruturas muito mais complexas que os íons em solução, de modo que suas regiões eletricamente carregadas podem se ligar a alguns sítio ativo, consequentemente bloqueando outros que possam estar disponíveis.

$$Q = \Lambda - \sum_{i=1}^N q_i (v_i + \alpha_i) \quad (3.14)$$

### 3.3. Materiais e métodos

#### 3.3.1. Produção de PspA4Pro

O processo de purificação foi aplicado à biomassa de *E. coli* produzida sob diferentes condições de cultivo. As condições de cultivo denominadas "A" e "B" foram aplicadas em *E. coli* que apresentava um fragmento do gene da proteína de superfície de pneumococo A (PspA) de *Streptococcus pneumoniae*. O domínio N-terminal e o primeiro bloco rico em prolina do gene PspA (GenBank: EF649969.1), aqui denominado PspA4Pro, foi inserido no plasmídeo pET37b + (Novagen) e expresso na *E. coli* BL21 (DE3) (Invitrogen, Carlsbad, CA, EUA) (FIGUEIREDO et al., 2017; MORENO et al., 2010). O cultivo A foi realizado a 27 °C, com indução por lactose (CAMPANI et al., 2016). O cultivo B foi realizado a 37 °C, com indução por IPTG (CAMPANI et al., 2015). Um terceiro cultivo, denominado "C", foi realizado a 37 °C, com indução por lactose, utilizando células *rE. coli* portadoras de pET37b + sem o inserto, portanto neste caso não foi produzida PspA4Pro recombinante.

As cepas recombinantes de *E. coli* foram gentilmente fornecidas pela Dra. Eliane Miyaji do Laboratório de Biologia Molecular, Instituto Butantan (São Paulo, Brasil). Todos os experimentos foram realizados em um biorreator de 5 L, com monitoramento e controle realizado pelo software SuperSys\_HCDC (HORTA et al., 2014). NH<sub>4</sub>OH (30%) foi utilizado para manter o pH dos cultivos em 6,7. O controlador ajustou automaticamente a velocidade do agitador entre 200 e 900 rpm para manter a concentração de oxigênio dissolvido em 30% de saturação. A vazão total do gás foi mantida em 4 L.min<sup>-1</sup> (CAMPANI et al., 2016; HORTA et al., 2014). Ao final dos cultivos, as biomassas foram separadas do caldo por centrifugação de 30 min, a 4 °C e 6700 g. Os pellets de células foram armazenados em sacos plásticos a -80 °C.

Antes de executar a cromatografia de troca aniônica (CTA), 100 ± 10 g das células foram descongeladas e ressuspensas em 1 L de tampão de lise. As células foram então rompidas através de circulação da suspensão em um homogeneizador contínuo de alta pressão por 8 minutos (4 °C, 500 bar e vazão de 1 L.min<sup>-1</sup>). Foi então adicionado 0,1% de detergente de brometo de cetiltrimetilamônio (CTAB), utilizado como um agente precipitante para remover detritos celulares, lipopolissacarídeos e impurezas proteicas da célula hospedeira. Após misturar por 60 min à temperatura ambiente a amostra foi centrifugada por 90 min a 17.696 g e 4 °C (FIGUEIREDO et al., 2017). O sobrenadante foi chamado de extrato celular clarificado (ECC) e apresentou PspA4Pro, assim como várias outras impurezas proteicas produzidas pelo

metabolismo celular. As amostras de ECC dos cultivos A (realizado com células produtoras de PspA4Pro induzidas por lactose, a 27 °C), B (realizado com células produtoras de PspA4Pro e induzido por IPTG, a 37 °C) e C (realizado com células não produtoras de PspA4Pro e induzido por lactose, a 37 °C) foram denominadas ECC<sub>A</sub>, ECC<sub>B</sub> e ECC<sub>C</sub>, respectivamente.

### 3.3.2. Cromatografia de troca aniônica

Foram executados quatro procedimentos diferentes de cromatografia de troca aniônica (designados CTA-I a CTA-IV). CTA-I, CTA-II e CTA-III foram aplicados a ECC<sub>A</sub>, ECC<sub>B</sub> e ECC<sub>C</sub>, respectivamente, e CTA-IV foi aplicado a uma solução de proteína pura (SPP) preparada com PspA4Pro previamente purificado (FIGUEIREDO et al., 2017). As cromatografias foram realizadas a resina Q-Sepharose FF empacotada em uma coluna XK 50/30, de 250 mL, que foi instalada em um cromatógrafo Äkta Avant 150. Todos os materiais cromatográficos foram fornecidos pela GE *Healthcare* (EUA). O fluxo de operação foi de 2,50 mL.min<sup>-1</sup>.cm<sup>-2</sup> (50 mL.min<sup>-1</sup>). O tampão de equilíbrio foi preparado com fosfato de sódio 10 mM (pH 6,5 e condutividade 1,2 mS.cm<sup>-1</sup>). O pH e a condutividade das amostras ECC<sub>A</sub>, ECC<sub>B</sub> e ECC<sub>C</sub> e SPP foram ajustados para os mesmos valores de pH e condutividade do tampão de equilíbrio. Os tampões de eluição foram preparados com NaCl nas concentrações de 150, 300 e 1000 mM e diluídos no tampão de equilíbrio. A cromatografia foi realizada de acordo com as seguintes etapas: (i) aplicação de 5 volumes de coluna (VC) do tampão de equilíbrio; (ii) aplicação de amostras ECC ou SPP; (iii) aplicação de 5 VC do tampão de equilíbrio para remoção de moléculas não retidas pela resina após a introdução da amostra; (iv) aplicação sequencial de 5 VC dos tampões de eluição contendo 150, 300 e 1000 mM de NaCl.

A fração não adsorvida, composta pelo material das etapas (ii) e (iii), foi denominada QNAds. As frações de eluição obtidas com os tampões de 150, 300 e 1000 mM foram denominadas Q150, Q300 e Q1000, respectivamente.

### 3.3.3. Quantificação de proteínas e determinação de pureza da PspA4Pro

Alíquotas de 10 mL das frações ECC, QNAds, Q150, Q300 e Q1000 foram coletadas durante cada CTA. Os volumes totais dessas frações foram medidos e as concentrações totais de proteínas foram determinadas pelo método de Lowry (LOWRY et al., 1951). A pureza de

PspA4Pro foi determinada por análise de densitometria de bandas usando SDS-PAGE a 12% contendo 2-mercaptoetanol (LAEMMLI et al., 1970), com um densitômetro Bio-Rad GS-800 calibrado. As imagens foram analisadas pelo software *Image Lab* v. 6.0. A pureza relativa da PspA4Pro foi calculada como a porcentagem da banda identificada em relação à soma das densidades de todas as outras bandas (FIGUEIREDO et al., 2017).

### 3.3.4. Modelagem e simulação

A Figura 3.1 mostra o procedimento de modelagem e simulação da CTA que foi realizado usando o software EMSO<sup>TM</sup> (*Environment for Modeling Simulation and Optimization*), um simulador de equações com linguagem orientada a objetos internos (SOARES et al., 2003). Tanto a isoterma de Langmuir modificada quanto a isoterma de ação em efeito estérico foram modeladas. As derivadas espaciais foram aproximadas pelo método das diferenças finitas, com  $N_t$  elementos, e as derivadas de tempo foram solucionadas pelo simulador utilizando um algoritmo de uso geral para sistemas de equações diferenciais-algébricas.

#### *A) Características de fluxo da coluna*

Os parâmetros  $L$  e  $u$  (Equação 3.2) foram obtidos a partir das condições de operação do sistema. Os parâmetros  $N_t$  e  $\epsilon_T$  foram determinados experimentalmente, utilizando um pulso de 0,5 mL de acetona a 1% (v.v<sup>-1</sup>) diluída em água ultra-pura, a uma vazão de 2,71 mL.min<sup>-1</sup>. O número de pratos teóricos ( $N_t$ ) foi calculado usando a Equação 3.3, enquanto que o parâmetro  $\epsilon_T$  foi determinado usando as Equações 3.4 e 3.5.

#### *B) Estimativas de parâmetros, simulações e validações experimentais*

As amostras ECC eram compostas por uma quantidade de proteínas diversificadas que apresentavam diferentes afinidades com a resina. Essas proteínas foram divididas em diferentes pools. Para modelagem e simulação, considerou-se que cada pool era composto por um grande número de proteínas desconhecidas com diferentes estruturas moleculares, mas com afinidade química semelhante em relação à resina Q-Sepharose, uma vez que foram

eluídas juntas. Apenas grupos proteicos adsorvidos pela resina foram modelados. As proteínas não adsorvidas (QNAds) não foram simuladas, pois não apresentaram afinidade com a resina.

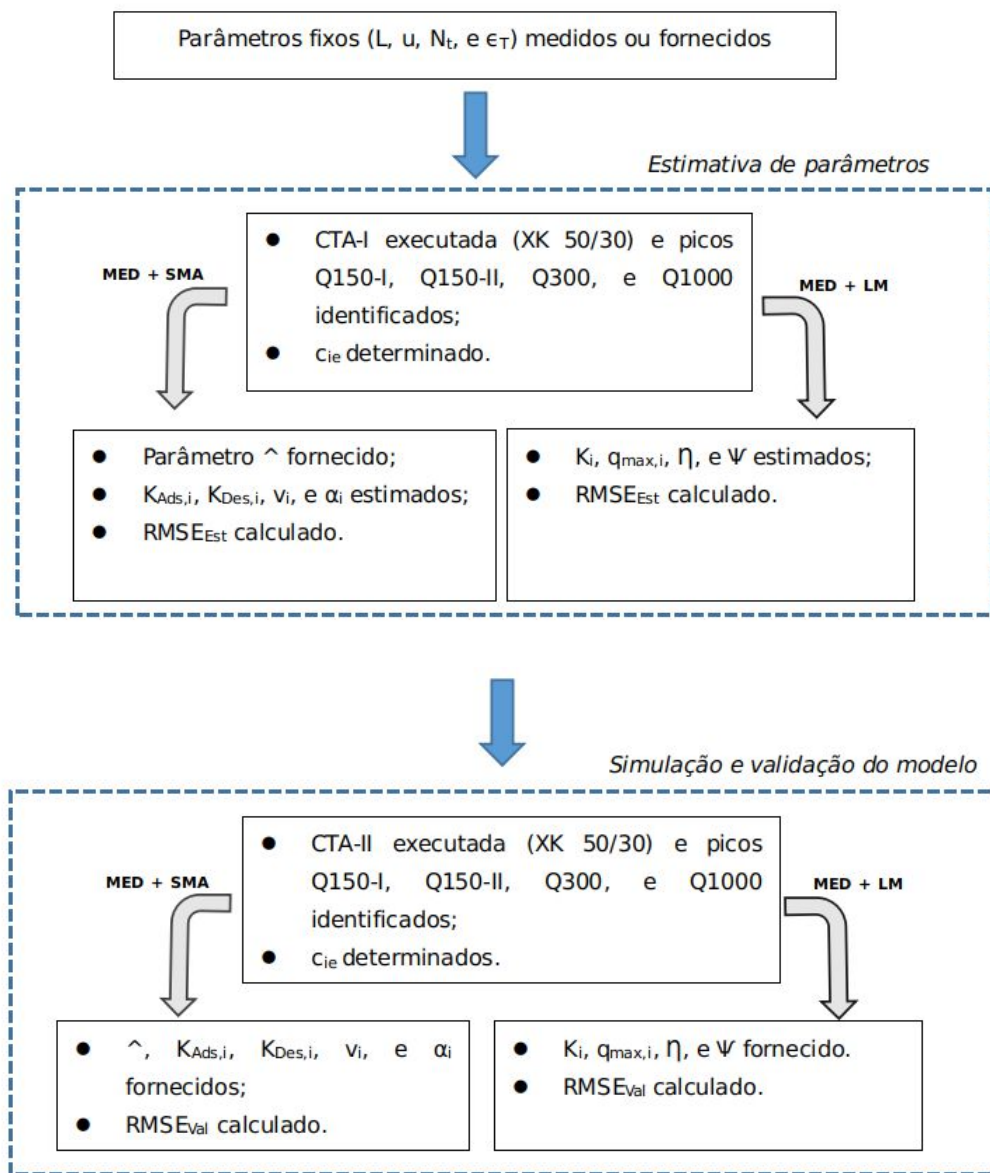


Figura 3.1. Descrição da estratégia de modelagem e simulação do processo utilizando o modelo do equilíbrio dispersivo (MED) em conjunto com as isotermas de ação em efeito estérico (SMA) ou Langmuir modificada (LM). Os parâmetros foram estimados a partir dos dados da CTA-I e a validação do modelo foi realizada com os dados da CTA-II.

Um total de quatro picos foram observados no cromatograma CTA-I: dois (Q150-I e Q150-II) para eluição realizada com 150 mM de NaCl e dois (Q300 e Q1000) para as outras frações de eluição. Esses picos podem ser visualizados nas cromatografias CTA-I, CTA-II e CTA-III na Figura 3.5. Para fins de modelagem, cada pico foi associado a um pool de proteínas “i” com características químicas específicas (CHAN et al., 2008). Em todas as simulações, a resina cromatográfica foi considerada inicialmente limpa, com todos os sítios ativos vinculados ao contra-íon padrão. As concentrações ( $c_{i,e}$ ) de cada um dos conjuntos de proteínas “i” (Q150-I, Q150-II, Q300 e Q1000) na ECC (solução de entrada) foram fornecidas, uma vez que as massas de proteínas para cada eluição foram quantificadas durante o experimento usando o método Lowry, e o volume das soluções de ECC eram conhecidos.

Para Q150, uma vez que Q150-I e Q150-II foram eluídos juntos, a massa de cada pool particular foi baseada na área de cada pico no cromatograma. A descrição matemática dos eventos que ocorreram dentro da coluna foi baseada no modelo do equilíbrio dispersivo (MED), juntamente com os modelos das isotermas de adsorção também necessários para complementar as equações do balanço de massa. O MED foi testado separadamente com os modelos da isoterma modificada de Langmuir (ML) e de ação em efeito estérico, ou “*steric mass action*” (SMA). Os parâmetros foram estimados a partir dos dados experimentais de CTA-I. Os valores de condutividade e absorvância a 280 nM foram fornecidos como as concentrações de NaCl e proteína, respectivamente. Para o primeiro caso, foi utilizada uma curva padrão construída com diferentes concentrações de NaCl e as condutividades correspondentes. Para o segundo caso, foi estabelecido um coeficiente de extinção molar médio para cada fração eluída, tornando as massas totais de proteínas associada aos picos cromatográficos as mesmas que os valores determinados pelo método de Lowry (LOWRY et al., 1951)<sup>3</sup>. Os parâmetros estimados foram  $K_i$ ,  $q_{\max,i}$ ,  $\Gamma$  e  $\Psi$  quando a isoterma ML foi testada, e  $k_{\text{Ads},i}$ ,  $k_{\text{Des},i}$ ,  $v_i$  e  $\alpha_i$  quando a isoterma SMA foi testada. O parâmetro  $\Lambda$  foi fornecido por HUUK et al. (2014). Para a isoterma SMA, todos os procedimentos de modelagem e simulação foram realizados usando “mol” como unidade de medida, seguido de conversão em massa usando a massa molar teórica de PspA4Pro (EXPASY, 2019).

---

<sup>3</sup> O Apêndice B explica de forma detalhada as estratégias utilizadas para disponibilizar os dados de condutividade e absorvância (280 nM) em forma de concentrações de NaCl e proteínas.

Os parâmetros foram estimados com a função objetiva dos mínimos quadrados. O problema de otimização foi resolvido no EMSO com o algoritmo simplex Nelder-Mead (NELDER, 1965). O erro quadrático médio da raiz, ou “*root mean square error*” (RMSE<sub>Est</sub>) para a estimativa de parâmetros (Equação 3.15) foi utilizado como métrica para avaliar a qualidade do ajuste do modelo, onde N é o número total de pontos experimentais,  $\sigma$  é o número de parâmetros estimados,  $y_i$  são os valores de saída do processo e  $\hat{y}_i$  são os valores de saída do modelo. Na próxima etapa, os dados da CTA-II foram usados para validar os modelos ML e SMA, e a Equação 3.15 foi usada novamente para calcular o RMSE<sub>Val</sub> (com  $\sigma = 0$ ).<sup>4</sup>

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - \sigma}} \quad (3.15)$$

### 3.3.5. Estimativa dos parâmetros e validações experimentais do extrato celular clarificado virtual

A Figura 3.2 mostra um esquema detalhado que explica como as equações MED e SMA foram usadas para descrever os dados obtidos por CTA-III, representando o pool de proteínas de *Escherichia coli* quando a proteína recombinante não foi expressa, e por CTA-IV, representando somente PspA4pro. Todos os parâmetros fixos e as estratégias de estimativa de parâmetros foram os mesmos que os apresentados na seção anterior. Esperava-se que esta abordagem fosse capaz de avaliar as propriedades particulares de PspA4Pro e das outras proteínas eluídas na fração Q300. Portanto, os conjuntos de proteínas Q150-I, Q150-II e Q1000 não foram simulados neste caso, e os parâmetros da isoterma SMA ( $k_{Ads,i}$ ,  $k_{Des,i}$ ,  $v_i$  e  $\alpha_i$ ) foram estimados separadamente para a PspA4Pro e o conjunto de outras proteínas eluídas em Q300. Após a estimativa desses parâmetros foi então criado um ECC virtual com 0,51 g de PspA4Pro e 0,18 g do pool de proteínas Q300, em um volume total de 0,9 L, e  $c_{i,e}$  foi determinado para cada caso, gerando um ECC virtual com pureza semelhante à pureza experimental normalmente observada nas eluições Q300 (entre 60% e 80%). A simulação foi

---

<sup>4</sup> O Apêndice C mostra um roteiro detalhado de simulação, bem como todas as equações envolvidas para modelagem matemática da coluna.

realizada utilizando uma coluna XK 26/20 e os parâmetros estimados de CTA-III e CTA-IV. O pico Q300 simulado foi dividido em seis subfrações, cada uma com 0,5 VC, e as áreas obtidas pela eluição de PspA4Pro e o conjunto de outras proteínas foram usadas para determinar as purezas de PspA4Pro e as massas totais de proteínas (MTP) em cada subfração.

A validação experimental foi realizada a partir de uma nova corrida cromatográfica, denominada CTA-V, que foi aplicada a um ECC obtido a partir de 30 g de células A (as mesmas utilizadas na CTA-I). A cromatografia de troca aniônica foi realizada usando 70 mL de resina Q-Sepharose FF empacotada em uma coluna XK 26/20 e instalada no sistema cromatográfico Akta Avant 150. A velocidade linear aplicada foi de  $2,50 \text{ mL} \cdot \text{min}^{-1} \cdot \text{cm}^{-2}$  ( $13,5 \text{ mL} \cdot \text{min}^{-1}$ ). Todos os outros procedimentos experimentais relativos ao equilíbrio da coluna, aplicação da amostra e eluições foram exatamente os mesmos que os aplicados quando a coluna XK 50/30 foi usada nas CTA-I a CTA-IV. Como na simulação, o pico Q300 também foi fracionado em seis subfrações, cada uma com 0,5 VC. Para todas elas, as MTP e purezas de PspA4Pro foram quantificadas para que os resultados experimentais da coluna XK 26/20 pudessem ser comparados à simulação.

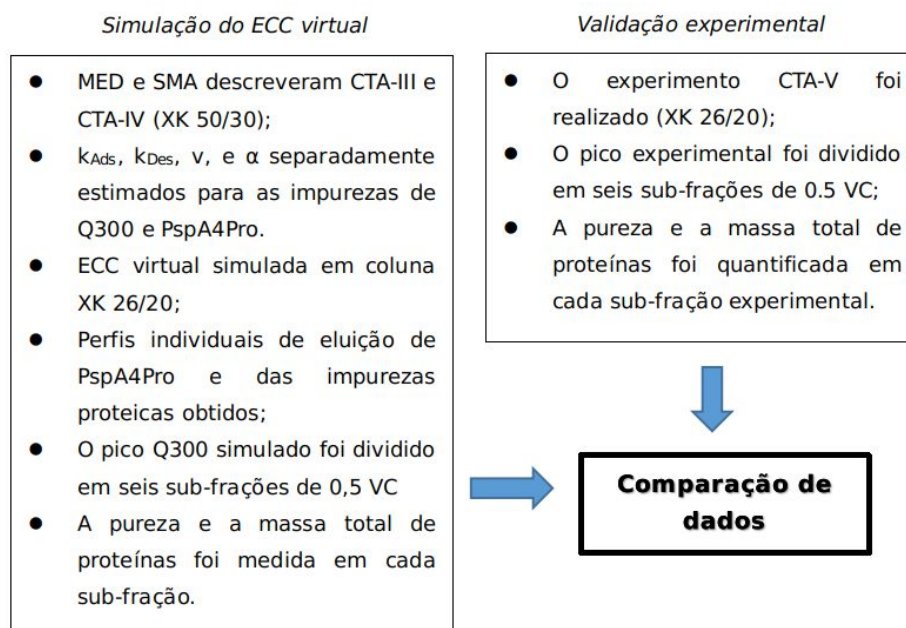


Figura 3.2. Descrição da estratégia de simulação do extrato celular clarificado (ECC) virtual utilizando o modelo do equilíbrio dispersivo (MED) em conjunto com a isoterma de ação em efeito estérico (SMA). Os parâmetros do pool de proteínas e PspA4Pro foram estimados a partir dos dados CTA-III e CTA-IV, respectivamente.

### 3.4. Resultados e Discussão

#### 3.4.1. Visão geral das cromatografias de troca aniônica

Os extratos celulares clarificados (ECC) e a solução pura de PspA4Pro (SPP) foram obtidos de acordo com os procedimentos descritos na Seção 3.3.1. Em seguida, foram realizadas as cromatografias de troca aniônica (CTA) e as amostras foram analisadas de acordo com os procedimentos descritos nas seções 3.3.2 e 3.3.3, respectivamente. A Tabela 3.1 mostra as características de cada CTA, incluindo os materiais de partida, além de fornecer informações sobre as frações QNAds, Q150, Q300 e Q1000.

Tabela 3.1. Características das condições de entrada, frações não adsorvidas (QNAds) e frações de eluição (Q150, Q300 e Q1000) para as cromatografias de troca aniônica (CTA) I a IV executadas em uma coluna XK 50/30.

		CTA		I	II	III	IV
		Material de partida		ECC <sub>A</sub>	ECC <sub>B</sub>	ECC <sub>C</sub>	SPP
<b>Condições de entrada</b>	MTP	(g)	9,54 ± 0,06	12,23 ± 0,19	28,45 ± 0,08	4,25 ± 0,08	
	Pureza	(%)	36,0	12,6	-	95	
	Volume	(L)	3,3	2,6	3,3	3,3	
<b>QNAds</b>	MTP	(g)	4,20 ± 0,17	6,30 ± 0,40	18,69 ± 0,04	0,20 ± 0,13	
	Pureza	(%)	5,8	5,8	-	0	
<b>Q150</b>	MTP	(g)	1,60 ± 0,01	2,31 ± 0,02	2,82 ± 0,01	0,21 ± 0,11	
	Pureza	(%)	8,8	5,4	-	0	
<b>Q300</b>	MTP	(g)	2,96 ± 0,05	2,02 ± 0,02	2,86 ± 0,01	4,38 ± 0,20	
	Pureza	(%)	61,8	36,2	-	95	
<b>Q1000</b>	MTP	(g)	0,62 ± 0,03	1,63 ± 0,41	2,82 ± 0,01	0,21 ± 0,12	
	Pureza	(%)	14,8	6,3	-	0	

\*MTP: Massa total de proteínas

\*ECC: Extrato celular clarificado

\*SPP: Solução pura de PspA4Pro

Os resultados de pureza mostram que a maior parte da PspA4Pro foi eluída na fração Q300, como já comprovado por FIGUEIREDO et al., 2017. Através de uma comparação entre CTA-I e CTA-II é possível perceber que a fração Q300 do primeiro experimento apresenta maior pureza, embora os mesmos procedimentos de lise e purificação

tenham sido executados em ambos os casos. Isso indica que o metabolismo celular responde de maneira particular às estratégias de cultivo aplicadas nos processos *upstream* das células A e B; portanto, para cada caso diferentes proteínas foram produzidas juntamente com PspA4Pro. Isso também fica evidente quando as purezas de entrada para CTA-I e CTA-II são analisadas (36,0 e 12,6%, respectivamente), o que mostra que a amostra ECC<sub>A</sub> apresenta mais PspA4Pro que ECC<sub>B</sub>. Esses diferentes perfis de proteínas para ECC<sub>A</sub> e ECC<sub>B</sub> podem ser explicados pelas condições de indução dos cultivos A e B, uma vez que a utilização de IPTG ou lactose leva a diferentes respostas metabólicas de *E. coli* (HEYLAND et al., 2011; MARISCH et al., 2013; LOPES et al., 2019). Portanto, condições mais brandas de indução com lactose a 27 °C favoreceram a síntese de PspA4Pro, enquanto que condições mais intensas de indução com IPTG a 37 °C acarretaram em uma alta carga metabólica e menor produção de PspA4Pro. Imagens dos géis (SDS-PAGE para) das amostras ECC<sub>A</sub>, ECC<sub>B</sub>, ECC<sub>C</sub> e SPP e suas respectivas frações Q300 (Figuras 3.3 e 3.4) também revelaram as diferentes composições do conjunto de impurezas proteicas em cada material de partida processado, bem como a maior composição do PspA4Pro na amostra ECC<sub>A</sub>.

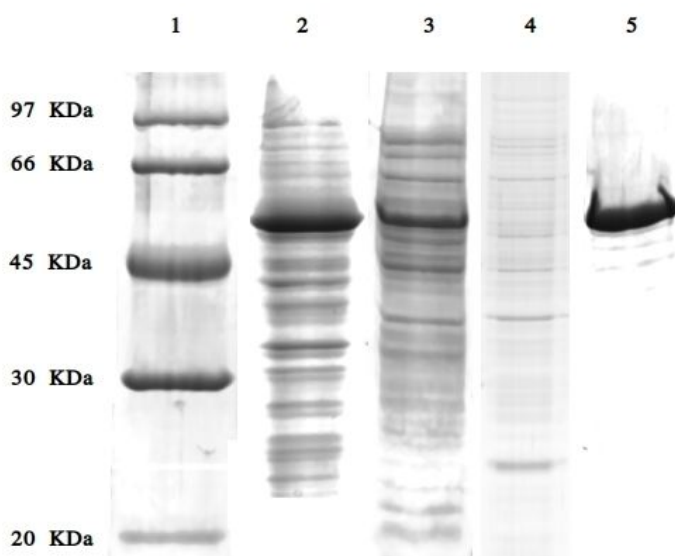


Figura 3.3. SDS-PAGE dos extratos celulares clarificados (ECC) com PspA4Pro recombinante e outras proteínas. 1) Marcador molecular contendo fosforilase (97 kDa), albumina (66 kDa), ovalbumina (45 kDa), anidro (30 kDa), tripsina (20 kDa) e lactalbumina (14,4 kDa). 2) ECC<sub>A</sub> produzido por células cultivadas com indução de lactose a 27 °C. 3) ECC<sub>B</sub> produzido por células cultivadas com indução de IPTG a 37 °C 4) ECC<sub>C</sub> produzido por células portadoras do vetor pET37b sem o inserto de PspA4Pro. 5) Amostra pura de PspA4Pro (SPP).

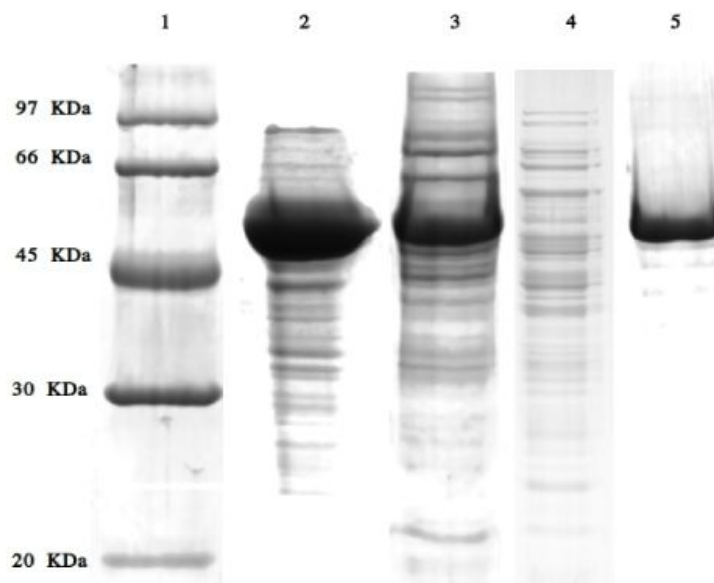


Figura 3.4. SDS-PAGE das frações de eluições Q300 produzidas pela cromatografia. 1) Marcador molecular contendo fosforilase (97 KDa), albumina (66 KDa), ovalbumina (45 KDa), anidro (30 KDa) e tripsina (20 KDa). 2) Fração Q300 produzida por células cultivadas com indução de lactose a 27 °C 3) Fração Q300 produzida por células cultivadas com indução de IPTG a 37 °C 4) Fração Q300 produzida por células portadoras do vetor pET37b sem o inserto de PspA4Pro. 5) Fração Q300 produzida por amostra pura de PspA4Pro (SPP).

As purezas das frações CTA-III não foram quantificadas, pois nesse caso as células não produziram PspA4Pro. A pureza de entrada da amostra CTA-IV foi de 95%, uma vez que era composta por PspA4Pro purificado (SPP), o que levou à eluição de quase toda a massa de proteína na fração Q300. Portanto, é importante destacar que os picos Q300 de CTA-IV e CTA-III representaram os perfis cromatográficos de PspA4Pro e do conjunto de demais proteínas que foram eluídas na mesma força iônica, respectivamente.

A Figura 3.5 fornece uma comparação entre os cromatogramas experimentais das CTAs descritas na Tabela 3.1. O tempo inicial foi definido como o início da aplicação da primeira eluição. Foram observados nos cromatogramas diferentes perfis de eluição Q150 para CTA-I, CTA-II e CTA-III, novamente indicando que modificações nas condições de cultivo resultaram em produções de conjuntos de proteínas com diferentes composições (Figura 3.3), gerando picos diferentes nas frações de eluição. As curvas de eluição mostradas na Figura 3.5 representam satisfatoriamente as condições reais do processo para a purificação de um extrato celular clarificado. A variabilidade existente nas características do material de partida foi uma

dificuldade adicional que teve que ser tratada pelo modelo, a fim de prever com precisão o perfil cromatográfico de cada mistura de proteínas.

Uma comparação entre os picos Q300 de CTA-I, CTA-II e CTA-IV mostra que picos mais largos foram obtidos nos dois primeiros casos, enquanto que um pico mais estreito foi observado em CTA-IV. Essas diferenças podem ser atribuídas à presença de proteínas de *Escherichia coli* nas eluições de CTA-I e CTA-II, as quais não estavam presentes em CTA-IV. Por outro lado, os picos Q300 e Q1000 de CTA-III apresentaram um perfil completamente diferente. Como nesse caso PspA4Pro não foi produzida, houve uma grande quantidade de proteínas com diferentes características de interação com a coluna eluídas, levando à produção de um pico mais largo.

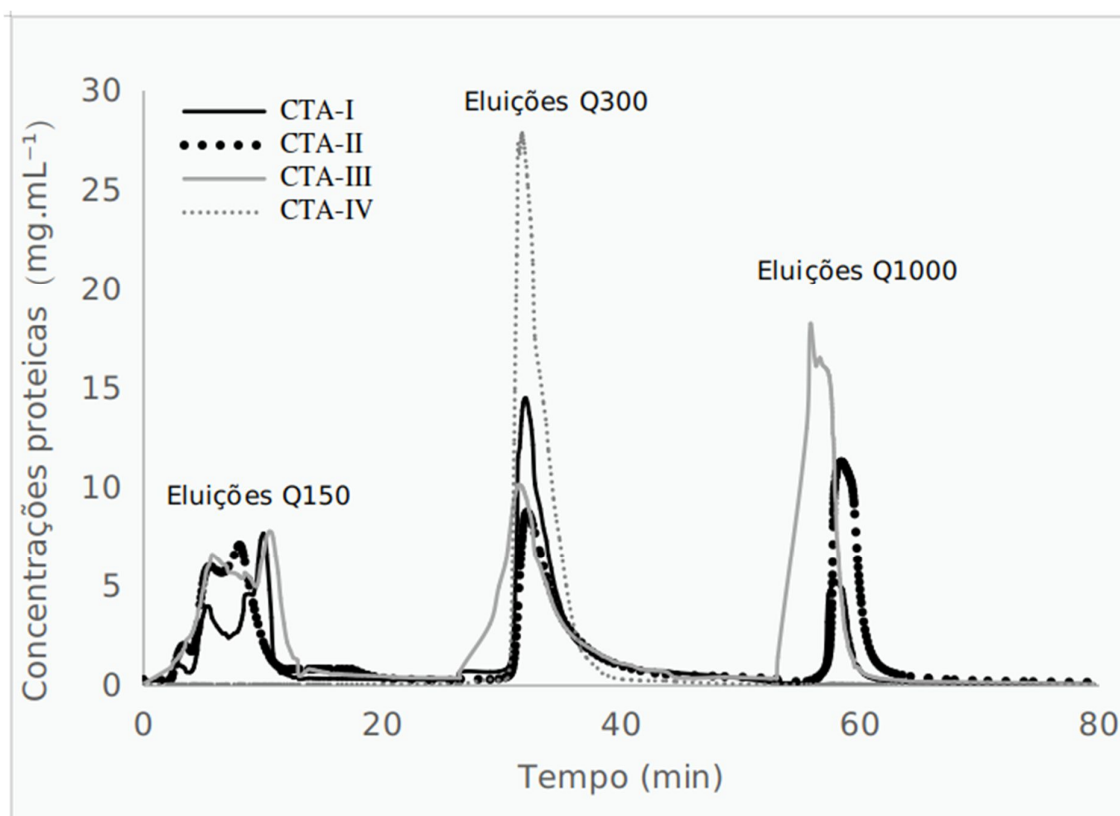


Figura 3.5. Comparação de cromatogramas experimentais: CTA-I realizada com extrato celular clarificado (ECC) do cultivo A (células induzidas por lactose a 27 °C), CTA-II realizada com ECC do cultivo B (células induzidas por IPTG a 37 °C), CTA-III com ECC do cultivo C (células portadoras do vetor pET37b sem inserto de PspA4Pro) e CTA-IV (PspA4Pro pura).

### 3.4.2. Estimativa de parâmetros, simulações e validações experimentais dos modelos de adsorção

Esta seção avalia os modelos das isotermas de adsorção utilizadas em conjunto com o MED para descrição da CTA. Os procedimentos de modelagem e simulação foram realizados conforme descrito pela Seção 3.3.4 e Figura 3.1. A Tabela 3.2 apresenta as condições iniciais e os parâmetros fixos, onde  $q_{i(1-N_t)}$  e  $c_{i(1-N_t)}$  são as concentrações iniciais de proteína nas fases estacionária e móvel, respectivamente, em todos os pratos teóricos,  $L$  e  $u$  são parâmetros fixos definidos pelo sistema, e  $\epsilon_T$  e  $N_t$  foram determinados conforme descrito na Seção 3.3.4.A.

Todos os parâmetros estimados são mostrados na Tabela 3.3. Para CTA-I, os parâmetros das isotermas SMA e LM foram estimados para todos os picos observados: Q150-I, Q150-II, Q300 e Q1000. Cada pico foi composto por um conjunto de proteínas que foram eluídas no mesmo momento, apresentando afinidade química semelhante com a resina, a qual foi representada por parâmetros específicos. Para CTA-II, os parâmetros não foram estimados, pois nesse caso os dados foram utilizados para validação. Para CTA-III e CTA-IV (cujos resultados de simulação serão discutidos posteriormente), os parâmetros de SMA foram estimados apenas para o pico Q300. Como a CTA-III foi aplicada na amostra ECC<sub>C</sub> que não possuía PspA4Pro, nesse caso os parâmetros foram estimados apenas para o pool de impurezas proteicas dessa eluição, enquanto a CTA-IV forneceu os parâmetros de PspA4Pro.

Tabela 3.2. Condições iniciais e parâmetros fixos usados para a simulação da cromatografia de troca aniônica.

Nome do parâmetro	Símbolo	Unidade	Valor
Proteínas adsorvidas iniciais	$q_{i(1-N_t)}$	(g.mL <sup>-1</sup> )	0
Proteínas não adsorvidas iniciais	$c_{i(1-N_t)}$	(g.mL <sup>-1</sup> )	0
Comprimento da coluna cromatográfica	$L$	(cm)	13
Velocidade da fase móvel	$u$	(cm.s <sup>-1</sup> )	0,042
Tamanho da partícula de resina	$d_p$	$\mu\text{m}$	90
Porosidade total da resina (XK 50/30)	$\epsilon_T$	-	0,9
Porosidade total da resina (XK 26/20)	$\epsilon_T$	-	0,906
Número total de pratos teóricos (XK 50/30)	$N_T$	-	287
Número total de pratos teóricos (XK 26/20)	$N_T$	-	90
Capacidade de saturação da resina <sup>A</sup>	$\Lambda$	(mol.L <sup>-1</sup> )	2,15

<sup>A</sup>Fornecido por Huuk, 2014

Tabela 3.3. Parâmetros estimados das isotermas de ação em efeito estérico (SMA) e Langmuir modificada (LM) durante as corridas cromatográficas CTA-I, CTA-III e CTA-IV.

CTA	Eluição	SMA					LM			
		$k_{Ads}$ (mL,g <sup>-1</sup> ,s <sup>-1</sup> )	$k_{Des}$ (mL,g <sup>-1</sup> ,s <sup>-1</sup> )	$K_{Eq}^A$	$\alpha$	$v$	$q_{max}$ (g,mL <sup>-1</sup> )	$K$ (g,mL <sup>-1</sup> )	$\eta$ (g,mL <sup>-1</sup> )	$\Psi$ (g,mL <sup>-1</sup> )
I <sup>B</sup>	Q150-I	0,0017	0,0105	0,16	21,4	1,0	15	0,05	0,07	0,0002
	Q150-II	0,0031	0,0088	0,35	0,02	1,2	23	0,06	0,10	0,0004
	Q300	0,0056	0,0087	0,65	25,1	1,9	36	0,04	0,25	0,001
	Q1000	0,0098	0,0022	4,42	24,0	1,6	999	0,78	0,11	0,002
III <sup>C</sup>	Q300	0,0064	0,0027	2,37	25,3	1,3	-	-	-	-
IV <sup>C</sup>	Q300	0,0015	0,0179	0,83	7,7	1,7	-	-	-	-

<sup>A</sup>  $K_{Eq}$  representa a constante de equilíbrio ( $K_{Eq} = k_{Ads} \cdot k_{Des}^{-1}$ ).

<sup>B</sup> Os Parâmetros de SMA e LM foram estimados para os picos Q150-I, Q150-II, Q300, e Q1000 na CTA-I, de acordo com os procedimentos descritos na Figura 3.1

<sup>C</sup> Os parâmetros SMA foram estimados para o pico Q300 das CTA-III e CTA-IV para descrever o pool de proteínas e PspA4Pro, respectivamente, no pico Q300, de acordo com os procedimentos descritos na Figura 3.2.

Os parâmetros SMA de CTA-I se mostraram adequados para a descrição fornecida pelas equações matemáticas, pois proteínas com maior afinidade com a resina apresentaram maiores constantes de adsorção ( $k_{Ads,i}$ ) e menores constantes de dessorção ( $k_{Des,i}$ ). O oposto ocorreu para proteínas com menor afinidade pela resina, indicando um menor tempo de residência na coluna. Os valores de  $v_i$  foram maiores para os grupos de proteínas com maior afinidade com a resina ( $v_{Q300}$  e  $v_{Q1000}$ ), mostrando que eles tendem a apresentar maiores quantidades de sítios ativos, explicando as ligações mais fortes com a coluna. Os valores de  $\alpha_i$  apresentaram diferentes ordens de magnitude, diferindo consideravelmente entre o grupo Q150-II (0,02) e os demais grupos (20 a 25). Uma diferença ainda maior foi observada por Faraji et al. (FARAJI et al., 2015), que estudou os perfis de adsorção de lactoperoxidase e lactoferrina em SP-Sepharose, obtendo valores de 1283 e 0,16, respectivamente. Considerando os outros parâmetros, estudos recentes relataram valores com ordens de magnitude semelhantes às obtidas neste estudo, com  $K_{Eq,i}$  variando de 0,01 a 13 e  $v_i$  variando de 0,06 a 3,0 (FARAJI et al., 2015; KARKOV et al., 2013; HUUK et al., 2014).

Os parâmetros estimados para a isoterma LM também foram adequados para descrição matemática do processo, uma vez que um maior  $K_i$  foi observado para as proteínas Q1000 ( $K_{Q1000} > K_{Q300} \approx K_{Q150-II} \approx K_{Q150-I}$ ), indicando maior afinidade desse pool de proteínas

com a resina e explicando o maior tempo de residência. Valores mais elevados de  $q_{\max,i}$  foram observados para eluições de proteínas com maiores afinidades pela resina ( $q_{\max,Q1000} > q_{\max,Q300} > q_{\max,Q150-I} > q_{\max,Q150-II}$ ), o que também explicou os diferentes tempos de residência. Apesar desses resultados, é importante destacar que os parâmetros LM são utilizados apenas para fins de modelagem, sem significado físico direto, uma vez que essa isoterma é empírica. Neste contexto, outros estudos relataram valores com diferentes ordens de magnitude utilizando a equação de Langmuir não modificada: GERONTAS et al. (2010) observaram valores de  $q_{\max}$  entre 0 e 0,5 g.mL<sup>-1</sup> para albumina bovina (BSA) e lactoferrina adsorvida em resinas cromatográficas. CLOSE et al. (2014) relataram um valor de  $q_{\max}$  de 0,00645 g.mL<sup>-1</sup>, enquanto OSUNA et al. (2012) obtiveram 0,073 g.mL<sup>-1</sup> para a adsorção de anticorpos por resinas cromatográficas. YU et al. (2012) relataram valores de  $q_{\max}$  entre 0,08 e 0,20 g.mL<sup>-1</sup> para BSA.

A Figura 3.6 mostra os dados experimentais da CTA-I e os modelos com as estimativas de parâmetros quando as isotermas SMA e LM foram usadas. As linhas verticais tracejadas indicam as eluições com diferentes gradientes de NaCl, as quais podem ser associadas aos picos cromatográficos. As raízes dos erros quadráticos médios para as estimativas (ou *root mean square errors*,  $RMSE_{Est}$ ) foram calculadas de acordo com a Equação 3.15. Os valores observados de  $RMSE_{Est}$  foram da mesma ordem de magnitude para as isotermas SMA e LM ( $8,1 \times 10^{-4}$  e  $7,9 \times 10^{-4}$ , respectivamente), mostrando que esses modelos forneceram descrições igualmente satisfatórias dos dados experimentais.

A Figura 3.7 mostra os dados experimentais da CTA-II e os resultados das simulações realizadas usando os parâmetros fornecidos pela Tabela 3.3. Novamente, os valores de  $RMSE$  para essa validação ( $RMSE_{Val}$ ) foram da mesma ordem de magnitude para SMA e LM ( $1,6 \times 10^{-3}$  e  $1,4 \times 10^{-3}$ , respectivamente). Além disso, os valores de  $RMSE_{Val}$  foram aproximadamente 2,0 vezes maiores que os valores de  $RMSE_{Est}$ . Isso era esperado, uma vez que  $RMSE_{Val}$  foi calculado usando os dados do perfil de eluição do ECC<sub>B</sub> (CTA-II) e as simulações foram executadas com as condições iniciais correspondentes para esse extrato celular clarificado (Tabelas 3.1 e 3.2), mas os parâmetros utilizados nesse caso (Tabela 3.3) foram os mesmos estimados a partir dos dados de ECC<sub>A</sub> (CTA-I). Como já mencionado, ECC<sub>A</sub> e ECC<sub>B</sub> foram produzidos a partir de células cultivadas sob diferentes condições. Portanto, diferentes misturas complexas de proteínas foram produzidas para cada caso (Figura 3.3), apresentando diferentes propriedades físico-químicas, e conseqüentemente prejudicando

o desempenho do modelo. No entanto, apesar desses obstáculos, os modelos das isothermas SMA e LM foram capazes de descrever satisfatoriamente os dados experimentais. É importante destacar que a análise realizada neste trabalho teve o objetivo de comparar a adequação das isothermas de adsorção na validação dos cromatogramas.

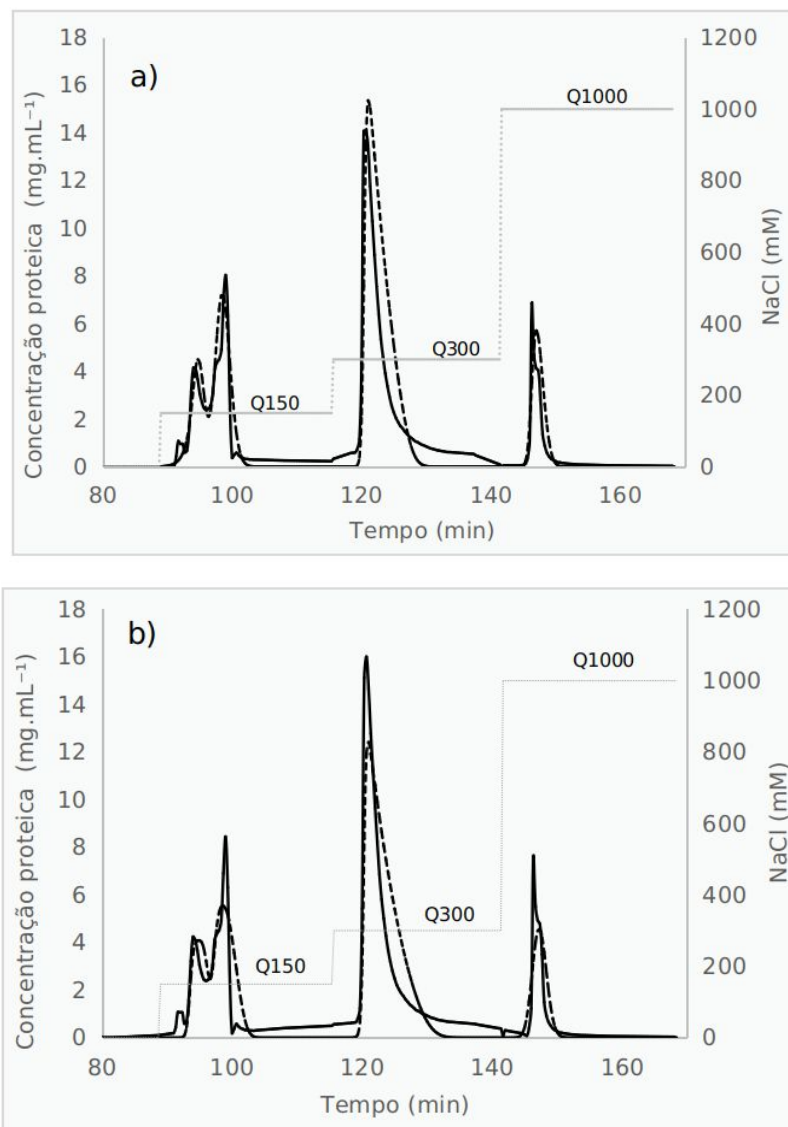


Figura 3.6. Perfis de eluição para as proteína totais dos extratos celulares ECC<sub>A</sub> na CTA-I (estimativa dos parâmetros do modelo): a) Dados experimentais (linha preta sólida), perfis simulados usando a isoterma de ação em efeito estérico (SMA) em conjunto com o modelo do equilíbrio dispersivo (MED) (linha preta tracejada) e perfis de eluição obtidos a partir de aplicação de NaCl a 150 mM, 300 mM e 1000 mM (linha cinza tracejada); b) Dados experimentais (linha preta sólida), perfis simulados usando a isoterma de Langmuir modificada (LM) em conjunto com o modelo do equilíbrio dispersivo (MED), e perfis de eluição obtidos a partir de aplicação de NaCl a 150 mM, 300 mM e 1000 mM (linha cinza tracejada)

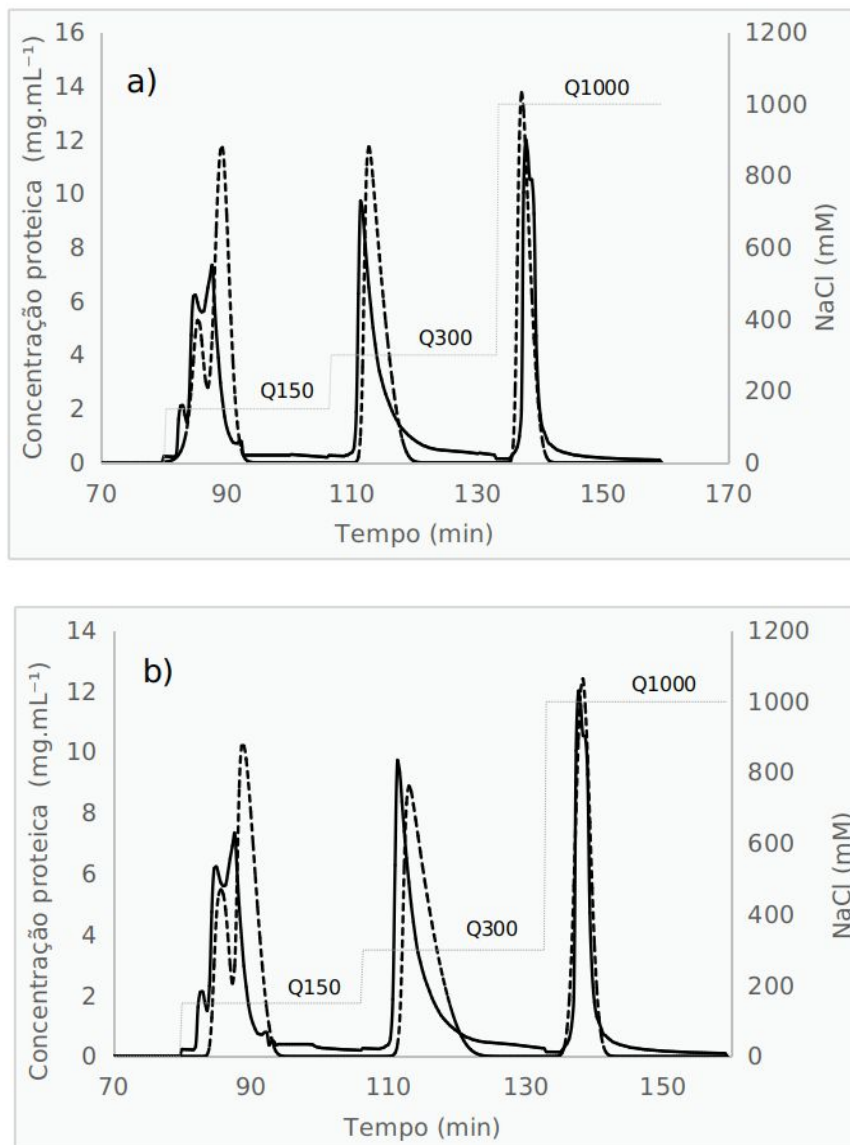


Figura 3.7. Perfis de eluição para as proteínas totais dos extratos celulares ECC<sub>B</sub> na CTA-II (validação do modelo): a) Dados experimentais (linha preta sólida), perfis simulados usando a isoterma de ação em efeito estérico (SMA) em conjunto com o modelo do equilíbrio dispersivo (MED) (linha preta tracejada) e perfis de eluição obtidos a partir de aplicação de NaCl a 150 mM, 300 mM e 1000 mM (linha cinza tracejada); b) Dados experimentais (linha preta sólida), perfis simulados usando a isoterma de Langmuir modificada (LM) em conjunto com o modelo do equilíbrio dispersivo (MED), e perfis de eluição obtidos a partir de aplicação de NaCl a 150 mM, 300 mM e 1000 mM (linha cinza tracejada).

Em todos os casos, o MED foi utilizado para descrever os balanços de massa na fase líquida, desconsiderando os fenômenos de transferência de massa no interior da resina, o que também pode ter sido um dos fatores responsáveis pelas baixas resoluções de algumas simulações. A utilização de modelos de balanço de massa mais complexos estava além do

escopo deste trabalho, uma vez que as proteínas simuladas foram obtidas por misturas complexas e não estavam disponíveis experimentalmente para permitir a determinação de parâmetros específicos de transferência de massa.<sup>5</sup>

### 3.4.3. Estimativa de parâmetros de extratos celulares virtuais, simulações e validações experimentais

Embora ambos os modelos (LM e SMA) tenham se adaptado bem aos dados experimentais, a isoterma SMA foi escolhida para esta próxima etapa do trabalho, uma vez que sua estrutura inclui os efeitos de competição entre os diferentes pools de proteínas (como pode ser observado na Equação 3.14), resultando em um modelo com poder descritivo superior, enquanto que o modelo LM apenas considera a competição entre a proteína e o sal, uma vez que as Equações empíricas 3.11 e 3.12 são usadas para modificar  $q_{\max,ie}$   $K_i$ , respectivamente, de acordo com a concentração de sal. Além disso, é importante destacar as equações da SMA foram usadas com sucesso anteriormente para descrever a cromatografia de troca iônica (FARAJI et al., 2015; KARKOV et al., 2013; HUUK et al., 2014; SHEN et al., 2005; GALLANT et al., 1995).

Portanto, os modelos SMA e MED foram usados para simular o ECC virtual, conforme descrito na Seção 3.3.5 e na Figura 3.2. Apenas o pico Q300 foi simulado, pois apresentou a maior pureza para PspA4Pro (Tabela 3.1). Os parâmetros da coluna  $\epsilon_T$  e  $N_T$  foram determinados novamente de acordo com o procedimento descrito na seção 3.2.1. Os valores encontrados foram 0,906 e 90, respectivamente (Tabela 3.2). Os parâmetros SMA para CTA-III e CTA-IV (Tabela 3.3) foram estimados para avaliar separadamente as características físico-químicas das proteínas contaminantes e PspA4Pro, respectivamente, eluídas na fração Q300. Os correspondentes perfis de eluição experimentais e simulados foram semelhantes, como mostra a Figura 3.8. O valor estimado de  $K_{Eq}$  foi maior para o pico da CTA-III em relação ao pico da CTA-IV (2,37 e 0,83, respectivamente; Tabela 3.3), sugerindo que a PspA4Pro apresentou menor afinidade com a resina quando comparada com o conjunto de impurezas proteicas eluídas no mesmo momento. Esta afirmação pode ser melhor visualizada na Figura 3.9, que apresenta separadamente o perfil de eluição de PspA4Pro e do pool de

---

<sup>5</sup> Além das isotermas apresentadas nesta seção, a isoterma de Langmuir competitiva também foi testada para descrever os dados das cromatografias CTA-I e CTA-II. Devido à menor qualidade do ajuste obtido, os dados não foram apresentados no desenvolvimento do artigo, mas podem ser verificados no apêndice D.

impurezas de um ECC simulado em uma coluna XK 26/20. Como esperado, a PspA4Pro foi eluída no início do pico Q300, antes do pool de demais proteínas, apresentando um menor tempo de residência<sup>6</sup>.

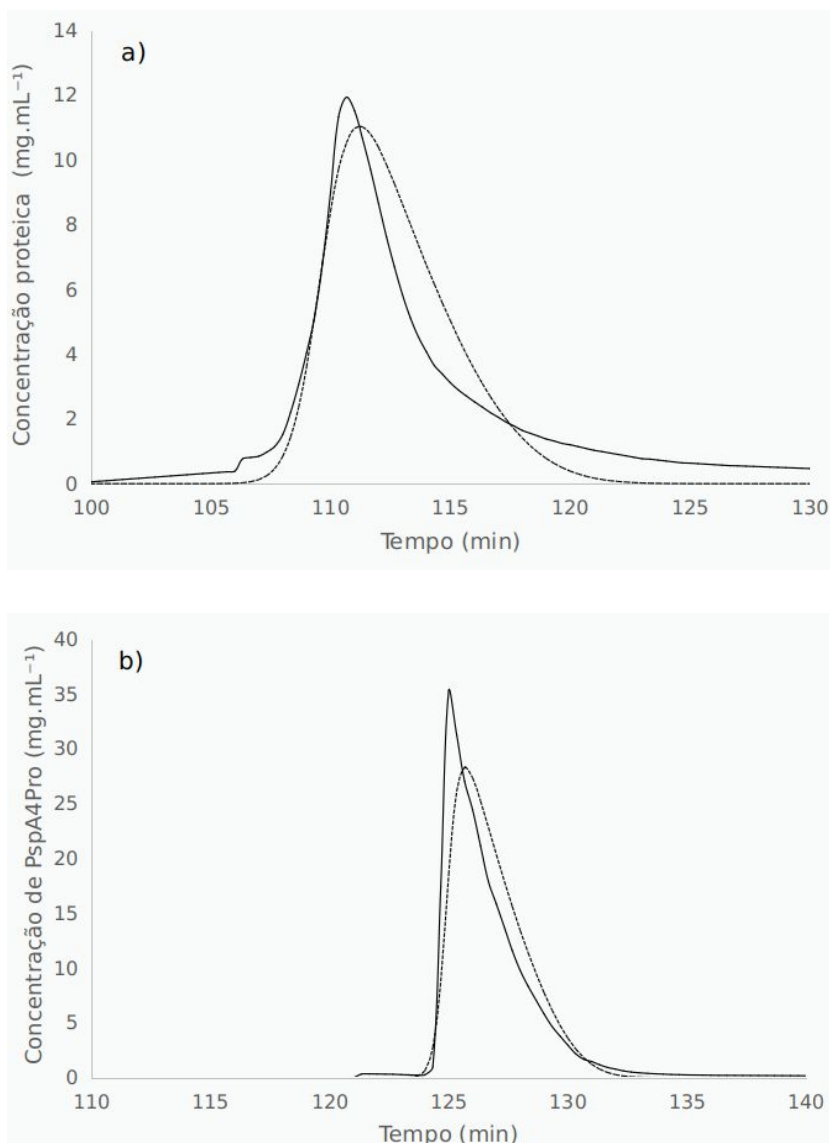


Figura 3.8. Modelo da isoterma de ação em efeito estérico (SMA) (linhas tracejadas) ajustado aos perfis de eluição experimental Q300 (linhas sólidas). a) cromatografia de troca aniônica (CTA-III) aplicada em extrato celular clarificado do cultivo C (células portadoras do vetor pET37b sem o gene de PspA4Pro); b) cromatografia de troca aniônica (CTA-IV) aplicada em solução de proteína pura (SPP).

<sup>6</sup> O código de simulação e os dados utilizados para construir a Figura 3.9 podem ser verificados no apêndice E.

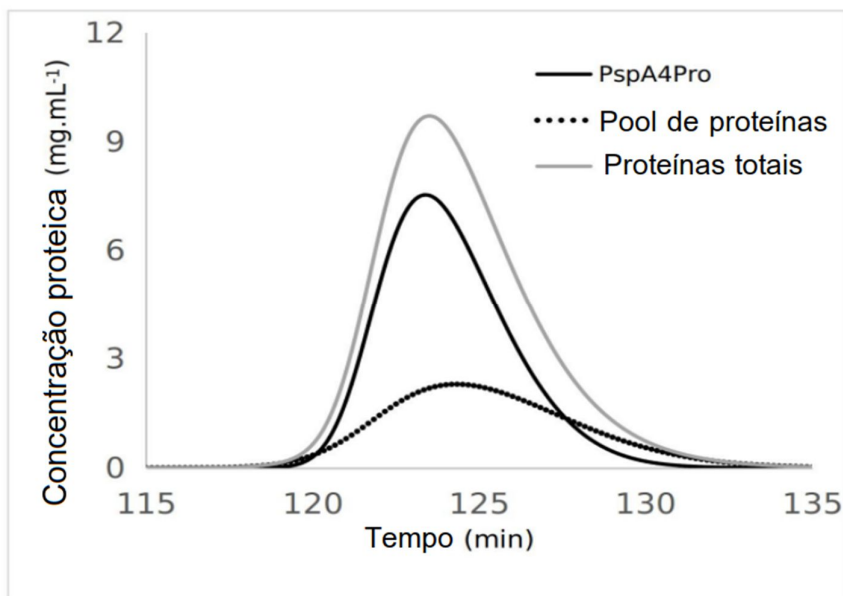


Figura 3.9. Perfis de eluição simulados para PspA4Pro, para o conjunto de impurezas proteicas e para a mistura de proteínas totais da cromatografia de troca aniônica (perfil de eluição obtido com aplicação de 300 mM de NaCl). As simulações foram realizadas utilizando a isoterma de ação em efeito estérico (SMA) em conjunto com o modelo do equilíbrio dispersivo (MED). Os parâmetros estimados estão expostos na Tabela 3.3.

A Tabela 3.4 apresenta uma comparação dos valores de MTP e pureza obtidos após a fragmentação dos picos virtuais e experimentais em seis subfrações de 0,5 VC. Os perfis de eluição experimental e simulado foram obtidos de acordo com os procedimentos descritos na Seção 3.3.5. É possível se visualizar uma clara semelhança entre os dois perfis. No entanto, alguns desvios ainda são observados nos resultados da simulação, os quais podem ser explicados pela precisão limitada dos métodos analíticos (Lowry e SDS-PAGE), bem como pelas diferenças nas características das proteínas dos materiais de partida utilizados para estimar os parâmetros das proteínas (SPP e ECC<sub>C</sub>) e validar a simulação (ECC<sub>A</sub>). Esses extratos celulares reais usados para estimar os parâmetros são misturas complexas compostas por proteínas com diferentes estruturas químicas. Assim, espera-se que as simulações não sejam tão precisas quanto em trabalhos que usaram misturas de proteínas conhecidas. Apesar disso, embora incertezas ocorram em certa medida, o esforço de modelagem e simulação de separações cromatográficas de misturas complexas é crucial para melhorar os processos reais de purificação.

Tabela 3.4. Valores simulados e experimentais da massa total de proteínas (MTP) e pureza para o perfil de eluição alternativo obtido com base no subfracionamento da eluição Q300 (CTA-V, realizado usando a coluna XK 26/20).

Sub- fragmento da eluição	VC			Pico Q300 simulado		Pico Q300 experimental	
				MTP (g)	Pureza (%)	MTP (g)	Pureza (%)
1	0,0	-	0,5	0,020	51,6	0,012	10,4
2	0,5	-	1,0	0,240	77,3	0,271	82,8
3	1,0	-	1,5	0,282	72,0	0,213	73,3
4	1,5	-	2,0	0,108	53,3	0,097	69,0
5	2,0	-	2,5	0,026	26,4	0,059	19,0
6	2,5	-	3,0	0,005	7,2	0,027	11,0

\*Dados obtidos com a divisão da fração Q300 em seis subfrações, cada uma com 0,5 volume de coluna (VC), usando os picos virtuais e experimentais.

Apesar dos desvios explicados acima, os resultados demonstraram claramente a precisão do modelo SMA e a adequação dessa metodologia de modelagem para a determinação numérica dos parâmetros. A confiabilidade dos dados simulados permitiu a identificação de modificações no processo de purificação, levando à obtenção de PspA4Pro com maior pureza, mas com menos MTP. De acordo com a simulação, a segunda subfração (0,5-1,0 VC) apresentou 35,2% da MTP eluída, com pureza 25,1% maior que a obtida para a fração Q300 da CTA-I (Tabela 3.1). Um comportamento semelhante foi observado nos dados experimentais para essa mesma subfração, que apresentou 39,9% da MTP eluída, com pureza 34,0% superior à da fração Q300 da CTA-I. Portanto, a abordagem proposta neste trabalho pode ser usada para avaliar os perfis de uma proteína de interesse durante a cromatografia, bem como para desenvolver modificações no processo, a fim de aumentar a pureza.

### 3.5. Conclusões

A abordagem descrita neste trabalho representa uma importante contribuição para a modelagem e otimização da cromatografia de troca iônica. Foi aplicada na purificação de uma proteína recombinante de pneumococo (PspA4Pro), a qual se encontra presente em extratos celulares clarificados reais obtidos a partir de biomassa de *E. coli*. Apesar das diferenças observadas entre os extratos celulares clarificados utilizados para estimativa e validação de parâmetros, os resultados da simulação mostraram que tanto a isoterma de ação em efeito estérico quanto a de Langmuir modificada foram capazes de representar os cromatogramas de

misturas complexas de proteínas. As simulações do modelo também foram capazes de prever a pureza da PspA4Pro em diferentes regiões das eluições cromatográficas, levando à identificação de uma subfração de eluição que proporcionou um aumento de 34,0% na pureza, em comparação com a fração completa obtida usando o protocolo original. É importante destacar que a aplicação do procedimento proposto não se restringe a um estudo de caso da PspA4Pro, já que poderia ser estendido para descrever a purificação de qualquer extrato celular por cromatografia de troca iônica, uma vez que os parâmetros do modelo poderiam ser estimados e atualizados de acordo com as características de outras misturas complexas de proteínas.

#### **4. OTIMIZAÇÃO DA CAPACIDADE DE PROCESSAMENTO DA CROMATOGRAFIA DE TROCA ANIÔNICA NA PURIFICAÇÃO DE PROTEÍNAS DE SUPERFÍCIE DE PNEUMOCOCO**

A proposta deste capítulo foi testar um pouco mais a capacidade preditiva dos modelos explorados no capítulo anterior. O modelo do equilíbrio dispersivo e a isoterma de ação em efeito estérico mais uma vez foram utilizados para simular o sistema cromatográfico. No entanto, foram realizadas variações nas forças iônicas das eluições e nas massas proteicas inseridas no sistema para avaliar a capacidade de previsão das simulações através de uma metodologia de validação cruzada. O modelo simulado foi utilizado para prever a capacidade de processamento da resina, o que possibilitou estimar a quantidade máxima de massa proteica que poderia ser processada sem perda de eficiência. A proposta se provou adequada tanto para o ajuste quanto para a validação das situações descritas. As simulações demonstraram que a coluna cromatográfica poderia processar até duas vezes mais proteína sem perda de qualidade da PspA4Pro eluída, o que também foi confirmado experimentalmente.

### ***Resumo***

O uso de modelos matemáticos para descrever a cromatografia de troca iônica é uma boa estratégia para desenvolver processos de purificação de proteínas recombinantes eficientes e otimizados. A principal limitação dessa abordagem está relacionada ao uso de misturas de proteínas puras como estratégia de extrair informações para caracterizar o processo, afastando as simulações de situações reais. Neste trabalho, dados cromatográficos de uma mistura não sintética de proteínas foram utilizados para a descrição matemática da cromatografia de troca aniônica na purificação de PspA4Pro, uma proteína produzida por *Escherichia coli* recombinante utilizada na fabricação de vacinas antipneumocócicas. O modelo do equilíbrio dispersivo foi utilizado em conjunto com a isoterma de ação em efeito estérico para simular todo o processo no EMSO (*Environment for Modeling, Simulation and Optimization*), um software orientado a equações com extensa capacidade de modelagem. Diferentes estratégias de operação foram testadas, envolvendo modificações nas forças iônicas das eluições e aumentos na carga de proteínas na entrada da coluna. Os parâmetros foram bem ajustados às situações descritas. As simulações demonstraram que a mesma eficiência e recuperação de purificação poderiam ser obtidas para uma massa de entrada de proteína duas vezes maior que a original, aumentando o rendimento específico por volume de coluna da purificação. Essa condição foi avaliada experimentalmente, fornecendo 0,022 g.mL<sup>-1</sup> de resina de PspA4Pro com 81,2% de pureza em comparação com 0,010 g.mL<sup>-1</sup> de resina de PspA4Pro com 73,9% de pureza para a menor carga alimentada. Esses resultados confirmam a eficiência da metodologia de modelagem e simulação descrita e reforçam a importância de sua aplicação para identificar condições de operação mais promissoras.

#### 4.1. Introdução

A cromatografia é um método de separação físico-química no qual uma fase móvel passa por uma fase estacionária e alguns componentes são seletivamente retidos na matriz sólida por adsorção, permitindo uma migração diferencial devido ao aumento da força iônica (GU, 2015). Os processos cromatográficos têm sido amplamente utilizados como etapas fundamentais na purificação de proteínas recombinantes terapêuticas. Nesses casos, o principal desafio é obter um produto com alta pureza, reduzindo as impurezas como ácidos nucleicos, pirogênicos, enzimas e proteínas contaminantes até os limites estabelecidos pelas agências reguladoras (MAMAT et al., 2015; FIGUEIREDO et al., 2017). Na maioria das situações, as etapas de purificação contribuem com até 80% do custo total do processo (SOFER et al., 1997), uma vez que são compostas por equipamentos de alta tecnologia e o custo da resina é muito alto (GU, 2015). Nesse contexto, a modelagem e simulação são ferramentas viáveis para reduzir o custo do processo e otimizá-lo.

Muitos estudos já abordaram o desenvolvimento de modelos matemáticos com o objetivo de prever o comportamento dos cromatogramas e aumentar a eficiência da purificação (KACZMARSKI et al., 2001; DONAT et al., 2018; CLOSE et al., 2014; TEOH et al., 2001; CHAN et al., 2008; ORELLANA et al., 2009; FARAJI et al., 2015; LIENQUEO et al., 2012; ORELLANA et al., 2009). Nos últimos anos, os avanços computacionais têm permitido a simulação de diferentes modelos que levam em consideração a cinética de adsorção, a transferência de massa, a dispersão axial e a difusividade na matriz sólida. A disponibilidade de tais modelos é muito adequada para fins de engenharia, pois permite estudos de otimização e escalonamento baseados em simulações de modelos (LIENQUEO et al., 2012). No entanto, a principal limitação dessa abordagem está relacionada ao uso de misturas sintéticas de proteínas puras como forma de extrair informações para identificação de parâmetros dos modelos. Se por um lado essa estratégia viabiliza a execução de experimentos em condições controladas, as quais são essenciais para a estimativa precisa dos parâmetros e adequação do modelo, por outro lado há pouca aplicabilidade observada em situações reais, uma vez que as proteínas recombinantes produzidas por uma célula hospedeira são misturadas com impurezas, incluindo proteínas contaminantes desconhecidas. Este pool de proteínas está diretamente relacionado às estratégias de cultivo utilizadas, o que torna a modelagem de processos de separação de misturas reais um desafio muito mais complexo.

Neste trabalho, dados cromatográficos de misturas não sintéticas foram modelados pelo modelo do equilíbrio dispersivo e pela isoterma de ação em efeito estérico para simular a cromatografia de troca aniônica na purificação de PspA4Pro, uma proteína A de superfície de pneumococo produzida por *Escherichia coli* recombinante e utilizada para fabricar vacinas contra *S. pneumoniae* (SENKOVICH et al., 2007; PERCIANI et al., 2013; MIYAJI et al., 2013; FIGUEIREDO et al., 2017). Uma metodologia de validação cruzada foi utilizada para estimar os parâmetros. As simulações foram realizadas no EMSO (*Environment for Modeling, Simulation and Optimization*), um software orientado a equações com capacidade de modelagem (SOARES et al., 2003; THOMBRE et al., 2015; BANERJEE et al., 2010; FURLAN, 2016) Espera-se que esse tipo de abordagem aproxime o processo da realidade e permita a otimização de aspectos relacionados à capacidade de processamento da resina, aumentando a eficiência da separação e reduzindo o custo de produção.

## 4.2. Materiais e métodos

### 4.2.1. Produção de PspA4Pro

A produção de PspA4Pro foi realizada em cultivos em batelada de *Escherichia coli* BL21 (DE3) contendo o plasmídeo pET37b (+) com um fragmento do gene da proteína de superfície de pneumococo A (PspA) de *Streptococcus pneumoniae* (MORENO et al., 2010). A expressão do gene PspA foi controlada pelos promotores lacUV5 e T7lac, os quais foram induzidos por meios suplementados com lactose ou isopropil- $\beta$ -D-tiogalactopiranosídeo (IPTG), fazendo com que a proteína recombinante se acumulasse no interior da célula. O clone utilizado nos cultivos foi gentilmente fornecido pela Dra. Eliane Miyaji do Laboratório de Biologia Molecular do Instituto Butantan (São Paulo, Brasil).

O cultivo foi realizado em um biorreator de 5 L monitorado e controlado pelo software SuperSys\_HCDC (CAMPANI et al., 2016)<sup>7</sup>. O meio utilizado para o cultivo era definido, apresentava glicerol como principal fonte de carbono e foi induzido por 20 g.L<sup>-1</sup> de lactose. Também foram aplicados três pulsos extras de glicerol e lactose durante a execução do cultivo, cujo tempo total foi de 35 horas, incluindo 17,5 horas de indução. O NH<sub>4</sub>OH (30%) foi utilizado para manter o pH em 6,7. A temperatura de operação e a concentração de oxigênio dissolvido foram de 27 °C e 30%, respectivamente. A transferência de massa de

---

<sup>7</sup> As células utilizadas em todos os procedimentos deste capítulo são as mesmas do cultivo A, capítulo 3.

oxigênio foi mantida por um agitador conectado a um controlador responsável por alterar automaticamente a velocidade do rotor entre 200 e 900 rpm. A vazão total de gás foi mantida em  $4 \text{ L}\cdot\text{min}^{-1}$  (SILVA, 2015). Após o cultivo, a biomassa foi separada do meio por 30 minutos de centrifugação, a  $4 \text{ }^\circ\text{C}$  e 6700 g. As células precipitadas foram armazenadas a  $-80 \text{ }^\circ\text{C}$ . Antes da realização das cromatografias de troca aniônica, as células foram descongeladas e submetidas a lise e clarificação com 0,1% do detergente brometo de cetiltrimetilamônio (CTAB), produzindo o extrato celular clarificado (ECC). Esses procedimentos foram descritos em detalhes por FIGUEIREDO et al., 2017.

#### 4.2.2. Cromatografia de troca aniônica

A cromatografia de troca aniônica foi realizada usando 70 mL de resina Q-Sepharose FF empacotada em uma coluna XK 26/20, a qual se encontrava montada no sistema cromatográfico Äkta Avant 15. A vazão linear utilizada foi de  $2,50 \text{ mL}\cdot\text{min}^{-1}\cdot\text{cm}^{-2}$  ( $13,5 \text{ cm}^3\cdot\text{min}^{-1}$ ). A coluna foi equilibrada com 5 volumes de coluna (VC) de tampão de fosfato de sódio 10 mM, pH 6,5 e condutividade  $1,2 \text{ mS}\cdot\text{cm}^{-1}$ . O pH e a condutividade do ECC (Seção 4.2.1) foram ajustados para esses mesmos valores. O sistema cromatográfico representado na Figura 4.1 descreve de forma adequada o processo de separação e foi usado como base para a configuração do modelo na plataforma EMSO. Inicialmente, as válvulas  $V_1$  e  $V_5$  foram abertas e a amostra de ECC contendo PspA4Pro foi inserida na coluna. Durante esta etapa, a fração não adsorvida (QNADs) com baixa pureza foi armazenada no recipiente 1 e as proteínas com afinidade pela resina foram adsorvidas. Após inserir a amostra, também foi aplicado um volume de 5 VC de tampão fosfato de sódio para remover proteínas residuais não adsorvidas. Após este procedimento, as válvulas  $V_1$  e  $V_5$  foram fechadas. Nas próximas etapas, iniciou-se o procedimento de dessorção: As proteínas foram dessorvidas através da aplicação de 5 VC de eluições de NaCl com forças iônicas fraca ( $Q_1$ ), média ( $Q_2$ ) e forte ( $Q_3$ ), nesta ordem, com abertura simultânea das válvulas  $V_2$  e  $V_6$ ,  $V_3$  e  $V_7$  e  $V_4$  e  $V_8$ , respectivamente. Proteínas com baixa, média e alta forças de ligação à resina foram respectivamente dessorvidas nas frações de eluição  $Q_1$ ,  $Q_2$  e  $Q_3$ . A maior parte da PspA4Pro foi eluída na fração  $Q_2$  (FIGUEIREDO et al., 2017). A Tabela 4.1 descreve os experimentos A a E realizados para testar diferentes cargas proteicas de processamento do ECC associadas a modificações de forças iônicas nos perfis de eluição da cromatografia de troca aniônica. Dados experimentais relacionados ao perfil de concentração de proteínas no cromatograma

foram fornecidos pelo detector UV 280 nM localizado no final da coluna. Dados de condutividade relacionados à concentração de NaCl também foram fornecidos. O algoritmo de Ramer-Douglas-Peucker foi utilizado para reduzir a quantidade de pontos fornecidos pelo software de cromatograma compondo uma curva semelhante com menos pontos (PRASAD et al., 2012).

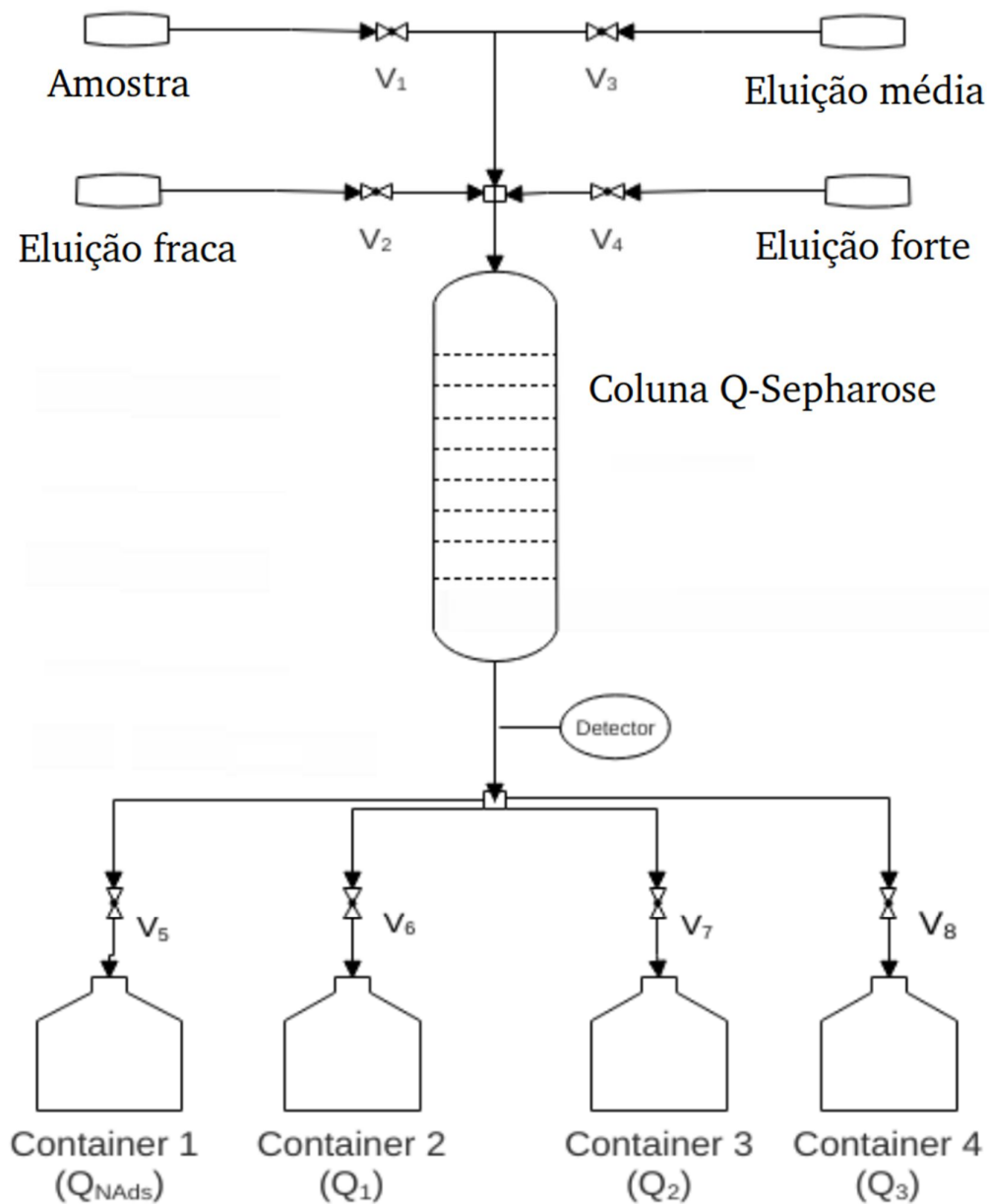


Figura 4.1. Descrição da cromatografia de troca aniônica. A amostra QNAds contém a fração não adsorvida. As amostras Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> contêm as eluições fraca, média e forte, respectivamente.

Tabela 4.1. Características do extrato celular clarificado e perfis de eluição dos experimentos realizados em uma coluna de 70 mL com resina Q-Sepharose.

<b>Experimento</b>		<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
<b>Extrato celular</b>	Volume de amostra $V_{in}$ (mL)	760	920	1000	950	2080
<b>clarificado</b>	Carga proteica ( $\text{g.mL}_{\text{coluna}}^{-1}$ )	0,034	0,061	0,034	0,031	0,080
	Concentração de NaCl em $Q_1$ (mM)	150	150	120	180	150
<b>Perfis de eluição</b>	Concentração de NaCl em $Q_2$ (mM)	300	300	250	350	300
	Concentração de NaCl em $Q_3$ (mM)	1000	1000	900	1000	1000

#### 4.2.3. Quantificação de proteínas, pureza de PspA4Pro e determinação da recuperação

As concentrações totais de proteína foram determinadas nos containers 1 a 4 (Figura 4.1) para os experimentos A e E pela metodologia Lowry (LOWRY et al., 1951). As massas totais de proteína ( $M_i$ ) foram obtidas já que os volumes de todas as frações "i" (ECC, QNAds,  $Q_1$ ,  $Q_2$  e  $Q_3$ ) foram medidos durante os experimentos. As purezas de PspA4Pro ( $P_i$ ) para todas as frações "i" foram determinadas por densitometria de bandas em SDS-PAGE 12% com 2-mercaptoetanol (LAEMMI, 1970) e calculadas como a porcentagem da banda de PspA4Pro em relação à soma das bandas das impurezas proteicas em cada amostra (FIGUEIREDO et al., 2017). As massas de PspA4Pro ( $M_{PspA4Pro,i}$ ) foram calculadas de acordo com a equação 4.1. As recuperações de PspA4Pro ( $R_i$ ) foram calculadas para as frações QNAds,  $Q_1$ ,  $Q_2$  e  $Q_3$  de todos os experimentos, de acordo com a equação 4.2, onde  $M_{PspA4Pro,ECC}$  é a massa total de PspA4Pro no extrato celular clarificado.

$$M_{PspA4Pro,i} = M_i \cdot P_i \quad (4.1)$$

$$R_i = \frac{M_{PspA4Pro,i}}{M_{PspA4Pro,ECC}} \quad (4.2)$$

#### 4.2.4. Determinação de parâmetros experimentais

O número de pratos teóricos ( $N_t$ ) e a porosidade total da coluna ( $\epsilon_T$ ) foram determinados por um pulso de 0,5 mL de solução de acetona a 1% (v.v<sup>-1</sup>) diluída em água ultrapura a uma vazão de 2,71 cm<sup>3</sup>.min<sup>-1</sup>.  $N_t$  foi calculado pela aplicação da equação 4.3 no pico obtido, onde  $t_o$  é o tempo de retenção da coluna e  $\Delta$  é a largura do pico (Amersham Biosciences, 2002).  $\epsilon_T$  foi definido pela razão entre o volume vazio ( $V_o$ ) e o volume da coluna ( $V_C$ ), e  $V_o$  foi calculado pelo tempo de retenção de acetona ( $t_o$ ) e pela vazão da fase móvel ( $F$ ), conforme demonstrado pelas equações 4.4 e 4.5 (CLOSE et al., 2014).

$$N_t = 5.54 \left( \frac{t_o}{\Delta} \right)^2 \quad (4.3)$$

$$\epsilon_T = \frac{V_o}{V_C} \quad (4.4)$$

$$V_o = t_o \cdot F \quad (4.5)$$

#### 4.2.5. Estratégias de simulação do processo

As simulações foram realizadas pelo software EMSO<sup>TM</sup> (*Environment for Modeling Simulation and Optimization*), um simulador orientado a equações com linguagem interna de modelagem orientada a objetos que permite a criação de novos modelos de acordo com a demanda do pesquisador (SOARES e SECCHI, 2003). Experiências anteriores identificaram um total de quatro picos no cromatograma (FIGUEIREDO et al., 2017), dois na eluição  $Q_1$  e dois nas eluições  $Q_2$  e  $Q_3$ . Neste trabalho, para que a simulação fosse possível, cada pico foi associado a um pool de proteínas "i". Esses pools foram denominados  $Q_1$ -I,  $Q_1$ -II,  $Q_2$  e  $Q_3$  e foram compostos por um grupo de proteínas com características químicas semelhantes, pois foram eluídas juntas (CHAN et al., 2008; BENEDINI et al., 2019). O sistema cromatográfico proposto pela Figura 4.1 foi descrito no simulador. O modelo do equilíbrio dispersivo (MED, Equações 4.6 e 4.7) e a isoterma de ação em efeito estérico (SMA, Equações 4.8 e 4.9) foram aplicados em cada prato teórico da coluna virtual para obter um sistema de equações diferenciais solucionado pelas rotinas numéricas do software.

O MED representa o balanço de massa em uma coluna de adsorção, onde  $c_i$  é a concentração de proteína “i” na fase móvel ( $\text{mol.cm}^{-3}$ ),  $q_i$  é a quantidade de adsorbato retido na estrutura sólida ( $\text{mol.cm}_{\text{resina}}^{-3}$ ),  $u$  é a velocidade da fase móvel ( $\text{cm.s}^{-1}$ ) e  $D_a$  é a influência da dispersão axial ( $\text{cm}^2.\text{s}^{-1}$ ). A aplicação desta equação considera adsorção uniforme em toda a coluna, ausência de gradiente de concentração radial e resistências à transferência de massa durante a adsorção (TEOH et al., 2001; KACZMARSKI et al., 2001). A dispersão axial  $D_a$  pode ser calculada pela equação 4.7, onde  $L$  é o comprimento da coluna e  $N_t$  é o número total de pratos teóricos (CLOSE et al., 2014). Como o EMSO não permite o desenvolvimento de diferenciais relacionados a variáveis espaciais, os diferenciais espaciais foram aproximados pelo método das diferenças finitas, com  $N_t$  elementos.

$$\frac{\partial c_i}{\partial t} + \frac{(1 - \epsilon_T)}{\epsilon_T} \frac{\partial q_i}{\partial t} + u \frac{\partial c_i}{\partial z} = D_a \frac{\partial^2 c_i}{\partial z^2} \quad (4.6)$$

$$D_a = \frac{Lu}{2N_t} \quad (4.7)$$

A isoterma SMA tem sido usada como estratégia para representar a competição entre proteínas e soluções iônicas na cromatografia de troca aniônica (GU, 2015).

$$\frac{\partial q_i}{\partial t} = K_{\text{Ads},i} c_i Q^{v_i} - K_{\text{Des},i} q_i c_{\text{salt}}^{v_i} \quad (4.8)$$

$$Q = \wedge - \sum_{i=1}^N q_i (v_i + \alpha_i) \quad (4.9)$$

A Equação 4.8 descreve que a quantidade de proteína “i” adsorvida depende das constantes de adsorção ( $K_{\text{Ads},i}$ ) e dessorção ( $K_{\text{Des},i}$ ) ( $\text{cm}^3.\text{mol}^{-1}.\text{s}^{-1}$ ), do número de sítios ativos ( $v_i$ ), da concentração de sal na fase líquida ( $c_{\text{salt}}$ ) e da carga livre resultante  $Q$  da resina ( $\text{mol.cm}^{-3}$ ). A equação 4.9 mostra que  $Q$  pode ser descrito pela capacidade iônica da resina  $\wedge$  ( $\text{mol.cm}^{-3}$ ) que representa o número total de sítios ativos disponíveis para ligação, e pela soma dos sítios ativos ocupados, nos quais o fator estérico de cada proteína adsorvida ( $\alpha_i$ ) representa o bloqueio molecular sobre a resina (SHEN et al., 2005; HUUK et al., 2014; KARKOV et al., 2013).

Para que a simulação fosse possível, cada pool de proteínas "i" descrito acima foi associado aos parâmetros SMA correspondentes  $K_{Ads,i}$ ,  $K_{Des,i}$ ,  $v_i$  e  $\alpha_i$ . Além disso as concentrações de proteína ( $c_{ie}$ ) e  $c_{NaCl}$  na entrada da coluna também foram informadas, uma vez que as massas totais de proteína para cada pool eram conhecidas (método Lowry), assim como o volume total de ECC. Como os grupos proteicos Q<sub>1</sub>-I e Q<sub>1</sub>-II foram eluídos na mesma fração, a massa particular de cada um foi estimada pela área dos picos cromatográficos. Os valores de absorvância a 280 nM foram transformados em concentrações de proteína através de um coeficiente de extinção molar geral estabelecido para cada fração eluída (Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub>), igualando a massa total de proteínas de cada pico cromatográfico ao valor determinado pelo Método Lowry (LOWRY, 1951). Os valores de condutividade foram transformados em concentrações de NaCl usando uma curva padrão construída com diferentes concentrações de NaCl e suas correspondentes condutividades<sup>8</sup>.

#### 4.2.6. Estimativas de parâmetros e validações experimentais

O diagrama apresentado na Figura 4.2 mostra a estratégia adotada para a estimativa de parâmetros (NELLES, 2001). A primeira caixa da figura explica que os pools de proteínas foram identificados, como já foi descrito na seção anterior. A segunda caixa mostra que os valores iniciais dos parâmetros foram determinados em uma primeira simulação na qual o efeito competitivo entre os conjuntos de proteínas identificados não foi considerado e os parâmetros foram ajustados aos dados de um único experimento, para cada pico em particular. Essa etapa foi necessária para identificar as regiões mais apropriadas para a estimativa de parâmetros, e esses resultados não serão apresentados neste trabalho. Os valores de  $K_{Ads,i}$ ,  $K_{Des,i}$ ,  $v_i$ ,  $\alpha_i$  estimados nesta etapa foram utilizados como suposições iniciais para as simulações I, II, III e IV, as quais foram realizadas para os experimentos A a D através da estratégia de validação cruzada descrita na região pontilhada da Figura 4.2: para a Simulação I, por exemplo, os valores de  $K_{Ads,i}$ ,  $K_{Des,i}$ ,  $v_i$  e  $\alpha_i$  foram estimados pelos pools de proteínas observados nos experimentos B, C e D, enquanto que a validação foi realizada com os dados do experimento A. Estratégias análogas foram utilizadas nas simulações II, III e IV. Em todos os casos, o parâmetro  $\hat{\phantom{x}}$  foi fornecido por HUUK et al. (2014). Para cada simulação, as raízes do erros quadráticos médios das estimativas (ou “*root mean square errors*”,  $RMSE_{Est}$ ,

<sup>8</sup> O Apêndice B explica de forma detalhada as estratégias utilizadas para disponibilizar os dados de condutividade e absorvância (280 nM) em forma de concentrações de NaCl e proteínas.

Equação 4.10) foram usadas para avaliar a qualidade dos ajustes de dados quando os parâmetros foram estimados, onde  $N$  é o número total de pontos experimentais,  $\sigma$  é o número de parâmetros estimados,  $y_i$  são os dados experimentais e  $\hat{y}_i$  são os dados obtidos pela modelagem. A equação 4.10 também foi usada para avaliar a qualidade das validações ( $RMSE_{val}$ ), com  $\sigma = 0$ . Após a conclusão dos procedimentos da Figura 4.2, todos os experimentos que forneceram bons resultados de validação foram utilizados novamente em uma nova simulação (Simulação V, não mostrada na Figura 4.2), na qual todos os parâmetros foram ajustados novamente aos dados experimentais, fornecendo suas estimativas finais

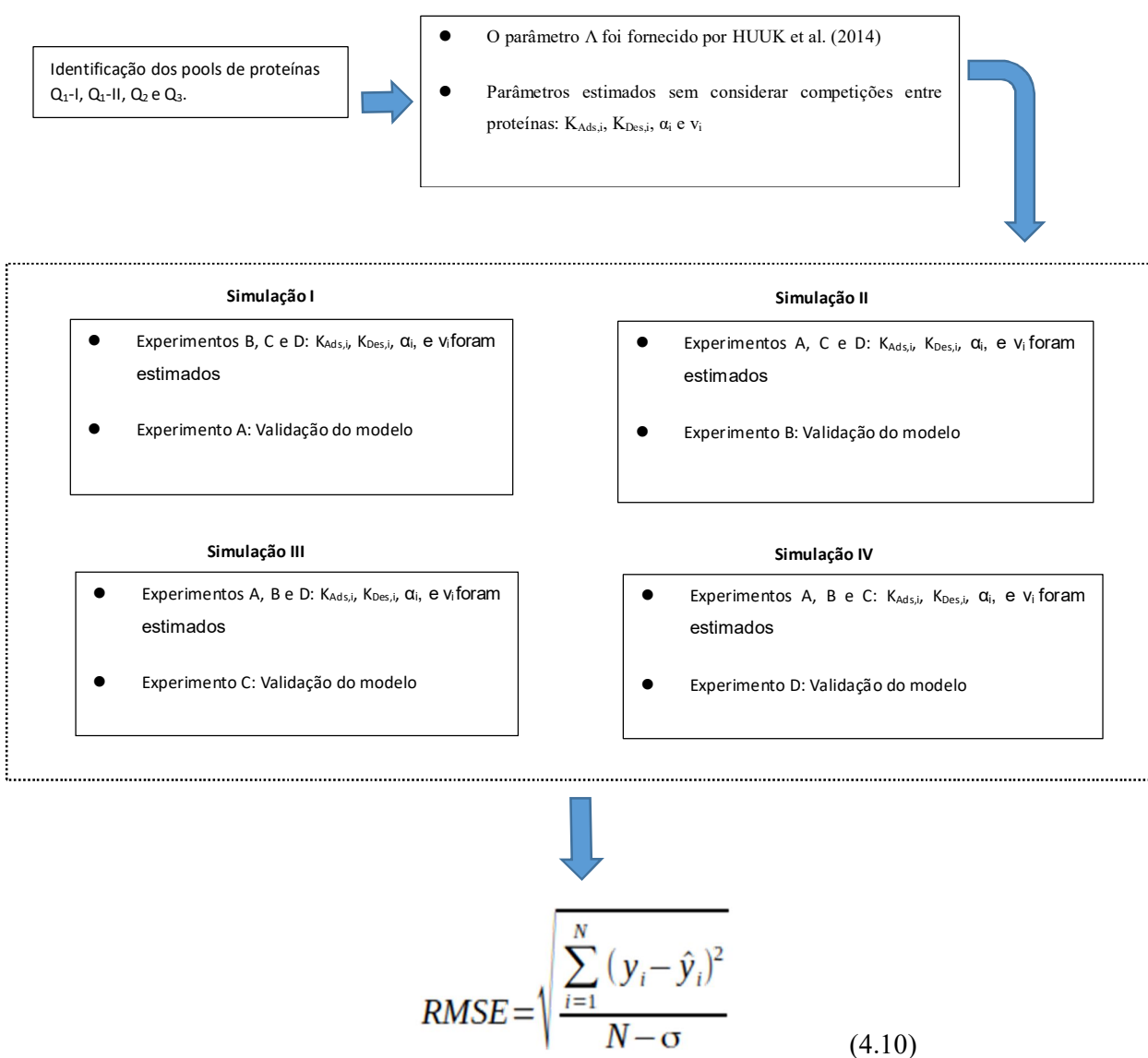


Figura 4.2. Procedimento numérico utilizado para estimar os parâmetros e simular o sistema descrito pela Figura 4.1. Foram utilizados dados dos experimentos A, B, C e D.

#### 4.2.7. Determinação da capacidade de processamento da resina

Após a estimativa dos parâmetros, conforme descrito na sessão anterior, simularam-se amostras virtuais do ECC, as quais apresentaram aumento gradual das massas dos grupos de proteínas (Q<sub>1-I</sub>, Q<sub>1-II</sub>, Q<sub>2</sub> e Q<sub>3</sub>). A soma das massas de proteínas do experimento A (M<sub>A</sub>) foi usada como referência, cujos valores eram conhecidos devido aos procedimentos analíticos, portanto  $M_A = M_{Q1-I} + M_{Q1-II} + M_{Q2} + M_{Q3}$ . Foram realizadas três simulações com massas de entrada de 1 x M<sub>A</sub>, 2 x M<sub>A</sub> e 3 x M<sub>A</sub>, que permitiram estudar o perfil de adsorção de proteínas no interior da coluna nas seguintes condições: Imediatamente após a aplicação do ECC; após 5 VC da aplicação de eluição Q<sub>1</sub>, com 150 mM de NaCl; e após 5 VC da aplicação de eluição Q<sub>2</sub>, com 300 mM de NaCl. O perfil de distribuição de proteínas obtido no interior da coluna nas condições descritas permitiu a realização de uma análise crítica para determinar a quantidade máxima de proteínas do ECC que a coluna poderia processar sem perder a eficiência. Após a determinação da capacidade máxima de processamento da resina, o experimento E foi realizado para validar as condições simuladas.

### 4.3. Resultados e discussão

#### 4.3.1. Visão geral das cromatografias de troca aniônica

O extrato celular clarificado (ECC) foi obtido de acordo com os procedimentos descritos pela seção 4.2.1. A seguir, foram realizadas as cromatografias de troca aniônica A e E e as amostras foram analisadas de acordo com os procedimentos descritos nas Seções 4.2.2 e 4.2.3. A Tabela 4.2 mostra as características das principais frações (ECC, QNAds, Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub>) de cada experimento. Quando os desvios padrão são considerados, os balanços de massa de proteínas totais se mostram viáveis para todos os experimentos ( $M_{Q1} + M_{Q2} + M_{Q3} + M_{QNAds} \leq M_{ECC}$ ). O mesmo ocorre para o balanço de massa de PspA4Pro ( $P_{Q1} \cdot M_{Q1} + P_{Q2} \cdot M_{Q2} + P_{Q3} \cdot M_{Q3} + P_{QNAds} \cdot M_{QNAds} \leq P_{ECC} \cdot M_{ECC}$ ).

De acordo com a Tabela 4.1, os experimentos A e B foram realizados com os mesmos perfis de eluição, mas a massa de entrada de ECC do experimento B foi duas vezes maior. Apesar dessa diferença, os resultados da Tabela 4.2 mostram que quantidades semelhantes de proteína foram eluídas nas frações Q<sub>1</sub> e Q<sub>2</sub>. Uma quantidade maior de proteínas estava presente nas frações QNAds do experimento B, explicando o destino da maior massa de entrada neste caso. Quando apenas as frações Q<sub>2</sub> são comparadas, a

recuperação de PspA4Pro e a pureza do experimento A superam B (62,4% e 73,9% contra 44,4% e 57,3%, respectivamente). Isso ocorreu devido a desvios nos perfis das frações ECC. Apesar de terem sido obtidas pela mesma metodologia (Seção 4.2.2), algumas variações não identificadas ocorreram durante a produção de ECC<sub>B</sub>, o que pode ser verificado quando a pureza dessa fração é comparada com a pureza média de outras frações de ECC (20,4% contra 30,5% - 33,3%, respectivamente), mostrando que a falta de qualidade de ECC<sub>B</sub> interferiu no desempenho cromatográfico do experimento B.

Quando os experimentos A, C e D são comparados, é possível visualizar os efeitos das variações nas forças iônicas das eluições no desempenho das cromatografias (consulte a Tabela 4.1 para verificar as condições experimentais), uma vez que em todos os casos as purezas e massas da ECC foram semelhantes. O experimento C apresentou as maiores massas de proteína nas eluições Q<sub>2</sub> e Q<sub>3</sub>. Isso pode ser explicado pela menor força iônica aplicada na eluição Q<sub>1</sub> neste caso, a qual não removeu de forma apropriada as proteínas fracamente ligadas à coluna, redirecionando parte delas principalmente para Q<sub>2</sub> e reduzindo a pureza dessa fração quando comparada aos experimentos A e D (45,3% contra 68,9% e 73,9%, respectivamente). Quando apenas os experimentos A e D são analisados, fica claro que em ambos os casos as eluições Q<sub>1</sub> foram eficientes para remover as impurezas fracamente ligadas, uma vez que as purezas e recuperações de PspA4Pro em Q<sub>2</sub> são mais altas. No entanto, a fração Q<sub>1</sub> do experimento D apresentou uma recuperação maior de PspA4Pro do que o experimento A (5,4% contra 1,9%, respectivamente), mostrando que a aplicação de uma eluição Q<sub>1</sub> com 180 mM de NaCl é eficiente na remoção de impurezas, mas também acarreta em uma maior perda de PspA4Pro.

Tabela 4.2. Massa ( $M_i$ ), pureza ( $P_i$ ) e recuperações de PspA4Pro ( $R_i$ ) nos extratos celulares clarificados (ECC) e frações QNADs,  $Q_1$ ,  $Q_2$  e  $Q_3$ . As amostras foram obtidas a partir da realização dos experimentos A a E para purificar PspA4Pro com cromatografia de troca aniônica.

	<b>Experimento</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
<b>ECC</b>	$M_{ECC}$ (g)	2,42 ± 0,32	4,30 ± 0,38	2,40 ± 0,32	2,20 ± 0,30	5,62 ± 0,34
	$P_{CCE}$	33,3%	20,4%	30,5%	30,5%	33,3%
<b>Q<sub>1</sub></b>	$M_1$ (g)	0,26 ± 0,02	0,24 ± 0,05	0,30 ± 0,01	0,32 ± 0,07	1,49 ± 0,14
	$R_1$	1,9% ± 0,4%	1,3% ± 0,3%	2,8% ± 0,5%	5,4% ± 1,9%	2,9% ± 0,6%
	$P_1$	5,2%	4,8%	6,9%	11,4%	3,7%
<b>Q<sub>2</sub></b>	$M_2$ (g)	0,68 ± 0,04	0,68 ± 0,06	0,85 ± 0,05	0,60 ± 0,03	1,54 ± 0,15
	$R_2$	62,4% ± 11,9%	44,4% ± 5,8%	52,6% ± 10,2%	61,6% ± 11,3%	66,8% ± 13,5%
	$P_2$	73,9%	57,3%	45,3%	68,9%	81,2%
<b>Q<sub>3</sub></b>	$M_3$ (g)	0,08 ± 0,02	0,17 ± 0,02	0,18 ± 0,01	0,02 ± 0,01	0,19 ± 0,02
	$R_3$	0,5% ± 0,2%	4,6% ± 0,5%	3,3% ± 0,6%	0,4% ± 0,2%	0,7% ± 0,1%
	$P_3$	5,5%	24,0%	13,3%	12,4%	6,5%
<b>Q<sub>NADs</sub></b>	$M_{QNADs}$ (g)	1,43 ± 0,08	2,95 ± 0,32	1,90 ± 0,47	0,40 ± 0,20	3,08 ± 1,12
	$R_{QNADs}$	13,0% ± 2,4%	18,8% ± 2,0%	4,3% ± 1,8%	7,5% ± 0,0%	12,0% ± 5,6%
	$P_{QNADs}$	7,3%	5,6%	16,5%	12,6%	7,3%

### 4.3.2. Estimativas de parâmetros, simulações e validações experimentais das cromatografias

Esta seção apresenta os resultados da estratégia de validação cruzada usada para estimar os parâmetros e simular as cromatografias de troca aniônica. Os procedimentos de modelagem e simulação foram realizados conforme descrito na Seção 4.2.6 e Figura 4.2. A Tabela 4.3 apresenta as condições iniciais e os parâmetros fixos da coluna, onde  $\epsilon_T$  e  $N_T$  foram determinados conforme descrito pela Seção 4.2.4;  $L$  e  $u$  são parâmetros fixos fornecidos pelo sistema; e  $q_{i(1-N_t)}$  e  $c_{i(1-N_t)}$  são as concentrações iniciais de proteína das fases estacionária e móvel, respectivamente.

Tabela 4.3. Condições iniciais e parâmetros fixos utilizados nas simulações da cromatografia de troca aniônica.

Nome do parâmetro	Símbolo	Unidade	Valor
Proteínas inicialmente adsorvidas	$q_{i(1-N_t)}$	(mol.mL <sup>-1</sup> )	0
Proteínas inicialmente não adsorvidas	$c_{i(1-N_t)}$	(mol.mL <sup>-1</sup> )	0
Altura da coluna	$L$	(cm)	13
Velocidade da fase móvel	$u$	(cm.s <sup>-1</sup> )	0,042
Porosidade total da resina	$\epsilon_T$	-	0,9
Número total de pratos teóricos	$N_T$	-	189
Capacidade de saturação da resina <sup>A</sup>	$\Lambda$	(mol.L <sup>-1</sup> )	2,15

<sup>A</sup>Dados fornecidos por Huuk.

A Tabela 4.4 mostra os parâmetros estimados para as simulações I a V e os valores  $RMSE_{Est}$  e  $RMSE_{Val}$ . Para as simulações I, II e III  $RMSE_{Est} > RMSE_{Val}$ , mostrando que as validações foram eficientes. No entanto, o  $RMSE_{Val}$  da simulação IV foi cinco vezes maior que o  $RMSE_{Est}$ , expondo uma diminuição do desempenho da simulação quando uma força iônica de eluição mais alta foi aplicada. Em todos os casos,  $K_{Eq,Q1-I} < K_{Eq,Q1-II} < K_{Eq,Q2} < K_{Eq,Q3}$ , um comportamento esperado, uma vez que os grupos de proteínas com  $K_{Eq}$  menor apresentaram menor afinidade com a resina e foram eluídos primeiro.  $K_{Eq}$  variou entre 0,3 e 11 para todos os conjuntos de proteínas. Outros trabalhos recentes forneceram valores com a mesma ordem de magnitude quando a resina Q-Sepharose foi utilizada para purificar proteínas: KARKOV et al. (2013) encontraram 2,17 e 0,13 para insulina e BSA, respectivamente, e HUUK et al. (2014) encontraram 0,22 para a ribonuclease. Quando o

parâmetro  $v$  é analisado, em todos os casos os pools de proteínas que são eluídos posteriormente ( $Q_2$  e  $Q_3$ ) apresentaram valores mais altos do que as proteínas dos grupos  $Q_1$ -I e  $Q_2$ -II, mostrando que os grupos mais fortemente ligados apresentaram maior quantidade de sítios ativos. Em todos os casos,  $v$  variou de 0,6 a 1,4, com a mesma ordem de magnitude dos valores fornecidos por KARKOV et al. (2013) para insulina e BSA (2,75 em ambos os casos) e por HUUK et al. (2014) para ribonuclease (1,5). Em relação ao parâmetro  $\alpha$ , os grupos de proteínas com maior afinidade pela resina ( $Q_2$  e  $Q_3$ ) apresentaram valores superiores aos grupos  $Q_1$ -I e  $Q_1$ -II, indicando que a interação dos primeiros deve ser mais complexa e forte, aumentando o fator estérico. Essas altas diferenças entre os fatores estéricos estimados também foram observadas por FARAJI et al. (2015), que encontraram valores de 1283 e 0,16 para lactoperoxidase e lactoferrina quando utilizaram resina SP-Sepharose.

Tabela 4.4. Parâmetros estimados das simulações I a V para os pools de proteínas ( $Q_1$ -I,  $Q_1$ -II,  $Q_2$  e  $Q_3$ ) observados durante a cromatografia de troca aniônica.

Pool de proteínas		Simulação				
		I	II	III	IV	V
<b><math>Q_1</math>-I</b>	$K_{Ads}$	171,6	175,2	178,2	174,0	175,3
	$K_{Des}$	478,3	485,6	478,3	480,1	481,0
	$K_{Eq}$	0,36	0,36	0,37	0,36	0,36
	$v$	0,61	0,65	0,74	0,84	0,64
	$\alpha$	9,15	10,40	8,74	9,91	9,59
<b><math>Q_1</math>-II</b>	$K_{Ads}$	330,5	348,2	358,0	354,2	346,9
	$K_{Des}$	267,1	264,0	263,4	267,3	267,1
	$K_{Eq}$	1,24	1,32	1,36	1,32	1,30
	$v$	0,83	0,85	0,86	0,85	0,83
	$\alpha$	0,58	1,04	0,76	0,60	0,76
<b><math>Q_2</math></b>	$K_{Ads}$	565,8	547,6	552,8	542,0	553,5
	$K_{Des}$	139,4	133,7	139,3	149,4	141,0
	$K_{Eq}$	4,06	4,09	3,97	3,63	3,92
	$v$	1,34	1,30	1,32	1,33	1,33
	$\alpha$	150	139	136	120	137
<b><math>Q_3</math></b>	$K_{Ads}$	417,4	414,0	411,4	413,5	415,2
	$K_{Des}$	40,1	40,0	40,2	40,7	40,3
	$K_{Eq}$	10,40	10,36	10,24	10,17	10,30
	$v$	1,30	1,27	1,32	1,26	1,28
	$\alpha$	206	224	204	209	211
<b>RMSE</b>	$RMSE_{Est} \times 10^{16}$	3,4	3,3	4,0	1,7	-
	$RMSE_{Val} \times 10^{16}$	2,9	2,8	1,7	8,3	-

\*  $K_{Eq} = K_{Ads} \cdot K_{Des}^{-1}$

As figuras 4.3 a 4.6 mostram os cromatogramas experimentais obtidos nas simulações I a IV. Os experimentos validados foram A, B, C e D, respectivamente. Embora a estratégia de validação cruzada considere que os pools de proteínas Q<sub>1</sub>-I e Q<sub>1</sub>-II apresentavam características diferentes, conforme mostrado nos dados experimentais, os mesmos foram eluídos em um único pico para descrever a simulação de Q<sub>1</sub>. Em todos os casos, é possível visualizar que o pico Q<sub>2</sub> apresenta uma qualidade de ajuste menor, uma vez que o pico simulado se afasta mais do experimental do que o Q<sub>1</sub> e Q<sub>3</sub>. Uma possível razão para isso é uma transferência de massa não desprezível que ocorre na partícula da resina, causada pela maior concentração de proteínas do pico Q<sub>2</sub>. Esses fenômenos não são considerados pelo MED, mas existem modelos mais complexos que os descrevem, como por exemplo o modelo do transporte dispersivo, que considera efeitos convectivos na superfície da resina (CANDY et al., 2012; FARAJI et al., 2015) e o modelo geral, que considera os efeitos convectivos na superfície e a difusividade da proteína no interior da partícula (KACZMARSKI et al., 2001; ORELLANA et al., 2009; LIENQUEO et al., 2012). A aplicação desses modelos não foi o foco deste trabalho.

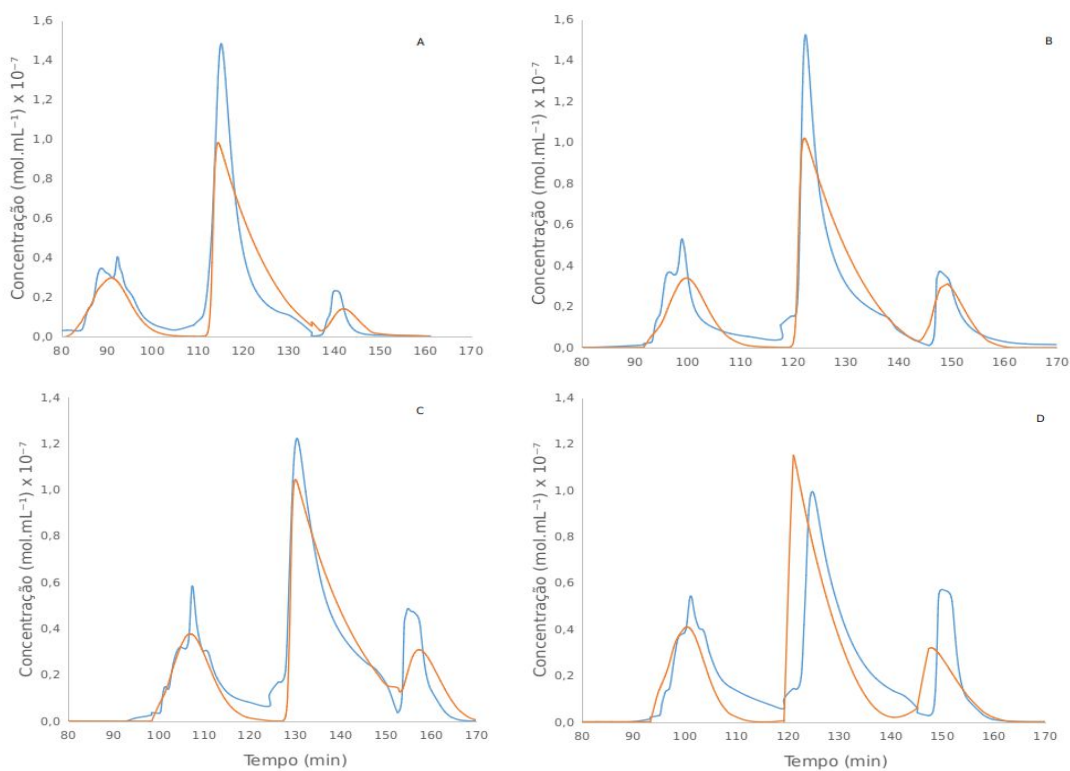


Figura 4.3. Curvas das eluições Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> simuladas (vermelhas) e ajustadas às curvas experimentais (azuis) dos experimentos B, C e D. A performance da simulação foi validada pelo experimento A.

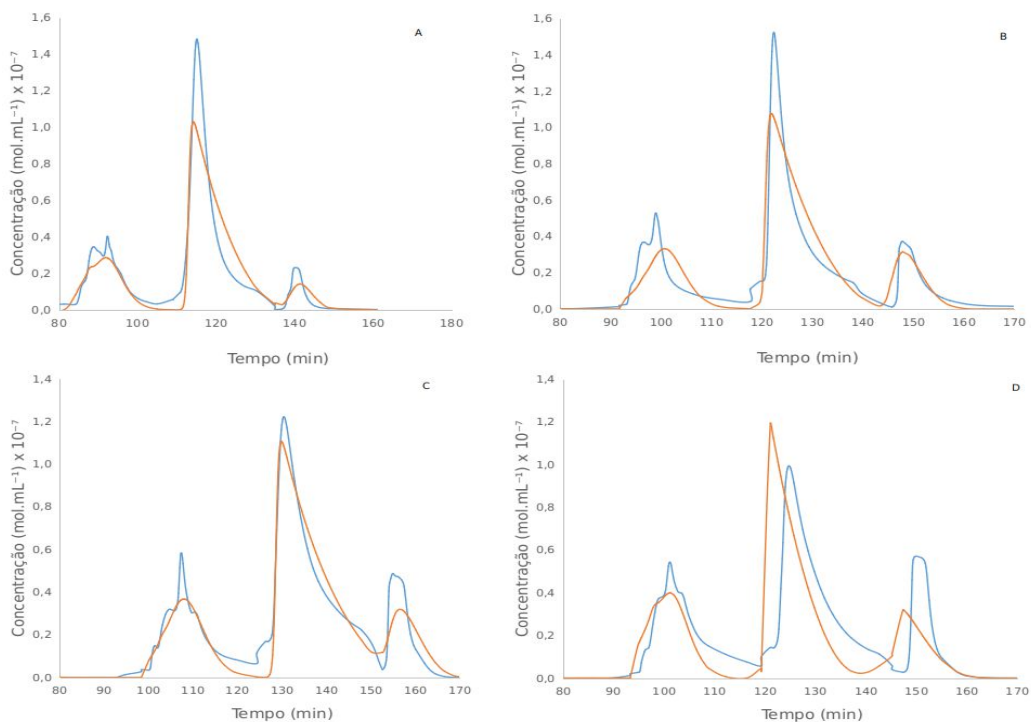


Figura 4.4. Curvas das eluições Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> simuladas (vermelhas) e ajustadas às curvas experimentais (azuis) dos experimentos A, C e D. A performance da simulação foi validada pelo experimento B.

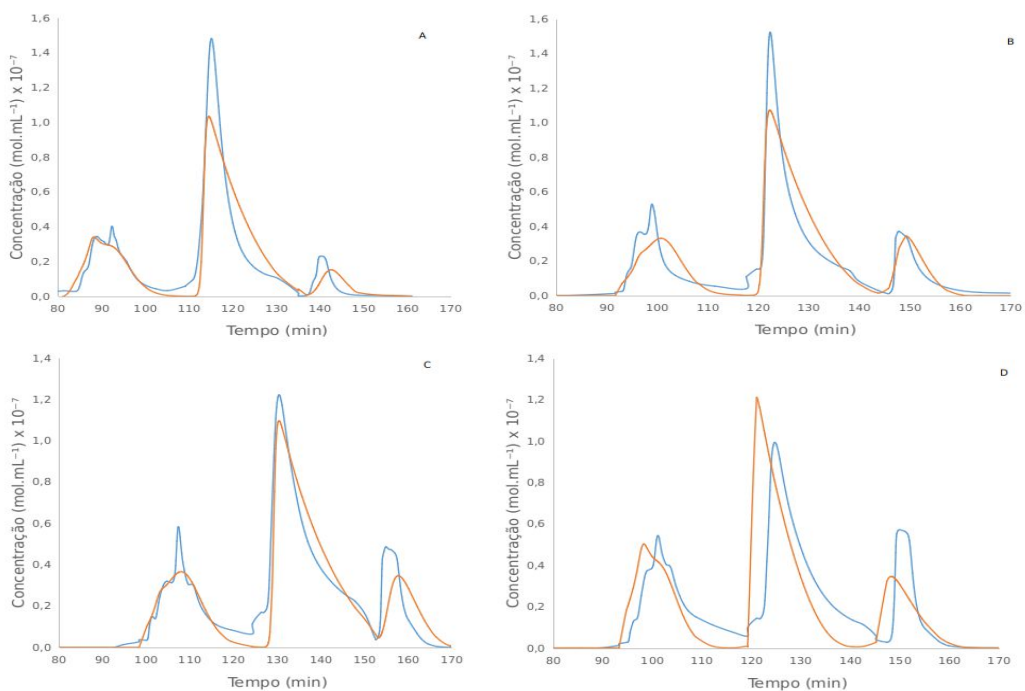


Figura 4.5. Curvas das eluições Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> simuladas (vermelhas) e ajustadas às curvas experimentais (azuis) dos experimentos A, B e D. A performance da simulação foi validada pelo experimento C.

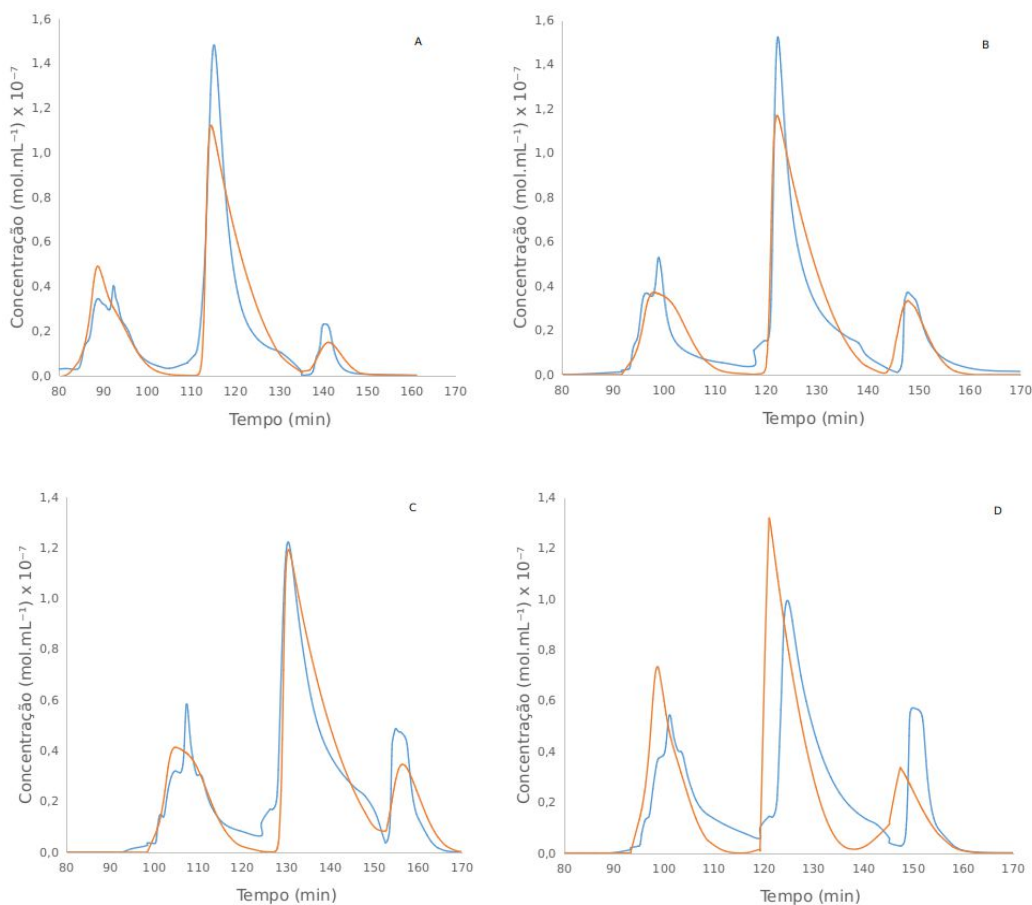


Figura 4.6. Curvas das eluições Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> simuladas (vermelhas) e ajustadas às curvas experimentais (azuis) dos experimentos A, B e C. A performance da simulação foi validada pelo experimento D.

Os resultados dessa seção mostram que a estratégia de modelagem utilizada fornece uma descrição adequada dos perfis de eluição de PspA4Pro e, portanto, pode ser utilizada para melhor compreender o processo de purificação da proteína de interesse.

### 4.3.3. Simulações dos ECC virtuais e determinação da capacidade de processamento da resina

Simulações de ECC virtuais foram executadas com as massas de entrada  $1 \times M_A$ ,  $2 \times M_A$  e  $3 \times M_A$ , de acordo com o procedimento descrito na Seção 4.2.7. Os parâmetros fixos da Tabela 4.3 e os parâmetros estimados da Tabela 4.4 (Simulação V) foram novamente utilizados para as simulações dessa seção. A Figura 4.7 mostra o perfil de adsorção dos grupos de proteínas  $Q_{1-I}$ ,  $Q_{1-II}$ ,  $Q_2$  e  $Q_3$  nos pratos teóricos da coluna em três diferentes condições operacionais: Imediatamente após as aplicações de ECC, da eluição  $Q_1$  e da eluição  $Q_2$ .

Os perfis de adsorção da linha A foram obtidos a partir do processamento das massas dos pools de proteínas quantificados no experimento A (Tabela 4.1), chamado  $M_A$ . Portanto, neste caso é possível analisar o que aconteceu dentro da coluna durante este experimento: Após a aplicação de ECC todas as proteínas permaneceram retidas nos primeiros 25 pratos da coluna. Após a aplicação de  $Q_1$ , foi observada a remoção completa dos pools de proteínas  $Q_{1-I}$  e  $Q_{1-II}$ , enquanto os pools  $Q_2$  e  $Q_3$  permaneceram retidos na coluna. O pool  $Q_2$  ficou retido entre os pratos 50 e 150, uma vez que apresenta menor afinidade com a resina, enquanto o pool  $Q_3$  permaneceu retido no início da coluna, entre os pratos 1 e 100. Após a aplicação da eluição  $Q_2$ , o pool de proteínas  $Q_2$  foi completamente removido. Enquanto isso, o pool  $Q_3$  permaneceu retido entre os pratos 50 e 150, uma vez que seria posteriormente removido pela aplicação da eluição  $Q_3$ .

As simulações representadas na linha B reproduzem o comportamento da coluna quando há processamento de duas vezes a massa de proteínas adsorvidas no experimento A ( $2 \times M_A$ ). Neste caso, os pools de proteínas permaneceram retidos nos primeiros 50 pratos da coluna. Mais uma vez, a aplicação da eluição  $Q_1$  removeu completamente os conjuntos de proteínas  $Q_{1-I}$  e  $Q_{1-II}$ . Uma pequena quantidade de  $Q_2$  também foi removida, indicando que talvez uma recuperação mais baixa desse pool fosse observada na eluição  $Q_2$ . A aplicação da eluição  $Q_2$  removeu completamente o pool  $Q_2$ , mas uma quantidade muito pequena do pool  $Q_3$  também foi removida, indicando que, neste caso, a pureza de PspA4Pro na fração  $Q_2$  poderia sofrer uma pequena contaminação.

As simulações mostradas na linha C reproduzem o comportamento da coluna quando são processadas três vezes a massa das proteínas adsorvidas no experimento A ( $3 \times M_A$ ). Após a aplicação de ECC, aproximadamente 100 pratos da coluna permaneceram

saturados. Após a aplicação de  $Q_2$ , é possível observar que uma grande quantidade do pool de proteínas  $Q_3$  elui junto com  $Q_2$ , indicando uma alta probabilidade de contaminação na PspA4Pro obtida.

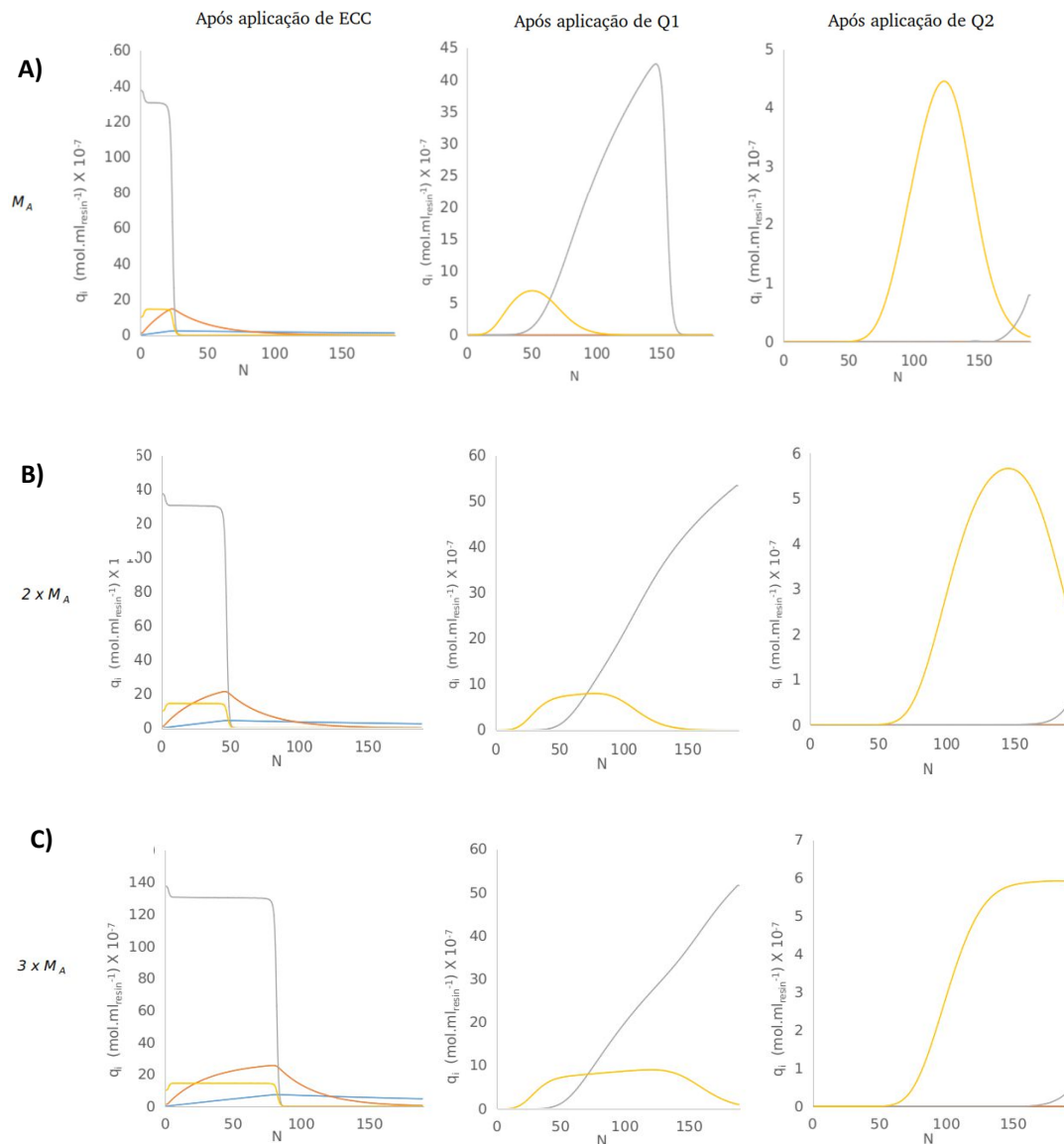


Figura 4.7. Perfis de eluição simulados ( $q_i$ ) dos pools de proteínas  $Q_1$ -I (linhas azuis),  $Q_1$ -II (linhas laranjas),  $Q_2$  (linha cinza) e  $Q_3$  (linha amarela) nos  $N$  pratos teóricos da coluna cromatográfica a partir do processamento da massa de proteínas adsorvidas no experimento A, chamada de  $M_A$  (linha A),  $2 \times M_A$  (linha B) e  $3 \times M_A$  (linha C). As condições simuladas foram apresentadas imediatamente após aplicações do extrato celular clarificado (ECC) e eluições  $Q_1$  e  $Q_2$ .

A partir dos resultados apresentados acima, pode-se considerar que o uso de um ECC com  $2 \times M_A$  poderia ser uma estratégia adequada para explorar uma capacidade de processamento limitante da cromatografia de troca aniônica. A inserção de uma massa proteica superior a  $2 \times M_A$  provavelmente reduziria a pureza da PspA4Pro eluída na fração  $Q_2$ , uma vez que esta seria contaminada por proteínas que deveriam ter sido eluídas em  $Q_3$ , como mostram as simulações. O experimento E foi realizado para reproduzir esta condição. A análise do perfil de eluição na Tabela 4.2 mostra que a purificação de um ECC com 5,62 g de proteínas ( $\approx 2 \times M_A$ ) levou à produção de 1,54 g de proteínas no pool  $Q_2$ , referente a  $0,022 \text{ g.mL}_{\text{resina}}^{-1}$  de proteína com uma pureza de PspA4Pro de 81,2%. Os resultados experimentais não mostraram contaminação com PspA4Pro ou perda de recuperação, como previsto pela simulação, o que pode ser explicado pelo desvio padrão do modelo associado aos parâmetros estimados. Os resultados do experimento E mostram um aumento da capacidade de processamento de cromatografia sem perda da qualidade do PspA4Pro. No experimento A, foram obtidas  $0,010 \text{ g.mL}_{\text{resina}}^{-1}$  de proteína com 73,9% de pureza e FIGUEIREDO et al. (2017) obtiveram  $0,007 \text{ g.mL}_{\text{resina}}^{-1}$  de proteína com 65,2%. Esses resultados mostram que a mesma coluna pode ser alimentada com até duas vezes a massa inicial de proteínas no ECC, apresentando purezas semelhantes ou maiores, o que aumenta a eficiência e reduz o custo operacional.

#### 4.4. Conclusões

As simulações deste trabalho forneceram uma boa descrição das corridas cromatográficas aplicadas em misturas não sintéticas de proteínas contendo PspA4Pro. Os parâmetros foram ajustados adequadamente através da metodologia de validação cruzada e os dados experimentais foram bem descritos, exceto por uma pequena diminuição na qualidade da simulação quando as forças iônicas das eluições aumentaram. As simulações foram capazes de prever que a coluna cromatográfica poderia processar duas vezes mais proteína sem perda da qualidade de PspA4Pro obtida. Essa condição foi testada experimentalmente, fornecendo  $0,022 \text{ g.mL}_{\text{resina}}^{-1}$  de PspA4Pro com 81,2% de pureza em comparação com  $0,010 \text{ g.mL}_{\text{resina}}^{-1}$  de PspA4Pro com 73,9% de pureza no processo original. A abordagem apresentada neste trabalho facilita a compreensão do processo de purificação de maneira sistemática e permite a aplicação de diversas melhorias técnicas.

## **5. UMA ABORDAGEM ROBUSTA PARA MODELAR E SIMULAR PROCESSOS CROMATOGRÁFICOS NA PURIFICAÇÃO DE MISTURAS COMPLEXAS**

O principal objetivo desta última etapa do trabalho foi apresentar uma proposta de simulação que melhorasse a abordagem desenvolvida no capítulo 3, possibilitando o desenvolvimento de um modelo mais completo que permitisse a identificação de alternativas viáveis para melhorar o desempenho da purificação. Diversas cromatografias foram realizadas e diferentes forças iônicas foram testadas. Mais uma vez, os parâmetros foram ajustados separadamente às cromatografias executadas com amostras sintéticas de PspA4Pro ou amostras com proteínas de células não produtoras de PspA4Pro. Para isso, utilizou-se a metodologia de validação cruzada descrita no capítulo 4. A maior quantidade de dados disponíveis, aliada ao procedimento apresentado, levaram a obtenção de um modelo mais robusto, possibilitando a simulação de esquemas alternativos de purificação que se basearam na fragmentação e modificação das forças iônicas das eluições principais. Mais uma vez, as simulações mostraram que a PspA4Pro tendia a eluir no início do pico. No entanto, diferentes perfis de eluição foram observados de acordo com a variação da força iônica, permitindo a identificação de condições que aumentaram a pureza e rendimento específico por volume de coluna. Os melhores resultados foram obtidos em menores forças iônicas, proporcionando um aumento de rendimento e pureza expressivos.

### *Resumo*

Os processos cromatográficos são considerados as etapas mais importantes na purificação de proteínas recombinantes. Recentemente, um grande esforço tem sido realizado para modelar e simular essas operações. No entanto, um problema frequente está relacionado à utilização de misturas sintéticas para identificação e validação de modelos, dificultando a aplicação das simulações em ensaios cromatográficos reais, uma vez que as proteínas de interesse geralmente estão presentes em misturas complexas de proteínas. Nesse contexto, este trabalho teve como objetivo melhorar uma abordagem desenvolvida anteriormente que visava lidar com a simulação de misturas complexas reais, aproximando a teoria da prática. A abordagem foi aplicada para purificar PspA4Pro, uma proteína recombinante de superfície de pneumococo utilizada para produzir vacinas. Em um primeiro momento, o modelo do equilíbrio dispersivo e a isoterma de ação em efeito estérico foram utilizados para descrever separadamente os perfis de eluição de PspA4Pro e das impurezas proteicas. Os parâmetros foram estimados por uma metodologia de validação cruzada que foi aplicada a um conjunto completo de corridas cromatográficas. Em um segundo momento, esquemas alternativos de eluição que aumentariam o desempenho do processo foram propostos por simulações realizadas em diferentes forças iônicas. Para validação, foram utilizados dados de extratos celulares clarificados derivados de células produtoras de PspA4Pro, com quantidades de proteínas e purezas determinadas em diferentes regiões dos picos de eluição. As simulações mostraram que PspA4Pro tendia a eluir no início do segundo pico. No entanto, perfis diferentes de eluição foram observados de acordo com a variação da força iônica, permitindo identificar condições que aumentavam a pureza sem redução do rendimento específico de PspA4Pro produzida por volume da coluna. Os melhores resultados foram obtidos com forças iônicas menos intensas, proporcionando um aumento de pureza e rendimento específico de 149% e 9,9%, respectivamente, quando comparados com a abordagem de operação original.

## 5.1. Introdução

A demanda de produção de proteínas recombinantes para aplicações industriais, criações de novos medicamentos e desenvolvimentos de pesquisas aumentou muito nos últimos anos. Uma quantidade relevante dessas proteínas é produzida por *Escherichia coli* recombinante, devido ao vasto conhecimento acumulado sobre a genética e metabolismo desse micro-organismo, à disponibilidade de métodos rápidos e precisos para a modificação do genoma (SHILOACH et al., 2005; SHOJA et al., 2008) e a possibilidade de produzir grandes quantidades de biomassa usando protocolos de cultivo bem estabelecidos (MAMAT et al., 2015). Apesar disso, as etapas de purificação ainda são um obstáculo para viabilizar a produção de proteínas, uma vez que muitas vezes são responsáveis por aumentar o custo e a complexidade do processo devido à demanda por um produto de alta qualidade, com contaminantes como ácidos nucleicos, pirogênios e enzimas em geral em concentrações abaixo dos limites estabelecidos pelas agências reguladoras (MAMAT et al., 2015; FIGUEIREDO et al., 2017).

As etapas de purificação mais importantes são geralmente compostas por corridas cromatográficas que exigem equipamentos de alta tecnologia e mão de obra qualificada (GU, 2015). Um esforço contínuo tem sido feito para modelar e simular esses processos com o objetivo de otimizar e aumentar eficiências (ORELLANA et al., 2009; LINQUEO et al., 2012; TEOH et al., 2001; KACZMARSKI et al., 2001; CHAN et al., 2008; ORELLANA et al., 2009; CLOSE et al., 2014; FARAJI et al., 2015; DONAT et al., 2018). No entanto, devido à dificuldade em simular misturas proteicas reais de extratos celulares clarificados, geralmente são utilizadas misturas sintéticas que apresentam um número limitado de proteínas puras para obter dados e caracterizar o processo. Essa estratégia é adequada para o planejamento experimental, estimativa de parâmetros do modelo e desenvolvimento de modelos teóricos robustos, mas inadequada para lidar com situações reais, pois os extratos celulares clarificados são misturas complexas de proteínas com características desconhecidas, apresentando variações na composição de cultivo para cultivo, conforme já discutido no capítulo 3 desta tese.

Por esse motivo, é importante que sejam desenvolvidos modelos capazes de descrever o comportamento de extratos celulares reais durante as cromatografias. Esses extratos são compostos por misturas complexas contendo a proteína de interesse e diversas outras proteínas contaminantes desconhecidas. No capítulo 3, foi proposta uma nova

abordagem para simular a separação cromatográfica de misturas reais não sintéticas. A proteína de interesse a ser purificada foi a PspA4Pro, uma proteína recombinante produzida por *Escherichia coli* BL21 (DE3) com o plasmídeo pET37b + PspA4Pro, expressando o gene do fragmento N-terminal da proteína A de superfície de pneumococo (PspA) (PERCIANI et al., 2013). A abordagem foi baseada em um procedimento que simulou separadamente a PspA4Pro e o conjunto de impurezas proteicas eluídas no mesmo momento, possibilitando identificar um perfil de eluição alternativo que aumentou a pureza em 34% em relação ao processo de purificação original. A abordagem apresentada no Capítulo 3 pode ser considerada um primeiro passo para desenvolver novas alternativas para lidar com misturas complexas. No entanto, apesar dos bons resultados, o aumento da pureza foi acompanhado por uma redução relevante no rendimento específico de proteína produzida por volume de coluna, e o modelo desenvolvido apresentou algumas incertezas devido à complexidade das misturas processadas e à quantidade reduzida de dados utilizados para estimar os parâmetros.

Nesse contexto, a principal motivação deste estudo é melhorar ainda mais a abordagem desenvolvida no Capítulo 3, fornecendo um modelo mais robusto para a simulação, e explorando sua capacidade de descrever o processo em diferentes forças iônicas de eluição, contribuindo para melhorar ainda mais as etapas do processo *downstream*. Em um primeiro momento, um conjunto completo de corridas cromatográficas foi aplicado em dois tipos diferentes de amostras: solução pura de PspA4Pro e extratos celulares clarificados reais derivados de células não produtoras de PspA4Pro. Diferentes forças iônicas foram usadas para eluir as proteínas e uma estratégia de validação cruzada<sup>9</sup> foi aplicada para estimar separadamente os parâmetros de adsorção de PspA4Pro e do pool de proteínas contaminantes. Os modelos foram validados por um novo conjunto de ensaios cromatográficos e modificações nas forças iônicas das eluições foram propostas como estratégia para aumentar a pureza de PspA4Pro sem perda de rendimento específico de proteína produzida por volume de coluna. Todas as simulações foram realizadas no EMSO (*Environment for Modeling, Simulation and Optimization*), um software orientado a equações (SOARES e SECCHI, 2003; FURLAN et al., 2016) que tem sido utilizado com sucesso para simular uma quantidade diversificada de processos relacionados a engenharia química (ALBARELLI et al., 2017; FONSECA et al., 2017; DA ROCHA NOVAES et al., 2017; ALMEIDA et al., 2018).

---

<sup>9</sup> A metodologia de validação cruzada é descrita em detalhes na seção 4.2.6. e Figura 4.2.

Espera-se que o procedimento proposto neste trabalho permita uma robusta melhoria na simulação de processos cromatográficos de misturas reais, incrementando a qualidade dos dados obtidos. Trata-se de um fator muito importante para melhor compreender e otimizar processos reais encontrados na indústria biofarmacêutica, uma vez que a produção de vários compostos com alta pureza depende de protocolos de purificação eficientes aplicados em misturas reais de proteínas.

## 5.2. Teoria

### 5.2.1. Modelo do equilíbrio dispersivo

O modelo do equilíbrio dispersivo (MED) é um modelo matemático usado para descrever a transferência de massa de proteínas dentro da coluna cromatográfica, assumindo um sistema isotérmico e adiabático, uma coluna radialmente homogênea e um coeficiente aparente de dispersão axial que depende das dimensões da coluna (CLOSE et al., 2014). O MED pode ser descrito pela Equação 5.1, onde  $c_i$  é a concentração de uma proteína “i” na fase móvel ( $\text{mol.mL}^{-1}$ ),  $q_i$  é a quantidade de adsorbato retido na estrutura sólida ( $\text{mol.mL}_{\text{resina}}^{-1}$ ),  $u$  é a velocidade da fase móvel ( $\text{cm.s}^{-1}$ ) e  $D_a$  é a dispersão axial ( $\text{cm}^2.\text{s}^{-1}$ ) que pode ser calculada usando a Equação 5.2, onde  $L$  é o comprimento da coluna e  $N_t$  é o número total de pratos teóricos (TEOH et al., 2001; CLOSE et al., 2014).

$$\frac{\partial c_i}{\partial t} + \frac{(1-\epsilon_T)}{\epsilon_T} \frac{\partial q_i}{\partial t} + u \frac{\partial c_i}{\partial z} = D_a \frac{\partial^2 c_i}{\partial z^2} \quad (5.1)$$

$$D_a = \frac{Lu}{2N_t} \quad (5.2)$$

O parâmetro  $N_t$  pode ser calculado pela Equação 5.3 para um pico cromatográfico de um componente que não apresente afinidade com a resina. Nesta equação,  $t_0$  é o tempo de retenção da coluna e  $\Delta$  é a largura do pico na metade de sua altura (AMERSHAM-BIOSCIENCES, 2004). A porosidade total ( $\epsilon_T$ ) pode ser calculada pelas equações 5.4 e 5.5, em que  $V_0$  e  $V_C$  representam os volumes vazio e total da coluna, respectivamente, e  $t_0$  e  $F$  representam o tempo de retenção do componente não reativo e a vazão, respectivamente (CLOSE et al., 2014).

$$N_i = 5.54 \left( \frac{t_o}{\Delta} \right)^2 \quad (5.3)$$

$$\epsilon_T = \frac{V_o}{V_C} \quad (5.4)$$

$$V_o = t_o \cdot F \quad (5.5)$$

### 5.2.2. A isoterma de ação em efeito estérico

A isoterma de ação em efeito estérico, ou “*steric mass action isotherm*” (SMA), é uma equação fenomenológica que representa a competição entre proteínas e soluções iônicas durante a cromatografia de troca aniônica (GU, 2015). Este modelo de adsorção considera que não há equilíbrio do processo de adsorção/dessorção, sendo que o sal nas soluções de eluição compete com as proteínas para se ligar à fase sólida. Além disso, o modelo considera que algumas proteínas apresentam estruturas complexas, tornando-as capazes de se ligar a mais de um sítio ativo, bem como bloquear vários outros (KARKOV et al., 2013; HUUK et al., 2014; GU, 2015). Este modelo pode ser representado pelas Equações 5.6 e 5.7. A primeira descreve que a quantidade de proteína “i” adsorvida depende das constantes de adsorção ( $k_{Ads,i}$ ) e dessorção ( $k_{Des,i}$ ) ( $\text{mL} \cdot \text{mol}^{-1} \cdot \text{s}^{-1}$ ), do número de sítios de ligação das proteínas ( $v_i$ ), da concentração de sal na fase líquida ( $c_{sal}$ ) e da carga livre resultante (Q) da resina ( $\text{mol} \cdot \text{mL}^{-1}$ ), a qual representa a neutralidade eletrostática do sistema, onde  $\Lambda$  ( $\text{mol} \cdot \text{mL}_{resina}^{-1}$ ) é a capacidade iônica total da resina e  $\alpha_i$  é o fator estérico, diretamente relacionado ao bloqueio molecular (KARKOV et al., 2013; HUUK et al., 2014).

$$\frac{\partial q_i}{\partial t} = K_{Ads,i} c_i Q^{v_i} - K_{Des,i} q_i c_{salt}^{v_i} \quad (5.6)$$

$$Q = \Lambda - \sum_{i=1}^N q_i (v_i + \alpha_i) \quad (5.7)$$

### 5.3. Materiais e métodos

#### 5.3.1. Produção de PspA4Pro

O processo de purificação de PspA4Pro foi aplicado à biomassa de *E. coli* produzida sob duas condições diferentes de cultivo. A condição de cultivo indicada como “A” foi realizada a 37 °C, induzida por lactose, utilizando células *rE. coli* portadoras de pET37b + sem inserto, de modo que nenhuma PspA4Pro recombinante foi produzida (Capítulo 3)<sup>10</sup>. A condição de cultivo indicada como “B” foi realizada a 37 °C, também com indução por lactose, mas utilizando *E. coli* contendo um fragmento do gene da proteína A de superfície de pneumococo (PspA) de *Streptococcus pneumoniae* (CAMPANI et al., 2016). O primeiro bloco rico em prolina do gene PspA (GenBank: EF649969.1), aqui denominado PspA4Pro, foi inserido no plasmídeo pET37b + (Novagen) e expresso na *E. coli* BL21 (DE3) (Invitrogen, Carlsbad, CA, EUA) (MORENO et al., 2010; FIGUEIREDO et al., 2017), portanto, as células desse cultivo produziram PspA4Pro. As células recombinantes de *E. coli* foram gentilmente fornecidas pela Dra. Eliane Miyaji do Laboratório de Biologia Molecular do Instituto Butantan (São Paulo, Brasil). Todos os experimentos foram realizados em um biorreator de 5 L, o qual foi monitorado e controlado pelo software SuperSys\_HCDC (HORTA et al., 2014). NH<sub>4</sub>OH (30%) foi utilizado para manter o pH em 6,7. O controlador ajustou automaticamente a velocidade do agitador entre 200 e 900 rpm a fim de manter a concentração de oxigênio dissolvido em 30% de saturação. A vazão total do gás foi mantida em 4 L.min<sup>-1</sup> (HORTA et al., 2014; CAMPANI et al., 2016). Ao final do cultivo, a biomassa foi separada do caldo por centrifugação durante 30 min, a 4 °C e 6700 g. Os pellets de células resultantes foram armazenados em sacos plásticos a -80 °C.

---

<sup>10</sup> As células do cultivo A utilizadas neste capítulo são as mesmas que as do cultivo C do capítulo 3.

Para executar as cromatografias de troca aniônica (CTA),  $100 \pm 10$  g das células A ou B foram descongeladas e suspensas em 1 L de tampão de lise. As células foram lisadas através de recirculação da suspensão em um homogeneizador contínuo de alta pressão (HCAP) por 8 min a  $4\text{ }^{\circ}\text{C}$ , 500 bar e vazão de  $1\text{ L}\cdot\text{min}^{-1}$ . Em seguida, foi adicionado 0,1% de detergente de brometo de cetiltrimetilamônio (CTAB) e as misturas foram agitada por 60 min em temperatura ambiente. As amostras A e B foram então centrifugadas por 90 min a  $17.696\text{ g}$  e  $4\text{ }^{\circ}\text{C}$  (FIGUEIREDO et al., 2017). O sobrenadante de cada amostra foi indicado como extrato celular clarificado (ECC). As amostras de ECC dos cultivos A e B foram denominadas  $\text{ECC}_A$  e  $\text{ECC}_B$ , respectivamente. É importante destacar que o  $\text{ECC}_A$  apresentava somente o conjunto de proteínas do metabolismo de *E.coli* enquanto que o  $\text{ECC}_B$  apresentava um conjunto semelhante de proteínas metabólicas misturadas com PspA4Pro.

### 5.3.2. Cromatografia de troca aniônica

Foram realizadas um total de nove cromatografias de troca aniônica (designadas CTA-I a CTA-IX): CTA-I, CTA-II e CTA-III foram aplicadas em  $\text{ECC}_A$ . CTA-IV, CTA-V e CTA-VI foram aplicadas em uma solução de proteína pura (SPP) preparada com PspA4Pro previamente purificada (FIGUEIREDO et al., 2017). CTA-VII, CTA-VIII e CTA-IX foram aplicadas em  $\text{ECC}_B$ . Todas as corridas cromatográficas foram realizadas usando 70 mL de resina Q-Sepharose FF empacotada em uma coluna XK 26/20 que foi instalada em um cromatógrafo Äkta Avant 150 (Instituto Butantan). Todos os materiais cromatográficos foram fornecidos pela GE Healthcare (EUA). A vazão linear utilizada foi de  $2,50\text{ mL}\cdot\text{min}^{-1}\cdot\text{cm}^{-2}$  ( $13,5\text{ cm}^3\cdot\text{min}^{-1}$ ). O tampão de equilíbrio foi preparado com fosfato de sódio 10 mM (pH 6,5 e condutividade de  $1,2\text{ mS}\cdot\text{cm}^{-1}$ ). O pH e a condutividade das amostras  $\text{ECC}_A$ ,  $\text{ECC}_B$  e SPP foram ajustados para os mesmos valores de pH e condutividade do tampão de equilíbrio. Três tampões de eluição foram preparados. O primeiro foi denominado  $Q_1$  e apresentou a menor força iônica, com 150 mM de NaCl. O segundo foi denominado  $Q_2$  e apresentou uma força iônica média, com concentrações de NaCl que variaram entre 250 mM e 350 mM. O terceiro foi denominado  $Q_3$  e apresentou a maior força iônica, com 1000 mM de NaCl. As características de  $Q_1$ ,  $Q_2$  e  $Q_3$  das CTAs podem ser visualizadas em detalhes na Tabela 5.1. Todas as cromatografias foram realizadas de acordo com as seguintes etapas: (i) aplicação de 5 volumes de coluna (VC) do tampão de equilíbrio; (ii) aplicação das amostras  $\text{ECC}_A$ ,  $\text{ECC}_B$  ou SPP; (iii) aplicação de 5 VC do tampão de equilíbrio para remoção de moléculas não

adsorvidas pela resina após a introdução da amostra; (iv) aplicação sequencial de 5 VC dos tampões de eluição Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub>.

De acordo com FIGUEIREDO et al. (2017), a PspA4Pro deveria ser eluída na fração Q<sub>2</sub>, juntamente com um conjunto de outras impurezas proteicas. Os resultados do capítulo 3 mostraram que a PspA4Pro e esse conjunto de proteínas apresentaram diferentes perfis de eluição no pico Q<sub>2</sub>, uma vez que o tempo de permanência de PspA4Pro na coluna é menor. Assim, para obter mais informações sobre o perfil cromatográfico deste pico, as eluições Q<sub>2</sub> das amostras ECC<sub>B</sub> foram fragmentadas em dez subfrações de 0,5 VC, conforme ilustrado pela Figura 5.1.

Tabela 5.1. Descrição das amostras de entrada e definição dos perfis de eluição Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> para as cromatografias de troca aniônica (CTA) I a IX realizadas em coluna Q-Sepharose de 70 mL.

<b>CTA</b>	<b>Amostra Entrada</b>	<b>Descrição</b>	<b>Perfis de eluição de NaCl de Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub> (mM)</b>
<b>I</b>	ECC <sub>A</sub>	Extrato celular clarificado obtido de células	150, 300, 1000
<b>II</b>	ECC <sub>A</sub>	não produtoras de PspA4Pro cultivadas a	150, 250, 1000
<b>III</b>	ECC <sub>A</sub>	37 °C com indução por lactose	150, 350, 1000
<b>IV</b>	SPP		150, 300, 1000
<b>V</b>	SPP	Amostra de PspA4Pro pura	150, 250, 1000
<b>VI</b>	SPP		150, 350, 1000
<b>VII</b>	ECC <sub>B</sub>	Extrato celular clarificado obtido de células	150, 300*, 1000
<b>VIII</b>	ECC <sub>B</sub>	produtoras de PspA4Pro cultivadas a	150, 250*, 1000
<b>IX</b>	ECC <sub>B</sub>	37 °C com indução por lactose	150, 350*, 1000

ECC: Extrato celular clarificado

SPP: Solução pura de PspA4Pro

\*Essas frações foram fragmentadas em dez subfrações de 0,5 VC, como representado pela Figura 5.1.

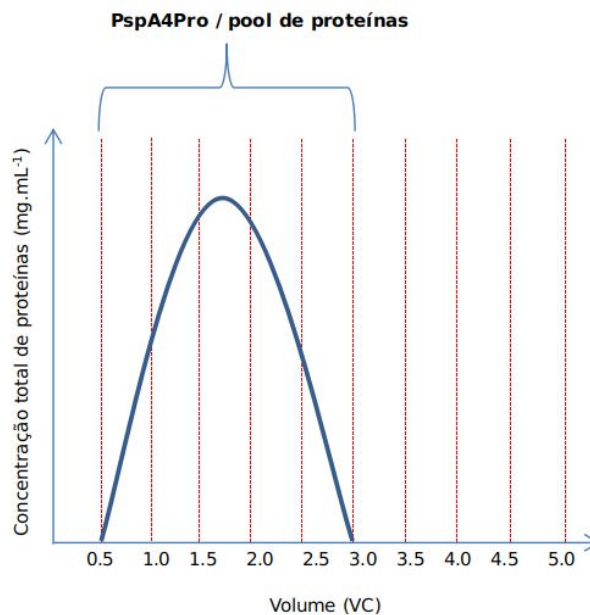


Figura 5.1. Representação do perfil de eluição  $Q_2$  para os experimentos realizados com  $ECC_B$ . O volume total eluído foi dividido em dez subfrações, cada uma com 0,5 volume de coluna.

### 5.3.3. Quantificação de proteínas e determinação de pureza de PspA4Pro

Alíquotas de 10 mL de  $ECC_A$ ,  $ECC_B$ ,  $Q_1$ ,  $Q_2$  e  $Q_3$  foram coletadas durante cada CTA executada. Os volumes totais dessas frações foram medidos e as concentrações totais de proteínas foram determinadas pelo método de Lowry (LOWRY et al., 1951), permitindo calcular a massa total de proteínas (MTP) de cada amostra. A pureza (P) de PspA4Pro foi determinada por análise densitométrica das bandas obtidas usando SDS-PAGE a 12% contendo 2-mercaptoetanol (LAEMMLI et al., 1970). Foi empregado um densitômetro Bio-Rad GS-800 calibrado. As imagens foram analisadas usando o software *Image Lab* v. 6.0. A pureza relativa de PspA4Pro foi calculada como a porcentagem da banda identificada em relação à soma das densidades de todas as outras bandas de proteínas observadas (FIGUEIREDO et al., 2017).

Os mesmos procedimentos analíticos descritos acima foram aplicados a cada subfração “i” obtida pelos subfracionamentos de  $Q_2$  em CTA-VII, CTA-VIII e CTA-IX, fornecendo os valores de  $MTP_i$  e  $P_i$ . Nesses casos, MTP e P das frações completas de  $Q_2$  foram calculados de acordo com as equações 5.8 e 5.9. A recuperação de PspA4Pro ( $REC_i$ ) também foi calculada para cada subfração, como mostra a equação 5.10, e a recuperação

global de PspA4Pro (G-REC) foi calculada por uma relação entre as massas de PspA4Pro na fração Q<sub>2</sub> e na solução de entrada (ECC<sub>B</sub>), conforme descrito pela equação 5.11.

$$MTP = \sum_{i=1}^{10} MTP_i \quad (5.8)$$

$$P = \frac{\sum_{i=1}^{10} MTP_i \cdot P_i}{MTP} \quad (5.9)$$

$$REC_i = \frac{MTP_i \cdot P_i}{MTP \cdot P} \quad (5.10)$$

$$G-REC = \frac{MTP \cdot P}{MTP_{ECC_B} \cdot P_{ECC_B}} \quad (5.11)$$

Perfis de eluição alternativos poderiam ser obtidos se as subfrações subseqüentes fossem misturadas, fornecendo pools de proteínas com purezas e recuperações médias. As Equações 5.12 e 5.13 mostram como as recuperações e as purezas de PspA4Pro foram calculadas para esses pools alternativos compostos por N subfrações.

$$REC_{1:N} = \frac{\sum_{i=1}^N MTP_i \cdot P_i}{MTP \cdot P} \quad (5.12)$$

$$P_{1:N} = \frac{\sum_{i=1}^N MTP_i \cdot P_i}{\sum_{i=1}^N MTP_i} \quad (5.13)$$

Como o fluxo da fase móvel era conhecido, também foi possível calcular o tempo de eluição ( $t_i$ ) para cada subfração “i”. Assim, o tempo total de eluição ( $\Delta t$ ) de um perfil de eluição alternativo com N elementos foi calculado como exposto na equação 5.14, e as equações 5.12 e 5.13 foram então representadas nos formatos das equações 5.15 e 5.16, onde “t” variou de 0 a  $\Delta t$  com intervalos de  $t_i$ . Numa situação hipotética em que  $t_i - t_{i-1} \approx 0$  é possível

dizer que a eluição  $Q_2$  seriam dividida em  $\theta$  subfragmentos, sendo que  $\theta \approx \infty$ . Então, as equações 5.15 e 5.16 poderiam ser representadas nos formatos das equações 5.17 e 5.18.

$$\Delta t = t_1 + t_2 + t_3 + \dots + t_N \quad (5.14)$$

$$REC_{0:\Delta t} = \frac{\sum_{t=0}^{\Delta t} MTP_i \cdot P_i}{MTP \cdot P} \quad (5.15)$$

$$P_{0:\Delta t} = \frac{\sum_{t=0}^{\Delta t} MTP_i \cdot P_i}{\sum_{t=0}^{\Delta t} MTP_i} \quad (5.16)$$

$$REC(t) = \frac{\int_0^t MTP(t) \cdot P(t) \cdot dt}{MTP \cdot P} \quad (5.17)$$

$$P(t) = \frac{\int_0^t MTP(t) \cdot P(t) \cdot dt}{\int_0^t MTP(t) \cdot dt} \quad (5.18)$$

#### 5.3.4. Modelagem e simulação

Os procedimentos para modelagem e simulação da CTA foram executados no software EMSO™ (*Environment for Modeling Simulation and Optimization*), um simulador orientado a equações (SOARES e SECCHI, 2003). Foram utilizadas as equações da isoterma de ação em efeito estérico (SMA) e do modelo do equilíbrio dispersivo (MED). As derivadas espaciais foram aproximadas pelo método das diferenças finitas, com 200 elementos, e as derivadas de tempo foram resolvidas pelo simulador utilizando um solucionador de uso geral para sistemas de equações diferenciais-algébricas de alta complexidade.

### *A) Estimativas de parâmetros de PspA4Pro e das impurezas proteicas*

A primeira caixa da Figura 5.2 mostra como os parâmetros foram estimados. De acordo com a etapa 1, os parâmetros fixos do sistema foram medidos ou fornecidos. Os parâmetros  $L$  e  $u$  foram obtidos a partir de condições de operação do sistema. Os parâmetros  $N_t$  e  $\epsilon_T$  foram determinados experimentalmente, utilizando um pulso de 0,5 mL de acetona a 1% (v.v<sup>-1</sup>) diluída em água ultrapura, a uma vazão de 2,71 mL.min<sup>-1</sup>. O número de pratos teóricos ( $N_t$ ) foi calculado usando a Equação 5.3, enquanto  $\epsilon_T$  foi determinada usando as Equações 5.4 e 5.5. O parâmetro  $\Lambda$  da isoterma SMA para a resina Q-Sepharose foi fornecido por HUUK et al. (2014).

De acordo com o passo 2, CTA-I, CTA-II e CTA-III processaram o ECC<sub>A</sub> e os dados obtidos foram utilizados para estimar os parâmetros dos pools de proteínas eluídas em Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub>. Um total de quatro picos foram identificados em cada cromatografia. Dois (chamados 1-I e 1-II) para a eluição Q<sub>1</sub> e dois (chamados 2 e 3) para as eluições Q<sub>2</sub> e Q<sub>3</sub>. Para fins de modelagem e simulação, foi considerado que cada pico era composto por um grande número de proteínas desconhecidas e distintas, mas que apresentavam a mesma afinidade química com a resina Q-Sepharose, uma vez que foram eluídas juntas. Assim, cada pico foi associado a um pool de “i” proteínas com características químicas particulares (CHAN et al., 2008). Em todas as simulações, a resina cromatográfica foi considerada inicialmente limpa, com todos os sítios ativos ligados ao contra íon padrão. As concentrações de entrada ( $c_{ie}$ ) de cada um dos pools de “i” proteínas no ECC<sub>A</sub> foram fornecidas, uma vez que as massas de proteínas de cada eluição foram medidas durante o experimento e o volume de ECC<sub>A</sub> também era conhecido. Para a eluição Q<sub>1</sub>, uma vez que os pools 1-I e 1-II foram eluídos na mesma fração, a massa de cada pool particular foi baseada na área obtida por cada pico no cromatograma. Os valores de condutividade e absorvância a 280 nM fornecidos pelo software foram transformados em concentrações de NaCl e proteínas totais, respectivamente. Uma curva padrão construída com diferentes concentrações de NaCl foi usada no primeiro caso. Para a determinação da concentração de proteínas totais foi estabelecido um coeficiente geral de extinção molar para cada fração eluída, de modo que a massa total de proteínas quantificada pelo método de Lowry (LOWRY et al., 1951) fosse similar à massa total de proteínas de cada pico cromatográfico<sup>11</sup>. Os parâmetros da isoterma SMA ( $k_{Ads,i}$ ,  $k_{Des,i}$ ,  $V_i$  e  $\alpha_i$ )

---

<sup>11</sup> O Apêndice B explica de forma detalhada as estratégias utilizadas para disponibilizar os dados de condutividade e absorvância (280 nM) em forma de concentrações de NaCl e proteínas.

dos pools proteicos 1-I, 1-II, 2 e 3 foram estimados a partir dos dados experimentais de CTA-I, CTA-II e CTA-III pela estratégia de validação cruzada<sup>12</sup>.

Os mesmos procedimentos descritos acima também foram utilizados para estimar os parâmetros de PspA4Pro a partir dos dados experimentais de CTA-IV, CTA-V e CTA-VI, conforme verificado na etapa 3 da Figura 5.2. Como esperado, apenas um pico foi identificado nas eluições Q<sub>2</sub> dessas cromatografias (FIGUEIREDO et al., 2017), as quais continham PspA4Pro com alta pureza. Nesse caso, o perfil de concentração de proteína no pico foi determinado pelo coeficiente de extinção molar da PspA4Pro (ExpASy).

### ***B) Simulações e validações experimentais***

A segunda caixa da Figura 5.2 mostra como a CTA foi simulada e validada experimentalmente com células produtoras de PspA4Pro. De acordo com a etapa 4, a CTA-VII foi realizada conforme descrito pela seção 5.3.2 e Tabela 5.1. As subfrações de eluição de Q<sub>2</sub> foram analisadas como descrito pela seção 5.3.3 e MTP<sub>i</sub> e P<sub>i</sub> foram quantificadas. As variáveis REC<sub>1:N</sub> e P<sub>1:N</sub> experimentais foram calculadas pelas equações 5.12 e 5.13 para os pools alternativos de eluição com N variando de 2 a 6, uma vez que o pico eluído foi observado nessas subfrações (nenhuma proteína foi detectada nas subfrações 1 e 7 a 10). Em seguida, a simulação de CTA-VII foi realizada como descrito pela etapa 5. Foram fornecidos os valores dos parâmetros da isoterma SMA e c<sub>ie</sub> da CTA-VII para os pools de proteínas 1-I, 1-II, 2 e 3 e PspA4pro. Uma vez que os parâmetros de PspA4Pro e dos pools de impurezas proteicas foram acessados separadamente, foi possível obter perfis de eluição separados de PspA4Pro e das demais proteínas também eluídas no pico Q<sub>2</sub>. Assim, o pico 2 simulado também foi dividido em dez subfrações e as variáveis REC<sub>i</sub> e P<sub>i</sub> simulados foram calculadas usando os valores das áreas de PspA4Pro e do pool de proteínas eluídas em cada subfração fornecida pelas simulações. REC<sub>1:N</sub> e P<sub>1:N</sub> simuladas foram calculadas pelas equações 5.12 e 5.13 para os pools alternativos de eluições. Mais uma vez, N variou de 2 a 6, pois nenhuma proteína foi detectada nas subfrações 1 e 7 a 10. A validação foi realizada quando os valores experimentais e simulados de REC<sub>1:N</sub> e P<sub>1:N</sub> foram comparados, conforme descrito pela etapa 6 .

---

<sup>12</sup> A estratégia de validação cruzada foi descrita na seção 4.2.7 do capítulo 4.

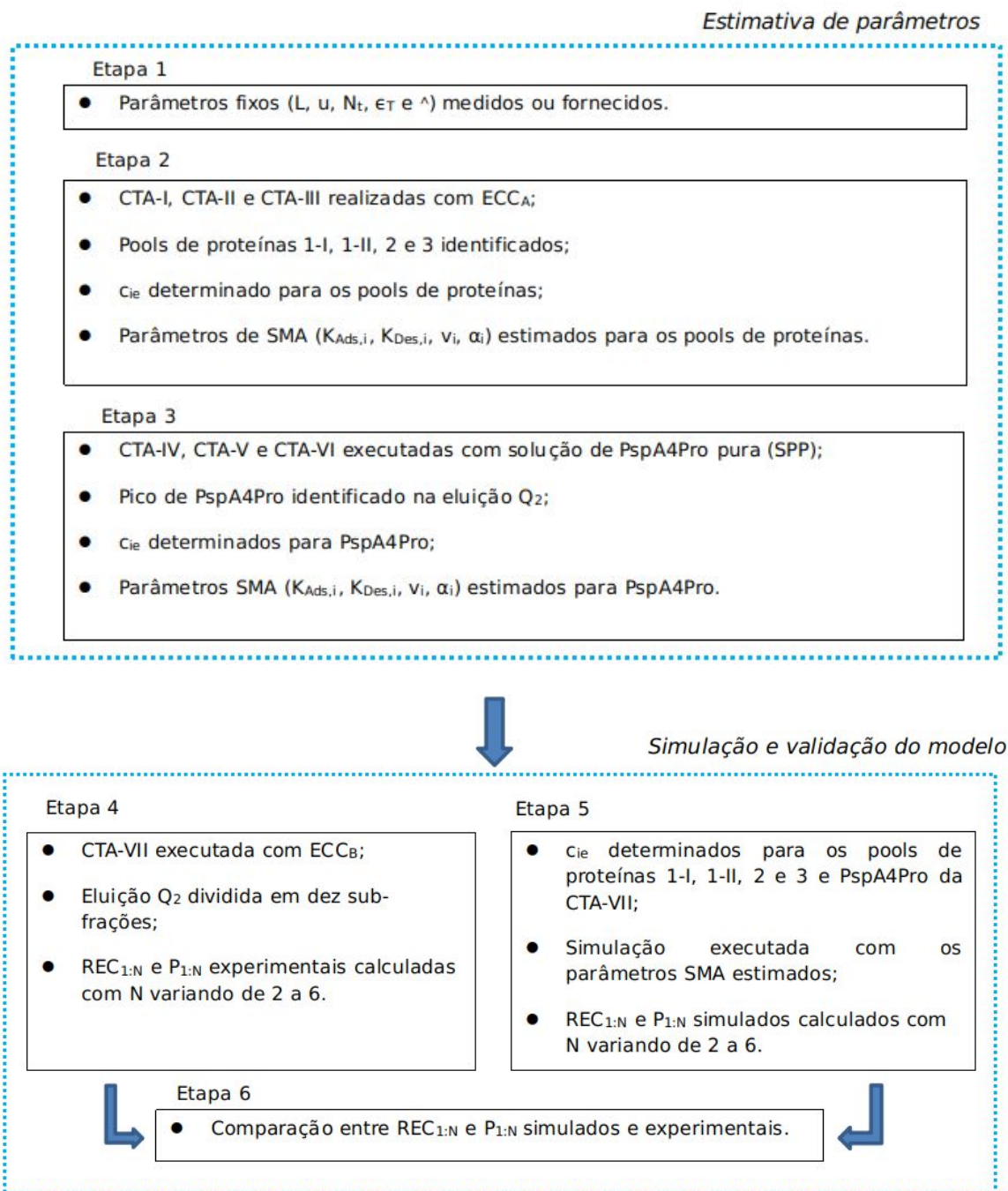


Figura 5.2. Estratégia experimental utilizada para identificar os parâmetros do modelo e simular a cromatografia de troca aniônica usando dados das corridas cromatográficas I a VII.

### C) Simulações dos perfis de eluições modificados e validações experimentais

A Figura 5.3 mostra como as condições modificadas da eluição  $Q_2$  foram testadas na simulação de CTA-VII. De acordo com o passo 1, foram aplicadas diferentes concentrações de NaCl em  $Q_2$ , variando entre 250 e 350 mM, fornecendo às simulações perfis de eluição alternativos para a PspA4Pro e o pool de proteínas do pico 2. Com essas informações, foi possível obter um perfil de  $REC(t)$  e  $P(t)$  em função da concentração de NaCl de  $Q_2$ , como mostram as equações 5.17 e 5.18. Em paralelo, de acordo com a etapa 2, as CTA-VIII e CTA-IX forneceram informações experimentais para avaliar algumas condições de operação sugeridas pelas simulações (consulte a Tabela 5.1). Para validação, as variáveis  $REC_{0:\Delta t}$  e  $P_{0:\Delta t}$  de CTA-VIII e CTA-IX foram calculadas pelas equações 5.15 e 5.16 para os pools alternativos de eluições com  $\Delta t$  variando de  $\Delta t(2)$  a  $\Delta t(6)$ , uma vez que novamente foi observado o pico eluído nas subfrações de 2 a 6. Como o tempo de aplicação de cada pool alternativo era conhecido, foi possível também comparar os dados experimentais  $REC_{0:\Delta t}$  e  $P_{0:\Delta t}$  de CTA-VIII e CTA-IX com dados simulados  $REC(t)$  e  $P(t)$  de CTA-VII, conforme descrito pela etapa 3.

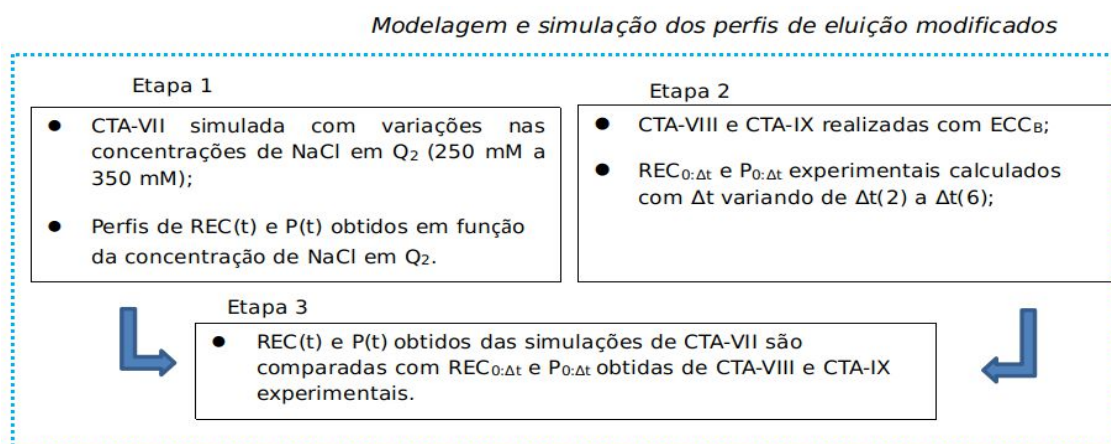


Figura 5.3. Estratégia experimental utilizada para simular o sistema a partir de variações na concentração de NaCl em  $Q_2$  (250 mM a 350 mM). Os dados utilizados pertenceram às cromatografias (CTAs) VII a IX.

## 5.4. Resultados e discussão

### 5.4.1. Visão geral das cromatografias de troca aniônica

Os extratos celulares clarificados (ECC) e a solução de PspA4Pro pura (SPP) foram obtidos de acordo com os procedimentos descritos pela Seção 5.3.1. A seguir, foram realizadas as cromatografias de troca aniônica (CTA) I a IX e as amostras foram analisadas de acordo com os procedimentos descritos pelas seções 5.3.2 e 5.3.3. A Tabela 5.1 mostra as características desses experimentos, incluindo informações sobre a massa total de proteínas (MTP) e a pureza para os diferentes materiais de partida e as frações Q<sub>1</sub>, Q<sub>2</sub> e Q<sub>3</sub>.

Tabela 5.2. Características dos materiais de partida e dos perfis de eluição (Q<sub>1</sub>, Q<sub>2</sub>, e Q<sub>3</sub>) para as corridas cromatográficas (CTA) I a IX.

CTA	Materiais de partida				Eluição Q <sub>1</sub>		Eluição Q <sub>2</sub>		Eluição Q <sub>3</sub>	
	Amostra	MTP (g)	Pureza (%)	Volume (mL)	MTP (g)	Pureza (%)	MTP (g)	Pureza (%)	MTP (g)	Pureza (%)
I		4,7 ± 0,0	-	433	0,74 ± 0,00	-	0,38 ± 0,02	-	0,064 ± 0,009	-
II	ECC <sub>A</sub>	4,7 ± 0,0	-	433	0,63 ± 0,04	-	0,35 ± 0,00	-	0,069 ± 0,002	-
III		4,7 ± 0,0	-	433	0,56 ± 0,02	-	0,40 ± 0,00	-	0,028 ± 0,001	-
IV		0,67 ± 0,02	96	350	-	-	0,51 ± 0,00	96	-	-
V	SPP	0,67 ± 0,01	95	350	-	-	0,42 ± 0,00	95	-	-
VI		0,67 ± 0,03	96	350	-	-	0,58 ± 0,01	96	-	-
VII		4,4 ± 0,2	44,4	433	0,081 ± 0,003	7,0	0,63 ± 0,01	78,4	0,57 ± 0,01	0
VIII	ECC <sub>B</sub>	4,4 ± 0,2	45,4	433	0,026 ± 0,002	11,7	0,87 ± 0,04	88,9	1,31 ± 0,03	0
IX		4,4 ± 0,2	46,2	433	0,132 ± 0,006	13,7	0,60 ± 0,03	80,1	0,34 ± 0,03	0

MTP: Massa total de proteínas

ECC: Extrato celular clarificado

SPP: Solução pura de PspA4Pro

Comparando as condições de entrada de CTA-I a CTA-IX, pode-se observar que o ECC<sub>A</sub> apresentou uma quantidade maior de MTP (4,7 g x 3) em comparação com o ECC<sub>B</sub> (4,4 g x 3), apesar da mesma quantidade de células (100 g) ter sido lisada em ambos os casos. Um comportamento semelhante foi observado no capítulo 3, no qual foram obtidos 28,45 g de MTP para células não produtoras de PspA4Pro contra 9,54 e 12,23 g para células produtoras de PspA4Pro.

As purezas das soluções de ECC<sub>A</sub> não foram medidas, pois eram compostas por um conjunto de proteínas sem PspA4Pro. A pureza média das soluções SPP foi de  $95,7 \pm 0,6$ , semelhante aos valores obtidos por FIGUEIREDO et al., 2017. A pureza média do ECC<sub>B</sub> foi de  $45,3 \pm 0,9$ , indicando um aumento quando comparada com as purezas dos extratos do capítulo 3, os quais apresentaram valores de 36,0% e 12,6% para frações provenientes de células cultivadas em batelada, mas induzidas de diferentes formas: com lactose (27 °C) e IPTG (37 °C), respectivamente (CAMPANI et al., 2016; CAMPANI et al., 2015). FIGUEIREDO et al. (2017) relataram uma pureza média de  $50,36 \pm 7,1$  para frações de ECC provenientes de células cultivadas em reatores em batelada e batelada alimentada (CAMPANI et al., 2016; HORTA et al., 2012), em ambos os casos usando lactose para indução. Este maior desempenho de pureza para células cultivadas com indução por lactose pode ser explicado pela assimilação metabólica mais lenta desse nutriente quando comparado com IPTG (HEYLAND et al., 2011; LOPES et al., 2019). Além disso, temperaturas mais altas de cultivo também contribuem para aumentar o metabolismo das células. Assim, espera-se que células cultivadas com IPTG em temperaturas mais altas (37 °C) cresçam mais rapidamente e forneçam maior produtividade de PspA4Pro, mas em contrapartida uma quantidade maior de contaminantes é produzida em paralelo, o que dificulta os processos de purificação subsequentes. Por outro lado, células cultivadas com lactose em temperaturas mais baixas (27 °C) apresentam menor produtividade de PspA4Pro, mas uma menor quantidade de impurezas proteicas é produzida.

Da comparação entre as eluições Q<sub>2</sub> para CTA-I a CTA-IX, é possível observar que a MTP média de CTA-I, CTA-II e CTA-III é de  $0,38 \pm 0,03$ , enquanto CTA-VII, CTA-VIII e CTA-IX apresentaram uma MTP média mais alta de  $0,70 \pm 0,15$ , indicando que a PspA4Pro foi eluída nestes casos, o que aumentou a concentração de proteínas na amostra e proporcionou uma recuperação global (G-REC) em torno de 30 a 32%. CTA-IV, CTA-V e CTA-VI apresentaram uma MTP média de  $0,50 \pm 0,08$  com 95-96% de pureza, proporcionando um G-REC em torno de 70 a 75%. Nestes últimos casos, a recuperação de PspA4Pro foi maior, uma vez que cargas menores de proteína (0,67 g) com solução de PspA4Pro pura (SPP) foram processadas pela coluna, reduzindo a competição pelos sítios ativos da Q-Sepharose. FIGUEIREDO et al. (2017) obtiveram uma recuperação de 66% nesta etapa. Nesse caso, o melhor desempenho pode ser explicado pelo maior volume da coluna XK 50/30 (em torno de 250 mL), aumentando também a quantidade de sítios ativos disponíveis.

Não foram observadas MTP nas eluições  $Q_1$  e  $Q_3$  de CTA-IV, CTA-V e CTA-VI, pois SPP foi purificada nesses casos e toda a PspA4Pro foi eluída na fração  $Q_2$ . A análise das frações  $Q_1$  e  $Q_3$  de CTA-I, CTA-II e CTA-III mostra que uma quantidade significativamente maior de MTP foi eluída para as frações  $Q_1$  quando comparadas com  $Q_3$  ( $0,64 \pm 0,09$  contra  $0,05 \pm 0,02$ ). Um comportamento oposto foi observado para CTA-VII, CTA-VIII e CTA-IX, as quais produziram  $0,08 \pm 0,05$  de MTP para  $Q_1$  contra  $0,74 \pm 0,51$  para  $Q_3$ . Essas diferenças entre os perfis de proteínas eluídas podem ser mais uma vez explicadas pelas diferenças genéticas entre as células A e B, o que faz com que apresentem rotas metabólicas diferentes e, conseqüentemente, diferentes perfis de proteínas no interior das células, gerando modificações nos ECCs. Quando as frações  $Q_3$  de CTA-I, CTA-II e CTA-III são comparadas, é possível observar que a MTP é inversamente proporcional à força iônica utilizada na eluição de  $Q_2$ , indicando que forças iônicas mais altas das eluições  $Q_2$  removeram mais impurezas proteicas, deixando menos impurezas a serem removidas pela terceira eluição ( $Q_3$ ). Um comportamento semelhante e ainda mais claro é observado para as frações  $Q_3$  de CTA-VII, CTA-VIII e CTA-IX. Esses resultados mostram mais uma vez que forças iônicas mais altas nas frações  $Q_2$  tendem a eluir mais impurezas proteicas além de PspA4Pro, reduzindo a pureza como já foi indicado.

#### **5.4.2. Estimativa de parâmetros e simulação da cromatografia de troca aniônica**

Esta seção mostra como os parâmetros de SMA foram estimados e como o modelo e a simulação foram realizados, conforme foi descrito pela Seção 5.3.4. e Figura 5.2. A Tabela 5.3 apresenta os parâmetros fixos e as condições iniciais utilizadas na modelagem, onde  $q_{i(1-N_t)}$  e  $c_{i(1-N_t)}$  são as concentrações iniciais de proteínas para em cada prato teórico nas fases estacionária e móvel, respectivamente,  $L$  e  $u$  são parâmetros fixos fornecidos pelo sistema, e  $\epsilon_T$  e  $N_t$  foram determinados conforme descrito pela Seção 5.2.1.

Tabela 5.3. Condições iniciais e parâmetros fixos utilizados para simular o processo de cromatografia de troca aniônica.

Nome do parâmetro	Símbolo	Unidade	Valor
Proteínas adsorvidas iniciais	$q_{i(1-N_T)}$	(g.mL <sup>-1</sup> )	0
Proteínas não adsorvidas iniciais	$C_{i(1-N_T)}$	(g.mL <sup>-1</sup> )	0
Altura do leito cromatográfico	L	(cm)	13
Velocidade da fase móvel	u	(cm.s <sup>-1</sup> )	0,042
Porosidade total da resina	$\epsilon_T$	-	0,75
Número total de pratos teóricos	$N_T$	-	190
Capacidade de saturação da resina <sup>A</sup>	$\Lambda$	(mol.L <sup>-1</sup> )	2,15

<sup>A</sup> Fornecido por HUUK et al., 2014

Os parâmetros de SMA foram fornecidos na Tabela 5.4. Eles foram estimados com os dados de CTA-I, CTA-II e CTA-III para os picos observados 1-I, 1-II, 2 e 3, cada um representando um pool de proteínas com afinidades químicas semelhantes à resina; e com os dados de CTA-IV, CTA-V e CTA-VI para o pico 2, que nesses casos eluíram somente PspA4Pro.

Tabela 5.4. Parâmetros estimados da isoterma de ação em efeito estérico (SMA) para os pools de proteínas (1-I, 1-II, 2 e 3) das cromatografias de troca aniônica CTA-I, CTA-II e CTA-III e para a PspA4Pro eluída nas corridas CTA-IV, CTA-V e CTA-VI.

Proteínas	$K_{Ads}$ (mL.g <sup>-1</sup> .s <sup>-1</sup> )	$K_{Des}$ (mL.g <sup>-1</sup> .s <sup>-1</sup> )	$K_{Eq}$	v	$\alpha$
Pool 1-I	0,0013	0,0056	0,24	1,07	2,5
Pool 1-II	0,0008	0,0034	0,24	0,87	41,7
PspA4Pro	0,0047	0,0062	0,75	1,26	6,3
Pool 2	0,0062	0,0050	1,25	1,11	151,4
Pool 3	0,0054	0,0030	1,82	1,5	169,2

\* $K_{Eq}$  representa a constante de equilíbrio ( $K_{Eq} = k_{Ads} \cdot k_{Des}^{-1}$ )

Os parâmetros de SMA se mostraram adequados para descrever os modelos matemáticos, uma vez que as proteínas com maior afinidade com a resina (Pools 2 e 3) tenderam a apresentar maiores constantes de adsorção ( $k_{Ads,i}$ ) e menores constantes de dessorção ( $k_{Des,i}$ ). O oposto foi observado para as proteínas com menor afinidade pela resina

(Pools 1-I e 1-II), explicando o menor tempo de residência na coluna. Quando  $K_{Eq}$  é analisado, nota-se um comportamento crescente de acordo com o tempo de residência das proteínas na coluna,  $K_{Eq,1A} = K_{Eq,1B} < K_{Eq,PspA4Pro} < K_{Eq,2} < K_{Eq,3}$ , mais uma vez destacando a maior afinidade das proteínas que apresentaram maiores tempos de residência. No capítulo 3 foram relatados valores de 0,16, 0,35, 0,65 e 4,42 para  $K_{Eq,1-I}$ ,  $K_{Eq,1-II}$ ,  $K_{Eq,2}$  e  $K_{Eq,3}$ , respectivamente. Esses valores variaram um pouco em relação aos apresentados nesse momento, uma vez que as células de *E. coli* de ambos os casos são geneticamente diferentes e foram cultivadas em condições distintas, o que levou à produção de proteínas com diferentes características químicas. Alguns outros trabalhos relataram valores de  $K_{Eqi}$  variando de 0,01 a 13 (KARKOV et al., 2013; HUUK et al., 2014; FARAJI et al., 2015). Quando  $K_{Eq,PspA4Pro}$  é comparado com  $K_{Eq,2}$  (o conjunto de proteínas eluídas no mesmo momento) é possível inferir que o tempo de residência de PspA4Pro é menor, uma vez que  $K_{Eq,PspA4Pro} < K_{Eq,2}$ . O capítulo 3 relatou um valor de  $K_{Eq,PspA4Pro}$  de 0,83, semelhante ao valor encontrado neste estudo (0,75). Essa similaridade é esperada, já que em ambos os casos foram utilizadas SPP com características semelhantes nas cromatografias. Os valores de  $v_i$  tenderam a ser mais altos para proteínas que apresentaram maiores tempos de residência na coluna, o que faz sentido, uma vez que essas proteínas apresentam um maior número de sítios ativos com a resina. Trabalhos recentes relataram valores de  $v_i$  variando entre 0,06 e 3,0 (KARKOV et al., 2013; HUUK et al., 2014; FARAJI et al., 2015; capítulo 3). Os valores de  $\alpha_i$  apresentaram diferentes ordens de magnitude, variando entre 2 e 170. Esse comportamento também foi observado no capítulo 3. Uma diferença ainda maior foi observada por FARAJI et al. (2015), relatando valores de 1283 e 0,16 para adsorção de lactoperoxidase e lactoferrina em SP-Sepharose.

A Figura 5.4 mostra os perfis experimentais e simulados de CTA-I, CTA-II e CTA-III utilizando os parâmetros das Tabelas 5.3 e 5.4. Quando os perfis do pico intermediário são comparados para cada CTA, é possível observar que a força iônica mais intensa (350 mM de NaCl) forneceu um pico mais estreito (CTA-III, gráfico “c”), enquanto a força iônica menos intensa (250 mM de NaCl) proporcionou um pico mais amplo (CTA-II, gráfico “b”). Um pico com largura intermediária (CTA-I) foi produzido quando uma força iônica intermediária (300 mM de NaCl, gráfico “a”) foi aplicada. Esse comportamento é esperado, uma vez que maiores forças iônicas aumentam a competição de NaCl com a proteína na ligação dos sítios ativos de Q-Sepharose, reduzindo o tempo de residência da proteína na coluna. Através da análise da Figura 5.4, é possível observar que o perfil completo comentado acima foi representado com sucesso pela simulação.

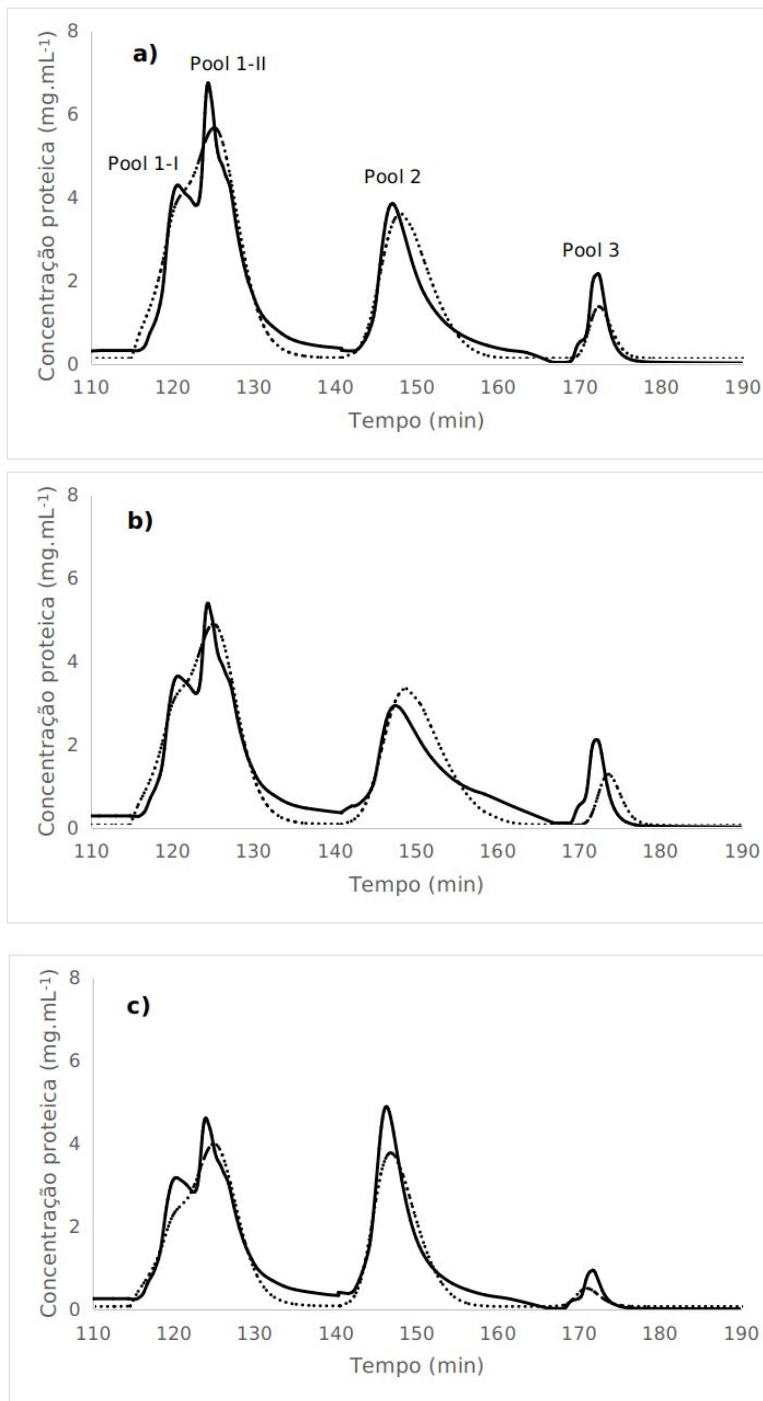


Figura 5.4. Comparações entre as cromatografias de troca aniônica CTA-I, CTA-II e CTA-III experimentais (linhas sólidas) e simuladas (linhas pontilhadas) representadas nos quadros “a)”, “b)” e “c)”, respectivamente: Os experimentos foram realizados com extrato celular clarificado (ECC) do cultivo A (células não produtoras de PspA4Pro cultivadas a 37 °C, com indução por lactose). Os primeiros e últimos picos de todas as corridas foram obtidos pela aplicação de 5 VC de eluições com 150 e 1000 mM de NaCl, respectivamente. Os segundos picos foram obtidos pela aplicação de 5 VC de eluições com 300, 250 e 350 mM de NaCl, respectivamente.

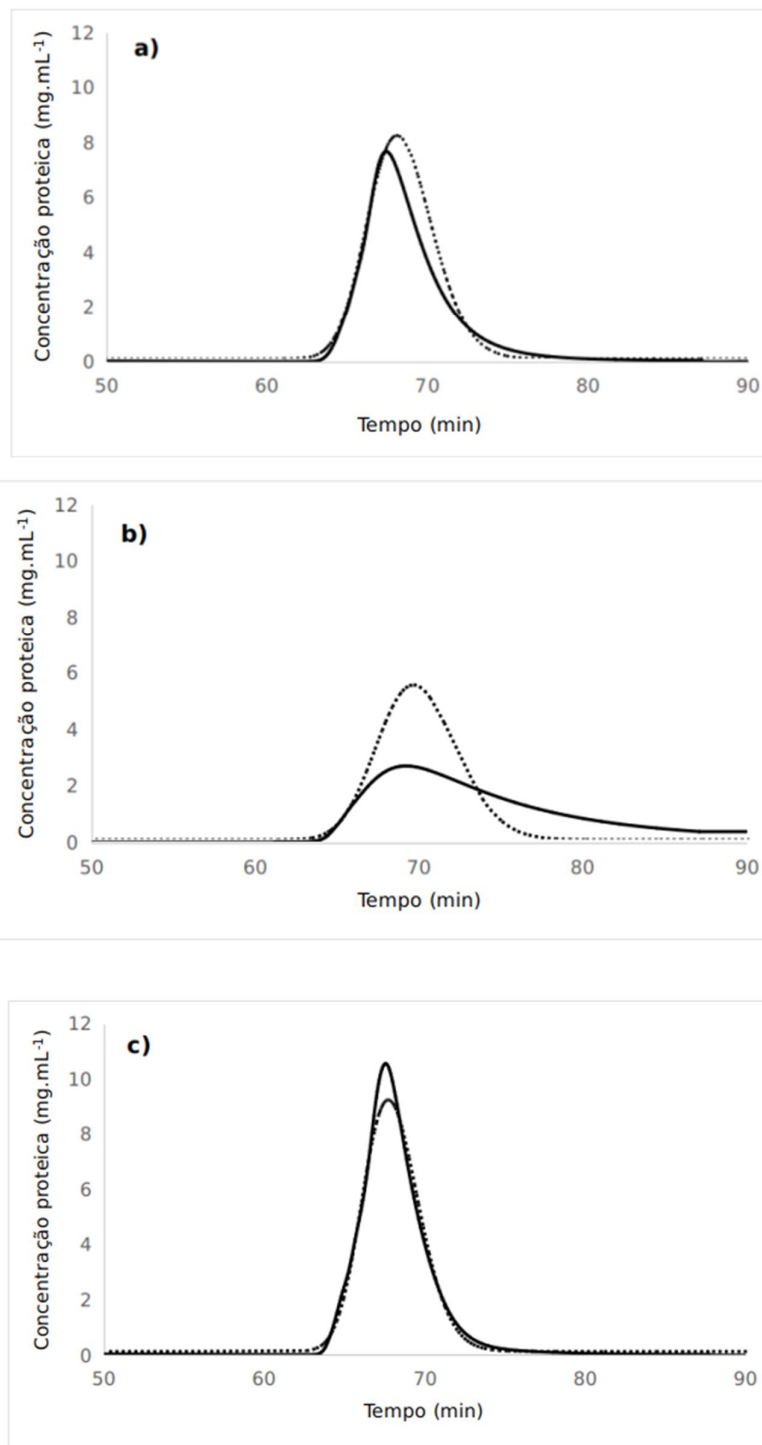


Figura 5.5. Comparação entre os cromatogramas experimentais (linhas sólidas) e simulados (linhas pontilhadas) das cromatografias de troca aniônica CTA-IV, CTA-V e CTA-VI: Os experimentos foram realizados com solução de PspA4Pro pura (SPP). As primeiras e últimas eluições não foram detectadas. As segundas eluições foram obtidas pela aplicação de 5 VC de solução com 300, 250 e 350 mM de NaCl, respectivamente.

A Figura 5.5 mostra o perfis experimentais e simulados de CTA-IV, CTA-V e CTA-VI usando os parâmetros de PspA4Pro das Tabelas 5.3 e 5.4. Como não foi observado MTP para o primeiro e o último pico (consulte a Tabela 5.2), apenas o segundo pico foi identificado. Através de análise da Figura 5.5 é possível verificar que o perfil completo de PspA4Pro foi simulado com sucesso. A Figura 5.6 mostra os dados experimentais de CTA-VII e os resultados para as simulações realizadas usando os parâmetros fornecidos nas Tabelas 5.3 e 5.4. A Tabela 5.5 faz uma comparação entre os tempos médios de residência e as concentrações máximas de proteína ( $C_{P,max}$ ) dos cromatogramas experimentais e simulados. É possível notar que o pico 2 apresentou os menores desvios padrões entre dados experimentais e simulados tanto para o tempos de residência quanto para  $C_{P,max}$ , o que indica a qualidade da simulação, já que esse é o pico que elui PspA4Pro. Maiores desvios padrões de  $C_{P,max}$  foram encontrados nos picos 1-I, 1-II e 3. Esse comportamento é esperado, uma vez que os parâmetros de SMA dos pools de proteínas foram estimados por células não produtoras de PspA4Pro (CTA-I, CTA-II e CTA- III), enquanto o ECC de CTA-VII foi obtido por células produtoras de PspA4Pro (ECC<sub>B</sub>). Essas diferenças entre as células de ECC<sub>A</sub> e ECC<sub>B</sub> podem levar a pools de proteínas contendo diferentes tipos impurezas proteicas. Essas condições aliadas aos erros dos procedimentos analíticos (Lowry e SDS-PAGE) foram provavelmente os principais fatores responsáveis por esses desvios.

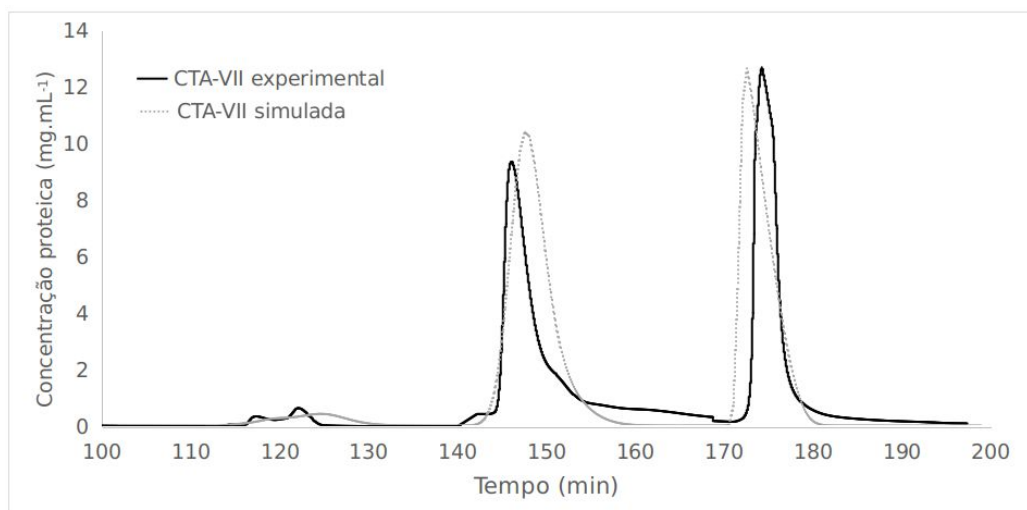


Figura 5.6. Comparação entre os cromatogramas experimentais e simulados de CTA-VII. Os dados experimentais foram obtidos com extrato celular clarificado (ECC) do cultivo B (células produtoras de PspA4Pro induzidas por lactose a 37 ° C). A simulação foi realizada considerando as MTPs e purezas informadas na Tabela 5.2 (CTA-VII), as condições iniciais e parâmetros fixos da Tabela 5.3 e os parâmetros estimados da Tabela 5.4 para todos os pools de proteínas e PspA4Pro.

Tabela 5.5. Comparação dos tempos de residências e concentração máxima de proteínas ( $C_{P,max}$ ) para os picos 1-I, 1-II, 2/PspA4Pro e 3 dos cromatogramas experimentais e simulados de CTA-VII.

<b>Pools de proteínas</b>		<b>1-I</b>	<b>1-II</b>	<b>2 / PspA4Pro</b>	<b>3</b>
<b>Tempo de residência (min)</b>	Experimental	117,4	122,1	146,1	174,3
	Simulado	119,0	123,5	146,0	171,0
	Desv. Pad. (%)	1,0	0,8	0,0	1,3
<b><math>C_{P,max}</math> (mg.mL<sup>-1</sup>)</b>	Experimental	0,37	0,66	9,35	12,69
	Simulado	0,24	0,35	8,87	10,74
	Desv. Pad. (%)	24,8	33,2	3,6	10,9

Tabela 5.6. Valores experimentais e simulados de recuperações ( $REC_{1:N}$ ) e purezas ( $P_{1:N}$ ) de CTA-VII seguidos pelos valores experimentais de recuperações ( $REC_{0:\Delta t}$ ) e purezas ( $P_{0:\Delta t}$ ) de CTA-VIII e CTA-IX para os perfis de eluições alternativos compostos por N subfrações.

<b>Perfil de eluição alternativo</b>	<b>N</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
	<b><math>\Delta t</math> (min)</b>	<b>5,70</b>	<b>8,6</b>	<b>11,4</b>	<b>14,25</b>	<b>17,1</b>
CTA-VII experimental	$REC_{1:N}$ (%)	25,2	83,0	98,3	100,0	100,0
	$P_{1:N}$ (%)	84,7	85,9	84,9	83,3	78,4
CTA-VII simulado	$REC_{1:N}$ (%)	45,5	87,8	99,3	99,8	99,9
	$P_{1:N}$ (%)	92,3	88,0	82,0	79,1	78,5
CTA-VIII experimental	$REC_{0:\Delta t}$ (%)	12,2	47,0	80,3	93,7	100,0
	$P_{0:\Delta t}$ (%)	76,2	84,1	88,8	89,0	88,9
CTA-IX experimental	$REC_{0:\Delta t}$ (%)	42,6	92,2	98,8	99,8	100,0
	$P_{0:\Delta t}$ (%)	80,3	82,7	81,5	80,6	80,1

\*  $\Delta t$  representa o tempo de eluição para de perfil de eluição alternativo, com N variando de 2 a 6.

\* Perfis de eluição alternativos de CTA-VIII e CTA-IX serão comentados na seção 5.4.3.

A Tabela 5.6 faz uma comparação entre os perfis de eluição alternativos do pico 2 (experimentais e simulados), os mesmos mostrados na Figura 5.6. Primeiro, os valores experimentais e simulados de  $MTP_i$  e  $P_i$  foram calculados de acordo com as seções 5.3.3 e 5.3.4.B, respectivamente. Em seguida, os perfis de eluição alternativos foram propostos para  $N$  variando de 2 a 6. Para cada  $\Delta t$ ,  $REC_{1:N}$  e  $P_{1:N}$  experimentais e simulados foram calculados pelas equações 5.12 e 5.13, respectivamente.

Quando os perfis de eluição alternativos experimentais de CTA-VII são analisados, é possível observar que quando  $N = 2$  uma pequena recuperação ( $REC_{1:2}$ ) é obtida, mostrando que  $MTP_2$  apresenta um baixo valor. Quando  $N = 3$ ,  $REC_{1:3}$  aumenta consideravelmente e  $P_{1:3}$  aumenta pouco, mostrando que  $MTP_3$  e  $P_3$  apresentam valores maiores que  $MTP_2$  e  $P_2$ . Quando  $N$  varia de 3 a 6, observa-se que  $REC_{1:N}$  ainda aumenta, mas  $P_{1:N}$  passa a diminuir consideravelmente. Esse comportamento é esperado, uma vez que a PspA4Pro elui no início do pico, enquanto que o pool de impurezas proteicas tende a eluir posteriormente. Como a MTP completa foi eluída entre as subfrações 2 a 6,  $P_{1:6}$  apresenta um valor semelhante à pureza média do pico 2 exposto na Tabela 5.2, referente à CTA-VII (78,4%). A partir da análise experimental de  $P_{1:N}$ , espera-se portanto a seguinte relação entre  $P_i$ :  $P_2 < P_3 > P_4 > P_5 > P_6$ , a qual pode ser verificada nas imagens de SDS-PAGE das subfrações 2 a 6 expostas na Figura 5.7. Os perfis de eluição alternativos simulados de CTA-VII não detectaram o comportamento da pureza experimental na transição entre  $N = 2$  e  $N = 3$ , uma vez que no caso experimental  $P_{1:3} > P_{1:2}$  e no caso da simulação  $P_{1:3} < P_{1:2}$ . Apesar disso, a estratégia de modelagem e simulação proposta fornece uma boa descrição do perfil de PspA4Pro e do conjunto de proteínas eluídas na cromatografia, uma vez que a simulação detectou de forma adequada o aumento nas recuperações,  $REC_{1:N}$ , quando  $N$  variou de 2 a 6, e a redução nas purezas,  $P_{1:N}$ , quando  $N$  variou de 3 a 6, deixando claro novamente que PspA4Pro elui antes. Além disso, a maioria dos valores simulados e experimentais de  $REC_{1:N}$  e  $P_{1:N}$  foram semelhantes.

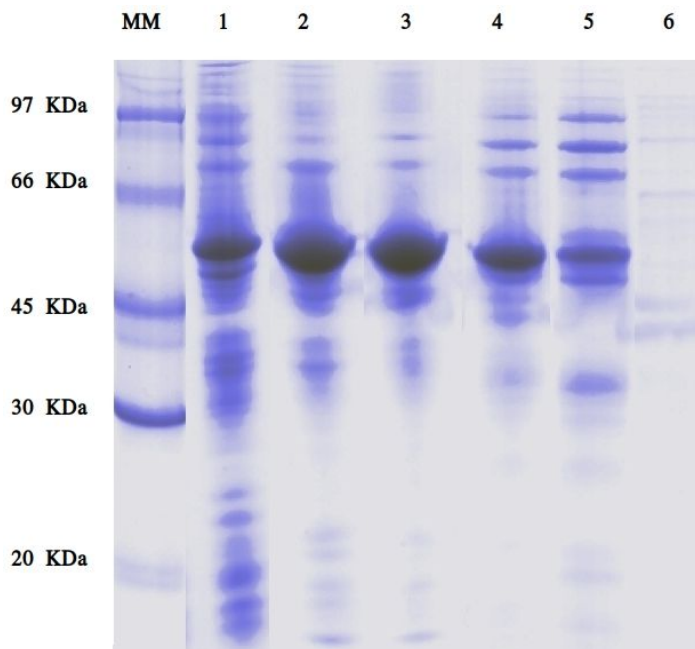


Figura 5.7: Géis SDS-PAGE do extrato celular clarificado ( $ECC_B$ ) que apresentava PspA4Pro recombinante junto com outras proteínas e dos subfragmentos 2 a 6 da eluição  $Q_2$  de CTA-VII. MM) Marcador molecular contendo fosforilase (97 KDa), albumina (66 KDa), ovalbumina (45 KDa), anidro (30 KDa) e tripsina (20 KDa). 1)  $ECC_B$  obtido por células produtoras de PspA4Pro cultivadas a 37 °C com indução por lactose. 2) subfragmento 2 de  $Q_2$ . 3) subfragmento 3 de  $Q_2$ . 4) subfragmento 4 de  $Q_2$ . 5) subfragmento 5 de  $Q_2$ . 6) subfragmento 6 de  $Q_2$ .

#### 5.4.3. Simulações e validações experimentais dos perfis de eluição modificados

Esta seção mostra como a simulação foi empregada para avaliar modificações nos perfis de eluição  $Q_2$  em um conjunto de CTAs virtuais, possibilitando a identificação de condições de operação mais promissoras. As condições iniciais ( $q_{ie}$  e  $c_{ie}$ ) e os parâmetros estimados da CTA-VII foram mantidos. Em seguida, novas simulações foram realizadas variando as concentrações de NaCl em  $Q_2$  de 250 a 350 mM, conforme descrito pela seção 5.3.4.C e Figura 5.3. A Figura 5.8 mostra como  $REC(t)$  e  $P(t)$ , representados pelas equações 5.17 e 5.18, foram descritos nessas condições. As Figuras 5.9 e 5.10 fornecem separadamente os mesmos dados de  $REC(t)$  e  $P(t)$  da Figura 5.8, mas em escalas diferentes para facilitar a visualização.

A análise da Figura 5.9 mostra que a recuperação  $REC(t)$  reduz drasticamente após 12 minutos de eluição, e que uma recuperação de 100% é obtida após 15 minutos (2,6 VC) da aplicação da eluição  $Q_2$  para todas as concentrações de NaCl simuladas. Além disso, também

é possível notar que em tempos menores de 15 minutos a variação da força iônica gera uma intensa modificação no perfil de  $REC(t)$ , já que concentrações mais altas de NaCl induzem a formação de perfis mais inclinados. Esse é um comportamento esperado, uma vez que já foi demonstrado que eluições com maiores forças iônicas produzem picos mais estreitos, incrementando a variação de  $MTP(t)$  e  $REC(t)$ , como mostra a equação 5.17.

De acordo com a Figura 5.10, foram observados dois perfis diferentes de  $P(t)$ . Quando o tempo é menor que 5,5 minutos,  $P(t)$  aumenta com o tempo de eluição e com a concentração de NaCl de  $Q_2$ . Quando o tempo de eluição é superior a 5,5 minutos,  $P(t)$  diminui com o tempo e com a concentração de NaCl de  $Q_2$ . A razão para esse comportamento pode ser melhor visualizada pelas simulações expostas na Figura 5.11, as quais descrevem separadamente o perfil de PspA4Pro e do pool de impurezas proteicas eluídos por duas concentrações diferentes de NaCl: 250 mM e 350 mM. Quando os picos simulados são avaliados antes de 5,5 minutos de eluição, observa-se um perfil de concentração maior de PspA4Pro (350 mM) do que de PspA4Pro (250 mM), proporcionando uma maior pureza para o primeiro caso. Quando o tempo é fixado em 5,5 minutos, é atingida a maior concentração de PspA4Pro (350 mM) (cerca de  $9 \text{ mg.mL}^{-1}$ ), enquanto a concentração de PspA4Pro (250 mM) ainda é baixa (cerca de  $5 \text{ mg.mL}^{-1}$ ). Após 5,5 minutos, a concentração de PspA4Pro (350 mM) diminui abruptamente, reduzindo a pureza muito rapidamente. Para o caso de 250 mM, há um atraso na diminuição da concentração de PspA4Pro, a qual começa a ocorrer somente após 7,5 minutos de eluição, seguido de uma diminuição da pureza em uma taxa menos intensa, uma vez que o pico é mais amplo. Assim, a pureza obtida por um pico de eluição de 250 mM reduz de forma mais lenta quando comparada com o pico de 350 mM. Esse perfil pode ser claramente visualizado na Figura 5.10, a qual representa uma rápida inversão de pureza (5,5 minutos) para eluições  $Q_2$  com concentrações mais altas de NaCl, enquanto que a inversão é menos intensa para as eluições  $Q_2$  com concentrações mais baixas de NaCl. Uma interpretação adicional da Figura 5.11 também mostra que as eluições  $Q_2$  com concentrações mais baixas de NaCl tendem a aumentar a distância entre os picos de PspA4Pro e do pool de proteínas eluídas no mesmo momento, permitindo acessar subfragmentos mais puros, uma vez que os picos são mais largos.

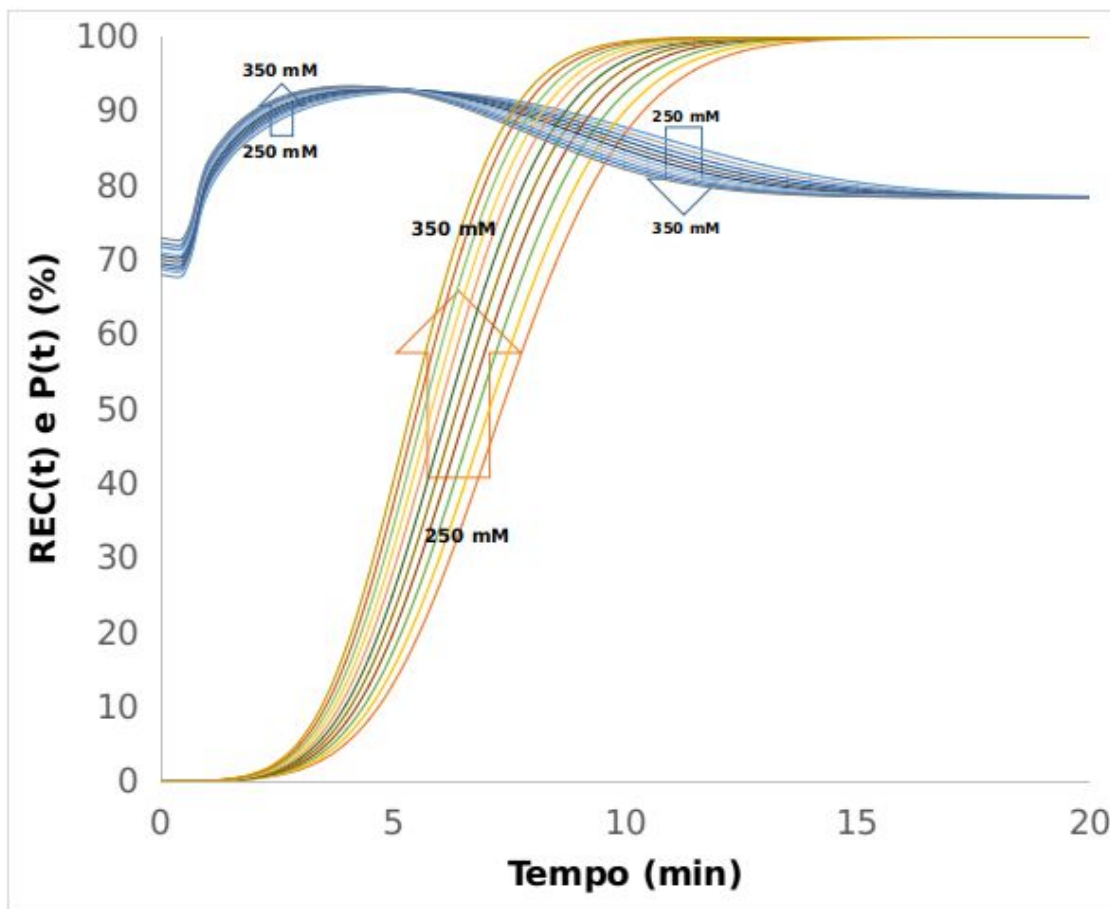


Figura 5.8. Valores simulados das recuperações de PspA4Pro (linhas amarelas) e purezas (linhas azuis) para os perfis alternativos de eluições baseados em variações da força iônica da eluição central (250 mM a 350 mM).

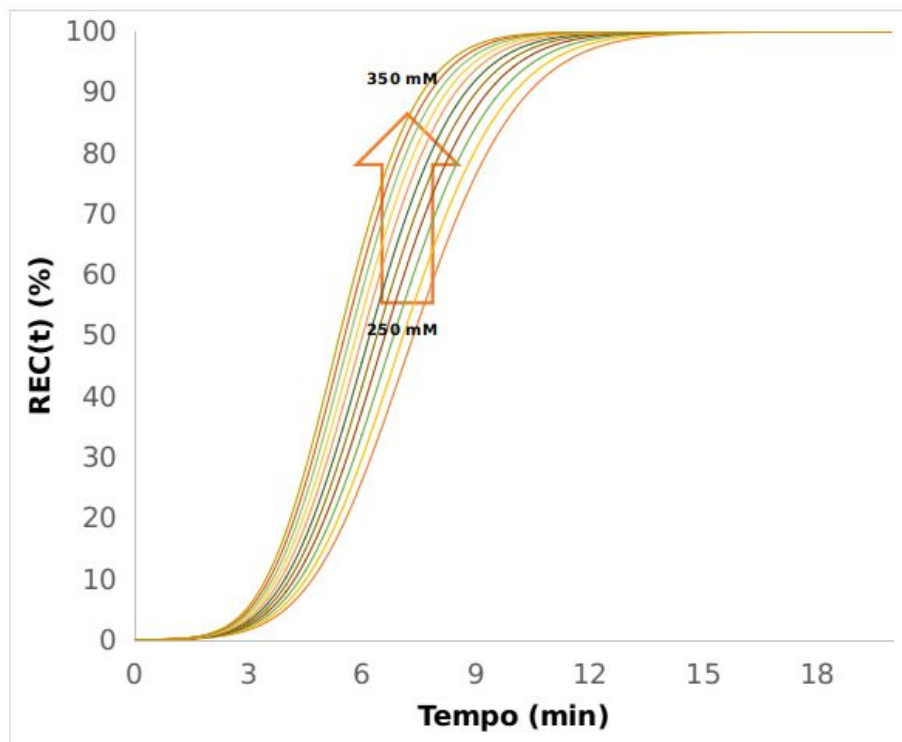


Figura 5.9. Valores simulados das recuperações de PspA4Pro REC(t) para os perfis alternativos de eluições baseados em variações da força iônica da eluição central (250 mM a 350 mM).

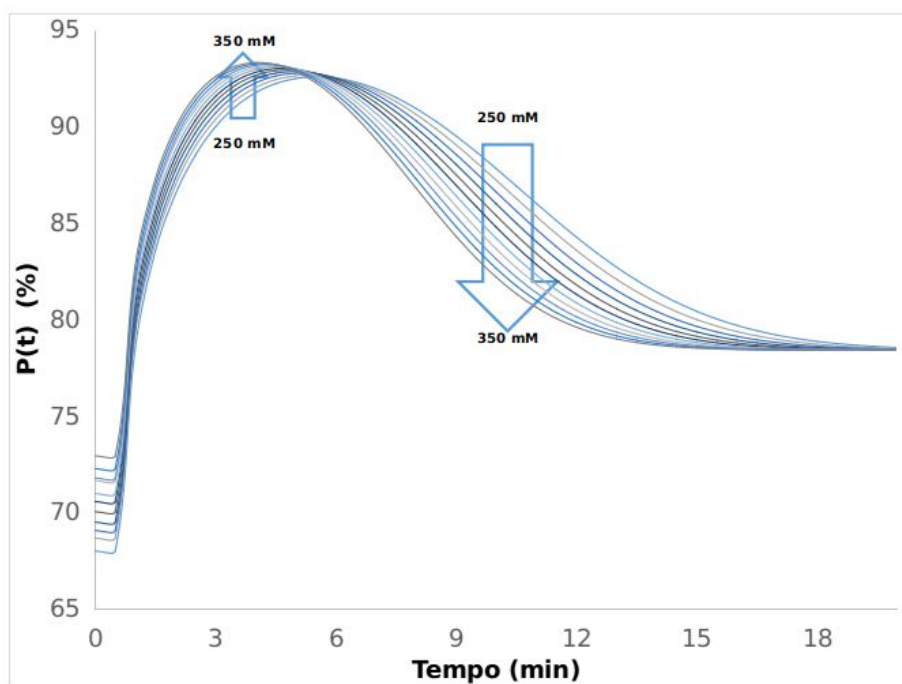


Figura 5.10. Valores simulados das purezas P(t) para os perfis alternativos de eluições baseados em variações da força iônica da eluição central (250 mM a 350 mM).

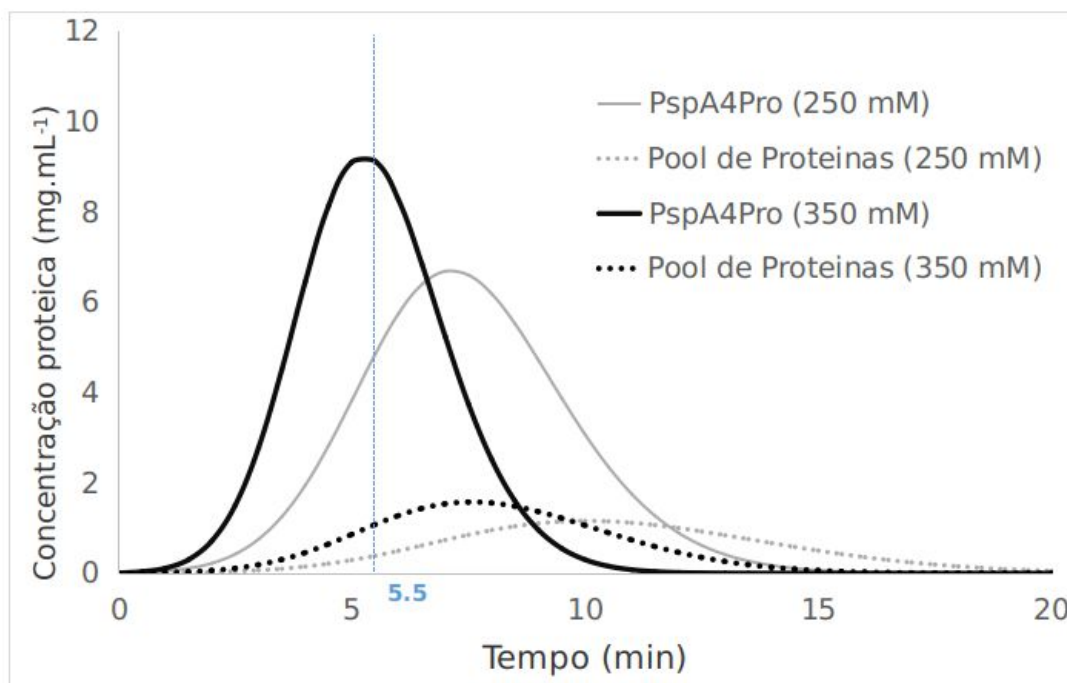


Figura 5.11. Comparação dos cromatogramas simulados de CTA-VII quando concentrações de 250 mM e 350 mM são aplicadas em Q<sub>2</sub>. Para cada caso, picos de PspA4Pro e do pool de proteínas eluídas no mesmo momento foram descritos separadamente.

As tendências dos perfis de eluição simulados representados nas Figuras 5.8 a 5.11 foram validadas executando CTA-VIII e CTA-IX, de acordo com o procedimento descrito pela seção 5.3.1 e pela Tabela 5.1. As concentrações de NaCl de Q<sub>2</sub> nesses casos foram 250 mM e 350 mM, respectivamente. Os dados dessas corridas cromatográficas são mostrados na Figura 5.12, onde é possível se observar as diferenças entre os perfis dos picos Q<sub>2</sub>. Como esperado, a CTA-VIII apresentou um pico Q<sub>2</sub> mais amplo que a CTA-IX. As Figuras 5.13 e 5.14 mostram as imagens SDS-PAGE das subfrações de Q<sub>2</sub> das CTA-VIII e CTA-IX, respectivamente. Quando comparadas, é possível observar que a PspA4Pro eluiu mais cedo para CTA-IX (350 mM de NaCl) do que para CTA-VIII (250 mM de NaCl). Essa condição foi prevista pelas simulações da Figura 5.11.

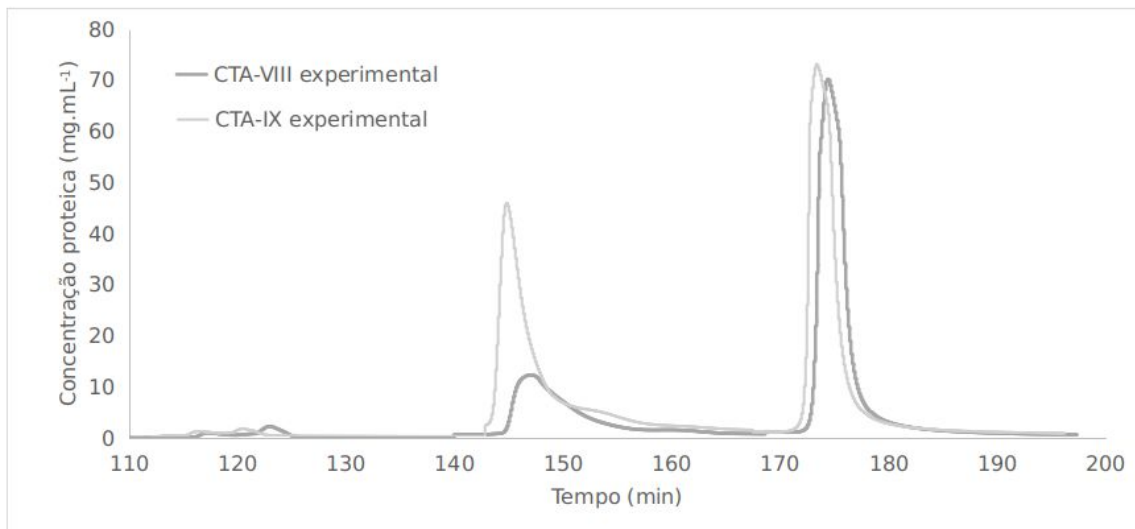


Figura 5.12. Comparação entre cromatogramas experimentais: CTA-VIII e CTA-IX realizados com extrato celular clarificado (ECC) do cultivo B (células produtoras de PspA4Pro cultivadas a 27 °C com indução por lactose). As eluições foram realizadas com 150 mM, 250 mM e 1000 mM de NaCl e 150 mM, 350 mM e 1000 mM de NaCl, respectivamente.

Os resultados da simulação também podem ser ratificados pelas eluições alternativas experimentais de CTA-VIII e CTA-IX mostradas na Tabela 5.6. Quando os perfis  $REC_{0:\Delta t}$  são comparados, maiores valores são observados para CTA-IX em todos os  $\Delta t$ . Trata-se de um comportamento esperado, uma vez que foi demonstrado por simulações que as eluições  $Q_2$  com concentrações mais altas de NaCl fornecem picos mais estreitos, aumentando a taxa de recuperação nesses casos. Quando os perfis  $P_{0:\Delta t}$  são comparados para  $N = 2$  (e  $\Delta t = 5,70$  min),  $P_{0:\Delta t}$  da CTA-IX é maior que a da CTA-VIII. Trata-se também de um comportamento esperado, já que nesse caso  $\Delta t$  é muito próximo a 5,5 minutos, referente à região da Figura 5.10 onde as purezas são maiores para as frações obtidas com maiores forças iônicas. Quando  $N > 2$  (e  $\Delta t > 5,70$  min) o limite de inversão de pureza representado pela Figura 5.10 é ultrapassado e, para todos os outros perfis de eluição alternativos, os valores de  $P_{0:\Delta t}$  de CTA-VIII são maiores que os de CTA-IX.

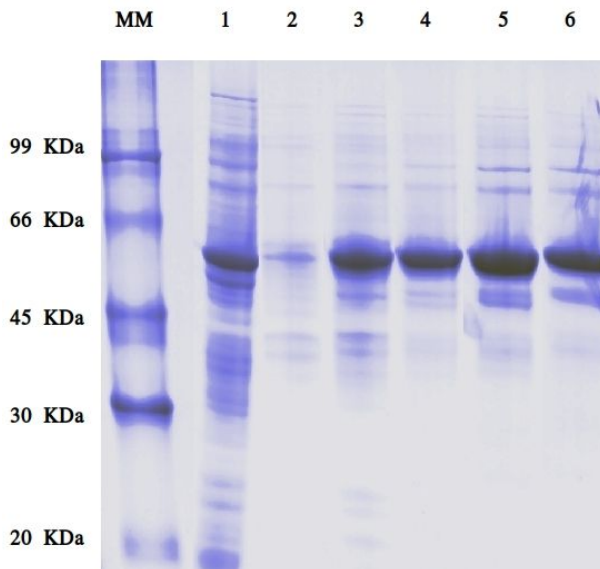


Figura 5.13: Géis SDS-PAGE do extrato celular clarificado ( $ECC_B$ ) que apresentava PspA4Pro recombinante junto com outras proteínas e dos subfragmentos 2 a 6 da eluição  $Q_2$  de CTA-VIII. MM) Marcador molecular contendo fosforilase (97 KDa), albumina (66 KDa), ovalbumina (45 KDa), anidro (30 KDa) e tripsina (20 KDa). 1)  $ECC_B$  obtido por células produtoras de PspA4Pro cultivadas a 37 °C com indução por lactose. 2) subfragmento 2 de  $Q_2$ . 3) subfragmento 3 de  $Q_2$ . 4) subfragmento 4 de  $Q_2$ . 5) subfragmento 5 de  $Q_2$ . 6) subfragmento 6 de  $Q_2$ .

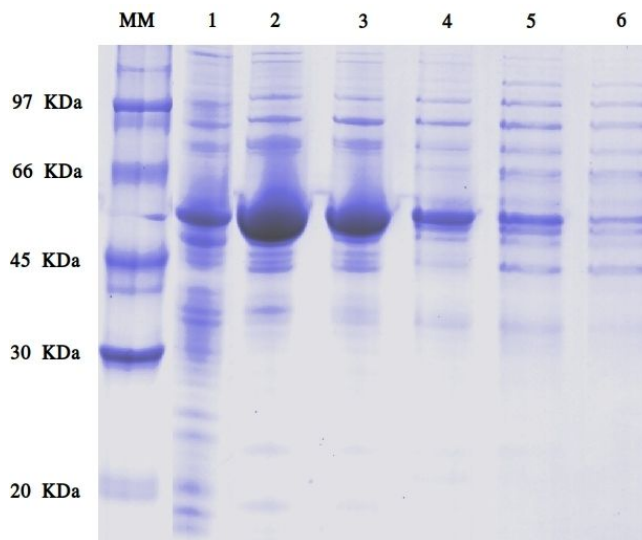


Figura 5.14: Géis SDS-PAGE do extrato celular clarificado ( $ECC_B$ ) que apresentava PspA4Pro recombinante junto com outras proteínas e dos subfragmentos 2 a 6 da eluição  $Q_2$  de CTA-IX. MM) Marcador molecular contendo fosforilase (97 KDa), albumina (66 KDa), ovalbumina (45 KDa), anidro (30 KDa) e tripsina (20 KDa). 1)  $ECC_B$  obtido por células produtoras de PspA4Pro cultivadas a 37 °C com indução por lactose. 2) subfragmento 2 de  $Q_2$ . 3) subfragmento 3 de  $Q_2$ . 4) subfragmento 4 de  $Q_2$ . 5) subfragmento 5 de  $Q_2$ . 6) subfragmento 6 de  $Q_2$ .

Uma vez que as simulações demonstraram que concentrações de NaCl menos intensas aumentam os valores de  $P_i$  em  $Q_2$ , é possível propor eluições alternativas para a CTA-IX com finalidades de aumentar ainda mais a pureza final. Por exemplo, se os subfragmentos 3 a 6 são selecionados,  $REC_{3-6}$  e  $P_{3-6}$  são 85,8% e 91,0%, respectivamente, representando um rendimento específico relativo de  $9,7 \times 10^{-3} \text{ g}_{\text{PspA4Pro}} \cdot \text{mLresin}^{-1}$ ; se os subfragmentos 4 a 6 são selecionados,  $REC_{4-6}$  reduz para 50,3% e  $P_{4-6}$  aumenta para 93,7%, representando um rendimento específico relativo de  $5,8 \times 10^{-3} \text{ g}_{\text{PspA4Pro}} \cdot \text{mLresin}^{-1}$ ; e se somente os subfragmentos 4 e 5 forem selecionados,  $REC_{4-5}$  reduz para 43,9% e  $P_{4-5}$  aumenta para 94,5%, representando um rendimento específico relativo de  $5,1 \times 10^{-3} \text{ g}_{\text{PspA4Pro}} \cdot \text{mLresin}^{-1}$ . A eluição alternativa 3-6 representa um aumento de rendimento específico e pureza de 149% e 9,9% quando comparados aos valores obtidos no capítulo 3, cujos melhores valores encontrados foram  $3,9 \times 10^{-3}$

Comparando essa eluição com as obtidas por FIGUEIREDO et al (2017), que produziram  $7,1 \times 10^{-3} \text{ g}_{\text{PspA4Pro}} \cdot \text{mL}^{-1}$  de resina com pureza de 65,2%, aumentos relativos de 35% e 39,6% para rendimento específico e pureza, respectivamente, são observados. As eluições alternativas 4-6 e 4-5 representam um aumento de rendimento específico e pureza de 49% e 13,2% e 30,7% e 14,1%, respectivamente, quando comparados aos valores relatados no capítulo 3.

## 5.5. Conclusões

Neste trabalho, uma abordagem já desenvolvida anteriormente para lidar com a separação de misturas complexas de proteínas foi aprimorada, representando uma importante contribuição para a modelagem e otimização da cromatografia de troca iônica. A abordagem foi aplicada no processo de purificação de PspA4Pro, uma proteína recombinante presente em extratos celulares de *E. coli*. A estratégia proposta se mostrou útil para estimar separadamente os parâmetros de PspA4Pro e das outras proteínas eluídas no processo. Os resultados das simulações foram validados com sucesso através de cromatografias executadas em diferentes forças iônicas, as quais processaram extratos celulares complexos obtidos por células recombinantes produtoras de PspA4Pro. O modelo também foi capaz de detectar forças iônicas mais apropriadas para eluir a PspA4Pro com menos impurezas, descrevendo com detalhes o perfil de eluição de PspA4Pro e do pool de outras proteínas em diferentes regiões do pico. Com essas informações, foi possível identificar um perfil de eluição alternativo que

proporcionou aumentos de rendimentos específicos e purezas de PspA4Pro de 149% e 9,9% quando comparados com a abordagem de simulação do capítulo 3 e de 35% e 39,6% quando comparado com o processo de purificação original. É importante destacar que o procedimento descrito neste trabalho pode ser aplicado à cromatografia de qualquer tipo de mistura complexa de proteínas, uma vez que os parâmetros do modelo podem ser determinados separadamente para outras proteínas de interesse.

## 6. CONSIDERAÇÕES FINAIS

O desenvolvimento de tecnologias para a produção de proteínas recombinantes é uma demanda presente em diversos nichos da biotecnologia. Nas últimas décadas, os avanços das técnicas de biologia molecular permitiram que células *E. coli* fossem utilizadas como fábricas celulares, apresentando altas produtividades de proteínas de interesse industrial e farmacêutico. Em paralelo, um esforço intensivo tem sido realizado no desenvolvimento de processos de purificação. Diversos trabalhos já abordaram a modelagem matemática e simulação de cromatografias, objetivando aumentar a eficiência e reduzir custos do processo *Downstream*. No entanto, poucos deles propuseram modelar misturas complexas de proteínas, restringindo a aplicação dos modelos desenvolvidos em situações reais.

Neste contexto, o presente trabalho propôs desenvolver estratégias para modelar e simular processos cromatográficos utilizados para purificar misturas complexas de proteínas. As metodologias foram desenvolvidas e aplicadas na etapa de cromatografia de troca aniônica do processo de purificação de PspA4Pro, uma proteína produzida por *E. coli* recombinante, isenta de cauda de histidina, e utilizada como agente carreador na formulação de vacinas conjugadas.

Em um primeiro momento, duas isotermas foram avaliadas para descrever a adsorção de proteínas durante as cromatografias: a isoterma de Langmuir competitiva e a isoterma de ação em efeito estérico (“*Steric Mass Action Isotherm*”). Ambas se mostraram eficientes para descrever os perfis de eluição, inclusive quando foram usados extratos celulares clarificados obtidos a partir de biomassas geradas sob diferentes condições de cultivo e indução. Para viabilizar a modelagem de misturas complexas, foi introduzida uma estratégia que visava estimar separadamente os parâmetros de PspA4Pro e das impurezas proteicas eluídas no mesmo momento. Como resultado prático desta modelagem preliminar foi identificado um modo alternativo de eluição que aumentou a pureza da PspA4Pro obtida em 34,0%, em relação à condição convencional de operação.

Apesar do aumento da pureza, verificou-se que a simulação apresentava alto desvio devido à insuficiência de dados experimentais e à estratégia de estimativa de parâmetros adotada. Assim, a evolução do trabalho levou à incorporação de uma metodologia de validação cruzada para ajustar o modelo a uma maior quantidade de dados experimentais e também tornar a validação mais robusta. As simulações possibilitaram a obtenção do perfil de distribuição da massa de proteínas no interior da coluna em diferentes momentos da cromatografia, tornando possível identificar condições de processamento que

permitiram dobrar o rendimento de PspA4Pro produzida em relação ao volume de resina, que passou de  $0,010 \text{ g.mL}^{-1}_{\text{resina}}$  na estratégia convencional para  $0,022 \text{ g.mL}^{-1}_{\text{resina}}$  na condição aprimorada, sem redução de pureza.

Em uma última etapa, a metodologia de validação cruzada foi utilizada para descrever separadamente os perfis de eluição de PspA4Pro e das impurezas proteicas eluídas no mesmo momento, possibilitando incorporar os efeitos da variação da força iônica das eluições na pureza da PspA4Pro obtida. Foi detectado um aumento expressivo de pureza com a redução da força iônica. As simulações realizadas permitiram mapear os perfis de recuperação e pureza para uma grande variedade de combinações de força iônica em função do tempo, levando à identificação de condições de eluição que produziram PspA4Pro com até 94,5% de pureza, ou seja, praticamente alcançando os requisitos mínimos exigidos pelos órgãos reguladores (95%) em apenas uma etapa cromatográfica. No processo convencional, etapas adicionais de crioprecipitação e cromatografia de troca catiônica são necessárias para alcançar essa faixa de pureza. Neste contexto, é importante observar que o potencial da aplicação dos conhecimentos gerados no presente trabalho vai muito além da otimização da etapa de cromatografia de troca aniônica em si, uma vez que os impactos se estendem ao processo de purificação como um todo, com excelentes perspectivas de efetiva diminuição do custo.

É ainda importante ressaltar que abordagem descrita neste trabalho é inédita, já que a modelagem de cromatografias de misturas complexas é um tema pouco explorado por grupos de pesquisa devido à alta complexidade das operações envolvidas.

Em trabalhos futuros, sugere-se que seja realizada uma otimização no processo de cromatografia estudado, incorporando aspectos operacionais e econômicos. Além disso, é importante destacar que os procedimentos de modelagem desenvolvidos nesta tese podem ser aplicados à purificação de outras misturas complexas, desde que novas constantes de adsorção sejam estimadas para as proteínas de interesse e as impurezas proteicas. Da mesma forma, o processo de purificação pode ser expandido na plataforma EMSO, permitindo a incorporação de outras operações de downstream geralmente presentes na purificação de proteínas recombinantes, possibilitando a criação de uma “Biorefinaria Virtual da Proteína Recombinante”, uma ferramenta poderosa para a otimização dos processos de purificação.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABCOMBI BIOSCIENCES INC. Abcombibio.com. Disponível em: <<https://www.abcombibio.com/pneumococcal>>. Acesso em: 10 maio 2018.
- ALBARELLI, Juliana Q. et al. Multi-objective optimization of a sugarcane biorefinery for integrated ethanol and methanol production. **Energy**, v. 138, p. 1281-1290, 2017.
- ALMEIDA, Rafael N.; SOARES, Rafael de P.; CASSEL, Eduardo. Fractionation process of essential oils by batch distillation. **Brazilian Journal of Chemical Engineering**, v. 35, n. 3, p. 1129-1140, 2018.
- ALONSODEVELASCO, E. et al. *Streptococcus pneumoniae*: virulence factors, pathogenesis, and vaccines. **Microbiol. Mol. Biol. Rev.**, v. 59, n. 4, p. 591-603, 1995.
- ASSENBERG, Rene et al. Advances in recombinant protein expression for use in pharmaceutical research. **Current opinion in structural biology**, v. 23, n. 3, p. 393-402, 2013.
- BANERJEE, Ipsita; PAL, Siladitya; MAITI, Spandan. Computationally efficient black-box modeling for feasibility analysis. **Computers & Chemical Engineering**, v. 34, n. 9, p. 1515-1521, 2010.
- BARAZZONE, Giovana C. et al. Production and purification of recombinant fragment of pneumococcal surface protein A (PspA) in *Escherichia coli*. **Procedia in Vaccinology**, v. 4, p. 27-35, 2011.
- BOGAERT, Debby et al. Pneumococcal vaccines: an update on current strategies. **Vaccine**, v. 22, n. 17-18, p. 2209-2220, 2004.
- BORBA, Carlos Eduardo. **Modelagem da remoção de metais pesados em coluna de adsorção de leito fixo. 2006. 163 f.** 2006. Tese de Doutorado. Dissertação (Mestrado)–Curso de Pós-graduação em Engenharia Química da Universidade Estadual de Campinas. Campinas.
- BRACEWELL, D. G. et al. Impact of clarification strategy on chromatographic separations: Pre processing of cell homogenates. **Biotechnology and bioengineering**, v. 100, n. 5, p. 941-949, 2008.
- BREMER, Maria GEG et al. Electrostatic interactions between immunoglobulin (IgG) molecules and a charged sorbent. **Colloids and Surfaces A: Physicochemical and Engineering Aspects**, v. 250, n. 1-3, p. 29-42, 2004.
- BRILES, David E. et al. Immunization of humans with recombinant pneumococcal surface protein A (rPspA) elicits antibodies that passively protect mice from fatal infection with

*Streptococcus pneumoniae* bearing heterologous PspA. **The Journal of infectious diseases**, v. 182, n. 6, p. 1694-1701, 2000.

BUYEL, Johannes Felix; FISCHER, Rainer. Downstream processing of biopharmaceutical proteins produced in plants: the pros and cons of flocculants. **Bioengineered**, v. 5, n. 2, p. 138-142, 2014.

CAMPANI JUNIOR, Gilson et al. Reator *airlift* operado em sobrepressão: construção, caracterização da transferência de oxigênio e aplicação em cultivos de *Escherichia coli* recombinante. 2014.

CAMPANI, Gilson et al. Oxygen transfer in a pressurized airlift bioreactor. **Bioprocess and biosystems engineering**, v. 38, n. 8, p. 1559-1567, 2015.

CAMPANI, Gilson et al. Oxygen transfer in a pressurized airlift bioreactor. **Bioprocess and biosystems engineering**, v. 38, n. 8, p. 1559-1567, 2015.

CAMPANI, Gilson et al. Recombinant protein production by engineered *Escherichia coli* in a pressurized airlift bioreactor: a techno-economic analysis. **Chemical Engineering and Processing: Process Intensification**, v. 103, p. 63-69, 2016.

CANDY KS et al. Design of high productivity antibody capture by protein A chromatography using an integrated experimental and modeling approach. **Journal of Chromatography B**, v. 899, p. 116-126, 2012.

CARPIO, Roymel Rodriguez; SECCHIB, Argimiro Resende; DE CAMPOS GIORDANOC, Roberto. Modeling, simulation and optimization of ethanol extractive distillation using glycerol. In: **EngOpt 2016-5th International Conference on Engineering Optimization**. E-papers Servicos Editoriais Ltda, 2016. p. 548-557.

CARVALHO, B. M. A. et al. Steric mass action model for lactoferrin adsorption in cryogel with immobilized copper ions. **Brazilian Journal of Chemical Engineering**, v. 33, n. 1, p. 215-223, 2016.

CASTRO-MARTÍNEZ, Claudia; LUNA-SUÁREZ, Silvia; PAREDES-LÓPEZ, Octavio. Overexpression of a modified protein from amaranth seed in *Escherichia coli* and effect of environmental conditions on the protein expression. **Journal of biotechnology**, v. 158, n. 1-2, p. 59-67, 2012.

CHAN, Sharon et al. A systematic approach for modeling chromatographic processes—Application to protein purification. **AIChE journal**, v. 54, n. 4, p. 965-977, 2008.

CHEN, Te; XU, Wenchun. A purification method for tag-free human cystatin C recombinant protein expressed in *Escherichia coli*. **Preparative Biochemistry and Biotechnology**, v. 47, n. 2, p. 123-128, 2017.

CILLÓNIZ, Catia; TORRES, Antoni. Entendimento da mortalidade em pneumonia pneumocócica bacterêmica. **Jornal Brasileiro de Pneumologia**, v. 38, n. 4, p. 419-421, 2012.

CLOSE, Edward J. et al. Modelling of industrial biopharmaceutical multicomponent chromatography. **Chemical Engineering Research and Design**, v. 92, n. 7, p. 1304-1314, 2014.

COLLINS, Carol H.; BRAGA, Gilberto Leite; BONATO, Pierina Sueli. Fundamentos de cromatografia. 2006.

DA ROCHA NOVAES, Leandro et al. Modeling, simulation and kinetic parameter estimation for diesel hydrotreating. **Fuel**, v. 209, p. 184-193, 2017.

DA SILVA, Míriam A. et al. Conjugation of PspA4Pro with capsular *Streptococcus pneumoniae* polysaccharide serotype 14 does not reduce the induction of cross-reactive antibodies. **Clin. Vaccine Immunol.**, v. 24, n. 8, p. e00118-17, 2017.

DEMANÈCHE, Sandrine et al. Dissimilar pH-dependent adsorption features of bovine serum albumin and  $\alpha$ -chymotrypsin on mica probed by AFM. **Colloids and Surfaces B: Biointerfaces**, v. 70, n. 2, p. 226-231, 2009.

DONAT, R.; LOPES UREÑA, S.; TORRES-LAPASÍO, J.R.; Gradient design for liquid chromatography using multi-scale optimization. **Journal of Chromatography**, Volume 1535, p. 32, 2018.

EXPASY, ExPASy Bioinformatics resource journal, (n.d.). <https://web.expasy.org/protparam>>. Acesso em: 14/01/2019.

FARAJI, Naeimeh; ZHANG, Yan; RAY, Ajay K. Determination of adsorption isotherm parameters for minor whey proteins by gradient elution preparative liquid chromatography. **Journal of Chromatography A**, v. 1412, p. 67-74, 2015.

FERRER-MIRALLES, Neus et al. Microbial factories for recombinant pharmaceuticals. **Microbial cell factories**, v. 8, n. 1, p. 17, 2009.

FIGUEIREDO, Douglas B. et al. Production and purification of an untagged recombinant pneumococcal surface protein A (PspA4Pro) with high-purity and low endotoxin content. **Applied microbiology and biotechnology**, v. 101, n. 6, p. 2305-2317, 2017.

FIGUEIREDO, Douglas Borges de. **Desenvolvimento do processo de purificação da proteína A de superfície de pneumococo do clado 4 (PspA4Pro)**. Tese de Doutorado. Universidade de São Paulo.

FINETTE, Gerard MS; MAO, Qui-Ming; HEARN, Milton TW. Comparative studies on the isothermal characteristics of proteins adsorbed under batch equilibrium conditions to ion-exchange, immobilised metal ion affinity and dye affinity matrices with different ionic

strength and temperature conditions. **Journal of Chromatography A**, v. 763, n. 1-2, p. 71-90, 1997.

FONSECA, G. C.; COSTA, C. B. B.; CRUZ, A. J. G. Comparing a Dynamic Fed-Batch and a Continuous Steady-State Simulation of Ethanol Fermentation in a Distillery to a Stoichiometric Conversion Simulation. **Brazilian Journal of Chemical Engineering**, v. 34, n. 4, p. 1121-1131, 2017.

FURLAN, Felipe F. et al. Process alternatives for second generation ethanol production from sugarcane bagasse. In: **Computer Aided Chemical Engineering**. Elsevier, 2015. p. 1349-1354.

FURLAN, Felipe Fernando et al. A simple approach to improve the robustness of equation-oriented simulators: Multilinear look-up table interpolators. **Computers & Chemical Engineering**, v. 86, p. 1-4, 2016.

GALLANT, Stuart R.; KUNDU, Amitava; CRAMER, Steven M. Modeling non-linear elution of proteins in ion-exchange chromatography. **Journal of Chromatography A**, v. 702, n. 1-2, p. 125-142, 1995.

GECIOVA, Jana; BURY, Dean; JELEN, Paul. Methods for disruption of microbial cells for potential use in the dairy industry—a review. **International Dairy Journal**, v. 12, n. 6, p. 541-553, 2002.

GERONTAS, Spyridon et al. Integration of scale-down experimentation and general rate modelling to predict manufacturing scale chromatographic separations. **Journal of Chromatography A**, v. 1217, n. 44, p. 6917-6926, 2010.

GHOSAL, Partha S.; GUPTA, Ashok K. Determination of thermodynamic parameters from Langmuir isotherm constant-revisited. **Journal of Molecular Liquids**, v. 225, p. 137-146, 2017.

GLOBAL DATA PLC HOME. GlobalData Plc. Disponível em: <<http://www.globaldata.com>>. Acesso em: 10 maio 2018.

GOEDDEL, David V. et al. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. **Proceedings of the National Academy of Sciences**, v. 76, n. 1, p. 106-110, 1979.

GRIMI, N. et al. Selective extraction from microalgae *Nannochloropsis* sp. using different methods of cell disruption. **Bioresource Technology**, v. 153, p. 254-259, 2014.

GRITTI, Fabrice; GUIOCHON, Georges. Calculated and experimental chromatograms for distorted gradients and non-linear solvation strength retention models. **Journal of Chromatography A**, v. 1356, p. 96-104, 2014.

GU, Tingyue. **Mathematical modeling and scale-up of liquid chromatography: With application examples**. Springer, 2015.

GUIOCHON, Georges; FELINGER, Attila; SHIRAZI, Dean GG. **Fundamentals of preparative and nonlinear chromatography**. Elsevier, 2006.

HARRISON, Susan TL. Bacterial cell disruption: a key unit operation in the recovery of intracellular products. **Biotechnology advances**, v. 9, n. 2, p. 217-240, 1991.

HEYLAND, Jan; BLANK, Lars M.; SCHMID, Andreas. Quantification of metabolic limitations during recombinant protein production in *Escherichia coli*. **Journal of biotechnology**, v. 155, n. 2, p. 178-184, 2011.

HILBRIG, Frank; FREITAG, Ruth. Protein purification by affinity precipitation. **Journal of chromatography B**, v. 790, n. 1-2, p. 79-90, 2003.

HORTA, A. C. L. et al. A supervision and control tool based on artificial intelligence for high cell density cultivations. **Brazilian journal of chemical engineering**, v. 31, n. 2, p. 457-468, 2014.

HORTA, Antonio Carlos Luperni et al. Intensification of high cell-density cultivations of *rE. coli* for production of *S. pneumoniae* antigenic surface protein, PspA3, using model-based adaptive control. **Bioprocess and biosystems engineering**, v. 35, n. 8, p. 1269-1280, 2012.

HORTA, Antonio Carlos Luperni et al. Sistema automático de supervisão e controle de cultivos de alta densidade celular de *E. coli* recombinante. 2011.

HUUK, Thiemo C. et al. Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography. **Separation and Purification Technology**, v. 136, p. 207-222, 2014.

Ion exchange chromatography. Uppsala: Amersham Biosciences, 2002.

INSIGHTS, COHERENT. Global Recombinant Therapeutic Antibodies and Proteins Market to Surpass US\$ 250 Billion by 2026. GlobeNewswire News Room. Disponível em: <<https://globenewswire.com/news-release/2018/07/23/1540521/0/en/Global-Recombinant-Therapeutic-Antibodies-and-Proteins-Market-to-Surpass-US-250-Billion-by-2026.html>>. Acesso em: 11 jan. 2019.

JANSON, Jan-Christer (Ed.). **Protein purification: principles, high resolution methods, and applications**. John Wiley & Sons, 2012.

KACZMARSKI, Krzysztof et al. Comparative modeling of breakthrough curves of bovine serum albumin in anion-exchange chromatography. **Journal of Chromatography A**, v. 925, n. 1-2, p. 1-17, 2001.

KAMIONKA, Mariusz. Engineering of therapeutic proteins production in *Escherichia coli*. **Current pharmaceutical biotechnology**, v. 12, n. 2, p. 268-274, 2011.

KARKOV, Hanne Sophie; SEJERGAARD, Lars; CRAMER, Steven M. Methods development in multimodal chromatography with mobile phase modifiers using the steric mass action model. **Journal of Chromatography A**, v. 1318, p. 149-155, 2013.

KHARE, Atul R.; PEPPAS, Nikolaos A. Swelling/deswelling of anionic copolymer gels. **Biomaterials**, v. 16, n. 7, p. 559-567, 1995.

KILIKIAN, B. V. et al. Process strategies to improve heterologous protein production in *Escherichia coli* under lactose or IPTG induction. **Process Biochemistry**, v. 35, n. 9, p. 1019-1025, 2000.

KRISHNAN, Anuradha et al. Single-column purification of the tag-free, recombinant form of the neuronal calcium sensor protein, hippocalcin expressed in *Escherichia coli*. **Protein expression and purification**, v. 123, p. 35-41, 2016.

KYE, Yoon-Chul et al. Intranasal immunization with pneumococcal surface protein A in the presence of nanoparticle forming polysorbitol transporter adjuvant induces protective immunity against the *Streptococcus pneumoniae* infection. **Acta biomaterialia**, v. 90, p. 362-372, 2019.

LAEMMLI, Ulrich K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **nat**, v. 227, n. 5259, p. 680, 1970.

LAN, Qingdao et al. A modified Langmuir model for the prediction of the effects of ionic strength on the equilibrium characteristics of protein adsorption onto ion exchange/affinity adsorbents. **Chemical Engineering Journal**, v. 81, n. 1-3, p. 179-186, 2001.

LANDER, Russel J. et al. Fractional precipitation of plasmid DNA from lysate by CTAB. **Biotechnology and bioengineering**, v. 79, n. 7, p. 776-784, 2002.

LE HA, Ngoc; UNGVARAI, Janos; KOVATS, Ervin S. Adsorption isotherm at the liquid-solid interface and the interpretation of chromatographic data. **Analytical Chemistry**, v. 54, n. 14, p. 2410-2421, 1982.

LEE, Chi-Jen; BANKS, Susan D.; LI, Jean P. Virulence, immunity, and vaccine related to *Streptococcus pneumoniae*. **Critical reviews in microbiology**, v. 18, n. 2, p. 89-114, 1991.

LEE, Sang Yup. High cell-density culture of *Escherichia coli*. **Trends in biotechnology**, v. 14, n. 3, p. 98-105, 1996.

LI, Bo et al. Protection elicited by nasal immunization with pneumococcal surface protein A (PspA) adjuvanted with bacterium-like particles against *Streptococcus pneumoniae* infection in mice. **Microbial pathogenesis**, v. 123, p. 115-119, 2018.

LI, Zhaopeng et al. Simple defined autoinduction medium for high-level recombinant protein production using T7-based *Escherichia coli* expression systems. **Applied microbiology and biotechnology**, v. 91, n. 4, p. 1203, 2011.

LIENQUEO, M. Elena et al. Mathematical modeling of protein chromatograms. **Chemical Engineering & Technology**, v. 35, n. 1, p. 46-57, 2012.

LIN, Fu-Yung et al. Microcalorimetric studies of the interaction mechanisms between proteins and Q-Sepharose at pH near the isoelectric point (pI): effects of NaCl concentration, pH value, and temperature. **Journal of Chromatography A**, v. 912, n. 2, p. 281-289, 2001.

LOPES, Camila et al. Improving the cost effectiveness of enhanced green fluorescent protein production using recombinant *Escherichia coli* BL21 (DE3): Decreasing the expression inducer concentration. **Biotechnology and applied biochemistry**, 2019.

LÓPEZ-UREÑA, S. et al. Gradient design for liquid chromatography using multi-scale optimization. **Journal of Chromatography A**, v. 1534, p. 32-42, 2018.

LOWRY, Oliver H. et al. Protein measurement with the Folin phenol reagent. **Journal of biological chemistry**, v. 193, p. 265-275, 1951.

MADHI, Shabir A.; PELTON, Stephen I. Epidemiology, diagnosis, and treatment of serious pneumococcal infections in children. In: **Pneumococcal Vaccines**. American Society of Microbiology, 2008. p. 95-116.

MAHMOUD, Dalia Khalid; MOHD, Mohamad Amran. Langmuir model application on solid-liquid adsorption using agricultural wastes: Environmental application.

MAMAT, Uwe et al. Detoxifying *Escherichia coli* for endotoxin-free production of recombinant proteins. **Microbial cell factories**, v. 14, n. 1, p. 57, 2015.

MANGTANI, Punam; CUTTS, Felicity; HALL, Andrew J. Efficacy of polysaccharide pneumococcal vaccine in adults in more developed countries: the state of the evidence. **The Lancet infectious diseases**, v. 3, n. 2, p. 71-78, 2003.

MARISCH, Karoline et al. Evaluation of three industrial *Escherichia coli* strains in fed-batch cultivations during high-level SOD protein production. **Microbial cell factories**, v. 12, n. 1, p. 58, 2013.

MCCABE, Warren Lee; SMITH, Julian Cleveland; HARRIOTT, Peter. **Unit operations of chemical engineering**. New York: McGraw-hill, 1993.

MEZZARI, Isabella Anna et al. Utilização de carvões adsorventes para o tratamento de efluentes contendo pesticidas. 2002.

MICOLI, F. et al. Production of a conjugate vaccine for *Salmonella enterica* serovar Typhi from *Citrobacter* Vi. **Vaccine**, v. 30, n. 5, p. 853-861, 2012.

MIYAJI, Eliane Namie et al. Serotype-independent pneumococcal vaccines. **Cellular and molecular life sciences**, v. 70, n. 18, p. 3303-3326, 2013.

MORENO, Adriana T. et al. Immunization of mice with single PspA fragments induces antibodies capable of mediating complement deposition on different pneumococcal strains and cross-protection. **Clin. Vaccine Immunol.**, v. 17, n. 3, p. 439-446, 2010.

MOROWVAT, Mohammad Hossein et al. Optimization of fermentation conditions for recombinant human interferon beta production by *Escherichia coli* using the response surface methodology. **Jundishapur journal of microbiology**, v. 8, n. 4, 2015.

NELDER, John A.; MEAD, Roger. A simplex method for function minimization. **The computer journal**, v. 7, n. 4, p. 308-313, 1965.

NELSON, David L.; COX, Michael M. **Princípios de Bioquímica de Lehninger-7**. Artmed Editora, 2018.

NORDE, Willem. Driving forces for protein adsorption at solid surfaces. In: **Macromolecular Symposia**. Basel: Hüthig & Wepf Verlag, 1996. p. 5-18.

ORELLANA, C. A.; SHENE, C.; ASENJO, J. A. Mathematical modeling of elution curves for a protein mixture in ion exchange chromatography applied to high protein concentration. **Biotechnology and bioengineering**, v. 104, n. 3, p. 572-581, 2009.

PERCIANI, Catia T. et al. Conjugation of polysaccharide 6B from *Streptococcus pneumoniae* with pneumococcal surface protein A: PspA conformation and its effect on the immune response. **Clin. Vaccine Immunol.**, v. 20, n. 6, p. 858-866, 2013.

PESSOA JUNIOR, Adalberto; KILIKIAN, Beatriz Vahan. Purificação de produtos biotecnológicos. 2005.

PRASAD, Dilip K. et al. A novel framework for making dominant point detection methods non-parametric. **Image and Vision Computing**, v. 30, n. 11, p. 843-859, 2012.

RABE, Michael; VERDES, Dorinel; SEEGER, Stefan. Understanding protein adsorption phenomena at solid surfaces. **Advances in colloid and interface science**, v. 162, n. 1-2, p. 87-106, 2011.

REDWAN, EL-Rashdy M. Cumulative updating of approved biopharmaceuticals. **Human Antibodies**, v. 16, n. 3-4, p. 137-158, 2007.

RIESENBERG, D ; GUTHKE, R. High-cell-density cultivation of microorganisms. **Applied microbiology and biotechnology**, v. 51, n. 4, p. 422-430, 1999.

RODRIGUES, Tasson C. et al. Mucosal immunization with PspA (Pneumococcal surface protein A)-adsorbed nanoparticles targeting the lungs for protection against pneumococcal infection. **PLoS one**, v. 13, n. 1, p. e0191692, 2018.

SAFI, Carl et al. Aqueous extraction of proteins from microalgae: effect of different cell disruption methods. **Algal Research**, v. 3, p. 61-65, 2014.

SCOPES, Robert K. **Protein purification: principles and practice**. Springer Science & Business Media, 2013.

SENKOVICH, Olga et al. Structure of a complex of human lactoferrin N-lobe with pneumococcal surface protein a provides insight into microbial defense mechanism. **Journal of molecular biology**, v. 370, n. 4, p. 701-713, 2007.

SHEN, Hong; FREY, Douglas D. Effect of charge regulation on steric mass-action equilibrium for the ion-exchange adsorption of proteins. **Journal of Chromatography A**, v. 1079, n. 1-2, p. 92-104, 2005.

SHEN, Hong; FREY, Douglas D. Effect of charge regulation on steric mass-action equilibrium for the ion-exchange adsorption of proteins. **Journal of Chromatography A**, v. 1079, n. 1-2, p. 92-104, 2005.

SHILOACH, Joseph; FASS, Rephael. Growing *E. coli* to high cell density—a historical perspective on method development. **Biotechnology advances**, v. 23, n. 5, p. 345-357, 2005.

SHOJA, ALSADATI SA et al. Recent advances in high cell density cultivation for production of recombinant protein. 2008.

SIAHPOOSH, Mohsen; FATEMI, Shohreh; VATANI, Ali. Mathematical modeling of single and multi-component adsorption fixed beds to rigorously predict the mass transfer zone and breakthrough curves. **Iranian Journal of Chemistry and Chemical Engineering (IJCCE)**, v. 28, n. 3, p. 25-44, 2009.

SILVA, Gabriel Gonçalves et al. Aspectos morfológicos, reológicos e fisiológicos dos cultivos de *Escherichia coli* recombinante. 2015.

SOARES, R. de P.; SECCHI, A. R. EMSO: A new environment for modelling, simulation and optimisation. In: **Computer Aided Chemical Engineering**. Elsevier, 2003. p. 947-952.

SOFER, Gail K.; HAGEL, Lars (Ed.). **Handbook of process chromatography: a guide to optimization, scale up, and validation**. Academic press, 1997.

SØRENSEN, Hans Peter; MORTENSEN, Kim Kusk. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. **Journal of biotechnology**, v. 115, n. 2, p. 113-128, 2005.

STIGTER, Dirk; DILL, Ken A. Charge effects on folded and unfolded proteins. **Biochemistry**, v. 29, n. 5, p. 1262-1271, 1990.

SUBSTECH SUBSTANCES AND TECHNOLOGY. Homogenization, 2018. Disponível em: <<http://www.substech.com/dokuwiki/doku.php?id=homogenization>>. Acesso em: Jun 2018.

TADA, Rui et al. Nasal vaccination with pneumococcal surface protein A in combination with cationic liposomes consisting of DOTAP and DC-chol confers antigen-mediated protective immunity against *Streptococcus pneumoniae* infections in mice. **International immunopharmacology**, v. 61, p. 385-393, 2018.

TEOH, H. K. et al. Experimental verification and optimisation of a detailed dynamic high performance liquid chromatography column model. **Computers & Chemical Engineering**, v. 25, n. 4-6, p. 893-903, 2001.

THOMBRE, Mandar N.; PREISIG, Heinz A.; ADDIS, Misganaw B. Developing surrogate models via computer based experiments. In: **Computer Aided Chemical Engineering**. Elsevier, 2015. p. 641-646.

TREYBAL, Robert E. Mass transfer operations. **New York**, v. 466, 1980.

VADESILHO, Cintia FM et al. Mapping of epitopes recognized by antibodies induced by immunization of mice with PspA and PspC. **Clin. Vaccine Immunol.**, v. 21, n. 7, p. 940-948, 2014.

WAEGEMAN, Hendrik; SOETAERT, Wim. Increasing recombinant protein production in *Escherichia coli* through metabolic and genetic engineering. **Journal of industrial microbiology & biotechnology**, v. 38, n. 12, p. 1891-1910, 2011.

WAKKEL, M. et al. Equilibrium and Kinetics of High Molecular Weight Protein Uptake in Ion Exchange Chromatography. **International Journal of Chemical Engineering and Applications**, v. 6, n. 5, p. 306, 2015.

WAUGH, David S. An overview of enzymatic reagents for the removal of affinity tags. **Protein expression and purification**, v. 80, n. 2, p. 283-293, 2011.

YU, Lin-Ling; SUN, Yan. Trace adsorption of positively charged proteins onto Sepharose FF and Sepharose FF-based anion exchangers. **Journal of Chromatography A**, v. 1253, p. 105-109, 2012.

YUSUFF, A. S. et al. Mathematical Modelling of Fixed Bed Adsorption Column for Liquid Phase Solute: Effect of Operating Variables. **International Journal of Scientific & Engineering Research**, v. 4, n. 8, p. 811-822, 2013

### APÊNDICE A: Exemplo de aplicação do EMSO

Para ilustrar melhor as definições do simulador EMSO descritas na seção 2.7, vamos considerar um sistema industrial que apresente uma reação de primeira ordem não reversível onde A se transforma em B. O processo é contínuo e composto por três reatores de mistura perfeita isotérmicos e em série, conforme mostra a figura A.1. As equações A.1 e A.2 descrevem o balanço de massa geral e o balanço de massa para o reagente A em cada reator, onde  $F_{in}$ ,  $C_{a,in}$ ,  $F_{out}$  e  $C_{a,out}$  são os fluxos molares totais e concentrações de A na entrada e saída, respectivamente,  $V$  é o volume do reator e  $r$  é a velocidade de reação.

$$F_{in} = F_{out} \quad (\text{A.1})$$

$$r_a V = F_{in} c_{a,in} - F_{out} c_{a,out} \quad (\text{A.2})$$

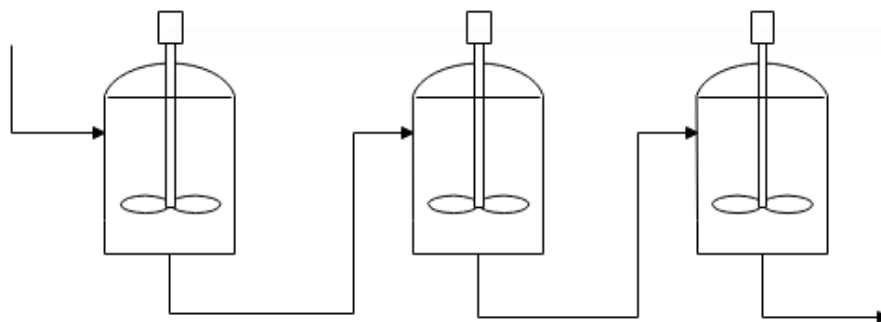


Figura A.1. Esquema de reatores de mistura perfeita para a conversão de A em B através de reação de primeira ordem.

Para descrição deste processo no EMSO, é necessário primeiramente que todos os parâmetros e variáveis sejam declarados ao simulador. Em um segundo momento devem ser definidos os modelos envolvidos no processo. A figura A.2 mostra a declaração de todas as variáveis e representa a descrição dos quatro modelos: O modelo de corrente composto por concentração e vazão molar, os modelos de alimentação e dreno, que são informações iniciais e finais do processo, respectivamente, com conexões de entrada ou saída em forma de corrente; e por fim o modelo de um reator químico isotérmico cujas variáveis são conexões de saída e entrada em forma de corrente, o volume e a velocidade de reação. Além disso, o

modelo do reator apresenta as equações base necessárias para que a reação química que acontece dentro do reator seja descrita de forma eficiente. As duas primeiras são as equações A.1 e A.2 e a terceira é a equação da cinética química de primeira ordem, considerando-se uma constante de reação de  $15 \text{ h}^{-1}$ .

Para finalizar, a figura A.3 descreve a estrutura de programação de um fluxograma. Inicialmente, são definidos os dispositivos a partir dos modelos descritos. Verifique que foi criado um total de cinco dispositivos: um para a alimentação (*al*), um para o dreno (*dre*) e três para os reatores envolvidos no processo (*rt1*, *rt2*, e *rt3*). Estes dispositivos foram conectados uns com os outros através das entradas e saídas que possuem, de tal forma que o processo completo represente o fluxograma da figura A.1. Por fim, algumas variáveis são especificadas e o processo pode ser simulado.

```
vazao_molar as Real (Unit='m^3/h');
concentracao as Real (Unit='kmol/m^3');
volume as Real (Unit='m^3');
velocidade_reacao as Real (Unit='kmol/m^3/h');
```

▼ Model corrente

VARIABLES

F as vazao\_molar;

C as concentracao;

end

▼ Model alimentacao

VARIABLES

out saida as corrente;

end

▼ Model dreno

VARIABLES

in entrada as corrente;

end

▼ Model reator

VARIABLES

in entrada as corrente;

out saida as corrente;

V as volume;

r as velocidade\_reacao;

EQUATIONS

entrada.F = saida.F;

$r * V = \text{entrada.F} * \text{entrada.C} - \text{saida.F} * \text{saida.C}$ ;

$r = 15 * 1/h * \text{saida.C}$ ;

end

Figura A.2. Definição dos modelos envolvidos em um processo que apresenta três reatores em série.

```

FlowSheet problema
DEVICES
rt1 as reator;
rt2 as reator;
rt3 as reator;
al as alimentacao;
dre as dreno;

CONNECTIONS
al.saida to rt1.entrada;
rt1.saida to rt2.entrada;
rt2.saida to rt3.entrada;
rt3.saida to dre.entrada;

SPECIFY
al.saida.F = 71 * 'm^3/h';
al.saida.C = 10 * 'kmol/m^3';
rt1.V = 10 * 'm^3';
rt2.V = 5 * 'm^3';
rt3.V = 3 * 'm^3';

end

```

Figura A.3. Desenvolvimento do fluxograma que representa o processo de três reatores em série.

A figura A.4 mostra os resultados do processo simulado, descrevendo a variação da concentração de saída de cada reator com o tempo. Como se trata de um processo contínuo a concentração de saída permanece constante, mas é possível verificar que  $C_{a,in} > C_{a1} > C_{a2} > C_{a3}$ , o que indica que a reação está ocorrendo e foi bem simulada.

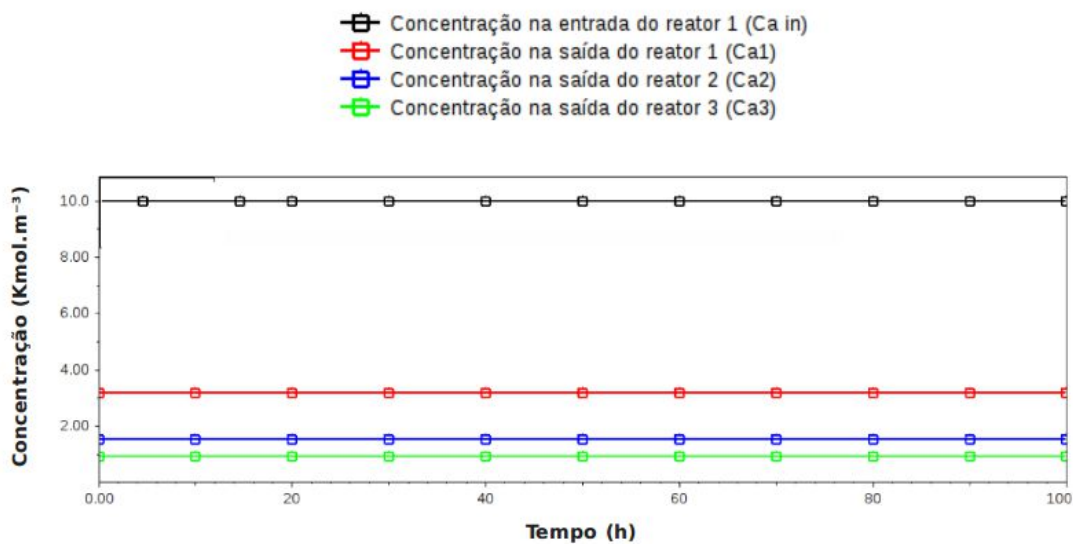


Figura A.4. Resultados do processo simulado composto por três reatores em série.

## APÊNDICE B: Disponibilização dos dados de condutividade e absorbância em forma de concentrações

O objetivo deste apêndice é explicar o procedimento adotado para disponibilização dos dados fornecidos pelo cromatógrafo em forma de concentrações de NaCl ( $C_{NaCl}$ ) e de proteínas (P), pois o sensor do equipamento fornece informações de porcentagem de condutividade e absorbância (280 nM). A aplicação desta metodologia é importante para a modelagem, pois informações mássicas são necessárias para o ajuste dos parâmetros. O perfil da cromatografia de troca aniônica é fornecido pelo equipamento conforme mostra a Figura B.1, onde os picos I e II representam a eluição Q150 e os picos III e IV representam as eluições Q300 e Q1000, respectivamente. É importante observar que os perfis representados na Figura B.1 são obtidos a partir de medições em diversos pontos com volumes fixos de amostra.

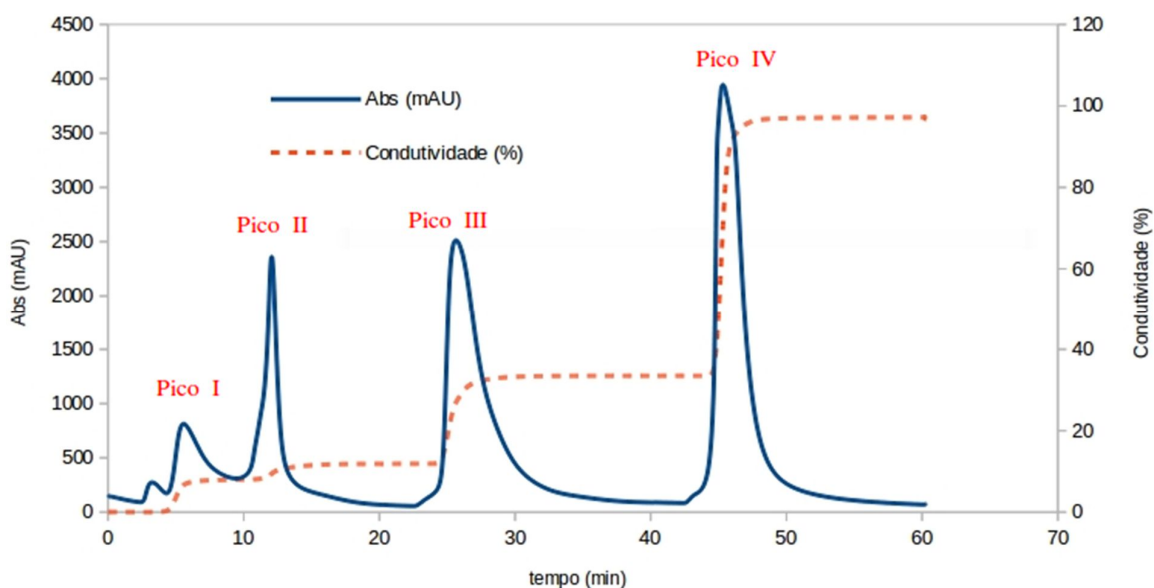


Figura B.1. Variação do perfil de absorbância (Abs) e porcentagem de condutividade (%C) durante o tempo no processo de cromatografia aniônica. Eluições: Q150 (0 a 20 minutos), Q300 (20 a 40 minutos) e Q1000 (40 a 60 minutos).

Sendo assim, os perfis de %C e Abs mostrados no gráfico devem ser transformados em perfis de  $C_{NaCl}$  e P, respectivamente. No primeiro caso, é construída uma curva padrão simples para que  $C_{NaCl}$  em cada ponto seja determinado a partir de %C. No segundo caso, utiliza-se a equação de Lambert-Beer para descrever o perfil de  $P_i$  para cada pico “i”,

conforme mostra a equação B.1, onde  $K_i$  é a constante de absorvidade molar, variável para cada pico identificado, pois cada um possui características químicas particulares, e  $D$  é a distância que a luz atravessa pelo corpo de medição, constante em todos os casos. Portanto, podemos reestruturar a equação B.1. nas formas das equações B.2 e B.3.

$$Abs_i = K_i \cdot D \cdot P_i \quad (B.1)$$

$$K'_i = K_i \cdot D \quad (B.2)$$

$$P_i = \frac{Abs_i}{K'_i} \quad (B.3)$$

Se considerarmos que o perfil de concentração ( $P_i$ ) de cada pico é composto por uma distribuição de  $N$  amostras, e que o volume de cada amostra é fixo ( $v$ ), podemos dizer que o somatório do produto  $v \cdot P_i$  em todos os pontos é igual a massa de proteínas totais  $M_{P,i}$  na fração completa do pico, conforme mostra a equação B.4. A partir das equações B.3 e B.4 é possível se obter a equação B.5.

$$M_{P,i} = \sum_{n=1}^N v \cdot P_{i,n} = v \sum_{n=1}^N P_{i,n} \quad (B.4)$$

$$M_{P,i} = \frac{V}{K'_i} \sum_{n=1}^N Abs_i \quad (B.5)$$

Sendo assim,  $K'_i$  é a única incógnita da equação B.5, uma vez que o valor de “ $v$ ”, o perfil de  $Abs_i$  e o número de pontos ( $N$ ) são conhecidos. A massa total de proteínas ( $M_{P,i}$ ) pode ser obtida a partir da concentração de proteínas totais ( $P_{T,i}$ ) obtida pelo método de Lowry (LOWRY, 1952) e pelo volume total ( $V_{T,i}$ ) da fração completa, conforme mostra a equação B.6.

$$M_{P,i} = P_{T,i} \cdot V_{T,i} \quad (B.6)$$

Finalmente, obtemos a equação B.7 a partir das equações B.2., B.5 e B.6, que mostra de forma clara como obter os valores de  $K_i$  a partir do perfil de  $Abs_i$  e de grandezas medidas no sistema. Através de análise dimensional podemos dizer que a constante  $K_i$ , também chamada de coeficiente de absorvidade molar, possui unidades de  $\text{mL} \cdot \text{mg}^{-1} \cdot \text{cm}^{-1}$  ou  $\text{M}^{-1} \cdot \text{cm}^{-1}$ .

$$K_i = \frac{V}{D \cdot P_{T,i} \cdot V_{T,i}} \sum_{n=1}^N Abs_i$$

(B.7)

Uma vez que as eluições são fracionadas separadamente (figura 3.1) é possível que a metodologia descrita seja aplicada para cada pico “i” identificado na figura B.1. Sendo assim, um valor de  $K_i$  é encontrado para cada pico e o perfil de P é obtido. Com isso, o conjunto de dados fornecido pela Figura B.1 pode ser disponibilizado conforme mostra a Figura B.2.

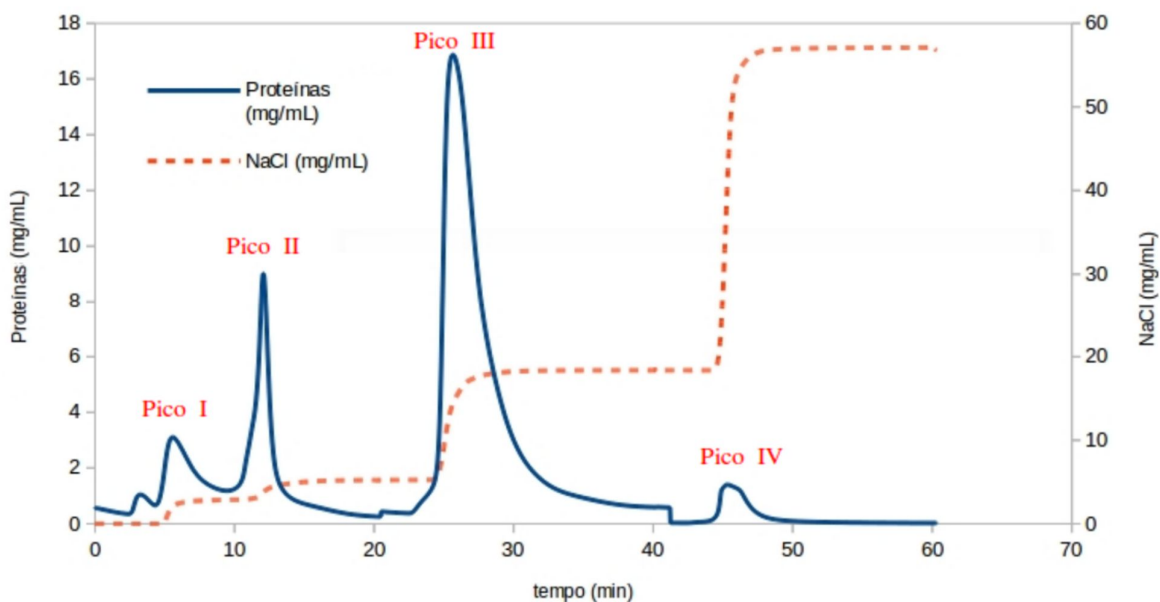


Figura B.2. Variação de concentração de proteínas (P) e NaCl ( $C_{\text{NaCl}}$ ) durante o tempo no processo de cromatografia aniônica.

Para que a eficiência da metodologia deste apêndice fosse testada, calculamos  $K_i$  para uma curva cromatográfica de PspA4Pro pura obtida experimentalmente (figura 3.8.b) e comparamos com o coeficiente de absorvidade molar teórico. O valor calculado foi de  $12291 \pm 910 \text{ M}^{-1} \text{ cm}^{-1}$ , enquanto o valor teórico é  $13410 \text{ M}^{-1} \text{ cm}^{-1}$  (ExpASy). A diferença observada entre os valores teórico e calculado pode ser explicada pelo desvio padrão do método de Lowry e pela presença de impurezas na amostra de PspA4Pro utilizada. Apesar dos desvios observados, acreditamos que a metodologia apresentada neste apêndice seja a forma mais adequada para descrever o perfil mássico de proteínas em cromatografias de misturas complexas, já que os grupos eluídos são formados por conjuntos de proteínas desconhecidas cujas propriedades não estão disponíveis na literatura.

### APÊNDICE C: Roteiro de modelagem da cromatografia de troca aniônica

Neste apêndice, pretendemos apresentar um roteiro detalhado que aborda a estratégia utilizada para a modelagem da cromatografia de troca aniônica. Para realização da modelagem consideramos que cada pico identificado no cromatograma corresponde a um grupo proteico com características químicas específicas. Portanto, os parâmetros das isotermas utilizadas foram ajustados para cada pico, conforme detalhado nos passos a seguir:

- Primeiramente, Os parâmetros fixos  $L$ ,  $u$ ,  $N_t$ ,  $\varepsilon_T$  e  $\hat{\epsilon}$  foram medidos ou fornecidos.
- A cromatografia de troca aniônica foi executada e as frações Q150, Q300 e Q1000 foram amostradas.
- Os picos Q150-I, Q150-II, Q300 e Q1000 foram identificados no cromatograma.
- A massa total de proteínas ( $M_{P,i}$ ) foi calculada para as frações do passo 2 a partir da concentração de proteínas totais ( $P_{T,i}$ ) (LOWRY, 1952) e volume total ( $V_{T,i}$ ) de cada fração, conforme mostram as equações C.1 a C.4. Para os picos Q150-I e Q150-II que foram eluídos na fração Q150, as massas totais particulares ( $M_{P,Q150-I}$  e  $M_{P,Q150-II}$ ) foram obtidas a partir das áreas de cada pico ( $A_{Q150-I}$  e  $A_{Q150-II}$ ) observadas no cromatograma.

$$M_{P,Q150-I} = A_{Q150-I} P_{T,Q150} \cdot V_{T,Q150} \quad (C.1)$$

$$M_{P,Q150-II} = A_{Q150-II} P_{T,Q150} \cdot V_{T,Q150} \quad (C.2)$$

$$M_{P,Q300} = P_{T,Q300} \cdot V_{T,Q300} \quad (C.3)$$

$$M_{P,Q1000} = P_{T,Q1000} \cdot V_{T,Q1000} \quad (C.4)$$

- Cada pico foi associado a um grupo proteico. Calculou-se as concentrações de cada grupo proteico ( $c_{i,e}$ ) no extrato celular clarificado (*clarified cell extract*), conforme mostram as equações C.5 a C.8, pois o volume desta fração ( $V_{CCE}$ ) foi medido durante o experimento. É importante lembrar que o extrato celular representa a mistura de proteínas que foi inserida no sistema cromatográfico. Essas informações devem estar disponíveis para que o sistema seja simulado.

$$c_{Q150-I,e} = \frac{M_{P,Q150-I}}{V_{CCE}} \quad (C.5)$$

$$c_{Q150-II,e} = \frac{M_{P,Q150-II}}{V_{CCE}} \quad (C.6)$$

$$c_{Q300,e} = \frac{M_{P,Q300}}{V_{CCE}} \quad (C.7)$$

$$c_{Q1000,e} = \frac{M_{P,Q1000}}{V_{CCE}} \quad (C.8)$$

- Uma vez conhecidos todos os parâmetros fixos ( $L$ ,  $u$ ,  $N_t$  e  $\epsilon_T$ ) e concentrações do extrato celular ( $c_{Q150-I,e}$ ,  $c_{Q150-II,e}$ ,  $c_{Q300,e}$ ,  $c_{Q1000,e}$ ) o sistema descrito pela seção 3.3.4 foi modelado.
- Como condições iniciais, foram consideradas nulas as concentrações dos “i” grupos proteicos no interior da coluna, tanto na fase sólida quanto na fase líquida. A coluna cromatográfica foi dividida em “ $N_t$ ” pratos teóricos.
- Para a modelagem no software EMSO, as equações C.9 a C.12 referentes ao modelo do equilíbrio dispersivo foram readaptadas, pois o programa não reconhece o elemento diferencial de distância  $dz$ . Por isso, as equações diferenciais parciais foram aproximadas por elementos finitos na variável distância ( $z$ ). A aproximação é descrita pelas equações C.13 a C.20 para o primeiro prato, onde  $c_{i,e}$  é a concentração de cada grupo “i” na entrada

do prato; C.21 a C.28 para os pratos 2 a  $N_T-1$  e C.29 a C.36 para o prato  $N_T$ . Em todos os casos  $\Delta Z$  é o comprimento de cada prato teórico, conforme mostra a equação C.37.

- Para complementar o modelo do equilíbrio dispersivo e fornecer o número de equações necessárias para simulação do sistema, as isotermas de ação em massa estérica (C.38 a C.42) e de Langmuir modificada (C.43 a C.54) foram utilizadas separadamente. No primeiro caso, os parâmetros  $K_{Ads,Q150-I}$ ,  $K_{Ads,Q150-II}$ ,  $K_{Ads,Q300}$ ,  $K_{Ads,Q1000}$ ,  $K_{Des,Q150-I}$ ,  $K_{Des,Q150-II}$ ,  $K_{Des,Q300}$ ,  $K_{Des,Q1000}$ ,  $V_{Q150-I}$ ,  $V_{Q150-II}$ ,  $V_{Q300}$ ,  $V_{Q1000}$ ,  $\alpha_{Q150-I}$ ,  $\alpha_{Q150-II}$ ,  $\alpha_{Q300}$  e  $\alpha_{Q1000}$  foram ajustados às curvas cromatográficas de cada pico específico. No segundo caso, os parâmetros  $q_{max,Q150-I}$ ,  $q_{max,Q150-II}$ ,  $q_{max,Q300}$ ,  $q_{max,Q1000}$ ,  $\Psi_{Q150-I}$ ,  $\Psi_{Q150-II}$ ,  $\Psi_{Q300}$ ,  $\Psi_{Q1000}$ ,  $K_{D,Q150-I}$ ,  $K_{D,Q150-II}$ ,  $K_{D,Q300}$ ,  $K_{D,Q1000}$ ,  $\eta_{Q150-I}$ ,  $\eta_{Q150-II}$ ,  $\eta_{Q300}$  e  $\eta_{Q1000}$  foram ajustados.

*Modelo do equilíbrio dispersivo:*

$$\frac{\partial c_{Q150-I}}{\partial t} + \frac{(1-\epsilon_T)}{\epsilon_T} \frac{\partial q_{Q150-I}}{\partial t} + u \frac{\partial c_{Q150-I}}{\partial z} = D_a \frac{\partial^2 c_{Q150-I}}{\partial z^2} \quad (C.9)$$

$$\frac{\partial c_{Q150-II}}{\partial t} + \frac{(1-\epsilon_T)}{\epsilon_T} \frac{\partial q_{Q150-II}}{\partial t} + u \frac{\partial c_{Q150-II}}{\partial z} = D_a \frac{\partial^2 c_{Q150-II}}{\partial z^2} \quad (C.10)$$

$$\frac{\partial c_{Q300}}{\partial t} + \frac{(1-\epsilon_T)}{\epsilon_T} \frac{\partial q_{Q300}}{\partial t} + u \frac{\partial c_{Q300}}{\partial z} = D_a \frac{\partial^2 c_{Q300}}{\partial z^2} \quad (C.11)$$

$$\frac{\partial c_{Q1000}}{\partial t} + \frac{(1-\epsilon_T)}{\epsilon_T} \frac{\partial q_{Q1000}}{\partial t} + u \frac{\partial c_{Q1000}}{\partial z} = D_a \frac{\partial^2 c_{Q1000}}{\partial z^2} \quad (C.12)$$

*Adaptações de equações diferenciais no primeiro prato teórico:*

$$\frac{\partial c_{Q150-I}}{\partial z} = \frac{c_{Q150-I}(1) - c_{Q150-I,e}}{\Delta Z} \quad (C.13)$$

$$\frac{\partial^2 c_{Q150-I}}{\partial z^2} = \frac{c_{Q150-I}(2) - 2 \cdot c_{Q150-I}(1) + c_{Q150-I,e}}{\Delta Z^2} \quad (C.14)$$

$$\frac{\partial c_{Q150-II}}{\partial z} = \frac{c_{Q150-II}(1) - c_{Q150-II,e}}{\Delta Z} \quad (C.15)$$

$$\frac{\partial^2 c_{Q150-II}}{\partial z^2} = \frac{c_{Q150-II}(2) - 2 \cdot c_{Q150-II}(1) + c_{Q150-II,e}}{\Delta Z^2} \quad (C.16)$$

$$\frac{\partial c_{Q300}}{\partial z} = \frac{c_{Q300}(1) - c_{Q300,e}}{\Delta Z} \quad (C.17)$$

$$\frac{\partial^2 c_{Q300}}{\partial z^2} = \frac{c_{Q300}(2) - 2 \cdot c_{Q300}(1) + c_{Q300,e}}{\Delta Z^2} \quad (C.18)$$

$$\frac{\partial c_{Q1000}}{\partial z} = \frac{c_{Q1000}(1) - c_{Q1000,e}}{\Delta Z} \quad (C.19)$$

$$\frac{\partial^2 c_{Q1000}}{\partial z^2} = \frac{c_{Q1000}(2) - 2 \cdot c_{Q1000}(1) + c_{Q1000,e}}{\Delta Z^2} \quad (C.20)$$

*Adaptações de equações diferenciais nos pratos teóricos 2 a  $N_i - 1$ :*

$$\frac{\partial c_{Q150-I}}{\partial z} = \frac{c_{Q150-I}(N+1) - c_{Q150-I}(N-1)}{2 \cdot \Delta Z} \quad (C.21)$$

$$\frac{\partial^2 c_{Q150-I}}{\partial z^2} = \frac{c_{Q150-I}(N+1) - 2 \cdot c_{Q150-I}(N) + c_{Q150-I}(N-1)}{\Delta Z^2} \quad (C.22)$$

$$\frac{\partial c_{Q150-II}}{\partial z} = \frac{c_{Q150-II}(N+1) - c_{Q150-II}(N-1)}{2 \cdot \Delta Z} \quad (C.23)$$

$$\frac{\partial^2 c_{Q150-II}}{\partial z^2} = \frac{c_{Q150-II}(N+1) - 2 \cdot c_{Q150-II}(N) + c_{Q150-II}(N-1)}{\Delta Z^2} \quad (C.24)$$

$$\frac{\partial c_{Q300}}{\partial z} = \frac{c_{Q300}(N+1) - c_{Q300}(N-1)}{2 \cdot \Delta Z} \quad (C.25)$$

$$\frac{\partial^2 c_{Q300}}{\partial z^2} = \frac{c_{Q300}(N+1) - 2 \cdot c_{Q300}(N) + c_{Q300}(N-1)}{\Delta Z^2} \quad (C.26)$$

$$\frac{\partial c_{Q1000}}{\partial z} = \frac{c_{Q1000}(N+1) - c_{Q1000}(N-1)}{2 \cdot \Delta Z} \quad (C.27)$$

$$\frac{\partial^2 c_{Q1000}}{\partial z^2} = \frac{c_{Q1000}(N+1) - 2 \cdot c_{Q1000}(N) + c_{Q1000}(N-1)}{\Delta Z^2} \quad (C.28)$$

*Adaptações de equações diferenciais no prato teórico  $N_t$ :*

$$\frac{\partial c_{Q150-I}}{\partial z} = \frac{c_{Q150-I}(N) - c_{Q150-I}(N-1)}{\Delta Z} \quad (C.29)$$

$$\frac{\partial^2 c_{Q150-I}}{\partial z^2} = \frac{c_{Q150-I}(N) - 2 \cdot c_{Q150-I}(N-1) + c_{Q150-I}(N-2)}{\Delta Z^2} \quad (C.30)$$

$$\frac{\partial c_{Q150-II}}{\partial z} = \frac{c_{Q150-II}(N) - c_{Q150-II}(N-1)}{\Delta Z} \quad (C.31)$$

$$\frac{\partial^2 c_{Q150-II}}{\partial z^2} = \frac{c_{Q150-II}(N) - 2 \cdot c_{Q150-II}(N-1) + c_{Q150-II}(N-2)}{\Delta Z^2} \quad (C.32)$$

$$\frac{\partial c_{Q300}}{\partial z} = \frac{c_{Q300}(N) - c_{Q300}(N-1)}{\Delta Z} \quad (C.33)$$

$$\frac{\partial^2 c_{Q300}}{\partial z^2} = \frac{c_{Q300}(N) - 2 \cdot c_{Q300}(N-1) + c_{Q300}(N-2)}{\Delta Z^2} \quad (C.34)$$

$$\frac{\partial c_{Q1000}}{\partial z} = \frac{c_{Q1000}(N) - c_{Q1000}(N-1)}{\Delta Z} \quad (C.35)$$

$$\frac{\partial^2 c_{Q1000}}{\partial z^2} = \frac{c_{Q1000}(N) - 2 \cdot c_{Q1000}(N-1) + c_{Q1000}(N-2)}{\Delta Z^2} \quad (C.36)$$

$$\Delta Z = \frac{L}{N_t} \quad (C.37)$$

*Isoterma de ação em massa estérica (Steric mass action isotherm)*

$$\frac{\partial q_{Q150-I}}{\partial t} = K_{Ads, Q150-I} c_{Q150-I} Q_{Q150-I}^{V_{Q150-I}} - K_{Des, Q150-I} q_{Q150-I} c_{salt}^{V_{Q150-I}} \quad (C.38)$$

$$\frac{\partial q_{Q150-II}}{\partial t} = K_{Ads, Q150-II} c_{Q150-II} Q_{Q150-II}^{V_{Q150-II}} - K_{Des, Q150-II} q_{Q150-II} c_{salt}^{V_{Q150-II}} \quad (C.39)$$

$$\frac{\partial q_{Q300}}{\partial t} = K_{Ads, Q300} c_{Q300} Q_{Q300}^{V_{Q300}} - K_{Des, Q300} q_{Q300} c_{salt}^{V_{Q300}} \quad (C.40)$$

$$\frac{\partial q_{Q1000}}{\partial t} = K_{Ads, Q1000} c_{Q1000} Q_{Q1000}^{V_{Q1000}} - K_{Des, Q1000} q_{Q1000} c_{salt}^{V_{Q1000}} \quad (C.41)$$

$$Q = \wedge - (q_{Q150-I} (v_{Q150-I} + \alpha_{Q150-I})) - (q_{Q150-II} (v_{Q150-II} + \alpha_{Q150-II})) - (q_{Q300} (v_{Q300} + \alpha_{Q300})) - (q_{Q1000} (v_{Q1000} + \alpha_{Q1000})) \quad (C.42)$$

*Isoterma de Langmuir modificada*

$$q_{Q150-I} = \frac{q'_{max, Q150-I} c_{Q150-I}}{(c_{Q150-I} + K'_{D, Q150-I})} \quad (C.43)$$

$$q'_{max, Q150-I} = \frac{q_{max, Q150-I}}{1 + \frac{c_{salt}}{\Psi_{Q150-I}}} \quad (C.44)$$

$$K'_{D, Q150-I} = \frac{\eta_{Q150-I} K_{D, Q150-I}}{\eta_{Q150-I} - c_{salt}} \quad (C.45)$$

$$q_{Q150-II} = \frac{q'_{\max,Q150-II} c_{Q150-II}}{(c_{Q150-II} + K'_{D,Q150-II})} \quad (C.46)$$

$$q'_{\max,Q150-II} = \frac{q_{\max,Q150-II}}{1 + \frac{c_{salt}}{\psi_{Q150-II}}} \quad (C.47)$$

$$K'_{D,Q150-II} = \frac{\eta_{Q150-II} K_{D,Q150-II}}{\eta_{Q150-II} - c_{salt}} \quad (C.48)$$

$$q_{Q300} = \frac{q'_{\max,Q300} c_{Q300}}{(c_{Q300} + K'_{D,Q300})} \quad (C.49)$$

$$q'_{\max,Q300} = \frac{q_{\max,Q300}}{1 + \frac{c_{salt}}{\psi_{Q300}}} \quad (C.50)$$

$$K'_{D,Q300} = \frac{\eta_{Q300} K_{D,Q300}}{\eta_{Q300} - c_{salt}} \quad (C.51)$$

$$q_{Q1000} = \frac{q'_{\max,Q1000} c_{Q1000}}{(c_{Q1000} + K'_{D,Q1000})} \quad (C.52)$$

$$q'_{\max,Q1000} = \frac{q_{\max,Q1000}}{1 + \frac{c_{salt}}{\psi_{Q1000}}} \quad (C.53)$$

$$K'_{D,Q1000} = \frac{\eta_{Q1000} K_{D,Q1000}}{\eta_{Q1000} - c_{salt}} \quad (C.54)$$

## APÊNDICE D: Simulação da cromatografia de troca aniônica com o modelo de Langmuir competitivo

Além das isotermas de adsorção apresentadas no capítulo 3, a isoterma de Langmuir competitiva também foi testada para simular a cromatografia de troca aniônica. Todos os procedimentos apresentados no capítulo 3 e Apêndice C foram os mesmos. No entanto, a equação 2.31 (seção 2.6.C) foi usada para complementar o modelo do equilíbrio dispersivo. A estrutura desta equação representa competição entre as “i” moléculas simuladas, ou seja, os “i” grupos proteicos representados pelos picos cromatográficos. Diferente das outras isotermas, neste caso não há nenhuma variável que represente a concentração de sal no sistema. Portanto, consideramos o NaCl como um agente competidor “i”, conforme pode ser verificado nas equações D.1 a D.5. Além disso, também acrescentamos à simulação uma equação do equilíbrio dispersivo para NaCl (D.6), a qual foi adaptada no formato das equações D.7, D.8 e D.9 nos pratos 1, 2 a Nt-1 e Nt da coluna, respectivamente. Os parâmetros ajustados foram  $q_{\max,Q150-I}$ ,  $q_{\max,Q150-II}$ ,  $q_{\max,Q300}$ ,  $q_{\max,Q1000}$ ,  $q_{\max,NaCl}$ ,  $K_{Q150-I}$ ,  $K_{Q150-II}$ ,  $K_{Q300}$ ,  $K_{Q1000}$  e  $K_{NaCl}$ .

$$q_{Q150-I} = q_{\max,Q150-I} K_{Q150-I} \frac{c_{Q150-I}}{(1 + K_{Q150-I} c_{Q150-I} + K_{Q150-II} c_{Q150-II} + K_{Q300} c_{Q300} + K_{Q1000} c_{Q1000} + K_{NaCl} c_{NaCl})} \quad (D.1)$$

$$q_{Q150-II} = q_{\max,Q150-II} K_{Q150-II} \frac{c_{Q150-II}}{(1 + K_{Q150-I} c_{Q150-I} + K_{Q150-II} c_{Q150-II} + K_{Q300} c_{Q300} + K_{Q1000} c_{Q1000} + K_{NaCl} c_{NaCl})} \quad (D.2)$$

$$q_{Q300} = q_{\max,Q300} K_{Q300} \frac{c_{Q300}}{(1 + K_{Q150-I} c_{Q150-I} + K_{Q150-II} c_{Q150-II} + K_{Q300} c_{Q300} + K_{Q1000} c_{Q1000} + K_{NaCl} c_{NaCl})} \quad (D.3)$$

$$q_{Q1000} = q_{\max,Q1000} K_{Q1000} \frac{c_{Q1000}}{(1 + K_{Q150-I} c_{Q150-I} + K_{Q150-II} c_{Q150-II} + K_{Q300} c_{Q300} + K_{Q1000} c_{Q1000} + K_{NaCl} c_{NaCl})} \quad (D.4)$$

$$q_{NaCl} = q_{\max,NaCl} K_{NaCl} \frac{c_{NaCl}}{(1 + K_{Q150-I} c_{Q150-I} + K_{Q150-II} c_{Q150-II} + K_{Q300} c_{Q300} + K_{Q1000} c_{Q1000} + K_{NaCl} c_{NaCl})} \quad (D.5)$$

$$\frac{\partial c_{NaCl}}{\partial t} + \frac{(1 - \epsilon_T)}{\epsilon_T} \frac{\partial q_{NaCl}}{\partial t} + u \frac{\partial c_{NaCl}}{\partial z} = D_a \frac{\partial^2 c_{NaCl}}{\partial z^2} \quad (D.6)$$

$$\frac{\partial c_{NaCl}}{\partial z} = \frac{c_{NaCl}(1) - c_{NaCl,e}}{\Delta Z} \quad (D.7)$$

$$\frac{\partial c_{\text{NaCl}}}{\partial z} = \frac{c_{\text{NaCl}}(N+1) - c_{\text{NaCl}}(N-1)}{2 \cdot \Delta Z} \quad (\text{D.8})$$

$$\frac{\partial c_{\text{NaCl}}}{\partial z} = \frac{c_{\text{NaCl}}(N) - c_{\text{NaCl}}(N-1)}{\Delta Z} \quad (\text{D.9})$$

A Tabela D.1 mostra os valores dos parâmetros ajustados. As Figuras D.1 e D.2 mostram os gráficos das simulações obtidas para o ajuste e para a validação do modelo, respectivamente. Mais uma vez, os valores de  $\text{RMSE}_{\text{Est}}$  e  $\text{RMSE}_{\text{Val}}$  apresentam a mesma ordem de grandeza ( $1,8 \times 10^{-4}$  e  $2,6 \times 10^{-4}$ , respectivamente), o que mostra que o modelo de Langmuir competitivo possui capacidade preditiva e pode ser utilizado para descrever a cromatografia de troca aniônica para misturas complexas. Apesar disso, na Figura D.2 é possível se observar maior desvio entre a simulação e os dados experimentais do pico Q1000, expondo uma falha de previsão do modelo nesta situação. Isso pode ter acontecido devido ao menor número de parâmetros ajustados no modelo de Langmuir competitivo (10 parâmetros) em relação aos outros dois modelos (16 parâmetros, veja o apêndice C), o que reduziu sua capacidade preditiva. Além disso, a maior quantidade em massa de proteínas Q1000 processada na amostra  $\text{CCE}_B$  em relação à amostra  $\text{CCE}_A$  (Tabela 3.1) também testou a capacidade do modelo e foi responsável pela redução de sua eficiência.

Tabela D.1. Parâmetros ajustados durante a simulação usando o modelo de Langmuir competitivo.

AEC	Eluição	K (mL.g <sup>-1</sup> )	q <sub>max</sub> (g.mL <sup>-1</sup> )
<b>I</b>	Q150-I	32,6	1,96
	Q150-II	92,0	1,76
	Q300	28,0	2,75
	Q1000	296,4	5,27
	NaCl	1037,2	0,004

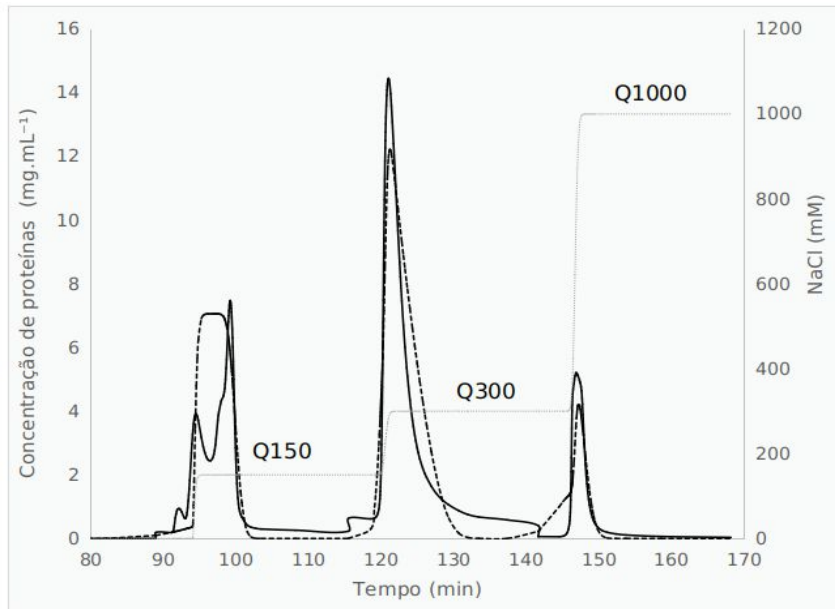


Figura D.1. Perfis de eluição de proteínas da mistura  $CCE_A$  na AEC-I (estimativa de parâmetros): Dados experimentais (linha sólida), perfis simulados usando a isoterma de Langmuir competitiva (linha pontilhada escura) e perfis de eluição usando 150 mM, 300 mM e 1000 mM de NaCl (linha pontilhada clara).

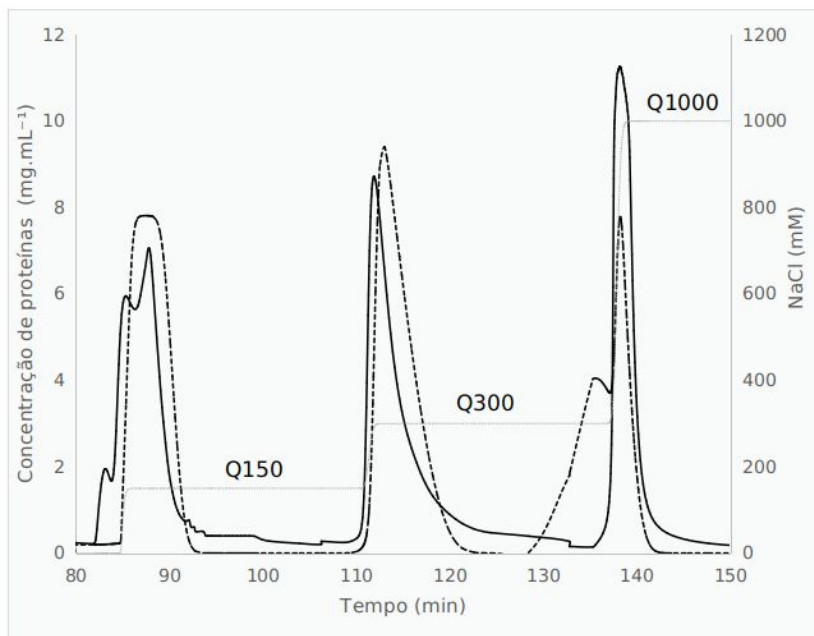


Figura D.2. Perfis de eluição de proteínas da mistura  $CCE_B$  na AEC-II (validação do modelo): Dados experimentais (linha sólida), perfis simulados usando a isoterma de Langmuir competitiva (linha pontilhada escura), e perfis de eluição usando 150 mM, 300 mM e 1000 mM de NaCl (linha pontilhada clara).

**APÊNDICE E: Material suplementar - Código da modelagem e simulação da cromatografia de troca aniônica**

**Modeling and simulation of anion exchange chromatography for purification of proteins in complex mixtures**

Leandro J. Benedini<sup>A,C\*</sup>, Douglas Figueiredo<sup>B</sup>, Joaquin Cabrera-Crespo<sup>B</sup>, Viviane M. Gonçalves<sup>B</sup>, Gabriel G. Silva<sup>A</sup>, Gilson Campani<sup>A,D</sup>, Teresa C. Zangirolami<sup>A</sup> and Felipe F. Furlan<sup>A</sup>

<sup>A</sup> Graduate Program in Chemical Engineering (PPGEQ), Federal University of São Carlos (UFSCar), Rod. Washington Luís, km 235 - SP-310, 13565-905, São Carlos, Brazil

<sup>B</sup> Butantan Institute, Special Vaccine Development Laboratory, Av. Vital Brasil, 1500, Butantã, 05503-900, São Paulo, Brazil

<sup>C</sup> Federal Institute of São Paulo, Av. Pastor José Dutra de Moraes, 239, Distrito Industrial Antônio Zácara, 158080-305, Catanduva, Brazil

<sup>D</sup> Department of Engineering, Federal University of Lavras, 37200-000, Lavras, MG, Brazil

## **Description**

This Supplementary Material aims to present the simulation code used to describe the anion exchange chromatography to purify PspA4Pro, a target protein that must be separated from a pool of protein impurities. Data obtained from simulation which was used to build Figure 3.9 of the related research article [1] is also presented here. The software EMSO™ was used to perform the simulation. All the performed experiments, procedures, modeling and simulations strategies are described in the related research article.

## **Purpose of the Data**

- This file provides the EMSO™ simulation code utilized to describe the anion exchange chromatography to purify a target protein.
- This information can help reproduction of modeling and simulation.
- The simulation code can be adapted to other situations where a target protein must be separated from a pool of impurities.
- Along with the simulation procedure described by the related document, the code provided here can be used to simulate ion exchange chromatography and enhance process performance.

## **Model and simulation strategy**

This section presents the EMSO [2,3] code utilized to describe the anion exchange chromatography. The software can be downloaded online [4]. Simulation code is divided in three parts which are indicated after hashtag (#). In Part 1 some essential variables and parameters are declared. In Part 2 the models of stream, source, valve, mixer and column are defined. Variables, parameters and equations are utilized to describe each model. Part 3 provides the description of the process full structure where devices are created according to

the developed models. These devices are connected, initial conditions are supplied and some parameters are specified. Finally, some important purification column parameters are set, such as porosity ( $\epsilon$ ), number of plates (N), length (L), mobile phase velocity (u), axial dispersion (Da) and resin saturation capacity ( $\lambda$ ). The main variables of this simulation are CA, CB and CS, which represent the concentrations of PspA4Pro, protein impurities (the Q300 pool of proteins) and salt (NaCl) in the fluid phase, respectively, and qA and qB, which represent the concentrations of PspA4Pro and protein impurities in the solid phase. Adsorption constants of A and B are also represented by  $K_{a_i}$ ,  $K_{d_i}$ ,  $x_i$  and  $v_i$ . If necessary, the code can be easily extended to include more components.

### **#PART 1: Declaration of variables and parameters**

```
using "types";
concentration as Real(Unit='(mol/cm^3)', Lower = -1e-6, Upper = 100, Default = 0.5);
constant as Real(Unit='cm^3/s/mol');
adsorbed_fraction as Real(Unit='mol/cm^3');
flow_rate as Real (Unit='cm^3/s');
volume as Real (Unit='cm^3');
length as Real (Unit='cm');
velocity as Real (Unit='cm/s');
porosity as Real;
number_plates as Integer (Default=12, Lower=0, Upper=12000);
ionic_charge as Real;
axial_dispersion as Real (Unit= 'cm^2/s');
ionic_capacity as Real (Unit= 'mol/cm^3');
```

### **#PART 2: Creation of models**

Model stream

VARIABLES

CA as concentration;

CB as concentration;

CS as concentration;

F as flow\_rate;

end

Model source

VARIABLES

out outlet as stream;

end

Model valve

VARIABLES

in inlet as stream;

out outlet as stream;

MaxFlow as flow\_rate(Brief="Maximum mass flow of fluid through the valve");

u as Real(Brief = "Periodical function", Protected = true);

n as Real(Brief = "Regularization function", Protected = true);

topen as time\_h(Brief = "Amount of time the valve stays open", Symbol = "t\_{open}");

tcycle as time\_h(Brief = "Duration of a cycle ", Symbol = "T\_{cycle}");

phase as Real(Brief = "Time phase in hours ahead", Default = 0, Symbol = "\phi", Unit = 'h');

PARAMETERS

eps as Real (Default=1e-6);

pi as angle(Protected = true);

MinFlow as flow\_rate(Brief="Minimum mass flow through the valve",Default = 1e-6);

SET

pi = acos(-1);

EQUATIONS

inlet.CA = outlet.CA;

inlet.CB = outlet.CB;

inlet.CS = outlet.CS;

"Periodical function"

u = sin( pi\*( 2\*time/tcycle + 1.5 - topen/tcycle + 2\*phase/tcycle ) ) + cos( pi \* topen/tcycle );

"Regularization function"

```

n = 0.5 * (1 + tanh(-u/eps));
"Inlet flow in the valve"
MaxFlow * n + MinFlow * (1-n) = inlet.F;
inlet.F = outlet.F;
end

```

Model mixer

VARIABLES

```

in inlet1 as stream;
in inlet2 as stream;
in inlet3 as stream;
in inlet4 as stream;
in inlet5 as stream;
out outlet as stream;

```

EQUATIONS

```

if outlet.F < 1e-6 * 'cm^3/s' then
inlet1.CA = outlet.CA;
inlet1.CB = outlet.CB;
inlet1.CS = outlet.CS;
else
inlet1.CA*inlet1.F+inlet2.CA*inlet2.F+inlet3.CA*inlet3.F+inlet4.CA*inlet4.F+inlet5.CA*inlet5.F = outlet.CA*outlet.F;
inlet1.CB*inlet1.F+inlet2.CB*inlet2.F+inlet3.CB*inlet3.F+inlet4.CB*inlet4.F+inlet5.CB*inlet5.F = outlet.CB*outlet.F;
inlet1.CS*inlet1.F+inlet2.CS*inlet2.F+inlet3.CS*inlet3.F+inlet4.CS*inlet4.F+inlet5.CS*inlet5.F = outlet.CS*outlet.F;
end
end

```

Model column

PARAMETERS

```

E as porosity;

```

A as Real (Unit = 'cm<sup>2</sup>', Lower = 1e-10, Upper = 100, Default = 5); #Sectional area of the column

L as length (Unit = 'cm'); #Column length

N as Integer (Default = 189);

Np as Integer ; #Number of theoretical plates

u as velocity (Unit = 'cm/s'); #Velocity of the mobile phase

Da as axial\_dispersion (Unit = 'cm<sup>2</sup>/s');

KaA as constant (Unit = 'cm<sup>3</sup>/s/mol'); #KAAds for PspA4Pro

KdA as constant (Unit = 'cm<sup>3</sup>/s/mol' ); #KDDes for PspA4Pro

KaB as constant (Unit = 'cm<sup>3</sup>/s/mol'); #KAAds for the pool of proteins

KdB as constant (Unit = 'cm<sup>3</sup>/s/mol' ); #KDDes for the pool of proteins

vA as ionic\_charge; #Number of binding sites of PspA4Pro

xA as ionic\_charge; #Steric factor of PspA4Pro

vB as ionic\_charge; #Number of binding sites of the pool of proteins

xB as ionic\_charge; #Steric factor of the pool of proteins

alfa as ionic\_capacity (Unit = 'mol/cm<sup>3</sup>' ); #Total ionic capacity of the resin

#### VARIABLES

Ct as concentration;

in inlet as stream;

out outlet as stream;

CA(N) as concentration;

CB(N) as concentration;

CS(N) as concentration;

qA(N) as adsorbed\_fraction;

qB(N) as adsorbed\_fraction;

nCA(N) as concentration;

nCB(N) as concentration;

Q(N) as ionic\_capacity;

#### EQUATIONS

#Total mass balance in the column:

u \* A = inlet.F;

outlet.F = inlet.F;

#Outlet conditions are equal to conditions of plate N:

outlet.CA = CA(N);

outlet.CB = CB(N);

outlet.CS = CS(N);

Ct = outlet.CA + outlet.CB; #Total protein

#Equilibrium dispersive model applied to plates 1 and 2:

nCA(1) = CA(1) + ((1-E)/E)\*qA(1);

nCA(2) = CA(2) + ((1-E)/E)\*qA(2);

nCB(1) = CB(1) + ((1-E)/E)\*qB(1);

nCB(2) = CB(2) + ((1-E)/E)\*qB(2);

diff(nCA(1))+(N\*u/L)\*(CA(1)- inlet.CA)= Da\*(inlet.CA - 2\*CA(1) + CA(2))\*(N^2/L^2);

diff(nCA(2))+(N\*u/2/L)\*(CA(2)- inlet.CA)= Da\*(inlet.CA - 2\*CA(1) + CA(2))\*(N^2/L^2);

diff(nCB(1))+(N\*u/L)\*(CB(1)- inlet.CB)= Da\*(inlet.CB - 2\*CB(1) + CB(2))\*(N^2/L^2);

diff(nCB(2))+(N\*u/2/L)\*(CB(2)- inlet.CB)= Da\*(inlet.CB - 2\*CB(1) + CB(2))\*(N^2/L^2);

CS(1) = inlet.CS;

CS(2) = CS(1);

#Steric mass action isotherm applied to plates 1 and 2:

diff(qA(1)) = KaA\*CA(1)\*(abs(Q(1))^vA) - KdA\*qA(1)\*(abs(CS(1))^vA);

diff(qA(2)) = KaA\*CA(2)\*(abs(Q(2))^vA) - KdA\*qA(2)\*(abs(CS(2))^vA);

diff(qB(1)) = KaB\*CB(1)\*(abs(Q(1))^vB) - KdB\*qB(1)\*(abs(CS(1))^vB);

diff(qB(2)) = KaB\*CB(2)\*(abs(Q(2))^vB) - KdB\*qB(2)\*(abs(CS(2))^vB);

Q(1) = alfa - qA(1)\*(vA+xA)- qB(1)\*(vB+xB);

Q(2) = alfa - qA(2)\*(vA+xA)- qB(2)\*(vB+xB);

#Equilibrium dispersive model applied to plates 3 to N-1:

for i in [3:(N-1)] do

nCA(i) = CA(i) + ((1-E)/E)\*qA(i);

nCB(i) = CB(i) + ((1-E)/E)\*qB(i);

```

diff(nCA(i))+(N*u/2/L)*(CA(i+1)- CA(i-1))= Da*(CA(i-1) - 2*CA(i) +
CA(i+1))*(N^2/L^2);
diff(nCB(i))+(N*u/2/L)*(CB(i+1)- CB(i-1))= Da*(CB(i-1) - 2*CB(i) + CB(i+1))*(N^2/L^2);
CS(i) = CS(i-1);

```

#Steric mass action isotherm applied to plates 3 to N-1:

```

diff(qA(i)) = KaA*CA(i)*(abs(Q(i))^vA) - KdA*qA(i)*(abs(CS(i))^vA);
diff(qB(i)) = KaB*CB(i)*(abs(Q(i))^vB) - KdB*qB(i)*(abs(CS(i))^vB);
Q(i) = alfa - qA(i)*(vA+xA)- qB(i)*(vB+xB);
end

```

for i in N do

#Mass balances applied to plate N:

```

nCA(i) = nCA(i-1);
nCB(i) = nCB(i-1);
CA(i) = CA(i-1);
CB(i) = CB(i-1);
CS(i) = CS(i-1);
diff(qA(i)) = KaA*CA(i)*(abs(Q(i))^vA) - KdA*qA(i)*(abs(CS(i))^vA);
diff(qB(i)) = KaB*CB(i)*(abs(Q(i))^vB) - KdB*qB(i)*(abs(CS(i))^vB);
Q(i) = Q(i-1);
end
end

```

### **#PART 3: Description of the process structure**

FlowSheet problem

DEVICES

SouA as source;

SouB as source;  
SouC as source;  
SouD as source;  
SouE as source;  
ValA as valve;  
ValB as valve;  
ValC as valve;  
ValD as valve;  
ValE as valve;  
Mix as mixer;  
Col as column;

#### CONNECTIONS

SouA.outlet to ValA.inlet;  
SouB.outlet to ValB.inlet;  
SouC.outlet to ValC.inlet;  
SouD.outlet to ValD.inlet;  
SouE.outlet to ValE.inlet;  
ValA.outlet to Mix.inlet1;  
ValB.outlet to Mix.inlet2;  
ValC.outlet to Mix.inlet3;  
ValD.outlet to Mix.inlet4;  
ValE.outlet to Mix.inlet5;  
Mix.outlet to Col.inlet;

#### INITIAL

Col.CA=0 \* 'mol/cm^3';  
Col.qA=0 \* 'mol/cm^3';  
Col.CB=0 \* 'mol/cm^3';  
Col.qB=0 \* 'mol/cm^3';

#### SPECIFY

#Sample application

```
SouA.outlet.CA = (0.0005233/47210) * 'mol/cm^3';
SouA.outlet.CB = (0.0002462/47210) * 'mol/cm^3';
SouA.outlet.CS = 0 * 'mol/cm^3';
ValA.MaxFlow = 0.225 * 'cm^3/s';
ValA.topen = 3945*'s';
ValA.tcycle = 10185*'s';
ValA.phase = 0*'s';

#Application of solution to remove non adsorbed proteins
SouB.outlet.CA = 0 * 'mol/cm^3';
SouB.outlet.CB = 0 * 'mol/cm^3';
SouB.outlet.CS = 0 * 'mol/cm^3';
ValB.MaxFlow = 0.225 * 'cm^3/s';
ValB.topen = 1555*'s';
ValB.tcycle = 10185*'s';
ValB.phase = -3945*'s';

#Q150 elution application
SouC.outlet.CA = 0 * 'mol/cm^3';
SouC.outlet.CB = 0 * 'mol/cm^3';
SouC.outlet.CS = (0.00876/58.44) * 'mol/cm^3';
ValC.MaxFlow = 0.225 * 'cm^3/s';
ValC.topen = 1557*'s';
ValC.tcycle = 10185*'s';
ValC.phase = -5500*'s';

#Q300 elution application
SouD.outlet.CA = 0 * 'mol/cm^3';
SouD.outlet.CB = 0 * 'mol/cm^3';
SouD.outlet.CS = (0.01752/58.44) * 'mol/cm^3';
ValD.MaxFlow = 0.225 * 'cm^3/s';
ValD.topen = 1561*'s';
ValD.tcycle = 10185*'s';
```

ValD.phase = -7057\*s';

#Q1000 elution application

SouE.outlet.CA = 0 \* 'mol/cm^3';

SouE.outlet.CB = 0 \* 'mol/cm^3';

SouE.outlet.CS = (0.0584/58.44) \* 'mol/cm^3';

ValE.MaxFlow = 0.225 \* 'cm^3/s';

ValE.topen = 1567\*s';

ValE.tcycle = 10185\*s';

ValE.phase = -8618\*s';

SET

Col.E = 0.90625;

Col.Np = 90;

Col.A = 5.31 \* 'cm^2';

Col.L = 13 \* 'cm';

Col.u = 0.042426\*'cm/s';

Col.Da = Col.L\*Col.u/(2\*Col.Np);

Col.alfa = 0.002117\*'mol/cm^3';

Col.N = 300;

Col.KaA = 675.711\*'cm^3/mol/s';

Col.KdA = 807.626\*'cm^3/mol/s';

Col.xA = 7.69;

Col.vA = 1.71403;

Col.KaB = 291.139\*'cm^3/mol/s';

Col.KdB = 119.341\*'cm^3/mol/s';

Col.xB = 25.3136;

Col.vB = 1.2929;

SET

ValB.MinFlow = 1e-6 \* 'cm^3/s';

ValC.MinFlow = 1e-6 \* 'cm^3/s';

ValD.MinFlow = 1e-6 \* 'cm^3/s';

```
ValE.MinFlow = 1e-6 * 'cm^3/s';
```

#### OPTIONS

```
TimeStart = 0;
```

```
TimeEnd = 10200;
```

```
TimeStep = 10;
```

```
TimeUnit = 's';
```

```
DAESolver (File = "dasslc",RelativeAccuracy=1e-4, AbsoluteAccuracy=1e-4);
```

```
end
```

#### Data

The code described in the previous section can be simulated by EMSO and information related to CA, CB and CS in column outlet can be provided. Simulated data can be extracted in “xls” format by clicking in “Result” tab and “save result”. After extraction, concentration data (CA and CB) was converted to  $\text{mg}\cdot\text{mL}^{-1}$  with PspA4Pro molar mass (42.7 KDa according to [5]) and data was provided as shown in Figure 6 of the related article [1].

#### References

- [1] L.J. Benedini, D. Figueiredo, J. Cabrera-Crespo, V.M. Gonçalves, G.G. Silva, G. Campani, T.C. Zangirolami and F.F. Furlan. Modeling and simulation of anion exchange chromatography for purification of proteins in complex mixtures. *J. Chromatogr. A* (2019).
- [2] R. de P. Soares, A.R. Secchi, EMSO: A new environment for modelling, simulation and optimisation, *Comput. Aided Chem. Eng.* (2003). doi:10.1016/S1570-7946(03)80239-0.
- [3] J.P. Henrique, R. de Sousa, A.R. Secchi, M.A.S.S. Ravagnani, C.B.B. Costa, Optimization of chemical engineering problems with EMSO software, *Comput. Appl. Eng. Educ.* (2018). doi:10.1002/cae.21867.
- [4] Projeto ALSOC, (n.d.). <https://www.enq.ufrgs.br/trac/alsoc> (accessed August 7, 2019).
- [5] ExPASy, ExPASy Bioinformatics resource journal, (n.d.). <https://web.expasy.org/protparam/> (accessed June 29, 2019).