

---

## TABLE DES MATIÈRES

---

I	INTRODUCTION GÉNÉRALE	1
II	ÉTUDE BIBLIOGRAPHIQUE	5
II.1	Les protéines : structure et propriétés . . . . .	5
II.1.1	Structure des protéines . . . . .	5
II.1.1.1	Les acides aminés . . . . .	5
II.1.1.2	Les quatre niveaux de structuration des protéines . . . . .	7
II.1.2	Propriétés physico-chimiques . . . . .	9
II.1.2.1	Masse moléculaire . . . . .	9
II.1.2.2	Charge nette et distribution des charges . . . . .	9
II.2	Les procédés chromatographiques pour la séparation de protéines . . . . .	12
II.2.1	Les différents types de procédés chromatographiques . . . . .	12
II.2.2	Fonctionnement du procédé . . . . .	13
II.3	La chromatographie d'échange d'ions . . . . .	15
II.3.1	Principaux mécanismes et phénomènes . . . . .	15
II.3.2	Les résines échangeuses d'ions . . . . .	16
II.3.2.1	Les matrices . . . . .	17
II.3.2.2	Les groupements fonctionnels . . . . .	18
II.3.2.3	Propriétés des résines . . . . .	18
II.3.3	L'équilibre d'échange d'ions . . . . .	21
II.3.3.1	Les isothermes d'adsorption . . . . .	21
II.3.3.2	L'isotherme de Langmuir . . . . .	23
II.3.3.3	La loi d'action de masse . . . . .	24
II.4	La simulation moléculaire appliquée à l'échange d'ions des protéines . . . . .	26

---

II.4.1	Modélisation moléculaire . . . . .	27
II.4.1.1	Construction des modèles 3D . . . . .	27
II.4.1.2	Modèles de représentation . . . . .	29
II.4.1.3	La mécanique moléculaire . . . . .	31
II.4.1.4	Échantillonnage de l'espace conformationnel . . . . .	33
II.4.1.5	Effets de solvatation . . . . .	37
II.4.1.6	Énergies libres de liaison . . . . .	39
II.4.2	Modélisation protéine-support . . . . .	41
II.4.2.1	Représentation des surfaces . . . . .	41
II.4.2.2	Simulations moléculaires protéine-support . . . . .	44
	Conclusion . . . . .	47
III	ÉTUDE EXPÉRIMENTALE . . . . .	49
	Introduction . . . . .	49
III.1	Matériels . . . . .	50
III.1.1	Support chromatographique . . . . .	50
III.1.2	Protéines . . . . .	52
III.1.2.1	$\alpha$ -chymotrypsine . . . . .	53
III.1.2.2	Lysozyme . . . . .	53
III.1.3	Solutions tampons . . . . .	54
III.2	Méthodes analytiques pour la quantification de protéines . . . . .	55
III.2.1	Analyse par spectrophotométrie UV-visible . . . . .	56
III.2.2	Analyse par HPLC-SEC . . . . .	57
III.2.3	Activité enzymatique . . . . .	62
III.3	Détermination des isothermes . . . . .	64
III.3.1	Expériences en <i>batch</i> . . . . .	64
III.3.1.1	Protocole expérimental . . . . .	64
III.3.1.2	Isothermes de l' $\alpha$ -chymotrypsine obtenues en <i>batch</i> à pH 5 . . . . .	65
III.3.2	Expériences en colonne . . . . .	68
III.3.2.1	Protocole expérimental . . . . .	68

---

III.3.2.2	Distribution de temps de séjour (DTS) de la colonne . . .	70
III.3.2.3	Isothermes de l' $\alpha$ -chymotrypsine à pH 5 . . . . .	71
III.3.2.4	Isothermes de l' $\alpha$ -chymotrypsine à pH 7 . . . . .	79
III.3.3	Equilibre d'échange d'ions en système multi-protéines . . . . .	82
Conclusion	. . . . .	86
IV	SIMULATIONS MOLÉCULAIRES . . . . .	89
Introduction	. . . . .	89
IV.1	Méthodologie . . . . .	91
IV.1.1	Construction du système . . . . .	91
IV.1.1.1	Protéines . . . . .	91
IV.1.1.2	Support chromatographique . . . . .	95
IV.1.1.3	Boîte de simulation . . . . .	96
IV.1.2	Préparation des simulations . . . . .	96
IV.1.2.1	Échantillonnage du système . . . . .	96
IV.1.2.2	Mode opératoire . . . . .	97
IV.1.2.3	Outils d'analyse . . . . .	99
IV.1.3	Méthodes de calcul des paramètres SMA . . . . .	100
IV.1.3.1	Charge caractéristique . . . . .	100
IV.1.3.2	Facteur stérique . . . . .	101
IV.2	Étude d'un système mono-protéine . . . . .	101
IV.2.1	Stabilité de l'adsorption . . . . .	102
IV.2.1.1	Critères de stabilité . . . . .	102
IV.2.1.2	Analyse de l'adsorption et comportement de la protéine	106
IV.2.2	Étude de l'interaction protéine-ligands . . . . .	109
IV.2.2.1	Identification des <i>patches</i> d'adsorption . . . . .	109
IV.2.2.2	Énergies libres de liaison . . . . .	111
IV.2.3	Comparaison des résultats expérimentaux et des simulations . . .	114
IV.2.3.1	Charge caractéristique et facteur stérique . . . . .	114
IV.2.3.2	Constante d'équilibre . . . . .	118

IV.2.4 Conclusion . . . . .	119
IV.3 Influence de l'environnement . . . . .	120
IV.3.1 Effet de la force ionique . . . . .	120
IV.3.2 Effet du pH . . . . .	122
IV.3.2.1 Configuration du système . . . . .	124
IV.3.2.2 Résultats . . . . .	124
IV.3.3 Conclusion . . . . .	127
IV.4 Étude d'un système bi-protéines . . . . .	128
IV.4.1 Simulations avec deux protéines . . . . .	128
IV.4.1.1 Configuration du système . . . . .	129
IV.4.1.2 Résultats . . . . .	130
IV.4.2 Simulations avec quatre protéines . . . . .	132
IV.4.2.1 Configuration du système . . . . .	132
IV.4.2.2 Résultats . . . . .	132
IV.4.3 Conclusion . . . . .	135
V CONCLUSION GÉNÉRALE	137
Bibliographie	143
Annexes	158
A FICHER PDB	159

---

## TABLE DES FIGURES

---

FIGURE II.1	Classification des acides aminés selon les propriétés physico-chimiques de leur chaîne latérale. . . . .	6
FIGURE II.2	Equation bilan la réaction de formation d'une liaison peptidique entre deux acides aminés. . . . .	7
FIGURE II.3	Les quatre niveaux de structuration des protéines. Issu de Abraham (2008). . . . .	8
FIGURE II.4	Illustration de la relation entre la charge nette et le pH. . . . .	11
FIGURE II.5	Surface de la CalB colorée en fonction du potentiel électrostatique aux pH 3, 7 et 10 (rouge = négatif; bleu = positif). Issu de Trodler et al. (2008), reproduit avec permission. Copyright (2008), Elsevier.	12
FIGURE II.6	Les 4 étapes de fonctionnement du procédé chromatographique. . . . .	14
FIGURE II.7	Représentation d'un grain de résine cationique. . . . .	15
FIGURE II.8	Classification des isothermes d'adsorption selon l'IUPAC. . . . .	22
FIGURE II.9	Illustration du phénomène d'encombrement stérique, avec une protéine possédant 4 points de fixation sur un support chromatographique. Issu de (Shen et Frey, 2005), reproduit avec permission. Copyright (2005), Elsevier. . . . .	26
FIGURE II.10	Les principaux modèles de représentation d'un système de simulation moléculaire. . . . .	30
FIGURE II.11	Représentation en Gros Grain de tous les acides aminés selon le champ de force MARTINI. Les différentes couleurs représentent les différents types de particules. Issu de Monticelli et al. (2008), reproduit avec permission. Copyright (2008), American Chemical Society. . . . .	32

---

FIGURE II.12	Représentation en deux dimensions des conditions périodiques aux limites, avec la boîte primaire de simulation encadrée en bleu. . . . .	36
FIGURE II.13	Représentation des différents modèles d'eau rigides et des sites chargés. Illustration issue de la thèse de Tesson (2016). . . . .	38
FIGURE II.14	Exemples de représentations de surfaces en simulation moléculaire : (a) Fibrinogène sur une surface d'or. Issu de Kharazian et al. (2018), reproduit avec permission (CC licence) ; (b) lysozyme sur une surface polyéthylène. Issu de Wei et al. (2011), reproduit avec permission. Copyright (2011) American Chemical Society ; (c) Illustration d'une surface de type SAM alkanéthiol avec une terminaison amine ( $\text{NH}_2$ ). Issu de Agashe et al. (2005), reproduit avec permission. Copyright (2005) American Chemical Society . . . . .	43
FIGURE II.15	Lysozyme sur deux adsorbants, avec (a) une distance entre ligands de 6,51 Å et (b) une distance entre ligands de 12,14 Å. Issu de Lang et al. (2015b), reproduit avec permission. Copyright (2015), Elsevier. . . . .	44
FIGURE II.16	(a) Projection cylindrique de points d'une protéine, puis <i>unwrap</i> et aplatissement du cylindre aboutissant à (b) une carte d'énergie ou d'interaction ( <i>energy map</i> ). Issu de Lang et al. (2015a), reproduit avec permission. Copyright (2015), Elsevier. . . . .	46
FIGURE III.1	Schéma de principe d'un système analytique HPLC-SEC. . . . .	58
FIGURE III.2	Courbe d'étalonnage de la colonne Superdex 75 Increase 10/300 GL, représentant la masse moléculaire des protéines en fonction de leur volume d'élution. . . . .	59
FIGURE III.3	Chromatogrammes HPLC-SEC de deux solutions d' $\alpha$ -chymotrypsine à pH 5 et à des concentrations en NaCl de 50 et 200 mmol/L. . . . .	60
FIGURE III.4	Chromatogramme HPLC-SEC d'une solution contenant 2 g/L de lysozyme ( $V_{\text{élution}}=15,15$ mL) et d' $\alpha$ -chymotrypsine ( $V_{\text{élution}}=12,62$ mL). . . . .	62

FIGURE III.5	(a) Représentation de la courbe de Michaelis-Menten et (b) courbe de l'évolution de la concentration de para-nitrophenol, représentative de l'évolution de la réaction, en fonction du temps. . . . .	63
FIGURE III.6	Isothermes d'adsorption de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF obtenues par la méthode en batch, à pH 5 (tampon citrate 50 mM) et $C_{NaCl} = 0$ mM. . . . .	66
FIGURE III.7	Représentation d'une courbe de percée, exprimée en fonction du volume passé à travers la colonne. . . . .	69
FIGURE III.8	(a) Courbe adimensionnelle $F(t)$ , correspondant à la réponse à une stimulation échelon positif sur la colonne HiTrap SP FF et (b) courbe de distribution du temps de séjour $E(t)$ . . . . .	72
FIGURE III.9	Courbes de percées de l' $\alpha$ -chymotrypsine sur colonne HiTrap SP FF 1 mL, en condition de pH 5 (tampon citrate 50 mM) sans ajout de sel. (a) $C_0 = 0,09$ g/L; (b) $C_0 = 0,19$ g/L; (c) $C_0 = 0,35$ g/L; (d) $C_0 = 0,54$ g/L; (e) $C_0 = 0,74$ g/L; (f) $C_0 = 0,95$ g/L. . . . .	73
FIGURE III.10	Isothermes d'adsorption de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF obtenues par la méthode en colonne, à pH 5 (tampon citrate pH 5). Le modèle SMA avec les paramètres ajustés est représenté en ligne continue. Les lignes discontinues représentent l'analyse de sensibilité ( $\pm 10\%$ sur les données expérimentales) du modèle. . . . .	75
FIGURE III.11	Gradient d'élution de l' $\alpha$ -chymotrypsine à pH 5. La ligne continue représente l'absorbance à 280 nm et la ligne discontinue la conductivité en sortie de colonne. . . . .	78
FIGURE III.12	Courbe de percée de l' $\alpha$ -chymotrypsine sur HiTrap SP FF 1 mL, à pH 5 et $C_{NaCl} = 400$ mM. . . . .	79
FIGURE III.13	Isothermes d'adsorption de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF obtenues par la méthode en colonne (a) à pH 7 (tampon phosphate) et (b) à pH 5 (tampon citrate) et 7. . . . .	81

---

FIGURE III.14	Isothermes d'adsorption du lysozyme et de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF à pH 5 (tampon citrate 50 mM). Chacune des isothermes a été mesurée indépendamment à partir d'une solution mono-protéine. . . . .	84
FIGURE III.15	Isothermes d'adsorption en système bi-protéines du lysozyme et de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF à pH 5 (tampon citrate 50 mM). Les symboles $\bullet$ et $\blacktriangle$ représentent les valeurs issues de l'essai n° 1 et les symboles $\circ$ et $\triangle$ représentent l'essai n°2. . .	85
FIGURE IV.1	Surface de l' $\alpha$ -chymotrypsine (PDB-ID : 1YPH) colorée en fonction du potentiel électrostatique à : (a) pH 5 (charge +7); (b) pH 7 (charge +5) et (c) pH 9 (charge 0). . . . .	93
FIGURE IV.2	Surface du lysozyme (PDB-ID : 1AKI) colorée en fonction du potentiel électrostatique à : (a) pH 5 (charge +10); (b) pH 8 (charge +7) et (c) pH 11 (charge 0). . . . .	94
FIGURE IV.3	(a) Représentation 3D "ball-and-stick" d'un ligand de la SP Sepharose FF et (b) la boîte de simulation primaire comprenant la surface chromatographique avec la couche de contre-ions (en bas), l' $\alpha$ -chymotrypsine (représentation "cartoon"), les ions chlorure (cyan) et sodium (bleu) et l'eau (surface bleu transparent). .	97
FIGURE IV.4	Représentation des six orientations initiales de la protéine étudiées. Chaque point sur la surface de la protéine (b) représente la face dirigée vers le support chromatographique en début de chaque simulation. Les différentes orientations sont obtenues à partir des rotations en -y et -z indiqués sur la carte (a). . . . .	98

FIGURE IV.5	Illustration du calcul du facteur stérique $\sigma$ par la méthode de l'enveloppe convexe. Les coordonnées cartésiennes -x et -y des atomes de la protéine (représentation cartoon avec surface transparente), représentées par des sphères blanches, sont projetées sur une surface plane (gris). L'enveloppe convexe de l'ensemble des atomes est dessinée par des lignes rouges. . . . .	102
FIGURE IV.6	Évolution de la distance minimale entre les ligands et l' $\alpha$ -chymotrypsine <i>PROT-LIG</i> au cours du temps, pour chaque orientation (O1-6, figures (a) à (f)) et les répétitions correspondantes (MD1 en rouge, MD2 en vert et MD3 en bleu). Les lignes en pointillé verticales correspondent au temps d'adsorption et les lignes en pointillé horizontales correspondent au seuil de 0,2 nm. . . . .	104
FIGURE IV.7	Évolution de la distance minimale entre les ligands et le centre de masse de l' $\alpha$ -chymotrypsine <i>COM-LIG</i> au cours du temps, pour chaque orientation (O1-6, figures (a) à (f)) et les répétitions correspondantes (MD1 en rouge, MD2 en vert et MD3 en bleu). Les lignes en pointillé verticales correspondent au temps d'adsorption et les lignes en pointillé horizontales correspondent au seuil de 3 nm. . . . .	105
FIGURE IV.8	Évolution du RMSD de l' $\alpha$ -chymotrypsine au cours du temps, pour chaque orientation (O1-6, figures (a) à (f)) et les répétitions correspondantes (MD1 en rouge, MD2 en vert et MD3 en bleu). . . . .	107
FIGURE IV.9	Visualisation au cours du temps de l'adsorption de l' $\alpha$ -chymotrypsine sur la SP Sepharose lors de la simulation MD2-O6. Pour chaque cliché, le temps de simulation et les distances PROT-LIG et COM-LIG sont données. Les résidus Lys169 et Lys170, correspondant aux principaux résidus impliqués dans l'adsorption, sont représentés sous forme de sphères, le reste de la protéine étant montrée sous une représentation <i>cartoon</i> . . . . .	109

---

FIGURE IV.10	Taux de présence durant les dernières 10 ns des simulations des résidus chargés positivement impliqués dans les adsorptions. . . .	110
FIGURE IV.11	Représentation de l' $\alpha$ -chymotrypsine adsorbée à la fin de 4 simulations, montrant ainsi les 4 <i>patches</i> d'adsorption identifiés. Les clichés sont issus des simulations : (a) MD3-O1, (b) MD3-O4, (c) MD1-O6 et (d) MD3-O2. Les 4 patches sont mis en évidence par des couleurs : P1 (magenta), P2 (vert), P3 (bleu) et P4 (orange). La Figure (e) montre la protéine et ses 4 <i>patches</i> sous les mêmes vues que sur la Figure IV.1. . . . .	112
FIGURE IV.12	Évolution de la distance minimale entre les ligands et la protéine <i>PROT-LIG</i> au cours du temps pour chaque orientation (O1-6, figures (a) à (f)). Cas d'étude à pH 5 et $C_{NaCl} = 500$ mM. . . . .	123
FIGURE IV.13	Illustration des deux configurations de départ pour les simulations en système bi-protéines, représentant l' $\alpha$ -chymotrypsine en bleu et le lysozyme en rouge, sur la surface chromatographique (plan gris). La configuration n°1 (Figure a) montre l' $\alpha$ Chy avec l'orientation initiale 1 et le lysozyme avec une orientation aléatoire et la configuration n°2 (Figure b) montrent les deux protéines avec une rotation en $180^\circ$ sur l'axe -y par rapport à la configuration 1. . .	129
FIGURE IV.14	Résultats des quatre simulations en système bi-protéines, comprenant l' $\alpha$ -chymotrypsine (en bleu cyan) et le lysozyme (en rouge), à pH 5 et $C_{NaCl} = 100$ mM. La dernière <i>frame</i> de chaque simulation, <i>i.e.</i> (a) MD1-CONF1, (b) MD2-CONF1, (c) MD1-CONF2 et (d) MD2-CONF2, sont présentées ainsi que les distances minimales Protéines–Ligands et Protéine–Protéine. . . . .	131

---

FIGURE IV.15	Illustration de la configuration de départ pour les simulations en système bi-protéines, comprenant deux $\alpha$ -chymotrypsines et deux lysozymes. La boîte de simulation est représentée par des lignes bleues et le support chromatographique par un plan gris. La dernière <i>frame</i> de la simulation MD1-CONF1 a été utilisée comme configuration initiale, à laquelle ont été ajoutées les deux protéines $\alpha$ Chy et lysozyme. . . . .	133
FIGURE IV.16	Résultats des trois simulations en système bi-protéines, comprenant deux $\alpha$ -chymotrypsines (en bleu cyan et vert) et deux lysozymes (en rouge et orange), à pH 5 et $C_{NaCl} = 100$ mM. La dernière <i>frame</i> de chaque simulation, <i>i.e.</i> (a) MD1-CONF1, (b) MD2-CONF1 et (c) MD3-CONF1, sont présentées ainsi que les distances minimales Protéines–Ligands et Protéine–Protéine. . .	134
FIGURE A.1	Extrait d’un fichier PDB du lysozyme (code PDB : 1AKI), représentant les informations sur la séquence en acide-aminés, les hélices $\alpha$ et les feuillets $\beta$ , les ponts disulfures, ainsi que les informations sur les 14 premiers atomes de la protéine. . . . .	159



---

## LISTE DES TABLEAUX

---

TABLEAU II.1	pKa des acides aminés (Nelson et al., 2008) . . . . .	10
TABLEAU II.2	Différents types de groupements fonctionnels ou ligands courants.	19
TABLEAU III.1	Tableau récapitulatif des principales propriétés de la résine SP Sepharose FF. . . . .	52
TABLEAU III.2	Tableau récapitulatif des principales propriétés de l' $\alpha$ -chymotrypsine et du lysozyme. . . . .	54
TABLEAU III.3	Conditions opératoires pour la détermination des isothermes d'ad- sorption de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF par la mé- thode en batch, à $T = 21^\circ\text{C}$ , pH 5 (tampon citrate 50 mM) et $C_{NaCl} = 0$ mM. . . . .	65
TABLEAU III.4	Expériences en batch correspondant à l'essai n°3 : points de l'iso- therme à hautes concentrations initiales en protéine, à pH 5 (tam- pon citrate 50 mM) et $C_{NaCl} = 0$ mM. . . . .	67
TABLEAU III.5	Paramètres de la loi SMA à pH 5, calculés par <i>fitting</i> sur données expérimentales (brutes, -10% et +10%). . . . .	77
TABLEAU III.6	Paramètres de la loi SMA à pH 7 et 5, calculés par fitting sur données expérimentales. . . . .	80
TABLEAU III.7	Conditions opératoires pour la détermination des isothermes d'échange en système bi-protéines ( $\alpha$ -chymotrypsine et lysozyme) sur la ré- sine SP Sepharose FF par la méthode en batch, à $T = 21^\circ\text{C}$ , pH 5 (tampon citrate 50 mM) et $C_{NaCl} = 0$ mM. . . . .	83

TABLEAU III.8	Conditions opératoires pour la détermination de l'isotherme d'adsorption du lysozyme sur la SP Sepharose FF par la méthode en batch, à $T = 21^{\circ}\text{C}$ , pH 5 (tampon citrate 50 mM) et $C_{\text{NaCl}} = 0$ mM. . . . .	83
TABLEAU IV.1	Tableau récapitulatif des principales propriétés de l' $\alpha$ -chymotrypsine et du lysozyme ainsi que des codes PDB correspondants. . . . .	95
TABLEAU IV.2	Temps de simulations totaux et temps d'adsorption pour chacune des 18 simulations. . . . .	108
TABLEAU IV.3	Valeurs d'énergies libres totales et des différentes contributions, issues des calculs par la méthode MM-PBSA pour chacune des 18 simulations. . . . .	113
TABLEAU IV.4	Charges caractéristiques $n$ et facteurs stériques $\sigma$ calculés à partir des simulations. Deux moyennes (non pondérée et pondérée sur l'énergie libre) ainsi que leurs écarts-types ont été calculés et comparés aux valeurs issues des expériences. . . . .	116
TABLEAU IV.5	Tableau récapitulatif des analyses des 18 simulations réalisées en système mono-constituants, pH 5 et concentration en NaCl de 100 mM. . . . .	117
TABLEAU IV.6	Tableau récapitulatif des analyses des 6 simulations réalisées en système mono-constituant, pH 5 et concentration en NaCl de 500 mM. . . . .	121
TABLEAU IV.7	Tableau récapitulatif des analyses des 12 simulations réalisées en système mono-constituant, pH 7 et concentration en NaCl de 100 mM. . . . .	125



---

## INTRODUCTION GÉNÉRALE

---

Les procédés chromatographiques restent à ce jour les opérations unitaires les plus utilisées et essentielles pour la purification des biomolécules. Pourtant, bien que leur efficacité pour obtenir des purifications élevées ne soit plus à prouver, leur coût représente toujours jusqu'à 80-90 % du coût total de production dans l'industrie des biotechnologies (Costa et al., 2014). Ces procédés sont toujours très largement utilisés car en plus d'obtenir des produits finis hautement purifiés et concentrés, les molécules ne sont généralement pas dénaturées par le procédé.

Parmi toutes les techniques de chromatographie existantes, la chromatographie d'échange d'ions est en particulier très appliquée à la purification des protéines. Son principe se base sur la particularité des protéines à présenter une charge globale variant selon le pH environnant, qui peuvent ainsi être séparées par rétention sur des résines échangeuses d'ions. Néanmoins, si ce procédé est particulièrement efficace lorsqu'il est appliqué à des milieux assez simples et pré-purifiés, ce n'est pas forcément le cas dans les milieux bruts et complexes, où plusieurs molécules dont d'autres protéines que celle d'intérêt peuvent être présentes. L'apparition de phénomènes propres à la composition complexe du milieu environnant, qui surviennent à une échelle locale, font que ces mécanismes s'avèrent difficiles à prédire et à contrôler. Aujourd'hui, les procédés industriels ont été essentiellement conçus en se basant sur des modèles empiriques basés sur des observations macroscopiques et se retrouvent pour la plupart sous-optimisés.

C'est donc dans un objectif d'optimisation et de réduction des coûts de production que la demande pour une meilleure compréhension des phénomènes locaux ne cesse de croître. En effet, une meilleure appréhension des mécanismes physico-chimiques survenant à l'échelle atomique pourrait permettre une progression dans les études d'optimisation du procédé. Cela peut également aider la création de nouveaux modèles prédictifs et mécanistiques, et ainsi de s'affranchir peu à peu des modèles empiriques, voire de concevoir des modèles hybrides. Ces informations à l'échelle de l'atome restent cependant encore difficiles d'accès dans les conditions du procédé, et les techniques expérimentales de microscopie actuelles sont difficiles à appliquer dans de telles situations. Toutefois, beau-

coup d'efforts ont été fournis ces dernières années dans ce domaine, avec notamment des méthodes analytiques et expérimentales appliquées notamment à la caractérisation des supports chromatographiques.

Outre par des méthodes expérimentales, il est également possible de tenter de répondre à ces problématiques en utilisant des méthodes computationnelles. Lors des dernières décennies, l'important développement des moyens informatiques a permis l'émergence de nombreuses méthodes de simulation numérique. A ces débuts, la simulation numérique était utilisée en grande partie en supplément à des expériences, dont la modélisation était parfaitement maîtrisée. Avec le temps et la croissance de la puissance informatique, les simulations ont très vite été utilisées afin d'acquérir de nouvelles connaissances, et en particulier lorsque les conditions étudiées ne sont pas réalisables à l'échelle du laboratoire. Aujourd'hui, elles contribuent même à l'élaboration de nouveaux modèles physiques (Kittelmann et al., 2017b).

Ainsi, parmi les approches actuellement proposées, la simulation moléculaire apparaît comme un outil pertinent et performant afin d'améliorer les connaissances au niveau moléculaire. Les simulations moléculaires ont été très largement utilisées pour l'étude des biomolécules, et en particulier des protéines. Le but de la modélisation moléculaire en particulier est de connaître les structures 3D des molécules au niveau atomique et d'étudier leurs interactions, et ainsi de comprendre le comportement des molécules. L'idée est ainsi d'appliquer ces méthodes à des cas d'études de chromatographie d'échange d'ions des protéines afin de mieux comprendre les mécanismes intermoléculaires, mais également de vérifier l'impact de certaines variables physico-chimiques telles que la force ionique et le pH sur ces phénomènes. La force de la simulation est également de pouvoir acquérir des données non accessibles expérimentalement dans des cas de conditions extrêmes et d'échelles de temps et d'espace non réalisables en laboratoire.

Si la dynamique moléculaire a été très largement utilisée dans l'étude des protéines en général, son application à l'étude des phénomènes locaux dans les procédés, et en particulier à la chromatographie reste assez limitée. Malgré tout, plusieurs groupes de recherche se sont déjà focalisés sur cette problématique et ont permis de montrer la pertinence de ce type d'approche pour une meilleure compréhension des procédés chromatographiques. Des protéines modèles, telles que le lysozyme, la ribonucléase A ou encore l'albumine de sérum bovin (BSA), ont notamment été utilisées dans la mise en place de ces méthodes de simulation. Plusieurs niveaux de description du système ont également été proposés, avec des représentations du système atome par atome (modèles tout-atome) ou des représentations plus simplifiées permettant notamment l'étude de systèmes plus complexes (modèles Gros Grains).

C'est donc dans cette démarche de contribuer à l'apport d'informations des phénomènes physico-chimiques à l'échelle atomique que ce travail a été proposé. L'objectif de cette étude est d'initier des études préliminaires grâce à un couplage de deux approches, expérimentales et numériques. Ainsi, la pertinence des simulations moléculaires peut être discutée par confrontation à des données expérimentales obtenues à l'échelle macroscopique. Le but n'est donc pas d'acquérir une connaissance exhaustive de toutes les interactions possibles, mais d'identifier des phénomènes clés qui permettent de mieux décrire les comportements à l'échelle du procédé.

Pour cela, la démarche appliquée dans ce travail est premièrement de définir un système modèle, composé d'une protéine et d'une résine échangeuse d'ions, et des conditions opératoires de référence. Vient ensuite la question de comment relier des observations macroscopiques à des simulations à l'échelle moléculaire : pour cela, à partir des pistes déjà abordées dans la littérature, la loi d'Action de Masse Stérique (loi SMA), qui est actuellement l'un des modèles les plus utilisés pour décrire des équilibres d'échange d'ions appliquées à des macromolécules, est utilisée. En effet, les paramètres de ce modèle peuvent être déterminés à la fois à l'aide d'expériences macroscopiques, mais également à partir des simulations moléculaires. La comparaison de ces paramètres calculés aux deux échelles peut alors servir de moyen de validation de la pertinence des simulations. Par ailleurs, en plus des études sur ce système de référence, l'influence de certaines conditions opératoires comme la force ionique ou le pH pourra être étudiée et son impact sur les mécanismes identifié. Enfin, une fois la méthodologie bien établie et validée pour un système mono-constituant, l'objectif ici est également d'observer le comportement de la protéine cible à l'échelle locale en présence d'une deuxième protéine et d'identifier des interactions spécifiques.

L'ensemble de ces résultats est présenté dans ce manuscrit sous la forme de trois chapitres, avec une introduction générale (**Chapitre I**) et une conclusion avec perspectives (**Chapitre V**).

Le **Chapitre II** est une synthèse bibliographique qui a pour objectif de poser les principales bases théoriques de cette étude. Tout d'abord, une première partie est consacrée à la présentation des protéines, de leur structure ainsi que de leurs propriétés physico-chimiques, importantes dans ce travail. Les procédés chromatographiques sont ensuite présentés de manière non exhaustive, avec une seconde partie centrée sur les bases de la chromatographie d'échange d'ions en particulier. La dernière partie de ce chapitre est consacrée à la simulation moléculaire où les bases théoriques de la modélisation moléculaire sont exposées, suivies d'un état de l'art axé sur la simulation moléculaire appliquée particulièrement à des cas d'interactions entre protéines et supports chromatographiques.

Le **Chapitre III** est entièrement consacré à l'étude expérimentale mise en place et réalisée lors de cette étude. Le choix du système de référence est exposé en même temps que la description des matériels utilisés. Les différentes méthodes analytiques utilisées sont ensuite présentées. Enfin, les résultats des expériences sont présentés dans une dernière partie, elle-même scindée en deux sous-parties présentant les deux méthodes de mesures mise en œuvre. Les résultats issus des études en système multi-constituants et avec des conditions de pH différentes sont également montrés.

Enfinement, le **Chapitre IV** expose toute la méthodologie mise en œuvre pour la réalisation des simulations moléculaires. Les principaux résultats obtenus avec les simulations sont décrits et discutés, et la mise en place d'un protocole pour la détermination des paramètres de la loi SMA est présentée. Les comparaisons entre expériences et simulations moléculaires sont également discutées dans ce chapitre. Les résultats obtenus avec le système de référence sont premièrement présentés, suivis par une étude de l'influence des conditions de force ionique et de pH, et pour finir avec une étude préliminaire d'un système composé de deux protéines.

Enfin, les principaux résultats de ce travail sont synthétisés et les observations les plus pertinentes mises en avant afin d'en extraire les premières conclusions concernant les systèmes étudiés, mais aussi la pertinence de l'approche, la validité des modèles ou les projections entre les deux échelles de travail. La complexité à laquelle ce travail s'est confronté et les réponses intéressantes, mais encore insuffisantes, permettent de décliner un grand nombre de perspectives. Celles considérées les plus nécessaires, ou encore les plus intéressantes, sont indiquées à la fin de ce mémoire.

## II.1 Les protéines : structure et propriétés

Les protéines sont des macromolécules biologiques constituées d'une séquence d'acides aminés. Elles sont présentes dans toutes les cellules vivantes et peuvent assurer une multitude de fonctions, comme par exemple la catalyse de réactions chimiques (enzymes), la transmission de signaux cellulaires (protéines membranaires) ou le maintien des structures cellulaires (protéines structurales) (Whitford, 2013). Il est d'usage de distinguer les peptides, qui possèdent en général moins de 100 acides aminés, et les protéines constituées d'une centaine d'acides aminés ou plus.

### II.1.1 Structure des protéines

#### II.1.1.1 Les acides aminés

Les acides aminés, ou  $\alpha$ -aminoacides, sont des composés organiques comprenant à la fois un groupe fonctionnel carboxyle ( $-\text{COOH}$ ) et amine ( $-\text{NH}_2$ ). Ce sont donc des espèces ampholytes possédant à la fois des propriétés acides et basiques. Il en existe 20 à entrer dans la composition des protéines, qui sont présentés sur la Figure II.1.

Ils possèdent une formule générale de la forme  $\text{R}-\text{CH}(\text{NH}_2)-\text{COOH}$  et diffèrent par la forme du radical R, que l'on appelle la chaîne latérale. Les acides aminés sont par ailleurs souvent classés en quatre groupes selon les propriétés de la chaîne latérale : acide, basique, polaire ou apolaire (Barrett, 2012). Par convention, le carbone auquel sont rattachés les groupements amine et carboxyle est appelé carbone- $\alpha$ , ou carbone asymétrique, ce qui induit une chiralité de la molécule (à l'exception de la glycine). Les acides aminés ont donc des propriétés optiques de forme L (levogyre), qui est la forme standard dans les protéines (Barrett, 2012).

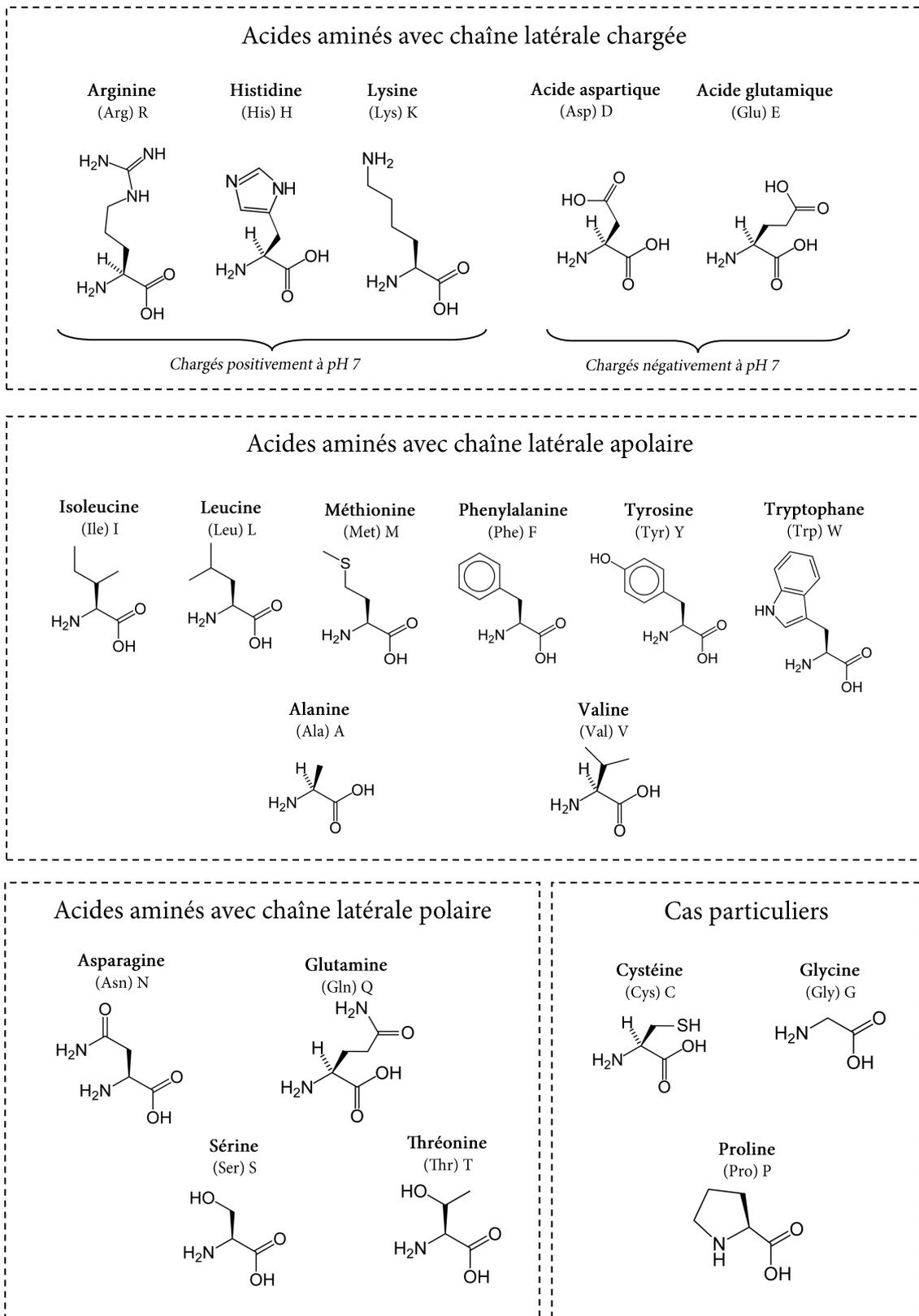


FIGURE II.1 – Classification des acides aminés selon les propriétés physico-chimiques de leur chaîne latérale.

Les acides aminés sont reliés entre eux au sein de la chaîne principale par une liaison peptidique entre le groupe carboxyle du premier et le groupe amine du prochain (voir Figure II.2) : ainsi liés, ils sont également appelés *résidus*.

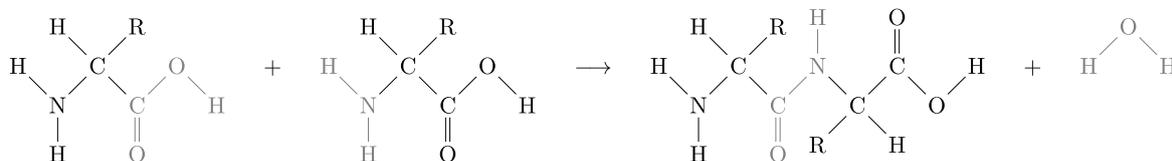


FIGURE II.2 – Equation bilan la réaction de formation d’une liaison peptidique entre deux acides aminés.

### II.1.1.2 Les quatre niveaux de structuration des protéines

Les protéines possèdent quatre niveaux de structure : primaire, secondaire, tertiaire et quaternaire, qui sont illustrés en Figure II.3.

#### STRUCTURE PRIMAIRE

Il s’agit de la séquence en acide aminés, ou enchaînement linéaire : on parle également de *chaîne polypeptidique*, ou encore de *squelette* de la protéine. Cette structure est essentielle, un seul changement pouvant modifier les propriétés de la protéine (Whitford, 2013).

#### STRUCTURE SECONDAIRE

La structure secondaire représente les repliements locaux des acides aminés dans une protéine. Ces repliements permettent de stabiliser la structure de la protéine à travers des liaisons hydrogènes (H) intramoléculaires. Il existe différentes catégories de structure secondaire, les plus connues étant :

- **Les hélices  $\alpha$**  : la chaîne polypeptidique s’enroule autour d’un axe, formant une structure hélicoïdale. Les hélices  $\alpha$  peuvent être représentées schématiquement par une hélice ou un cylindre.
- **Les brins  $\beta$  (feuilletts)** : le brin  $\beta$  possède une conformation très étendue (linéaire). Il n’est pas stable seul car il ne possède aucune liaison H : il n’est stable que dans les feuilletts  $\beta$  (des liaisons H s’établissent entre 2 brins  $\beta$  différents, parallèles ou antiparallèles). Les feuilletts sont représentés conventionnellement par des flèches. Deux brins d’un feuillet sont reliés par des coudes, impliquant 3 à 4 résidus (Petsko et al., 2008).
- **Les boucles** : elles relient les différentes structures secondaires entre elles (hélice/hélice, hélice/brin ou encore deux brins n’appartenant pas au même feuillet).

Souvent situées à la surface de la protéine, elles sont à l'origine des changements de direction de la chaîne principale. Elles représentent environ un tiers des acides aminés (Petsko et al., 2008).

### STRUCTURE TERTIAIRE

La structure tertiaire décrit l'agencement spatial des éléments de structure secondaire en une conformation particulière. Elle est stabilisée par de multiples liaisons dont, par ordre de stabilité décroissante : les liaisons covalentes (ponts disulfures), électrostatiques (pont salin), hydrogènes, hydrophobes et par les interactions de Van der Waals. Les protéines possèdent un ou plusieurs *domaines* qui correspondent à des arrangements de structures tertiaires déterminés d'une partie de la chaîne polypeptidique. Un même domaine peut être retrouvé dans plusieurs protéines différentes (Petsko et al., 2008).

### STRUCTURE QUATERNAIRE

Pour les protéines possédant plusieurs chaînes polypeptidiques, il est possible de distinguer un quatrième niveau de structure. Ainsi, la structure quaternaire décrit l'agencement dans l'espace de plusieurs chaînes polypeptidiques (sous-unités) ainsi que les interactions qui les lient.

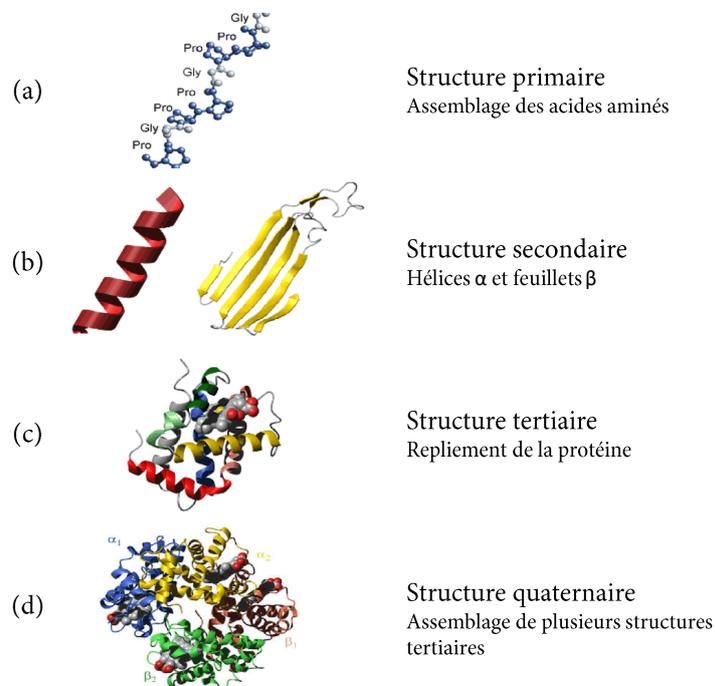


FIGURE II.3 – Les quatre niveaux de structuration des protéines. Issu de Abraham (2008).

## II.1.2 Propriétés physico-chimiques

### II.1.2.1 Masse moléculaire

La masse moléculaire d'une protéine se mesure en Dalton (Da), qui est une unité de masse atomique correspondant à la masse d'un atome d'hydrogène ( $1 \text{ Da} = 1 \text{ g/mol}$ ). Elle peut aller de 2 kDa à 3.000 kDa – comme par exemple la titine, qui est la plus grande protéine connue chez l'être humain (Fulton et Isaacs, 1991). La masse moléculaire est un paramètre très important à prendre en compte en chromatographie car elle influe notamment sur la diffusion des molécules au sein des échangeurs (Hunter et Carta, 2002; Langford et al., 2006). Par exemple, lorsqu'une protéine atteint une certaine taille, il peut y avoir un phénomène d'exclusion de taille qui s'ajoute aux phénomènes déjà présents (comme l'échange d'ions ou l'hydrophobicité).

### II.1.2.2 Charge nette et distribution des charges

La charge nette, ou charge globale, d'une protéine représente la charge moyenne de tous les résidus la composant. En effet, les acides aminés possèdent différents pKa et peuvent se retrouver protonés ou déprotonés selon le pH environnant. Ayant à la fois des propriétés basiques et acides, leur charge varie selon le milieu : lorsque  $\text{pH} < \text{pKa}$ , le groupement  $\text{NH}_2$  est ionisé ( $\text{NH}_3^+$ ) alors que lorsque  $\text{pH} > \text{pKa}$ , le groupement  $\text{COOH}$  est ionisé ( $\text{COO}^-$ ). Même s'il existe des valeurs de pKa connues pour chaque acide aminé (voir Tableau II.1), le pKa réel de chaque résidu peut légèrement varier en fonction de plusieurs paramètres comme la force ionique, les résidus à proximité spatiale ou encore la température.

Cependant, la charge globale de la protéine va surtout dépendre de l'état de protonation des groupements sur les chaînes latérales, les groupements amines et carboxyles ne pouvant être ionisés qu'aux extrémités C- et N-terminales. Seuls certains résidus possèdent un radical -R pouvant être ionisé : ce sont les résidus titrables *i.e.* Arginine (Arg), Lysine (Lys), Histidine (His), Acide Aspartique (Asp), Acide Glutamique (Glu), Cystéine (Cys) et Tyrosine (Tyr).

Il existe un certain pH pour lequel la balance des charges négatives et positives de la protéine s'équilibre : on parle alors de point isoélectrique, ou pI, de la protéine. Ainsi dans ces conditions, la charge nette est nulle (voir Figure II.4).

Dans la plupart des cas, la charge globale de la protéine n'est pas une information suffisante pour déterminer le potentiel de rétention électrostatique d'une protéine. En effet, les charges de chaque résidu titrable peuvent ne pas être réparties équitablement ou uniformément sur la surface de la protéine. Le fait d'obtenir une charge nette nulle signifie uniquement que les charges se compensent. Ainsi, une protéine dans une solution dont le

TABLEAU II.1 – pKa des acides aminés (Nelson et al., 2008)

Acide aminé	Abréviations		pKa			pI
			Carboxyle	Amine	Chaîne latérale	
Acide aspartique	Asp	D	1,88	9,60	3,65	2,77
Acide glutamique	Glu	E	2,19	9,67	4,25	3,22
Alanine	Ala	A	2,34	9,69	-	6,00
Arginine	Arg	R	2,17	9,04	12,48	10,76
Asparagine	Asn	N	2,02	8,80	-	5,41
Cystéine	Cys	C	1,96	10,13	8,18	5,07
Glutamine	Gln	Q	2,17	9,13	-	5,65
Glycine	Gly	G	2,34	9,60	-	5,97
Histidine	His	H	1,82	9,17	6,00	7,59
Isoleucine	Ile	I	2,36	9,60	-	6,02
Leucine	Leu	L	2,36	9,60	-	5,98
Lysine	Lys	K	2,18	8,95	10,53	9,74
Méthionine	Met	M	2,28	9,21	-	5,74
Phénylalanine	Phe	F	1,83	9,13	-	5,48
Proline	Pro	P	1,99	10,60	-	6,30
Sérine	Ser	S	2,21	9,15	-	5,58
Thréonine	Thr	T	2,09	9,10	-	5,60
Tryptophane	Trp	W	2,83	9,39	-	5,89
Tyrosine	Tyr	Y	2,20	9,11	10,07	5,66
Valine	Val	V	2,32	9,62	-	5,96

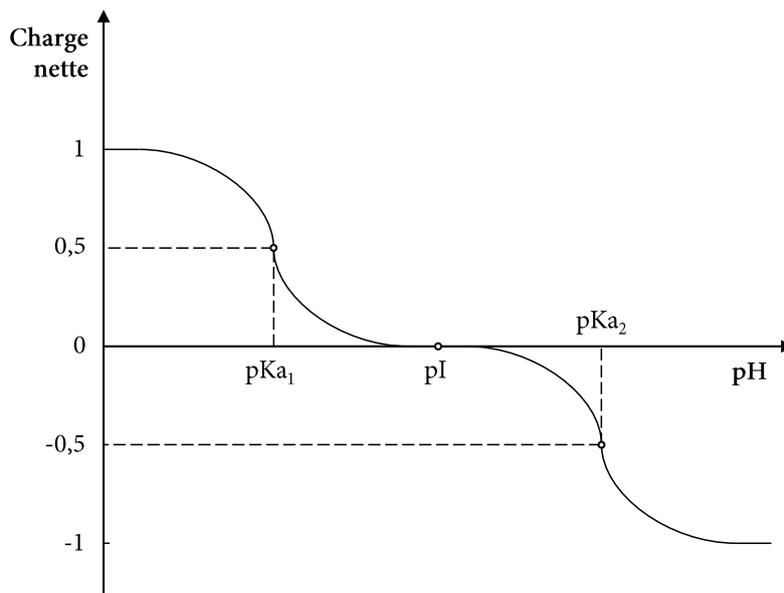


FIGURE II.4 – Illustration de la relation entre la charge nette et le pH.

pH est proche du pI peut potentiellement interagir avec un support si la distribution des charge à sa surface est asymétrique (DePhillips et Lenhoff, 2001; Whitley et al., 1989). En revanche, une protéine dans un milieu proche du pI peut présenter des problèmes de stabilité car c'est à ce pH que sa solubilité est minimale (Tanford, 1961). Il peut également arriver que certaines charges ne soient tout simplement pas accessibles de par la taille ou la conformation d'une protéine, notamment dans les cas où les charges ne se situent pas à sa surface.

Dans certains cas très particuliers, il est également possible qu'une protéine ne possède pas de point isoélectrique, mais plutôt une "région" isoélectrique. C'est notamment le cas de CalB, qui est la lipase de *Candida antarctica* B, et dont il a été montré que sa purification par échange d'ions est difficile (Trodler et al., 2008). Dans des pH allant de 5 à 8 environ (voir Figure II.5), cette enzyme présente une charge nette nulle car elle ne possède aucun résidu titrable dont le pKa est compris entre ces valeurs, à l'exception d'une seule histidine. De ce fait, un changement de pH dans cette gamme ne changera pas la charge globale de l'enzyme. Ainsi, il peut s'avérer essentiel de bien connaître la répartition de charge sur une protéine à un pH donné lorsque le mode de rétention principal repose sur les forces électrostatiques.

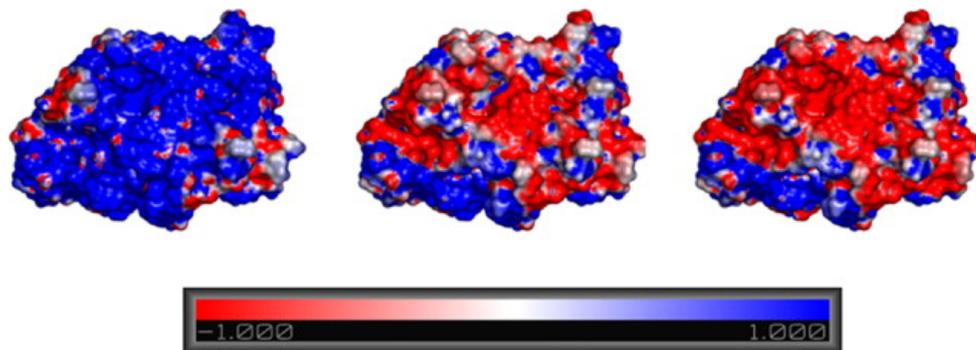


FIGURE II.5 – Surface de la CalB colorée en fonction du potentiel électrostatique aux pH 3, 7 et 10 (rouge = négatif; bleu = positif). Issu de Trodler et al. (2008), reproduit avec permission. Copyright (2008), Elsevier.

## II.2 Les procédés chromatographiques pour la séparation de protéines

### II.2.1 Les différents types de procédés chromatographiques

Les procédés chromatographiques sont de plus en plus utilisés dans l'industrie. C'est une opération unitaire de transfert de matière basée sur la différence d'affinité des solutés en solution (phase mobile) par rapport à un solide (phase stationnaire). Il en existe plusieurs types, qui peuvent être classés selon certains critères dont notamment :

- Le mode d'opération
- Le mécanisme de rétention
- Le mode de contact

#### MODE D'OPÉRATION

Une distinction peut être faite en fonction de l'agencement de la phase stationnaire. Il existe la chromatographie planaire, ou chromatographie sur papier (méthode uniquement analytique), et la chromatographie en colonne, dans laquelle la phase stationnaire sera disposée dans un tube cylindrique. Ce mode de fonctionnement peut être aussi bien utilisé en chromatographie analytique qu'en chromatographie préparative. Dans cette thèse, la chromatographie planaire ne sera pas abordé.

## MÉCANISME DE RÉTENTION

Au sein d'une colonne de chromatographie, la séparation des molécules peut se faire par plusieurs mécanismes : (i) les molécules peuvent être plus ou moins retenues en fonction de leur taille, il s'agit de la chromatographie d'exclusion de taille, ou stérique (SEC, *Steric Exclusion Chromatography*) ou (ii) les molécules peuvent interagir avec le support via différents mécanismes, impliquant l'électrostatique (chromatographie d'échange d'ions), l'affinité (chromatographie d'affinité), l'hydrophobicité (chromatographie d'interaction hydrophobe) ou l'adsorption. Il arrive fréquemment que plusieurs de ces types d'interactions soient mis en œuvre simultanément afin d'améliorer l'efficacité de la séparation : on parle alors de chromatographie multi-modale.

## MODE DE CONTACT

Différents modes de contact sont possibles entre le soluté et le solide. Les trois principaux fonctionnements sont les colonnes en *lit fixe*, en *lit fluidisé* ou encore en *lit expansé*.

Le fonctionnement en lit fixe est le plus utilisé car c'est celui qui présente la meilleure résolution, l'écoulement de type piston étant forcé à travers le lit. En revanche, cela engendre d'une part de fortes pertes de charge, qui limitent la hauteur de la colonne et la productivité, et d'autre part un risque de colmatage élevé.

Afin de limiter ces phénomènes, le fonctionnement en lit fluidisé est utilisé et préconisé dans certains cas d'application (capture à partir de milieux bruts). A l'inverse du fonctionnement en lit fixe, ce fonctionnement induit un fort mélange dans la colonne et donc une perte de résolution ; c'est pour cette raison que le lit fluidisé est rarement utilisé dans des cas de purification de protéines.

Il existe également un fonctionnement en lit expansé, qui est un compromis entre les deux autres fonctionnements. Une stratification du lit en fonction de la taille des particules permet la réduction du mélange.

## II.2.2 Fonctionnement du procédé

Le procédé chromatographique est fondamentalement un procédé discontinu. En général, il y a quatre étapes comme le montre la Figure II.6.

Il y a tout d'abord l'étape d'*équilibration*, qui est une étape de préparation de la colonne aux conditions de travail : une solution tampon, au pH et à la force ionique fixés, passe à travers le lit de résine. Cette étape intervient le plus souvent après un autre cycle de fonctionnement.

Vient ensuite l'étape de *charge*, qui vise à retenir la molécule d'intérêt sur la phase stationnaire : la colonne est alimentée par la solution contenant la (ou les) molécule(s) d'intérêt. Celle(s)-ci se fixe(nt) sur la résine selon le mécanisme de rétention propre au procédé. Cette étape s'arrête lorsque la molécule d'intérêt sort de la colonne, ce qui signifie qu'elle n'est plus retenue et que la résine commence à être saturée.

Une fois l'étape précédente stoppée, il y a l'étape de *lavage* où le volume poreux de la colonne est éliminé par passage d'une solution tampon de capacité ionique faible (le plus souvent uniquement de l'eau déminéralisée). Cela permet d'évacuer la solution mère et les impuretés présentes entre les grains de résine.

Finalement, il y a l'*élution*, qui permet de récupérer la molécule d'intérêt purifiée, préalablement fixée sur la résine lors de la charge. Pour cela, une solution tampon de force ionique de plus en plus élevée est injectée dans la colonne afin de décrocher la molécule. Le but de cette étape est d'affaiblir la liaison entre les molécules et le support en modifiant l'équilibre. Une fois la totalité des molécules récupérée, le procédé peut repartir pour un nouveau cycle de fonctionnement avec une nouvelle étape d'équilibration.

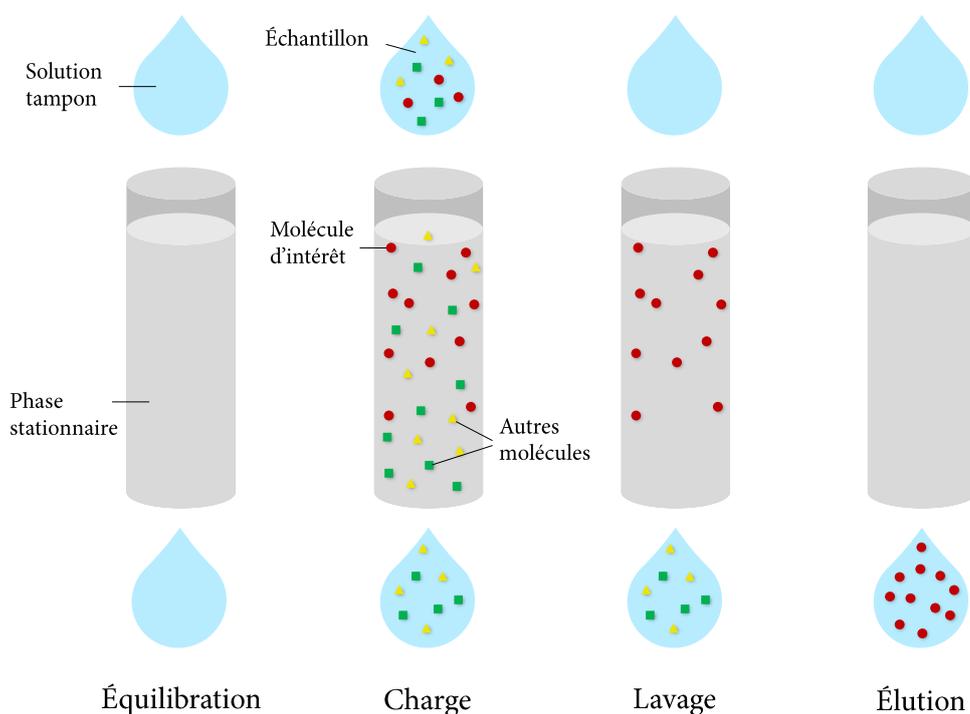


FIGURE II.6 – Les 4 étapes de fonctionnement du procédé chromatographique.

## II.3 La chromatographie d'échange d'ions

### II.3.1 Principaux mécanismes et phénomènes

La chromatographie d'échange d'ions est aujourd'hui l'une des techniques les plus utilisées pour purifier les protéines (Przybycien et al., 2004; Jungbauer, 2013; Lenhoff, 2016). Son mécanisme de rétention principal repose sur l'interaction électrostatique entre une protéine chargée et une surface de charge opposée. En effet, selon le pH, une protéine peut donc être globalement chargée positivement ( $\text{pH} < \text{pI}$ ) ou négativement ( $\text{pH} > \text{pI}$ ). La phase stationnaire en chromatographie d'échange d'ions est le plus souvent composée de petites billes de résines extrêmement poreuses – les résines échangeuses d'ions – dont la surface présente une multitude de ligands (ou sites) chargés, qui sont fixés sur la matrice de la résine et donc immobiles. Ces sites chargés vont attirer des ions de charge opposée, que l'on appelle les contre-ions (voir Figure II.7). Par convention, une résine est dite "cationique" lorsqu'elle échange des cations et "anionique" lorsqu'elle échange des anions.

Pendant le passage du mélange à purifier à travers la phase stationnaire, la protéine se fixe sur le support chromatographique en déplaçant les contre-ions initialement présents sur les sites de l'échangeur. Cette fixation est réversible et la compétition entre la protéine et les contre-ions, qui sont de même charge, est toujours présente. Dans des solutions tampons de faible force ionique, la compétition est faible et la protéine passe la plupart du temps fixée sur le support. Cependant, une protéine, même fortement attachée, est entraînée et continue de descendre le long de la colonne. Le temps que mettra une protéine à sortir de la colonne dépend du temps qu'elle passe en solution : par exemple, si une protéine descend à 10% de la vitesse du tampon (ce qui signifie qu'elle passe 90% du temps accrochée au support), il faudra 10 volumes de colonne pour faire sortir la protéine. Ainsi, les protéines possédant les plus faibles charges nettes électriques devraient sortir les premières.

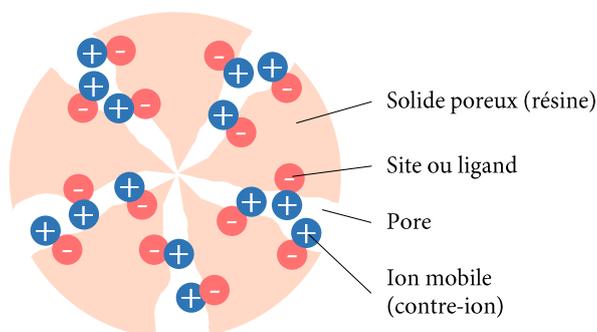


FIGURE II.7 – Représentation d'un grain de résine cationique.

Une fois la protéine fixée sur la surface de la résine et l'étape de charge finie, il existe plusieurs manières d'éluer la protéine. Le but étant de modifier l'équilibre, la première solution est d'effectuer une élution par gradient de sel. L'ajout de sel en solution augmente le nombre de contre-ions en solution et donc la compétition entre les ions et la protéine. A partir d'une certaine force ionique, la protéine ira préférentiellement en solution et sera remplacée par les contre-ions à la surface de la résine et pourra ainsi être récupérée en sortie de colonne.

Une autre solution pour décrocher la protéine de la surface chargée est de modifier le pH de la solution environnante et ainsi modifier sa charge nette. Si celle-ci décroît, son attraction avec le support diminue également. En pratique, cette solution est rarement utilisée car les changements de pH dans une colonne, et en particulier à l'intérieur des pores de la résine, sont difficiles à maîtriser. Il a été mis en évidence qu'à tout moment, le pH réel environnant les échangeurs peut présenter jusqu'à 1 unité d'écart avec le pH de la solution tampon : il s'agit là d'une conséquence de l'effet Donnan (Gregor, 1951). Lors d'un échange cationique par exemple, les sites négatifs du support solide attirent les protons et repoussent les ions hydroxydes, ce qui engendre un pH plus acide aux alentours des sites de fixation. Le phénomène inverse est observé dans le cas d'une résine anionique.

Un autre aspect important en échange d'ions est le choix des solutions tampons. Il permet le maintien du pH en solution et conditionne la charge nette et la stabilité de la protéine. On appelle un système tampon un milieu dont le pH varie peu par ajout d'acide, de base ou par dilution. Pour un couple acido-basique donné, une solution tampon correspond au mélange de l'acide et de sa base conjuguée de sorte à ce que le pH soit compris entre  $pK_a - 2$  et  $pK_a + 2$ . Le pouvoir tampon maximal est obtenu pour un mélange équimolaire et de forte concentration, lorsque le pH est voisin du  $pK_a$ . Dans le cas de l'échange d'ions, il est important de chercher à minimiser la force ionique du milieu (plus la force ionique est faible, plus l'interaction de la protéine avec le support est forte).

### II.3.2 Les résines échangeuses d'ions

Les résines échangeuses d'ions sont de petites billes poreuses composées d'un matériau solide, que l'on appelle la matrice. Elles sont définies notamment selon les groupements fonctionnels (ou ligands, ou sites) présents à la surface. D'une manière plus générale, leur choix est critique lorsqu'il s'agit de séparer des protéines par échange d'ions et l'ensemble de leurs propriétés sont à prendre en compte, *i.e.* le caractère hydrophile/hydrophobe, la porosité, la stabilité et la résistance mécanique ou encore le gonflement.

### II.3.2.1 Les matrices

Il existe plusieurs types de matrices dans le commerce, dont certaines sont particulièrement recommandées (et utilisées) dans le domaine de la séparation de biomolécules telles que les protéines. Le choix de la matrice dépendra non seulement des propriétés (comme une bonne résistance à l'usure), mais également de l'usage prévu : en effet, en agroalimentaire par exemple, certains matériaux peuvent ne pas être autorisés. Trois grandes classes de matrices, qui sont les plus répandues dans le commerce, sont présentées ci-dessous sous forme d'une liste non-exhaustive.

#### LES RÉSINES ORGANIQUES HYDROPHOBES

Elles sont généralement à base de polymères (styrène, acrylique, métacrylique) et d'un agent de réticulation, dont le rôle est de rendre la matrice plus rigide (le plus courant étant le divinylbenzène). Elles sont très utilisées dans le domaine de la purification de l'eau.

Avantages : elles sont rigides et possèdent une grande capacité ionique.

Inconvénients : leur petite taille de pores fait qu'elles sont rarement utilisées pour la purification de macromolécules.

Les résines commerciales les plus courantes sont les Source (GE Healthcare), POROS (Thermo Fisher Scientific), Toyopearl (TOSOH Bioscience) ou Fractogel (EMD millipore)

#### LES RÉSINES ORGANIQUES HYDROPHILES

Elles sont essentiellement à base d'agarose, qui est un polymère de la famille des agars et possède un fort pouvoir gélifiant, de cellulose et plus rarement à base de dextrane. Il existe plusieurs degrés de réticulation (2, 4 ou 6%), un degré élevé induisant une plus forte rigidité.

Avantages : elles possèdent des capacités élevées pour la rétention de protéines. La cellulose et l'agarose étant considérés comme des produits naturels, ces résines sont compatibles avec des usages agroalimentaire ou pharmaceutique.

Inconvénients : étant peu rigides, il faut privilégier un fort degré de réticulation afin de limiter la compressibilité de la résine en colonne. Leur caractère hydrophile peut également provoquer des gonflements difficiles à maîtriser.

Les résines commerciales à base d'agarose et de cellulose sont très nombreuses, comme par exemple les gammes des Sepharose, Capto ou Streamline (GE Healthcare) pour l'agarose et des Cellufine (JNC corporation) ou HyperCel (PALL Life Sciences) pour la cellulose.

## LES RÉSINES À BASE DE SILICE

Bien qu'il existe des résines composées de silice pure, elles sont le plus souvent greffées avec un autre matériau (polymères). Elles peuvent également être céramisées (la silice est transformée en céramique) afin de limiter l'usure des particules de résine.

Avantages : la structure de la matrice n'est pas modifiée selon la solution utilisée (pas de gonflements).

Inconvénients : elles s'usent mécaniquement lorsqu'elles sont soumises à de forts gradients de pression en colonne.

Ces résines sont généralement peu utilisées pour la purification de protéines du fait de leur manque de stabilité en milieu alcalin, mais de nouvelles résines à base de silice à usage unique sont par exemple développées pour l'industrie pharmaceutique (Gu et al., 2015).

De nos jours, de nouveaux supports chromatographiques toujours plus performants sont développés afin de palier certaines limites dues aux propriétés des matériaux. Des résines composites ont notamment été créées à base de plusieurs matériaux, mélangeant ainsi leurs propriétés, comme par exemple la résine HyperD (Fernandez et Carta, 1996; Lenhoff, 2011).

### II.3.2.2 Les groupements fonctionnels

Les groupements fonctionnels, appelés également ligands ou sites, sont irréversiblement liés à la matrice des résines échangeuses d'ions. Ils tapissent la surface des échangeurs et ce sont eux qui vont permettre l'échange des ions par interaction électrostatique. Il en existe plusieurs types, à savoir des anioniques forts ou faibles, et des cationiques forts ou faibles (voir le Tableau II.2). A l'instar des acides et des bases, un ligand fort sera toujours ionisé en solution quel que soit le pH ; à l'inverse, un ligand faible ne sera ionisé que dans une certaine gamme de pH.

### II.3.2.3 Propriétés des résines

#### CAPACITÉS D'ÉCHANGE

Les échangeurs sont caractérisés le plus souvent par leur capacité d'échange (ou capacité ionique), qui correspond à la concentration molaire de sites par unité de volume (ou de masse) de résine, et qui dépend de la densité de ligands présents à la surface. Cette capacité est aussi dite *totale* car elle correspond au nombre maximal de sites où l'échange peut se produire.

TABLEAU II.2 – Différents types de groupements fonctionnels ou ligands courants.

Type d'échangeur	Abr.	Groupement fonctionnel	
		Nom	Structure
Anionique fort	Q	Trimethyl aminomethyl	$-\text{O}-\text{CH}_2-\text{N}^+(\text{CH}_3)_3$
	TMAE	Trimethyl aminoethyl	$-\text{O}-\text{CH}_2-\text{CH}_2-\text{N}^+(\text{CH}_3)_3$
	QAE	Diethyl-(2-hydroxypropyl) aminoethyl	$-\text{O}-\text{CH}_2-\text{CH}_2-\text{N}^+(\text{CH}_2-\text{CH}_3)_2(\text{CH}_2-\text{CHOH}-\text{CH}_3)$
	QA	Trimethylamino-hydroxypropyl	$-\text{O}-\text{CH}_2-\text{CHOH}-\text{CH}_2-\text{N}^+(\text{CH}_3)_3$
Anionique faible	DMAE	Dimethyl aminoethyl	$-\text{O}-\text{CH}_2-\text{CH}_2-\text{NH}^+(\text{CH}_3)_2$
	DEAE	Diethyl aminoethyl	$-\text{O}-\text{CH}_2-\text{CH}_2-\text{NH}^+(\text{CH}_2-\text{CH}_3)_2$
Cationique faible	CM	Carboxymethyl	$-\text{O}-\text{CH}_2-\text{COO}^-$
Cationique fort	SP	Sulfopropyl	$-\text{O}-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{SO}_3^-$
	SE	Sulfoethyl	$-\text{O}-\text{CH}_2-\text{CH}_2-\text{SO}_3^-$
	S	Sulfomethyl	$-\text{O}-\text{CH}_2-\text{SO}_3^-$

En pratique, il est rare que tous les sites soient occupés même lorsque la résine est saturée, du fait de l'inaccessibilité des sites dues à la taille des pores, à la taille des protéines ou encore à la limitation par le transfert de matière. Le terme de capacité d'échange d'utile, ou apparente, est alors utilisé pour caractériser le nombre de sites où l'échange s'est réellement produit.

Dans les procédés de séparation de protéines, il est également d'usage de définir une capacité dynamique ou capacité utile (Hardin et al., 2009; Janakiraman et al., 2018). En effet, afin de limiter la perte de protéines lors de l'étape de charge, celle-ci est arrêtée lorsque la concentration en protéines en sortie de colonne atteint 5 à 10% de la concentration d'alimentation. La quantité de protéine retenue à ce moment-là, exprimée par quantité de résine (masse ou volume), est donc appelée capacité dynamique à 5 (ou 10) %. Pour la plupart des résines commerciales, cette capacité a été mesurée comme par exemple dans les études de Cabanne et al. (2004), qui ont mesuré les capacités dynamiques à 10% de deux résines composites (Streamline Q XL et Q HyperZ) à l'aide d'une solution de BSA.

## DENSITÉ DE LIGANDS

L'efficacité de la rétention de la protéine par la résine dépend très fortement de sa capacité d'échange car elle caractérise entre autres la densité de groupements fonctionnels au sein de la résine. Ainsi, une résine possédant une forte densité de ligands sera capable de créer des fixations multipoints (plusieurs charges qui vont intervenir simultanément sur la protéine) et ainsi induire une meilleure rétention de la protéine (Huang et al., 2010). A l'inverse, une densité faible fournira une liaison moins forte avec la molécule et donc une capacité de rétention réduite. Plusieurs études (Kisley et al., 2014) ont montré que la densité de ligands joue un rôle essentiel dans l'échange d'ions et la modifier est un bon moyen d'améliorer la rétention.

A partir de la densité de groupements fonctionnels, il est possible de faire une approximation de la distance moyenne entre deux ligands. Selon Wu et Walters (1992), cette distance  $S$  peut être estimée selon :

$$S = \left( \frac{A}{LD \cdot N_A} \right)^{\frac{1}{2}} \quad (\text{II.1})$$

Avec  $A$  la surface spécifique du support chromatographique ( $\text{m}^2/\text{mL}$ ),  $LD$  la densité de groupements fonctionnels ( $\text{mol}/\text{mL}$ ) et  $N_A$  la constante d'Avogadro ( $\text{mol}^{-1}$ ).

Il faut noter que la surface spécifique du support n'est généralement pas une donnée fournie par le fabricant, il faudra probablement la déterminer expérimentalement si on ne la trouve pas dans la littérature. Chang et Lenhoff (1998) ont déterminé la surface spécifique de la résine Sepharose FF comme étant égale à  $42,5 \text{ m}^2/\text{mL}$ . Huang et al. (2010) ont utilisé cette donnée dans l'équation II.1 afin de calculer la distance entre les ligands de la DEAE Sepharose FF, qui serait comprise entre  $0,67$  et  $2,04 \text{ nm}$ . Cela signifie entre autres que si le rayon de Stokes de la molécule à échanger est supérieur à cette distance, elle peut potentiellement se fixer sur plusieurs sites simultanément. Néanmoins, la détermination expérimentale de cette surface spécifique reste très délicate à mettre en œuvre, et notamment lorsqu'il s'agit d'une résine composée majoritairement d'eau comme dans le cas des matrices en agarose. Traditionnellement, les surfaces spécifiques sont mesurées à l'aide de la méthode B.E.T. (Brunauer et al., 1938), qui nécessite le séchage de la résine. Une résine type agarose, contenant en moyenne  $85\%$  d'eau, devient quasiment impossible à exploiter par cette technique du fait de l'altération de sa structure par séchage. C'est pour cette raison que DePhillips et Lenhoff (2000) ont suggéré d'utiliser la chromatographie d'exclusion de taille inverse pour tenter de mesurer cette donnée.

## TAILLE DES PORES

Le type de matrice induit un paramètre important qui est la taille des pores, élément essentiel à prendre en considération lorsque l'on travaille avec des grosses molécules comme les protéines. Dans le cas de macromolécules, il faut privilégier des tailles de pores les plus grandes possibles afin de permettre aux protéines d'occuper le plus de surface chromatographique possible et ne pas bloquer le transport. Si les pores sont trop petits, il y aura un phénomène d'exclusion de taille qui va se superposer au phénomène d'échange d'ions. Malheureusement, les fabricants de résines fournissent rarement cette information. Dans les cas où elle est donnée, la valeur indique seulement un diamètre moyen des pores. Il existe en revanche des techniques expérimentales qui permettent de déterminer la distribution de taille de pores. Dans la littérature, de nombreux auteurs (DePhillips et Lenhoff, 2000; Yao et Lenhoff, 2004, 2006; Angelo et al., 2013; Coquebert de Neuville et al., 2013) ont mesuré la distribution de taille de pores pour les résines commerciales les plus courantes. La technique utilisée est la chromatographie d'exclusion de taille inverse (ISEC, Inverse Size-Exclusion Chromatography). Le but de cette technique est de déterminer le temps de séjour dans la résine de solutés de tailles différentes (le plus souvent des particules de dextrane) et ainsi en déduire une distribution de taille de pores.

### II.3.3 L'équilibre d'échange d'ions

La modélisation des procédés chromatographiques nécessite la connaissance des phénomènes physico-chimiques mis en jeu (Mollerup, 2008; Lenhoff, 2016). Les deux aspects fondamentaux de l'échange d'ions sont l'équilibre et la cinétique du système étudié.

L'équilibre d'échange décrit la répartition du soluté entre la phase mobile et la phase stationnaire (résine) dans des conditions définies. De nombreux modèles dans la littérature permettent de décrire les équilibres d'adsorption et sont capables de prendre en compte divers phénomènes (Lan et al., 2001; Leško et al., 2015). Il existe des modèles très simplifiés permettant de décrire une adsorption, comme le modèle de Langmuir et ses diverses formes (Olsen et Watanabe, 1957; Misak, 1993; Jeppu et Clement, 2012), et des modèles plus complexes, développés pour des cas particuliers et qui possèdent un caractère prédictif (Raje et Pinto, 1997; Bosma et Wesselingh, 2004; Su et Sun, 2006; Creasy et al., 2015; Marek et al., 2018). Ces types de modèles ont connu un fort développement ces dernières années, malgré l'utilisation encore très fréquentes des modèles "simples".

#### II.3.3.1 Les isothermes d'adsorption

Les isothermes d'adsorption permettent de caractériser l'affinité d'un soluté pour un adsorbant, et dans certains cas de caractériser un matériau adsorbant (mesure de la sur-

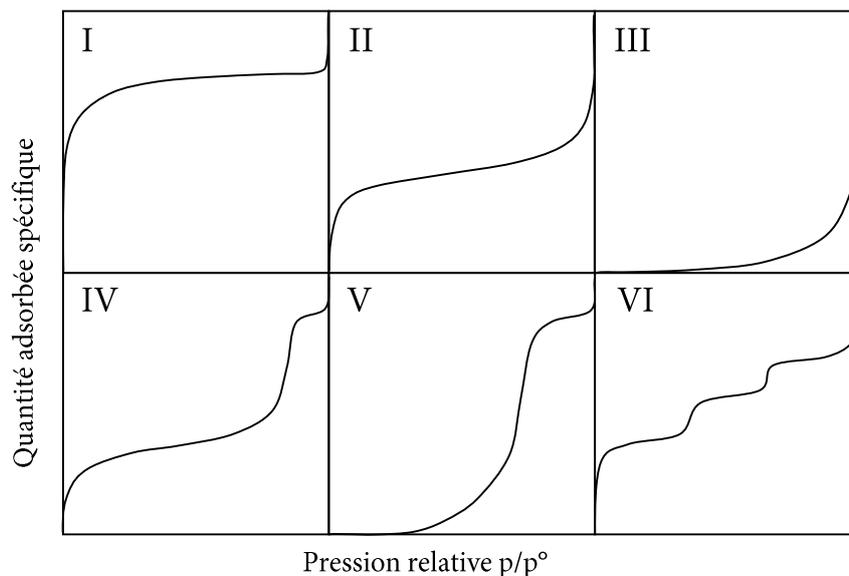


FIGURE II.8 – Classification des isothermes d'adsorption selon l'IUPAC.

face spécifique par la méthode BET par exemple (Brunauer et al., 1938)). Une isotherme représente, à température donnée, la quantité de soluté adsorbée sur le matériau poreux en fonction de la quantité de soluté restée en solution (phase liquide) ou en fonction de la pression relative (phase gazeuse). Une classification des types d'isothermes a été établie par l'IUPAC (*International Union of Pure and Applied Chemistry*) et permet de distinguer six formes caractéristiques d'isothermes d'adsorption (Figure II.8).

**Type I :** Ces isothermes montrent une augmentation rapide de la quantité adsorbée, suivie d'un palier approximativement horizontal (saturation de l'adsorbant). En phase gazeuse, cette isotherme est généralement attribuée à l'adsorption sur une surface contenant des micropores ( $< 2$  nm). Elles peuvent également traduire la formation d'une mono-couche sur la surface.

**Type II :** Ces isothermes sont les plus rencontrées et traduisent l'adsorption sur des surfaces non-poreuses. L'inflexion correspond au moment où la surface est saturée en mono-couche et où l'adsorption commence à se faire en multi-couche.

**Type III :** La forme convexe de cette isotherme traduit en général une faible affinité entre l'adsorbant et le soluté.

**Type IV :** Ces isothermes se retrouvent lorsque les adsorbants présentent des pores de petits diamètres (mésopores, entre 2 et 50 nm). En adsorption en phase gazeuse, il se produit un phénomène de condensation capillaire (représentée par une inflexion de la courbe) à l'intérieur des pores.

**Type V :** Similaire à l'isotherme de type IV, seule la première partie de l'isotherme correspond à une forme convexe de type III, qui indique une faible interaction adsorbant-soluté.

Type VI : Cette isotherme correspond à un milieu poreux dans lequel l'adsorption se fait en multi-couches successives.

Expérimentalement, il est rare que les isothermes obtenues correspondent rigoureusement à l'un de ces types d'isothermes. En revanche, elles permettent d'interpréter les différentes portions de l'isotherme, comme par exemple la courbure caractéristique de l'apparition d'une multi-couche. Bien que ces modèles d'isothermes soient développés principalement pour l'étude de l'adsorption en phase gazeuse, ils peuvent aussi être utilisés pour décrire l'adsorption de protéines (Carta et Jungbauer, 2010). Cependant, lors de la mesure d'une isotherme en phase liquide, seule la concentration en protéine restante en solution est mesurée, la quantité adsorbée est déduite par bilan de matière. Ainsi, une forte augmentation de cette valeur calculée peut donc traduire plusieurs phénomènes, comme l'adsorption en multi-couches, mais également une agrégation ou une précipitation de la protéine.

### II.3.3.2 L'isotherme de Langmuir

L'équilibre d'échange d'ions est encore aujourd'hui très souvent décrit par des modèles traduisant le phénomène d'adsorption, mais qui ne représentent pas les forces électrostatiques qui sont à la base du mécanisme d'échange d'ions.

Parmi les modèles les plus utilisés en adsorption, ceux basés sur l'isotherme de Langmuir sont les plus simples et permettent de décrire les phénomènes d'adsorption mono-constituant. Le modèle de Langmuir a été développé selon les hypothèses suivantes :

- Surface du support uniforme (un seul type de sites, identiques et indépendants)
- Une seule molécule adsorbée par site (monocouche)
- Pas d'interaction entre les molécules adsorbées
- Une seule espèce adsorbable (pas de compétition)
- Nombre de sites fini
- Adsorption réversible

L'équation de Langmuir (II.2) s'écrit :

$$q = \frac{Q \cdot K_L \cdot C}{1 + K_L \cdot C} \quad (\text{II.2})$$

Où  $q$  est la quantité de protéine adsorbée par la résine (g/L),  $C$  est la concentration de protéine en solution (g/L),  $Q$  est la capacité maximale d'adsorption de la résine (g/L) et  $K_L$  est la constante d'équilibre (L/g). Ce modèle est très utilisé car il représente mathématiquement la forme concave d'une isotherme, que l'on appelle isotherme favorable (isotherme de type I selon l'IUPAC, Figure II.8).

### II.3.3.3 La loi d'action de masse

Afin de décrire le mécanisme d'échange d'ions, des modèles plus spécifiques ont été développés, le plus simple et le plus largement répandu étant la loi d'action de masse. L'hypothèse principale de cette loi considère que l'échange d'ions est une réaction stoechiométrique (voir équation II.3), de manière à préserver l'électro-neutralité de chaque phase.



où  $R$  représente la résine,  $I$  le contre-ion monovalent,  $P^n$  la protéine de charge caractéristique  $n$ .

La loi d'action de masse permet d'exprimer la constante d'équilibre  $K_{P^n/I}$  selon :

$$K_{P^n/I} = \frac{(\bar{P}^n) \cdot (I)^n}{(\bar{I})^n \cdot (P^n)} = \frac{[\bar{P}^n] \cdot [I]^n \gamma_{\bar{P}^n} \cdot \gamma_I^n \bar{C}^{\circ n-1}}{[\bar{I}]^n \cdot [P^n] \bar{\gamma}_I^n \cdot \gamma_{P^n} C^{\circ n-1}} \quad (\text{II.4})$$

où  $[\bar{I}]$  représente la concentration du contre-ion dans la résine et  $[I]$  dans la solution ;  $[\bar{P}^n]$  représente la concentration de la protéine dans la résine et  $[P^n]$  dans la solution ;  $\gamma_I$  le coefficient d'activité de l'ion I en solution et  $\gamma_{P^n}$  le coefficient d'activité de la protéine en solution, et  $C^\circ$  les concentrations de l'état de référence. Pour une réaction donnée, la constante d'équilibre thermodynamique peut également être définie selon :

$$\Delta_r G^\circ = -RT \ln K \quad (\text{II.5})$$

avec  $\Delta_r G^\circ$  l'enthalpie libre standard de réaction,  $R$  la constante des gaz parfaits et  $T$  la température absolue.

Dans le cas de l'échange d'ions, cette constante d'équilibre traduit l'affinité des ions en solution pour une résine. Généralement, les coefficients d'activité sont supposés égaux à 1 car il est difficile d'accéder à ces valeurs au sein d'une résine échangeuse d'ions. Ainsi, le terme  $\frac{\gamma_{\bar{P}^n} \cdot \gamma_I^n \bar{C}^{\circ n-1}}{\bar{\gamma}_I^n \cdot \gamma_{P^n} C^{\circ n-1}}$  est considéré égal à 1 et la constante d'équilibre  $K$  apparente s'écrit :

$$K = \frac{[\bar{P}^n] \cdot [I]^n}{[\bar{I}]^n \cdot [P^n]} \quad (\text{II.6})$$

ou encore en utilisant  $C$  pour les concentrations en solution et  $q$  pour les concentrations dans la résine :

$$K = \frac{q \cdot C_I^n}{q_I^n \cdot C} = \frac{q/C}{(q_I/C_I)^n} \quad (\text{II.7})$$

D'autre part, la quantité maximale de sites  $q_R$ , qui représente la capacité maximale d'échange, peut s'écrire (toujours dans le cas où I est un ion monovalent) :

$$q_R = nq + q_I \quad (\text{II.8})$$

Ainsi, la combinaison des équations II.7 et II.8 donne :

$$C = \frac{q \cdot C_I^n}{K \cdot (q_R - n \cdot q)^n} \quad (\text{II.9})$$

Ce modèle a été construit selon les hypothèses suivantes :

- Particules adsorbantes sphériques et uniformes (taille, densité)
- Groupes fonctionnels également distribués sur la surface de la particule
- Paramètres du modèle constants et indépendants de la concentration en sel
- Facteurs influençant la structure 3D de la protéine et l'effet de la concentration en sel négligés
- Nature multipoint de l'interaction protéine-support décrite par les charges caractéristiques

#### LOI D'ACTION DE MASSE STÉRIQUE

La principale limitation de la loi d'action de masse est son incapacité à prendre en compte le blocage de certains sites de l'échangeur dû à la grande taille des protéines. Une extension de cette loi a donc été mise au point par Brooks et Cramer (1992) : il s'agit du **modèle d'action de masse stérique** (SMA, Steric Mass Action).

Ce modèle prend en compte un paramètre supplémentaire qui est le facteur stérique (noté  $\sigma$ ). Il traduit l'encombrement stérique d'une partie des sites de l'échangeur par une macromolécule (voir Figure II.9). Ainsi, la quantité de sites inaccessibles  $\hat{q}_I$  exprimée par unité de volume d'échangeur s'écrit :

$$\hat{q}_I = \sigma \cdot q \quad (\text{II.10})$$

A l'équilibre, le nombre total de contre-ions à la surface de l'échangeur est égal à la somme des contre-ions échangeables et des contre-ions non-échangeables :

$$q_{I,total} = \hat{q}_I + q_I \quad (\text{II.11})$$

La quantité totale de sites de l'échangeur  $q_R$  peut donc s'écrire :

$$q_R = q_I + (n + \sigma)q \quad (\text{II.12})$$

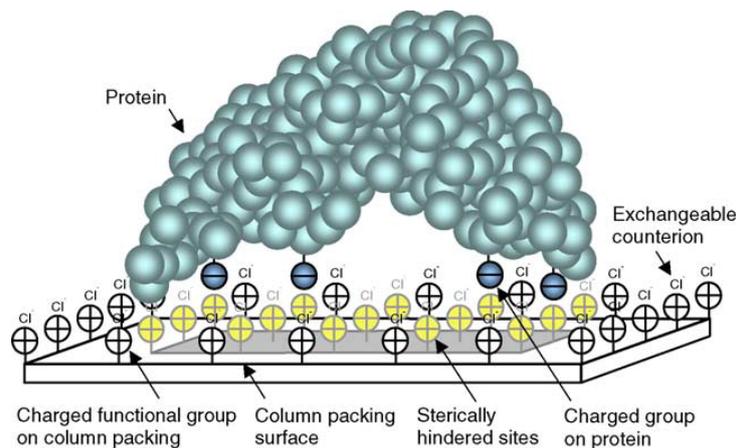


FIGURE II.9 – Illustration du phénomène d’encombrement stérique, avec une protéine possédant 4 points de fixation sur un support chromatographique. Issu de (Shen et Frey, 2005), reproduit avec permission. Copyright (2005), Elsevier.

Ainsi, la loi d’action de masse stérique s’écrit sous la forme :

$$C = \left( \frac{q}{K} \right) \left( \frac{C_I}{q_R - (n + \sigma)q} \right)^n \quad (\text{II.13})$$

où  $K$  est la constante d’équilibre,  $n$  la charge caractéristique de la protéine et  $\sigma$  le facteur stérique.

Brooks et Cramer (1992) ont également exprimé l’équation SMA pour une application en multi-constituants. Elle permet d’exprimer la concentration en solution de la protéine  $i$  en fonction de sa concentration dans la résine selon :

$$C_i = \left( \frac{q_i}{K_i} \right) \left( \frac{C_I^n}{q_R - \sum_{i=1}^{N_c} (n_i + \sigma_i)q_i} \right)^n \quad (\text{II.14})$$

Néanmoins, l’utilisation de ce modèle sous-entend que les protéines n’ont pas d’interactions entre elles, et ne permet pas de représenter des phénomènes telles que la formation d’une multi-couche par exemple.

## II.4 La simulation moléculaire appliquée à l’échange d’ions des protéines

Le comportement des protéines à l’intérieur d’une particule de résine échangeuse d’ions reste encore aujourd’hui assez mal connu (Lenhoff, 2016). En effet, de nombreux modèles

utilisés sont empiriques car basés sur des données expérimentales, et assez peu de modèles prédictifs ont été développés (Mattrey et al., 2017). Cela est dû en grande partie au manque d'information sur les mécanismes physico-chimiques mis en jeu à l'échelle atomique. Les techniques expérimentales à l'échelle microscopique restent aujourd'hui assez peu utilisées en chromatographie d'échange d'ions car elles sont difficilement adaptables, mais aussi coûteuses et intrusives. En effet, des techniques comme la microscopie de force atomique (ou AFM, Atomic Force Microscopy) ont déjà été utilisées (Johnson et al., 2000; Demanèche et al., 2009) mais les mesures intraparticulaires restent compliquées. Il en va de même avec la microscopie confocale à balayage laser (ou CLSM, Confocal Laser Scanning Microscopy), où cette fois les mesures étaient limitées par l'atténuation de la lumière au sein des adsorbants (Yang et al., 2006). Dans ce contexte, un des moyens d'acquérir des connaissances sur les mécanismes de rétention au niveau atomistique est d'utiliser des simulations de dynamique moléculaire dans le but d'étudier les interactions protéine-surface (Latour, 2008; Zhang et Sun, 2010).

La simulation numérique est en général utilisée en complément d'expériences afin d'interpréter et de mieux comprendre des données expérimentales. Les simulations moléculaires permettent entre autres d'étudier les effets de certains paramètres sur un mécanisme ou un phénomène, de donner des informations sur des données non accessibles expérimentalement en raison d'échelles de temps ou d'espace non réalisables. La force de la simulation moléculaire réside également dans sa capacité de prédiction des phénomènes physico-chimiques. En effet, il est possible de réaliser des simulations *in silico* dans des conditions particulières et ainsi de s'affranchir des expériences, soit parce qu'elles sont trop coûteuses en matière ou en équipements, soit parce que les conditions de travail étudiées sont extrêmes et non réalisables en laboratoire. Cependant, il est nécessaire de s'assurer au préalable de la robustesse et de la fiabilité des simulations. Pour cela, la méthode doit être confrontée à des données expérimentales représentatives du phénomène étudié avant d'être utilisée à but prédictif.

## II.4.1 Modélisation moléculaire

### II.4.1.1 Construction des modèles 3D

#### REPRÉSENTATION 3D DES PROTÉINES

Les protéines sont des macromolécules biologiques complexes dont la structure 3D doit être connue afin de pouvoir correctement simuler leur comportement. La structure 3D d'une protéine donne énormément d'informations sur sa fonction biochimique comme par exemple la forme de la surface, l'association de sous-unités, l'agencement du site actif ou

encore les propriétés électrostatiques. Il existe de nombreuses techniques expérimentales permettant de déterminer avec exactitude cette structure.

L'une des techniques les plus employées aujourd'hui est la cristallographie aux rayons X (*X-ray*), basée sur la diffraction des rayons X des cristaux de macromolécules (Drenth, 2007). A travers l'analyse de la densité électronique de la molécule, cette technique permet ainsi de connaître l'arrangement spatial des atomes de la molécule composant le cristal et d'en obtenir les coordonnées cartésiennes. L'étape limitante de cette méthode est en général l'étape de cristallisation, car les conditions physico-chimiques appropriées doivent être identifiées mais également car le cristal peut contenir une trop grande quantité de solvant aqueux (Smyth, 2000). En revanche, lorsque la qualité des cristaux produits est suffisante, cette technique permet de reconstituer la structure d'une macromolécule avec de hautes résolutions ( $< 3.5 \text{ \AA}$ , DiMaio et Chiu (2016)).

D'autres techniques peuvent être employées, comme par exemple la résonance magnétique nucléaire (RMN), qui utilise les propriétés magnétiques des noyaux des atomes afin d'en déduire des informations telles que les distances et, de manière indirecte, les angles dans la molécule (Wüthrich, 1990). Souvent appliquée à l'étude de petites protéines ( $< 30\text{-}40 \text{ kDa}$ ), cette technique est parfois couplée à la cristallographie aux rayons X afin d'obtenir des structures les plus complètes possible (Snyder et al., 2005).

La cryo-microscopie électronique (cryo-EM) est également une technique de plus en plus utilisée pour déterminer des structures de protéines avec une résolution proche de la résolution atomique (Zhou, 2011; Rossmann et al., 2011; DiMaio et Chiu, 2016). Cette technique permet de congeler très rapidement des biomolécules et utilise des électrons pour les "imager" et ainsi révéler leurs structures moléculaires. Aujourd'hui encore, la technique de cryo-EM est complémentaire aux techniques de cristallographie aux rayons X car elle permet l'étude de molécules complexes et/ou de grande taille, alors que la cristallographie aux rayons X est plus adaptée à des échantillons petits et ordonnés.

Enfin, il arrive que la structure d'une protéine ne puisse pas, ou du moins pas totalement, être déterminée via des techniques expérimentales : dans ces cas-là, la modélisation moléculaire peut être utilisée. Afin de reconstruire des structures ou d'effectuer des prédictions structurales, des calculs sont réalisés à l'aide de différentes techniques, comme par exemple des calculs *ab initio* (prédiction de structures 3D à partir de principes physiques), des méthodes d'enfilage ou *threading* (mise en correspondance d'une séquence et d'un repliement), ou encore des modélisations comparatives ou par homologie (déduction de la séquence à partir de protéines similaires partageant au moins 30 % d'identité, ou de protéines homologues déjà connues).

Depuis le début des années 70, les coordonnées des structures de nombreuses protéines ont été répertoriées dans des banques de données structurales, dont la plus complète et la plus utilisée est la *Protein Data Bank* (PDB). Les coordonnées de chaque atome, ainsi

que diverses informations sur la structure secondaire (hélices, feuilletts) et les connectivités entre atomes (pont disulfures), sont stockées dans un fichier texte (.pdb, voir Annexe A.1). Certaines structures de protéines ont été résolues plusieurs fois, dans des conformations, des conditions de cristallisation ou des degrés de résolution différents, ou même en complexe avec un ou plusieurs ligands. Ainsi, à chacune de ces structures résolues est associé un code PDB.

Le mécanisme d'échange d'ions étant régi par les interactions électrostatiques, l'un des aspects les plus importants lors de la préparation du système de simulation est l'assignation des charges sur les différents atomes de la protéine. Au cours d'une simulation de dynamique moléculaire classique, le pH de la solution n'est pas explicitement représenté car il n'y a pas de protons dissociables. Des méthodes permettant des simulations à pH constants sont toutefois développées (Goh et al., 2014), mais elles nécessitent des puissances de calculs élevées et restent pour le moment en développement. Usuellement, lors de simulations de systèmes protéiques, la meilleure approximation est d'attribuer un état de protonation dominant à un pH donné pour chacun des résidus titrables.

Ainsi, les états de protonation des résidus sont préalablement calculés. Pour cela, il existe de nombreux serveurs en ligne, comme par exemple H++ (Gordon et al., 2005; Myers et al., 2006; Anandakrishnan et al., 2012), DelPhiPKa (Wang et al., 2015b; Pahari et al., 2018) ou PROPKA (Olsson et al., 2011; Søndergaard et al., 2011), qui permettent le calcul des pKa des acides-aminés en fonction du pH d'intérêt et de la structure 3D des protéines. Ces serveurs utilisent les calculs des pKa et les coordonnées cartésiennes des atomes pour assigner les hydrogènes nécessaires ainsi que la charge des atomes dans de nouveaux fichiers de coordonnées. La plupart de ces serveurs en ligne, dont H++ et DelPhi, résolvent l'équation de Poisson-Boltzmann, dérivée de l'équation de Poisson, qui décrit l'effet du solvant sur le champ électrostatique autour de la protéine (Bashford et Karplus, 1990; Wang et al., 2015a). Il existe également d'autres méthodes pour le calcul des pKa, basées sur des modèles empiriques comme par exemple le serveur PROPKA (Li et al., 2005; Bas et al., 2008; Olsson et al., 2011).

### II.4.1.2 Modèles de représentation

Les systèmes, composés d'un ensemble de particules, peuvent être décrits selon différents niveaux de précision. Il existe par exemple des modèles de représentation à l'échelle quantique et d'autres modèles plus ou moins simplifiés, choisis notamment en fonction de la précision requise, du type de système étudié et des puissances de calcul disponibles. Les principaux modèles de représentation sont présentés en Figure II.10.

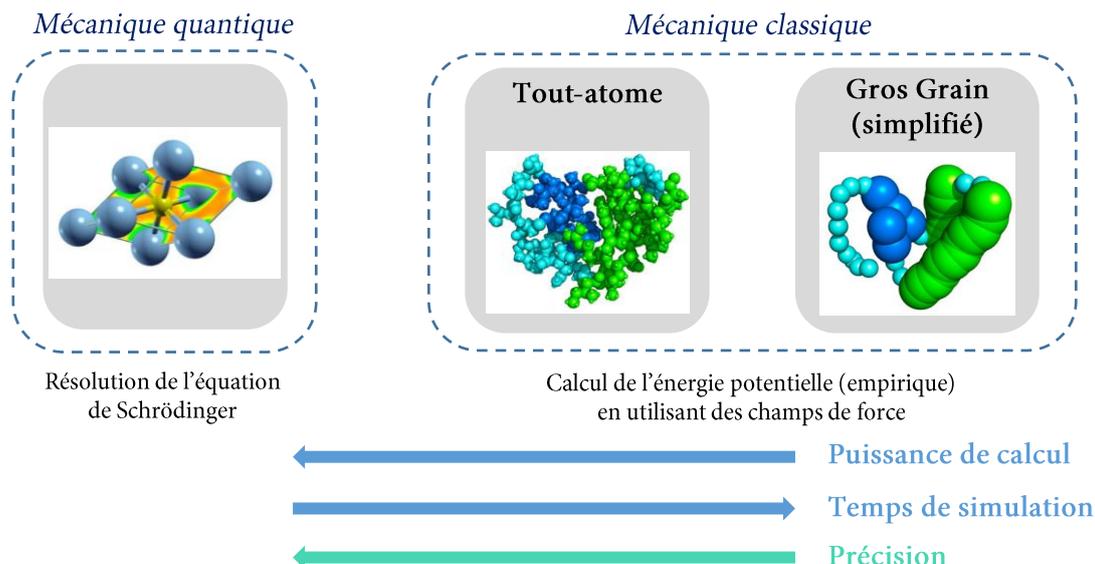


FIGURE II.10 – Les principaux modèles de représentation d'un système de simulation moléculaire.

### LES MODÈLES *ab initio*

Les méthodes quantiques considèrent explicitement les électrons du système. La description des interactions à un niveau *ab initio* est basée sur la résolution de l'équation de Schrödinger et le formalisme des orbitales moléculaires. De ce fait, ces calculs sont très précis, ne nécessitent aucun ajustement de paramètres car basés sur des principes physiques et permettent notamment de décrire des réactions chimiques. En revanche, ils sont aussi extrêmement coûteux en temps de calcul. En effet, la résolution numérique de l'équation de Schrödinger est compliquée à mettre en œuvre. Pour cela, il existe différentes méthodes permettant sa résolution, dont trois sont généralement les plus utilisées : (i) les méthodes de Monte-Carlo quantique, basées sur des algorithmes stochastiques (Ceperley et Alder, 1986; Hammond et al., 1994; Ceperley, 2010), (ii) les méthodes traditionnelles de chimie quantique, comme par exemple les méthodes Hartree-Fock (Møller et Plesset, 1934) et post-Hartree-Fock ou Moller-Plesset (MP2)(Szabo et Ostlund, 2012), et (iii) la théorie de la fonctionnelle de la densité (DFT), proposée par Hohenberg et Kohn (1964).

### LES MODÈLES TOUT-ATOME

Les modèles en **tout-atome** considèrent chaque atome du système comme étant les particules élémentaires. Ils utilisent des champs de force (Ponder et Case, 2003), qui sont des équations empiriques décrivant les énergies entre tous les atomes sur la base de leur configuration (voir paragraphe II.4.1.3). Un atome est représenté comme étant une

sphère possédant un rayon, une masse et une charge, et la molécule comme étant une association de plusieurs atomes. Contrairement aux méthodes quantiques, la variation de la structure électronique, comme dans un cas de réaction chimique (rupture de liaisons, transfert électronique, etc.), ne peut pas être aisément décrite par cette méthode, entre autres car les simulations en tout-atome se déroulent principalement à nombre d'atomes constants. Le plus souvent, ces modèles sont utilisés pour décrire les interactions protéine-ligand, entre protéines ou entre une protéine et une surface (MacKerell et al., 1998).

## LES MODÈLES GROS GRAIN

Les modèles dits **Gros Grain** sont des méthodes empiriques simplifiées et ont été développés afin de pouvoir simuler des systèmes de taille conséquente, en général lorsque les simulations en tout-atome nécessitent des temps de calcul trop importants (Chen et al., 2006; Takada, 2012; Kmiecik et al., 2016). L'intérêt majeur est également de pouvoir étudier des phénomènes biologiques qui se produisent sur des échelles de temps trop longues (de l'ordre de la microseconde ou même de la milliseconde) pour être étudiés avec des modèles atomistiques. Le but de cette méthode est de réduire le nombre de particules élémentaires dans le système, qui ne sont plus des atomes mais des groupements d'atomes. Cependant, des approximations sont faites et ces calculs sont forcément moins précis que les modèles tout-atome, notamment à cause de la perte de flexibilité de la protéine due aux regroupements d'atomes. Dans le système, chaque groupe d'atomes est considéré comme une sphère possédant un rayon, une masse et une charge. De nombreux champs de force spécifiques aux études en gros grains ont été développés (Kmiecik et al., 2016; Monticelli et al., 2008), comme par exemple le champ de force de MARTINI (Marrink et al., 2007), PRIMO (Kar et al., 2013) ou encore UNRES (Liwo et al., 2002, 2007). Comme le montre la Figure II.11, dans le champ de force de MARTINI, chaque acide aminé est représenté par un groupement d'atomes et le nombre de particules est bien plus restreint. L'une des difficultés à utiliser ce type de modèle est de bien définir les groupements d'atomes avec les paramètres correspondants, et d'autant plus si le système étudié n'est pas déjà connu et paramétrisé.

### II.4.1.3 La mécanique moléculaire

La modélisation des interactions au sein d'un système peut se faire en utilisant des approches de **mécanique classique**, basées sur des descriptions physiques des biomolécules. Ces méthodes de calculs utilisent les lois de la mécanique newtonienne et permettent d'approximer l'énergie potentielle d'interaction entre deux atomes, et diffèrent de la mécanique quantique car les structures électroniques ne sont pas prises en comptes. Ces méthodes sont empiriques et les calculs des forces se basent sur des paramètres de

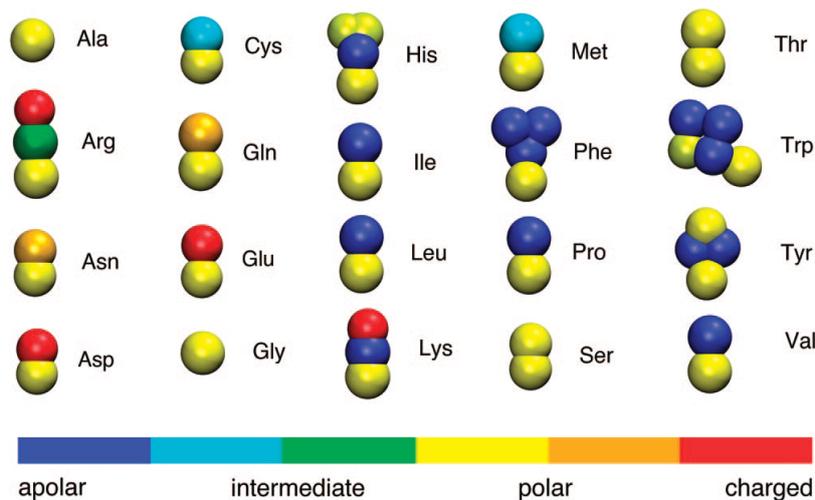


FIGURE II.11 – Représentation en Gros Grain de tous les acides aminés selon le champ de force MARTINI. Les différentes couleurs représentent les différents types de particules. Issu de Monticelli et al. (2008), reproduit avec permission. Copyright (2008), American Chemical Society.

référence issues d'expériences et/ou de calculs quantiques. L'ensemble de ces paramètres, ou constantes de forces, constituent ce que l'on appelle un champ de force (*force-field*).

Un champ de force permet de décrire la structure de l'énergie potentielle de chacune des particules élémentaires du système. De manière générale, les champs de force sont choisis en fonction du niveau de description requis (tout-atome, Gros Grain) ainsi que du domaine d'application. Un grand nombre de champs de force ont été développés pour diverses utilisations (Monticelli et Tieleman, 2012; Lopes et al., 2014). Les plus courants dans l'étude des protéines, parmi les modèles en tout-atome, sont les champs de force AMBER, CHARMM, OPLS et GROMOS (Ponder et Case, 2003).

L'énergie totale d'une protéine est définie comme étant la somme des énergies liées (ou covalentes) et non-liées (ou non-covalentes). Les énergies covalentes comprennent les énergies de liaison dues à l'étirement des liaisons ( $E_{liaison}$ ), au fléchissement ou à la courbure des angles ( $E_{angle}$ ) et à la torsion ( $E_{dièdre}$ ). Les énergies non covalentes décrivent les interactions de type van der Waals ou électrostatique (coulombique). Ainsi, l'énergie totale peut s'écrire :

$$E_{tot} = E_{liaison} + E_{angle} + E_{dièdre} + E_{vdw} + E_{elec} \quad (\text{II.15})$$

avec l'énergie de déformation de liaison :

$$E_{liaison} = \sum \frac{1}{2} k_b (b - b_0)^2 \quad (\text{II.16})$$

l'énergie de déformation d'angles :

$$E_{angle} = \sum \frac{1}{2} k_0 (\theta - \theta_0)^2 \quad (\text{II.17})$$

l'énergie de torsion d'angles :

$$E_{dièdre} = \sum \frac{1}{2} k_\varphi (1 + \cos(n\varphi - \varphi_0)) \quad (\text{II.18})$$

l'interaction de van der Waals (potentiel de Lennard-Jones) :

$$E_{vdw} = \sum_{i=1}^N \sum_{j=i+1}^{N-1} 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (\text{II.19})$$

et l'énergie potentielle électrostatique (loi de Coulomb) :

$$E_{elec} = \sum_{i=1}^N \sum_{j=i+1}^{N-1} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (\text{II.20})$$

Dans ces équations, les paramètres  $b$ ,  $\theta$ ,  $\varphi$  et  $r$  décrivent respectivement les longueurs, les angles, les angles dièdres et les distances entre atomes. Tous les autres paramètres, *i.e.*  $(k_b; b_0)$ ,  $(k_0; \theta_0)$  et  $(k_\varphi; \varphi_0)$ , sont fixés par le champ de force choisi et dépendent du système étudié.

#### II.4.1.4 Échantillonnage de l'espace conformationnel

Les résultats de simulations moléculaires dépendent de plusieurs facteurs : du type de modèle de représentation employé, des champs de force permettant de décrire les interactions mais également de l'échantillonnage statistique des configurations du système. Ainsi, l'échantillonnage de l'espace conformationnel d'un système peut être réalisé à l'aide de différentes techniques dont la méthode de Monte-Carlo et la dynamique moléculaire.

#### LA MÉTHODE DE MONTE-CARLO

La méthode de Monte-Carlo permet une exploration stochastique des configurations du système. L'échantillonnage se fait de manière aléatoire dans le but de déterminer un ensemble de conformations représentatives du système. Les paramètres calculés sont donc une moyenne d'ensemble et non temporelle et seules des propriétés thermodynamiques statiques peuvent être calculées. Le principe des méthodes de Monte-Carlo se base sur

une génération de nombres aléatoires et des probabilités statistiques pour permettre de résoudre des problèmes complexes tels que des équations. Chaque configuration du système est générée à partir de la précédente par des mouvements aléatoires et arbitraires, acceptés ou non selon un critère préalablement défini : c'est l'algorithme de Metropolis (Metropolis et al., 1953; Pharr et al., 2017).

## LA DYNAMIQUE MOLÉCULAIRE

La dynamique moléculaire permet une exploration des configurations d'un système en simulant son évolution au cours du temps. Les trajectoires de toutes les particules d'un système sont calculées à partir d'une configuration initiale, par résolution des équations du mouvement de Newton. Ainsi, la dynamique moléculaire permet de résoudre l'équation classique du mouvement pour un système de  $N$  particules soumises à un champ de force, chaque particule possédant une masse  $m_i$  et une vitesse de déplacement  $v_i$ . L'équation simplifiée de Newton peut s'écrire :

$$\vec{f}_i = m_i \frac{d\vec{v}_i}{dt} \quad (\text{II.21})$$

Lors d'une simulation de dynamique moléculaire, chaque particule élémentaire, dont les coordonnées cartésiennes sont fixées, se voit attribuer une vitesse initiale, générée aléatoirement dans la plupart des cas. Les nouvelles positions des particules à un temps  $t + \delta t$  sont calculées par un algorithme utilisant les calculs de force (énergie potentielle) et de vitesse (énergie cinétique), ce qui génère un nouveau jeu de coordonnées. Le processus est ensuite réitéré jusqu'à atteindre un temps de simulation total défini, permettant ainsi de générer les trajectoires de chaque particule du système. Plusieurs algorithmes sont utilisés pour intégrer numériquement l'équation du mouvement, dont les plus courants sont les algorithmes de Verlet (Verlet, 1967), Leap-frog (Hockney et Eastwood, 1981) ou encore Beeman (Beeman, 1976).

À l'origine, les simulations en dynamique moléculaire se déroulaient dans l'ensemble thermodynamique microcanonique, c'est-à-dire un ensemble NVE (nombre de particules, volume et énergie totale du système constants). Cependant, ces conditions ne sont pas forcément représentatives des systèmes que l'on souhaite modéliser. En laboratoire, les systèmes sont plutôt soumis à des températures constantes et non des énergies constantes, et il peut être plus pertinent de travailler à pression constante plutôt qu'à volume constant. L'ensemble NPT (nombre de particules, pression et température constants) devient l'ensemble le plus communément utilisé pour les dynamiques moléculaires de systèmes protéiques, ainsi que les ensembles NVT (nombre de particules, volume et température constants) qui sont aussi couramment utilisés. Plusieurs méthodes de contrôle de la tem-

pérature et de la pression ont été développées, dont les plus courantes sont les thermostats et barostats de Berendsen (Berendsen et al., 1984), d'Andersen (Andersen, 1980), Langevin (Feller et al., 1995) ou encore de Nosé-Hoover (Martyna et al., 1992; Nosé, 1984; Hoover, 1985).

#### *Effets de bords et conditions périodiques*

L'étude d'un système de simulation est conduite en créant une boîte de simulation d'un certain volume et contenant un nombre fini de particules. Cette boîte primaire peut être de forme cubique, mais dans des cas particuliers, cela peut être plus pertinent de travailler avec des formes différentes comme un octaèdre tronqué par exemple (Wassenaar et Mark, 2005). Le cube est la forme la plus simple à mettre en œuvre mais son utilisation peut s'avérer plus coûteuse en raison de la présence de particules (essentiellement le solvant) dans les coins de la boîte. Le choix de la boîte primaire dépend donc essentiellement de la taille et de la forme du système étudié, mais également de la puissance de calcul disponible. Cependant, le fait de travailler dans une boîte aux dimensions restreintes engendre des effets de bords importants car de nombreuses particules vont se retrouver en surface de la boîte de simulation.

Pour palier ces effets de surface, des conditions périodiques (*periodic boundary conditions*, PBC) sont appliquées. Le but est de répliquer la boîte primaire par translation dans les trois directions de l'espace, comme l'illustre la Figure II.12. Ainsi, un atome de la boîte primaire peut interagir avec tous les atomes de sa boîte, mais également avec toutes les images de ces atomes. Idéalement, le système doit être construit de sorte à ce qu'un atome ne puisse pas interagir avec sa propre image. Cette méthode de réplication à l'infini permet donc de s'affranchir des effets de taille finie du système. En revanche, cela n'a physiquement pas de sens de répliquer cette boîte à l'infini et cela ne peut pas être représenté expérimentalement. Le but ici est donc de bien définir le système et la boîte de simulation afin de s'affranchir des effets de bords et pouvoir calculer des interactions à longues portées sans pour autant influencer les résultats de simulation. Afin d'éviter l'apparition d'artefacts de simulations, des contraintes sont requises. Il est par exemple fortement recommandé de laisser une distance minimale de 1 nm (qui est en général le seuil limite pour le calcul des interactions non liées) entre la molécule étudiée et les bords de la boîte (de Souza et Ornstein, 1997).

#### *Interactions à longues portées*

Les interactions non liées correspondent aux interactions de type van der Waals et électrostatique. Ces dernières peuvent être composées de deux termes, (i) les interactions de courtes portées, et (ii) les interactions à longues portées.

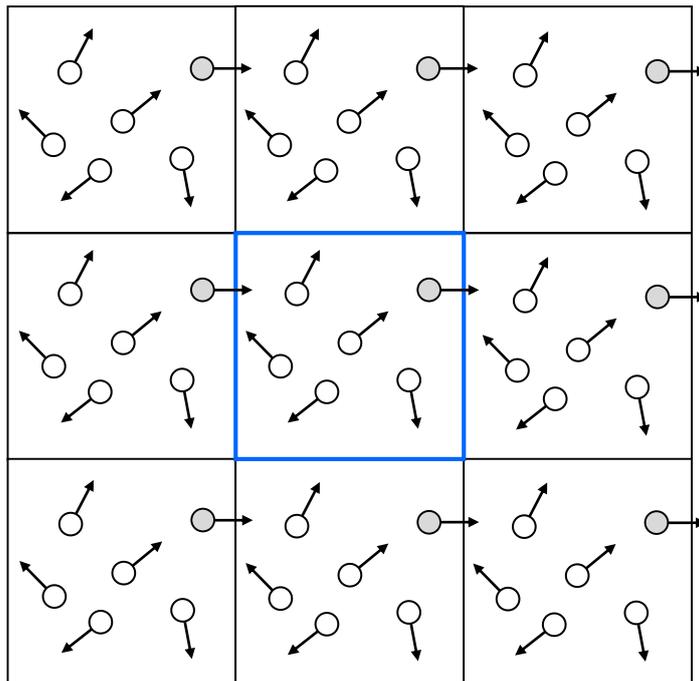


FIGURE II.12 – Représentation en deux dimensions des conditions périodiques aux limites, avec la boîte primaire de simulation encadrée en bleu.

Les interactions de courtes portées sont définies, pour les deux types van der Waals et électrostatique, par un rayon de coupure (ou *cut-off*) afin de limiter le nombre d'interactions atome-atome lors des calculs. Au delà de ce rayon de coupure, appelé également distance de troncation, différentes méthodes sont appliquées pour estimer les énergies de ces interactions à longues portées. Tout d'abord, afin d'éviter de potentiels artefacts dans les simulations dus à la coupure brutale des forces à partir du *cut-off*, certaines fonctions sont appliquées afin de ramener cette force à une valeur de zéro de manière plus douce : ce sont le plus souvent des fonctions *switch* ou *shift* (Loncharich et Brooks, 1989; Steinbach et Brooks, 1994). Ces fonctions sont en général utilisées pour traiter les interactions de type van der Waals au-delà du *cut-off*, car les forces associées diminuent rapidement en fonction de la distance.

Cependant, il a été montré que dans certains cas, notamment les systèmes membranaires, le traitement des effets à longues portées pour les interactions de type van der Waals serait bénéfique pour simuler certaines propriétés comme l'aire par lipide ou le paramètre d'ordre (Huang et García, 2014)). Concernant les interactions de longues portées électrostatiques, ces dernières sont calculées sur la base de la sommation d'Ewald (Ewald, 1921; Komeiji, 2000), qui nécessite une périodicité comme décrite précédemment (PBC). L'algorithme standard le plus utilisé en simulation moléculaire, grâce à son efficacité en temps de calcul, est l'algorithme PME (*Particle Mesh Ewald*, Darden et al. (1993);

Petersen (1995)). Depuis quelques années, cet algorithme a également été implémenté et appliqué aux interactions de type van der Waals pour des systèmes membranaires (Wennberg et al., 2013).

#### II.4.1.5 Effets de solvation

La représentation du solvant, qui est l'eau dans la plupart des cas, et de ses effets sur l'adsorption d'une protéine est un élément très important à prendre en compte. En effet, une simulation d'une interaction protéine-surface sans solvant représenterait un comportement moléculaire dans le vide, alors que les échange d'ions des solutions de protéines se font en phase aqueuse. De ce fait, les effets de solvation doivent être correctement représentés car les paramètres physiques de la couche de solvation tels que la densité, la constante diélectrique de l'eau ou encore l'énergie libre gouvernent en grande partie le phénomène d'adsorption.

En simulation tout-atome, la méthode la plus directe pour représenter les effets de solvation est de considérer explicitement toutes les molécules de solvant dans le système. De nombreux modèles d'eau ont été développés afin de décrire le comportement des molécules d'eau dans le système. Trois grands types de modèles ont été développés : (i) les modèles rigides, où les positions atomiques sont fixes, (ii) les modèles flexibles, où la position des atomes n'est pas restreinte et (iii) les modèles polarisables, qui incluent un terme de polarisation explicite. Parmi les modèles rigides, il existe des modèles d'eau de 3 à 6 sites : sur une molécule d'eau, les sites chargés peuvent se situer aussi bien sur les atomes que sur d'autres sites spécifiques (voir Figure II.13). Les modèles d'eau rigides à 3 sites sont les modèles les plus couramment utilisés en raison de leur simplicité et de leur performance dans les calculs, et plus particulièrement en dynamique moléculaire. Les plus connus sont les modèles SPC et SPC/E (Berendsen et al., 1987), TIP3P (Jorgensen et al., 1983). D'autres modèles, comme TIP4P, TIP4P/EW ou TIP5P (Jorgensen et al., 1983; Jorgensen et Madura, 1985; Mahoney et Jorgensen, 2000) sont également utilisés lorsque le nombre de sites considéré est supérieur à 3. Enfin, le choix des modèles d'eau se fait également en fonction du champ de force choisi pour les simulations car ils sont très souvent développés pour fonctionner avec des champs de force spécifiques (exemples : SPC pour Gromos, TIP3P pour CHARMM ou AMBER).

Le principal avantage de la méthode en solvant explicite est sa précision car elle fournit des informations détaillées sur chacune des molécules d'eau. De plus, celles-ci peuvent directement interagir avec les groupements fonctionnels des résidus de la protéine, mais également avec ceux du support chromatographique. En contrepartie, utiliser ce type de modèle demande une plus grande puissance de calcul, ou des temps de simulation plus longs, car le nombre d'atomes représentant l'ensemble du solvant est très important.

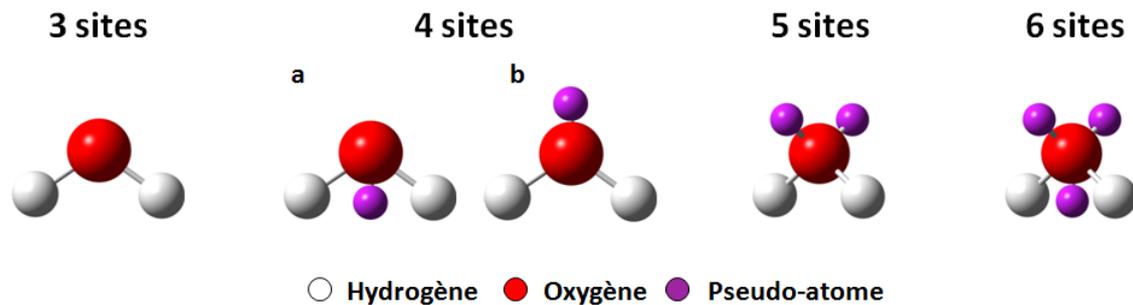


FIGURE II.13 – Représentation des différents modèles d'eau rigides et des sites chargés. Illustration issue de la thèse de Tesson (2016).

Afin de pouvoir réduire les temps de calculs, le solvant est parfois considéré de manière implicite lors des simulations. Dans ces types de modèles, l'eau n'est pas considéré comme un ensemble de molécules mais plutôt comme un milieu continu. Il s'agit d'un milieu uniforme polarisable possédant une constante diélectrique fixe. Ainsi, le système de simulation comprend beaucoup moins d'atomes et les calculs sont plus rapides. En revanche, comme pour tout système simplifié, des approximations sont faites et les méthodes en solvant implicite sont moins précises qu'en solvant explicite.

Lors de la solvatation, il y a deux effets importants qui doivent être correctement représentés : (i) les contributions électrostatiques, liées aux effets de solvatation autour des groupes fonctionnels chargés et polaires et (ii) les effets hydrophobes, liés aux effets de solvatation autour des groupes non polaires. La méthode la plus simple pour exprimer les effets électrostatiques est de définir une constante diélectrique relative (calculée par rapport à la constante diélectrique du vide), qui est incorporée à la loi de Coulomb dans le champ de force (équation II.20). Pour l'eau, cette constante  $\epsilon$  vaut 80 (Archer et Wang, 1990) à 298 K, et la constante diélectrique relative du vide est donc égale à 1. Cependant, cette méthode n'est pas recommandée car elle néglige totalement la compétition entre les molécules d'eau et les groupes chargés des résidus de la protéine, et également ne présente aucun moyen de prendre en compte les effets hydrophobes. Une méthode plus appropriée et bien plus précise est l'étude des effets électrostatiques par la résolution de l'équation de Poisson-Boltzmann (PB), qui sera détaillée dans le paragraphe suivant II.4.1.6. En revanche, cette méthode est en général trop précise, c'est à dire trop coûteuse en temps de calcul, pour être utilisée en dynamique moléculaire. Ainsi, la méthode de Born généralisée (*generalized-Born*, GB) est une procédure alternative pour approximer l'équation PB, et qui a déjà été très largement utilisée. Il faut cependant noter qu'aucune de ces deux méthodes ne permet de représenter les effets hydrophobes, qui sont en général calculés séparément. Le plus souvent, ils sont estimés par des calculs de surface accessible au solvant (*solvent-accessible surface area*, SASA), qui tiennent compte de la tension de surface du solvant. Cependant, l'estimation des effets hydrophobes reste

encore à améliorer, ces calculs étant des approximations car ils nécessitent des paramètres d'ajustement (Latour, 2008).

#### II.4.1.6 Énergies libres de liaison

L'estimation des énergies libres de liaison est un moyen de quantifier l'interaction entre une protéine  $P$  et un ligand  $L$ , ou éventuellement entre une protéine avec une surface, selon :



où  $PL$  représente le complexe protéine-ligand. D'une manière générale, l'enthalpie libre standard  $\Delta G$ , usuellement appelée énergie libre, est définie comme :

$$\Delta G = \Delta H - T\Delta S \quad (\text{II.23})$$

où  $\Delta H$  est la variation d'enthalpie,  $\Delta S$  la variation d'entropie et  $T$  la température absolue.

De nombreuses méthodes numériques ont été développées afin d'estimer ces valeurs d'énergies libres, qui peuvent être scindées en deux catégories principales : les méthodes *pathway* et les méthodes *end-point* (Åqvist et al., 2002). Les méthodes *pathway* consistent à échantillonner les conformations existantes entre les formes libres et liées du complexe protéine-ligand. Ainsi, le ligand est graduellement séparé de la protéine et l'énergie libre de liaison est calculée en fonction de toutes les contributions énergétiques entre l'état initial et l'état final. Deux méthodes courantes utilisant cette approche sont la méthode de l'intégration thermodynamique (TI) et la méthode de perturbation d'énergie libre (*Free-Energy Perturbation*, FEP) (Rao et al., 1987; Deng et Roux, 2009). Les méthodes *end-point* se basent sur un échantillonnage séparé entre les conformations libres et liées du complexe. Les contributions entrant dans le calcul de l'énergie libre (l'énergie interne, l'entropie et l'énergie de solvatation) sont déterminées distinctement pour chacune des conformations et l'énergie libre de liaison est ensuite déduite par différence. Cette technique pour le calcul des énergies libres est utilisée dans de nombreuses méthodologies comme par exemple la méthode d'énergie d'interaction linéaire (LIE), l'approximation de réponse linéaire (LRA) ou encore les méthodes MM-PBSA/MM-GBSA (Molecular Mechanics Poisson-Boltzmann Surface Area / Molecular Mechanics generalized-Born Surface Area). Les méthodes *pathway* sont les plus précises car elles prennent en compte les états de transition du système, mais ce sont aussi celles qui demandent le plus de puissance de calcul. Ainsi, les méthodes *end-point* telles que MM-PBSA ou MM-GBSA sont à ce jour les plus communément utilisées (Genheden et Ryde, 2015).

La méthode MM-PBSA, développée par Srinivasan et al. (1998) est devenue une des méthodes la plus employée pour estimer les interactions protéine-ligand (Homeyer et Gohlke, 2012). Elle est par exemple très utilisée afin d'évaluer la stabilité relative de structures biomoléculaires (Ferrari et al., 2007; Brice et Dominy, 2011). Elle permet également de décomposer les contributions énergétiques et ainsi de mieux comprendre en détail les mécanismes d'interactions principaux mis en jeu (Gohlke et al., 2003).

Dans cette approche, trois termes énergétiques sont combinés : (i) le premier terme correspond à l'énergie potentielle dans le vide, comprenant les énergies liées et non-liées (voir le paragraphe sur la mécanique moléculaire et la description des champs de force) ; (ii) le second correspond à l'énergie de solvation des espèces (énergies polaires et non-polaires) et (iii) le troisième est le terme entropique.

L'expression générale de l'énergie libre de liaison entre une protéine et un ligand peut s'écrire :

$$\Delta G_{liaison} = G_{complexe} - (G_{protéine} + G_{ligand}) \quad (\text{II.24})$$

où  $G_{complexe}$  représente l'énergie libre totale du complexe protéine-ligand et  $G_{protéine}$  et  $G_{ligand}$  les énergies libres totales de la protéine et du ligand, respectivement. Pour chacune de ces entités, l'énergie libre peut être définie comme :

$$G_x = \langle E_{MM} \rangle - TS + \langle G_{solvation} \rangle \quad (\text{II.25})$$

où  $x$  représente la protéine, le ligand ou le complexe protéine-ligand,  $\langle E_{MM} \rangle$  l'énergie potentielle moyenne dans le vide,  $TS$  la contribution entropique à l'énergie libre dans le vide (entropie et température) et  $\langle G_{solvation} \rangle$  l'énergie libre de solvation. Il faut noter que dans ces calculs, le terme entropique est négligé, et ce sont en réalité des enthalpies libres qui sont déterminées. Cependant, pour être en accord avec la littérature, le terme d'énergie libre sera systématiquement employé dans ce manuscrit.

Comme défini dans le paragraphe II.4.1.3, l'énergie potentielle  $E_{MM}$ , basée sur les paramètres de champs de force de mécanique moléculaire, inclue les énergies liées et non-liées. Dans le calcul MM-PBSA, l'hypothèse est faite de considérer  $\Delta E_{liée} = 0$  car les conformations de la protéine et du ligand à l'état libre ou dans le complexe sont considérés identiques. Ainsi, le terme  $E_{MM}$  est calculé uniquement selon les interactions non-liées, *i.e.* van der Waals (potentiel de Lennard-Jones, équation II.19) et électrostatiques (loi de Coulomb, équation II.20).

L'énergie libre de solvation représente l'énergie nécessaire pour transférer un soluté à partir du vide dans un solvant. Cette énergie est décomposée en deux termes selon l'équation :

$$G_{solvation} = G_{polaire} + G_{non-polaires} \quad (\text{II.26})$$

où  $G_{polaire}$  et  $G_{non-polaire}$  représentent les contributions électrostatiques et non-électrostatiques respectivement.

Le calcul du terme  $G_{polaire}$  est basé sur la résolution de l'équation de Poisson-Boltzmann (PB) :

$$\nabla \cdot [\varepsilon(r)\nabla \cdot \varphi(r)] - \varepsilon(r)\kappa(r)^2 \sinh[\varphi(r)] + \frac{4\pi\rho^f(r)}{kT} = 0 \quad (\text{II.27})$$

où  $\varphi(r)$  représente le potentiel électrostatique,  $\varepsilon(r)$  la constante diélectrique,  $\rho^f(r)$  la densité de charge et  $\kappa(r)^2$  est une valeur liée à la réciproque de la longueur de Debye et est dépendante de la force ionique.

Le terme non-électrostatique  $G_{non-polaire}$  peut être décrit à l'aide de nombreux modèles (Kumari et al., 2014), dont le plus utilisé est le modèle de la surface accessible au solvant (*solvent accessible surface area*, SASA). L'hypothèse principale de ce modèle est de considérer que l'énergie  $G_{non-polaire}$  est linéairement dépendante de la surface accessible au solvant. L'énergie peut ainsi être calculée comme étant :

$$G_{non-polaire} = \gamma A + b \quad (\text{II.28})$$

où  $A$  est la surface accessible au solvant,  $\gamma$  un coefficient relié à la tension de surface et  $b$  un paramètre d'ajustement.

La méthode MM-GBSA est une méthode alternative à la méthode MM-PBSA, dont la résolution numérique est plus rapide mais moins précise. En effet, cette méthode est basée sur la résolution de l'équation de Born-généralisée (GB) pour le calcul de l'énergie de solvatation, qui est une approximation de l'équation PB (Eq. II.27). La résolution de l'équation GB est de plus en plus employée car en plus d'être avantageuse en terme de coûts et de temps de calculs, elle présente des résultats similaires à ceux issus d'une résolution de l'équation de PB (Genheden et Ryde, 2015).

## II.4.2 Modélisation protéine-support

### II.4.2.1 Représentation des surfaces

L'adsorption de molécules, en particulier des protéines, sur divers types de surfaces a déjà fait l'objet de nombreuses études utilisant la simulation moléculaire (Ozboyaci et al., 2016). L'apport de la simulation moléculaire dans l'étude des interactions protéine-surface permet d'obtenir des informations à l'échelle atomique difficilement accessibles expérimentalement. Pour cela, les surfaces étudiées doivent être correctement représentées et paramétrisées afin de reproduire au mieux les phénomènes locaux.

De nombreux matériaux sont utilisés pour l'adsorption des protéines et ont fait l'objet d'une modélisation à l'échelle moléculaire (des exemples sont présentés sur la Figure II.14). Parmi eux, il y a les surfaces composées en majorité de métal (et alliages), qui servent le plus souvent comme matériaux inertes dans les capteurs et sondes, ou encore en tant qu'électrode. Ainsi, l'or (Kharazian et al., 2018), le cuivre (Chen et al., 2010), l'argent (Ghosh et al., 2012) ou encore le fer et les aciers (Imamura et al., 2003; Zhang et al., 2009b) ont déjà fait l'objet de nombreuses études numériques. Les surfaces composées de métaux oxydés, très utilisées pour leurs propriétés catalytiques et leur stabilité mécanique (Andreescu et al., 2012), ont également été étudiées (Ozboyaci et al., 2016). Il existe d'autres matériaux tels que les nanomatériaux à base de carbone (Scida et al., 2011) et divers polymères (Nie et Kumacheva, 2008) qui ont également été largement étudiés. Ces surfaces sont composées de plusieurs groupements greffés à une matrice par chimisorption et organisés en mono-couche (Schreiber, 2004). Ces groupements sont composés d'une tête, qui est la partie accrochée à la matrice, et d'une queue, qui est la partie fonctionnelle participant à l'adsorption. Elle peut être chargée positivement ou négativement, hydrophobe (neutre, non-polaire) ou encore hydrophile (neutre, polaire). Finalement, une grande partie des surfaces modélisées sont de type mono-couches auto-assemblées (ou *self-assembled monolayers*, SAMs).

#### CAS PARTICULIER : LES SURFACES CHROMATOGRAPHIQUES

En chromatographie d'échange d'ions, les surfaces des adsorbants sont composés d'une multitude de groupements fonctionnels chargés fixés à la matrice. La représentation des groupements fonctionnels, ou ligands, en simulation moléculaire nécessite dans un premier temps la connaissance de plusieurs paramètres essentiels, à savoir la structure chimique d'un ligand ainsi que la répartition sur la surface (ou densité de ligands). Généralement, la surface chromatographique est représentée uniquement par une couche de ligand où la matrice n'est pas explicitement décrite afin de limiter la taille du système de simulation. Par exemple, Dismar et Hubbuch (2010) ont reconstruit numériquement une surface de SP Sepharose FF à partir de la structure moléculaire d'un seul ligand, qui a été dupliqué en -x et -y. Afin de représenter le point de fixation du ligand à la matrice, les positions des premiers atomes du ligand ont été restreintes. C'est également le cas dans les études de Liang et al. (2012, 2015b), où la résine Q Sepharose FF a été construite en Gros Grain, sans représentation de la matrice en agarose.

Dans ces études, la surface est le plus souvent considérée comme plane, mais il peut arriver que la surface soit représentée sous forme cylindrique. En effet, un pore de résine ne possède pas une surface plane mais plutôt incurvée par endroit du fait de la tortuosité. Cependant, à l'échelle de la simulation moléculaire, la surface peut être considérée comme plane : la surface étudiée est en général extrêmement faible comparée à la surface totale de

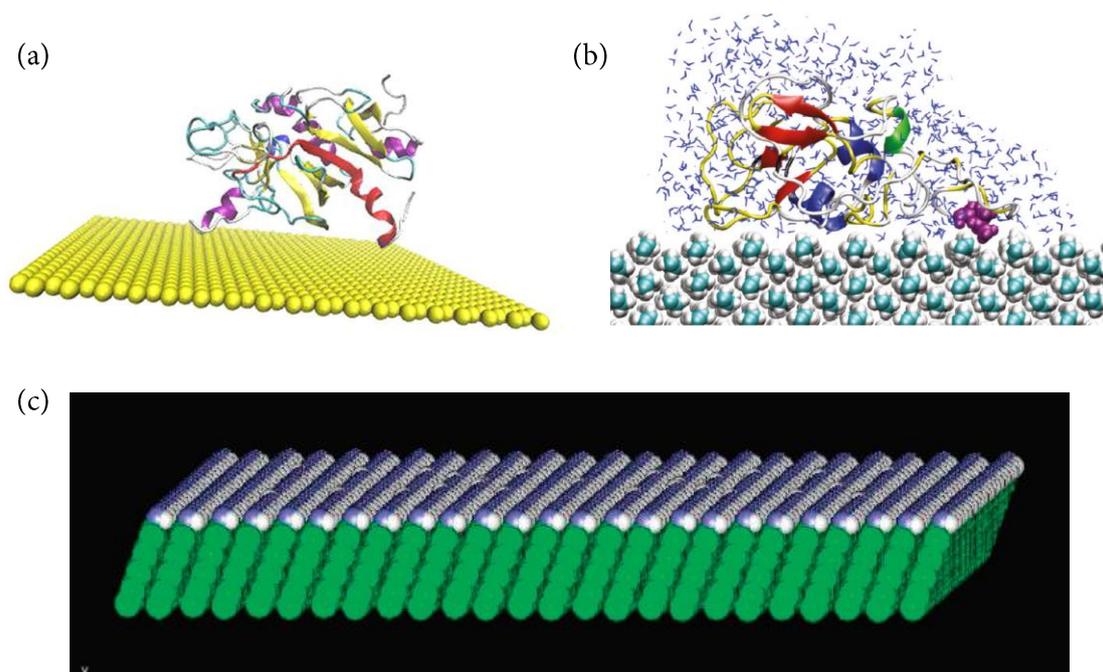


FIGURE II.14 – Exemples de représentations de surfaces en simulation moléculaire : (a) Fibrinogène sur une surface d'or. Issu de Kharazian et al. (2018), reproduit avec permission (CC licence); (b) lysozyme sur une surface polyéthylène. Issu de Wei et al. (2011), reproduit avec permission. Copyright (2011) American Chemical Society; (c) Illustration d'une surface de type SAM alkanéthiol avec une terminaison amine ( $\text{NH}_2$ ). Issu de Agashe et al. (2005), reproduit avec permission. Copyright (2005) American Chemical Society

l'adsorbant et la courbure sur une si petite surface peut être considérée comme négligeable. Néanmoins, dans certaines études comme celles de Riccardi et al. (2008), la simulation moléculaire est utilisée comme outil afin d'étudier et de créer de nouvelles surfaces. Dans cet exemple, une surface sphérique en agarose (explicitement représentée) est modélisée. Zhang et al. (2009a) ont également travaillé sur l'étude de surfaces cylindriques. Dans ces cas particuliers de *design* de nouvelles surfaces, il peut donc être intéressant de modifier la morphologie et/ou la topologie de la surface afin d'étudier l'impact sur la rétention. Enfin, l'un des intérêts de l'apport de la simulation est également de pouvoir étudier l'influence de la densité de ligands sur l'échange, comme l'ont fait Lang et al. (2015b), où des simulations ont été réalisées en utilisant deux surfaces de densité différentes (voir Figure II.15). L'intérêt de cette étude est d'autant plus grand lorsque l'on connaît les difficultés à obtenir expérimentalement des valeurs correctes de surface spécifique avec des résines en matériau de type agarose (voir paragraphe II.3.2.3). Il est également possible d'étudier l'influence de la répartition ou distribution des ligands sur la surface. Celle-ci est souvent considérée uniforme comme sur la Figure II.15, mais il peut être intéressant de varier les

points de fixation des ligands, comme l'ont fait Zhang et al. (2010) qui ont pris en compte la forme hexaédrique caractéristique d'une matrice en agarose.

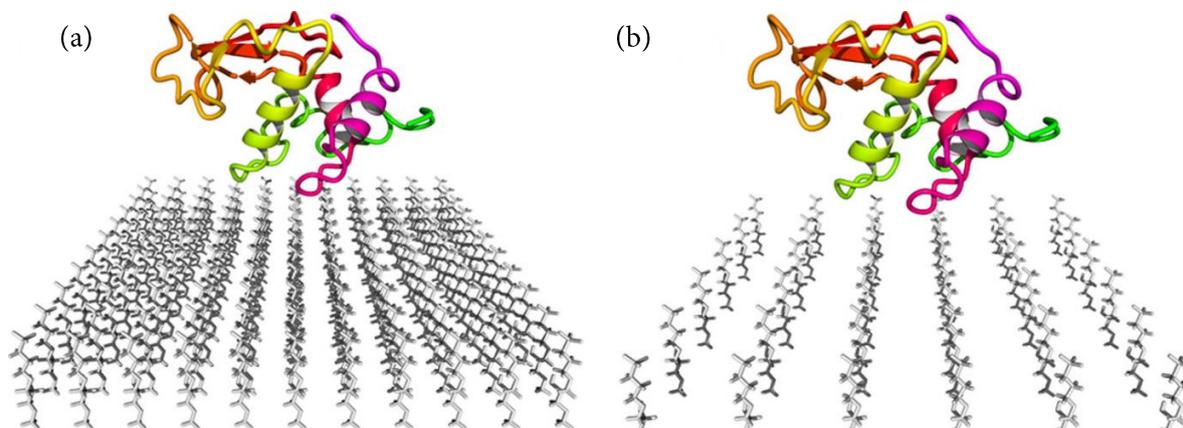


FIGURE II.15 – Lysozyme sur deux adsorbants, avec (a) une distance entre ligands de 6,51 Å et (b) une distance entre ligands de 12,14 Å. Issu de Lang et al. (2015b), reproduit avec permission. Copyright (2015), Elsevier.

### II.4.2.2 Simulations moléculaires protéine-support

La compréhension des phénomènes locaux lors de l'échange d'ions a fait l'objet de plusieurs études combinant des expériences à des simulations moléculaires. Pour cela, diverses protéines ont été étudiées, le plus souvent des protéines dites "modèles", qui sont très bien connues et facilement disponibles commercialement sous forme purifiée. Le lysozyme, la ribonucléase A, le cytochrome c ou encore la BSA (albumine de sérum bovin) peuvent être cités à titre d'exemple parmi les principales protéines usuellement utilisées.

Il existe plusieurs approches dans la littérature afin de relier des simulations à l'échelle atomique avec des expériences à l'échelle macroscopique. Deux stratégies principales peuvent être distinguées : (i) la première se base sur des simulations en tout-atome et s'applique essentiellement à des petits systèmes (peptides ou petites protéines < 30-40 kDa) alors que (ii) la deuxième exploite des simulations plus simplifiées de type Gros Grain afin de pouvoir étudier de plus gros systèmes et sur une échelle de temps plus importante. Cette partie sera consacrée à ces deux approches, qui sont à ce jour les plus abouties en terme de simulations en dynamique moléculaire dans le cas spécifique de la chromatographie d'échange d'ions de protéines. Néanmoins, il existe d'autres travaux, comme ceux de Latour et al (Latour, 2008; Collier et al., 2009; Yancey et al., 2010; Latour, 2014), qui se focalisent sur le développement de méthodes permettant la bonne description des interactions et des phénomènes de solvatation dans des cas d'interactions peptide-surface.

## SIMULATIONS TOUT-ATOME

Afin d'effectuer une comparaison entre échelle atomique et échelle macroscopique, il est important de déterminer des paramètres de référence possédant un sens physique réel, mesurables directement ou indirectement via les deux approches.

C'est le cas notamment des travaux dirigés par J. Hubbuch durant ces 10 dernières années. Sa recherche s'est notamment focalisée sur l'utilisation de simulations en tout-atome appliquées à de petits systèmes, dans un but de développement d'outils prédictifs. Dimer et Hubbuch (2010) ont premièrement étudié l'adsorption du lysozyme sur la résine SP Sepharose FF. Ils ont démontré que le lysozyme présentait plusieurs surfaces d'adsorption possibles, appelés *patches*, qui correspondent à différents résidus chargés positivement à la surface accessible de la protéine. Ainsi, différentes orientations du lysozyme adsorbé sur la surface chromatographique ont été identifiées expérimentalement à différents pH (Dimer et Hubbuch, 2007; Dimer et al., 2008), puis confrontées aux résultats des simulations moléculaires. Dans cette étude, de très courtes simulations (100 ps) ont été réalisées pour 62 orientations du lysozyme, dans le but de calculer l'énergie d'interaction entre la protéine et le support. Une carte d'énergie (*energy map*) représentant la surface projetée de la protéine a été réalisée (comme l'illustre la Figure II.16) et les différents *patches* d'adsorption identifiés correspondent bien à ceux précédemment identifiés par les expériences. Dans leurs études, Dimer et Hubbuch (2010) ont également comparé l'énergie d'interaction avec une donnée expérimentale aisément mesurable : la conductivité nécessaire à l'élution de la protéine. Ainsi, ils ont montré une corrélation entre ces deux valeurs, qui peut s'expliquer par le fait que chacun de ces deux paramètres représente la « force » de l'adsorption. En effet, plus l'énergie d'interaction est basse et plus l'interaction est favorable car énergétiquement stable.

Par la suite, Lang et al. (2015a) ont développé une méthodologie permettant de prédire les interactions, et en particulier les énergies d'interactions électrostatiques, applicables à plusieurs protéines et sur différents supports chromatographiques (notamment la Q et SP Sepharose FF). Cette méthodologie a été testée et validée par confrontation avec des données expérimentales (Kittelman et al., 2017a), comme l'ont fait Dimer et Hubbuch (2010). Cette technique permet également d'identifier les résidus participant à l'adsorption pour la plupart des protéines en dessinant des *energy maps*, représentant l'énergie électrostatique sur toute la surface accessible de la protéine (voir Figure II.16). Lang et al ont réutilisé cette méthodologie dans le but d'étudier diverses densités de ligands et ainsi évaluer l'impact sur l'adsorption (Lang et al., 2015b). Par ailleurs, leur étude a montré que si la densité de ligands est un paramètre important à prendre en compte, d'autres paramètres tels que le diamètre des pores de la résine et la surface réellement accessible sont également des critères importants, qui influencent la capacité de rétention d'une résine échangeuse d'ions.

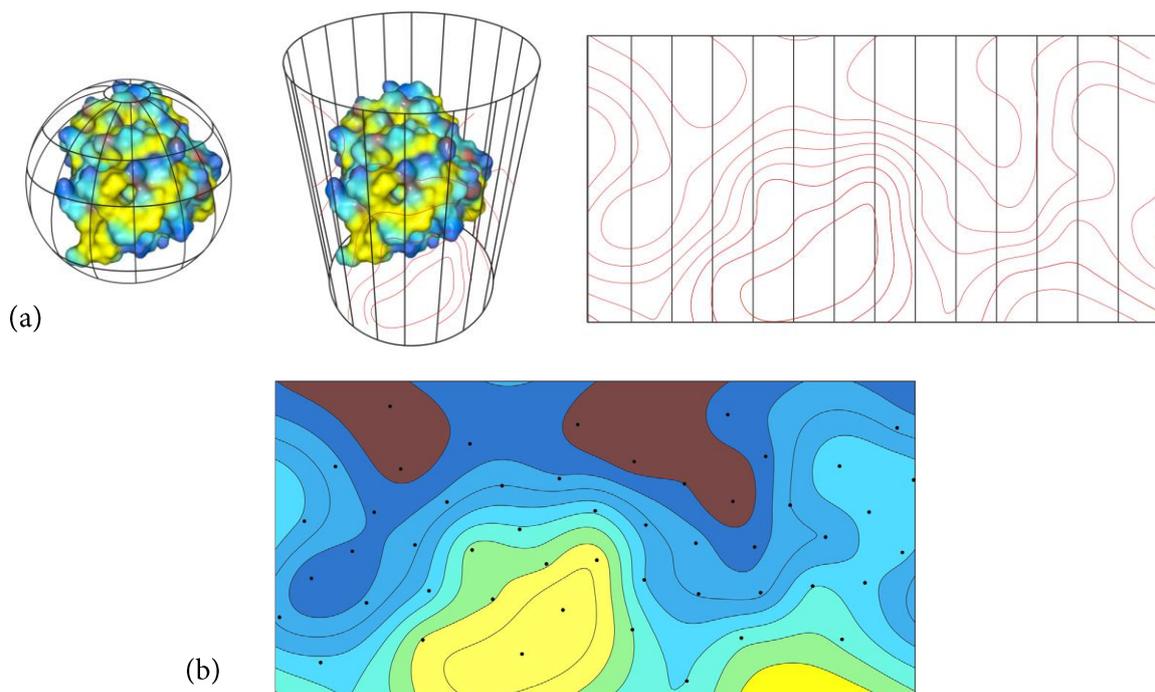


FIGURE II.16 – (a) Projection cylindrique de points d'une protéine, puis *unwrapping* et aplatissement du cylindre aboutissant à (b) une carte d'énergie ou d'interaction (*energy map*). Issu de Lang et al. (2015a), reproduit avec permission. Copyright (2015), Elsevier.

Steudle et Pleiss (2011) ont également étudié la rétention du lysozyme sur la résine Source 15S via des simulations de dynamique moléculaire. Leurs travaux ont montré que le lysozyme ne possède pas une seule orientation d'adsorption préférentielle mais plusieurs, ce qui rejoint les conclusions des études de Dismar et Hubbuch (2010), mais également de Kubiak et Mulheran (2009). De plus, il semblerait que la flexibilité du modèle utilisé ait une influence sur la qualité de la prédiction, c'est-à-dire qu'il est préférable de ne pas restreindre la position de la protéine (et des chaînes latérales) lors des simulations et de conduire de longues simulations afin de permettre à la protéine de trouver son orientation la plus favorable à l'adsorption.

## SIMULATIONS GROS GRAIN

Assez peu de travaux dans la littérature s'intéressent aux interactions et à la compétition protéine-protéine lors de l'échange d'ions. En effet, des simulations en multi-protéines requièrent une taille de système importante et donc une plus grande puissance de calcul. Pour cette raison, Liang et al (Liang et al., 2012, 2015b,a) ont étudié la rétention d'un système bi-protéines, couplant la BSA/HSA avec la bHb, sur une résine anionique (Q

Sepharose FF) en réalisant des simulations en gros grains. Ces deux protéines sont de grandes tailles et utiliser des modèles en tout-atome conduiraient à des systèmes avec un nombre conséquent d'atomes. En effet, en plus de considérer la taille des protéines, il faut construire une boîte de simulation assez grande pour éviter les effets de bords (voir paragraphe II.4.1.4 sur les conditions périodiques) et la remplir de molécules d'eau et d'ions, ce qui augmente très fortement le nombre d'atomes total. Ainsi, les simulations en gros grains permettent de réduire le nombre de particules mais également de conduire des simulations plus longues qu'en tout-atome (400 ns à 2  $\mu$ s). Dans l'étude de l'adsorption, il est indispensable de réaliser de longues simulations pour pouvoir observer le phénomène d'adsorption, qui peut mettre du temps à se produire selon les simulations. Liang et al ont étudié l'adsorption des deux protéines tout d'abord séparément, puis en adsorption simultanée (les deux protéines sont placées dans la boîte au début des simulations) et séquentielle (une protéine est placée dans la boîte, puis l'autre). Ils ont montré tout d'abord que la BSA (ou HSA) a une meilleure affinité avec la Q Sepharose FF que la bHb car elle possède de nombreux sites de fixation sur la surface chromatographique. Elle s'adsorbe également plus rapidement que la bHb.

Deux paramètres SMA, la charge et le facteur stérique, ont été calculés pour chaque simulation puis moyennés, et les valeurs ont été comparées à celles issues des expériences. Dans cette étude, le calcul de ces paramètres est la méthode de comparaison entre les simulations et les expériences macroscopiques choisie, les deux paramètres ayant une signification physique. Ainsi, la charge a été calculée comme étant le nombre de résidus de la protéine participant à l'interaction (distance entre résidus et ligands inférieure à 0,45 nm) alors que le facteur stérique a été calculé comme étant le nombre de ligands dont la distance est comprise entre 0,45 et 2,5 nm (rayon de gyration de la protéine).

Les simulations ont montré une très forte interaction entre les deux protéines, ce qui a influencé l'adsorption de deux manières : (i) les deux protéines se sont "gênées" dans leurs mouvements dans la boîte, modifiant ainsi potentiellement les orientations préférentielles d'adsorption et (ii) la HSA a montré une plus forte affinité pour la résine et la bHb a préféré interagir avec la HSA plutôt qu'avec le support chromatographique, formant ainsi une adsorption multicouche (Liang et al., 2015b). En revanche, ce phénomène n'est pas nettement visible expérimentalement, les points des isothermes en multiconstituants étant trop dispersés pour être correctement exploités.

## Conclusion

La chromatographie d'échange d'ions des protéines est un domaine qui a été très largement étudié au cours des dernières décennies et sur lequel de nombreuses connaissances ont été acquises (modèles d'équilibre, cinétiques, modes d'opération). Malgré tout, le

manque de connaissances des phénomènes intervenant lors de la rétention de la protéine à l'échelle atomique et parfois leur faible caractère prédictif, a conduit à envisager de nouvelles approches afin de pallier ce manque d'informations. Parmi elles, une stratégie émergente ces dernières années et qui semble avoir un fort potentiel est l'utilisation de simulations moléculaires afin de mieux comprendre les mécanismes d'échange.

Finalement, relativement peu d'études ont été réalisées sur ce sujet, mais les premiers travaux combinant des simulations moléculaires avec des expériences semblent très prometteurs. Deux approches principales ont été identifiées dans la littérature, proposant des niveaux de description des systèmes différents. En effet, si la première stratégie a pour but de simuler des petits systèmes contenant des molécules de petite taille (moins de 30 kDa environ) dans un but de développement d'outils prédictifs, la deuxième stratégie, utilisant des modèles simplifiés et donc moins précis, permet d'étudier de plus gros systèmes mais également d'atteindre des échelles de temps plus grandes. Ces études ont également soulevé l'intérêt de l'utilisation de la dynamique moléculaire pour mieux comprendre les interactions au sein de systèmes multi-protéines.

# III

---

## ÉTUDE EXPÉRIMENTALE

---

### Introduction

Afin de pouvoir juger de la pertinence des simulations moléculaires pour décrire les phénomènes en chromatographie d'échange d'ions à l'échelle atomique, il est nécessaire de les confronter à des expériences, qui feront l'objet de ce chapitre. Il est d'usage en chromatographie d'échange d'ions d'étudier expérimentalement ces phénomènes à l'échelle macroscopique, les techniques expérimentales microscopiques étant chères et compliquées à mettre en œuvre (Zhang et Sun, 2010). En général, l'étude préliminaire de l'échange d'ions consiste à mesurer des isothermes d'adsorption. Celles-ci sont décrites par de nombreux modèles dans la littérature, mais assez peu permettent de décrire réellement le phénomène d'échange d'ions qui se produit à l'intérieur de l'échangeur. En effet, les phénomènes mis en jeu sont particulièrement complexes à décrire avec des modèles mathématiques. La loi d'Action de Masse Stérique (SMA) est le modèle le plus utilisé à ce jour car il permet de prendre en compte le déplacement des ions. Sa particularité repose aussi sur le fait qu'il est applicable à l'étude de molécules de grande taille telles que les protéines car le phénomène d'encombrement stérique qui en résulte est également décrit.

L'objectif de ce chapitre est dans un premier temps de présenter les matériels et méthodes utilisés afin d'obtenir des isothermes d'adsorption sur un système modèle, composé d'une résine couramment employée (SP Sepharose FF) et de deux protéines modèles ( $\alpha$ -chymotrypsine et lysozyme). Dans une seconde partie seront présentés et discutés les résultats obtenus. L'affinité protéine-support sera étudiée, tout d'abord en système mono-protéine puis en système bi-protéines afin d'observer de potentiels phénomènes de compétition entre protéines. L'influence de paramètres tels que la force ionique et le pH sera également discuté.

## III.1 Matériels

### III.1.1 Support chromatographique

La résine SP Sepharose FF, fabriquée et fournie par GE Healthcare (Uppsala, Suède), est une résine à base d'agarose réticulé très utilisée en chromatographie d'échange d'ions. C'est une résine cationique forte possédant des groupements sulfopropyl à la surface (voir paragraphe II.3.2.2, Tableau II.2).

#### MESURE DE LA CAPACITÉ IONIQUE

La capacité ionique totale d'une résine échangeuse d'ions correspond au nombre total de sites de liaisons disponibles, exprimée généralement en quantité de sites (*mmol* ou *meq*) par quantité de résine (volumique ou massique). Il est possible de déterminer expérimentalement la capacité totale de manière assez simple, surtout dans le cas d'une résine cationique telle que la SP Sepharose FF. Le principe est de saturer la surface de la résine avec un ion, de préférence monovalent et de petite taille. Cet ion est alors élué en totalité par échange avec un autre ion. L'éluat est récupéré et dosé : la quantité de cet ion initialement fixé est alors égale à la quantité de sites présents à la surface de la résine (dans le cas d'un ion monovalent, un ion correspond à un site de fixation).

Lors de la mesure de la capacité ionique, la résine est préalablement packée dans une colonne de 1 mL (colonne HiTrap SP FF, 1 mL). Les sites de la résine étant chargés négativement ( $\text{SO}_3^-$ ), la résine est préalablement équilibrée avec des ions  $\text{H}^+$ . Pour cela, une grande quantité d'acide chlorhydrique (HCl) concentré est injecté dans la colonne, afin de s'assurer de la saturation totale de la résine. L'intérêt de travailler avec des ions de très petite taille est de pouvoir recouvrir la totalité de la surface, même lorsque les pores de la résine deviennent extrêmement petits. En effet, certaines molécules trop grosses peuvent être bloquées rapidement et ne pas s'accrocher sur la totalité des sites disponibles à l'intérieur des pores. Il est également possible de contrôler facilement la concentration en  $\text{H}^+$  à la sortie de la colonne : si elle est égale à celle en entrée, cela signifie qu'il n'y a plus d'ions  $\text{H}^+$  retenus par la résine et donc que celle-ci est saturée.

Après l'étape de charge, une étape de rinçage à l'eau ultra-pure est nécessaire afin d'éliminer tout les ions interstitiels qui ne sont pas fixés sur un site. Une fois cette étape réalisée, il faut élué la totalité des ions  $\text{H}^+$  de la résine. Pour cela, entre 25 et 50 mL d'une solution concentrée de chlorure de sodium (NaCl) à 1 M est injectée dans la colonne. L'éluat contenant les ions  $\text{H}^+$  à doser est récupéré dans une fiole jaugée de 100 mL et le volume est ajusté avec de l'eau ultra-pure.

La solution récupérée est ensuite dosée par titrage colorimétrique avec une solution de soude (NaOH). L'indicateur coloré utilisé est le bleu de bromothymol (BBT). À noter qu'il est possible d'estimer le volume équivalent et donc la concentration en NaOH adéquate en considérant les valeurs de capacité totale issues du fournisseur (entre 180 et 250 meq/mL (GE-Healthcare, 2014)). Cela permet d'optimiser la mesure et de choisir une concentration en NaOH conduisant à un volume équivalent le plus grand possible, et ce afin de minimiser les incertitudes de mesure.

Finalement, la capacité ionique totale de la résine notée  $q_R$ , exprimée en volume de lit de résine, est de  $230.2 \pm 3.3$  mmol/L.

### MESURE DU TAUX D'HUMIDITÉ

Du fait de sa matrice à base d'agarose extrêmement hygroscopique, il est important de mesurer le taux d'humidité de la résine avant chaque expérience afin d'obtenir des résultats comparables et reproductibles. Toutes les mesures sont effectuées en triplicata. Une masse de résine humide connue  $m_{coup+rés,h}$  est introduite dans une coupelle en aluminium, dont la masse  $m_{coup}$  est également connue. La coupelle est ensuite mise à l'étuve à  $105^\circ\text{C}$  pendant au minimum 2 heures afin d'éliminer l'eau présente. La coupelle contenant la résine sèche est ensuite pesée ( $m_{coup+rés,s}$ ) et l'humidité de la résine peut être calculée selon :

$$\text{Humidité} = \frac{m_{coup+rés,h} - m_{coup+rés,s}}{m_{coup+rés,h} - m_{coup}} \quad (\text{III.1})$$

Les valeurs des trois mesures sont moyennées. Dans le cas de la SP Sepharose FF, le taux d'humidité moyen est aux alentours de 85%. Il correspond à la quantité de liquide présente entre les particules de résine mais également à celle présente dans les pores de la résine.

### STRUCTURE PHYSIQUE

Le diamètre des particules de SP Sepharose FF est d'environ  $90 \mu\text{m}$  selon le fournisseur (GE-Healthcare, 2014). Cette valeur est en réalité une moyenne et on a plus affaire à une distribution de taille de particules allant de  $45$  à  $165 \mu\text{m}$  (DePhillips et Lenhoff, 2000).

Le diamètre des pores est une information rarement fournie par les fabricants de résines. En revanche, quelques études dans la littérature proposent des valeurs issues d'expériences, mettant en œuvre généralement de la chromatographie d'exclusion de taille en utilisant des particules neutres de tailles différentes comme le dextrane (DePhillips et Lenhoff, 2000; Yao et Lenhoff, 2004, 2006). Dans le cas de la SP Sepharose FF, qui a été

très étudiée, les travaux de Yao et Lenhoff (2006) sur la caractérisation de la distribution des tailles de pores suggèrent une valeur moyenne de 58 nm.

Les porosités des résines, intra- et extra-particulaire, font également l'objet de nombreuses études expérimentales afin de les caractériser. Selon Ubiera et Carta (2006), la porosité intraparticulaire de la SP Sepharose FF est d'environ 0,90. Sa porosité extraparticulaire est quant à elle estimée à 0,35, qui est une valeur de porosité (ou de "degré de vide") standard dans des cas de lit de particules sphériques monodisperses.

Les principales caractéristiques de la SP Sepharose FF sont présentées dans le tableau récapitulatif III.1.

TABLEAU III.1 – Tableau récapitulatif des principales propriétés de la résine SP Sepharose FF.

Propriétés	SP Sepharose FF
Matrice	Agarose réticulé à 6%
Groupe fonctionnel	Sulfopropyl (SP) : $-O-CH_2-CH_2-CH_2-SO_3^-$
Capacité ionique	230,2 meq/L (entre 180 et 250 meq/L selon le fournisseur)
Capacité dynamique*	70 mg/mL
Diamètre moyen des particules	90 $\mu$ m
Diamètre moyen des pores	58 nm
Porosité	Intraparticulaire : 0,90 Extraparticulaire : 0,35
pH de travail	4 à 13

\* Capacité à 50% avec une solution de Ribonucléase A dans une solution tampon acetate 0,1M, NaCl 2M (pH 5)

### III.1.2 Protéines

Initialement, lors du choix d'un système modèle pour l'étude en système mono-constituant, il avait été décidé de travailler avec une autre enzyme, la CAL-B (Lipase B de *Candida antarctica*). Cette protéine possédait l'avantage d'être bien connue et couramment utilisée. Cependant, cette protéine n'est pas disponible purifiée commercialement et sa purification est difficile, comme expliqué en partie dans le paragraphe II.1.2.2. De plus, cette protéine étant habituellement vendue commercialement dans un milieu brut, sa purification en quantité suffisante pour la réalisation des expériences a été très inefficace, malgré les ef-

forts réalisés durant plusieurs mois de travail. Pour l'ensemble de ces raisons, les résultats préliminaires obtenus avec cette protéine ne seront pas présentés dans ce manuscrit.

Finalement, l' $\alpha$ -chymotrypsine a été utilisée comme protéine modèle pour les études en système mono-constituant, ainsi que le lysozyme qui a été utilisé en supplément de l' $\alpha$ -chymotrypsine pour l'étude du système en multi-constituants.

### III.1.2.1 $\alpha$ -chymotrypsine

L' $\alpha$ -chymotrypsine (notée  $\alpha$ Chy) est une enzyme qui peut être utilisée en tant que protéine modèle en chromatographie d'échange d'ions, de par sa petite taille et sa bonne stabilité (Xu et Glatz, 2009; Zhao et al., 2012; Huuk et al., 2014). C'est plus particulièrement une protéase digestive présente dans le pancréas qui catalyse les protéolyses. Elle est tout d'abord synthétisée sous forme de chymotrypsinogène, qui est inactive et stockée dans le pancréas. Son activation suite à plusieurs hydrolyses forme la chymotrypsine, qui devient une protéine globulaire compacte composée de trois chaînes reliées par deux ponts disulfures. Il existe deux formes de l' $\alpha$ Chy, A et B, qui possèdent 80% de similitudes mais ont des propriétés différentes comme par exemple le pH isoélectrique (pI). C'est la forme A qui a été utilisée dans cette thèse.

Dans sa forme monomérique, l' $\alpha$ Chy est composée de 241 acides aminés et possède une masse moléculaire d'environ 25,6 kDa. Son point isoélectrique se situe aux alentours de 8,3 (Laskowski, 1955), ce qui signifie que la protéine est globalement de charge positive à des pH inférieurs à cette valeur. Il a été montré que l' $\alpha$ Chy a tendance à former un dimère dans des conditions particulières de force ionique et de pH (Aune et Timasheff, 1971; Aune et al., 1971). En effet, les conditions les plus favorables à sa dimérisation sont un pH de 4 et une concentration en sel supérieure à 100 mM. Afin de s'assurer que la protéine reste bien sous sa forme monomérique, une analyse de sa taille moléculaire sera ultérieurement effectuée dans les conditions de travail étudiées.

L' $\alpha$ Chy utilisée dans les expériences est une enzyme commerciale issue du pancréas bovin (Sigma C-4129, pureté  $\geq 85\%$ ). Conditionnée sous forme de poudre, elle a été utilisée sans aucune purification préalable.

### III.1.2.2 Lysozyme

Le lysozyme est une protéine globulaire présente chez de nombreuses espèces d'animaux. Il se trouve essentiellement dans le blanc d'œuf. Cette protéine, qui est également une enzyme capable de lyser des bactéries est impliquée dans la défense contre les infections bactériennes. Elle est de taille relativement petite : elle possède une chaîne principale composée de 129 acides aminés et sa masse moléculaire est de 14,3 kDa. Du fait de sa taille et de son abondance, elle est peu onéreuse sous sa forme purifiée et est l'une des plus utilisées dans la littérature dans l'étude des systèmes en échange cationique. En effet, son

point isoélectrique haut, égal à 11,3, fait qu'elle se sépare très bien par chromatographie d'échange d'ions et des puretés assez élevées peuvent être atteintes.

Dans cette étude, le lysozyme utilisé provient de chez Sigma (L-6876, pureté  $\geq 90\%$ ) et correspond au lysozyme dit de type C que l'on trouve dans le blanc d'œuf de poule. Comme pour l' $\alpha$ Chy, la protéine est conditionnée sous forme de poudre et a été utilisée sans aucune purification préalable.

TABLEAU III.2 – Tableau récapitulatif des principales propriétés de l' $\alpha$ -chymotrypsine et du lysozyme.

Propriétés	$\alpha$ -chymotrypsine	Lysozyme
Masse moléculaire	25,6 kDa	14,3 kDa
Acides aminés (chaînes)	241 (3)	129 (1)
Point isoélectrique (pI)	8,3	11,3
Référence Sigma	C-4129	L-6876
Pureté	$\geq 85\%$	$\geq 90\%$

### III.1.3 Solutions tampons

Le choix de la solution tampon lors de l'échange d'ions détermine le succès de la séparation car ce sont ces solutions qui conditionnent le pH et la force ionique de fonctionnement. En effet, la stabilité et la charge des protéines en sont dépendantes.

En échange d'ions, les concentrations en ions ne doivent pas être trop élevées car cela induit une augmentation de la force ionique. Plus la force ionique en solution est importante et moins la force de liaison entre la protéine et le support chromatographique est grande. Il est donc important de trouver un compromis entre force ionique suffisante pour maintenir l'effet tampon, mais pas trop élevée pour ne pas inhiber la séparation.

Il existe un grand nombre de solutions tampons utilisées en chromatographie d'échange d'ions. Dans cette thèse, deux pH différents seront étudiés : pH 5 et pH 7. A pH 5, la solution tampon choisie est un tampon citrate, composé de citrate de sodium (ou citrate trisodique) et de son acide conjuguée, l'acide citrique ( $pK_{a1} = 3,13$  ;  $pK_{a2} = 4,76$  ;  $pK_{a3} = 6,40$ ) selon le protocole par Gomori (1955). La solution est préparée en pesant exactement les masses nécessaires pour atteindre le pH voulu. Si besoin, le pH est réajusté par ajout d'acide ou de base en contrôlant sa valeur à l'aide d'un pHmètre. Dans ce tampon, l'ion

positif qui va venir neutraliser le support chromatographique est l'ion sodium  $\text{Na}^+$ . A pH 7, pour obtenir un pouvoir tampon important, d'autres tampons comme les tampons à base de phosphate sont couramment utilisés. Le critère principal pour le choix de la nouvelle solution tampon est la présence de l'ion  $\text{Na}^+$  et une force ionique similaire. Ainsi, à pH 7, le tampon choisi est un tampon phosphate composé de dihydrogénophosphate de sodium ( $\text{NaH}_2\text{PO}_4$ ) et d'hydrogénophosphate de sodium ( $\text{Na}_2\text{HPO}_4$ ) selon le protocole par Sørensen (Gomori, 1955). Afin de conserver exactement la même quantité d'ions sodium entre les deux tampons, un peu de  $\text{NaCl}$  est ajouté par la suite au tampon phosphate afin de réajuster la concentration.

Ici, la quantité d'ions  $\text{Na}^+$  est souhaitée identique entre les deux solutions tampons. Néanmoins, cela ne veut pas dire que la force ionique de ces deux solutions soit la même. La force ionique d'une solution, notée  $I$ , est définie comme :

$$I = \frac{1}{2} \sum C_i \cdot z_i^2 \quad (\text{III.2})$$

avec  $C_i$  la concentration molaire de l'ion  $i$  et  $z_i$  sa valence. Ainsi, une solution tampon citrate 50 mM à pH 5 sans ajout de  $\text{NaCl}$  possède une force ionique égale à 0,22 M alors qu'une solution tampon phosphate 50 mM à pH 7 avec ajout de 0,018 M de  $\text{NaCl}$  possède une force ionique égale à 0,13 M.

Tous les sels utilisés pour les expériences proviennent de chez Sigma et possèdent une qualité analytique. Toutes les solutions tampons sont préparées par dilution dans de l'eau ultra-pure.

## III.2 Méthodes analytiques pour la quantification de protéines

De nombreuses méthodes d'analyse ont été mises au point afin de doser les protéines (Noble et Bailey, 2009; Grotefend et al., 2012). Parmi elles, il existe (i) des méthodes globales non sélectives, où la concentration en protéines totale est mesurée ; (ii) des méthodes plus spécifiques visant à doser différentes protéines dans un milieu, le plus souvent par des méthodes séparatives. Dans cette section seront présentées uniquement les méthodes d'analyse utilisées dans le cadre de la thèse, à savoir la spectrophotométrie UV-visible, l'HPLC-SEC et la mesure de l'activité enzymatique.

### III.2.1 Analyse par spectrophotométrie UV-visible

L'une des méthodes les plus simples à mettre en œuvre est la lecture directe de l'absorbance en UV d'une solution de protéines. Les acides aminés tryptophane (Trp) et tyrosine (Tyr), ou plus spécifiquement leurs cycles aromatiques, absorbent fortement les ultra-violets à une longueur d'ondes de 280 nm. Il est donc possible de déterminer la concentration en protéines en solution en mesurant l'absorbance à 280 nm.

Les avantages de cette méthode sont sa simplicité et sa rapidité de mise en œuvre. C'est également une méthode non destructive car l'échantillon n'est pas altéré lors de la mesure. En revanche, il est nécessaire de connaître le coefficient d'extinction de la protéine et/ou d'effectuer une courbe d'étalonnage à partir de solutions de protéines pures et de concentrations connues. Cette méthode est également non sélective et ne permet pas de connaître les concentrations de chacune des protéines, mais uniquement la concentration globale. Elle est donc utilisée principalement dans l'analyse de systèmes mono-protéine. Le dernier inconvénient à prendre en compte est la possible interférence avec d'autres composés absorbant également à 280 nm (ou à longueurs d'ondes proches), comme les acides nucléiques qui absorbent les UV à 260 nm.

#### QUANTIFICATION DE L' $\alpha$ -CHYMOTRYPSINE

L'absorbance en laboratoire a été mesurée à l'aide d'un spectrophotomètre à double faisceau de type JASCO V630. Des cuves en quartz macroscopiques (2,5 mL), préalablement rincées avec la solution à analyser avant chaque mesure, ont également été utilisées. Enfin, chaque mesure correspond à la moyenne de trois valeurs mesurées à 5 secondes d'intervalle.

La courbe d'étalonnage de l' $\alpha$ Chy, permettant de relier l'absorbance à 280 nm à sa concentration, a été réalisée pour une gamme de concentration allant de 0,1 à 1 g/L, dans un tampon citrate 50 mM à pH 5. La solution mère d' $\alpha$ Chy a été préparée à partir de la protéine sous forme solide et purifiée de chez Sigma. Un spectre d'absorbance de la solution a été réalisé en amont afin de s'assurer qu'il n'y ait pas d'autres espèces susceptibles d'interférer le signal. Ainsi, la concentration en  $\alpha$ Chy ( $C_{\alpha Chy}$ ) peut être exprimée selon l'absorbance à 280 nm (ou densité optique D.O.) :

$$D.O. = 2,1893.C_{\alpha Chy} \quad (R^2 = 0,9999) \quad (III.3)$$

avec  $C_{\alpha Chy}$  exprimée en g/L.

## QUANTIFICATION DU LYSOZYME

Le lysozyme a été quantifié par UV-vis de la même manière que l' $\alpha$ Chy, pour une gamme de concentration allant de 0,1 à 1 g/L. Sa relation entre concentration et absorbance à 208 nm est :

$$D.O. = 2,3768.C_{Lys} \quad (R^2 = 0,9998) \quad (\text{III.4})$$

avec  $C_{Lys}$  exprimée en g/L.

## III.2.2 Analyse par HPLC-SEC

La chromatographie d'exclusion de taille (ou *Size Exclusion Chromatography*) est une technique de séparation de molécules basée sur la différence de taille. Cette technique permet donc de séparer les protéines contenues dans une solution en fonction de leur taille et donc de leur masse moléculaire; il est ainsi possible de les identifier, à condition de connaître au préalable la masse moléculaire des différentes protéines et de faire un étalonnage de la colonne. Lors du passage d'une solution dans une colonne remplie d'un gel approprié, les molécules de grande taille vont être exclues des particules et sortiront les premières à un volume d'élution  $V_0$ , soit le volume mort de la colonne. A l'inverse, les plus petites molécules vont pénétrer à l'intérieur des pores du gel et sortiront les dernières à un volume d'élution  $V_T$ , soit le volume total du liquide dans la colonne (intra- et inter-particulaire). En sortie de colonne, la concentration en protéine est monitorée via un détecteur UV en ligne. Ainsi, le signal obtenu permet d'identifier les volumes de rétention de chacune des protéines de la solution, les volumes dépendant directement de leur poids moléculaire. En plus de permettre l'identification de protéines, cette technique permet également d'en déterminer la quantité en mesurant la surface des pics après étalonnage.

## DISPOSITIF EXPÉRIMENTAL ET PROTOCOLE

Lors des expériences, une colonne Superdex 75 Increase 10/300 GL (GE Healthcare) a été utilisée et connectée à un système Äkta Purifier (GE Healthcare), selon le schéma de principe représenté sur la Figure III.1. Le système permet l'automatisation des mesures et possède deux pompes permettant de régler finement le débit, un injecteur pour introduire la solution à analyser, un capteur de pression, un conductimètre en ligne ainsi qu'un détecteur UV. Le système est connecté à un ordinateur et l'acquisition des données se fait via le logiciel Äkta UNICORN 5.11.

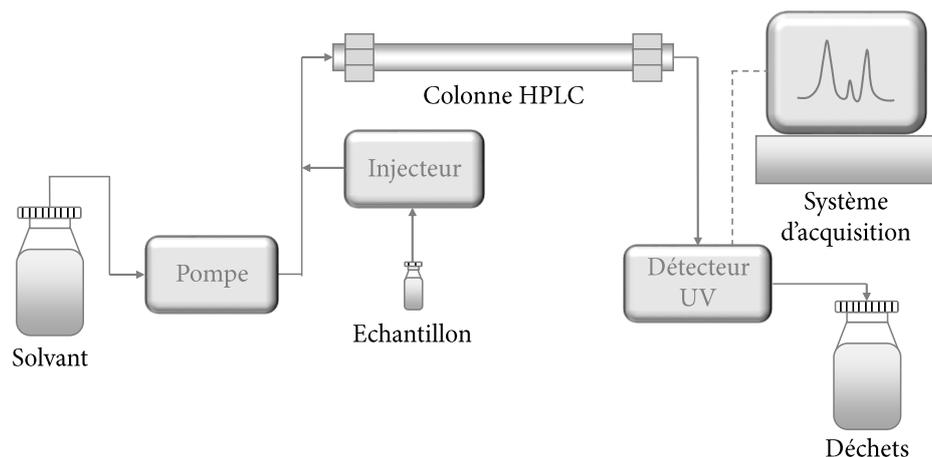


FIGURE III.1 – Schéma de principe d'un système analytique HPLC-SEC.

Lors d'une analyse HPLC-SEC, la solution à analyser est injectée manuellement à l'aide d'une seringue dans une boucle d'injection (100  $\mu\text{L}$ ), qui est ensuite transportée dans la colonne grâce à un débit de solution tampon citrate 50 mM pH 5 (phase mobile), fixé à 0,8 mL/min. La colonne est préalablement équilibrée avec cette même solution tampon, qui a été filtrée sur Büchner avec des membranes en cellulose (seuil de coupure de 0,45  $\mu\text{m}$ ). De même, tous les échantillons sont préalablement filtrés avec des filtres seringue à base de cellulose régénérée (seuil de coupure de 0,45  $\mu\text{m}$ ) afin de ne pas obstruer la colonne avec des impuretés. Afin de pouvoir identifier les protéines sur les chromatogrammes, il est également nécessaire de réaliser au préalable une courbe d'étalonnage. Pour cela, une solution de plusieurs protéines globulaires de masse moléculaires connues est injectée dans la colonne, et le volume d'élution pour chacune d'elles est relevé. La courbe d'étalonnage est présentée en Figure III.2 et permet ainsi de connaître le volume d'élution d'une protéine quelconque à partir de sa masse moléculaire.

#### QUANTIFICATION DE L' $\alpha$ -CHYMOTRYPSINE

Les solutions d' $\alpha\text{Chy}$  précédemment préparées pour la courbe de calibration UV-vis, de concentrations connues, sont réutilisées afin de réaliser la courbe de calibration de l'HPLC-SEC. Selon la Figure III.2, le volume d'élution de l' $\alpha\text{Chy}$  est de 12,7 mL. Pour chaque échantillon, l'aire du pic est mesurée grâce au logiciel UNICORN et la concentration en protéine en fonction de l'aire du pic peut être décrite selon la relation :

$$\text{Aire} = 200,74 \cdot C_{\alpha\text{Chy}} \quad (R^2 = 0,9996) \quad (\text{III.5})$$

avec  $C_{\alpha\text{Chy}}$  exprimée en g/L.

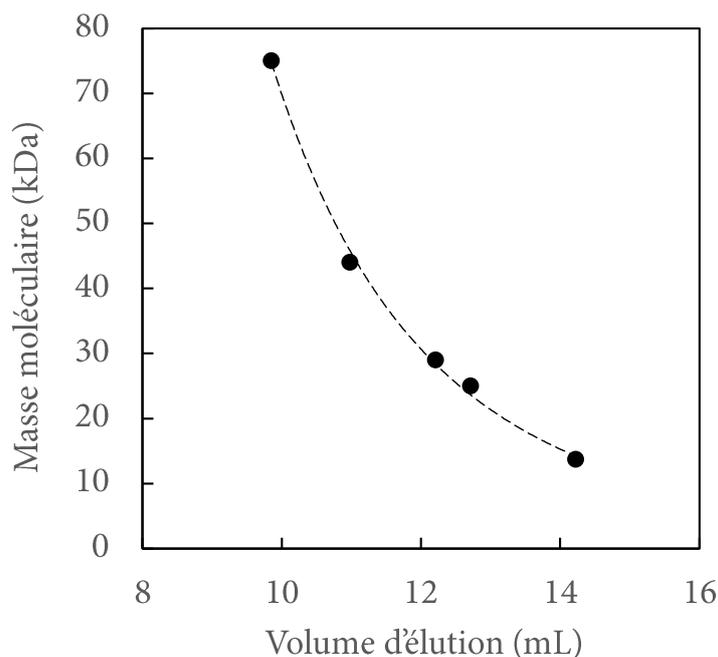


FIGURE III.2 – Courbe d'étalonnage de la colonne Superdex 75 Increase 10/300 GL, représentant la masse moléculaire des protéines en fonction de leur volume d'élution.

Comme expliqué précédemment dans le paragraphe III.1.2.1, l' $\alpha$ Chy, comme la plupart des protéines, peut former un dimère dans des conditions particulières de pH, de température et/ou de force ionique. Dans cette étude, le pH de travail sera fixé à 5, qui est un pH suffisamment éloigné du pI de la protéine, mais qui reste différent du pH de dimérisation. Afin de s'assurer que l' $\alpha$ Chy est bien présente majoritairement sous sa forme monomérique lors des expériences, des analyses HPLC-SEC préalables ont été réalisées.

Le dimère de l' $\alpha$ Chy possède une masse moléculaire de 51,2 kDa. Ainsi, le dimère peut être identifié par analyse HPLC-SEC d'un échantillon dans les conditions opératoires requises. Selon la courbe d'étalonnage tracée sur la Figure III.2, une molécule d'environ 50 kDa devrait être éluee à un volume égal à 10,8 mL, alors que la forme monomérique de l' $\alpha$ Chy est éluee à un volume de 12,7 mL.

La Figure III.3 représente les chromatogrammes issus des analyses HPLC-SEC de deux solutions d' $\alpha$ Chy : (i) la première solution est composée d'un tampon citrate à pH 5 à laquelle est ajouté 50 mM de NaCl et (ii) la deuxième solution contient 200 mM de NaCl. Les deux solutions contiennent la même quantité de protéine (environ 1 g/L). Sur ces deux chromatogrammes, il y a un grand pic dont le volume d'élution correspond bien au monomère de l' $\alpha$ Chy. Il faut également noter qu'il existe un léger décalage entre les deux pics : en effet, il a été montré (Hong et Fountain, 2011) que la présence de sel en solution peut modifier le comportement des protéines et ainsi conduire à différents temps de rétention au sein de la colonne. Il est visible ici que plus la force ionique augmente et plus le volume d'élution augmente également. Dans des cas d'analyses de solutions complexes,

afin d'identifier correctement les protéines, il faudrait réaliser une courbe d'étalonnage avec une force ionique comparable à celle de la solution à analyser. Cependant, dans notre cas, cela n'est pas nécessaire car les volumes restent proches et la solution est assez pure pour que l'identification des protéines soit possible.

Sur le deuxième chromatogramme de la Figure III.3 ( $C_{NaCl} = 200$  mM), un petit pic apparaît pour un volume d'élution légèrement inférieur à 11 mL : ce pic pourrait correspondre au dimère de l' $\alpha$ Chy. En effet, cette protéine est le seul composé en solution pouvant absorber les UV à 280 nm. La présence de pic confirme probablement la dimérisation de l' $\alpha$ Chy. En revanche, la quantité de dimère en solution est très largement inférieure à la quantité de monomère, la différence relative entre les aires de ces pics étant seulement de 2 %. Dans les expériences, la concentration la plus forte utilisée lors de la mesure des isothermes sera de 100 mM NaCl. Il n'est donc pas aberrant de considérer que l' $\alpha$ Chy est présente sous sa forme monomérique en solution, car dans ces conditions de pH et de force ionique, la quantité de dimère est négligeable. En conclusion, seule la forme monomérique de l' $\alpha$ Chy sera considérée lors des mesures des isothermes.

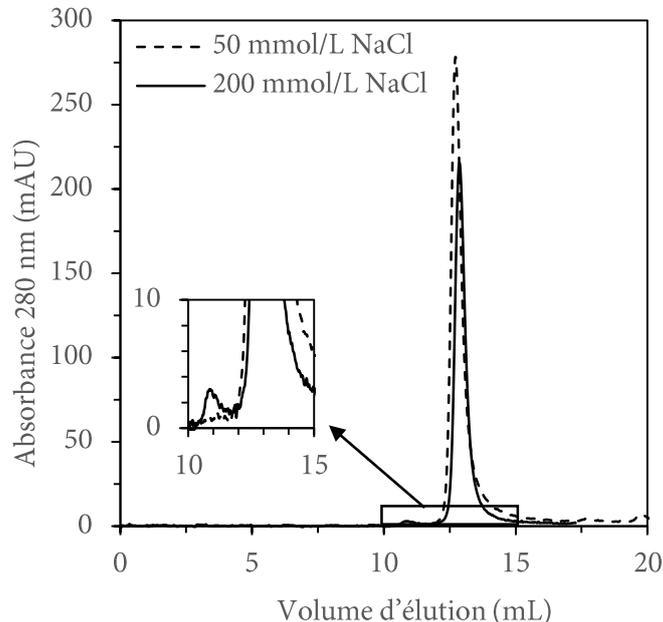


FIGURE III.3 – Chromatogrammes HPLC-SEC de deux solutions d' $\alpha$ -chymotrypsine à pH 5 et à des concentrations en NaCl de 50 et 200 mmol/L.

## QUANTIFICATION DU LYSOZYME

Le lysozyme a été quantifié par HPLC-SEC de la même manière que l' $\alpha$ -chymotrypsine. Son volume d'éluion est égal à 14,2 selon la Figure III.2. Sa relation entre concentration et aire du pic est :

$$\text{Aire} = 277,61.C_{Lys} \quad (R^2 = 0,9967) \quad (\text{III.6})$$

avec  $C_{Lys}$  exprimée en g/L.

## ANALYSES EN SYSTÈME BI-PROTÉINES

En système multi-constituants, il est nécessaire de choisir une méthode permettant la différenciation des protéines en solution, au préalable à leur quantification. Une méthode comme la spectroscopie UV n'est donc pas applicable car elle ne permet la mesure que d'une quantité globale de protéine sans aucune distinction. La méthode analytique la plus utilisée dans des cas de solutions multi-protéines est l'HPLC-SEC, car, comme déjà présenté, c'est également une méthode séparative.

La séparation étant basée sur la différence de taille entre les protéines, il faut s'assurer en premier lieu que les pics soient bien résolus et bien séparés afin de pouvoir correctement les analyser, le lysozyme et l' $\alpha$ -chymotrypsine possédant des masses moléculaires plutôt proches (cf. Tableau III.2). La Figure III.4 montre un chromatogramme issu d'une analyse d'un échantillon contenant environ 2 g/L de chacune des deux protéines. Les deux pics sont bien séparés et la concentration de chacune des protéines peut donc être estimée selon les équations III.5 et III.6, déterminées selon une calibration effectuée à partir de la solution mère contenant 2 g/L de chaque protéine. Le facteur de résolution  $R_{HPLC}$  peut être calculé selon l'équation III.7 et est égal à 1,8.

$$R_{HPLC} = 2 \frac{t_{r2} - t_{r1}}{\omega_1 + \omega_2} \quad (\text{III.7})$$

avec  $t_{ri}$  les temps (ou volumes) de rétention de chaque protéine  $i$  et  $\omega_i$  les largeurs des pics. En général, les pics sont considérés comme bien résolus lorsque  $R_{HPLC}$  est supérieur à 1,5, ce qui est le cas ici. Toutefois, même si les pics sont bien résolus, le pic correspondant au lysozyme présente une traînée et de petits pics apparaissent à des volumes d'éluion proche du volume total de la colonne. De ce fait, il est un peu plus compliqué de déterminer correctement la concentration du lysozyme que de l' $\alpha$ Chy. Afin de permettre une bonne reproductibilité des analyses, la ligne de base pour le calcul des intégrales des pics est

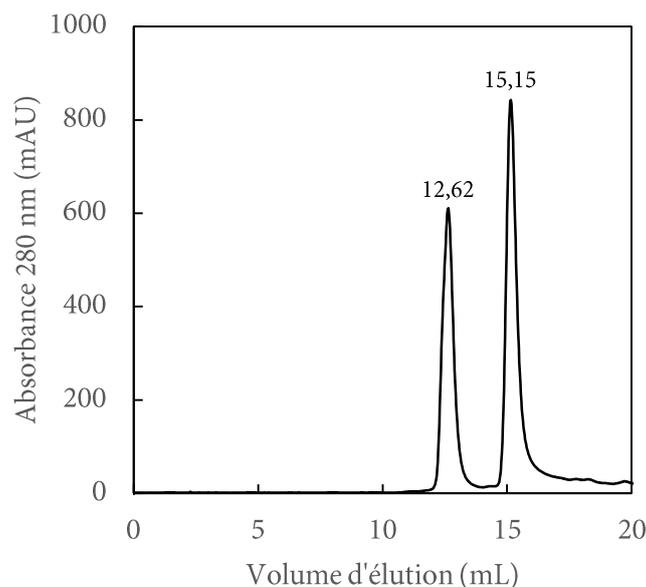


FIGURE III.4 – Chromatogramme HPLC-SEC d'une solution contenant 2 g/L de lysozyme ( $V_{\text{élution}}=15,15$  mL) et d' $\alpha$ -chymotrypsine ( $V_{\text{élution}}=12,62$  mL).

remise à zéro à chaque fois et les bornes pour l'interprétation du pic sont fixées à 14,66 et 16,90 mL, imposées par les caractéristiques de la colonne utilisée.

### III.2.3 Activité enzymatique

Les enzymes peuvent être caractérisées grâce à leur activité, c'est-à-dire leur capacité à catalyser une réaction. Mesurer l'activité enzymatique permet non seulement de vérifier si une enzyme est toujours active, mais également d'en estimer la quantité. Cela permet entre autres de quantifier la protéine sous sa forme native. Cette activité est exprimée en unité arbitraire notée  $U$  qui correspond à la quantité nécessaire d'enzyme pour convertir 1  $\mu\text{mol}$  de substrat en une minute. Il existe des conditions optimales de pH et de température pour lesquelles l'activité est la plus élevée. Ce sont dans ces conditions que la mesure d'activité enzymatique se déroule.

En pratique, la cinétique de la réaction suit la loi de Michaelis et Menten, qui exprime la vitesse stationnaire initiale de la réaction en fonction de la quantité de substrat initiale (voir Figure III.5.a). La constante  $K_M$  correspond à la quantité de substrat pour laquelle la vitesse est égale à la moitié de la vitesse maximale. Afin de calculer l'activité enzymatique, le but sera de mesurer la vitesse maximale de la réaction. Celle-ci est atteinte lorsque l'enzyme est en présence d'une concentration en substrat dite saturante (concentration très grande pour laquelle  $[S_0] \gg K_M$ ). Ainsi, dans des conditions de pH, température et  $[S_0]$  optimales, la méthode utilisée ici (méthode cinétique) consiste à suivre l'évolution de

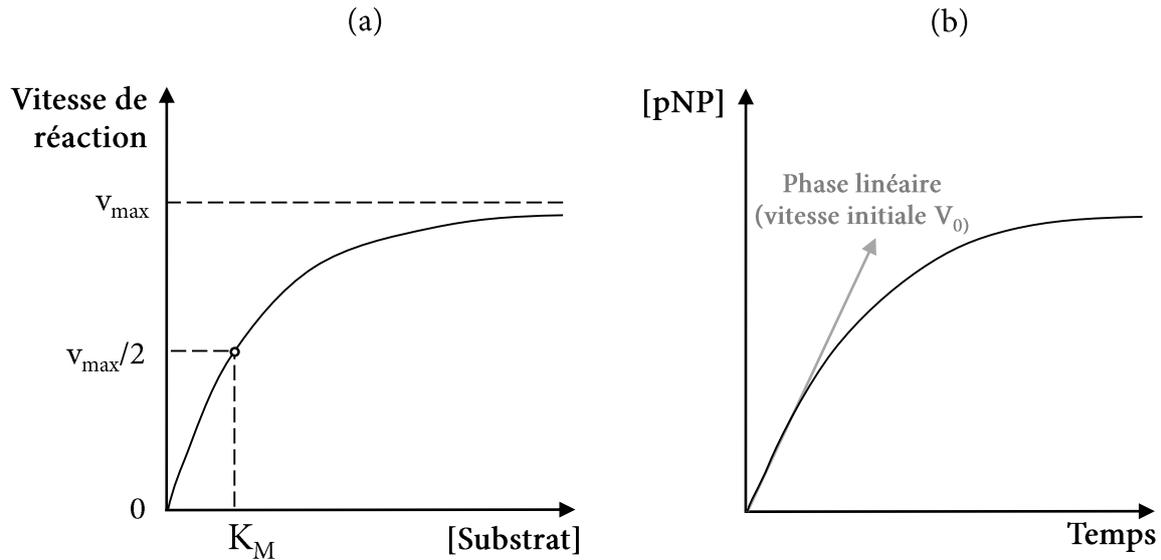


FIGURE III.5 – (a) Représentation de la courbe de Michaelis-Menten et (b) courbe de l'évolution de la concentration de para-nitrophenol, représentative de l'évolution de la réaction, en fonction du temps.

la réaction au cours du temps (voir Figure III.5.b). La vitesse maximale est ainsi mesurée lors des premières minutes de la réaction, dans la zone linéaire de la courbe.

Dans la littérature, la mesure de l'activité de l' $\alpha$ -chymotrypsine a déjà été étudiée (Abuin et al., 2011; Verma et Ghosh, 2013) et les conditions optimales de pH, température et  $[S_0]$  ont déjà été définies. Le substrat utilisé dans ce protocole est le para-nitrophenyl acetate (pNPA), qui possède l'avantage de former du para-nitrophenol (pNP). En effet, le pNP est de couleur jaune et sa concentration peut aisément être mesurée directement par spectrophotométrie visible à 405 nm. Sa courbe d'étalonnage est préalablement établie et son coefficient d'extinction est égal à  $14200 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$ , soit :

$$D.O. = 14200 \cdot C_{pNP} \quad (R^2 = 0,9994) \quad (\text{III.8})$$

avec  $C_{pNP}$  exprimée en mol/L.

L'activité de l' $\alpha$ -chymotrypsine a été mesurée en suivant la conversion du substrat pendant 5 minutes à  $20^\circ\text{C}$  à partir du protocole de Verma et Ghosh (2013). 20  $\mu\text{L}$  d'une solution d'enzyme a été mélangé avec 290  $\mu\text{L}$  d'une solution de pNPA (7,3 mg dans 10 mL d'éthanol pur) et 2,69 mL de solution tampon Tris-HCl (100 mM, pH 7,75). La production de pNP a été monitorée en mesurant l'absorbance à 405 nm dans une cuve macroscopique et à l'aide d'un spectrophotomètre JASCO V630 UV-vis. Afin d'exprimer l'activité en fonction de la quantité d'enzyme ( $\text{U}/\text{g}_{\alpha\text{Chy}}$ ), la concentration de la solution

d'enzyme a été préalablement mesurée par la méthode spectroscopie UV. Au final, en tenant compte de sa définition, l'activité enzymatique se calcule selon :

$$A = \frac{v_{max} \times V_{MR} \times 1000}{14,2 \times V_E \times C_{\alpha Chy}} \quad (\text{III.9})$$

avec :

- $A$  l'activité enzymatique en U/g $_{\alpha Chy}$
- $v_{max}$  la vitesse maximale de la réaction en  $\text{min}^{-1}$
- $V_{MR}$  le volume du milieu réactionnel en L
- $V_E$  le volume de solution d'enzyme en L
- $C_{\alpha Chy}$  la concentration d'enzyme en solution en g $_{\alpha Chy}$ /L

### III.3 Détermination des isothermes

#### III.3.1 Expériences en *batch*

##### III.3.1.1 Protocole expérimental

Une isotherme d'adsorption permet de décrire l'état d'équilibre d'un soluté en présence d'un solide et sa répartition entre la phase liquide et la phase solide (ici, la résine échangeuse d'ions). Ce protocole décrit la mesure des isothermes au moyen d'expériences en système fermé dit en *batch*.

Une masse connue de résine conditionnée  $m_R$  est introduite dans un tube, puis mise en contact avec une solution de protéine, de volume  $V$  et de concentration  $C_0$ . En général, plusieurs tubes sont préparés en parallèle, chacun contenant des ratios liquide/solide différents. Les essais sont réalisés en duplicata (ou triplicata). Au préalable, la résine (initialement stockée dans de l'éthanol à 20%) est rincée à l'eau ultra-pure puis équilibrée avec le tampon. La résine est ensuite filtrée sur Büchner à l'aide d'une pompe à vide avant d'être utilisée dans les tubes. L'humidité de la résine est mesurée à ce moment-là (cf paragraphe III.1.1). La solution de protéine est préparée à partir de protéine purifiée sous forme de poudre et dissoute dans du tampon. Cette solution est également filtrée sur Büchner, sur membrane avec des pores de 0,45  $\mu\text{m}$ . De manière générale, toutes les solutions en contact avec la résine sont filtrées avant utilisation.

Une fois la solution introduite dans les tubes, ces derniers sont mis à l'agitation pendant une certaine durée (entre 2 et 8h selon le système) afin d'atteindre l'état d'équilibre. Une fois cet état atteint, le surnageant est prélevé et la concentration en protéine  $C$  est mesurée

par spectrophotométrie UV à 280 nm. Ainsi, la quantité de protéine fixée sur la résine  $q$  peut être calculée par bilan de matière :

$$q = (C_0 - C) \frac{V}{V_R} \quad (\text{III.10})$$

avec  $V_R$  le volume de la résine, calculé à partir de la masse volumique de la résine humide égale à 1,131 g/mL (Pyle, 2012),  $C_0$  la concentration en protéine initiale et  $V$  le volume de solution.

Ce protocole peut également être appliqué pour un système multi-protéines. Dans ce cas, la solution introduite dans les tubes est préparée à partir d'un mélange des protéines étudiées, et dont la concentration de chacune sera préalablement mesurée. La manipulation reste identique, il n'y a que la méthode d'analyse du surnageant qui change. En effet, une analyse HPLC-SEC sera privilégiée car elle permet de séparer les molécules avant leur quantification.

### III.3.1.2 Isothermes de l' $\alpha$ -chymotrypsine obtenues en *batch* à pH 5

Les isothermes d'adsorption de l' $\alpha$ Chy à pH 5 ont tout d'abord été réalisées en *batch*. Lors de cette expérience, une seule force ionique a été étudiée, soit l'équivalent d'une solution tampon citrate 50 mM sans ajout de NaCl ( $I = 0,22$  M). La Figure III.6 montre deux isothermes, réalisées dans des conditions similaires à quelques jours d'intervalle. Seule la concentration initiale en protéine est différente car il est difficile d'obtenir des solutions de départ avec exactement les mêmes concentrations. Pour le reste, les expériences ont été réalisées selon le protocole décrit dans le paragraphe III.3.1 et dans les conditions décrites dans le Tableau III.3.

TABLEAU III.3 – Conditions opératoires pour la détermination des isothermes d'adsorption de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF par la méthode en *batch*, à  $T = 21^\circ\text{C}$ , pH 5 (tampon citrate 50 mM) et  $C_{NaCl} = 0$  mM.

Conditions	Essai n°1	Essai n°2
Concentration initiale protéine	0,85 g/L	1,15 g/L
Volume de solution dans chaque tube	5 mL	5 mL
Masse de résine dans chaque tube	0,01 à 0,50 g (humide)	0,01 à 0,65 g (humide)
Ratio protéine/résine	0,009 à 0,425	0,012 à 0,575

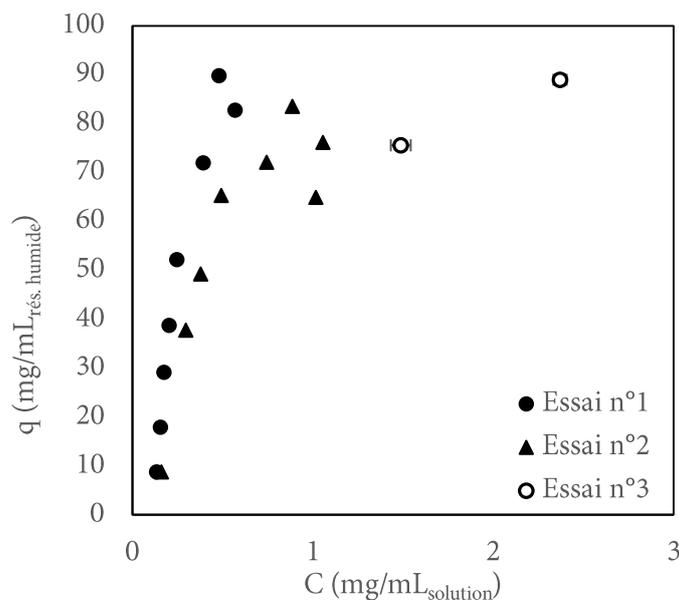


FIGURE III.6 – Isothermes d'adsorption de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF obtenues par la méthode en batch, à pH 5 (tampon citrate 50 mM) et  $C_{NaCl} = 0$  mM.

Malgré des conditions opératoires similaires, les deux isothermes présentent un léger décalage alors qu'elles devraient théoriquement se superposer. Ce résultat peut être expliqué de plusieurs façons, la première étant les incertitudes de mesure. En effet, lors de telles manipulations, les masses de résine introduites dans les tubes peuvent être extrêmement faibles (environ 0,01 g). Malgré l'utilisation d'une balance de précision (à 0,0001 g), une faible valeur mesurée induit une incertitude très grande et peut impacter grandement le résultat final. A ces incertitudes sur la masse peuvent s'ajouter également les incertitudes de mesure liées à la lecture par spectrophotométrie UV, qui sont difficilement quantifiables, ou encore à la lecture des volumes lors de la préparation des solutions. De plus, la quantité de protéine adsorbée  $q$  est exprimée par volume de résine, ce qui nécessite la conversion de la masse de résine en volume. Or, si la masse de résine sèche ne varie pas, le volume peut varier en fonction de la teneur en eau contenue dans celle-ci. Ainsi, l'humidité de la résine a été mesurée à chaque expérience et les isothermes ont également été étudiées en prenant en compte  $q$  exprimé par masse de résine sèche. Cependant, le même décalage est observé entre les deux isothermes, ce qui tend à confirmer qu'il existe des incertitudes de mesures difficilement quantifiables lors de ces expériences en batch, et qui sont indépendantes de l'incertitude sur le volume de la résine.

Il faut également noter que les points des isothermes pour une concentration supérieure à 0,4 mg/mL environ correspondent aux échantillons possédant de très faibles masses de résine. Or, ce sont bien sur ces points-là pour lesquels il y a une forte dispersion des résultats, ce qui confirme le fort impact des incertitudes de mesure. A l'inverse, pour des concentrations inférieures à 0,4 mg/mL, les points semblent plus reproductibles d'une

expérience à l'autre. Il est cependant important de souligner les limites que peut imposer l'utilisation de la méthode en *batch*. Lors des expériences, dans chaque échantillon (ou tube) contenant des ratios liquide-solide différents, un état d'équilibre est atteint. Or, il peut arriver que les conditions finales ne soient pas exactement les mêmes dans chaque tube, et en particulier la force ionique qui peut varier selon la quantité d'ions échangée. De ce fait, les points mesurés, bien que représentatif d'un état d'équilibre, n'appartiennent pas forcément à la même isotherme car les conditions de chaque échantillon ne sont pas strictement identiques. Il peut également arriver que des variations de pH soient observées, mais ce phénomène reste très limité grâce à l'utilisation de solutions tampons.

Indépendamment de ces isothermes, des essais en batch ont été réalisés dans les mêmes conditions avec deux concentrations initiales bien plus grandes afin de vérifier si un plateau était atteint (comme dans le cas d'une isotherme de Langmuir). Il faut toutefois noter que les masses de résine dans les tubes restent très petites afin de pouvoir augmenter le ratio solide-liquide et ainsi représenter les derniers points de l'isotherme. Les résultats obtenus sont présentés dans le Tableau III.4 où chaque point a été réalisé trois fois, puis moyenné.

TABLEAU III.4 – Expériences en batch correspondant à l'essai n°3 : points de l'isotherme à hautes concentrations initiales en protéine, à pH 5 (tampon citrate 50 mM) et  $C_{NaCl} = 0$  mM.

<i>Ratio protéine/résine</i>	$C_0 = 1,71$ g/L		$C_0 = 2,60$ g/L	
	<i>0,53 (moyenne)</i>		<i>0,92 (moyenne)</i>	
	C (g/L)	q (g/L)	C (g/L)	q (g/L)
Répétition n°1	1,55	74,84	2,40	87,70
Répétition n°2	1,46	76,02	2,38	89,29
Répétition n°3	1,45	75,81	2,33	89,89
<b>Moyenne</b>	<b>1,49</b>	<b>75,56</b>	<b>2,37</b>	<b>88,96</b>
Ecart type	0,06	0,63	0,04	1,13

Si l'on représente ces deux points sur le même graphique que les isothermes (Figure III.6, représentés par des cercles vides), il semblerait qu'il n'y ait plus de forte évolution de la masse adsorbée et que l'isotherme présente bien une forme favorable de type Langmuir. Cela signifie en pratique que la résine est saturée et ne dispose plus de sites disponibles pour l'adsorption car malgré l'ajout de protéine en solution, la quantité de protéine au sein de l'adsorbant n'augmente que très peu. Ainsi, au regard de ces isothermes, la capacité

maximale de la résine est de l'ordre de 90 mg/mL, pour l' $\alpha$ Chy et dans ces conditions de travail.

En conclusion, l' $\alpha$ Chy présente une bonne affinité avec la résine SP Sepharose FF à pH 5, les isothermes étant de type favorable. En revanche, les incertitudes de mesure liées à la méthode induisent une forte dispersion dans les résultats et les différents essais montrent que les résultats ne sont pas très reproductibles. Ainsi, la seconde méthode dite *en colonne* a été mise en place afin de déterminer les isothermes d'adsorption, qui permet de représenter plus précisément l'état d'équilibre entre la protéine et la résine.

## III.3.2 Expériences en colonne

### III.3.2.1 Protocole expérimental

Une autre méthode pour la détermination d'isothermes d'échange consiste à travailler en système ouvert sur colonne (lit fixe) en déterminant des courbes de percées. Lorsqu'une colonne est alimentée en continu par une solution de protéine, il est possible de suivre l'évolution de la concentration en sortie de colonne en fonction du temps (ou du volume passé) : la courbe obtenue est appelée courbe de percée (voir Figure III.7). Elle représente le front de concentration de la protéine dans la colonne et permet de voir le moment de la percée pour lequel la résine commence à ne plus retenir totalement la protéine.

En pratique, lors d'une étape de purification, l'alimentation est stoppée lorsque la concentration en sortie atteint environ 5 à 10% de la concentration d'alimentation afin de ne pas perdre trop de protéine. Lors de la mesure d'une courbe de percée, l'alimentation est maintenue jusqu'à l'apparition d'un plateau, correspondant au moment où la concentration en sortie est égale à la concentration d'alimentation. Lorsque cet état est atteint, il y a un état d'équilibre entre la résine et la solution d'alimentation de concentration  $C_0$ . Dans un diagramme  $C/C_0 = f(V)$ , l'aire entre la courbe et la ligne pour laquelle  $C = C_0$  correspond à la fraction de protéine alimentée qui a été retenue et est donc calculée par bilan de matière. Ainsi, une courbe de percée obtenue avec une concentration d'alimentation donnée  $C$  correspond à un point d'équilibre sur une isotherme d'adsorption, la quantité de protéine fixée sur la résine  $q$  pouvant être calculée grâce à la courbe. En réalisant plusieurs courbes de percées à des concentrations d'alimentation différentes, il est donc possible de tracer une isotherme complète.

Ce processus demande plus de temps et de matière première que la méthode en batch, mais il est aussi plus représentatif de l'équilibre. En effet, la résine est en équilibre avec une solution d'alimentation bien connue, et ainsi la concentration en ions et le pH sont bien définis pour tous les points de l'isotherme. La méthode en *batch* est également limitée par le fait que les ratios protéine/solide soient fixés dans chaque tube. De plus,

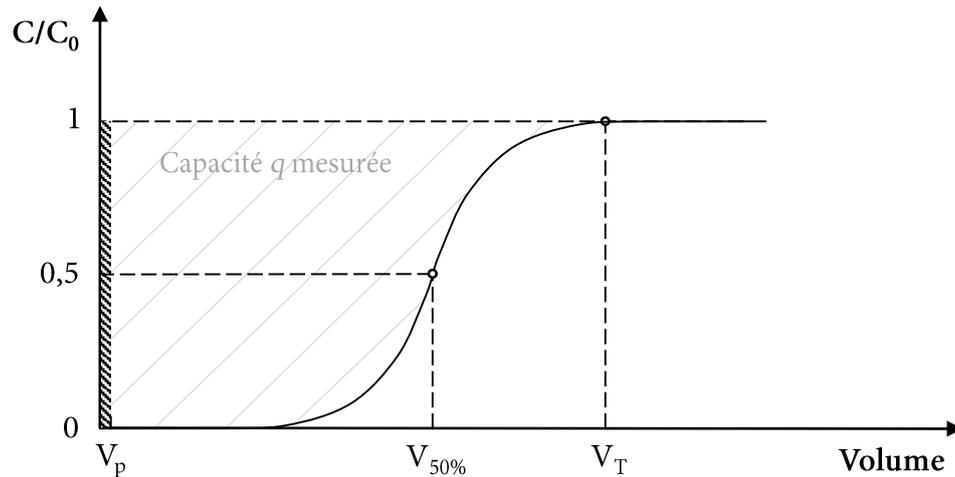


FIGURE III.7 – Représentation d’une courbe de percée, exprimée en fonction du volume passé à travers la colonne.

les conditions finales, et notamment la force ionique, ne sont pas forcément les mêmes du fait de l’échange d’ions. En revanche, cette méthode en colonne est limitée par la mesure de la concentration en protéine, qui se fait en ligne via un détecteur UV. En pratique, il est difficile d’atteindre des concentrations supérieures à 1 g/L car le signal se retrouve saturé. Une solution pourrait être par exemple d’implémenter un système de dilution en sortie de colonne.

Le protocole appliqué lors des expériences est le suivant. Une colonne HiTrap SP FF (GE Healthcare) de 1 mL est connectée au système Äkta Purifier. Une colonne de petit volume est préférable pour ce type de manipulation afin de réduire au maximum la durée de l’expérience et la quantité de protéine à utiliser. La colonne est tout d’abord lavée avec environ 5 volumes de colonne (CV) d’eau ultra-pure, puis équilibrée avec au minimum 3 CV de solution tampon de travail. En parallèle, la solution de protéine est préparée en mélangeant une certaine quantité de protéine commerciale avec la solution tampon de travail, puis filtrée sur Büchner. La solution est ensuite chargée en continu dans la colonne à un débit de 1 mL/min et la concentration en sortie est mesurée via le détecteur UV en ligne. La méthode est gérée grâce au logiciel UNICORN et la manipulation est arrêtée lorsque la concentration en sortie ne varie plus.

Afin d’économiser la matière première, la colonne n’est pas éluée directement après la première charge et une nouvelle solution de protéine plus concentrée que la précédente est injectée. Il s’agit d’un fonctionnement par palier, où la concentration d’alimentation est de plus en plus forte afin de saturer à chaque fois un peu plus la résine. Ainsi, la colonne continue d’être chargée jusqu’au prochain équilibre qui sera plus rapidement atteint. Lors des calculs de bilans matière pour la détermination de la capacité  $q$ , il faudra donc prendre en compte la quantité de protéine fixée lors de la (ou des) précédente(s) charge(s). A la fin

des expériences, la colonne est éluée avec une solution tampon contenant 1 M de chlorure de sodium.

L'analyse des courbes de percée a été réalisée sur le logiciel MS Excel. La méthode des trapèzes a été appliquée afin de déterminer l'aire entre la courbe et la ligne  $C = C_0$ . Une calibration a également été réalisée afin de convertir l'absorbance en concentration de protéines. Finalement, la quantité de protéine retenue  $q$  dans la résine est calculée selon :

$$q = \frac{C_0 \times (V_T - V_p) - \int_{V_p}^{V_T} C.dV}{V_R} \quad (\text{III.11})$$

avec  $q$  en g/L,  $C_0$  la concentration d'alimentation en g/L,  $V_T$  le volume total passé dans la colonne,  $V_p$  le volume poreux et  $V_R$  le volume de lit de résine. Il est en effet nécessaire de ne pas tenir compte du volume poreux  $V_p$  de la colonne dans le calcul de la masse de protéine retenue, car elle reste uniquement dans ce volume sans être retenue par la résine. Le calcul de ce volume fera l'objet du paragraphe suivant.

### III.3.2.2 Distribution de temps de séjour (DTS) de la colonne

Comme indiqué précédemment lors de l'analyse des courbes de percée, il est plus rigoureux de ne pas considérer le volume mort (ou poreux) de la colonne pour le calcul de  $q$ . Pour cela, mais également dans le but de vérifier le bon état de *package* de la colonne, une caractérisation préalable de l'écoulement dans la colonne est nécessaire via une mesure de la distribution des temps de séjour (DTS). Pour caractériser cet écoulement, un traceur est injecté à l'entrée de la colonne et sa réponse en sortie est mesurée. Il doit s'agir d'une molécule facilement mesurable en ligne et inerte, c'est-à-dire qu'elle ne doit pas interagir avec l'absorbant afin de représenter uniquement l'écoulement dans les pores. Il faudrait idéalement que le traceur soit de la même taille que la protéine afin d'être dans les mêmes conditions du point de vue du phénomène d'exclusion de taille. Il peut donc être intéressant de travailler avec la protéine (l' $\alpha$ -chymotrypsine) dans des conditions de non rétention (forte force ionique ou pH égal au pI par exemple).

L'injection lors de la mesure de la DTS peut se faire de deux manières : par impulsion ou par échelon. Ici, la détermination de la DTS a été réalisée à partir d'un échelon positif en entrée de colonne. Comme expliqué plus tard dans la partie III.3.2.3, une courbe de percée de l' $\alpha$ Chy en présence de 400 mM de NaCl a été réalisée et l'éluution a montré que la protéine ne s'était pas fixée sur la résine à cause de la forte salinité. Ainsi, c'est cette courbe de percée qui a été utilisée afin de déterminer la DTS de la colonne, la molécule n'ayant pas interagi avec la résine. Il est également possible d'effectuer un échelon de conductivité car cela est facilement mesurable en sortie de colonne. En revanche, les ions  $\text{Na}^+$  et  $\text{Cl}^-$  sont de taille extrêmement petite comparé à la protéine, leur écoulement peut

donc différer de celle-ci et la détermination de  $V_p$  se verrait ainsi fortement perturbée. Les conditions de débit et le dispositif expérimental sont les mêmes que pour la mesure des courbes de percée.

La courbe obtenue est premièrement normée par  $F(t) = \frac{C(t)}{C_{max}}$ ,  $F(t)$  étant l'intégrale de  $E(t)$  (Villiermaux, 1993).  $E(t)$  est donc obtenue en dérivant numériquement la fonction  $F(t)$  par rapport à  $t$ . Le temps de séjour moyen  $\bar{t}_s$  est égal à :

$$\bar{t}_s = \int_0^{\infty} tE(t)dt \quad (\text{III.12})$$

Finalement, connaissant le débit, le volume poreux peut être calculé. Dans notre cas, l'entrée et la sortie de la colonne sont constitués de capillaires, ce qui signifie que le système est en configuration fermé-fermé à la dispersion et donc que, dans un cas idéal,  $\bar{t}_s = \frac{V_p}{Q} = \tau$ ,  $Q$  étant le débit et  $\tau$  le temps de passage.

La Figure III.8 représente la courbe adimensionnelle  $F(t)$ , qui est la réponse à un échelon positif (Fig. a), et la courbe de distribution du temps de séjour  $E(t)$  (Fig. b) obtenue pour la colonne HiTrap SP FF de 1 mL. Cette DTS expérimentale permet dans un premier temps de constater qu'elle ne présente pas de pic caractéristique d'un court-circuit. En revanche, une légère traînée est observée ce qui peut suggérer du fluide stagnant dans le système. Le temps de passage dans la colonne est d'environ 1 min pour un débit de 1 mL/min, et le temps de séjour moyen calculé est de 1,54 min, soit un volume poreux de  $V_p$  de 1,54 mL. A noter que la différence peut également s'expliquer par le fait qu'il est assez compliqué expérimentalement de maîtriser le moment exact de pénétration de la solution dans la colonne. Malgré les précautions prises, il existe un faible volume dû au capillaire à l'entrée de colonne qui peut expliquer une légère surestimation du temps de séjour moyen.

### III.3.2.3 Isothermes de l' $\alpha$ -chymotrypsine à pH 5

Les isothermes d'échange de l' $\alpha$ -chymotrypsine ont été déterminées en colonne, à pH 5 et quatre forces ioniques différentes. Pour chaque isotherme, six courbes de percées au minimum ont été réalisées, avec des concentrations initiales en protéine allant de 0,1 à 1 g/L. La Figure III.9 représente six courbes de percées réalisées pour l'obtention de l'isotherme de l' $\alpha$ Chy à pH 5, sous tampon citrate 50 mM sans ajout de NaCl. Lors des mesures, l'acquisition des données a été arrêtée lorsque la concentration mesurée en sortie de colonne atteignait la valeur de  $C_0$  (concentration d'alimentation). Dans certains cas, comme par exemple sur la Figure III.9.a, le plateau n'est pas tout à fait atteint au moment de l'arrêt et la concentration continue d'augmenter très lentement, ce qui est dû à une limitation par le transfert. Le choix a été fait d'arrêter un peu plus tôt la manipulation afin d'économiser la matière première, au moment où la concentration en sortie mesurée est très proche de celle en entrée. Dans ces cas-là, la quantité de protéine restante à

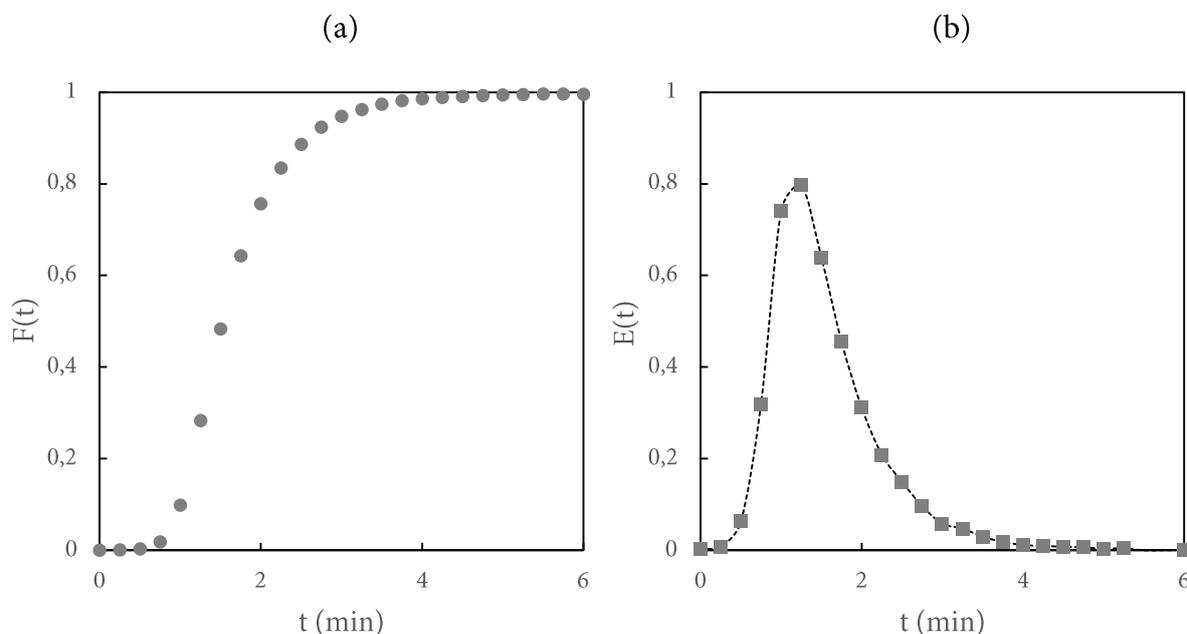


FIGURE III.8 – (a) Courbe adimensionnelle  $F(t)$ , correspondant à la réponse à une stimulation échelon positif sur la colonne HiTrap SP FF et (b) courbe de distribution du temps de séjour  $E(t)$ .

fixer pour saturer la résine a été calculée par extrapolation de la courbe, même si cette quantité s'est toujours avérée négligeable en comparaison de la quantité déjà adsorbée. Par ailleurs, la courbe de percée représentées sur la Figure III.9.f illustre la saturation du signal UV pour des concentrations en protéines atteignant les 1 g/L.

Pour chacune des courbes de percées mesurées, la quantité de protéine adsorbée  $q$  a été calculée selon l'équation III.11. Lors d'une mesure de courbe de percée (dans le cas  $C_0 = 0,95$  g/L), afin de s'assurer que cette quantité calculée correspond bien à la quantité fixée, une élution de la protéine a été réalisée une fois l'étape de charge terminée. Dans cette expérience, la quantité calculée  $q_{calc}$  était égale à 64,3 mg de protéine pour un volume de 1 mL de résine. Ensuite, l'élution a été réalisée en mode isocratique avec une solution tampon contenant 1 M de NaCl, puis l'éluat récupéré a été dosé par spectrophotométrie UV. La quantité de protéine récupérée  $q_{éluat}$  a été de 70 mg. Cette quantité est légèrement supérieure à celle calculée, ce qui peut être expliqué d'une part par la présence potentielle de protéine résiduelle dans les tuyaux qui se retrouve dans l'éluat malgré l'étape de rinçage préalable. D'autre part, la courbe d'étalonnage en UV n'a pas été réalisée dans les mêmes conditions (concentration en NaCl différente), ce qui peut influencer sur la mesure. Malgré cela, les résultats obtenus restent dans le même ordre de grandeur et montrent bien que le calcul de  $q$  par les courbes de percées est représentatif de la quantité de protéine réellement fixée (et récupérée), mais aussi que des erreurs expérimentales de l'ordre de 10 % restent possibles.

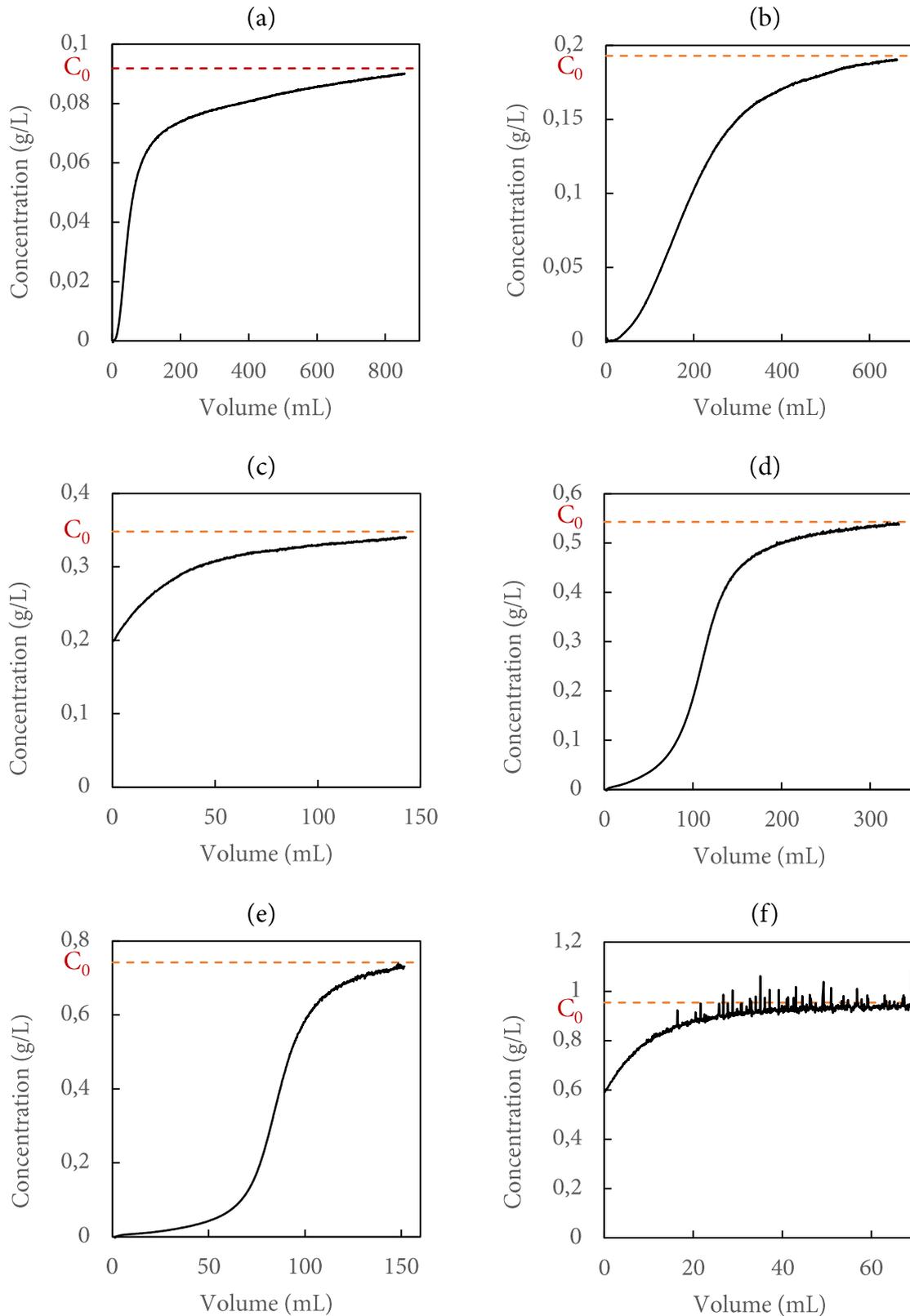


FIGURE III.9 – Courbes de percées de l' $\alpha$ -chymotrypsine sur colonne HiTrap SP FF 1 mL, en condition de pH 5 (tampon citrate 50 mM) sans ajout de sel. (a)  $C_0 = 0,09$  g/L; (b)  $C_0 = 0,19$  g/L; (c)  $C_0 = 0,35$  g/L; (d)  $C_0 = 0,54$  g/L; (e)  $C_0 = 0,74$  g/L; (f)  $C_0 = 0,95$  g/L.

Afin de caractériser l'influence de la force ionique, des concentrations en chlorure de sodium de 25, 50 et 100 mmol/L ont été ajoutées à la première isotherme, qui a été réalisée avec uniquement une solution tampon citrate à 50 mM. Les isothermes sont présentées en Figure III.10. A noter que les concentrations sont exprimées en mmol/L et non en g/L, par cohérence avec l'expression de la constante d'équilibre  $K$  qui est l'un des paramètres du modèle SMA utilisé.

Toutes les isothermes possèdent bien une forme concave favorable, comme pour celles obtenues avec les expériences en batch. Les points des isothermes sont cependant moins dispersés que ceux des isothermes en batch. A noter également que le point  $C = 0,6$  mg/mL de la première isotherme ( $C_{NaCl} = 0$  mM) est en réalité une valeur moyenne issue de deux manipulations identiques, réalisées plusieurs semaines après le reste de l'isotherme. Ce point s'inscrit bien dans la continuité de l'isotherme, ce qui montre que cette méthode est bien plus reproductible que la méthode batch et est moins sensible aux incertitudes de mesure.

Cependant, la forme des isothermes suggèrent que le plateau n'est pas encore atteint et que la résine n'est pas complètement saturée lorsque  $C = 1$  g/L, soit 0,039 mmol/L, ce qui correspond à la concentration maximale pouvant être mesurée avec le dispositif expérimental utilisé. De plus, lors de la préparation des solutions d' $\alpha$ Chy, il est difficile de préparer des solutions de concentrations supérieures à 3 g/L car la protéine a du mal à se solubiliser. Néanmoins, il est tout à fait possible de travailler avec ces données expérimentales, qui représentent bien l'équilibre d'adsorption et peuvent être confrontés aux modèles théoriques.

L'étude de l'effet de la concentration en sel sur l'équilibre montre que plus la force ionique est grande et plus l'affinité protéine-résine décroît. Ce résultat est attendu car l'augmentation de la force ionique implique une forte présence d'ions en solution qui vont entrer en compétition avec la protéine à la surface chromatographique. De ce fait, si la concentration en sel est encore augmentée lors de la mesure des courbes de percée, il ne sera plus possible d'observer une rétention de la protéine sur le support car la compétition protéine-ions sera devenue trop défavorable pour la protéine.

## DÉTERMINATION DES PARAMÈTRES SMA

Afin de déterminer les paramètres de la loi d'action de masse stérique (SMA) expérimentalement, un minimum de trois isothermes à des forces ioniques différentes est fortement recommandé, l'équation présentant trois inconnues. En effet, plus le nombre d'isothermes est grand et plus l'effet de la force ionique sur l'équilibre peut être représenté précisément. La particularité de ce modèle est que les trois paramètres à déterminer, *i.e.* la constante d'équilibre  $K$ , la charge caractéristique  $n$  et le facteur stérique  $\sigma$ , ont des valeurs identiques pour décrire les différentes isothermes, seule la concentration en contre-ions en so-

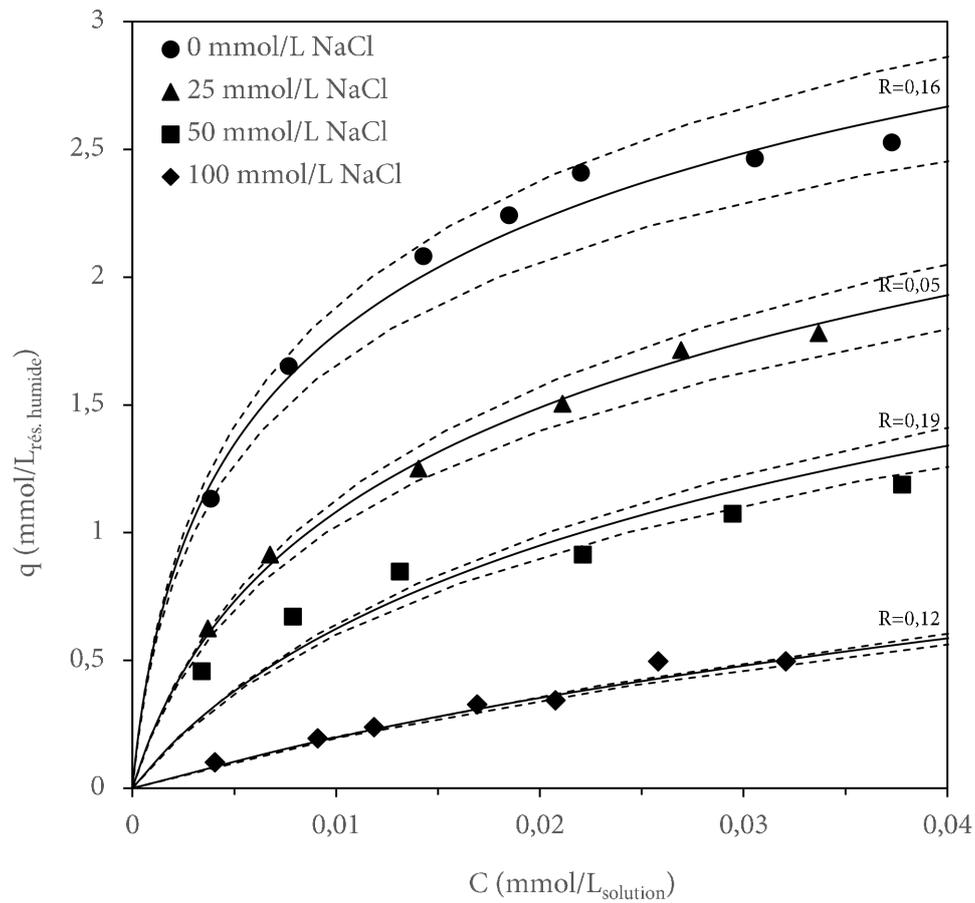


FIGURE III.10 – Isothermes d'adsorption de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF obtenues par la méthode en colonne, à pH 5 (tampon citrate pH 5). Le modèle SMA avec les paramètres ajustés est représenté en ligne continue. Les lignes discontinues représentent l'analyse de sensibilité ( $\pm 10\%$  sur les données expérimentales) du modèle.

lution change d'une isotherme à l'autre. Cependant, considérer  $K$  constante pour toutes les isothermes signifie que ce sont les concentrations qui sont prises en compte et non les activités, l'influence de la force ionique sur le comportement "idéal" de la dissolution est ainsi occultée.

La méthode la plus simple à mettre en place pour la détermination des paramètres SMA est un ajustement de paramètres d'un modèle à des données expérimentales (ou *fitting*). A partir de la série de données  $(C_i; q_i)$ , il est possible d'estimer les valeurs des paramètres du modèle pour lesquelles la fonction se rapproche au mieux des valeurs expérimentales. Pour cela, la méthode des moindres carrés est l'une des approches les plus utilisées. Cette méthode permet de minimiser l'écart entre les données expérimentales et le modèle théorique en calculant un facteur de fiabilité  $R$  (equation III.13) :

$$R = \sqrt{\frac{\sum(C_{i,exp} - C_{i,cal})^2}{\sum C_{i,exp}^2}} \quad (\text{III.13})$$

avec  $C_{exp}$  et  $C_{calc}$  représentant respectivement les concentrations dans la phase liquide expérimentale et calculée via le modèle.

Les paramètres SMA qui correspondent le mieux à ces données expérimentales sont présentés dans la Tableau III.5 et permettent d'obtenir un facteur de fiabilité  $R$  égal à 0,14. Par ailleurs, le facteur de fiabilité a ensuite été recalculé pour chacune des isothermes indépendamment à partir des paramètres ajustés, et les valeurs sont indiquées sur la Figure III.10. Malheureusement, il est assez compliqué d'estimer un intervalle de confiance pour ces paramètres car les incertitudes sur les données expérimentales sont difficilement quantifiables. En revanche, il est possible d'étudier la sensibilité du modèle à des variations sur les données. La Figure III.10 représente les isothermes expérimentales ainsi que le modèle SMA. Les traits continus représentent le modèle calculé avec les paramètres obtenus précédemment, alors que les traits discontinus représentent le modèle avec les paramètres obtenus en appliquant un incrément de  $\pm 10\%$  sur les données expérimentales (ces paramètres sont également présentés dans le Tableau III.5).

L'étude de ces paramètres montre tout d'abord que, comme attendu, la protéine possède une bonne affinité avec le support chromatographique car elle possède en moyenne 5,1 sites de fixation à sa surface. Le facteur stérique quant à lui est plutôt faible, ce qui paraît cohérent compte tenu de la faible taille de la protéine. Ces valeurs sont par ailleurs en accord avec des paramètres publiés dans la littérature (Ladiwala et al., 2005). En revanche, les valeurs des constantes d'équilibre sont très différentes par rapport à celle de la littérature. La constante  $K$  calculée ici n'est pas définie de la même manière que l'ont fait les auteurs de l'article : celle-ci pourrait être calculée comme étant  $K_{SMA}(q_R/C_{sel})^n$  selon Ladiwala et al. (2005). En effet, en reprenant le  $K_{SMA}$  de cet article et en faisant

TABLEAU III.5 – Paramètres de la loi SMA à pH 5, calculés par *fitting* sur données expérimentales (brutes, -10% et +10%).

Paramètres SMA	Fitting avec données exp.	Fitting avec données exp. -10%	Fitting avec données exp. +10%
Charge caractéristique $n$	5,1	5,1	5,1
Facteur stérique $\sigma$	28,3	32,0	25,2
Constante d'équilibre $K$	11	11	11

le calcul ci-dessus, la constante calculée est de 18,7, ce qui correspond mieux à l'ordre de grandeur de la constante que nous avons déterminée.

L'étude de sensibilité montre qu'une variation des données expérimentales influe majoritairement sur l'écart entre les différentes isothermes et non sur leur forme. Ainsi, lors du réajustement des paramètres, c'est uniquement le facteur stérique qui change alors que la charge et la constante restent identiques. Il serait toutefois plus adéquat d'utiliser une autre méthode que celle des moindres carrés afin d'estimer les paramètres SMA. Par exemple, une méthode statistique telle que Monte Carlo permettrait d'obtenir une estimation des paramètres avec leur intervalle de confiance. Cependant, à cause de la non réversibilité de l'équation SMA, cette méthode n'a pas pu être appliquée à notre cas de figure.

#### ELUTION DE L' $\alpha$ -CHYMOTRYPSINE ET ACTIVITÉ ENZYMATIQUE

Lors de l'étape d'élution, l'équilibre est modifié au sein de la résine afin de permettre à la protéine de se décrocher du support. En échange d'ions, la méthode la plus courante et la plus simple à contrôler est d'appliquer un changement de force ionique pour réduire l'affinité protéine-support. Afin de déterminer la concentration en sel minimale nécessaire pour éluer la protéine, un gradient d'élution a été appliqué sur la colonne. Pour cela, deux solutions tampons ont été préparées, l'une contenant uniquement du tampon (tampon citrate 50 mM pH 5, appelé tampon A) et la seconde contenant le tampon A ainsi qu'une concentration de 1 M en NaCl (appelé tampon B). L'élution par gradient a donc été réalisée à l'aide de deux pompes, permettant de délivrer en continu une solution dont la composition varie de 0% à 100% de tampon B. Le chromatogramme obtenu ainsi que l'évolution de la conductivité en sortie de colonne sont représentés en Figure III.11.

La conductivité au point d'élution, qui se mesure lorsque le pic est le plus haut, est de 22,2 mS/cm, ce qui correspond à une concentration en sel d'environ 270 mM. Cependant,

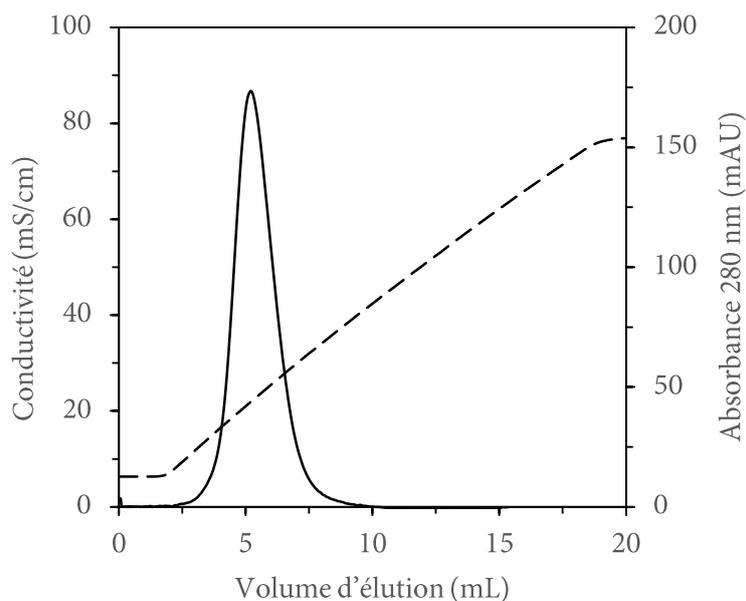


FIGURE III.11 – Gradient d'élution de l' $\alpha$ -chymotrypsine à pH 5. La ligne continue représente l'absorbance à 280 nm et la ligne discontinue la conductivité en sortie de colonne.

la largeur du pic est relativement grande : l'élution débute à des conductivités bien plus faibles et se termine lorsque la conductivité dépasse 42 mS/cm (soit  $C_{NaCl} = 500$  mM). Il faut tout de même noter que cela est majoritairement dû à l'hydrodynamique de la colonne et/ou au transfert de matière. Par sécurité, il vaut mieux se placer à des concentrations en NaCl supérieures ou égales à 500 mM afin d'être certain de pouvoir récupérer toute la protéine. Ce chromatogramme permet également de conclure qu'au delà d'une salinité de 270 mM, la protéine ne devrait pas s'adsorber sur la résine et donc qu'il n'est pas possible de mesurer des isothermes à de telles concentrations.

De ce fait, une expérience de rétention de la protéine a été effectuée sur la colonne à une salinité de 400 mM et la courbe de percée a été tracée (voir Figure III.12). Celle-ci semble montrer que la protéine est retenue en partie par la résine car il est possible de calculer l'aire entre la courbe et la droite d'équation  $C = C_0$ . Or, ce front de concentration, qui représente un retard de la colonne sans aucune rétention de la protéine sur la résine, peut également être dû uniquement à la limitation par le transfert et à l'hydrodynamique de la colonne. Afin de s'assurer que la protéine n'est effectivement pas retenue par la résine, l'élution isocratique (1 M de NaCl) a été réalisée et aucun pic d'élution n'a été identifié. Cela confirme que la protéine ne s'est pas fixée sur la colonne dans ces conditions de haute salinité, ou en tout cas pas en quantité suffisante pour être détectable par ces techniques expérimentales.

Il peut arriver qu'une protéine, lors de l'étape de charge puis d'élution, se retrouve dénaturée du fait de l'échange d'ions et du changement d'environnement. Pour cette raison,

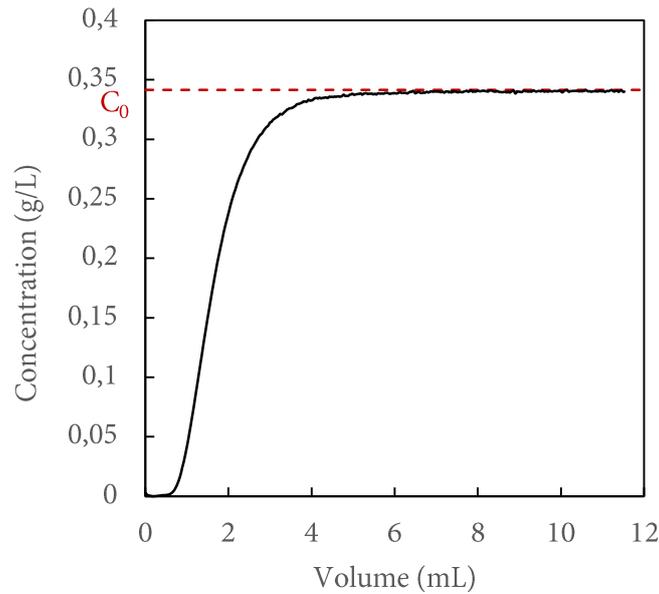


FIGURE III.12 – Courbe de percée de l' $\alpha$ -chymotrypsine sur HiTrap SP FF 1 mL, à pH 5 et  $C_{NaCl} = 400$  mM.

il est intéressant d'étudier l'activité enzymatique d'une solution de protéine avant et après l'échange d'ions, l'activité enzymatique représentant la capacité d'une enzyme à convertir un substrat. Or, si celle-ci subit un changement conformationnel dû au mécanisme d'échange d'ions, son activité enzymatique devrait se retrouver altérée.

Lors d'une expérience de charge puis d'élution, un échantillon de la solution initiale a été analysé, ainsi qu'un échantillon de l'éluat. Afin que les résultats soient comparables, la concentration en protéines de chacun de ces échantillons est mesurée afin de rapporter l'activité enzymatique en fonction de la quantité de protéine. La solution initiale de protéine présentait une activité enzymatique égale à  $93,5$  mU/mg $_{\alpha Chy}$ . Après élution, l'activité enzymatique a été estimée à  $96,0$  mU/mg $_{\alpha Chy}$ . La différence relative entre de ces deux valeurs est très faible ( $< 3\%$ ) et est très probablement majoritairement dues aux incertitudes de mesure. En effet, il n'y a pas de raison physique que l'activité augmente après l'échange. En conclusion, ce résultat suggère que le mécanisme d'échange d'ions n'altère pas la conformation de l' $\alpha Chy$ , et que celle-ci peut-être entièrement récupérée grâce à une élution par changement de force ionique.

#### III.3.2.4 Isothermes de l' $\alpha$ -chymotrypsine à pH 7

Le pH d'une solution est l'un des critères les plus importants en échange d'ions car il conditionne la charge globale d'une protéine. L' $\alpha Chy$  s'adsorbe bien à pH 5 sur la SP Sepharose FF, car elle est globalement chargée positivement à ce pH de travail. Son pI étant de 8,3, plus le pH se rapproche de cette valeur et plus la charge globale de la protéine

se rapproche de la neutralité. Ainsi, à un pH 7, l' $\alpha$ Chy devrait montrer une moins bonne rétention par la résine.

Pour étudier ce phénomène, des isothermes obtenues par la méthode en colonne ont été réalisées à pH 7. Le tampon utilisé est un tampon phosphate comme décrit dans la partie III.1.3. Deux isothermes ont été mesurées, avec des concentrations en NaCl de 0 et 25 mM. Les résultats sont présentés en Figure III.13, où les isothermes à pH 5 sont également représentées à titre de comparaison.

Comme attendu, à ce pH, l' $\alpha$ Chy montre une moins bonne affinité avec la SP Sepharose. Les isothermes avec des concentrations en NaCl supérieures ou égales à 50 mM n'ont pas été réalisées car du fait du pH et de la salinité, la quantité de protéine adsorbée sera très faible et difficilement détectable lors des analyses. De plus, les points des isothermes à pH 7 montrent une plus grande dispersion des résultats, en grande partie due au fait que les quantités mesurées soit plus faibles, ce qui induit de plus grandes incertitudes de mesure.

Les paramètres SMA ont été calculés à pH 7 (voir Tableau III.6). Etant donné que seulement deux isothermes ont été réalisées, il peut s'avérer délicat de réussir à trouver un bon ajustement entre le modèle et les données expérimentales lorsque trois paramètres sont inconnus. Ainsi, une hypothèse de départ a été introduite, la valeur du facteur stérique a été considérée identique à celle obtenue à pH 5, ce paramètre dépendant essentiellement de la forme et de la taille de la protéine qui est inchangée. Il est donc intéressant de voir qu'après *fitting*, la charge caractéristique reste également identique et que le changement vient principalement de la constante d'équilibre, qui est représentative de la force d'interaction. L' $\alpha$ Chy paraît donc avoir le même mode d'interaction avec la résine qu'à pH 5 (même nombre de sites de fixation), mais c'est l'énergie d'interaction en elle-même qui est réduite. Par ailleurs, ce résultat est cohérent avec les travaux de Gadam et al. (1993), qui ont déterminé une charge de l' $\alpha$ -chymotrypsinogène, dont la structure est quasi-identique à celle de l' $\alpha$ Chy (voir paragraphe III.1.2.1), égale à 4,8 à pH 6 (tampon phosphate). Le facteur stérique a quant à lui été estimé égal à 49,2 avec une résine possédant une capacité ionique deux fois supérieure à celle de la SP Sepharose FF.

TABLEAU III.6 – Paramètres de la loi SMA à pH 7 et 5, calculés par fitting sur données expérimentales.

Paramètres SMA	pH 7	pH 5
Charge caractéristique $n$	5,0	5,1
Facteur stérique $\sigma$	28,3	28,3
Constante d'équilibre $K$	0,6	11

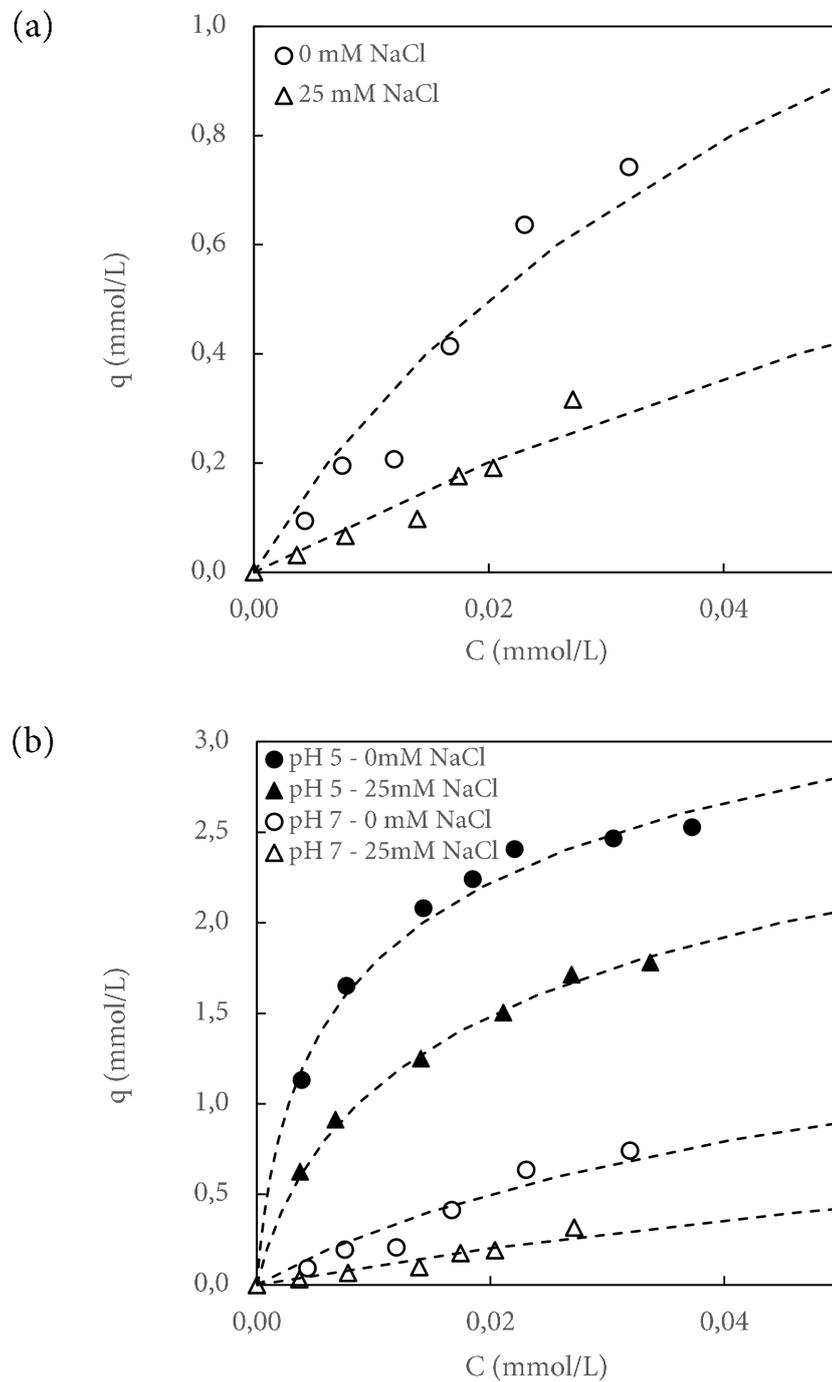


FIGURE III.13 – Isothermes d'adsorption de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF obtenues par la méthode en colonne (a) à pH 7 (tampon phosphate) et (b) à pH 5 (tampon citrate) et 7.

Il faut également prendre en compte que la composition de la solution tampon utilisée lors de ces expériences n'est pas la même qu'à pH 5. En effet, bien que la quantité de contre-ions  $\text{Na}^+$  ait été conservée, l'environnement de la protéine n'est plus exactement le même et cela peut avoir des effets sur l'adsorption. C'est notamment le cas des forces ioniques qui sont différentes ( $I_{\text{citrate}} = 0,22 \text{ M}$  et  $I_{\text{phosphate}} = 0,13 \text{ M}$ ). Il est par ailleurs possible qu'à forces ioniques égales, l'affinité de la protéine pour la résine à pH 7 soit encore plus faible, la force ionique du tampon phosphate étant actuellement deux fois plus faible que celle du tampon citrate.

### III.3.3 Equilibre d'échange d'ions en système multi-protéines

#### CONDITIONS OPÉRATOIRES

Les isothermes en système bi-protéines, comprenant le lysozyme et l' $\alpha$ -chymotrypsine, ont été réalisées à l'aide de la méthode batch. En effet, la méthode en colonne est difficilement applicable car le suivi se fait par mesure de l'absorbance UV, qui n'est pas sélective. En sortie de colonne, seule une valeur globale de concentration en protéines serait mesurée et non la concentration de chacune d'elles. Il serait également possible de récupérer des échantillons en sortie de colonne pour les analyser par HPLC-SEC, mais cette procédure s'avère longue à mettre en place. Ainsi, le choix s'est porté sur la méthode batch, qui nécessite moins de préparation et surtout moins de matière première.

Pour cela, le protocole appliqué est sensiblement le même qu'en système mono-constituant. La principale différence réside dans la solution de protéines initiale introduite dans les tubes, qui ne contient plus une mais deux protéines. Deux essais différents ont été réalisés, dont les conditions opératoires sont présentées dans le Tableau III.7.

Lors de l'essai n°2, de plus grandes masses de résine ont notamment été utilisées afin de diminuer l'incertitude de mesure liée à la pesée de la résine humide et ainsi réduire la dispersion des résultats. De plus, lors de cet essai, deux répétitions de mesure ont été réalisées dans des conditions très similaires afin d'obtenir une valeur moyenne. Les surnageants ont été analysés par HPLC-SEC, le lysozyme et l' $\alpha$ Chy présentant des pics bien résolus sur le chromatogramme (voir Figure III.4) et permettant ainsi leur quantification.

L'isotherme du lysozyme seul a également été réalisée selon les conditions opératoires décrites dans le Tableau III.8, afin de pouvoir identifier son comportement en l'absence de l' $\alpha$ Chy.

TABLEAU III.7 – Conditions opératoires pour la détermination des isothermes d'échange en système bi-protéines ( $\alpha$ -chymotrypsine et lysozyme) sur la résine SP Sepharose FF par la méthode en batch, à  $T = 21^\circ\text{C}$ , pH 5 (tampon citrate 50 mM) et  $C_{NaCl} = 0$  mM.

Conditions	Essai n°1		Essai n°2	
	Lysozyme	$\alpha$ -chymotrypsine	Lysozyme	$\alpha$ -chymotrypsine
Concentration initiale protéine	1,93 g/L	2,01 g/L	1,98 g/L	1,97 g/L
Volume de solution	10 mL		150 mL	
Masse de résine (humide)	0,01 à 0,85 g		0,25 et 0,26 g	
Ratio protéine/résine	0,023 à 1,930	0,024 à 2,010	1,14 et 1,19	1,14 et 1,18

TABLEAU III.8 – Conditions opératoires pour la détermination de l'isotherme d'adsorption du lysozyme sur la SP Sepharose FF par la méthode en batch, à  $T = 21^\circ\text{C}$ , pH 5 (tampon citrate 50 mM) et  $C_{NaCl} = 0$  mM.

Conditions	Essai n°1
Concentration initiale protéine	0,93 g/L
Volume de solution dans chaque tube	10 mL
Masse de résine dans chaque tube	0,02 à 0,84 g (humide)
Ratio protéine/résine	0,011 à 0,361

## RÉSULTATS

Tout d'abord, l'isotherme du lysozyme seul a été tracée et comparée à l'isotherme de l' $\alpha$ Chy seule afin d'identifier leur comportement au contact de la SP Sepharose FF, sans que des phénomènes de compétition protéine-protéine ne soient mis en jeu. Les deux isothermes sont présentées sur la Figure III.14. Il est nécessaire de noter l'allure de la courbe du lysozyme, et plus particulièrement le dernier point qui semble augmenter brusquement. De la même manière que ce qui a été observé pour l' $\alpha$ Chy, il est très probable que les derniers points de l'isotherme mesurée en batch présentent une plus grande dispersion du fait de la grande incertitude sur la masse de résine qui est extrêmement faible. Lorsque ces deux isothermes sont comparées, il apparaît que le lysozyme possède une meilleure

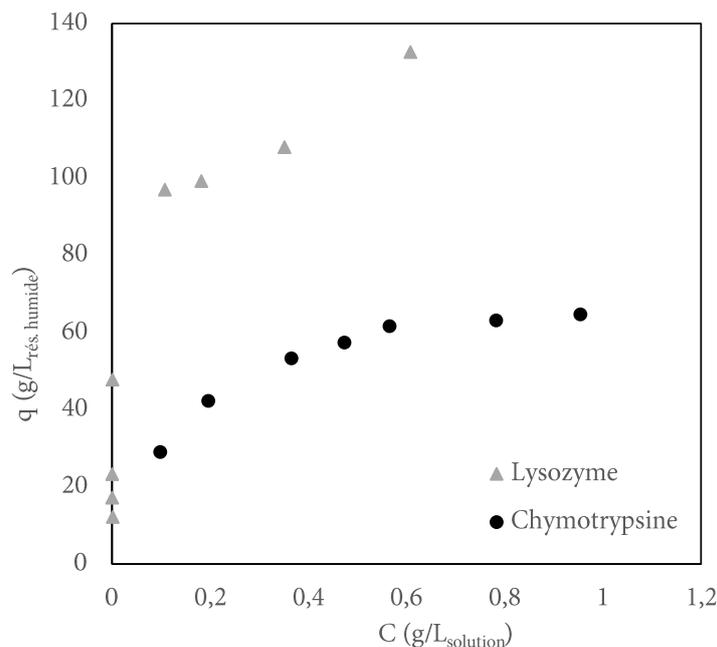


FIGURE III.14 – Isothermes d'adsorption du lysozyme et de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF à pH 5 (tampon citrate 50 mM). Chacune des isothermes a été mesurée indépendamment à partir d'une solution mono-protéine.

affinité avec la résine que l' $\alpha$ Chy de par la forme de l'isotherme, mais également au regard de la capacité maximale atteinte. Ce résultat semble logique pour deux raisons : la première est que le lysozyme est plus petit que l' $\alpha$ Chy et il peut donc diffuser plus facilement et plus profondément à l'intérieur des pores. La deuxième raison est que le point isoélectrique du lysozyme qui est bien plus élevé que celui de l' $\alpha$ Chy (Tableau III.2) et donc plus éloigné du pH de travail. Il est donc très probable que la charge globale du lysozyme soit fortement positive, ce qui favorise les interactions électrostatiques avec la résine.

Les isothermes obtenues à partir du mélange des deux protéines ont été ensuite tracées et étudiées afin d'observer le comportement du système et du couple protéine-protéine. Ces isothermes sont représentées sur la Figure III.15, où les points issus des essais n°1 et 2 (voir Tableau III.7) sont présentés.

La première observation est que le lysozyme semble toujours avoir une meilleure affinité que l' $\alpha$ Chy, en tout cas à faibles concentrations en solution. En effet, une plus grande quantité de lysozyme s'adsorbe sur la résine car à ce moment-là, il est très probable que la compétition entre protéines ne soit pas très forte du fait des faibles concentrations. En revanche, lorsque les concentrations augmentent, la capacité de rétention de l' $\alpha$ Chy augmente brusquement, jusqu'à atteindre des capacités supérieures par rapport à l'isotherme mono-constituant. Quant au lysozyme, sa capacité semble plutôt diminuer. Comme discuté précédemment, ces points de l'isotherme sont déterminés avec des masses de résine

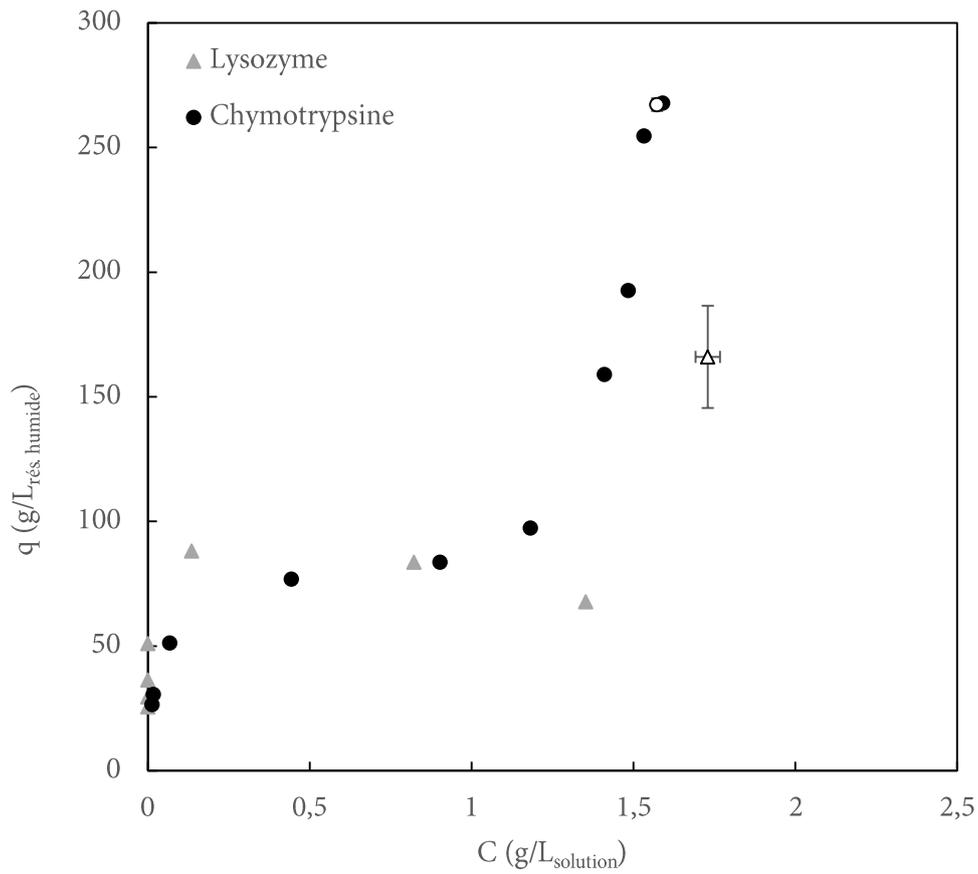


FIGURE III.15 – Isothermes d'adsorption en système bi-protéines du lysozyme et de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF à pH 5 (tampon citrate 50 mM). Les symboles ● et ▲ représentent les valeurs issues de l'essai n° 1 et les symboles ○ et △ représentent l'essai n°2.

extrêmement faibles, et donc sont affectés d'une plus grande incertitude. Cela impacte fortement le calcul de  $q$  par bilan de matière, générant des erreurs plus importantes. C'est donc dans le but de réduire ces incertitudes de mesure que l'essai n°2 a été réalisé. Pour chaque protéine, la moyenne des deux répétitions de mesure est représentée sur la Figure III.15 sous forme de cercle/triangle vides. Les écarts-types sont également tracés, bien qu'ils ne soient pas visibles car trop faibles dans le cas de l' $\alpha$ Chy. Ces résultats permettent dans un premier temps de valider l'allure de l'isotherme de l' $\alpha$ Chy. En revanche, les résultats concernant le lysozyme montrent bien qu'il y a vraisemblablement eu des problèmes de mesures et que l'isotherme ne décroît pas à des concentrations  $C$  supérieures à 1 g/L, comme le suggéraient les premiers résultats. Finalement, les points de l'isotherme restent dans le même ordre de grandeur que ceux déterminés en système mono-protéine. Il faut également prendre en compte qu'en plus de l'incertitude sur les masses de résine, l'analyse en HPLC-SEC s'est avérée plus compliquée que pour l' $\alpha$ Chy (cf. paragraphe III.2.2).

Que ce soit au niveau de la forme de l'isotherme ou même des valeurs de capacité de rétention atteintes, le lysozyme paraît se comporter de la même manière en système mono-constituant et multi-constituants. Ce n'est cependant pas le cas de l' $\alpha$ Chy, qui présente une forme d'isotherme assez particulière. Lorsque les concentrations des deux protéines en solution sont élevées, c'est-à-dire ici à partir de 1 g/L environ, la forme de l'isotherme subit un changement fort d'allure. C'est donc à ce moment-là que les protéines semblent entrer en compétition et interagir, car en deçà de cette valeur, elles ont des comportements similaires à des cas de protéines isolées. En général, des formes d'isothermes de ce type (type VI selon l'IUPAC, voir partie II.3.3.1) sont le plus souvent considérées comme une traduction d'une adsorption en multi-couches. Cependant, le changement brusque de pente de l'isotherme de l' $\alpha$ Chy est uniquement la traduction d'une forte disparition de la protéine en solution, car lors des expériences, seule la concentration  $C$  est directement mesurée,  $q$  est calculé par bilan de matière. L'hypothèse émise lors du calcul de  $q$  est donc que la quantité de protéine qui n'est plus en solution est retenue sur la support chromatographique. Or, il peut également s'agir d'autres phénomènes comme par exemple une agrégation ou une précipitation de la protéine, qui expliqueraient également une diminution de la quantité de protéine solubilisée (Carta et Jungbauer, 2010). En solution, aucune précipitation de la protéine n'a pu être observée à l'œil nu, mais il est également possible que ces phénomènes puissent avoir lieu à l'intérieur de la résine, ce qui, dans ce cas, est difficilement identifiable expérimentalement.

## Conclusion

L' $\alpha$ -chymotrypsine et le lysozyme sont deux protéines modèles, souvent utilisées dans la littérature dans l'étude et le développement de nouvelles méthodes appliquées à la chro-

matographie d'échange d'ions. Dans cette étude, des expériences d'échange d'ions sur une résine cationique, la SP Sepharose FF, ont été réalisées afin d'observer et de mieux comprendre le comportement de ces deux protéines. Dans le cas de l' $\alpha$ -chymotrypsine, elles ont également permis d'estimer les paramètres principaux de la loi d'action de masse stérique (constante d'équilibre, charge caractéristique et facteur stérique) afin de pouvoir dans un second temps les comparer aux valeurs obtenues via les simulations moléculaires. Premièrement, le comportement de l' $\alpha$ Chy seule a été étudié, puis l'impact de la compétition protéine-protéine sur l'échange a également été observé en ajoutant le lysozyme. Enfin, l'influence du pH et de la force ionique a également été discutée.

L' $\alpha$ Chy montre une bonne affinité pour la résine car elle présente des isothermes d'adsorption favorables. La méthode de mesure par courbes de percées a été privilégiée car elle a permis d'obtenir des résultats plus reproductibles et représentatifs de l'état d'équilibre qu'avec la méthode en batch, bien qu'elle soit plus difficile à mettre en œuvre. De plus, la loi SMA a permis de décrire correctement les allures des isothermes après ajustement des paramètres. La charge caractéristique suggère par ailleurs que la protéine possède bien une interaction multi-points, c'est-à-dire que la protéine se fixe sur la résine en occupant plusieurs ligands (5,1 en moyenne). Quant au facteur stérique, qui représente l'ensemble des ligands cachés par la protéine et donc indisponibles pour l'adsorption, sa valeur est relativement faible compte tenu de la petite taille de la protéine. Il est également intéressant de constater que lorsque la solution subit un changement de pH, comme ici à pH 7, ces deux paramètres de la loi SMA ne varient que très peu. En effet, bien que les isothermes montrent une moins bonne affinité, le *fitting* le plus adéquat montre que la charge et le facteur stérique sont les mêmes. En revanche, la constante d'équilibre, représentative de la force d'interaction, est bien plus faible qu'à pH 5.

Des premières expériences en multi-constituants ont également été réalisées via la mesure d'isothermes d'un mélange  $\alpha$ Chy-lysozyme, le lysozyme étant une protéine plus petite que l' $\alpha$ Chy et dont le pI est plus élevé. Ainsi, comme attendu, le lysozyme seul montre une meilleure affinité pour la résine car sa charge globale est plus grande. Cependant, l'étude en mélange bi-protéines révèle un comportement particulier : lorsque les protéines sont présentes en solution à de fortes concentrations (supérieures à 1 g/L), l' $\alpha$ Chy pourrait se fixer en formant une multi-couche. Le lysozyme quant à lui présente un comportement sensiblement similaire seul ou en présence de l' $\alpha$ Chy.

Finalement, les expériences ont permis de mettre en évidence des phénomènes visibles et/ou quantifiables à l'échelle macroscopiques (paramètres SMA, influence du pH, possible adsorption multi-couche), ce qui permettra une étude des simulations moléculaires plus ciblée dans le but de comprendre les phénomènes observés.



# IV

---

## SIMULATIONS MOLÉCULAIRES

---

### Introduction

La détermination des isothermes d'adsorption de l' $\alpha$ -chymotrypsine a permis de mesurer des comportements globaux qui permettront le dimensionnement empirique d'un procédé chromatographique pour sa purification. Néanmoins, leur détermination a été coûteuse et leur application ne peut pas se faire au-delà des conditions opératoires explorées expérimentalement. Ainsi, l'objectif de ce chapitre est d'étudier comment l'utilisation de simulations moléculaires permettraient de déterminer le même type de résultats, tout en apportant une connaissance complémentaire des phénomènes de rétention ayant lieu à l'échelle locale de la protéine.

Pour cela, la méthode de la Dynamique Moléculaire a été privilégiée pour l'échantillonnage de l'espace conformationnel, mais également pour pouvoir observer l'évolution du système au cours du temps. Le système de référence choisi lors de l'étude expérimentale, composé de l' $\alpha$ -chymotrypsine et de la résine SP Sepharose FF, a été modélisé selon des modèles en tout-atome. Ces modèles, plus précis que les modèles de type Gros Grains dans la description du système, permettent une représentation plus minutieuse des mouvements des atomes en leur laissant une plus grande flexibilité. De plus, les moyens de calcul disponibles ont permis la représentation du solvant de manière explicite, les représentations implicites étant moins précises. La surface chromatographique a été modélisée à partir de la structure du ligand, et elle a été représentée avec une couche de contre-ions  $\text{Na}^+$  afin de la neutraliser.

La première partie de ce chapitre permet tout d'abord de poser la méthodologie développée pour la réalisation des simulations, ainsi que les modèles utilisés pour l'exploitation des résultats. Dans un premier temps, ils sont appliqués à l'analyse d'un système contenant une seule protéine. La comparaison des paramètres de rétention déterminés expérimentalement et par la simulation permettra en outre d'examiner l'intérêt de la démarche et la validité des approches utilisés. L'influence de variables essentielles dans les procédés chromatographiques, pH et force ionique, sur la rétention de la protéine

sera aussi discutée. Enfin, la dernière partie de ce travail sera consacrée à une étude préliminaire impliquant deux ( $\alpha$ Chy et lysozyme) puis quatre protéines (2  $\alpha$ Chy et 2 lysozymes), permettant ainsi d'initier des études de systèmes multi-constituants, beaucoup moins exploités dans la littérature, tant expérimentalement que dans les simulations.

## IV.1 Méthodologie

### IV.1.1 Construction du système

#### IV.1.1.1 Protéines

##### $\alpha$ -CHYMOTRYPSINE

Les coordonnées atomiques de l' $\alpha$ -chymotrypsine sont issues de la banque de données PDB (PDB-ID : 1YPH) et ont été obtenues par cristallographie aux rayons X, avec une résolution de 1,34 Å. La protéine est composée de trois chaînes indépendantes liées par des ponts disulfures, dont deux inter-chaînes et trois intra-chaînes. Seules les chaînes B, D et F ont été extraites et exploitées car la structure résolue est celle de la forme dimérique de l' $\alpha$ Chy. De plus, les chaînes A, C et E ne sont pas complètement résolues car il y a des résidus manquants dans le fichier PDB. La forme monomérique est ainsi composée au total de 241 acides aminés dont 12 se trouvent dans la chaîne B, 141 dans la chaîne D et 97 dans la chaîne F. Les dimensions en 3D de l' $\alpha$ Chy sont de 4,5 nm  $\times$  3,8 nm  $\times$  3,5 nm (Matthews et al., 1967; Yampolskaya et Platikanov, 2006).

Les états de protonation des résidus titrables ont été déterminés en utilisant PROPKA à partir du serveur en ligne PDB2PQR (Dolinsky et al., 2004). Comme décrit précédemment dans le chapitre bibliographique (section II.4.1.1), ce programme permet de calculer les pKa de tous les résidus titrables à un pH fixé à partir de la structure 3D. Il est également possible de retrouver le point isoélectrique de la protéine car le programme PROPKA calcule la charge nette de la protéine en fonction du pH. Ainsi, dans le cas de l' $\alpha$ Chy, le pI calculé est de 8,6, ce qui est cohérent avec la valeur issue de la littérature de 8,3 (Hudáky et al., 1999).

La distribution des charges à la surface de l' $\alpha$ Chy a été déterminée numériquement à l'aide de l'outil APBS (*Adaptive Poisson Boltzmann Solver*, Baker et al. (2001); Jurrus et al. (2017)) et visualisée grâce au logiciel de visualisation moléculaire PyMOL (Schrödinger, LLC, 2015). Ainsi, il est possible de colorer la surface de la protéine, et plus particulièrement ici la surface accessible au solvant, en fonction de la charge comme le montre la Figure IV.1, le bleu représentant les charges positives et le rouge les charges négatives. Cette figure représente l' $\alpha$ Chy à trois pH différents : (a) pH 5, qui est le pH de référence utilisé dans cette thèse, (b) pH 7, qui sera également utilisé afin d'étudier l'influence du pH sur l'échange d'ions, et (c) pH 9, qui est un pH légèrement supérieur au point isoélectrique. Cette représentation permet de visualiser et d'identifier assez rapidement les zones à la surface de la protéine susceptibles de participer à l'interaction, mais également de voir comment les charges se distribuent à la surface. A pH 5, l' $\alpha$ Chy

présente une grande région chargée positivement sur une de ses faces alors que la charge est globalement neutre voire légèrement négative de l'autre côté. La charge globale de la protéine à ce pH est estimée à +7.

Plus le pH se rapproche du pI et plus ces deux zones se différencient, comme à pH 7 où malgré la présence de la région très fortement positive, la face opposée devient de plus en plus négative (la charge globale est estimée à +5 dans ces conditions). De ce fait, à pH 9, qui est un pH légèrement plus grand que le pI, il y a assez clairement une face positive et une autre négative, ce qui conduit à une charge globale de la protéine neutre. Pour autant, il n'apparaît pas impossible pour l' $\alpha$ Chy de s'adsorber sur une surface chromatographique cationique lorsque le pH est égal au pI, si l'on considère uniquement les effets électrostatiques. En effet, le mécanisme d'échange d'ions est complexe et bien qu'il soit régi majoritairement par les effets électrostatiques, d'autres effets non négligeables comme par exemple la solvatation de la protéine doivent être considérés. Ainsi, ces calculs de distribution de charge ne suffisent pas à eux-seuls à prédire l'adsorption de la protéine sur un support chromatographique mais elles permettent d'en obtenir une première approximation.

## LYSOZYME

La structure cristallographique du lysozyme du blanc d'œuf est issue de la banque de données PDB (PDB-ID : 1AKI) et possède une résolution de 1,5 Å. C'est une protéine souvent considérée comme modèle qui ne possède qu'une seule chaîne de 129 acides aminés. Ses dimensions 3D sont de 2,8 nm  $\times$  2,8 nm  $\times$  5,0 nm (Krigbaum et Kuegler, 1970).

De la même manière que pour l' $\alpha$ Chy, les états de protonation des résidus titrables ont été calculés à l'aide de PROPKA et les fichiers PQR ont été générés à l'aide du serveur PDB2PQR. Ces calculs ont permis d'estimer le pI du lysozyme à 11, qui est proche de la valeur issue de la littérature (11,3 Wetter et Deutsch (1951)).

La Figure IV.2 représente la surface du lysozyme colorée en fonction du potentiel électrostatique à différents pH. La sous-figure (a) montre la distribution des charges à pH 5, qui sera le pH de travail pour l'étude en multi-constituants ; la (c) est à pH 11, qui est un pH égal au point isoélectrique et (b) à pH 8, qui est un pH intermédiaire entre ces deux valeurs. Le lysozyme apparaît très fortement chargé positivement à pH 5, de manière uniforme sur toute la surface malgré quelques zones neutres. La charge globale a été estimée à +10. La protéine est toujours positivement chargée à pH 8 (charge globale égale à +7) car cela reste un pH encore suffisamment éloigné du point isoélectrique. A pH 11, la charge globale est neutre avec des zones positives, négatives et majoritairement neutres. En revanche, la répartition semble assez équilibrée à la différence de l' $\alpha$ Chy, qui présente clairement deux grandes zones distinctes positive et négative.

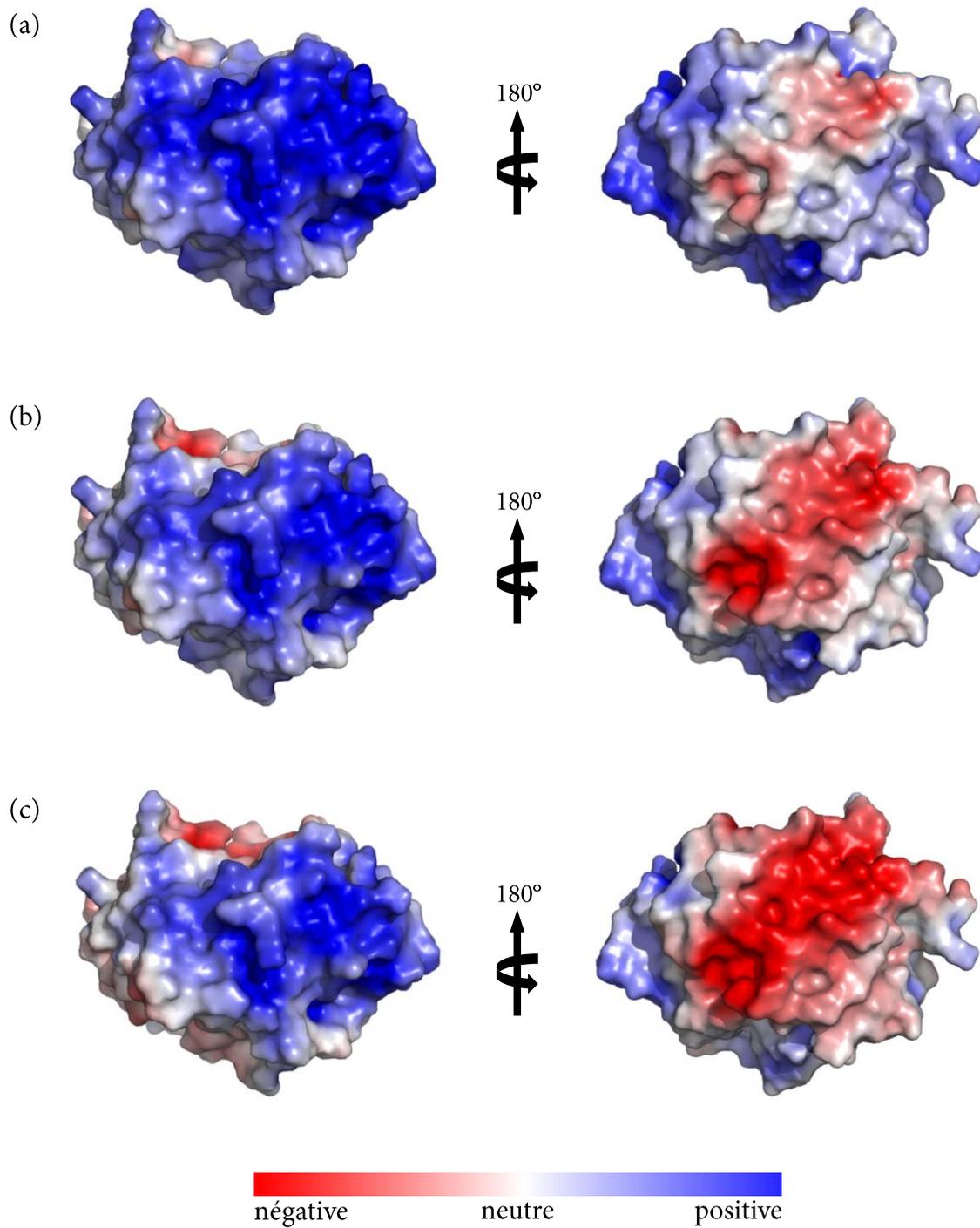


FIGURE IV.1 – Surface de l' $\alpha$ -chymotrypsine (PDB-ID : 1YPH) colorée en fonction du potentiel électrostatique à : (a) pH 5 (charge +7) ; (b) pH 7 (charge +5) et (c) pH 9 (charge 0).

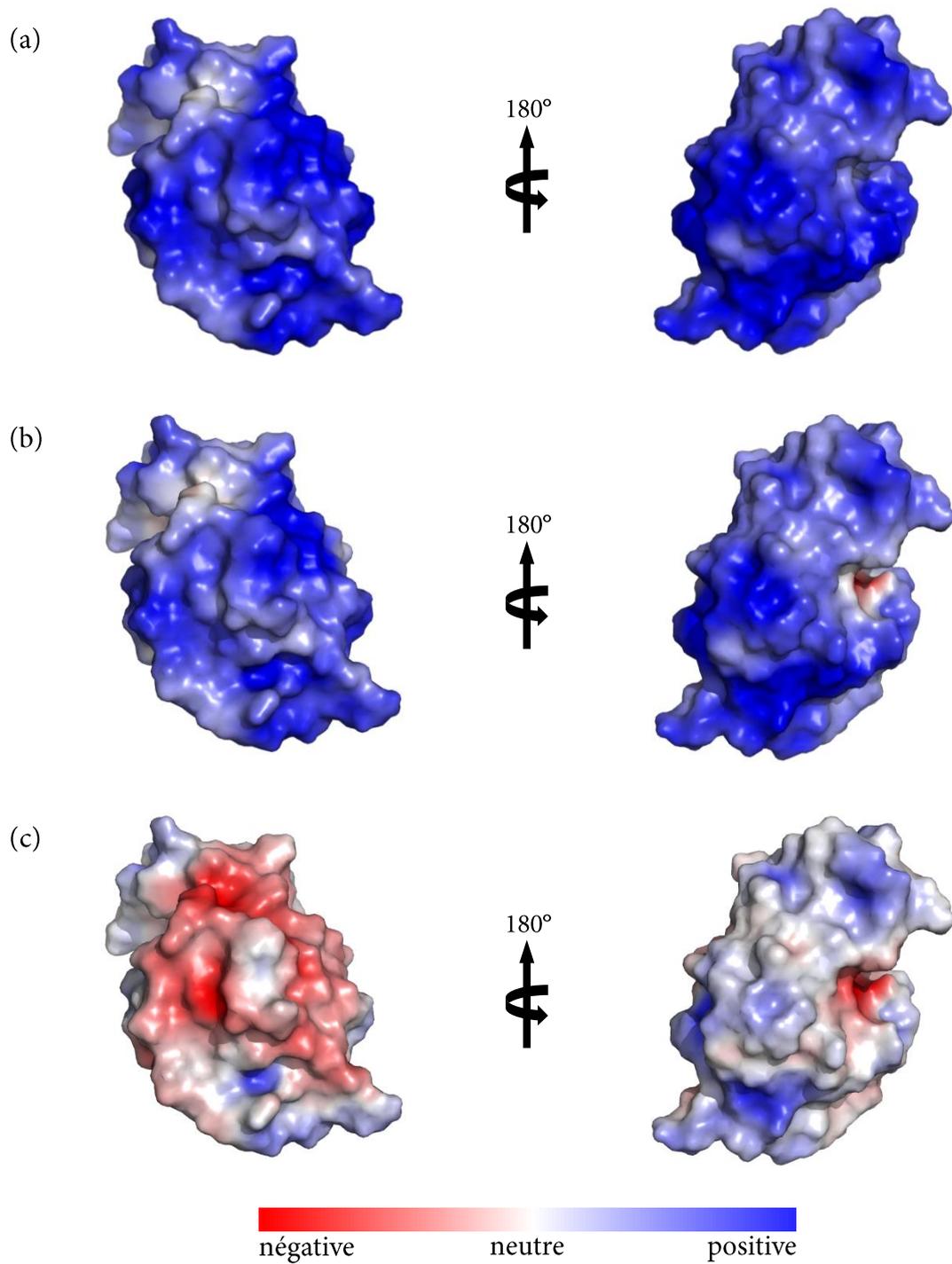


FIGURE IV.2 – Surface du lysozyme (PDB-ID : 1AKI) colorée en fonction du potentiel électrostatique à : (a) pH 5 (charge +10); (b) pH 8 (charge +7) et (c) pH 11 (charge 0).

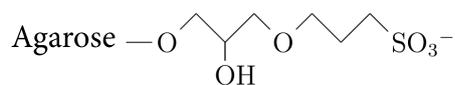
Le Tableau IV.1 récapitule les principales propriétés de l' $\alpha$ -chymotrypsine et du lysozyme présentées dans cette section.

TABLEAU IV.1 – Tableau récapitulatif des principales propriétés de l' $\alpha$ -chymotrypsine et du lysozyme ainsi que des codes PDB correspondants.

Propriétés	$\alpha$ -chymotrypsine	Lysozyme
Code PDB (résolution)	1YPH (1,34 Å)	1AKI (1,5 Å)
Masse moléculaire	25,6 kDa	14,3 kDa
Acides aminés (chaînes)	241 (3)	129 (1)
Dimensions 3D	4,5 nm x 3,8 nm x 3,5 nm	2,8 nm x 2,8 nm x 5,0 nm
Point isoélectrique pI (selon le calcul de PROPKA)	8,6 (littérature : 8,3)	11,0 (littérature : 11,3)
Charge globale (selon le calcul de PROPKA)	pH 5 : +7 pH 7 : +5	pH 5 : +10 pH 8 : +7

#### IV.1.1.2 Support chromatographique

Le support chromatographique choisi pour l'étude, la SP Sepahrose FF, est composée d'une matrice en agarose réticulée à laquelle sont greffés des groupes sulfopropyl, qui est un groupement fonctionnel c (voir paragraphe II.3.2.2). La structure chimique d'un ligand est définie selon Johansson et al. (2004) comme étant :



le premier atome d'oxygène appartenant à la matrice. La structure 3D d'un ligand (voir Figure IV.3.a) a été construit à l'aide du logiciel Avogadro (Hanwell et al., 2012) et paramétrisé en utilisant un champ de force générique CGenFF (Vanommeslaeghe et al., 2010; Yu et al., 2012) ainsi que des algorithmes automatiques issu du site web CGenFF (Vanommeslaeghe et Jr., 2012; Vanommeslaeghe et al., 2012).

Dans cette étude, la matrice en agarose n'est pas explicitement représentée afin de limiter le nombre d'atomes total dans le système. Afin de simuler le lien entre les groupements fonctionnels et la matrice, un atome de carbone est ajouté (lié au premier oxygène, voir Figure IV.3.a) et la position des trois premiers atomes sera restreinte en bas de la boîte de simulation. Ainsi, la flexibilité des ligands sera conservée.

La surface chromatographique est ensuite construite par duplication de la structure d'un ligand, qui seront ensuite distribués uniformément dans les directions -x et -y sur une surface de 10 nm × 10 nm. La densité de ligands, exprimée en nombre de ligands par nm<sup>2</sup>, a été calculée à partir de la capacité totale de la résine mesurée expérimentalement (230,2 mmol/L), du nombre d'Avogadro et de la surface spécifique issue de la littérature (42 m<sup>2</sup>/mL, selon la méthode de DePhillips et Lenhoff (2000) explicitée dans le paragraphe II.3.2.3). Ainsi, la surface chromatographique contiendrait au total 324 ligands pour une surface de 100 nm<sup>2</sup>.

### IV.1.1.3 Boîte de simulation

Tout d'abord, la surface chromatographique est placée en bas de la boîte de simulation dans le plan -x et -y (voir Figure IV.3.b). 324 ions sodium sont ajoutés et positionnés au-dessus de chaque ligand afin de les neutraliser, ce qui permet de représenter l'étape d'équilibration propre à l'échange d'ions. Une couche de vide d'épaisseur 1 nm est conservée en dessous de la surface chromatographique, et créée par translation du support dans la direction -z. Cette couche est nécessaire afin d'empêcher la diffusion des molécules d'eau. Des conditions périodiques aux limites sont imposées en -x et -y pour la continuité du système dans ce plan (voir paragraphe II.4.1.4).

La deuxième étape consiste à ajouter la protéine dans la boîte et la positionner au dessus de la surface, comme dans le cas de l' $\alpha$ Chy illustré sur la Figure IV.3.b. Le choix a été fait de placer la protéine à une distance d'environ 0,75 nm du support, qui est une distance suffisante pour permettre à la protéine de se réorienter et de se déplacer dans la boîte lors des simulations, tout en restant inférieur au *cutoff* de 1 nm défini pour le calcul des interactions électrostatiques (voir paragraphe IV.1.2.2). Une fois la protéine correctement positionnée dans l'axe des -z, elle est centrée en -x et -y.

La boîte, d'une taille finale de 10,00 nm × 10,00 nm × 11,25 nm, est ensuite remplie de molécules d'eau. Les molécules de solvant qui se trouvent entre les ligands et en dessous sont supprimées afin de conserver la couche initiale de vide. Finalement, des ions sodium et chlorure sont ajoutés en fonction de la concentration en chlorure de sodium désirée afin de maintenir le système électriquement neutre et pour se placer dans les conditions expérimentales choisies.

## IV.1.2 Préparation des simulations

### IV.1.2.1 Échantillonnage du système

Afin de pouvoir comparer des valeurs observables aux échelles macroscopique et atomique, il est nécessaire de produire un nombre suffisant de simulations, et ce afin de

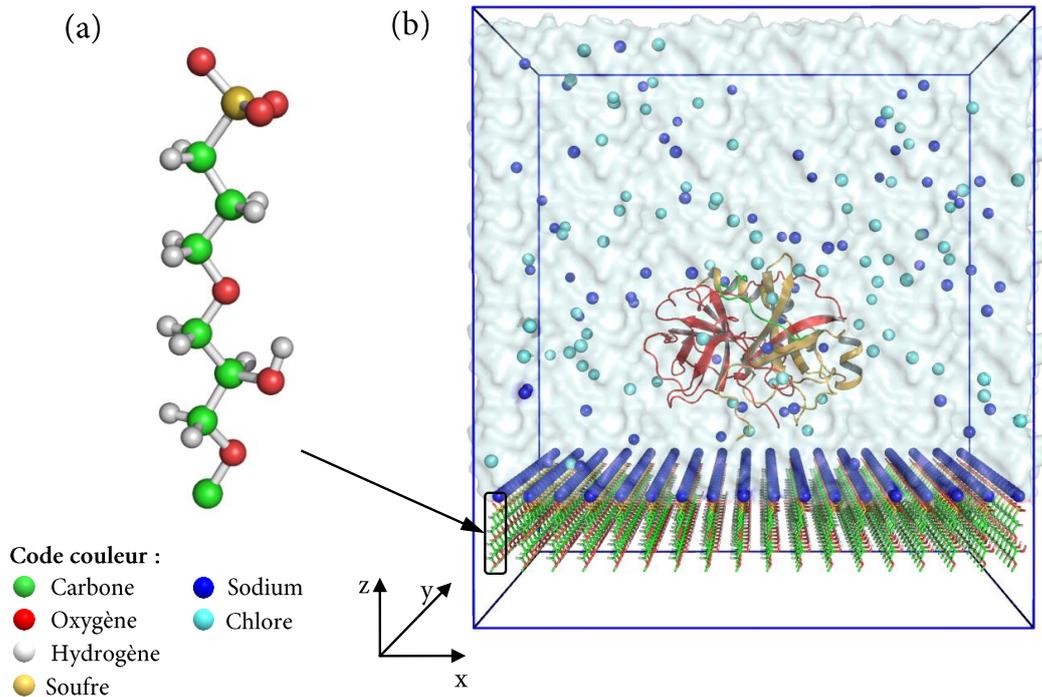


FIGURE IV.3 – (a) Représentation 3D "ball-and-stick" d'un ligand de la SP Sepharose FF et (b) la boîte de simulation primaire comprenant la surface chromatographique avec la couche de contre-ions (en bas), l' $\alpha$ -chymotrypsine (représentation "cartoon"), les ions chlorure (cyan) et sodium (bleu) et l'eau (surface bleu transparent).

générer plusieurs comportements locaux possibles qui peuvent permettre de comprendre une observation globale.

Ainsi, dans le but d'éviter de biaiser les résultats des simulations en orientant initialement la protéine favorablement face à la surface chromatographique, plusieurs rotations ont été effectuées sur cette dernière afin de présenter différentes faces de la protéine vers le support. Au final, six orientations initiales de la protéine dans la boîte de simulation ont été étudiées, comme le représente la Figure IV.4. Pour ce faire, une première rotation de la protéine est effectuée à  $90^\circ$  autour de l'axe des  $-z$ , puis ensuite autour de l'axe des  $-y$ . Après cette étape de rotation, la protéine est de nouveau centrée et la distance protéine-support est ajustée (voir paragraphe IV.1.1.3).

#### IV.1.2.2 Mode opératoire

##### LOGICIEL, MODÈLES ET CALCUL DES INTERACTIONS

Toutes les simulations de dynamique moléculaire ont été réalisées à l'aide du logiciel GROMACS version 5.0 (Abraham et al., 2015). Le champ de force tout-atome choisi dans cette étude est le CHARMM36 (Huang et MacKerell, 2013) pour protéines et solvant (eau

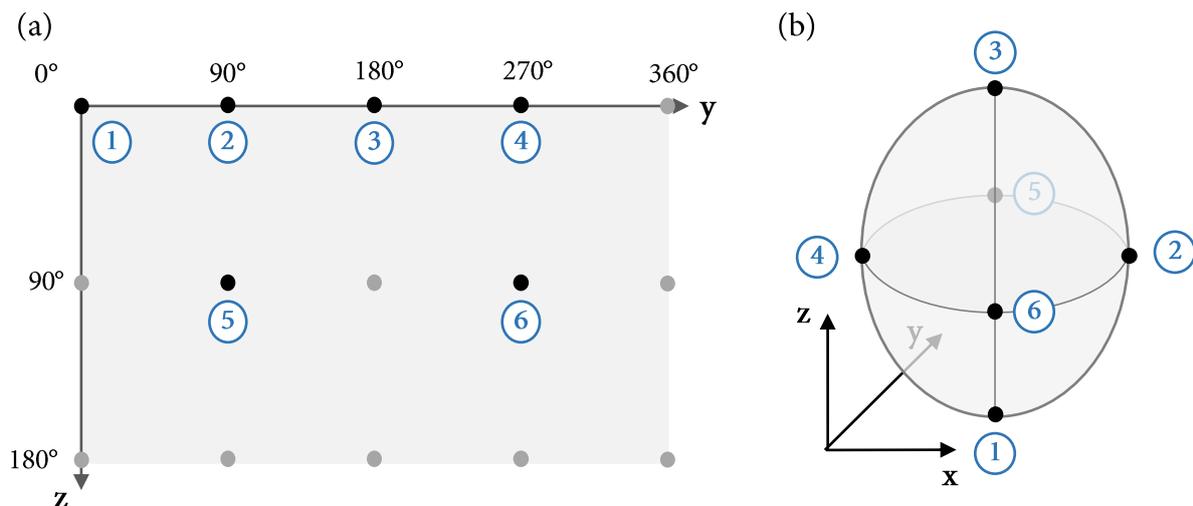


FIGURE IV.4 – Représentation des six orientations initiales de la protéine étudiées. Chaque point sur la surface de la protéine (b) représente la face dirigée vers le support chromatographique en début de chaque simulation. Les différentes orientations sont obtenues à partir des rotations en -y et -z indiqués sur la carte (a).

et ions). Enfin, le modèle d'eau TIP3P (Durell et al., 1994) a été utilisé pour simuler le comportement des molécules d'eau (voir section II.4.1.5).

Pour le calcul des interactions non liées à courte portée, la distance de troncation (*cut-off*) a été défini à 1,2 nm. Afin d'éviter une coupure brutale des forces (van der Waals) au niveau du *cut-off*, une fonction *switch* a été appliquée afin d'amener la force à 0 de manière plus douce. Les interactions à longue portée (électrostatiques) ont été traitées par la méthode PME (*Particle Mesh Ewald*, avec une interpolation cubique et un espacement de Fourier de 0,16 nm (Essmann et al., 1995)) et un algorithme 3dc prenant en compte la périodicité en -xy (Yeh et Berkowitz, 1999).

Les liaisons covalentes impliquant les atomes d'hydrogène ont été contraintes en utilisant l'algorithme LINCS et ses paramètres par défaut (Hess et al., 1997). Pour simuler la matrice du milieu chromatographique, un mur (*wall*) a été placé dans la partie inférieure de la boîte de simulation avec les types d'atomes définis comme CG331, pour modéliser implicitement les atomes de carbone de l'agarose. De plus, afin d'éviter la diffusion des molécules d'eau et d'être compatible avec la pression en -xy et la PME, un autre *wall* a été créé avec des atomes définis comme OT afin d'imiter l'oxygène de l'eau. Les interactions entre les murs ont été définies en utilisant le potentiel direct de Lennard-Jones et un seuil de potentiel linéaire à 0,3 nm.

La température a été contrôlée à l'aide du thermostat Berendsen modifié (Bussi et al., 2007) avec une constante de temps pour le couplage de 0,1 ps et des groupes définis

comme protéine, ligands et solvants. La pression semi-isotrope à 1 bar a été appliquée en utilisant le barostat de Berendsen (Berendsen et al., 1984) avec une constante de temps pour le couplage de pression de 5.0 ps et une compressibilité de  $4e^{-5}$  bar<sup>-1</sup> sur l'axe -z seulement.

## ÉTAPES

Pour toutes les étapes du protocole de simulation, les trois premiers atomes des ligands de la SP Sepharose ont été spatialement contraints avec une constante de force de 2000 kJ.mol<sup>-1</sup>.nm<sup>-2</sup>.

Tout d'abord, une étape de minimisation d'énergie du système a été réalisée et arrêtée lorsque la force maximale de convergence est devenue inférieure à 1000 kJ.mol<sup>-1</sup>.nm<sup>-1</sup>. Ensuite, une étape de chauffage de 100 ps en ensemble NVT<sup>1</sup> a été effectuée pour atteindre 293 K, puis une courte étape d'équilibration de 200 ps en NPT<sup>2</sup> a été réalisée. Uniquement pendant le chauffage et l'équilibration NPT, des contraintes de position supplémentaires de 1000 kJ.mol<sup>-1</sup>.nm<sup>-2</sup> ont été fixées pour le squelette de la protéine et les autres atomes des ligands.

Enfin, la phase de production s'est déroulée en NPT avec seulement des contraintes de position sur les trois premiers atomes des ligands SP Sepharose. Les durées des simulations ont varié de 90 à ~ 300 ns selon le temps nécessaire à l'adsorption de la protéine, avec un pas de temps fixé à 2 fs.

### IV.1.2.3 Outils d'analyse

La plupart des analyses ont été réalisées à l'aide des outils d'analyse intégrés au logiciel GROMACS. Ces outils permettent entre autres d'analyser les fichiers de trajectoire issus des simulations et d'en extraire des informations au cours du temps, comme par exemple la distance minimale entre deux groupes d'atomes (*mindist*, *distance* ou *pairdist*), le rayon de gyration de la protéine (*gyrate*), le rmsd (*rmsd*, qui décrit la déviation quadratique moyenne des positions atomiques) ou encore les coordonnées cartésiennes des atomes et/ou résidus (*traj*). Afin d'automatiser les analyses, des scripts en bash, python ou encore awk ont également été créés.

Les calculs d'énergies libres de liaison ont été effectués à l'aide du programme *g\_mmpbsa* (Kumari et al., 2014), développé afin de réaliser des calculs MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) en utilisant GROMACS et APBS (Baker et al., 2001).

---

1. Nombre d'atomes, volume et température constants.  
2. Nombre d'atomes, volume et pression constants.

Afin de visualiser les simulations dynamiques, le logiciel VMD 1.9.3 a été utilisé (Humphrey et al., 1996) et les images ont été produites à l'aide du logiciel PyMOL version 1.7 ou plus (Schrödinger, LLC, 2015). Enfin, les graphes et analyses statistiques ont été réalisés en utilisant MS Excel, gnuplot ou R.

### IV.1.3 Méthodes de calcul des paramètres SMA

Le calcul des paramètres SMA, *i.e.* la charge caractéristique et le facteur stérique, est issu des analyses des 10 dernières nanosecondes de chaque simulation. Ainsi, les charges et facteurs stériques calculés pour chacune des simulations sont une moyenne sur ces 10 ns. Enfin, les valeurs finales des paramètres SMA, qui seront comparées aux valeurs issues des expériences, représentent la moyenne des paramètres calculés pour l'ensemble des simulations réalisées.

#### IV.1.3.1 Charge caractéristique

Selon la définition de Brooks et Cramer (1992), la charge caractéristique représente le nombre de ligands impliqués dans l'interaction entre la protéine et le support chromatographique. Dans les études de Liang et al. (2012, 2015a), la charge caractéristique est calculée comme étant le nombre de résidus en interaction avec le support, et non comme étant le nombre de ligands. Cette façon de calculer s'avère pertinente dans un cas d'étude en gros grain, mais ce n'est pas le cas en tout atome. En effet, en tout atome, un résidu chargé (+ 1) peut être en interaction avec plusieurs oxygènes appartenant à deux ligands différents. Cela est dû à la structure chimique du groupement fonctionnel qui présente trois oxygènes chargés à - 0,55 (voir la Figure IV.3.a). En gros grain, le groupement chargé est représenté comme une seule particule possédant une charge de +/- 1, ce qui fait qu'un résidu ne pourra interagir qu'avec un seul ligand et non plusieurs. Dans ce cas-là, considérer le nombre de résidus en interaction revient au même que de considérer le nombre de ligands car il y a bien une équivalence  $1 \text{ résidu} = 1 \text{ ligand}$ . Ainsi, afin de correspondre au mieux à la définition de la charge caractéristique  $n$  dans le modèle SMA posée par Brooks et Cramer, il a été décidé que la charge caractéristique serait calculée en fonction du nombre de ligands en interaction uniquement.

Comme défini précédemment, la protéine et les ligands sont considérés en interaction lorsque la distance minimale PROT-LIG est inférieure à 0,2 nm. La charge caractéristique a donc été calculée comme étant le nombre de ligands qui possèdent une distance minimale avec les résidus chargés positivement de la protéine inférieure à 0,2 nm. A noter que seule l'interaction entre les ligands et les résidus chargés positivement est considérée ici afin

d'exclure les interactions de type liaisons hydrogènes qui peuvent apparaître, et ainsi ne prendre en compte que les interactions de type électrostatique.

### IV.1.3.2 Facteur stérique

Le facteur stérique  $\sigma$  représente le nombre de ligands "cachés" par la protéine qui ne sont plus disponibles pour interagir avec d'autres protéines, mais qui ne sont pas en interaction avec la protéine considérée. Cela se traduit ici par le nombre de ligands se trouvant sous la protéine, mais possédant une distance PROT-LIG supérieur à 0,2 nm.

Afin de déterminer si un ligand est caché ou pas par la protéine, plusieurs techniques sont applicables. La première est issue de Liang et al. (2012) et consiste à considérer l'ensemble des ligands dont la distance est comprise entre le seuil défini pour déterminer s'il y a interaction (ici 0,2 nm) et le rayon de gyration de la protéine (environ 1,7 nm). L'avantage de cette méthode est qu'elle est plutôt rapide à mettre en œuvre car seuls les outils d'analyses de GROMACS sont nécessaires pour l'appliquer. En revanche, l'inconvénient majeur de ce type de méthodes résulte dans le fait de devoir choisir un seuil pour déterminer si un ligand est caché ou pas. Si ce seuil peut être aisément déterminé dans le cas de la charge caractéristique, ce n'est pas le cas ici où le choix du rayon de gyration comme seuil est très arbitraire.

C'est pour cette raison qu'une seconde technique a été développée et appliquée. L'idée est de représenter la surface projetée de la protéine sur le support chromatographique et de quantifier le nombre de ligands compris dans cette surface. Cette méthode est illustrée par la Figure IV.5. Pour cela, les coordonnées cartésiennes en -x et -y de tous les atomes de l' $\alpha$ Chy ont été extraites pour chaque *frame* sur les 10 dernières ns de la trajectoire, soit 11 frames par simulation. Ensuite, un calcul de l'enveloppe convexe de l'ensemble de ces points a été réalisé à l'aide de la classe Python *scipy.spatial.ConvexHull* (Barber et al., 1996), et l'aire de cette surface a été calculée. Ainsi, en connaissant la densité de ligands au  $\text{nm}^2$ , il est possible d'estimer le nombre de ligands compris dans cette surface. Finalement, le facteur stérique a été déterminé comme étant la différence entre ce nombre de ligands et la charge caractéristique.

## IV.2 Étude d'un système mono-protéine

Les études numériques ont été effectuées premièrement à partir d'un système simple mono-protéine, composé de l' $\alpha$ -chymotrypsine et de la résine SP Sepharose FF. Les conditions opératoires sont un pH égal à 5, une concentration chlorure de sodium égale à 100 mM, afin d'être en accord avec les conditions choisies lors de l'étude expérimentale. Cependant, afin de ne pas complexifier le système, le choix de ne pas représenter explicitement

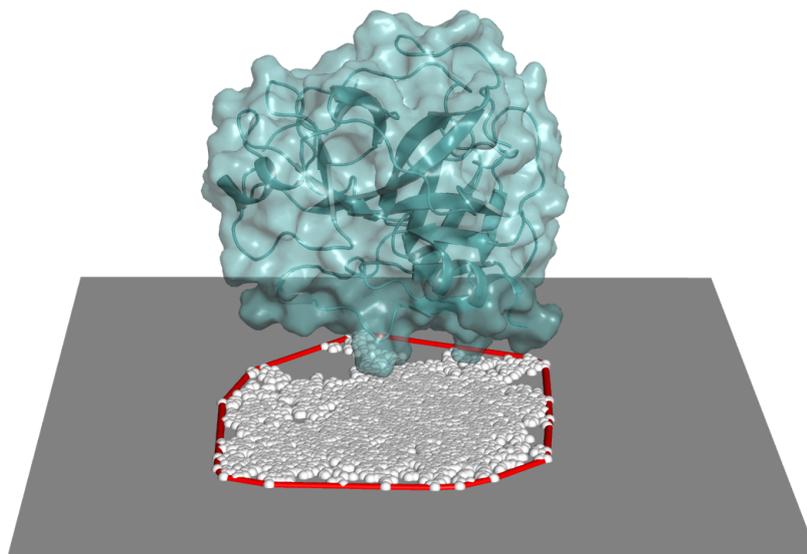


FIGURE IV.5 – Illustration du calcul du facteur stérique  $\sigma$  par la méthode de l’enveloppe convexe. Les coordonnées cartésiennes  $-x$  et  $-y$  des atomes de la protéine (représentation cartoon avec surface transparente), représentées par des sphères blanches, sont projetées sur une surface plane (gris). L’enveloppe convexe de l’ensemble des atomes est dessinée par des lignes rouges.

les molécules du tampon citrate dans les simulations a été fait. Un total de 18 simulations a été réalisé, soit 3 simulations différentes pour chaque orientation initiale de la protéine.

## IV.2.1 Stabilité de l’adsorption

### IV.2.1.1 Critères de stabilité

Toutes les simulations ont initialement été lancées pour une durée minimale de 90-100 ns (selon la puissance de calcul disponible). Des analyses ont été effectuées afin de déterminer si une adsorption de la protéine sur la surface a eu lieu, et si oui, si l’adsorption était stable. Si ces deux conditions n’étaient pas remplies, la simulation était prolongée jusqu’à atteindre l’adsorption de la protéine, avec une durée minimale de stabilité. Pour déterminer cela, l’évolution de plusieurs observables a été étudiée à l’aide des outils d’analyse de GROMACS.

#### DISTANCE MINIMALE ENTRE LA PROTÉINE ET LES LIGANDS

Afin de vérifier s’il y a une adsorption, et donc une interaction entre la protéine et le support chromatographique, une des méthodes les plus simples et directes est d’observer l’évolution de la distance minimale entre les deux. L’outil d’analyse *mindist* permet de

calculer, à partir des coordonnées cartésiennes de tous les atomes du système au cours du temps, la distance minimale entre deux groupes d'atomes, le premier groupe étant la protéine et le deuxième l'ensemble des ligands. Les résultats sont présentés en Figure IV.6. Pour l'ensemble des simulations, la distance minimale PROT-LIG finit par atteindre des valeurs oscillant entre 0,15 et 0,17 nm. Ces distances correspondent aux distances minimales possibles entre deux atomes en tenant compte de leur rayon de Van der Waals (explicité dans le champ de force). Ainsi, selon ces résultats, il est considéré que la protéine est en interaction avec les ligands lorsque la distance minimale PROT-LIG est inférieure à 0,2 nm.

#### DISTANCE MINIMALE ENTRE LE CENTRE DE MASSE DE LA PROTÉINE ET LES LIGANDS

De la même manière que *mindist*, l'outil de GROMACS *pairdist* permet entre autres de calculer des distances minimales entre deux groupes d'atomes. Son intérêt réside dans le fait de pouvoir sélectionner les positions de référence pour le calcul des distances. L'une des possibilités est donc de considérer uniquement le centre de masse (COM) d'un groupe d'atomes, comme ici le centre de masse de l' $\alpha$ Chy. Ainsi, le calcul permet de visualiser l'évolution de la position du centre de masse de la protéine par rapport aux ligands, où plus précisément par rapport au ligand le plus proche. Les résultats sont présentés en Figure IV.7. L'intérêt de l'observation de la position du COM de la protéine est de mettre en évidence les réorientations et rotations de la protéine au cours de la simulation. En effet, comme l'illustre la Figure IV.7, certaines simulations montrent que même après l'adsorption de la protéine (lorsque PROT-LIG < 0,20 nm), la distance COM-LIG fluctue toujours, ce qui traduit une réorientation de la protéine à la surface, qui tente de trouver son orientation la plus favorable à une adsorption stable. Cela explique également le fait que ces deux observables ne convergent pas au même moment, le COM-LIG semble mettre plus de temps à se stabiliser. En revanche, il est difficile de définir un seuil minimal comme pour PROT-LIG, le COM ne convergeant pas vers la même valeur selon les simulations. Toutefois, si la distance COM-LIG se trouve généralement en dessous de 3 nm, elle ne peut pas être inférieure au rayon de gyration de la protéine, qui est d'environ 1,7 nm.

#### RMSD DE LA PROTÉINE

Une protéine possède une certaine flexibilité et peut se déformer au cours du temps selon l'environnement auquel elle est soumise. Cette flexibilité est visible durant les simulations en tout-atome, la protéine n'étant pas considérée comme un bloc rigide mais comme un assemblage d'atomes non restreints. Le calcul du RMSD *Root Mean Square Deviation* permet de visualiser cette déformation au cours du temps. Le calcul est appliqué la plupart du temps sur le squelette de la protéine car une forte déformation du

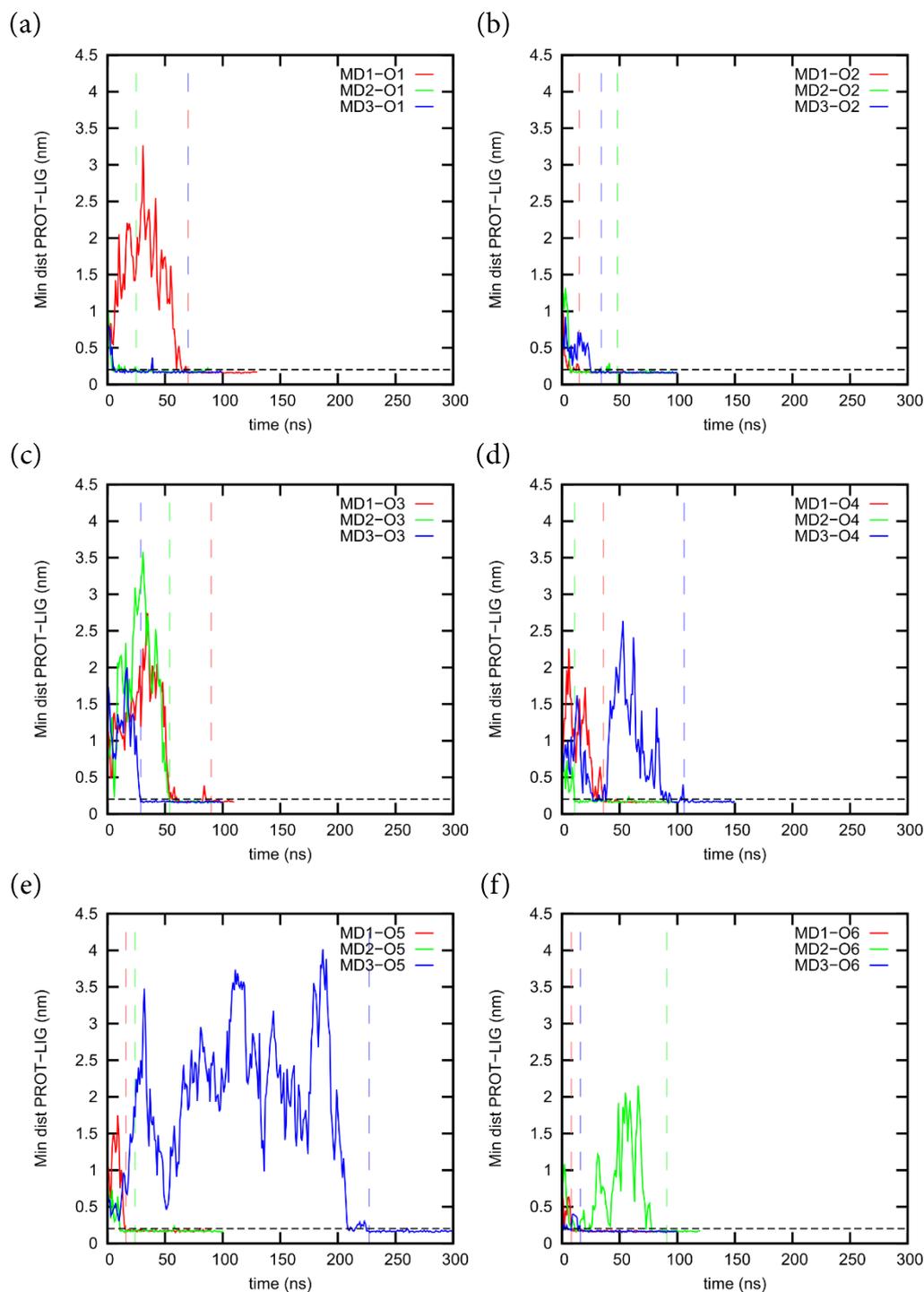


FIGURE IV.6 – Évolution de la distance minimale entre les ligands et l' $\alpha$ -chymotrypsine *PROT-LIG* au cours du temps, pour chaque orientation (O1-6, figures (a) à (f)) et les répétitions correspondantes (MD1 en rouge, MD2 en vert et MD3 en bleu). Les lignes en pointillé verticales correspondent au temps d'adsorption et les lignes en pointillé horizontales correspondent au seuil de 0,2 nm.

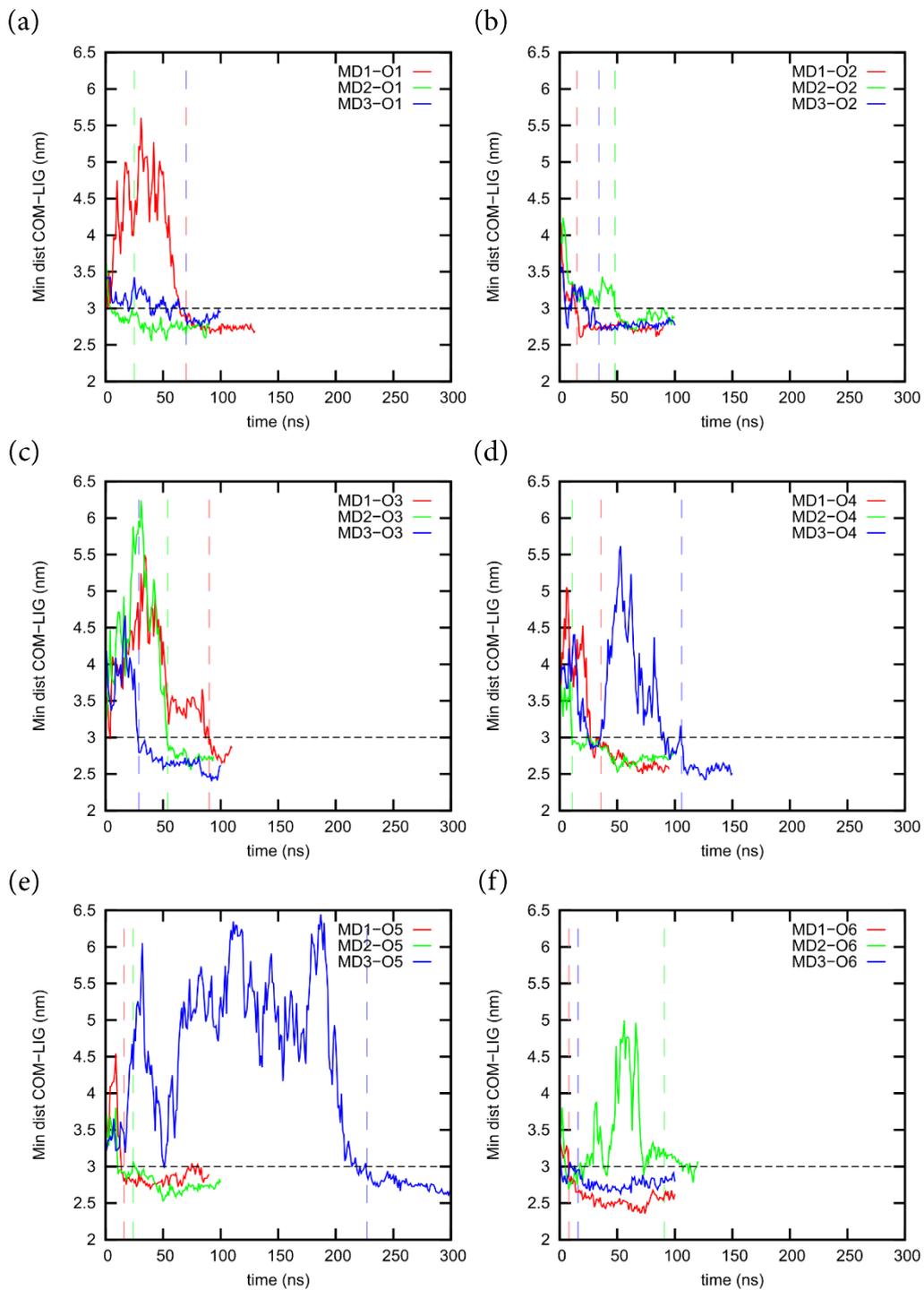


FIGURE IV.7 – Évolution de la distance minimale entre les ligands et le centre de masse de l' $\alpha$ -chymotrypsine *COM-LIG* au cours du temps, pour chaque orientation (O1-6, figures (a) à (f)) et les répétitions correspondantes (MD1 en rouge, MD2 en vert et MD3 en bleu). Les lignes en pointillé verticales correspondent au temps d'adsorption et les lignes en pointillé horizontales correspondent au seuil de 3 nm.

squelette peut traduire un changement conformationnel important de la protéine. En pratique, pour chaque *frame* étudiée, les coordonnées cartésiennes des atomes du squelette sont comparées aux coordonnées de la structure de référence de la protéine. Le RMSD représente l'écart avec la référence au cours du temps et permet d'obtenir des graphes comme illustré en Figure IV.8. En général, la protéine est considérée comme stable lorsque le RMSD atteint un plateau.

Il a été observé que l'étude du RMSD pour déterminer la stabilité de l'adsorption ne semble pas pertinente dans cette étude. Pour certaines simulations, le RMSD dépasse le seuil de 0,2 nm au moment de l'adsorption, et cela malgré une stabilité des deux autres observables PROT-LIG et COM-LIG. Ce phénomène est dû majoritairement à la structure particulière de l' $\alpha$ -chymotrypsine, qui est composée de trois chaînes indépendantes. Ainsi, après adsorption et stabilisation de la protéine, les extrémités des chaînes ne participant pas à l'adsorption peuvent bouger car elles possèdent une liberté de mouvement. L'augmentation du RMSD est donc due à cette flexibilité des extrémités, mais elle ne traduit pas pour autant un changement conformationnel de la protéine.

En conclusion, l'adsorption de la protéine lors des simulations sera considérée comme stable lorsque (i) la distance minimale PROT-LIG sera inférieure à 0,20 nm et (ii) la distance COM-LIG ne fluctuera plus significativement. Ces deux conditions devront être respectées pendant un minimum de 10 ns consécutives en fin de simulation. Le Tableau IV.2 répertorie les temps de simulation et les temps d'adsorption pour chacune des 18 simulations.

#### IV.2.1.2 Analyse de l'adsorption et comportement de la protéine

Les 18 simulations ont toutes abouti à une adsorption de l' $\alpha$ Chy sur la surface chromatographique. Comme l'illustrent les Figures IV.6 et IV.7, la majorité d'entre elles ont montré une adsorption stable avant 100 ns, le temps moyen d'adsorption étant de 54 ns. Il semblerait qu'aucune des six orientations initiales ne soient défavorables et aient empêché l'adsorption de la protéine sur le support. Cependant, certaines orientations, comme par exemple l'orientation 2, paraissent particulièrement favorables à l'adsorption car les trois simulations ont montré une adsorption rapide et sans réorientation préalable de la protéine dans la boîte. Dans ce cas de figure, les résidus participant à l'adsorption semblent avoir été orientés directement vers le support chromatographique.

Les fluctuations de la distance COM-LIG montrent que dans la plupart des cas, la protéine se réoriente dans la boîte de simulation avant l'adsorption afin de trouver son orientation la plus favorable. Il y a aussi des cas où la protéine se réoriente après l'adsorption, comme par exemple la simulation MD1-O3 où malgré une adsorption au bout de 50 ns, la position de la protéine ne se stabilise uniquement qu'à partir de 90 ns.

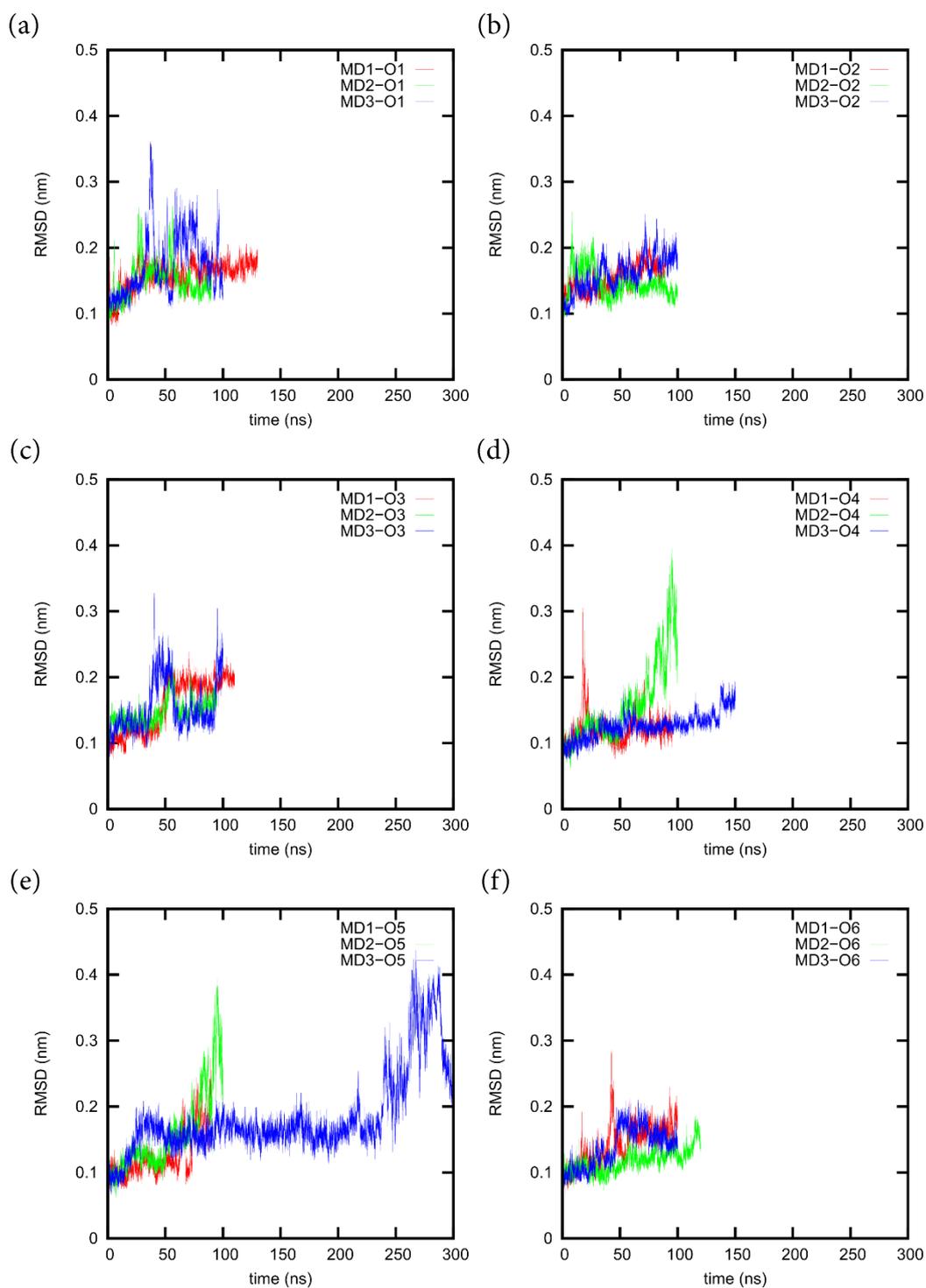


FIGURE IV.8 – Évolution du RMSD de l' $\alpha$ -chymotrypsine au cours du temps, pour chaque orientation (O1-6, figures (a) à (f)) et les répétitions correspondantes (MD1 en rouge, MD2 en vert et MD3 en bleu).

TABLEAU IV.2 – Temps de simulations totaux et temps d’adsorption pour chacune des 18 simulations.

Orientation	Répétition MD	Temps de simulation total (ns)	Temps d’adsorption (ns)
1	1	130	70
	2	90	25
	3	100	70
2	1	90	15
	2	100	48
	3	100	34
3	1	110	90
	2	95	54
	3	100	29
4	1	95	36
	2	100	11
	3	150	106
5	1	90	16
	2	100	24
	3	300	227
6	1	100	8
	2	120	91
	3	100	16

Deux comportements particuliers ont également été identifiés : (i) la simulation MD3-O5 a dû être prolongée jusqu’à 300 ns, car la protéine a subi naturellement des translations et orientations dans la boîte du fait de l’agitation thermique, avant de trouver son orientation optimale, et (ii) les simulations MD3-O4 et MD2-O6, qui montrent que la protéine s’est approchée du support puis éloignée, avant de finalement s’adsorber de nouveau. La Figure IV.9 illustre en détail la simulation MD2-O6 en représentant 8 clichés ou *snapshots*, qui montrent les phases de diffusion et d’adsorption de la protéine au cours de la simulation.

Ces résultats permettent également d’illustrer le caractère aléatoire de ces simulations, trois simulations produites à partir d’un même point de départ ne donnent pas le même résultat. Cela est particulièrement visible pour l’orientation 5, dont l’une des simulations a nécessité un temps de simulation bien plus long pour aboutir à une adsorption.

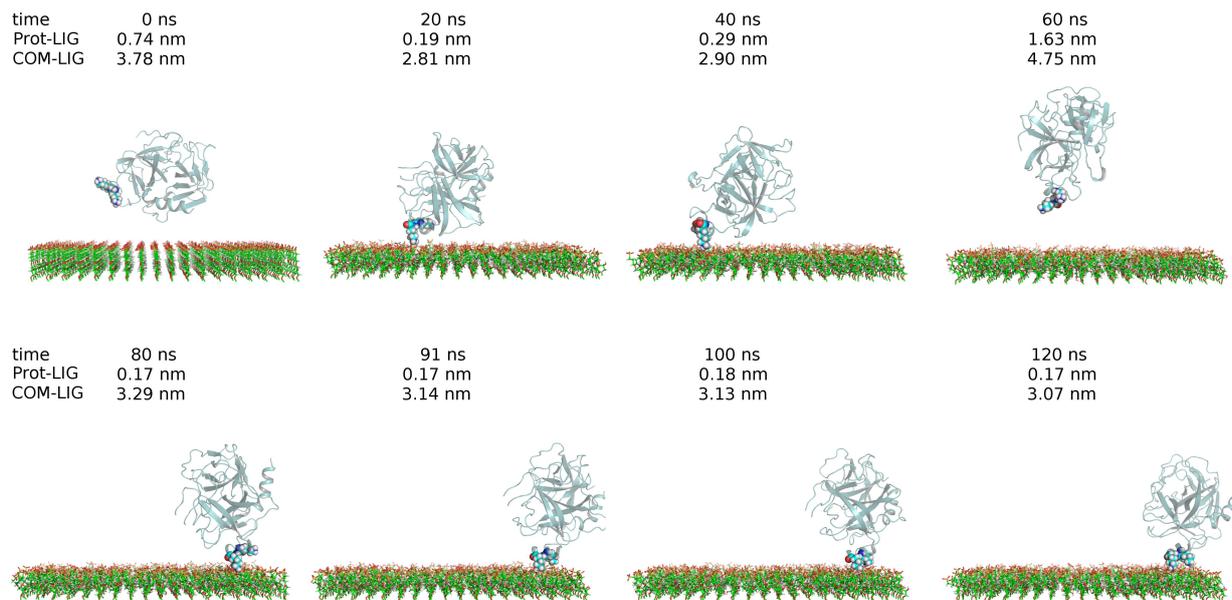


FIGURE IV.9 – Visualisation au cours du temps de l’adsorption de l’ $\alpha$ -chymotrypsine sur la SP Sepharose lors de la simulation MD2-O6. Pour chaque cliché, le temps de simulation et les distances PROT-LIG et COM-LIG sont données. Les résidus Lys169 et Lys170, correspondant aux principaux résidus impliqués dans l’adsorption, sont représentés sous forme de sphères, le reste de la protéine étant montrée sous une représentation *cartoon*.

## IV.2.2 Étude de l’interaction protéine-ligands

### IV.2.2.1 Identification des *patches* d’adsorption

Étant donné que toutes les simulations réalisées ont conduit à une adsorption de la protéine sur la surface chromatographique, il est possible de définir les résidus (chargés positivement) impliqués dans l’interaction. Pour cela, durant les 10 dernières nanosecondes de chaque simulation, la distance entre chaque résidu et l’ensemble des ligands est calculée. Ainsi, les résidus possédant une distance inférieure à 0,2 nm de la surface chromatographique sont considérés comme en interaction. L’ensemble des résidus impliqués est appelé un *patch* d’adsorption, ou encore *binding patch*. A pH 5, les résidus titrables positivement chargés sont principalement les Lysines et les Arginines.

La Figure IV.10 présente les différents résidus positifs identifiés comme participant à l’interaction. Leur taux de présence à proximité des ligands, qui correspond à la fraction de temps pendant laquelle la distance résidu-ligands est inférieure à 0,2 nm, est donné. En effet, il peut arriver qu’au cours de la simulation, certains résidus se rapprochent des ligands sans pour autant rester en interaction. D’autres résidus ont également été identifiés comme étant proches des ligands pendant la phase d’adsorption mais ne donnaient pas lieu à des interactions électrostatiques. Par exemple, la Sérine-92, qui n’est pas chargée, se

retrouve à une distance inférieure à 0,2 nm des ligands à cause de sa proximité immédiate de la Lysine-93, qui est une lysine participant à l'interaction. Il arrive également que des résidus puissent former une liaison hydrogène entre leur groupe hydroxyle et les atomes d'oxygène du groupe sulfopropyl des ligands. Tous ces résidus ne sont pas considérés comme résidus liants. En revanche, il faut noter que l'Alanine-149 a été impliquée dans certaines adsorptions et comptée comme résidu liant. En effet, Ala149 correspond au N-terminal de la troisième chaîne de l' $\alpha$ -chymotrypsine qui se retrouve chargé positivement au pH de travail et forme un *patch* d'adsorption avec l'Arginine-145.

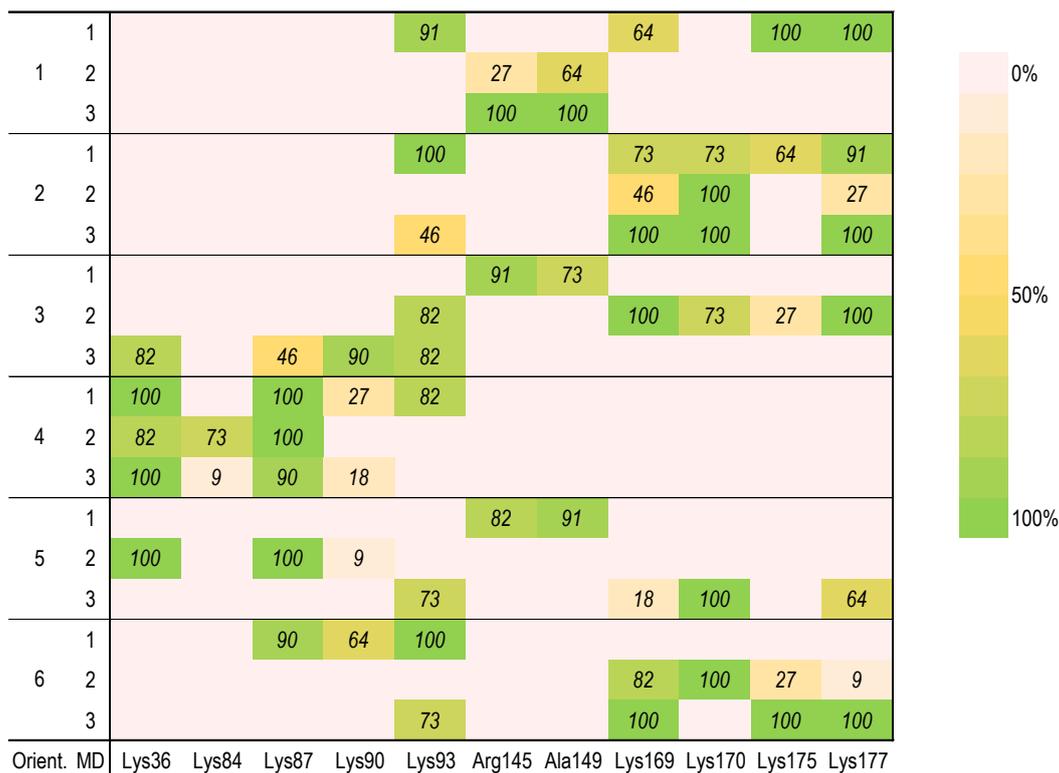


FIGURE IV.10 – Taux de présence durant les dernières 10 ns des simulations des résidus chargés positivement impliqués dans les adsorptions.

L'analyse de ces taux de présence souligne la présence de 4 *patches* récurrents, supposés les plus favorables à l'adsorption. Ces *patches* (notés P) peuvent être définis comme :

- P1 : Arg145 - Ala149
- P2 : Lys36 - Lys84 - Lys87 - Lys90
- P3 : Lys93
- P4 : Lys169 - Lys170 - Lys175 - Lys177

Ces quatre *patches* sont représentés en Figure IV.11, qui montre la protéine adsorbée à la fin (dernière *frame*) de quatre simulations. Il faut noter que plusieurs *patches* peuvent être impliqués dans une même adsorption. En effet, le choix a été fait de faire apparaître Lys93 dans un *patch* à part car du fait de sa localisation entre P2 et P4, il peut participer à

l'adsorption en même temps que P2 ou P4. Cela dépend majoritairement de la topologie de la protéine et de l'orientation que celle-ci a adopté sur la surface. Par exemple, la simulation MD1-O1 possède un *binding patch* comprenant à la fois P3 et P4. En revanche, du fait de la géométrie de la protéine, il n'est pas possible d'impliquer à la fois P2 et P4 dans une adsorption. De plus, le patch P1 est différencié des autres de par sa forme particulière, les résidus étant situés aux extrémités de deux chaînes. Malgré le fait que ce *patch* ne comprenne que deux résidus, il semblerait qu'il puisse avoir une certaine interaction avec les ligands car sa forme lui permet de pénétrer entre les ligands.

L'ensemble des résidus impliqués dans l'adsorption sont globalement localisés sur une même face de la protéine comme le montre la Figure IV.11.e. Cette face correspond d'ailleurs à la zone chargée positivement qui a été identifiée sur la Figure IV.1.a. Ainsi, comme attendu, cette zone de la surface de la protéine semble être la plus favorable à une interaction électrostatique avec le support chromatographique.

### IV.2.2.2 Énergies libres de liaison

Afin d'évaluer la force de l'interaction entre la protéine et les ligands durant la phase d'adsorption, des calculs d'énergies libres de liaison ont été réalisés. Pour cela, la méthode de calcul MM-PBSA (présentée dans le chapitre bibliographique, paragraphe II.4.1.6) a été utilisée. Pour chaque simulation, les valeurs d'énergies libres de liaison totales sont présentées dans le Tableau IV.3 ainsi que les différentes contributions, *i.e.* les énergies de van der Waals, électrostatiques, de solvation polaire et non-polaire.

Premièrement, il est important de noter que ces valeurs d'énergies libres seront essentiellement utilisées à but comparatif. En effet, il est reporté dans la littérature (Latour, 2008; Ozboyaci et al., 2016) que les valeurs numériques calculées par la méthode MM-PBSA sont en général surestimées. Néanmoins, cela permet de pouvoir comparer des simulations, des états du système différents et de déterminer lesquels paraissent les plus énergétiquement stables (autrement dit, lorsque l'énergie est la plus basse).

Une première analyse des valeurs d'énergies totales permet de constater qu'il existe effectivement des *patches* préférentiels, présentés dans les Tableaux IV.3 et IV.5. Ainsi, il est clairement visible que les *patches* identifiés précédemment contenant des lysines présentent les énergies les plus faibles, et donc les plus favorables. Il existe cependant quelques exceptions, comme la simulation MD1-O4, qui montre une valeur d'énergie plus haute que MD1-O6 malgré des *patches* d'adsorption similaires. Il faut également noter que les calculs d'énergies ont été effectués uniquement sur la dernière *frame* de chaque simulation en raison du temps de calcul élevé de la méthode MM-PBSA. Il aurait été idéal de pouvoir calculer les énergies durant les 10 dernières nanosecondes de chaque simulation avec un pas de temps réduit, afin d'en extraire une moyenne avec ses écarts-types. Il peut donc arriver que pour certaines simulations, l'énergie calculée à ce temps

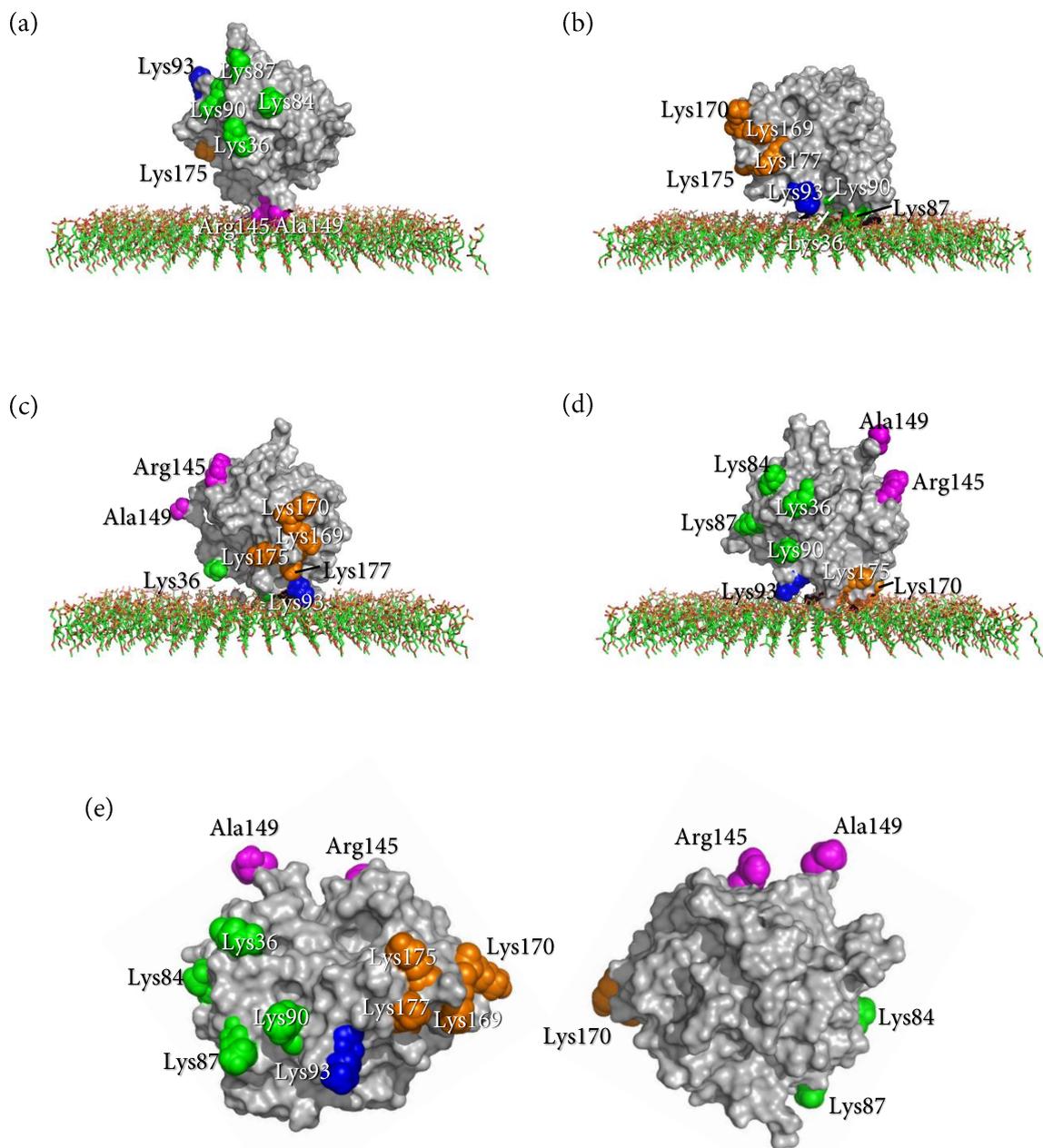


FIGURE IV.11 – Représentation de l' $\alpha$ -chymotrypsine adsorbée à la fin de 4 simulations, montrant ainsi les 4 *patches* d'adsorption identifiés. Les clichés sont issus des simulations : (a) MD3-O1, (b) MD3-O4, (c) MD1-O6 et (d) MD3-O2. Les 4 *patches* sont mis en évidence par des couleurs : P1 (magenta), P2 (vert), P3 (bleu) et P4 (orange). La Figure (e) montre la protéine et ses 4 *patches* sous les mêmes vues que sur la Figure IV.1.

TABLEAU IV.3 – Valeurs d'énergies libres totales et des différentes contributions, issues des calculs par la méthode MM-PBSA pour chacune des 18 simulations.

Orientation	Répétition MD	Patches d'adsorption	Energie libre totale (kJ/mol)	Energie de VDW (kJ/mol)	Energie électrostatique (kJ/mol)	Energie de solvation polaire (kJ/mol)	Energie de solvation non-polaire (kJ/mol)
1	1	P3 – P4	-2913	-135	-5907	3106	24
	2	P1	-2225	-76	-3747	1576	22
	3	P1	-2589	-83	-4145	1607	32
2	1	P3 – P4	-4430	-178	-7485	3225	8
	2	P4	-2764	-44	-4867	2126	21
	3	P3 – P4	-3290	-162	-6243	3101	14
3	1	P1	-2128	-76	-3715	1606	57
	2	P3 – P4	-4830	-142	-7756	3045	23
	3	P2 – P3	-3564	-70	-5389	1874	21
4	1	P2 – P3	-1834	-106	-4513	2777	9
	2	P2	-2477	-85	-4905	2494	19
	3	P2	-2554	-138	-4786	2360	10
5	1	P1	-1540	-70	-3204	1708	25
	2	P2	-1914	-130	-4258	2439	35
	3	P3 – P4	-2559	-192	-5094	2704	23
6	1	P2 – P3	-4921	-121	-7242	2423	19
	2	P4	-3100	-91	-4884	1861	13
	3	P3 – P4	-2830	-85	-5469	2687	38
Valeurs moyennes			-2915 ± 980	-110 ± 41	-5200 ± 1303	2373 ± 567	23 ± 12

final ne soit pas totalement représentatif de la phase d'adsorption. Enfin, il est également possible que malgré la stabilité apparente de l'adsorption selon les critères définis dans le paragraphe IV.2.1.1, ce ne soit pas tout à fait le cas et qu'une autre orientation plus favorable existe. Cette observation montre également l'importance d'un échantillonnage correct du système afin de pouvoir observer des tendances globales. En effet, les *patches* d'adsorption les plus probables semblent être une combinaison du patch P3 avec P4 voire P2.

Le Tableau IV.3, qui présente également les différentes contributions à l'énergie libre de liaison, montre que les effets électrostatiques sont, comme attendu, les interactions dominantes. Le terme électrostatique, calculé grâce à la loi de Coulomb, représente les interactions électrostatiques dans le vide entre la protéine et les ligands et apporte la plus grande contribution à l'énergie totale. L'énergie de solvatation polaire quant à elle permet de contre-balancer ces effets car cela traduit le passage dans le solvant à partir du vide du complexe protéine-ligand. Pour autant, les interactions électrostatiques restent suffisamment importantes pour rester prédominantes dans l'adsorption. Même s'il paraît logique que lors du mécanisme d'échange d'ions, la contribution électrostatique à l'énergie de liaison soit dominante, il est également possible que ces effets soient largement surestimés du fait du nombre importants de ligands dans le système. Traditionnellement, les méthodes MM-PBSA ont été développées afin d'étudier l'interaction entre une protéine et un seul ligand, comme cela est décrit dans le chapitre bibliographique (section II.4.1.6). Ici, lors de l'utilisation de l'outil de calcul *g\_mmpbsa*, ce n'est pas un ligand qui a été considéré mais l'ensemble des ligands avec les contre-ions. Le choix a été fait de conserver les contre-ions  $\text{Na}^+$  se trouvant à proximité immédiate de la surface chromatographique afin de neutraliser les ligands, et ainsi minimiser leur impact dans le calcul de l'énergie électrostatique. Une autre solution aurait été de calculer les énergies de liaisons uniquement entre la protéine et l'ensemble des ligands en interaction. Au final, cette technique a permis de réduire d'un facteur 10 environ la valeur calculée de l'énergie électrostatique.

## IV.2.3 Comparaison des résultats expérimentaux et des simulations

### IV.2.3.1 Charge caractéristique et facteur stérique

Les deux paramètres de la loi SMA, estimés à la fois grâce à des expériences macroscopiques et à des simulations moléculaires, ont été comparés. Cela est possible grâce à la signification physique de ces paramètres qui permet leur détermination à l'échelle macroscopique et à l'échelle moléculaire. Il s'agit donc d'un moyen de pouvoir confronter

les résultats issus des simulations et de vérifier la robustesse des calculs, afin de pouvoir par la suite utiliser les simulations dans un but prédictif.

Les résultats des calculs des paramètres SMA pour l' $\alpha$ Chy, à pH 5 et  $C_{NaCl} = 100$  mM, sont présentés dans le Tableau IV.4. Deux moyennes ont été calculées à partir des valeurs issues des simulations : une moyenne non pondérée et une moyenne pondérée basée sur les valeurs d'énergies libres. Cette moyenne permet de prendre en compte la probabilité d'apparition d'un *patch* d'adsorption, en supposant ainsi que les *patches* possédant les énergies libres les plus favorables soient les orientations prédominantes d'adsorption. De plus, afin de pouvoir mieux étudier les valeurs des paramètres SMA, le Tableau IV.5 récapitule les différentes données issues des analyses des simulations, à savoir les résidus participant à l'adsorption, les charges caractéristiques, les facteurs stériques et les énergies libres totales.

La charge caractéristique  $n$ , estimée à 5,1 à partir des isothermes déterminées expérimentalement, présente une bonne cohérence avec la valeur moyenne non pondérée, qui est de  $4,7 \pm 1,6$ . L'écart-type semble important, mais il est majoritairement dû à certaines exceptions présentant des *patches* particuliers, comme par exemple le P2 qui présente une interaction plus faible que les autres *patches*, ou encore certaines adsorptions impliquant de nombreux résidus. Malgré cela, la valeur moyenne semble bien représenter la majorité des charges calculées, ce qui signifie qu'il existe bien des orientations préférentielles d'adsorption. Cette observation est confirmée par le calcul de la moyenne pondérée avec l'énergie, qui tend à se rapprocher encore plus de la valeur issue des expériences (4,9). Il n'est donc pas incohérent de supposer que l'orientation optimale pour la protéine lors de l'adsorption montre l'énergie de liaison la plus faible, il semblerait également que ce soit la plus probable. Ainsi, selon les calculs d'énergies (Tableau IV.3), les adsorptions impliquant P3 et P4 paraissent les plus stables.

Les valeurs du facteur stérique  $\sigma$  issues des simulations sont en revanche plus éloignées de celle issue des expériences. L'ordre de grandeur entre les deux méthodes est toutefois le même et les résultats sont en bon accord. Comme dans le cas de la charge caractéristique, la moyenne pondérée est plus proche de la valeur expérimentale même si le changement est ici bien moins significatif. La différence entre ces deux valeurs de  $\sigma$  pourrait s'expliquer en grande partie par l'incertitude sur la surface spécifique de la SP Sepharose FF. En effet, lors de la préparation des simulations, la surface chromatographique a été construite selon une densité de ligands calculée à partir de cette surface spécifique, dont la valeur est issue de la littérature (DePhillips et Lenhoff, 2000). Or, il est probable que cette valeur soit éloignée de la réalité, et ce pour plusieurs raisons. La première est que la surface spécifique d'une résine est usuellement mesurée par des méthodes expérimentales telles que la méthode BET (Brunauer-Emmett-Teller). Comme indiqué dans le paragraphe II.3.2.3, cette méthode est difficilement applicable à des matrices de type agarose qui peuvent subir des altérations importantes lors de l'étape de séchage. C'est pour cela

TABLEAU IV.4 – Charges caractéristiques  $n$  et facteurs stériques  $\sigma$  calculés à partir des simulations. Deux moyennes (non pondérée et pondérée sur l'énergie libre) ainsi que leurs écarts-types ont été calculés et comparés aux valeurs issues des expériences.

Orientation	Répétition MD	Charge	Facteur stérique
1	1	6,8	42,8
	2	1,0	56,6
	3	4,6	52,8
2	1	6,5	45,5
	2	3,3	44,0
	3	5,9	45,1
3	1	2,6	52,7
	2	6,4	44,3
	3	5,6	51,5
4	1	6,0	49,4
	2	4,2	52,5
	3	4,2	53,5
5	1	2,6	55,3
	2	4,4	50,7
	3	4,4	46,9
6	1	4,6	51,4
	2	4,1	47,7
	3	7,0	41,2
Valeurs moyennes issues des simulations	Non pondérée	$4,7 \pm 1,6$	$49,1 \pm 4,6$
	Pondérée	$4,9 \pm 1,5$	$48,5 \pm 4,3$
Valeurs issues des expériences		5,1	28,3

que DePhillips et Lenhoff (2000) ont tenté d'appliquer une méthode de chromatographie d'exclusion de taille inverse, permettant d'obtenir la valeur utilisée dans ce travail. Ainsi, il est difficile de conclure de la pertinence de cette valeur de surface spécifique pour le calcul de la densité des ligands. La deuxième raison peut résulter du fait que les mesures de surface ont été réalisées il y a deux décennies et que la fabrication de la résine ait légèrement évolué. Enfin, il est également probable que lors de ces mesures, la surface accessible par les molécules utilisées ne correspondent pas à celle réellement accessible par une protéine, qui possède une taille différente. Il semblerait donc que la densité

TABLEAU IV.5 – Tableau récapitulatif des analyses des 18 simulations réalisées en système mono-constituants, pH 5 et concentration en NaCl de 100 mM.

Orientation	Répétition MD	Patches d'adsorption	Charge	Facteur stérique	Energie libre totale (kJ/mol)
1	1	Lys93 – Lys169 – Lys175 – Lys177 (P3-P4)	6,8	42,8	-2913
	2	Arg145 – Ala149 (P1)	1,0	56,6	-2225
	3	Arg145 – Ala149 (P1)	4,6	52,8	-2589
2	1	Lys93 – Lys169 – Lys170 – Lys175 – Lys177 (P3-P4)	6,5	45,5	-4430
	2	Lys169 – Lys170 – Lys177 (P4)	3,3	44,0	-2764
	3	Lys93 – Lys169 – Lys170 – Lys177 (P3-P4)	5,9	45,1	-3290
3	1	Arg145 – Ala149 (P1)	2,6	52,7	-2128
	2	Lys93 – Lys169 – Lys170 – Lys175 – Lys177 (P3-P4)	6,4	44,3	-4830
	3	Lys36 – Lys87 – Lys90 – Lys93 (P2-P3)	5,6	51,5	-3564
4	1	Lys36 – Lys87 – Lys93 (P2-P3)	6,0	49,4	-1834
	2	Lys36 – Lys84 – Lys87 (P2)	4,2	52,5	-2477
	3	Lys36 – Lys87 (P2)	4,2	53,5	-2554
5	1	Arg145 – Ala149 (P1)	2,6	55,3	-1540
	2	Lys36 – Lys87 (P2)	4,4	50,7	-1914
	3	Lys93 – Lys170 – Lys177 (P3-P4)	4,4	46,9	-2559
6	1	Lys87 – Lys90 – Lys93 (P2-P3)	4,6	51,4	-4921
	2	Lys169 – Lys170 (P4)	4,1	47,7	-3100
	3	Lys93 – Lys169 – Lys175 – Lys177 (P3-P4)	7,0	41,2	-2830

de ligands calculée pour la construction numérique du support chromatographique soit surestimée. Une autre source d'erreur peut également provenir de la mesure de la capacité ionique totale de la résine, qui a été déterminée expérimentalement (cf paragraphe III.1.1). Cependant, l'écart-type issu de ces mesures de capacité suggèrent une bonne répétabilité des expériences et ne semblent pas être la source la plus importante d'incertitudes. Les travaux de Gadam et al. (1993) montrent par ailleurs que la capacité ionique est un paramètre important à prendre en compte car elle influe grandement sur la valeur du facteur stérique. En effet, leur valeur de facteur stérique, déterminée avec une résine possédant une capacité deux fois supérieure à celle de la SP Sepharose, est également deux fois plus importante.

A la fois pour  $n$  et  $\sigma$ , une autre source d'erreur peut être liée à la méthode de détermination des paramètres SMA par la méthode des moindres carrés (cf paragraphe III.3.2.3). Comme discuté dans le chapitre expérimental, cette méthode ne permet pas d'obtenir un intervalle de confiance sur les valeurs ajustées. C'est une méthode mathématique qui mène au meilleur ajustement possible, mais d'autres paramètres peuvent également montrer un bon ajustement en considérant toutes les incertitudes de mesures. De plus, les résultats issus des simulations sont également dépendants du champ de force choisi, de l'échantillonnage conformationnel mais également de l'environnement considéré. Les sels composant la solution tampon ne sont par exemple pas pris en compte ici car ce sont des molécules complexes, et la force ionique a été traduite par une concentration équivalent en NaCl.

#### IV.2.3.2 Constante d'équilibre

Selon la définition de l'énergie libre présentée dans le paragraphe II.5, il est possible de calculer une valeur d'énergie libre à partir des données expérimentales. Ainsi, à pH 5, la constante d'équilibre de la réaction a été mesurée comme étant égale à 11, l'énergie libre peut donc être estimée selon :

$$\Delta G^\circ = -RT \ln K = -8,314 \times 293 \times \ln 11 = -5,841 \text{ kJ/mol} \quad (\text{IV.1})$$

Malgré toutes les hypothèses prises en compte (notamment les activités en solution égales à 1 dans le calcul de  $K$ ), les valeurs d'énergies libres calculées par le biais des expériences et celles issues des simulations sont radicalement différentes (écart d'un facteur 1000 environ). Outre le fait que les calculs de MM-PBSA sont généralement surestimés, il semblerait que ces calculs issus des simulations ne correspondent pas tout à fait à la réalité physique du système. Lors d'un échange d'ions, une multitude de protéines se trouvent en solution et/ou sur l'adsorbant et peuvent influencer la force de l'interaction. Or, tout cet environnement n'est pas réellement pris en compte dans les calculs autrement que par le biais des conditions périodiques, ce qui reste limité et assez peu représentatif des

conditions réelles. De plus, la constante d'équilibre estimée expérimentalement représente bien un équilibre global du système, alors que les calculs de MM-PBSA sont focalisés sur une seule molécule adsorbée.

Finalement, si ces calculs d'énergies libres semblent être pertinents pour mieux comprendre le mécanisme d'adsorption d'une protéine sur un support, ils ne paraissent en revanche pas représentatifs de l'équilibre global lors d'un échange d'ions.

## IV.2.4 Conclusion

Des simulations de dynamique moléculaire ont été réalisées afin de mieux comprendre le comportement de l' $\alpha$ -chymotrypsine lors de l'échange d'ions sur la résine SP Sepharose FF, dans des conditions de pH 5 et avec une concentration en chlorure de sodium égale à 100 mM. Afin d'échantillonner au mieux le système et d'acquérir des données représentatives de l'échelle macroscopique, six orientations initiales de la protéine ont été étudiées et trois simulations ont été réalisées pour chacune d'elles. Ainsi, les analyses de ces 18 simulations ont permis d'estimer les paramètres de la loi d'action de masse stérique, qui ont également été mesurées expérimentalement à l'échelle macroscopique.

D'une part, quatre *patches* d'adsorption (résidus participant à l'adsorption) ont pu être identifiés comme étant la forme prédominante d'adsorption, ce qui a été confirmé par des calculs d'énergies libres de liaison en utilisant la méthode MM-PBSA. D'autre part, la comparaison des paramètres SMA a montré une bonne concordance, spécialement pour la charge caractéristique et le facteur stérique, même si la prédiction de ce dernier reste à améliorer. En effet, une incertitude existe quant à la détermination de la densité de ligands sur la surface chromatographique, qui se répercute sur la valeur du facteur stérique. Enfin, il apparaît que la constante d'équilibre de la loi SMA ne puisse pas être prédite par les simulations, et en particulier par les calculs d'énergies libres de liaison. Les valeurs d'énergies calculées par la méthode MM-PBSA sont très surestimées et ne permettent pas en l'état de retrouver une valeur de la constante d'équilibre similaire à celle obtenue expérimentalement.

Au vu de ces premiers résultats, il semblerait que les simulations à l'échelle atomique permettent de mieux comprendre les mécanismes d'adsorption, et en particulier de prédire certains paramètres macroscopiques directement liés à la protéine. En effet, des trois paramètres de la loi SMA, la charge caractéristique a été mieux prédite car elle dépend quasi exclusivement de la protéine, qui est bien connue et modélisée. Le facteur stérique quant à lui dépend non seulement de la protéine mais également de la surface chromatographique, et même si les valeurs comparées sont du même ordre de grandeur, il reste encore des ajustements à faire. Finalement, le paramètre le plus compliqué à prédire est la constante d'équilibre, probablement parce qu'elle ne dépend pas seulement d'une pro-

téine, mais de tout un environnement complexe (multitude de protéines en solution, force ionique, etc.) qui est difficile à représenter à une si petite échelle.

## IV.3 Influence de l'environnement

Lors de cette étude, le système mono-constituant composé de l' $\alpha$ -chymotrypsine et de la résine SP Sepharose FF a été conservé ainsi que toute la méthodologie mise en place précédemment. Seuls la force ionique et le pH ont été modifiés consécutivement afin de pouvoir identifier les effets d'un changement de ces deux paramètres sur l'adsorption de l' $\alpha$ Chy.

### IV.3.1 Effet de la force ionique

En chromatographie d'échange d'ions, il est primordial de contrôler la force ionique de la solution car elle influe grandement sur la capacité de rétention d'une protéine sur une résine. Comme cela a été montré dans le chapitre expérimental, la capacité de rétention décroît avec l'augmentation de la force ionique, jusqu'au moment où la compétition entre la protéine et les ions en solution est trop forte et la protéine ne peut plus se fixer. Dans le cas du système  $\alpha$ Chy – SP Sepharose, le tracé de la courbe de percée dans des conditions de pH 5 (tampon citrate 50 mM) avec ajout de 400 mM de NaCl (voir paragraphe III.3.2.3) a montré que la protéine ne s'était pas fixée, ou du moins pas en quantité suffisante pour être détectable lors des analyses. De plus, la mesure du gradient d'élution et de la conductivité au point d'élution a confirmé cette observation, le pic d'élution se trouvant à une concentration en NaCl d'environ 270 mM.

Pour cette raison, quelques simulations ont été mises en place avec une concentration en ions sodium  $\text{Na}^+$  égale à 500 mM, ce qui correspond à la totalité des ions  $\text{Na}^+$  présents dans le sel (400 mM) et dans le tampon citrate (100 mM). Six simulations ont été réalisées, soit une simulation par orientation initiale de la protéine, et le temps de simulation total a été fixé à 200 ns pour chacune d'elles. Les principaux résultats sont présentés dans le Tableau IV.6, *i.e.* le temps d'adsorption, la charge caractéristique, le facteur stérique et l'énergie libre totale de liaison.

La première constatation est que sur les six simulations réalisées, la moitié a abouti à une adsorption de la protéine sur la surface avant 200 ns. Il faut cependant noter que dans le Tableau IV.6, les charges, facteurs stériques et énergies libres sont calculés uniquement à titre indicatif et comparatif car ils ne peuvent pas être confrontés à des données expérimentales, aucune adsorption n'ayant pu être mesurée. De plus, les moyennes de ces paramètres ont été calculées sur seulement trois simulations, qui peuvent ne pas être

TABLEAU IV.6 – Tableau récapitulatif des analyses des 6 simulations réalisées en système mono-constituant, pH 5 et concentration en NaCl de 500 mM.

Orientation	Répétition MD	Patches d'adsorption	Temps d'adsorption (ns)	Charge	Facteur stérique	Energie libre totale (kJ/mol)
1	1	NA	NA	NA	NA	NA
2	1	Lys93 – Lys169 – Lys175 – Lys177 (P3-P4)	5	8,7	39,8	-544
3	1	NA	NA	NA	NA	NA
4	1	Lys36 – Lys84 – Lys87 (P2)	60	6,3	49,1	-102
5	1	Lys93 – Lys169 – Lys170 – Lys175 – Lys177 (P3- P4)	150	7,5	42,5	-1108
6	1	NA	NA	NA	NA	NA
Valeurs moyennes				7,5 ± 1,2	43,8 ± 4,8	-585 ± 505

représentatives du système global. Il est néanmoins intéressant de remarquer qu'une augmentation de la concentration en sel dans la boîte de simulation influe bien sur la capacité de rétention de la protéine, qui a du mal à se fixer. En effet, pour certaines simulations, l'analyse de la distance minimale PROT-LIG au cours du temps (Figure IV.12) montre que l' $\alpha$ Chy a tenté de s'adsorber mais s'est finalement éloignée de la surface, peut-être à cause de la trop forte compétition avec les ions en solution. Contrairement à ce que montrent les résultats expérimentaux, la protéine réussit tout de même à se fixer sur la surface chromatographique dans la moitié des cas.

D'une part, cela semble confirmer que les effets électrostatiques sont potentiellement surestimés lors des simulations, ce qui peut conduire à une surévaluation des probabilités d'adsorption. En effet, l'effet de l'environnement ne peut être représenté correctement avec une seule protéine, cette simplification pouvant amplifier les effets électrostatiques sur la protéine considérée. Par ailleurs, les énergies libres de liaison, bien que supérieures à celles obtenues à 100 mM, restent favorables à une interaction. D'autre part, il est probable que la protéine soit effectivement retenue en très petite quantité, mais qui serait trop faible pour être mesurable expérimentalement. Il est également possible que la protéine interagisse avec la surface mais que le système obtenu ne soit pas assez stable et donc que la protéine finisse par se désorber du support. Enfin, il faut également souligner qu'il existe un aspect qui n'est pas pris en compte dans les simulations qui est le transfert de matière de la protéine jusqu'au support chromatographique. Lors des simulations, la protéine est placée à une distance très proche de la surface de manière à étudier uniquement le mécanisme de rétention, et le transport de la protéine est négligé. Or, expérimentalement, la grande quantité d'ions en solution écrantant le champ électrique généré par le support peuvent entraver la bonne diffusion des protéines et ainsi les empêcher de se rapprocher suffisamment de la surface chromatographique et donc de s'adsorber (empêchant l'effet Donnan de se produire (Gregor, 1951)).

### IV.3.2 Effet du pH

Le pH de la solution est un paramètre clé en chromatographie ionique car il conditionne les états de protonation des résidus titrables de la protéine. Il peut donc être pertinent d'étudier l'effet du pH à l'échelle locale sur le mécanisme d'adsorption d'une protéine.

Pour rappel, à ce pH, la protéine est globalement moins chargée qu'à pH 5 (charge de +5 au lieu de +7) mais semble présenter une face toujours très chargée positivement, comme discuté dans le paragraphe IV.1.1.1. Expérimentalement, l' $\alpha$ Chy a montré une moins bonne affinité avec la SP Sepharose FF à pH 7, ce qui s'est traduit au niveau des paramètres SMA par une forte baisse de la constante d'équilibre  $K$ . Par ailleurs, afin d'obtenir un bon ajustement des paramètres SMA avec seulement deux isothermes,

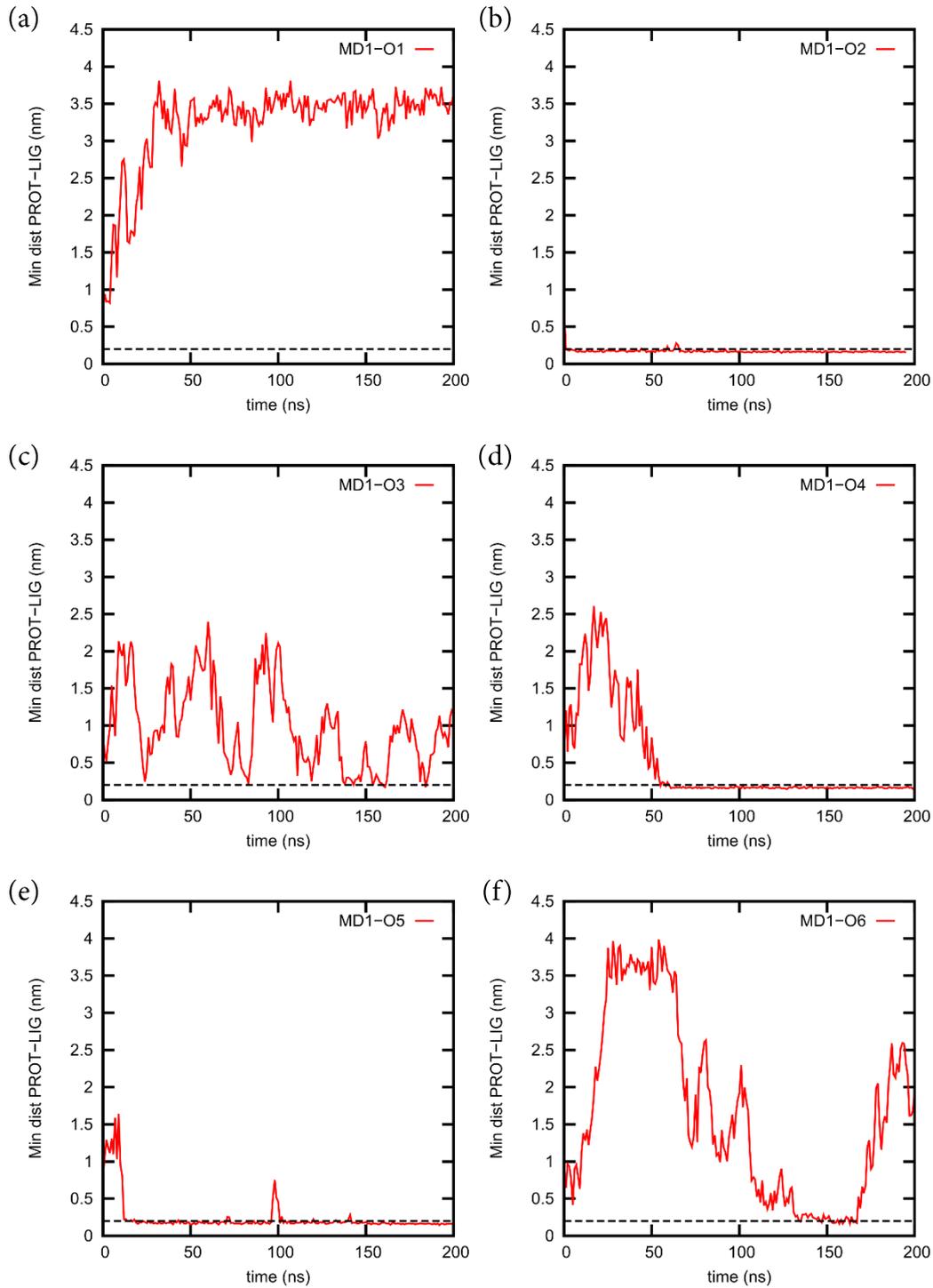


FIGURE IV.12 – Évolution de la distance minimale entre les ligands et la protéine *PROT-LIG* au cours du temps pour chaque orientation (O1-6, figures (a) à (f)). Cas d'étude à pH 5 et  $C_{NaCl} = 500$  mM.

l'hypothèse de considérer le facteur stérique à pH 7 égal à celui à pH 5 a été posée. Ainsi, la charge caractéristique calculée dans ces conditions s'est avérée très similaire à celle à pH 5.

#### IV.3.2.1 Configuration du système

Douze simulations avec l' $\alpha$ -chymotrypsine ont été réalisées à pH 7, qui est un pH compris entre le pH de référence fixé à 5 et le pI de la protéine. Les douze simulations correspondent à deux simulations différentes pour chacune des six orientations initiales définies précédemment. Une concentration en NaCl de 100 mM a été conservée, afin de pouvoir comparer les résultats avec ceux obtenus à pH 5. Ainsi, le protocole appliqué pour les simulations à pH 5 est conservé, la seule différence réside dans la structure de la protéine et l'assignation des hydrogènes afin de respecter les états de protonation des résidus. Toutes les durées de simulations ont été fixées à 100 ns.

#### IV.3.2.2 Résultats

Les principaux résultats obtenus par la simulation à pH 7, *i.e.* le temps d'adsorption, la charge caractéristique, le facteur stérique et l'énergie libre totale de liaison, sont présentés dans le Tableau IV.7. Tout d'abord, à l'exception d'une simulation qui n'a pas abouti à une adsorption durant les 100 ns (MD1-O3), la protéine ne semble pas avoir de difficulté à s'adsorber malgré le changement de pH. Les *patches* d'adsorption identifiés précédemment sont conservés, ce qui est un résultat logique compte-tenu de l'état de protonation des lysines. En effet, en dessous d'un pH de 9,74 (voir Tableau II.1), les lysines se trouvent sous une forme protonée. Or, elles composent les trois *patches* d'adsorption P2-P3-P4 qui paraissent être les patches préférentiels d'adsorption. Il est cependant intéressant de noter que contrairement aux résultats à pH 5, le *patch* P2 semble être prédominant (8 simulations sur les 12 au total). Il semblerait également que l'orientation de départ de la protéine dans la boîte influence la position de la protéine finale, car ce sont essentiellement les *patches* P3-P4 qui sont retrouvés pour les trois premières orientations. Enfin, l'orientation 3 semble ne pas être favorable à l'adsorption car l'une des deux simulations n'a pas abouti à une adsorption. Cette simulation n'ayant pas été prolongée, la durée fixée de 100 ns n'a pas été suffisante pour que la protéine puisse correctement se réorienter pour ensuite s'adsorber.

Les paramètres SMA calculés à partir des simulations ont été comparés à ceux obtenus expérimentalement. De la même manière qu'à pH 5, les charges caractéristiques calculées à partir des deux méthodes sont en bon accord. Ce n'est pas nécessairement le cas pour le facteur stérique pour l'ensemble des raisons évoquées dans le paragraphe IV.2.3.1, même si les valeurs restent malgré tout dans le même ordre de grandeur.

TABLEAU IV.7 – Tableau récapitulatif des analyses des 12 simulations réalisées en système mono-constituant, pH 7 et concentration en NaCl de 100 mM.

Orientation	Répétition MD	Patches d'adsorption	Temps d'adsorption (ns)	Charge	Facteur stérique	Energie libre totale (kJ/mol)
1	1	Lys93 – Lys169 – Lys175 – Lys177 (P3-P4)	20	7,7	41,8	-1805
	2	Lys87 – Lys93 (P2-P3)	10	4,7	53,0	-1060
2	1	Lys93 – Lys169 – Lys170 (P3-P4)	30	3,3	47,1	-1958
	2	Lys93 – Lys169 – Lys177 (P3-P4)	10	6,0	42,4	-1331
3	1	NA	NA	NA	NA	NA
	2	Lys87 – Lys90 – Lys93 (P2-P3)	85	3,1	50,9	-1380
4	1	Lys36 – Lys63 – Lys79 – Lys87 (P2)	25	4,6	55,0	-380
	2	Lys36 – Lys63 – Lys79 – Lys87 (P2)	30	6,6	48,6	151
5	1	Lys36 – Lys87 (P2)	20	6,2	51,0	-652
	2	Lys36 – Lys79 – Lys84 – Lys87 (P2)	20	4,5	47,3	259
6	1	Lys36 – Lys87 (P2)	15	4,9	51,0	-772
	2	Lys87 – Lys90 – Lys93 (P2-P3)	50	7,6	55,2	-1192
Valeurs moyennes			29 ± 22	5,4 ± 1,6	49,4 ± 4,5	-920 ± 1227

Lors de l'étude expérimentale, il a également été mis en évidence que la charge et le facteur stérique étaient similaires pour les deux pH étudiés. Ce résultat semble être confirmé par les résultats des simulations, les valeurs des paramètres SMA montrant finalement peu de différences avec les valeurs précédemment obtenues à pH 5. La charge caractéristique est plus élevée, mais pas de façon significative compte-tenu de l'écart-type et du nombre plus limité de simulations (12 au lieu de 18). Au regard des *patches* d'adsorption impliqués, il paraît finalement logique que la charge soit quasiment la même car ce sont les mêmes lysines qui entrent en interaction avec le support, et interagissent en moyenne avec un même nombre de ligands. De même, le facteur stérique est très semblable à celui obtenu à pH 5, sa valeur dépendant essentiellement de la géométrie et de l'orientation de la protéine, qui reste inchangée. D'ailleurs, ce résultat tend à confirmer l'hypothèse posée lors du calcul des paramètres SMA expérimentaux.

Il est intéressant de constater que dans ce cas, la simulation moléculaire a permis de comprendre les phénomènes mis en jeu (*patches* d'adsorption identiques) afin de mieux comprendre les observations macroscopiques. Il serait en effet logique de penser que plus le pH se rapproche du pI et plus la charge, qui représente la force de liaison entre la protéine et le support, décroît. Or, si cela peut être vrai dans certaines conditions, ce n'est pas nécessairement le cas comme cela a été montré ici. Il est en effet important de considérer la répartition des charges à la surface de la protéine, et pas seulement un paramètre global comme le pI.

Malgré tout, il est nettement visible expérimentalement que l' $\alpha$ Chy possède moins d'affinité avec la résine à ce pH, ce qui se traduit notamment par une très faible valeur de la constante d'équilibre. Les simulations moléculaires ont quant à elles permis de montrer une différence de valeurs d'énergies libres (-2915 kJ/mol à pH 5 et -920 kJ/mol à pH 7), ce qui traduit bien la perte d'affinité entre la protéine et la résine avec l'augmentation du pH. Il existe cependant une forte dispersion des résultats car l'écart-type est très grand (1227 kJ/mol) et des énergies positives ont été calculées pour certains *patches*. Comme discuté précédemment, ces valeurs ne peuvent pas être exploitées telles quelles car elles sont en général largement surestimées. Il serait donc intéressant de pouvoir calculer plus précisément ces valeurs d'énergies et ainsi les relier aux valeurs de constantes d'équilibre. De la même manière que dans le paragraphe IV.2.3.2, il est possible de calculer théoriquement l'énergie libre de liaison à partir de la constante d'équilibre expérimentale  $K$ . Cependant, dans ce cas à pH 7, la constante  $K$  est inférieure à 1 ce qui conduirait à une énergie positive et donc défavorable à une interaction. Si pour certains *patches*, des énergies positives ont effectivement été calculées, l'écart-type est trop grand pour pouvoir conclure correctement de la pertinence de ces valeurs.

Ces différences observées entre les résultats expérimentaux et les analyses des simulations moléculaires peuvent s'expliquer de plusieurs raisons. Tout d'abord, il est possible

que la constante  $K$  issu de l'expérience ne soit pas exacte car elle représente uniquement l'ajustement de paramètres le plus adéquat, l'intervalle de confiance sur cette valeur n'ayant pas pu être calculé. Une autre raison pourrait être que les *patches* induisant une énergie de liaison positive soient les *patches* dominants lors de l'adsorption à pH 7, bien qu'ils ne soient pas les plus favorables énergétiquement. Enfin, il ne faut pas oublier que lors des simulations, l'effet du pH n'est pris en compte que par le biais de l'état de protonation des résidus de la protéine et le pH du solvant n'est pas explicitement modélisé, ce qui peut également conduire à des différences entre simulations et expériences à l'échelle macroscopique.

### IV.3.3 Conclusion

L'influence d'un changement de pH et de force ionique sur les mécanismes de rétention de l' $\alpha$ -chymotrypsine sur la SP Sepharose FF a été étudiée. Pour cela, plusieurs simulations ont été réalisées, dans un premier temps à pH 5 et avec une concentration en chlorure de sodium de 500 mM, et dans un second temps à pH 7 et  $C_{NaCl} = 100$  mM. Expérimentalement, lorsqu'une forte hausse de concentration en NaCl a été appliqué, aucune rétention n'a pu être mesurée. Le changement de pH, quant à lui, a eu pour effet d'affaiblir l'interaction entre la protéine et le support et donc de montrer que l'affinité est plus faible qu'à pH 5. L'objectif des simulations a donc été de vérifier la pertinence des simulations pour représenter les effets de l'environnement sur la rétention de la protéine.

En ce qui concerne l'effet de la concentration en sel, les simulations ont montré qu'une augmentation de la force ionique conduit bien à des adsorptions moins favorables. Malgré tout, l' $\alpha$ Chy s'est adsorbée dans environ 50% des cas, alors qu'aucune rétention n'a été observée expérimentalement. Dans tous les cas, plus de simulations seraient nécessaires pour mieux évaluer l'influence de cette variable, dont les effets sur l'échange d'ions sont divers. A pH 7, les résultats des simulations ont montré une adsorption quasi-systématique de la protéine, malgré des énergies d'interactions plus faibles qu'à pH 5. Le mode d'interaction, représenté par les valeurs de charge caractéristique et de facteur stérique, est d'ailleurs très similaires à celui observé à pH 5. Ce résultat va dans le sens des observations macroscopiques, qui ont permis de constater que ces deux paramètres SMA étaient très proches aux deux pH, et que seule la constante d'équilibre était bien plus faible à pH 7. Si cette constatation peut paraître surprenante, les analyses des simulations ont permis de mettre en évidence que les résidus participant à l'interaction sont les mêmes à ces deux pH car ils restent chargés positivement. Il est donc très intéressant de constater que le mécanisme d'adsorption est très dépendant de la structure et de la répartition des charges de la protéine. La perte d'affinité constatée macroscopiquement pourrait alors être attribuée à d'autres facteurs plus globaux liés à la complexité de l'environnement,

comme par exemple la quantité importante de protéines en solution, dont les effets ne sont actuellement pas pris en compte durant les simulations.

## IV.4 Étude d'un système bi-protéines

La chromatographie d'échange d'ions en système multi-protéines peut impliquer l'apparition de nouveaux phénomènes tels que la compétition entre protéines, qui dépendent essentiellement du système en question et peuvent difficilement être prédit en extrapolant des résultats issus d'études en mono-constituants. Le potentiel des simulations moléculaires appliquées à ces cas apparaît donc particulièrement important. Pour cette raison, des études ont été réalisées en système bi-protéines, où le lysozyme a été ajouté en plus du système  $\alpha$ -chymotrypsine – SP Sepharose FF. Cette étude est scindée en deux parties avec pour objectif d'étudier deux cas limites mis en évidence par les expériences : (i) une situation de faible compétition entre protéines et (ii) une situation de forte compétition, lorsque les protéines sont présentes en forte concentration.

### IV.4.1 Simulations avec deux protéines

Cette première étude se focalise sur la représentation du système  $\alpha$ Chy – Lysozyme dans un cas où la compétition entre protéines est relativement peu importante, c'est-à-dire un cas où les concentrations en protéines sont relativement faibles. Ce cas correspond globalement à la première partie des isothermes préalablement tracées (paragraphe III.3.3), où les protéines semblent se comporter de la même manière qu'en système mono-constituant.

Les paramètres habituellement calculés, à savoir la charge caractéristique, le facteur stérique et les énergies libres de liaison n'ont pas été calculés ici et seule une analyse qualitative a été réalisée. La loi SMA peut être transposable à des systèmes multi-protéines sous l'hypothèse que les protéines n'interagissent pas entre elles. Or, les mesures d'isothermes ont montré que ce n'était pas le cas ici avec l' $\alpha$ Chy et le lysozyme, qui pourraient s'adsorber en formant une multi-couche à partir d'une certaine concentration en protéines en solution. De ce fait, les paramètres SMA n'ont pas été déterminés expérimentalement. Les simulations peuvent en revanche permettre d'identifier les potentielles interactions pouvant avoir lieu à la surface du support chromatographique.

Enfin, il est nécessaire de préciser que le lysozyme seul n'a pas fait l'objet de simulations préalables au cours de ce travail. Les analyses ont été focalisées principalement sur les interactions entre les deux protéines et l'impact sur la rétention.

### IV.4.1.1 Configuration du système

Tout d'abord, la méthodologie appliquée aux cas d'études en système mono-constituant a été conservée, *i.e.* un pH égal à 5 et une concentration en chlorure de sodium égale à 100 mM. La différence principale entre les deux systèmes réside dans le nombre de protéines initialement introduit dans la boîte de simulation, et donc le nombre d'atomes total. Ainsi, en plus de l' $\alpha$ -chymotrypsine, le lysozyme a été placé dans la boîte de simulation, dont la taille et la forme sont inchangées. La distance entre chaque protéine et la surface chromatographique ainsi que la distance entre les protéines ont été fixées à 0,8 nm afin de ne pas privilégier l'adsorption de l'une des deux. La Figure IV.13 représente les deux configurations de départ qui ont été utilisées. Dans la configuration n°1 (Figure IV.13.a), l' $\alpha$ Chy est orientée selon l'orientation 1 définie précédemment (paragraphe IV.1.2.1) et le lysozyme est orientée de manière aléatoire. Dans la configuration n°2 (Figure IV.13.b), les deux protéines ont subi une rotation à  $180^\circ$  sur l'axe -y par rapport aux orientations définies dans la configuration 1. Au total, quatre simulations ont été réalisées, soit deux simulations par configuration. Toutes les durées de simulation ont été fixées à 200 ns.

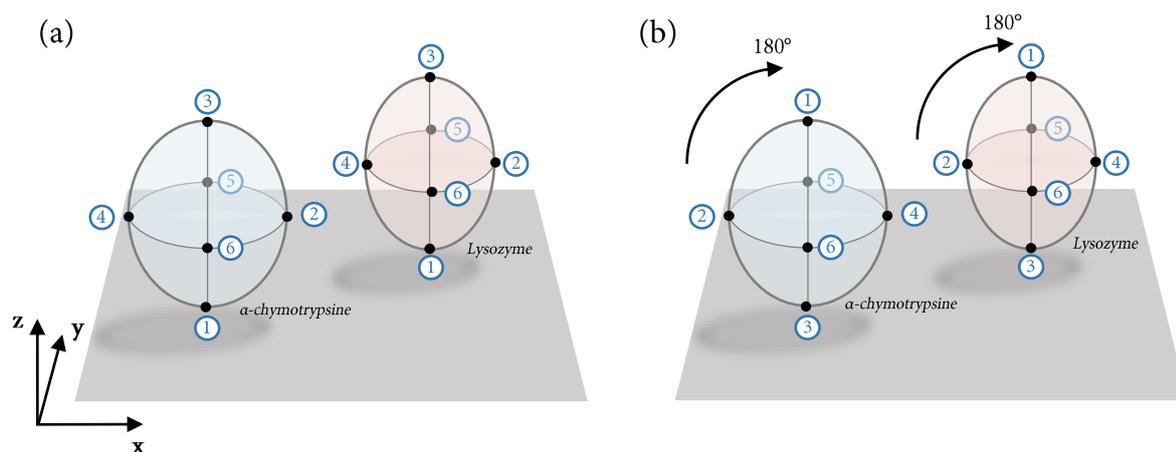


FIGURE IV.13 – Illustration des deux configurations de départ pour les simulations en système bi-protéines, représentant l' $\alpha$ -chymotrypsine en bleu et le lysozyme en rouge, sur la surface chromatographique (plan gris). La configuration n°1 (Figure a) montre l' $\alpha$ Chy avec l'orientation initiale 1 et le lysozyme avec une orientation aléatoire et la configuration n°2 (Figure b) montrent les deux protéines avec une rotation en  $180^\circ$  sur l'axe -y par rapport à la configuration 1.

Il faut noter qu'il est assez contraignant d'explorer l'ensemble des orientations possibles, comme cela a été fait avec l' $\alpha$ Chy seule : en effet, une exploration exhaustive des six orientations possibles pour chacune des protéines nécessiterait une combinaison de  $6 \times 6 = 36$  configurations différentes. Afin de limiter le nombre de simulations, le choix a été fait d'étudier uniquement ces deux configurations de départ afin d'obtenir une première idée du comportement de ces deux protéines.

### IV.4.1.2 Résultats

Les principaux résultats des quatre simulations sont présentées en Figure IV.14, qui montrent la dernière *frame* de chaque simulation. De plus, les distances minimales Protéine–Ligands et Protéine–Protéine ont été tracées et sont présentées sur cette même figure.

La Figure IV.14 montre tout d’abord que les deux protéines se sont adsorbées au cours des simulations de 200 ns. Cependant, si le lysozyme s’est adsorbé sur la surface chromatographique dans tous les cas, l’ $\alpha$ Chy s’est fixée deux fois sur la surface et deux fois sur le lysozyme déjà adsorbé, formant ainsi une multicouche. Ce résultat est très intéressant et semble confirmer la tendance à l’adsorption multicouche déjà supposée au regard de la forme des isothermes. Il faut également noter que c’est l’ $\alpha$ Chy qui semble se fixer sur le lysozyme et non l’inverse : le lysozyme s’est toujours fixé sur les ligands comme le montre l’évolution la distance Lysozyme–Ligands (courbes vertes sur la Figure IV.14). Cela semble logique étant donné la forte distribution de charges positives sur la surface du lysozyme. Expérimentalement, ce dernier semble également avoir plus d’affinité avec la résine que l’ $\alpha$ Chy.

D’autre part, bien que la charge globale de l’ $\alpha$ Chy soit de +7 à pH 5, il a été montré dans le paragraphe IV.1.1.1 que les charges ne sont pas réparties équitablement à la surface de la protéine, avec une face majoritairement neutre/négative et la face opposée majoritairement positive. Ainsi, cela explique en partie la capacité que possède cette protéine à former une multicouche avec le lysozyme. En effet, ce sont bien des résidus situés sur la face neutre/négative qui sont en interaction avec le lysozyme. Cependant, les graphiques de la Figure IV.14, et en particulier l’évolution de la distance minimale entre l’ $\alpha$ Chy et le lysozyme (courbes bleues), montrent que cette interaction n’est pas très stable dans le temps, contrairement à l’interaction entre protéines et ligands. A plusieurs reprises, lors des simulations MD2-CONF1 et MD1-CONF2, les deux protéines se sont brièvement éloignées avant de se rapprocher à nouveau.

Ce résultat paraît en accord avec les observations faites à l’échelle macroscopique par l’étude des isothermes : lorsque qu’il y a assez peu de protéines en solution, l’ $\alpha$ Chy se comporte comme en système mono-protéine et interagit avec le support chromatographique. En revanche, lorsque les concentrations en protéines augmentent et que les deux protéines se retrouvent en situation de compétition, le comportement de l’ $\alpha$ Chy change et il y a très probablement la formation d’une multicouche. Il est néanmoins difficile de se placer en réelle situation de compétition protéine-protéine durant les simulations car malgré la présence de deux protéines, il reste suffisamment de place sur la surface pour que les deux protéines puissent s’adsorber indépendamment l’une de l’autre.

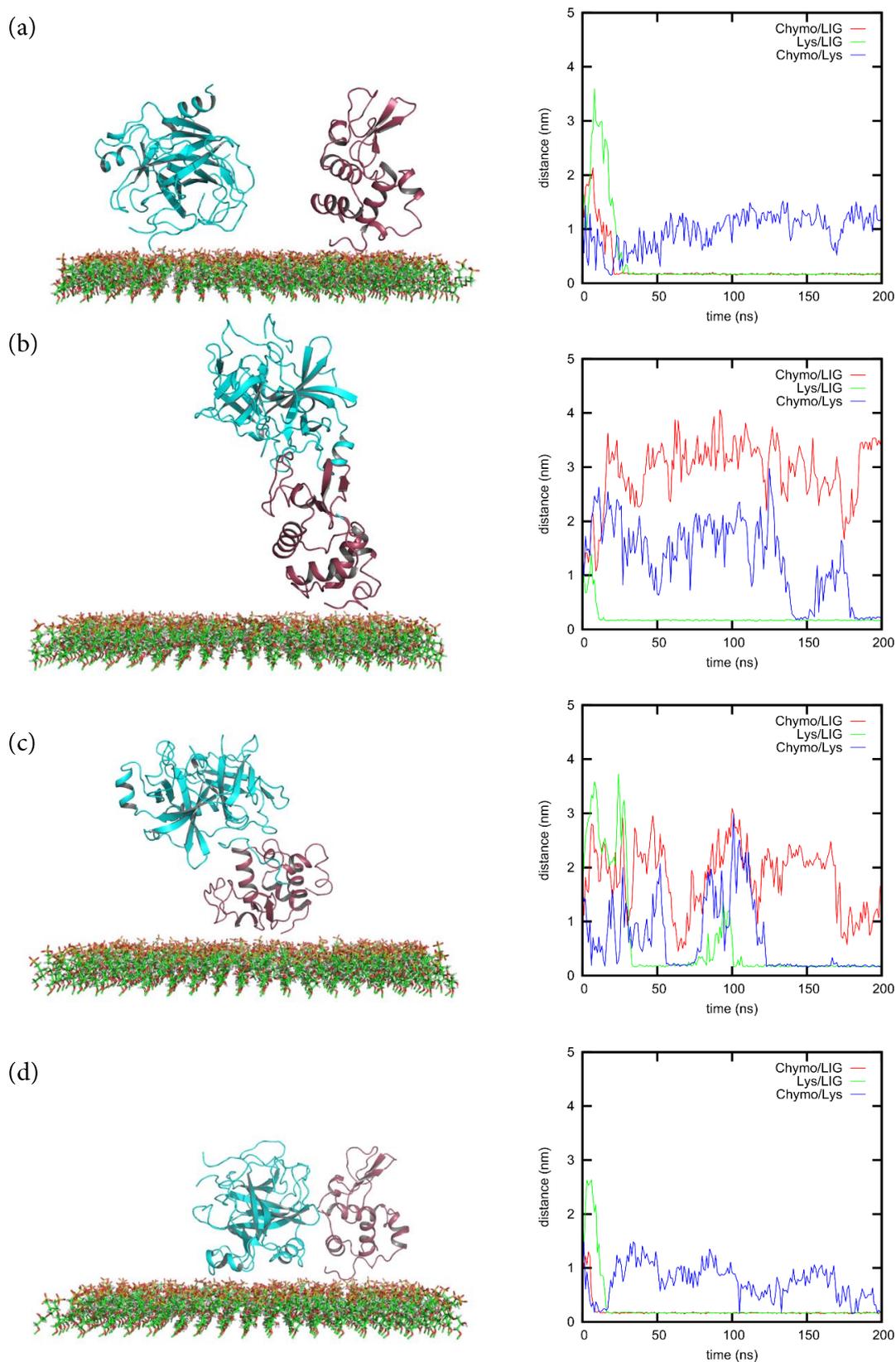


FIGURE IV.14 – Résultats des quatre simulations en système bi-protéines, comprenant l' $\alpha$ -chymotrypsine (en bleu cyan) et le lysozyme (en rouge), à pH 5 et  $C_{NaCl} = 100$  mM. La dernière *frame* de chaque simulation, *i.e.* (a) MD1-CONF1, (b) MD2-CONF1, (c) MD1-CONF2 et (d) MD2-CONF2, sont présentées ainsi que les distances minimales Protéines-Ligands et Protéine-Protéine.

## IV.4.2 Simulations avec quatre protéines

Cette deuxième partie de l'étude en système multi-constituants vise à représenter un cas où la compétition protéine–protéine est très forte. Pour ce faire, quatre protéines, plus précisément deux  $\alpha$ Chy et deux lysozymes, ont été introduites dans la boîte de simulation dont deux étant déjà préalablement adsorbées sur le support. L'intérêt des simulations comprenant quatre protéines dans une boîte de simulation restreinte est de simuler un cas de forte concentration en protéines correspondant à la partie de l'isotherme où une adsorption multicouche semble se produire (voir paragraphe III.3.3). Ce phénomène a déjà été mis en évidence avec les précédentes simulations, malgré le fait qu'il n'y ait que deux protéines dans la boîte de simulation.

### IV.4.2.1 Configuration du système

Un total de trois simulations a été réalisé avec les 4 protéines dans la boîte de simulation : 2 protéines déjà adsorbées ( $\alpha$ Chy et lysozyme) et 2 autres protéines ajoutées dans la boîte de simulation. Pour cela, la dernière *frame* de la simulation MD1-CONF1, où les deux protéines se sont chacune adsorbées indépendamment de l'autre sur la surface chromatographique, a été extraite et réutilisée comme point de départ. Les deux autres protéines ont ensuite été ajoutées manuellement dans la boîte de simulation, avec des distances suffisantes pour ne pas rentrer en conflit avec les protéines déjà adsorbées. La Figure IV.15 montre la configuration de départ des trois simulations avec les quatre protéines dans la boîte. Les conditions opératoires sont les mêmes que précédemment, c'est-à-dire un pH égal à 5, une concentration en chlorure de sodium égal à 100 mM et des temps de simulations fixés à 200 ns.

### IV.4.2.2 Résultats

Les résultats des trois simulations sont présentés sur la Figure IV.16, où les dernières *frames* de chaque simulation sont illustrées et les graphiques traduisant l'évolution des différentes distances minimales sont tracés. A noter que les trois figures ont été orientées selon différents points de vue de manière à pouvoir observer correctement les quatre protéines simultanément.

En situation de forte compétition entre protéines, l' $\alpha$ Chy s'est adsorbée sur les autres protéines pour chacune des simulations du fait de l'espace disponible restreint. En effet, même s'il y aurait techniquement assez de surface disponible pour les quatre protéines, la disposition à la surface de celles déjà adsorbées limitent les possibilités. Ainsi, à deux reprises, la deuxième  $\alpha$ Chy s'est fixée sur la première déjà adsorbée initialement, ce qui

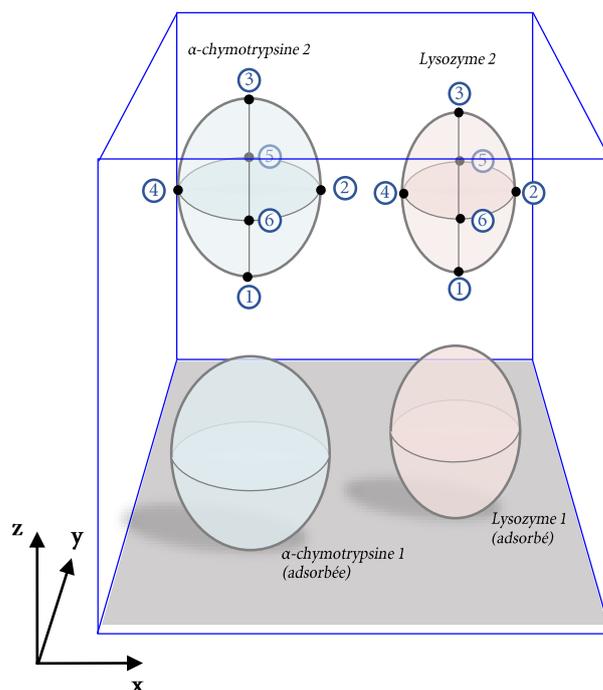


FIGURE IV.15 – Illustration de la configuration de départ pour les simulations en système bi-protéines, comprenant deux  $\alpha$ -chymotrypsines et deux lysozymes. La boîte de simulation est représentée par des lignes bleues et le support chromatographique par un plan gris. La dernière *frame* de la simulation MD1-CONF1 a été utilisée comme configuration initiale, à laquelle ont été ajoutées les deux protéines  $\alpha$ Chy et lysozyme.

pourrait cependant être majoritairement dû à la position de départ des deux  $\alpha$ Chy, assez proches l'une de l'autre.

Ces résultats semblent montrer la tendance selon laquelle le lysozyme possède bien un avantage sur l' $\alpha$ Chy et se fixe préférentiellement sur la surface. Dans cette situation où l' $\alpha$ Chy ne possède finalement que peu de surface disponible à l'adsorption, elle aura donc tendance à former une multicouche, soit avec le lysozyme, soit avec elle-même. Ce dernier résultat est d'ailleurs surprenant car lorsque l' $\alpha$ Chy n'est pas en présence du lysozyme, elle ne semble pas pour autant se fixer en multicouche d'après les observations des isothermes d'adsorption. Cependant, comme montré précédemment, les interactions protéine-protéine semblent moins stables au cours du temps que les interactions protéine-support, mais elles semblent malgré tout présentes lorsqu'il n'y a plus de surface disponible sur le support.

Il est cependant nécessaire de souligner que ces résultats de simulations permettent de mettre en évidence l'apparition d'un phénomène de multicouche lorsque les protéines se trouvent déjà à proximité de la surface. Cela n'exclut donc pas la possibilité que les protéines puissent éventuellement interagir avant leur adsorption, même si cela semble

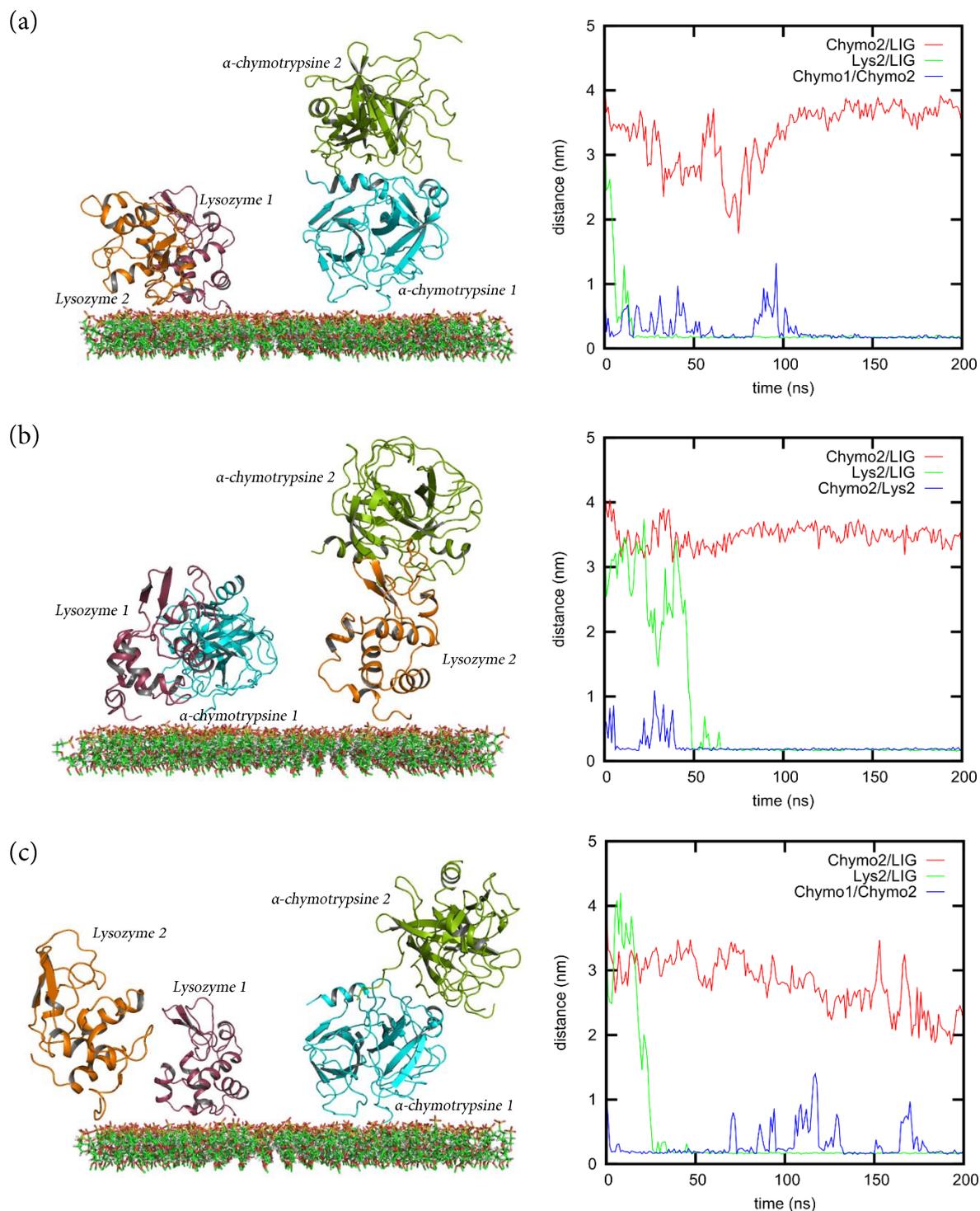


FIGURE IV.16 – Résultats des trois simulations en système bi-protéines, comprenant deux  $\alpha$ -chymotrypsines (en bleu cyan et vert) et deux lysozymes (en rouge et orange), à pH 5 et  $C_{NaCl} = 100$  mM. La dernière *frame* de chaque simulation, *i.e.* (a) MD1-CONF1, (b) MD2-CONF1 et (c) MD3-CONF1, sont présentées ainsi que les distances minimales Protéines–Ligands et Protéine–Protéine.

peu probable au regard de la stabilité de l'interaction. Il serait d'ailleurs intéressant de transposer les méthodes de calculs d'énergies de liaison à ces cas d'interaction  $\alpha$ Chy-lysozyme ou  $\alpha$ Chy- $\alpha$ Chy afin de quantifier la force de l'interaction. Par ailleurs, si les protéines semblent interagir entre-elles à l'intérieur des pores de la résine, il est assez peu probable que ce soit le cas en solution même à fortes concentrations car les analyses HPLC-SEC réalisées auraient permis de mettre en évidence un agglomérat, ce qui n'était pas le cas.

### IV.4.3 Conclusion

Les études conduites en système bi-protéines, avec comme cas d'étude l' $\alpha$ -chymotrypsine et le lysozyme sur la résine SP Sepharose FF, semblent prometteuses, et cela malgré le peu de simulations réalisées qui ne permettent pas une étude exhaustive de toutes les configurations possibles du système. Deux cas limites ont été représentés : (i) des simulations ont été réalisées avec deux protéines dans la boîte de simulation afin de représenter une faible concentration en protéines, puis (ii) des simulations ont été réalisées avec un total de quatre protéines, soit deux de chaque, afin de représenter un cas de forte compétition entre protéines.

En conclusion, il semblerait que lorsque la surface disponible à l'adsorption le permet, les deux protéines interagissent avec le support chromatographique. Le lysozyme semble cependant posséder une meilleure affinité avec la résine que l' $\alpha$ Chy, ce qui a également été suggéré par les résultats des expériences. En effet, le lysozyme s'est toujours adsorbé préférentiellement sur la surface, alors que dans des cas de forte compétition où la surface d'adsorption se retrouve finalement restreinte, l' $\alpha$ Chy s'est adsorbé en formant une multicouche. Ce phénomène a par ailleurs été supposé lors de l'étude expérimentale des isothermes de rétention, qui présentaient une forme pouvant traduire une rétention en multicouche. L' $\alpha$ Chy semble toutefois posséder une interaction plus stable avec la surface chromatographique qu'avec les autres protéines et s'adsorber préférentiellement sur les ligands lorsque l'espace disponible le permet. Cette adsorption multicouche est en grande partie possible grâce à la répartition dissymétrique des charges à la surface de l' $\alpha$ Chy, qui présente une face majoritairement positive et l'autre négative/neutre. Ainsi, cela lui permet de pouvoir interagir avec le lysozyme, dont les charges positives sont équitablement réparties à sa surface, ou bien avec elle-même par interaction entre ses deux faces de charges opposées.

L'intérêt des simulations moléculaires a ainsi été mis en évidence dans ces études en système multi-constituants et ont notamment permis d'identifier des phénomènes locaux pouvant être à l'origine des comportements observés à l'échelle macroscopique. En re-

vanche, cette étude reste préliminaire et un nombre plus important de simulations serait nécessaire pour mieux comprendre ces interactions entre protéines.



---

## CONCLUSION GÉNÉRALE

---

L'optimisation des procédés chromatographiques appliqués à la purification des protéines nécessite une meilleure compréhension des phénomènes à l'échelle atomique. Malgré leur intérêt et les nouvelles techniques émergentes, ces procédés restent encore sous-optimisés et peu de modèles prédictifs ont été mis au point. C'est donc dans l'optique d'acquérir des connaissances sur les comportements des protéines lors de la rétention à l'intérieur des résines échangeuses d'ions que ce travail de thèse a été initié. A ce jour, les techniques expérimentales à l'échelle microscopique s'avèrent généralement difficiles à mettre en œuvre et nécessitent entre autres des équipements sophistiqués et coûteux. Par conséquent, c'est dans l'idée de contourner ces difficultés techniques que certaines études ont été conduites en utilisant des méthodes numériques pour obtenir des informations sur les mécanismes à l'échelle atomique. Ainsi, la simulation moléculaire apparaît comme un outil intéressant afin de pouvoir étudier *in silico* les phénomènes d'échange d'ions des protéines, pour ensuite mieux comprendre les observations faites à l'échelle macroscopique.

Dans ce contexte, un système modèle, l' $\alpha$ -chymotrypsine, a tout d'abord été choisi afin de servir de référence pour la mise en œuvre des méthodes numériques et expérimentales. Ainsi, l' $\alpha$ -chymotrypsine présentait plusieurs intérêts : (i) disponibilité dans le commerce pour la réalisation des expériences à l'échelle macroscopique, (ii) structure cristallographique disponible dans la banque de données PDB pour la mise en œuvre des simulations, (iii) protéine connue pour sa rétention sur la résine standard cationique SP Sepharose FF, (iv) activité enzymatique afin de vérifier l'absence de dénaturation après élution.

La première étape du travail a été de définir comment relier des expériences à l'échelle macroscopique avec des simulations à l'échelle atomique. Pour cela, et à partir de certains travaux issus de la littérature, la loi d'Action de Masse Stérique (SMA), qui est un modèle très répandu pour la description de l'équilibre d'échange entre une macromolécule et une résine, a été appliquée et utilisée. Ce choix s'est montré particulièrement pertinent du fait de la définition des paramètres de cette loi, qui peuvent être déterminés expérimentalement via des mesures d'isothermes d'équilibre, mais qui possèdent également un sens physique et peuvent ainsi être déterminés à l'échelle atomique. Par conséquent, la

charge caractéristique, représentant le nombre de sites d'interaction du support chromatographique avec la protéine, et le facteur stérique, représentant le nombre de sites cachés (sans interaction et non disponibles) par la protéine ont pu être déterminés via les deux approches.

La deuxième étape a donc été de réaliser des expériences à l'échelle macroscopique. Des isothermes d'adsorption ont été mesurées dans des conditions de travail définies, à savoir un pH égal à 5 (avec un tampon citrate à 50 mM) et des concentrations en chlorure de sodium variant de 0 à 100 mM. Dans ces conditions, la méthode par mesure de courbes de percée a été privilégiée. Finalement, le modèle SMA a été appliqué et les paramètres, *i.e.* la charge caractéristique, le facteur stérique mais également la constante d'équilibre, ont été ajustés par comparaison entre le modèle non-linéaire et les données expérimentales. L' $\alpha$ -chymotrypsine a par ailleurs montré une bonne affinité avec la résine SP Sepharose FF et ne semble pas connaître de problèmes de précipitation, de dimérisation ou de dénaturation lors de l'échange d'ions dans les conditions de travail définies.

En parallèle de ces expériences, un système *in silico* mono-constituant été créé pour modéliser l'environnement local d'une protéine dans une résine chromatographique. Compte tenu de la taille relativement petite de la protéine choisie (25 kDa), le modèle de représentation appliqué a été le tout-atome. De plus, la méthode de Dynamique Moléculaire a été choisie afin d'échantillonner l'espace conformationnel et de pouvoir étudier l'évolution du système au cours du temps, notamment sa rétention sur le support. La surface chromatographique a été construite et est composée uniquement des ligands de la résine SP Sepharose FF, ainsi que d'une couche de contre-ions  $\text{Na}^+$ . En effet, le choix a été fait de ne pas représenter la matrice de la résine (gel d'agarose) afin de limiter la taille des systèmes et ainsi réduire le temps de calculs. Des simulations ont alors été réalisées, dans des orientations initiales différentes de la protéine afin de ne pas favoriser l'adsorption en fonction de l'orientation de la protéine au départ.

Une première étude de la répartition des charges sur la surface de la protéine selon le pH a permis d'identifier deux régions distinctes pour l' $\alpha$ -chymotrypsine. En effet, cette protéine possède la particularité de présenter une face globalement positive et une autre face neutre voire négative par endroit. Elle possède malgré tout une charge nette positive de +7 à pH 5 et peut donc être échangée sur une résine cationique. Lors des analyses des simulations, des sites d'interaction préférentiels ont été identifiés et les résidus participant essentiellement à l'adsorption ont été mis en évidence. Ces *patches* d'adsorption ont été proposés à partir de leur distance au ligand. L'estimation des énergies libres de liaison a permis de comparer qualitativement leurs contributions énergétiques. Cependant, l'écart entre ces valeurs et celles attendues à partir de la valeur expérimentale de la constante d'équilibre suggère une inadéquation du modèle à la physique du système. Ensuite, les deux autres paramètres de la loi SMA, la charge et le facteur stérique, ont été calculés

pour chaque simulation en moyennant les valeurs instantanées sur une période durant laquelle la protéine reste adsorbée. Pour l'ensemble des simulations, une moyenne a été aussi déterminée afin de tenir compte de divers comportements locaux possibles, pouvant contribuer à une observation macroscopique. Ces premiers résultats sont en bon accord avec ceux obtenus expérimentalement, et montrent tout particulièrement la pertinence de la démarche *in silico*. Dans le cas de l'estimation de la charge caractéristique, qui est une donnée dépendante essentiellement de la protéine et de ses résidus chargés en interaction, la concordance des valeurs est très bonne. Le facteur stérique, qui dépend grandement de la densité de ligands sur la surface chromatographique, présente une plus grande différence mais reste toutefois dans le même ordre de grandeur. Des ajustements seraient nécessaires, et en particulier une vérification de la densité de ligands qui a été en partie issue de la littérature et qui est difficilement quantifiable expérimentalement.

Il semble qu'une étude plus approfondie serait nécessaire pour permettre l'utilisation des simulations moléculaires pour prédire des paramètres globaux. En effet, lorsque les paramètres étudiés dépendent de la protéine uniquement comme par exemple la charge caractéristique, les simulations montrent un bon caractère prédictif. En revanche, lorsqu'il s'agit d'estimer un paramètre plus global comme la constante d'équilibre, qui ne dépend pas que d'une protéine mais de tout un environnement complexe (quantité de protéines en solution, force ionique, pH, etc.), les études numériques à une échelle locale trop restreinte ne permettent pas pour l'instant de bien modéliser ces effets.

L'influence de variables tels que le pH et la force ionique de la solution a également été étudiée et discutée. Pour cela, le système de référence a été conservé et ces deux variables (pH et force ionique) ont été modifiées successivement. Tout d'abord, une expérience de détermination de la courbe de percée à pH 5 et forte concentration en sel (400 mM) a été conduite et a montré que l' $\alpha$ -chymotrypsine n'est pas retenue sur la résine dans ces conditions du fait de la trop grande compétition avec les ions en solution. Lors des simulations dans ces mêmes conditions, il a été observé à quelques reprises une adsorption de la protéine malgré la forte concentration en sel. Cependant, comme attendu, les conditions ont été nettement moins favorables à l'adsorption et vont dans le sens des observations à l'échelle macroscopique. Ensuite, des expériences et des simulations ont été réalisées avec un pH 7, qui est un pH proche du point isoélectrique (pI) de la protéine, égal à 8,3. Expérimentalement, la protéine a montré moins d'affinité avec la résine qu'à pH 5, ce qui est un résultat attendu généralement attribué au fait que la charge nette de la protéine se rapproche de la neutralité lorsque le pH se rapproche du pI. La simulation ayant montré que les *patches* de fixation sont similaires à pH 5 et 7, la valeur du facteur stérique déterminé par la simulation à pH 5 a été utilisée pour le *fitting* du modèle SMA aux résultats expérimentaux. Dans ces conditions, il a été déterminé expérimentalement que la charge caractéristique dans la loi SMA est similaire à celle obtenue à pH 5, résultat confirmé par

les simulations. Cela s'explique en grande partie par le fait qu'aux deux pH, les lysines interagissant majoritairement avec le support sont toujours chargées. La répartition des charges laisse également supposer que, malgré une charge nette qui baisse, la face chargée positivement est toujours capable d'entrer en interaction avec la résine. Cependant, les constantes d'équilibre de la loi SMA déterminées expérimentalement aux deux pH sont très différentes, et les énergies libres de liaison calculées par le modèle le sont aussi, même si comme indiqué précédemment, les valeurs obtenues par les simulations et l'expérience ne peuvent pas être comparées, du fait peut-être de la complexité de l'environnement réel qui n'est pas assez pris en compte dans la simulation. Ce résultat est très intéressant à plusieurs niveaux. Il permet de constater que le mécanisme d'adsorption est extrêmement dépendant de la structure et de la répartition des charges de la protéine. Alors qu'il serait logique de penser qu'en se rapprochant du pI, des paramètres tels que la charge caractéristique soient plus faibles, ce résultat montre que ce n'est pas nécessairement le cas car cela dépend essentiellement de la protéine elle-même. Ainsi, la perte d'affinité est sûrement attribuable à d'autres facteurs liés à l'environnement, peut-être par effet de la non-idéalité des solutions dans les pores de la résine.

Une fois les méthodes de simulation et d'analyses validées, quelques études préliminaires sur un système multi-constituants ont été réalisées. Le système de référence a été repris et une autre protéine a été ajoutée, le lysozyme, qui est l'une des protéines modèles les plus couramment étudiées en chromatographie d'échange d'ions. Les conditions de référence ont également été conservées (pH 5 et  $C_{sel} = 100$  mM). Les isothermes déterminées en système bi-protéines ont tout d'abord relevé l'adsorption préférentielle du lysozyme par rapport à l' $\alpha$ -chymotrypsine, et une plus grande capacité de rétention, qui s'explique notamment par sa petite taille et son point isoélectrique plus éloigné du pH de travail. Ensuite, d'après les isothermes obtenues en système bi-protéine, il convient de distinguer deux situations différentes :

- lorsque les protéines en solution sont peu concentrées (situation faiblement compétitive), les isothermes en système bi-protéines sont sensiblement similaires à celles obtenues avec les protéines seules en solution, malgré des capacités par protéine logiquement inférieures puisqu'une partie du support est occupée par l'autre protéine ;
- lorsque les protéines en solution sont relativement concentrées (situation d'adsorption compétitive), une isotherme de l' $\alpha$ Chy en forme sigmoïdale est observée, laquelle peut être caractéristique d'une rétention multicouche.

Ce dernier phénomène a également été observé lors des simulations, et en particulier dans le cas où la surface disponible à l'adsorption s'est retrouvée restreinte du fait de la présence d'autres protéines. Comme observé sur les isothermes, l'adsorption du lysozyme sur le support chromatographique semble plus favorable que celle de l' $\alpha$ -chymotrypsine.

Cette dernière a montré la capacité d'interagir avec le support ou avec d'autres protéines déjà adsorbées. Ce phénomène pourrait être lié à la répartition des charges non isotrope à la surface de l' $\alpha$ Chy, dont les deux faces lui permettent de s'adsorber aussi bien sur des surfaces chargées négativement, comme le support chromatographique, que sur des surfaces chargées positivement, comme le lysozyme ou elle-même. Il faut cependant noter que cette interaction protéine-protéine semble bien moins stable qu'une adsorption sur le support, ce qui laisse penser que l' $\alpha$ Chy ne s'adsorbe en multicouche uniquement lorsqu'il n'y a plus de place sur le support.

L'ensemble de ces résultats permettent de mettre en évidence l'intérêt des études de simulation moléculaire dans le but de mieux comprendre les mécanismes de rétention des protéines au sein d'une résine échangeuse d'ions. Néanmoins, certaines limitations ont également pu être mises en évidence par la confrontation de ces méthodes avec les résultats expérimentaux. Finalement, ce travail ouvre ainsi la porte à plusieurs perspectives.

L'une des problématiques essentielles soulevées par cette étude qui mériterait d'être plus amplement étudiée est la correcte estimation des énergies libres de liaison. En effet, il a été montré que la méthode utilisée, la méthode MM-PBSA, ne semble pas être adaptée et des méthodes telles que l'*Umbrella Sampling* ou la *Steered Molecular Dynamics* pourraient apporter une meilleure estimation de ces énergies d'interaction. Il serait alors intéressant de les comparer à nouveau avec les valeurs de constantes d'équilibre issues des expériences. De même, il pourrait être pertinent d'étudier plus en détail l'effet de protéines supplémentaires sur l'évolution de la valeur de ces énergies. Si dans cette étude, les énergies ont été calculées sur une seule protéine isolée, il serait intéressant de vérifier l'impact de l'ajout de protéines dans la boîte de simulation sur ces valeurs.

Bien qu'elles soient difficiles à mettre en place, des techniques expérimentales microscopiques peuvent être envisagées, et en particulier afin d'acquérir des données expérimentales directes des valeurs d'énergies. La technique de l'AFM (microscopie de force atomique) semble notamment prometteuse et pourrait permettre d'obtenir des estimations d'énergies de liaison entre une protéine et un support chromatographique. Il est également possible d'envisager cette technique afin d'obtenir des informations sur la topographie de la surface chromatographique et en particulier pour acquérir des données sur la répartition des ligands, et ce dans le but de mieux représenter les surfaces lors des simulations moléculaires.

L'étude préliminaire conduite sur un système multi-constituants a soulevé le fort potentiel des méthodes de simulation moléculaire pour mieux comprendre les phénomènes d'interactions entre protéines. Une étude plus approfondie serait nécessaire, notamment par la mise en œuvre de nouvelles simulations avec de nouvelles orientations de départ afin de mieux échantillonner le système. Des analyses plus poussées et plus représentatives

du système global pourraient ainsi être effectuées. Cette approche pourrait être aussi appliquée à d'autres molécules en compétition avec la protéine pour le support, ou pouvant interagir avec elle.

Enfin, une dernière perspective qui paraît pertinente serait de transposer ces études à des cas de systèmes plus complexes, autant par la taille des molécules que par le nombre. Pour cela, il serait intéressant dans un premier temps de reproduire ce travail avec des méthodes numériques simplifiées type Gros Grain, et ainsi valider la concordance des deux méthodes. Enfin, une fois cette étape terminée, l'application à de plus gros systèmes peut être envisageable, ce qui permettrait entre autres l'étude de grosses molécules d'intérêt ou encore de pouvoir agrandir la boîte de simulation et travailler avec un plus grand nombre de protéines afin d'étudier des situations de compétition entre elles.

---

## BIBLIOGRAPHIE

---

- Abraham, A.L., 2008. Caractérisation et analyse évolutive des répétitions intragéniques : une étude au niveau des gènes, des séquences protéiques et des structures tridimensionnelles. Ph.D. thesis. Université Pierre et Marie Curie.
- Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015. GROMACS : High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2, 19–25. doi :10.1016/j.softx.2015.06.001.
- Abuin, E., Lissi, E., Bridi, R., 2011. Kinetics of p - nitrophenyl acetate hydrolysis catalyzed by  $\alpha$ - Chymotrypsin in presence of polyethylene glycol. *Journal of the Chilean Chemical Society* 56, 948–950. doi :10.4067/s0717-97072011000400028.
- Agashe, M., Raut, V., Stuart, S.J., Latour, R.A., 2005. Molecular simulation to characterize the adsorption behavior of a fibrinogen gamma-chain fragment. *Langmuir* 21, 1103–1117. doi :10.1021/la0478346.
- Anandakrishnan, R., Aguilar, B., Onufriev, A.V., 2012. H++ 3.0 : automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research* 40, W537–W541. doi :10.1093/nar/gks375.
- Andersen, H.C., 1980. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics* 72, 2384–2393. doi :10.1063/1.439486.
- Andreescu, S., Ornatska, M., Erlichman, J.S., Estevez, A., Leiter, J.C., 2012. Biomedical Applications of Metal Oxide Nanoparticles. Springer US, Boston, MA. chapter 3. pp. 57–100. doi :10.1007/978-1-4614-0379-1\_3.
- Angelo, J.M., Cvetkovic, A., Gantier, R., Lenhoff, A.M., 2013. Characterization of cross-linked cellulosic ion-exchange adsorbents : 1. structural properties. *Journal of Chromatography A* 1319, 46–56. doi :10.1016/j.chroma.2013.10.003.
- Åqvist, J., Luzhkov, V.B., Brandsdal, B.O., 2002. Ligand binding affinities from MD simulations. *Accounts of Chemical Research* 35, 358–365. doi :10.1021/ar010014p.
- Archer, D.G., Wang, P., 1990. The dielectric constant of water and debye-hückel limiting law slopes. *Journal of Physical and Chemical Reference Data* 19, 371–411. doi :10.1063/1.555853.
- Aune, K.C., Goldsmith, L.C., Timasheff, S.N., 1971. Dimerization of alpha-chymotrypsin. II. ionic strength and temperature dependence. *Biochemistry* 10, 1617–1622. doi :10.1021/bi00785a018.
- Aune, K.C., Timasheff, S.N., 1971. Dimerization of alpha-chymotrypsin. I. pH dependence in the acid region. *Biochemistry* 10, 1609–1617. doi :10.1021/bi00785a017.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A., 2001. Electrostatics of nanosystems : Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences* 98, 10037–10041. doi :10.1073/pnas.181342398.

- Barber, C.B., Dobkin, D.P., Huhdanpaa, H., 1996. The quickhull algorithm for convex hulls. *ACM Transactions On Mathematical Software* 22, 469–483.
- Barrett, G., 2012. *Chemistry and Biochemistry of the Amino Acids*. Springer Netherlands.
- Bas, D.C., Rogers, D.M., Jensen, J.H., 2008. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins : Structure, Function, and Bioinformatics* 73, 765–783. doi :10.1002/prot.22102.
- Bashford, D., Karplus, M., 1990. pKa's of ionizable groups in proteins : atomic detail from a continuum electrostatic model. *Biochemistry* 29, 10219–10225. doi :10.1021/bi00496a010.
- Beeman, D., 1976. Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics* 20, 130–139. doi :10.1016/0021-9991(76)90059-0.
- Berendsen, H.J.C., Grigera, J.R., Straatsma, T.P., 1987. The missing term in effective pair potentials. *The Journal of Physical Chemistry* 91, 6269–6271. doi :10.1021/j100308a038.
- Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R., 1984. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* 81, 3684–3690. doi :10.1063/1.448118.
- Bosma, J.C., Wesselingh, J.A., 2004. Available area isotherm. *AIChE Journal* 50, 848–853. doi :10.1002/aic.10080.
- Brice, A.R., Dominy, B.N., 2011. Analyzing the robustness of the MM/PBSA free energy calculation method : Application to DNA conformational transitions. *Journal of Computational Chemistry* 32, 1431–1440. doi :10.1002/jcc.21727.
- Brooks, C.A., Cramer, S.M., 1992. Steric mass-action ion exchange : Displacement profiles and induced salt gradients. *AIChE Journal* 38, 1969–1978. doi :10.1002/aic.690381212.
- Brunauer, S., Emmett, P.H., Teller, E., 1938. Adsorption of gases in multimolecular layers. *Journal of the American Chemical Society* 60, 309–319. doi :10.1021/ja01269a023.
- Bussi, G., Donadio, D., Parrinello, M., 2007. Canonical sampling through velocity rescaling. *Journal of Chemical Physics* 126, 014101. doi :10.1063/1.2408420.
- Cabanne, C., Noubhani, A., Dieryck, W., Hocquellet, A., Santarelli, X., 2004. Evaluation of three expanded bed adsorption anion exchange matrices with the aid of recombinant enhanced green fluorescent protein overexpressed in escherichia coli. *Journal of Chromatography B* 808, 91–97. doi :10.1016/j.jchromb.2004.02.009.
- Carta, G., Jungbauer, A., 2010. *Protein Chromatography*. Wiley VCH Verlag GmbH.
- Ceperley, D., Alder, B., 1986. Quantum monte carlo. *Science* 231, 555–560. doi :10.1126/science.231.4738.555.
- Ceperley, D.M., 2010. An overview of quantum monte carlo methods. *Reviews in Mineralogy and Geochemistry* 71, 129–135. doi :10.2138/rmg.2010.71.6.
- Chang, C., Lenhoff, A.M., 1998. Comparison of protein adsorption isotherms and uptake rates in preparative cation-exchange materials. *Journal of Chromatography A* 827, 281–293. doi :10.1016/s0021-9673(98)00796-1.
- Chen, C., Depa, P., Sakai, V.G., Maranas, J.K., Lynn, J.W., Peral, I., Copley, J.R.D., 2006. A comparison of united atom, explicit atom, and coarse-grained simulation models for poly(ethylene oxide). *The Journal of Chemical Physics* 124, 234901. doi :10.1063/1.2204035.

- Chen, S.K., Wang, G, B.C., Huang, W.Z., Gao, S., 2010. Molecular dynamics simulation of the interaction between maps and cu(100) surface., in : The 2nd Conference on Environmental Science and Information Application Technology, pp. 533–536.
- Collier, G., Vellore, N.A., Latour, R.A., Stuart, S.J., 2009. Development of molecular simulation methods to accurately represent protein-surface interactions : Method assessment for the calculation of electrostatic effects. *Biointerphases* 4, 57–64. doi :10.1116/1.3266417.
- Costa, H.B., Fernandes, P.M., Romão, W., Ventura, J.A., 2014. A new procedure based on column chromatography to purify bromelain by ion exchange plus gel filtration chromatographies. *Industrial Crops and Products* 59, 163–168. doi :10.1016/j.indcrop.2014.04.042.
- Creasy, A., Barker, G., Yao, Y., Carta, G., 2015. Systematic interpolation method predicts protein chromatographic elution from batch isotherm data without a detailed mechanistic isotherm model. *Biotechnology Journal* 10, 1400–1411. doi :10.1002/biot.201500089.
- Darden, T., York, D., Pedersen, L., 1993. Particle mesh ewald : An  $n \cdot \log(n)$  method for ewald sums in large systems. *The Journal of Chemical Physics* 98, 10089–10092. doi :10.1063/1.464397.
- Demaneche, S., Chapel, J.P., Monrozier, L.J., Quiquampoix, H., 2009. Dissimilar pH-dependent adsorption features of bovine serum albumin and  $\alpha$ -chymotrypsin on mica probed by AFM. *Colloids and Surfaces B : Biointerfaces* 70, 226–231. doi :10.1016/j.colsurfb.2008.12.036.
- Deng, Y., Roux, B., 2009. Computations of standard binding free energies with molecular dynamics simulations. *The Journal of Physical Chemistry B* 113, 2234–2246. doi :10.1021/jp807701h. PMID : 19146384.
- DePhillips, P., Lenhoff, A.M., 2000. Pore size distributions of cation-exchange adsorbents determined by inverse size-exclusion chromatography. *Journal of Chromatography A* 883, 39–54. doi :10.1016/S0021-9673(00)00420-9.
- DePhillips, P., Lenhoff, A.M., 2001. Determinants of protein retention characteristics on cation-exchange adsorbents. *Journal of Chromatography A* 933, 57–72. doi :https://doi.org/10.1016/S0021-9673(01)01275-4.
- DiMaio, F., Chiu, W., 2016. Tools for model building and optimization into near-atomic resolution electron cryo-microscopy density maps, in : *Methods in Enzymology*. Elsevier, pp. 255–276. doi :10.1016/bs.mie.2016.06.003.
- Dismer, F., Hubbuch, J., 2007. A novel approach to characterize the binding orientation of lysozyme on ion-exchange resins. *Journal of Chromatography A* 1149, 312–320. doi :10.1016/j.chroma.2007.03.074.
- Dismer, F., Hubbuch, J., 2010. 3D structure-based protein retention prediction for ion-exchange chromatography. *Journal of Chromatography A* 1217, 1343–1353. doi :10.1016/j.chroma.2009.12.061.
- Dismer, F., Petzold, M., Hubbuch, J., 2008. Effects of ionic strength and mobile phase pH on the binding orientation of lysozyme on different ion-exchange adsorbents. *Journal of Chromatography A* 1194, 11–21. doi :10.1016/j.chroma.2007.12.085.

- Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., Baker, N.A., 2004. PDB2PQR : An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic acids research* 32, W665–7. doi :10.1093/nar/gkh381.
- Drenth, J., 2007. *Principles of Protein X-Ray Crystallography*. SpringerLink : Springer e-Books, Springer New York. URL : <https://books.google.fr/books?id=Jobr7svN0IIC>.
- Durell, S.R., Brooks, B.R., Ben-Naim, A., 1994. Solvent-induced forces between two hydrophilic groups. *Journal of Physical Chemistry* 98, 2198–2202. doi :10.1021/j100059a038.
- Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G., 1995. A smooth particle mesh Ewald method. *Journal of Chemical Physics* 103, 8577–8593. doi :10.1063/1.470117.
- Ewald, P.P., 1921. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik* 369, 253–287. doi :10.1002/andp.19213690304.
- Feller, S.E., Zhang, Y., Pastor, R.W., Brooks, B.R., 1995. Constant pressure molecular dynamics simulation : The langevin piston method. *The Journal of Chemical Physics* 103, 4613–4621. doi :10.1063/1.470648.
- Fernandez, M., Carta, G., 1996. Characterization of protein adsorption by composite silica-polyacrylamide gel anion exchangers i. equilibrium and mass transfer in agitated contactors. *Journal of Chromatography A* 746, 169–183. doi :[https://doi.org/10.1016/0021-9673\(96\)00337-8](https://doi.org/10.1016/0021-9673(96)00337-8).
- Ferrari, A.M., Degliesposti, G., Sgobba, M., Rastelli, G., 2007. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Bioorganic & Medicinal Chemistry* 15, 7865–7877. doi :10.1016/j.bmc.2007.08.019.
- Fulton, A.B., Isaacs, W.B., 1991. Titin, a huge, elastic sarcomeric protein with a probable role in morphogenesis. *BioEssays* 13, 157–161. doi :10.1002/bies.950130403.
- Gadam, S.D., Jayaraman, G., Cramer, S.M., 1993. Characterization of non-linear adsorption properties of dextran-based polyelectrolyte displacers in ion-exchange systems. *Journal of Chromatography A* 630, 37–52. doi :10.1016/0021-9673(93)80440-j.
- GE-Healthcare, 2014. Sepharose Fast Flow ion exchange media and prepacked column formats. GE Healthcare Bio-Sciences AB. URL : [www.gelifesciences.com/protein-purification](http://www.gelifesciences.com/protein-purification). data file 18-1177-22 AE.
- Genheden, S., Ryde, U., 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* 10, 449–461. doi :10.1517/17460441.2015.1032936.
- Ghosh, S., Jana, S., Guchhait, N., 2012. Domain specific association of small fluorescent probe trans-3-(4-monomethylaminophenyl)-acrylonitrile (MMAPA) with bovine serum albumin (BSA) and its dissociation from protein binding sites by ag nanoparticles : Spectroscopic and molecular docking study. *The Journal of Physical Chemistry B* 116, 1155–1163. doi :10.1021/jp2094752.

- Goh, G.B., Hulbert, B.S., Zhou, H., Brooks, C.L., 2014. Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins : Structure, Function, and Bioinformatics* 82, 1319–1331. doi :10.1002/prot.24499.
- Gohlke, H., Kiel, C., Case, D.A., 2003. Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the ras–raf and ras–RalGDS complexes. *Journal of Molecular Biology* 330, 891–913. doi :10.1016/s0022-2836(03)00610-7.
- Gomori, G., 1955. [16] preparation of buffers for use in enzyme studies, in : *Methods in Enzymology*. Elsevier, pp. 138–146. doi :10.1016/0076-6879(55)01020-3.
- Gordon, J.C., Myers, J.B., Folta, T., Shoja, V., Heath, L.S., Onufriev, A., 2005. H++ : a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Research* 33, W368–W371. doi :10.1093/nar/gki464.
- Gregor, H.P., 1951. Gibbs-donnan equilibria in ion exchange resin systems. *Journal of the American Chemical Society* 73, 642–650. doi :10.1021/ja01146a042.
- Grotefend, S., Kaminski, L., Wroblewitz, S., Deeb, S.E., Kühn, N., Reichl, S., Limberger, M., Watt, S., Wätzig, H., 2012. Protein quantitation using various modes of high performance liquid chromatography. *Journal of Pharmaceutical and Biomedical Analysis* 71, 127–138. doi :10.1016/j.jpba.2012.08.024.
- Gu, F., Chodavarapu, K., McCreary, D., Plitt, T.A., Tamoria, E., Ni, M., Burnham, J.J., Peters, M., Lenhoff, A.M., 2015. Silica-based strong anion exchange media for protein purification. *Journal of Chromatography A* 1376, 53–63. doi :https://doi.org/10.1016/j.chroma.2014.11.082.
- Hammond, B.L., Lester, W.A., Reynolds, P.J., 1994. *Monte Carlo Methods in Ab Initio Quantum Chemistry*. WORLD SCIENTIFIC. doi :10.1142/1170.
- Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., Hutchison, G.R., 2012. Avogadro : an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics* 4, 17. doi :10.1186/1758-2946-4-17.
- Hardin, A.M., Harinarayan, C., Malmquist, G., Axén, A., van Reis, R., 2009. Ion exchange chromatography of monoclonal antibodies : Effect of resin ligand density on dynamic binding capacity. *Journal of Chromatography A* 1216, 4366–4371. doi :10.1016/j.chroma.2008.08.047.
- Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., 1997. LINCS : a linear constraint solver for molecular simulations. *Journal of Computational Chemistry* 18, 1463–1472. doi :10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
- Hockney, R., Eastwood, J., 1981. *Computer Simulation Using Particles*. McGraw–Hill.
- Hohenberg, P., Kohn, W., 1964. Inhomogeneous electron gas. *Physical Review* 136, B864–B871. doi :10.1103/physrev.136.b864.
- Homeyer, N., Gohlke, H., 2012. Free energy calculations by the molecular mechanics poisson-boltzmann surface area method. *Molecular Informatics* 31, 114–122. doi :10.1002/minf.201100135.
- Hong, P., Fountain, K.J., 2011. Method Development for Size-Exclusion Chromatography of Monoclonal Antibodies and Higher Order Aggregates. Application note. Waters Corporation.

- Hoover, W.G., 1985. Canonical dynamics : Equilibrium phase-space distributions. *Physical Review A* 31, 1695–1697. doi :10.1103/physreva.31.1695.
- Huang, J., MacKerell, A.D., 2013. CHARMM36 all-atom additive protein force field : Validation based on comparison to nmr data. *Journal of Computational Chemistry* 34, 2135–2145. doi :10.1002/jcc.23354.
- Huang, K., García, A.E., 2014. Effects of truncating van der waals interactions in lipid bilayer simulations. *The Journal of Chemical Physics* 141, 105101. doi :10.1063/1.4893965.
- Huang, Y., Bi, J., Zhao, L., Ma, G., Su, Z., 2010. Regulation of protein multipoint adsorption on ion-exchange adsorbent and its application to the purification of macromolecules. *Protein Expression and Purification* 74, 257–263. doi :10.1016/j.pep.2010.07.002.
- Hudáky, P., Kaslik, G., Venekei, I., Gráf, L., 1999. The differential specificity of chymotrypsin A and B is determined by amino acid 226. *European Journal of Biochemistry* 259, 528–533. doi :10.1046/j.1432-1327.1999.00075.x.
- Humphrey, W., Dalke, A., Schulten, K., 1996. VMD : Visual molecular dynamics. *Journal of Molecular Graphics* 14, 33–38. doi :10.1016/0263-7855(96)00018-5.
- Hunter, A.K., Carta, G., 2002. Protein adsorption on novel acrylamido-based polymeric ion-exchangers : Iv. effects of protein size on adsorption capacity and rate. *Journal of Chromatography A* 971, 105 – 116. doi :[https://doi.org/10.1016/S0021-9673\(02\)01027-0](https://doi.org/10.1016/S0021-9673(02)01027-0).
- Huuk, T.C., Hahn, T., Osberghaus, A., Hubbuch, J., 2014. Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography. *Separation and Purification Technology* 136, 207–222. doi :10.1016/j.seppur.2014.09.012.
- Imamura, K., Kawasaki, Y., Awadzu, T., Sakiyama, T., Nakanishi, K., 2003. Contribution of acidic amino residues to the adsorption of peptides onto a stainless steel surface. *Journal of Colloid and Interface Science* 267, 294–301. doi :10.1016/s0021-9797(03)00700-8.
- Janakiraman, V.N., Solé, M., Maria, S., Pezzini, J., Cabanne, C., Santarelli, X., 2018. Comparative study of strong cation exchangers : Structure-related chromatographic performances. *Journal of Chromatography B* 1080, 1–10. doi :10.1016/j.jchromb.2018.02.014.
- Jeppu, G.P., Clement, T.P., 2012. A modified langmuir-freundlich isotherm model for simulating pH-dependent adsorption effects. *Journal of Contaminant Hydrology* 129-130, 46–53. doi :10.1016/j.jconhyd.2011.12.001.
- Johansson, B.L., Andersson, M., Lausmaa, J., Sjövall, P., 2004. Chemical characterisation of different separation media based on agarose by static time-of-flight secondary ion mass spectrometry. *Journal of Chromatography A* 1023, 49–56. doi :10.1016/j.chroma.2003.10.008.
- Johnson, C.A., Yuan, Y., Lenhoff, A.M., 2000. Adsorbed layers of ferritin at solid and fluid interfaces studied by atomic force microscopy. *Journal of Colloid and Interface Science* 223, 261–272. doi :10.1006/jcis.1999.6680.
- Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79, 926–935. doi :10.1063/1.445869.

- Jorgensen, W.L., Madura, J.D., 1985. Temperature and size dependence for monte carlo simulations of TIP4p water. *Molecular Physics* 56, 1381–1392. doi :10.1080/00268978500103111.
- Jungbauer, A., 2013. Continuous downstream processing of biopharmaceuticals. *Trends in Biotechnology* 31, 479–492. doi :10.1016/j.tibtech.2013.05.011.
- Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D.W., Dolinsky, T., Konecny, R., Koes, D.R., Nielsen, J.E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G.W., Holst, M.J., McCammon, J.A., Baker, N.A., 2017. Improvements to the APBS biomolecular solvation software suite. *Protein Science* 27, 112–128. doi :10.1002/pro.3280.
- Kar, P., Gopal, S.M., Cheng, Y.M., Predeus, A., Feig, M., 2013. PRIMO : A transferable coarse-grained force field for proteins. *Journal of Chemical Theory and Computation* 9, 3769–3788. doi :10.1021/ct400230y.
- Kharazian, B., Lohse, S.E., Ghasemi, F., Raoufi, M., Saei, A.A., Hashemi, F., Farvadi, F., Alimohamadi, R., Jalali, S.A., Shokrgozar, M.A., Hadipour, N.L., Ejtehadi, M.R., Mahmoudi, M., 2018. Bare surface of gold nanoparticle induces inflammation through unfolding of plasma fibrinogen. *Scientific Reports* 8. doi :10.1038/s41598-018-30915-7.
- Kisley, L., Chen, J., Mansur, A.P., Dominguez-Medina, S., Kulla, E., Kang, M.K., Shuang, B., Kourentzi, K., Poongavanam, M.V., Dhamane, S., Willson, R.C., Landes, C.F., 2014. High ionic strength narrows the population of sites participating in protein ion-exchange adsorption : A single-molecule study. *Journal of Chromatography A* 1343, 135–142. doi :10.1016/j.chroma.2014.03.075.
- Kittelmann, J., Lang, K.M., Ottens, M., Hubbuch, J., 2017a. Orientation of monoclonal antibodies in ion-exchange chromatography : A predictive quantitative structure–activity relationship modeling approach. *Journal of Chromatography A* 1510, 33–39. doi :10.1016/j.chroma.2017.06.047.
- Kittelmann, J., Lang, K.M., Ottens, M., Hubbuch, J., 2017b. An orientation sensitive approach in biomolecule interaction quantitative structure–activity relationship modeling and its application in ion-exchange chromatography. *Journal of Chromatography A* 1482, 48–56. doi :10.1016/j.chroma.2016.12.065.
- Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A.E., Kolinski, A., 2016. Coarse-grained protein models and their applications. *Chemical Reviews* 116, 7898–7936. doi :10.1021/acs.chemrev.6b00163.
- Komeiji, Y., 2000. Ewald summation and multiple time step methods for molecular dynamics simulation of biological molecules. *Journal of Molecular Structure : THEOCHEM* 530, 237–243. doi :10.1016/s0166-1280(99)00314-0.
- Krigbaum, W.R., Kuegler, F.R., 1970. Molecular conformation of egg-white lysozyme and bovine  $\alpha$ -lactalbumin in solution. *Biochemistry* 9, 1216–1223. doi :10.1021/bi00807a024.
- Kubiak, K., Mulheran, P.A., 2009. Molecular dynamics simulations of hen egg white lysozyme adsorption at a charged solid surface. *The Journal of Physical Chemistry B* 113, 12189–12200. doi :10.1021/jp901521x.

- Kumari, R., Kumar, R., Lynn, A., 2014. *g\_mmpbsa*—a GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling* 54, 1951–1962. doi :10.1021/ci500020m.
- Ladiwala, A., Rege, K., Breneman, C.M., Cramer, S.M., 2005. A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems. *Proceedings of the National Academy of Sciences* 102, 11710–11715. doi :10.1073/pnas.0408769102.
- Lan, Q., Bassi, A.S., Zhu, J.X.J., Margaritis, A., 2001. 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* 81, 179–186. doi :10.1016/s1385-8947(00)00197-2.
- Lang, K.M., Kittelmann, J., Dürr, C., Osberghaus, A., Hubbuch, J., 2015a. A comprehensive molecular dynamics approach to protein retention modeling in ion exchange chromatography. *Journal of Chromatography A* 1381, 184–193. doi :10.1016/j.chroma.2015.01.018.
- Lang, K.M., Kittelmann, J., Pilgram, F., Osberghaus, A., Hubbuch, J., 2015b. Custom-tailored adsorbents : A molecular dynamics study on optimal design of ion exchange chromatography material. *Journal of Chromatography A* 1413, 60–67. doi :10.1016/j.chroma.2015.08.021.
- Langford, J., Schure, M., Yao, Y., Maloney, S., Lenhoff, A., 2006. Effects of pore structure and molecular size on diffusion in chromatographic adsorbents. *Journal of Chromatography A* 1126, 95–106. doi :10.1016/j.chroma.2006.06.060.
- Laskowski, M., 1955. [2] chymotrypsinogens and chymotrypsins, in : *Methods in Enzymology*. Elsevier, pp. 8–26. doi :10.1016/s0076-6879(55)02166-6.
- Latour, R.A., 2008. Molecular simulation of protein-surface interactions : Benefits, problems, solutions, and future directions (review). *Biointerphases* 3, FC2–FC12. doi :10.1116/1.2965132.
- Latour, R.A., 2014. Perspectives on the simulation of protein–surface interactions using empirical force field methods. *Colloids and Surfaces B : Biointerfaces* 124, 25–37. doi :10.1016/j.colsurfb.2014.06.050.
- Lenhoff, A.M., 2011. Protein adsorption and transport in polymer-functionalized ion-exchangers. *Journal of Chromatography A* 1218, 8748–8759. doi :10.1016/j.chroma.2011.06.061.
- Lenhoff, A.M., 2016. Ion-exchange chromatography of proteins : the inside story. *Materials Today : Proceedings* 3, 3559–3567. doi :10.1016/j.matpr.2016.10.038.
- Leško, M., Åsberg, D., Enmark, M., Samuelsson, J., Fornstedt, T., Kaczmarski, K., 2015. Choice of model for estimation of adsorption isotherm parameters in gradient elution preparative liquid chromatography. *Chromatographia* 78, 1293–1297. doi :10.1007/s10337-015-2949-0.
- Li, H., Robertson, A.D., Jensen, J.H., 2005. Very fast empirical prediction and rationalization of protein pKa values. *Proteins : Structure, Function, and Bioinformatics* 61, 704–721. doi :10.1002/prot.20660.
- Liang, J., Fieg, G., Jakobtorweihen, S., 2015a. Ion-Exchange Adsorption of Proteins : Experiments and Molecular Dynamics Simulations. *Chemie Ingenieur Technik* 87, 903–909. doi :10.1002/cite.201400095.

- Liang, J., Fieg, G., Jakobtorweihen, S., 2015b. Molecular Dynamics Simulations of a Binary Protein Mixture Adsorption onto Ion-Exchange Adsorbent. *Industrial & Engineering Chemistry Research* 54, 2794–2802. doi :10.1021/ie504374x.
- Liang, J., Fieg, G., Keil, F.J., Jakobtorweihen, S., 2012. Adsorption of Proteins onto Ion-Exchange Chromatographic Media : A Molecular Dynamics Study. *Industrial & Engineering Chemistry Research* 51, 16049–16058. doi :10.1021/ie301407b.
- Liwo, A., Arlukowicz, P., Czaplewski, C., Oldziej, S., Pillardy, J., Scheraga, H.A., 2002. A method for optimizing potential-energy functions by a hierarchical design of the potential-energy landscape : Application to the UNRES force field. *Proceedings of the National Academy of Sciences* 99, 1937–1942. doi :10.1073/pnas.032675399.
- Liwo, A., Khalili, M., Czaplewski, C., Kalinowski, S., Ołdziej, S., Wachucik, K., Scheraga, H.A., 2007. Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. i. temperature dependence of the effective energy function and tests of the optimization method with single training proteins. *The Journal of Physical Chemistry B* 111, 260–285. doi :10.1021/jp065380a.
- Loncharich, R.J., Brooks, B.R., 1989. The effects of truncating long-range forces on protein dynamics. *Proteins : Structure, Function, and Genetics* 6, 32–45. doi :10.1002/prot.340060104.
- Lopes, P.E.M., Guvench, O., MacKerell, A.D., 2014. Current status of protein force fields for molecular dynamics simulations, in : *Methods in Molecular Biology*. Springer New York, pp. 47–71. doi :10.1007/978-1-4939-1465-4\_3.
- MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., Karplus, M., 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins†. *The Journal of Physical Chemistry B* 102, 3586–3616. doi :10.1021/jp973084f.
- Mahoney, M.W., Jorgensen, W.L., 2000. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics* 112, 8910–8922. doi :10.1063/1.481505.
- Marek, W.K., Sauer, D., Dürauer, A., Jungbauer, A., Piątkowski, W., Antos, D., 2018. Prediction tool for loading, isocratic elution, gradient elution and scaling up of ion exchange chromatography of proteins. *Journal of Chromatography A* 1566, 89–101. doi :10.1016/j.chroma.2018.06.057.
- Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., de Vries, A.H., 2007. The MARTINI force field : coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* 111, 7812–7824. doi :10.1021/jp071097f.
- Martyna, G.J., Klein, M.L., Tuckerman, M., 1992. Nosé–hoover chains : The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics* 97, 2635–2643. doi :10.1063/1.463940.

- Matthews, B.W., Sigler, P.B., Henderson, R., Blow, D.M., 1967. Three-dimensional structure of tosyl- $\alpha$ -chymotrypsin. *Nature* 214, 652–656. doi :10.1038/214652a0.
- Mattrey, F.T., Makarov, A.A., Regalado, E.L., Bernardoni, F., Figus, M., Hicks, M.B., Zheng, J., Wang, L., Schafer, W., Antonucci, V., Hamilton, S.E., Zawatzky, K., Welch, C.J., 2017. Current challenges and future prospects in chromatographic method development for pharmaceutical research. *TrAC Trends in Analytical Chemistry* 95, 36–46. doi :10.1016/j.trac.2017.07.021.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21, 1087–1092. doi :10.1063/1.1699114.
- Misak, N.Z., 1993. Langmuir isotherm and its application in ion-exchange reactions. *Reactive Polymers* 21, 53–64. doi :10.1016/0923-1137(93)90054-j.
- Møller, C., Plesset, M.S., 1934. Note on an approximation treatment for many-electron systems. *Physical Review* 46, 618–622. doi :10.1103/physrev.46.618.
- Mollerup, J.M., 2008. A review of the thermodynamics of protein association to ligands, protein adsorption, and adsorption isotherms. *Chemical Engineering & Technology* 31, 864–874. doi :10.1002/ceat.200800082.
- Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P., Marrink, S.J., 2008. The MARTINI coarse-grained force field : Extension to proteins. *Journal of Chemical Theory and Computation* 4, 819–834. doi :10.1021/ct700324x.
- Monticelli, L., Tieleman, D.P., 2012. Force fields for classical molecular dynamics, in : *Methods in Molecular Biology*. Humana Press, pp. 197–213. doi :10.1007/978-1-62703-017-5\_8.
- Myers, J., Grothaus, G., Narayanan, S., Onufriev, A., 2006. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins : Structure, Function, and Bioinformatics* 63, 928–938. doi :10.1002/prot.20922.
- Nelson, D., Lehninger, A., Cox, M., 2008. *Lehninger Principles of Biochemistry*. Lehninger Principles of Biochemistry, W. H. Freeman.
- Coquebert de Neuville, B., Tarafder, A., Morbidelli, M., 2013. Distributed pore model for bio-molecule chromatography. *Journal of Chromatography A* 1298, 26–34. doi :10.1016/j.chroma.2013.04.074.
- Nie, Z., Kumacheva, E., 2008. Patterning surfaces with functional polymers. *Nature Materials* 7, 277–290. doi :10.1038/nmat2109.
- Noble, J.E., Bailey, M.J., 2009. Chapter 8 quantitation of protein, in : *Methods in Enzymology*. Elsevier, pp. 73–95. doi :10.1016/s0076-6879(09)63008-1.
- Nosé, S., 1984. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* 81, 511–519. doi :10.1063/1.447334.
- Olsen, S.R., Watanabe, F.S., 1957. A method to determine a phosphorus adsorption maximum of soils as measured by the langmuir isotherm1. *Soil Science Society of America Journal* 21, 144. doi :10.2136/sssaj1957.03615995002100020004x.

- Olsson, M.H.M., Søndergaard, C.R., Rostkowski, M., Jensen, J.H., 2011. PROPKA3 : Consistent treatment of internal and surface residues in empirical pKa predictions. *Journal of Chemical Theory and Computation* 7, 525–537. doi :10.1021/ct100578z.
- Ozboyaci, M., Kokh, D.B., Corni, S., Wade, R.C., 2016. Modeling and simulation of protein–surface interactions : achievements and challenges. *Quarterly Reviews of Biophysics* 49. doi :10.1017/S0033583515000256.
- Pahari, S., Sun, L., Basu, S., Alexov, E., 2018. DelPhiPKa : Including salt in the calculations and enabling polar residues to titrate. *Proteins : Structure, Function, and Bioinformatics* 86, 1277–1283. doi :10.1002/prot.25608.
- Petersen, H.G., 1995. Accuracy and efficiency of the particle mesh ewald method. *The Journal of Chemical Physics* 103, 3668–3679. doi :10.1063/1.470043.
- Petsko, G.A., Ringe, D., Sanlaville, C., Charmot, D., 2008. Structure et fonction des protéines. *Biologie cellulaire et moléculaire, De Boeck Supérieur.*
- Pharr, M., Jakob, W., Humphreys, G., 2017. Monte carlo integration, in : *Physically Based Rendering.* Elsevier, pp. 747–802. doi :10.1016/b978-0-12-800645-0.50013-0.
- Ponder, J.W., Case, D.A., 2003. Force fields for protein simulations, in : *Protein Simulations.* Elsevier, pp. 27–85. doi :10.1016/s0065-3233(03)66002-x.
- Przybycien, T.M., Pujar, N.S., Steele, L.M., 2004. Alternative bioseparation operations : life beyond packed-bed chromatography. *Current Opinion in Biotechnology* 15, 469–478. doi :10.1016/j.copbio.2004.08.008.
- Pyle, D., 2012. *Separations for Biotechnology 2.* Springer Netherlands.
- Raje, P., Pinto, N.G., 1997. Combination of the steric mass action and non-ideal surface solution models for overload protein ion-exchange chromatography. *Journal of Chromatography A* 760, 89–103. doi :10.1016/S0021-9673(96)00812-6.
- Rao, S.N., Singh, U.C., Bash, P.A., Kollman, P.A., 1987. Free energy perturbation calculations on binding and catalysis after mutating asn 155 in subtilisin. *Nature* 328, 551–554. doi :10.1038/328551a0.
- Riccardi, E., Wang, J.C., Liapis, A.I., 2008. Rational surface design for molecular dynamics simulations of porous polymer adsorbent media. *The Journal of Physical Chemistry B* 112, 7478–7488. doi :10.1021/jp800078v.
- Rossmann, M.G., Battisti, A.J., Plevka, P., 2011. Future prospects, in : *Recent Advances in Electron Cryomicroscopy, Part B.* Elsevier, pp. 101–121. doi :10.1016/b978-0-12-386507-6.00005-1.
- Schreiber, F., 2004. Self-assembled monolayers : from simple model systems to biofunctionalized interfaces. *Journal of Physics : Condensed Matter* 16, R881–R900. doi :10.1088/0953-8984/16/28/r01.
- Schrödinger, LLC, 2015. *The PyMOL Molecular Graphics System, Version 1.7.*
- Scida, K., Stege, P.W., Haby, G., Messina, G.A., García, C.D., 2011. Recent applications of carbon-based nanomaterials in analytical chemistry : Critical review. *Analytica Chimica Acta* 691, 6–17. doi :10.1016/j.aca.2011.02.025.

- Shen, H., Frey, D.D., 2005. Effect of charge regulation on steric mass-action equilibrium for the ion-exchange adsorption of proteins. *Journal of Chromatography A* 1079, 92–104. doi :10.1016/j.chroma.2005.02.086.
- Smyth, M.S., 2000. x ray crystallography. *Molecular Pathology* 53, 8–14. doi :10.1136/mp.53.1.8.
- Snyder, D.A., Chen, Y., Denissova, N.G., Acton, T., Aramini, J.M., Ciano, M., Karlin, R., Liu, J., Manor, P., Rajan, P.A., Rossi, P., Swapna, G.V.T., Xiao, R., Rost, B., Hunt, J., Montelione, G.T., 2005. Comparisons of NMR spectral quality and success in crystallization demonstrate that NMR and x-ray crystallography are complementary methods for small protein structure determination. *Journal of the American Chemical Society* 127, 16505–16511. doi :10.1021/ja053564h.
- Søndergaard, C.R., Olsson, M.H.M., Rostkowski, M., Jensen, J.H., 2011. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *Journal of Chemical Theory and Computation* 7, 2284–2295. doi :10.1021/ct200133y.
- de Souza, O., Ornstein, R., 1997. Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with particle-mesh ewald method. *Biophysical Journal* 72, 2395–2397. doi :10.1016/s0006-3495(97)78884-2.
- Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A., Case, D.A., 1998. Continuum solvent studies of the stability of dna, rna, and phosphoramidate dna helices. *Journal of the American Chemical Society* 120, 9401–9409. doi :10.1021/ja981844+.
- Steinbach, P.J., Brooks, B.R., 1994. New spherical-cutoff methods for long-range forces in macromolecular simulation. *Journal of Computational Chemistry* 15, 667–683. doi :10.1002/jcc.540150702.
- Stedle, A., Pleiss, J., 2011. Modelling of lysozyme binding to a cation exchange surface at atomic detail : The role of flexibility. *Biophysical Journal* 100, 3016–3024. doi :10.1016/j.bpj.2011.05.024.
- Su, X.L., Sun, Y., 2006. Thermodynamic model for nonlinear electrostatic adsorption equilibrium of protein. *AIChE Journal* 52, 2921–2930. doi :10.1002/aic.10900.
- Szabo, A., Ostlund, N., 2012. *Modern Quantum Chemistry : Introduction to Advanced Electronic Structure Theory*. Dover Books on Chemistry, Dover Publications.
- Takada, S., 2012. Coarse-grained molecular simulations of large biomolecules. *Current Opinion in Structural Biology* 22, 130–137. doi :10.1016/j.sbi.2012.01.010.
- Tanford, C., 1961. *Physical chemistry of macromolecules*. John Wiley and Sons, Inc., New York.
- Tesson, S., 2016. Un champ de force polarisable pour l'étude des argiles à l'échelle moléculaire. Ph.D. thesis. Université Pierre et Marie Curie. doi :10.13140/RG.2.2.25096.83201.
- Trodler, P., Nieveler, J., Rusnak, M., Schmid, R.D., Pleiss, J., 2008. Rational design of a new one-step purification strategy for *Candida antarctica* lipase B by ion-exchange chromatography. *Journal of Chromatography A* 1179, 161–167. URL : <http://www.sciencedirect.com/science/article/pii/S0021967307020705>, doi :10.1016/j.chroma.2007.11.108.
- Ubiera, A.R., Carta, G., 2006. Radiotracer measurements of protein mass transfer : Kinetics in ion exchange media. *Biotechnology Journal* 1, 665–674. doi :10.1002/biot.200600023.

- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., Jr., A.D.M., 2010. CHARMM general force field : A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force field. *Journal of Computational Chemistry* 31, 671–690. doi :10.1002/jcc.21367.
- Vanommeslaeghe, K., Jr., A.D.M., 2012. Automation of the CHARMM general force field (CGenFF) I : bond perception and atom typing. *Journal of Chemical Information and Modeling* 52, 3144–3154. doi :10.1021/ci300363c.
- Vanommeslaeghe, K., Raman, E.P., Jr., A.D.M., 2012. Automation of the CHARMM general force field (CGenFF) II : Assignment of bonded parameters and partial atomic charges. *Journal of Chemical Information and Modeling* 52, 3155–3168. doi :10.1021/ci3003649.
- Verlet, L., 1967. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical Review* 159, 98–103. doi :10.1103/physrev.159.98.
- Verma, S.K., Ghosh, K.K., 2013. Activity, stability and kinetic parameters for  $\alpha$ -chymotrypsin catalysed reactions in aot/isooctane reverse micelles with nonionic and zwitterionic mixed surfactants. *Journal of Chemical Sciences* 125, 875–882. doi :10.1007/s12039-013-0434-6.
- Villiermaux, J., 1993. Génie de la réaction chimique : conception et fonctionnement des réacteurs. Génie des procédés de l'École de Nancy, Tec & Doc.
- Wang, L., Li, L., Alexov, E., 2015a. pKa predictions for proteins, RNAs, and DNAs with the gaussian dielectric function using DelPhi pKa. *Proteins : Structure, Function, and Bioinformatics* 83, 2186–2197. doi :10.1002/prot.24935.
- Wang, L., Zhang, M., Alexov, E., 2015b. DelPhiPKa web server : predicting pKa of proteins, RNAs and DNAs. *Bioinformatics* 32, 614–615. doi :10.1093/bioinformatics/btv607.
- Wassenaar, T.A., Mark, A.E., 2005. The effect of box shape on the dynamic properties of proteins simulated under periodic boundary conditions. *Journal of Computational Chemistry* 27, 316–325. doi :10.1002/jcc.20341.
- Wei, T., Carignano, M.A., Szleifer, I., 2011. Lysozyme adsorption on polyethylene surfaces : Why are long simulations needed ? *Langmuir* 27, 12074–12081. doi :10.1021/la202622s.
- Wennberg, C.L., Murtola, T., Hess, B., Lindahl, E., 2013. Lennard-jones lattice summation in bilayer simulations has critical effects on surface tension and lipid properties. *Journal of Chemical Theory and Computation* 9, 3527–3537. doi :10.1021/ct400140n.
- Wetter, L.R., Deutsch, H.F., 1951. Immunological studies on egg white proteins. iv. immunological and physical studies of lysozyme. *The Journal of biological chemistry* 192, 237–242.
- Whitford, D., 2013. *Proteins : Structure and Function*. Wiley.
- Whitley, R., Wachter, R., Liu, F., Wang, N.H., 1989. Ion-exchange equilibria of lysozyme myoglobin and bovine serum albumin : Effective valence and exchanger capacity. *Journal of Chromatography A* 465, 137–156. URL : <http://www.sciencedirect.com/science/article/pii/S0021967301926536>, doi :[https://doi.org/10.1016/S0021-9673\(01\)92653-6](https://doi.org/10.1016/S0021-9673(01)92653-6).
- Wüthrich, K., 1990. Protein structure determination in solution by nmr spectroscopy. *The Journal of biological chemistry* 265, 22059–22062.

- Wu, D., Walters, R.R., 1992. Effects of stationary phase ligand density on high-performance ion-exchange chromatography of proteins. *Journal of Chromatography A* 598, 7–13. doi :10.1016/0021-9673(92)85108-6.
- Xu, L., Glatz, C.E., 2009. Predicting protein retention time in ion-exchange chromatography based on three-dimensional protein characterization. *Journal of Chromatography A* 1216, 274–280. doi :10.1016/j.chroma.2008.11.075.
- Yampolskaya, G., Platikanov, D., 2006. Proteins at fluid interfaces : Adsorption layers and thin liquid films. *Advances in Colloid and Interface Science* 128-130, 159–183. doi :10.1016/j.cis.2006.11.018.
- Yancey, J.A., Vellore, N.A., Collier, G., Stuart, S.J., Latour, R.A., 2010. Development of molecular simulation methods to accurately represent protein-surface interactions : The effect of pressure and its determination for a system with constrained atoms. *Biointerphases* 5, 85–95. doi :10.1116/1.3493470.
- Yang, K., Shi, Q.H., Sun, Y., 2006. Modeling and simulation of protein uptake in cation exchanger visualized by confocal laser scanning microscopy. *Journal of Chromatography A* 1136, 19–28. doi :10.1016/j.chroma.2006.09.036.
- Yao, Y., Lenhoff, A.M., 2004. Determination of pore size distributions of porous chromatographic adsorbents by inverse size-exclusion chromatography. *Journal of Chromatography A* 1037, 273–282. doi :10.1016/j.chroma.2004.02.054.
- Yao, Y., Lenhoff, A.M., 2006. Pore size distributions of ion exchangers and relation to protein binding capacity. *Journal of Chromatography A* 1126, 107–119. doi :10.1016/j.chroma.2006.06.057.
- Yeh, I.C., Berkowitz, M.L., 1999. Ewald summation for systems with slab geometry. *Journal of Chemical Physics* 111, 3155–3162. doi :10.1063/1.479595.
- Yu, W., He, X., Vanommeslaeghe, K., Jr., A.D.M., 2012. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry* 33, 2451–2468. doi :10.1002/jcc.23067.
- Zhang, L., Sun, Y., 2010. Molecular simulation of adsorption and its implications to protein chromatography : A review. *Biochemical Engineering Journal* 48, 408–415. doi :10.1016/j.bej.2009.12.003.
- Zhang, L., Zhao, G., Sun, Y., 2009a. Molecular insight into protein conformational transition in hydrophobic charge induction chromatography : A molecular dynamics simulation. *The Journal of Physical Chemistry B* 113, 6873–6880. doi :10.1021/jp809754k.
- Zhang, L., Zhao, G., Sun, Y., 2010. Effects of ligand density on hydrophobic charge induction chromatography : Molecular dynamics simulation. *The Journal of Physical Chemistry B* 114, 2203–2211. doi :10.1021/jp903852c.
- Zhang, Z., Chen, S., Feng, Y., Ding, Y., Zhou, J., Jia, H., 2009b. Electrochemical and molecular simulation studies on the corrosion inhibition of l-glutamine monolayers on an iron surface. *Journal of the Serbian Chemical Society* 74, 407–415. doi :10.2298/jsc0904407z.

- 
- Zhao, K., Bai, Q., Song, C., Wang, F., Yang, F., 2012. Preparation of weak cation exchange packings for chromatographic separation of proteins using "click chemistry". *Journal of Separation Science* 35, 907–914. doi :10.1002/jssc.201101097.
- Zhou, Z.H., 2011. Atomic resolution cryo electron microscopy of macromolecular complexes, in : *Recent Advances in Electron Cryomicroscopy, Part B*. Elsevier, pp. 1–35. doi :10.1016/b978-0-12-386507-6.00001-4.



# A

## FICHER PDB

```

SEQRES 1 A 129 LYS VAL PHE GLY ARG CYS GLU LEU ALA ALA ALA MET LYS
SEQRES 2 A 129 ARG HIS GLY LEU ASP ASN TYR ARG GLY TYR SER LEU GLY
SEQRES 3 A 129 ASN TRP VAL CYS ALA ALA LYS PHE GLU SER ASN PHE ASN
SEQRES 4 A 129 THR GLN ALA THR ASN ARG ASN THR ASP GLY SER THR ASP
SEQRES 5 A 129 TYR GLY ILE LEU GLN ILE ASN SER ARG TRP TRP CYS ASN
SEQRES 6 A 129 ASP GLY ARG THR PRO GLY SER ARG ASN LEU CYS ASN ILE
SEQRES 7 A 129 PRO CYS SER ALA LEU LEU SER SER ASP ILE THR ALA SER
SEQRES 8 A 129 VAL ASN CYS ALA LYS LYS ILE VAL SER ASP GLY ASN GLY
SEQRES 9 A 129 MET ASN ALA TRP VAL ALA TRP ARG ASN ARG CYS LYS GLY
SEQRES 10 A 129 THR ASP VAL GLN ALA TRP ILE ARG GLY CYS ARG LEU
FORMUL 2 HOH *78(H2 O)
HELIX 1 1 ARG A 5 ARG A 14 1 10
HELIX 2 2 TYR A 20 GLY A 22 5 3
HELIX 3 3 LEU A 25 SER A 36 1 12
HELIX 4 4 CYS A 80 LEU A 84 5 5
HELIX 5 5 THR A 89 ASP A 101 1 13
HELIX 6 6 GLY A 104 ALA A 107 5 4
HELIX 7 7 VAL A 109 ARG A 114 1 6
HELIX 8 8 VAL A 120 TRP A 123 5 4
SHEET 1 A 2 THR A 43 ARG A 45 0
SHEET 2 A 2 THR A 51 TYR A 53 -1 N ASP A 52 O ASN A 44
SSBOND 1 CYS A 6 CYS A 127 1555 1555 1.97
SSBOND 2 CYS A 30 CYS A 115 1555 1555 2.00
SSBOND 3 CYS A 64 CYS A 80 1555 1555 1.99
SSBOND 4 CYS A 76 CYS A 94 1555 1555 2.02
ATOM 1 N LYS A 1 35.365 22.342 -11.980 1.00 22.28 N
ATOM 2 CA LYS A 1 35.892 21.073 -11.427 1.00 21.12 C
ATOM 3 C LYS A 1 34.741 20.264 -10.844 1.00 16.85 C
ATOM 4 O LYS A 1 33.945 20.813 -10.081 1.00 18.94 O
ATOM 5 CB LYS A 1 36.872 21.435 -10.306 1.00 20.78 C
ATOM 6 CG LYS A 1 37.453 20.248 -9.565 1.00 18.47 C
ATOM 7 CD LYS A 1 38.688 20.649 -8.775 1.00 20.32 C
ATOM 8 CE LYS A 1 39.057 19.508 -7.837 1.00 24.76 C
ATOM 9 NZ LYS A 1 40.423 19.771 -7.299 1.00 28.27 N
ATOM 10 N VAL A 2 34.739 18.961 -11.042 1.00 19.96 N
ATOM 11 CA VAL A 2 33.903 17.998 -10.333 1.00 18.10 C
ATOM 12 C VAL A 2 34.800 17.312 -9.294 1.00 19.39 C
ATOM 13 O VAL A 2 35.759 16.605 -9.665 1.00 22.14 O
ATOM 14 CB VAL A 2 33.140 17.034 -11.232 1.00 16.81 C

```

N° atome – Type d'atome – Résidu – Chaîne – N° séquence – Coordonnées x y z – Occupation – Facteur B

FIGURE A.1 – Extrait d'un fichier PDB du lysozyme (code PDB : 1AKI), représentant les informations sur la séquence en acide-aminés, les hélices  $\alpha$  et les feuillets  $\beta$ , les ponts disulfures, ainsi que les informations sur les 14 premiers atomes de la protéine.





## Résumé

La compréhension des mécanismes d'échange d'ions à l'échelle locale a connu un fort intérêt ces dernières décennies. En effet, des informations à l'échelle atomique des mécanismes physico-chimiques pourraient permettre l'optimisation des procédés chromatographiques à l'échelle industrielle, qui restent à ce jour sous-optimisés car largement basés sur des méthodes empiriques. De plus, certains comportements observés en système mono-constituant ne peuvent être transposés à des cas plus complexes du fait de l'apparition de phénomènes tels que les interactions protéines-support ou encore la compétition protéine-protéine. Dans ce contexte, une approche particulièrement prometteuse est d'utiliser la simulation moléculaire pour étudier ces phénomènes locaux à l'intérieur des adsorbants (résines échangeuses d'ions) en complément aux techniques expérimentales sophistiquées, qui s'avèrent difficiles à mettre en œuvre et généralement très coûteuses. La simulation moléculaire peut ainsi permettre d'étudier les interactions d'une protéine dans son environnement, notamment avec la résine chromatographique, mais aussi d'identifier d'éventuels changements conformationnels. Dans ce travail, l'utilisation de la simulation moléculaire est proposée et discutée en confrontant les données obtenues numériquement à des données d'expériences réalisées à l'échelle macroscopique. En particulier, l'équilibre d'échange d'ions a été étudiée avec mesure des isothermes d'adsorption et application de la loi d'Action de Masse Stérique, afin d'évaluer la pertinence de cette approche. Les premiers résultats de simulations moléculaires obtenus sur un système modèle ont ainsi permis d'acquérir une vision moléculaire du mécanisme de rétention de la protéine étudiée sur la surface chromatographique et notamment de mettre en évidence des orientations préférentielles. De plus, la comparaison de deux paramètres physiques calculés à partir des deux approches (in silico et expérimentale) a montré un bon accord, indiquant que l'utilisation de ce type de méthode numérique est prometteuse dans ce domaine. La modélisation des effets de l'environnement (pH, force ionique, protéines en compétition) semble également prometteuse, mais nécessiterait une plus grande investigation et l'utilisation d'approches numériques plus adaptées à la complexité du système.

**Mots clés :** Simulation moléculaire, protéines, chromatographie d'échange d'ions,  $\alpha$ -chymotrypsine, lysozyme, loi SMA.

## Abstract

The interest for a better understanding of ion-exchange mechanisms at the atomic level has strongly increased over the past decades. Indeed, molecular-level information about physic-chemical mechanisms could help optimizing chromatographic processes for protein purification, which are sub-optimized and still based on empirical methods. Furthermore, behaviors observed in a single-component system cannot be adapted to more complex systems due to the occurrence of phenomena such as protein-protein competition or protein-support interactions. In this context, a promising approach is the use of molecular simulations to investigate local phenomena inside the adsorbents (ion-exchange resin) as a complement to sophisticated experimental techniques, which are difficult to implement and tend to be expensive. Thus, molecular simulations may allow to study protein interactions in its environment, especially with the chromatographic resin, but also to identify possible conformational changes. In this work, the use of molecular simulations is proposed and discussed by comparing simulations against data from macroscopic scale experiments. In particular, ion-exchange equilibrium was studied through adsorption isotherms experiments and application of the Steric Mass Action law in order to evaluate the reliability of this strategy. The first results from molecular simulations obtained with a model system allowed to acquire a molecular insight of retention mechanism of the protein on the chromatographic surface and in particular to identify preferential orientations. Moreover, the comparison of two physical parameters obtained from both approaches (in silico and experiments) showed a good agreement, emphasizing that the use of this kind of numerical method is useful in the field. Modeling the environment effects (pH, ionic strength, competitive proteins) seems also promising, though a deeper investigation and the use of other numerical approaches dedicated to the system complexity should be carried out.

**Keywords :** Molecular simulation, proteins, ion-exchange chromatography,  $\alpha$ -chymotrypsin, lysozyme, SMA law.