

THÈSE

présentée à

L'UNIVERSITÉ PIERRE ET MARIE CURIE

par

Bruno DAUNAY

pour obtenir le diplôme de

**DOCTORAT DE L'UNIVERSITÉ
PIERRE ET MARIE CURIE**

Spécialité : Mécanique – Robotique

**Couplage haptique pour des applications
de *docking* moléculaire**

Soutenue le 4 décembre 2007

JURY

M.	N. CHAILLET	Professeur à l'Université de Besançon	Rapporteur
Mme	A. VODA	Maître de conférences (HDR) à l'Université de Grenoble	Rapporteur
M.	A. FERREIRA	Professeur à l'Ecole Nationale Supérieure d'Ingénieurs de Bourges	Examineur
M.	A. MICAELLI	Co-directeur de thèse Directeur de recherche au Commissariat à l'Energie Atomique	Examineur
Mme	V. PERDEREAU	Professeur à l'Université Pierre et Marie Curie	Examineur
M.	S. REDON	Chargé de recherche à l'Institut National de Recherche en Informatique et en Automatique	Examineur
M.	S. REGNIER	Directeur de thèse Maître de conférences (HDR) à l'Université Pierre et Marie Curie	Examineur

INVITE

M.	H. MINOUX	Ingénieur de recherche à Sanofi-Aventis	Examineur
----	-----------	---	-----------

THÈSE

présentée à

L'UNIVERSITÉ PIERRE ET MARIE CURIE

par

Bruno DAUNAY

pour obtenir le diplôme de

**DOCTORAT DE L'UNIVERSITÉ
PIERRE ET MARIE CURIE**

Spécialité : Mécanique – Robotique

**Couplage haptique pour des applications
de *docking* moléculaire**

Soutenue le 4 décembre 2007

JURY

M.	N. CHAILLET	Professeur à l'Université de Besançon	Rapporteur
Mme	A. VODA	Maître de conférences (HDR) à l'Université de Grenoble	Rapporteur
M.	A. FERREIRA	Professeur à l'Ecole Nationale Supérieure d'Ingénieurs de Bourges	Examineur
M.	A. MICAELLI	Co-directeur de thèse Directeur de recherche au Commissariat à l'Energie Atomique	Examineur
Mme	V. PERDEREAU	Professeur à l'Université Pierre et Marie Curie	Examineur
M.	S. REDON	Chargé de recherche à l'Institut National de Recherche en Informatique et en Automatique	Examineur
M.	S. REGNIER	Directeur de thèse Maître de conférences (HDR) à l'Université Pierre et Marie Curie	Examineur

INVITE

M.	H. MINOUX	Ingénieur de recherche à Sanofi-Aventis	Examineur
----	-----------	---	-----------

COUPLAGE HAPTIQUE POUR DES APPLICATIONS DE DOCKING MOLÉCULAIRE

Résumé

Réaliser une tâche de *docking* moléculaire avec un retour d'effort à six degrés de liberté suppose de résoudre trois difficultés majeures. La première difficulté concerne le temps de réponse du changement conformationnel de la protéine qui n'est pas temps réel. En conséquence, la commande du système peut être instable. La deuxième difficulté réside dans la différence d'échelle entre les efforts et les déplacements de l'environnement de l'utilisateur et de celui de la molécule ainsi que dans la forte raideur des efforts moléculaires ce qui pose un réel problème d'interprétation. Enfin, la troisième difficulté réside dans le fait que le simulateur utilise une énergie d'interaction alors que l'interface haptique requiert un effort d'interaction. La conversion en effort doit donc être réalisée.

La plate-forme de téléopération mise en place et autorisant un retour d'effort à six degrés de liberté prend en compte ces difficultés et propose :

- la réutilisation d'un logiciel de simulation moléculaire intégrant un processus de minimisation de l'énergie calculée à partir d'un champ de force quelconque, interfacé avec un organe haptique de la classe *Virtuose* ;
- une commande en variables d'onde capable d'atténuer la grande amplitude des efforts et de rendre passive la communication retardée par le calcul du changement de conformation de la protéine. Des coefficients homothétiques font l'adéquation entre les grandeurs physiques des deux mondes ;
- une conversion de l'énergie d'interaction en effort à ressentir assurée par un modèle paramétrable utilisant la méthode des moindres carrés pour l'actualisation de ses paramètres. Sa dérivée analytique fournit les forces et les moments de l'interaction.

HAPTIC FEEDBACK FOR MOLECULAR DOCKING APPLICATIONS

Abstract

Carrying out a molecular docking task with a force feedback with six degrees of freedom presupposes addressing three major difficulties. Using software designed for pharmaceutical applications requires that it be first adapted. The initial difficulty therefore consists in the protein's conformational change response time, which is not real time. The commands therefore risk being unstable. The scale difference between the forces and the displacements of the user's and the molecule's environment, as well as the high stiffness of the intermolecular forces pose a real problem with regard to the interpretation of the stress. Further, the molecular simulator assesses interaction energy. Consideration must therefore be given to its conversion into force.

The remote operation platform set up and authorising a force feedback with six degrees of freedom, takes into account these difficulties and puts forward :

- the reuse of molecular simulation software integrating an energy minimisation process calculated from any force field interfaced with a haptic organ of the *Virtuose* class ;
- a wave-variable command capable of attenuating the considerable amplitude of the stress and making communication passive, delayed by the calculation of the protein's conformation change. Homothetic coefficients provide the concordance between the physical magnitudes of the two worlds ;
- a conversion of the interaction energy into stress to be felt. The latter is provided by a parameterisable model using the least squares method to update its parameters. Its analytical derivative provides the interaction stress.

REMERCIEMENTS

Le travail présenté dans ce manuscrit est le fruit d'une collaboration entre plusieurs laboratoires de recherche. Si la mise en œuvre et la rédaction sont personnelles, l'économie d'ensemble en revient à une équipe au sein de laquelle je tiens à remercier plus particulièrement trois personnes :

Tout d'abord *Stéphane Régnier* qui a été successivement un enseignant, un encadrant puis mon directeur de thèse. Depuis le début, c'est lui qui a accepté la lourde charge de me guider, qui a su trouver les mots justes et parfois sévères pour m'encourager et qui s'est révélé un appui sûr et solide tant au plan moral que scientifique tout au long de ces trois années de recherche. A la fois garde-fou et boîte à idées, il est en grande partie responsable de l'aboutissement de mes travaux.

Alain Micaelli ensuite qui est devenu au fil du temps un de mes encadrants. En cinq minutes d'explications, il a ce don de relancer pour six mois les recherches des thésards qui trouvent refuge auprès de lui. Sa rigueur et sa vision scientifique de la problématique de ma thèse sont à l'origine des résultats majeurs qui s'y trouvent.

Florian Schramm enfin, avec qui j'ai partagé les moments d'enthousiasme et les périodes difficiles qu'induisent nécessairement trois années de recherche dans nos équipes respectives. Il m'a fait découvrir la rigueur du travail allemand quand je lui ai montré ma vision de la recherche française. Son soutien scientifique durant certaines périodes laborieuses a été grandement appréciable.

Trois années d'enseignement et de recherche, de découverte des arcanes de l'université et du monde de l'entreprise se terminent aujourd'hui. J'y ai appris la *théorie des cafards* que *Lucile* et moi avons pu vérifier. *Lucile*, qui est l'instigatrice des fous rires des pauses-café. Experte dans une certaine imitation, je ne me lassais pas - et ne me lasserai jamais - d'écouter ses expériences. La *G. S.*, dont fait notamment partie *Jérémie*, se termine finalement bien et une autre aventure commence pour elle.

Par la force des choses, ma rédaction s'est intégralement déroulée sur une mezzanine aux mœurs particulières. l'ambiance qui y régnait lorsque les chaleurs caniculaires de l'été sévissaient était certainement peu propice aux études... *Mathieu, Léa, Guigui* et *Barth'* furent ainsi, tour à tour, l'objet d'« expériences » au gré de leurs absences estivales. Les premiers mois furent agités mais toujours studieux. L'entraide permanente me fut salutaire. C'est ainsi qu'*Ahlem* s'est vue dépossédée de son bureau pour que je puisse mener à bien toutes les simulations et créer toutes les images présentées ici. *Aude* m'a apporté toute son expertise à l'entame du troisième chapitre et *Guillaume*, une aide précieuse pour la communication « sockétiale ». *Barth'* me fut d'un grand secours en relisant régulièrement ma prose et en me pourvoyant en corrigés pour que mes étudiants n'aient pas à souffrir de ma charge de travail.

Les correcteurs de mes épreuves ont aussi été d'un recours précieux. En sus de la qualité scientifique améliorée au fil des pages grâce à la profusion d'encre rouge disséminée par *Stéphane* et *Alain*, la grammaire et l'orthographe ont nécessité de nombreux appoints. En dehors du soutien indéfectible et permanent de *Tsübermaxa*, en osant la référence, où, cette fois, *ber* remplace *bru* et c'est tout à mon honneur, *Anne* m'a fait très largement profiter de ses compétences de rédactrice pour que, à défaut de ressembler à un mensuel, ce manuscrit ressemble à une thèse. Les corrections que vous tous avez apportées l'ont été dans des temps records, votre implication personnelle m'a beaucoup touché.

Sophie, Franz, Dawid, Jon, PhB², YBB, Seb' et consorts avec qui j'ai effectué un parcours parallèle de rêve durant ces années de formation à l'*ESPP* ont enduré mes longues tirades, lors des soirées prolongées, sur la haute idée que je me fais de l'encadrement à l'université et en *P*. Ils m'ont tout bonnement supporté, c'est donc un clin d'œil que je leur adresse. Collègues et compagnons de galère, toujours motivés par l'aventure, ils ont été une source inépuisable de confiance et de réconfort.

N'ayant pas cité tout le monde, loin s'en faut, je me dois tout de même de mentionner *David L.* pour son appréciable investissement quant à la relecture de mes articles et *Gilbert* qui s'est impliqué dans l'éradication des fautes d'orthographe.

Je conclurai cette page en évoquant *Clara*, dont la simple présence, toujours joyeuse, m'a fait survoler ces trois années. Mélange de culture australienne, néo-zélandaise et désormais calédonienne et nivane, elle me permet de parcourir le monde grâce à ses récits de thèse. Les miens sont encore bien maigres et j'espère un jour pouvoir raconter le mouvement d'un robot en Antarctique, l'agitation d'une molécule à la Réunion ou la manipulation d'un quelconque objet submicronique sur la Lune.

TABLE DES MATIÈRES

Table des matières	i
Table des figures	v
Notations	ix
Introduction générale	1
1 Haptique et simulations moléculaires	5
1 Généralités	6
1.1 Du virus au médicament	6
1.1.1 Exemple simplifié du virus du sida (VIH)	6
1.1.2 Vers un médicament	6
1.2 Biologie structurale	7
1.2.1 Les protéines	7
1.2.2 Les acides aminés et peptides	8
1.2.3 Liaison peptidique	8
1.2.4 Conclusion	9
1.2.5 Représentation des protéines en mécanique moléculaire	10
1.2.5.1 Représentation en coordonnées cartésiennes	10
1.2.5.2 Représentation en coordonnées internes	10
1.2.5.3 Conclusion	11
2 Simulation moléculaire	12
2.1 Méthodes issues du génie biologique	13
2.1.1 Modélisation des interactions	13
2.1.1.1 Mécanique quantique	13
2.1.1.2 Méthodes hybrides	14
2.1.1.3 Mécanique moléculaire ou méthodes empiriques	15
2.1.2 Simulation du comportement protéique	24
2.1.2.1 Méthodes de minimisation du potentiel	24
2.1.2.2 La dynamique moléculaire	26
2.2 Méthodes issues de la robotique	27

	2.2.1	Modélisation par chaîne cinématique	27
	2.2.2	Planification de la trajectoire moléculaire	30
	2.2.3	Conclusion	31
	2.3	Bilan	31
3		Haptique moléculaire	32
	3.1	Approche rigide	33
	3.2	Approche flexible/rigide	36
	3.3	Bilan	38
4		Approche proposée	39
	4.1	Ligand flexible et protéine flexible	40
	4.2	Logiciel de simulation du <i>docking</i> moléculaire	41
	4.3	Retour d'effort	41
5		Conclusion	43
2		Etude d'un couplage simple pour le docking	45
1		Couplage haptique : généralités	46
	1.1	Introduction	46
	1.2	Passivité	46
	1.3	Couplage idéal sans retard	48
	1.3.1	Couplage idéal	48
	1.3.2	Commande du ligand	48
	1.3.3	Principe	49
	1.3.4	Facteurs d'échelle	50
	1.4	L'environnement	50
	1.5	Transparence	51
2		Système mis en place	52
	2.1	Architecture de la plate-forme de téléopération	52
	2.2	Optimisation du simulateur moléculaire	53
	2.2.1	La méthode de minimisation	54
	2.2.2	Contraintes internes	55
	2.2.3	Conclusion	56
	2.3	Description d'une manipulation	56
	2.3.1	Action de l'opérateur	57
	2.3.2	Protocole UDP	57
	2.3.3	Facteur de déplacement	58
	2.3.4	Evaluation de l'énergie	59
	2.3.5	Efforts à ressentir	60
	2.3.6	Coefficient en effort	62
	2.3.7	Minimisation	63
	2.4	Commande d'un ligand flexible	64
	2.4.1	Repères de travail	64
	2.4.1.1	Repères associés au <i>Virtuose</i>	65
	2.4.1.2	Simulateur moléculaire	65
	2.4.2	Cas du ligand flexible	67
3		Retour d'effort	71
	3.1	Influence du facteur d'effort	71
	3.1.1	Système rigide	71
	3.1.2	Système flexible	71

3.2	Influence du facteur de déplacement	73
3.2.1	Facteur de déplacement faible	73
3.2.2	Facteur de déplacement élevé	74
3.3	Conclusion	75
4	Stabilité du système	76
5	Conclusion	78
3	Couplage d'un système retardé à grande dynamique	79
1	Variables d'onde	80
1.1	Introduction	80
1.2	Présentation	81
1.3	Principe	82
1.4	Passivité	83
1.5	Conclusion	85
2	Etude d'une commande en variables d'onde	86
2.1	Commande du <i>Virtuose</i> dans le cas idéal	86
2.2	Utilisation des variables d'onde	87
2.3	Filtrage	88
2.4	Transparence	90
2.5	Conclusion	91
3	Variables d'onde et amortisseur virtuel	91
3.1	Schéma de commande	92
3.2	Codage et décodage des variables	92
3.3	Résultats	94
3.3.1	Faible coefficient d'amortissement	95
3.3.2	Coefficient d'amortissement élevé	95
3.4	Conclusion	96
4	Variables d'onde sans amortissement	97
4.1	Simulateur moléculaire	97
4.1.1	Intégration de la vitesse	97
4.1.2	Expression de la position du ligand	98
4.2	Analyse des résultats	100
5	Conclusion	101
4	Estimation des efforts à partir d'un modèle d'énergie paramétrable	103
1	Estimation directe de la force par un observateur	103
1.1	Introduction	103
1.2	Méthode	104
1.3	Simulation	106
1.4	Conclusion	107
2	Estimation des efforts à partir d'un modèle d'énergie paramétrable	108
2.1	Principe en trois dimensions	108
2.1.1	Evaluation de l'énergie	108
2.1.1.1	Détermination d'une fonction approchant localement l'énergie	108
2.1.1.2	Méthode des moindres carrés	109
2.1.2	De l'énergie modélisée à la force d'interaction	115
2.2	Conclusion	116

3	Modélisation de l'énergie d'interaction ligand-protéine	117
3.1	Détermination d'une forme de l'estimation	117
3.1.1	Forme de l'estimation pour la translation	117
3.1.1.1	Validation en simulation	119
3.1.1.2	Validation expérimentale	120
3.1.1.3	Conclusion	121
3.1.2	Forme de l'estimation pour la rotation	122
3.1.2.1	Formulation	122
3.1.2.2	Estimation des paramètres	123
3.1.2.3	Validation en simulation	125
3.1.2.4	Validation expérimentale	126
3.1.2.5	Conclusion	127
3.2	Formulation générale	127
3.3	Validation expérimentale	127
3.4	Conclusion	129
4	De la modélisation d'une énergie au calcul des efforts	129
4.1	Calcul des forces d'interaction	129
4.1.1	Validation expérimentale	130
4.1.2	Conclusion	131
4.2	Calcul des moments d'interaction	131
4.2.1	Validation expérimentale	132
4.2.2	Conclusion	133
5	Conclusion	133
	Conclusion et perspectives	137
	Liste des publications	143
	Bibliographie	145

TABLE DES FIGURES

1.1	Interaction du virus du sida avec un lymphocyte T4.	7
1.2	Propriétés structurales d'un acide aminé.	8
1.3	Formation de la liaison peptidique.	8
1.4	Caractéristiques d'un dipeptide.	9
1.5	Représentation d'une molécule en coordonnées cartésiennes.	10
1.6	Représentation d'une molécule en coordonnées internes	11
1.7	Représentation schématique du procédé QM/MM.	15
1.8	Energie potentielle d'une liaison covalente.	17
1.9	Angle de valence.	18
1.10	Angle de torsion.	19
1.11	Energie de torsion.	19
1.12	Angle impropre.	20
1.13	Forces de van der Waals.	21
1.14	Surface d'énergie conformationnelle.	25
1.15	Représentation biologique et robotique d'une molécule.	28
1.16	Représentation en arbre binaire d'une chaîne articulée.	29
1.17	Haptique moléculaire : approche par grille volumique.	34
1.18	Haptique moléculaire : changement de conformation d'un ligand au contact d'un récepteur avec minimisation de l'énergie.	36
1.19	Haptique moléculaire : passage d'une molécule au travers d'une protéine membranaire.	42
1.20	Interface à retour d'effort <i>Virtuose</i> couplée au logiciel MOE en salle de réalité virtuelle du CEA-LIST.	42
2.1	Principe du couplage virtuel mis en place.	49
2.2	Principe du couplage homothétique.	50
2.3	Architecture générale de la plate-forme de téléopération.	52
2.4	Temps de calcul de la partie moléculaire intégrant la minimisation de 250 atomes au cours d'une simulation de <i>docking</i> moléculaire.	53
2.5	Résultat de deux minimisations superposées pour la molécule de streptavidine.	53

2.6	Comparaison de la méthode adoptée, pour la minimisation de la protéine streptavidine et de son ligand, par rapport à une minimisation idéale.	54
2.7	L'espace des conformations à explorer dans le cas d'un squelette fixe est moins important que s'il était mobile.	55
2.8	Schéma de couplage.	56
2.9	Synopsis d'une opération de <i>docking</i> moléculaire.	57
2.10	Evaluation de l'énergie.	60
2.11	Dynamique moléculaire effectuée sur la molécule streptavidine et son ligand. . .	63
2.12	Repères associés aux différentes entités du système.	65
2.13	Changement de conformation du ligand pour une consigne en position nulle. . .	68
2.14	Changement de conformation du ligand au cours de sa manipulation.	69
2.15	Récepteur flexible au cours de la manipulation.	70
2.16	Manipulation de la biotine dans le site actif de la streptavidine à \vec{z} constant. La minimisation n'est pas activée.	71
2.17	Manipulation de la biotine dans le site actif de la streptavidine à \vec{z} constant. La minimisation est activée.	72
2.18	Manipulation de la biotine dans le site actif de la streptavidine à \vec{z} constant. Limitation des efforts.	73
2.19	Manipulation de la biotine dans le site actif de la streptavidine suivant une consigne en position linéaire.	74
2.20	Manipulation de la biotine dans le site actif de la streptavidine suivant une consigne en position linéaire.	75
2.21	Couplage sans retard du <i>Virtuose</i> sur un ressort virtuel.	76
2.22	Couplage avec retard du <i>Virtuose</i> sur un ressort virtuel.	77
3.1	Transformation en variables d'onde.	81
3.2	Couplage en variables d'onde d'une interface haptique avec une simulation. . . .	82
3.3	Codage et décodage des variables d'onde.	83
3.4	Commande de la simulation dans le cas idéal sans retard.	86
3.5	Modélisation de la transformation en variables d'onde à l'aide de blocs élémentaires. .	87
3.6	Chemins de rétroactions possibles lors d'un couplage en variables d'onde.	88
3.7	Codage et décodage d'une information en variables d'onde soumise à un retard au travers d'un filtre passe-bas.	89
3.8	Schéma de couplage utilisant les variables d'onde comme un amortisseur virtuel. . .	92
3.9	Codage et décodage des variables d'onde.	93
3.10	Comportement de la force d'interaction biotine-streptavidine selon l'axe \vec{x} . $B = 0$. . .	95
3.11	Comportement de la force d'interaction biotine-streptavidine selon l'axe \vec{x} . $B = 50$. . .	96
3.12	Schéma bloc des variables d'onde intégrées après leur décodage.	97
3.13	Comportement de la force d'interaction biotine-streptavidine selon l'axe \vec{x}	100
4.1	Comparaison de la dérivation analytique d'un potentiel donné avec l'estimation de la dérivée par l'utilisation de projecteurs.	106
4.2	Effet de la variation du gain ε sur la précision de l'estimation.	107
4.3	Représentation d'une surface de paramètres inconnus, assimilée à une énergie d'interaction entre un ligand et une protéine, en trois dimensions.	109
4.4	Obtention des paramètres de la fonction d'estimation afin d'approcher une surface d'énergie inconnue en deux positions.	113
4.5	Comparaison de la précision de la modélisation pour différentes tailles de la matrice \mathbf{A}	114

4.6	Comparaison de la précision de la modélisation pour différentes valeurs du gain G .	114
4.7	Approximation de l'énergie par la méthode des moindres carrés et calcul des forces associées.	116
4.8	Profils d'énergies obtenus pour des rotations et des translations du ligand dans le site actif d'une protéine selon les trois axes de coordonnées. Ces profils sont comparés à une fonction quadratique. Trois champs de forces différents sont comparés.	118
4.9	Estimation des paramètres du modèle d'énergie pour la translation.	120
4.10	Estimation des paramètres du modèle d'énergie pour la translation à partir de la position d'énergie minimum $\mathbf{X}_{e_{opt}}$	121
4.11	Représentation schématique du principe de minimisation de la fonction modèle dans le repère $(e_{\chi_x}^e, e_{\chi_y}^e, e_{\chi_z}^e)$ par rapport à la fonction réelle dans son repère $(e_{\chi_x}, e_{\chi_y}, e_{\chi_z})$	122
4.12	Estimation des paramètres du modèle d'énergie pour la rotation à partir d'une énergie consigne.	125
4.13	Estimation des paramètres du modèle d'énergie pour la rotation à partir de l'énergie d'interaction biotine/streptavidine.	126
4.14	Estimation des paramètres du modèle d'énergie faisant intervenir la rotation et la translation.	128
4.15	Estimation des paramètres du modèle d'énergie pour la translation et la rotation à partir de la position d'énergie minimum.	130
4.16	Estimation des paramètres du modèle d'énergie pour la translation et la rotation à partir de la position d'énergie minimum. Au cours de la manipulation, le ligand est soumis à une minimisation du champ de force.	132
4.17	Manipulation de la biotine aux abords du site actif de la streptavidine. Une commande en variables d'onde assure la stabilité du système. Les efforts sont calculés grâce à un modèle paramétrable.	134

NOTATIONS

Cette page regroupe l'ensemble des notations utilisées dans le présent manuscrit. Elles sont valables pour l'ensemble de celui-ci sauf s'il est précisé une notation au sein d'une équation.

t	Temps
k	Itération
P	Puissance
E	Energie
W	Travail d'une force
K_h	Coefficient de déplacement
\mathbf{K}_D	Matrice des coefficients de déplacement $[1 \times 3]$
\mathbf{K}_H	Matrice $[4 \times 4]$ des coefficients de déplacement et de rotation
K_f	Coefficient en effort
\mathbf{P}	Matrice immitance $[2 \times 2]$
\mathbf{X}_{indice}	Matrice de position. En général, l'indice indique l'objet représenté.
\vec{X}_{indice}	Vecteur position d'une particule, d'un objet indicé
$\vec{\dot{X}}_{indice}$	Vecteur vitesse d'une particule, d'un objet indicé

$\dot{\mathbf{X}}_{indice}$	Matrice des composantes de la vitesse
\vec{F} / \mathbf{F}	Force / Matrice des composantes tridimensionnelles de la force
\vec{M} / \mathbf{M}	Moment / Matrice des composantes tridimensionnelles du moment
$K - K_l - K_\theta$	Raideur d'un ressort - Raideur spécifique d'une liaison covalente - Raideur spécifique d'un angle de valence
C	Amortissement
Γ	Couples articulaires
\mathbf{J}	Matrice Jacobienne
\mathbf{I}	Matrice d'inertie
Z	Impédance
\mathbf{H}_h	Matrice homogène $[4 \times 4]$ du virtuose
\mathbf{H}_l	Matrice homogène $[4 \times 4]$ du ligand
\mathbf{R}	Matrice de rotation $[3 \times 3]$
\hbar	Constante de Planck
$\Psi(\vec{\mathbf{X}}, t)$	Fonction d'onde
\mathcal{T}	Fonction temporelle
H_ϕ	Opérateur de Fock
ϕ	Fonction d'orbitale spatiale / angle de rotation
r	Distance d'interaction entre deux atomes
l	Longueur d'une liaison covalente
ω_v	Fréquence de vibration
ω_T	Angle de torsion
ϵ_0, ϵ	Permittivité diélectrique du vide, permittivité diélectrique
q	Charge d'un atome
k_B	Constante de Boltzmann

T	Température
\mathbf{T}_v	Matrice homogène en vitesse $[4 \times 4]$
\mathbf{U}	Variable d'onde sortante $[6 \times 1]$
\mathbf{V}	Variable d'onde entrante $[6 \times 1]$
b	Raideur caractéristique d'une variable d'onde
τ	Retard de communication
\mathbf{W}	Matrice des retards de communication $[6 \times 6]$
λ	Fréquence de coupure d'un filtre

INTRODUCTION GÉNÉRALE

Un des premiers volets de la conception des médicaments concerne l'étude des cibles thérapeutiques responsables d'une maladie. En général, ces cibles sont un constituant d'une entité biologique associée à la maladie, tels un virus, une bactérie ou un parasite. Ce sont principalement des protéines ou une sous-classe comme les enzymes. Les protéines sont donc souvent les cibles choisies dans l'élaboration d'un médicament. Il faut identifier de façon spécifique leur structure tridimensionnelle, qui exprime la maladie. Une fois la protéine identifiée et sa structure tridimensionnelle déterminée, par exemple par cristallographie, il s'agit de bloquer son action.

Les protéines ont un ou plusieurs sites actifs dans leur conformation native. Le médicament (ligand) doit être choisi pour que son interaction avec la cible, à des endroits spécifiques, modifie cette conformation tridimensionnelle native et, par voie de conséquence, sa fonctionnalité nominale. Le changement de conformation peut engendrer une inhibition irréversible (la protéine est chimiquement détruite) ou réversible (la protéine est simplement liée au médicament sans changement de conformation). L'opération d'arrimage du ligand dans le site actif de la protéine est appelée *docking* moléculaire.

Au même titre que la modélisation tridimensionnelle de la protéine, il est possible de modéliser les interactions (ou liaisons) entre le ligand candidat et la protéine. Celles-ci, au niveau moléculaire entre le médicament et la cible, sont définies selon les groupements de molécules impliquées. Elles peuvent être des liaisons ioniques, hydrogènes, de van der Waals, ou une somme de plusieurs de ces liaisons. Les protéines et autres composés organiques sont dans un état stable dans leur environnement. Simuler le changement conformationnel d'une protéine revient à simuler les effets géométriques provoqués par une perturbation des liaisons, lors d'une quelconque excitation, en l'occurrence l'approche d'un ligand. Les énergies d'interaction sont modélisées, selon la précision recherchée, par la mécanique quantique ou par des modèles cherchant à réduire la complexité des phénomènes mis en jeu. Cette simplification ne trouve son intérêt que dans la réduction du temps de simulation du système. L'état natif d'une protéine étant énergétiquement stable, dans un puits de potentiel, simuler son évolution conformationnelle induite, par exemple, par une interaction extérieure revient à minimiser son énergie totale

mise en jeu.

La grande diversité des protéines et des propriétés physico-chimiques à simuler engendre une grande diversité de modèles d'interaction, chacun étant spécifique d'une particularité du système à modéliser. Une fois l'énergie totale du système modélisée, selon l'application envisagée, plusieurs classes d'algorithmes calculent le changement de conformation. Parmi celles-ci, la dynamique moléculaire explore l'espace des conformations possibles et les méthodes de minimisation optimisent la conformation initiale afin que l'énergie totale soit minimale.

Les logiciels de *docking* de molécules cherchent à identifier la position et l'orientation d'un ligand, maximisant ses interactions avec la protéine. Au cours de l'évolution du ligand dans le site actif, l'énergie totale du système est minimisée pour garantir la stabilité énergétique du système moléculaire. A l'aide de la description géométrique et chimique de la protéine, il est possible de cribler les molécules d'une librairie de ligands candidats qui peuvent interagir avec le site actif de la cible. Une première étape élimine de cette librairie toutes les molécules ne détenant pas les propriétés physico-chimiques nécessaires pour une interaction avec le site actif de la cible. L'étape suivante consiste à déterminer, parmi les molécules restantes, si un attachement avec le site actif est possible. Si cette condition est vérifiée, les algorithmes de *docking* quantifient l'affinité de l'attachement par une fonction coût. La plupart des algorithmes inclut en effet deux composantes. La première est une technique de recherche de l'emplacement optimal du ligand dans la cavité du site actif. La seconde est une fonction pour quantifier chaque emplacement possible afin de classer et comparer les ligands candidats entre eux. Les ligands candidats sont, par la suite, sélectionnés selon de nouveaux critères, notamment la solvabilité (le médicament est-il transportable par le sang?), la cinétique de dissolution et la toxicité.

Les outils logiciels actuels proposent la visualisation d'un modèle tridimensionnel du ligand et de la protéine. Cette visualisation permet de déterminer empiriquement les structures géométriques de ligands qui complètent le site catalytique, de façon optimale d'un point de vue de l'utilisateur. Lors de la conception du ligand, l'utilisateur peut améliorer une géométrie de ligand candidat en ajoutant ou retirant des atomes.

Une telle approche n'est cependant pas satisfaisante. L'aide visuelle fournie par l'ordinateur dans cette recherche de complémentarité ne rend pas directement compte des forces moléculaires s'exerçant entre le ligand et la protéine. Ainsi, la modification empirique de la géométrie du ligand suppose une bonne connaissance, *a priori*, des zones attractives et répulsives du site actif.

Cette difficulté peut être résolue par l'emploi d'un retour haptique pour les forces d'interaction. En couplant le modèle informatique du ligand à un organe à retour d'effort, l'opérateur devrait ressentir les forces intermoléculaires exercées par la protéine et déterminer si la procédure d'insertion du *docking* est réalisable. Un couplage haptique à six degrés de liberté autorise ainsi l'opérateur à ressentir à la fois les forces lors de la translation et les moments lors d'une rotation. Une fois la procédure d'insertion validée, la comparaison de l'affinité des différents candidats avec le site actif devient possible. Le retour d'effort est donc un élément de comparaison supplémentaire pouvant servir à la discrimination des différentes géométries candidates.

Cette approche émergente, les systèmes interactifs haptiques, se heurte encore à plusieurs obstacles majeurs. Les simulations moléculaires, de par la complexité algorithmique liée au calcul des conformations énergétiquement stables, ne sont pas temps réel. La stabilité d'un couplage bilatéral est donc directement affectée. Tous les modèles d'interactions atomiques

décrivent les liaisons en termes d'énergie. Les algorithmes d'optimisations conformationnelles utilisant l'énergie du système comme critère de performance, il n'est donc pas envisageable de dériver analytiquement le modèle d'interaction choisi pour obtenir les forces et les moments de l'interaction. Enfin, la grande diversité des modèles d'interaction et des méthodes d'optimisation de l'énergie rend la plate-forme de téléopération dépendante du choix initial d'un modèle d'interaction et d'une méthode d'optimisation.

Ces propriétés intrinsèques, des temps importants de simulation moléculaire, un système aux dimensions nanométriques, une approche énergétique des interactions et une grande diversité des méthodes de simulation ont conduit à imaginer une plate-forme de manipulation d'un ligand, avec six degrés de liberté en efforts, utilisant pleinement l'existant en terme de simulation moléculaire, un couplage bilatéral d'un système nanométrique original, stable et indépendante des outils logiciels choisis. Ce choix garantit une utilisation générique du couplage proposant ainsi une interactivité avec l'utilisateur en termes d'algorithmes.

Le présent manuscrit est composé de quatre chapitres, chacun dédié à un aspect spécifique du nanomonde et plus particulièrement à celui de la simulation moléculaire et de son couplage bilatéral avec le macromonde.

Le premier chapitre présente succinctement les propriétés structurales des protéines puis montre la diversité des modèles d'interactions et des simulations conformationnelles de ces systèmes. Un exposé de l'état de l'art en haptique moléculaire montre que très peu d'études concernent un retour d'effort avec six degrés de liberté. En revanche, elles constatent toutes que l'évaluation de l'énergie d'interaction n'est pas compatible avec une contrainte temps réel. En outre, elles mettent en avant l'instabilité des forces intermoléculaires. Les systèmes moléculaires présentés ont subi une réduction de leur espace des conformations en fixant, par exemple, certaines chaînes de la protéine et du ligand. La recherche conformationnelle en est ainsi simplifiée.

A partir de l'exploration de cet état de l'art, la plate-forme de téléopération est présentée. Elle se compose d'une partie dédiée à la simulation moléculaire, utilisant un logiciel existant, et du *Virtuose*, une interface capable d'assurer un retour de forces et de moments. L'objectif initial est de réaliser un *docking* avec un retour d'effort à six degrés de liberté pour un ligand et une protéine quelconques, sans réduction de l'espace des conformations.

Le second chapitre concerne l'étude du comportement de l'interface haptique et d'un ligand dans un site actif de protéine avec un couplage bilatéral homothétique force-position. L'objectif est, dans un premier temps, d'identifier les difficultés liées à la commande d'un système moléculaire. L'adéquation entre les déplacements macroscopiques de l'interface et ceux, nanométriques, du ligand est réalisée par un coefficient de déplacement fixe au cours de la manipulation. Par ailleurs, les efforts de l'ordre du nano-Newton sont exprimés dans le monde de l'opérateur par un coefficient en effort déterminé par la dynamique moléculaire. La commande du ligand, sans réduction de son espace des conformations, n'est pas temps réel. Une première approximation consistant à réduire les degrés de liberté de la protéine et du ligand est proposée. Des simulations valident cette approche réductrice. Cependant, les temps de simulation excèdent les limites acceptables pour la stabilité d'une communication entre une interface haptique et une simulation. De plus, les forces d'interaction calculées au cours du *docking* varient très fortement. Ce constat amène à envisager une nouvelle loi de commande pour réduire l'effet de la raideur des forces de van der Waals puis rendre stable la communication.

Le troisième chapitre prend en compte les problèmes de stabilité et de raideur des forces

mis en exergue au chapitre précédent. Un nouveau couplage réalisé avec les variables d'onde est présenté. Les variables d'onde sont étudiées comme ressort et amortisseur virtuels entre la simulation moléculaire et l'interface haptique, afin d'atténuer la forte variation des forces et des moments. L'avantage d'un tel couplage réside dans le fait qu'il s'affranchit de l'instabilité engendrée par les retards de simulation. Cette commande, nouvelle pour cette gamme d'application, doit garantir la stabilité de la plate-forme de téléopération et atténuer la grande variation des forces intermoléculaires observée au chapitre précédent. Cet objectif de stabilité et de filtrage des efforts est recherché afin de pouvoir conduire une réflexion sur les métaphores haptiques à mettre en place pour une bonne compréhension des interactions.

La simulation conformationnelle des systèmes moléculaires s'effectue grâce à une énergie calculée, par exemple, à partir d'un champ de force. Les modèles d'interactions représentés par la mécanique moléculaire, les méthodes hybrides et les champs de force, dépendent de la fonction biochimique à simuler ou, plus simplement, de la taille de la protéine. Cette diversité peut être un inconvénient si les forces à renvoyer à l'interface haptique sont directement calculées à partir d'une de ces méthodes. En effet, il ne serait dès lors plus possible d'appliquer un autre modèle d'interactions sans revoir le couplage. De plus, l'énergie ne peut être dérivée suivant la variation de position de l'interface haptique. Le quatrième et dernier chapitre s'affranchit des modèles d'interaction existants en modélisant l'énergie fournie par le simulateur par un modèle d'énergie dépendant de paramètres à estimer. La dérivée analytique des termes du modèle fournit les forces et les moments de l'interaction ligand-protéine. La solution analytique ne présente pas les instabilités jusqu'ici remarquées, notamment du fait de la lente convergence du modèle. Le couplage du nouveau modèle d'interaction avec les variables d'onde précédemment décrites est réalisé, finalisant ainsi la plate-forme de retour haptique pour le *docking* moléculaire.

La conclusion montre que les solutions apportées par cette thèse ne peuvent conduire, en l'état actuel des recherches, à un produit finalisé. Cependant, les perspectives envisagées, notamment un travail approfondi sur la transparence des commandes en variables d'onde et l'ajout de métaphores haptiques ou visuelles, amènent à penser que la plate-forme proposée pourrait atteindre ses objectifs initiaux.

Les perspectives présentées montrent notamment que le système conçu dans le cadre de cette thèse peut être appliqué, grâce à son indépendance par rapport aux méthodes énergétiques, à d'autres systèmes basés sur une évaluation de l'énergie. Ainsi, une plate-forme de téléopération initialement dédiée au *docking* moléculaire, pourrait être réutilisée dans un cadre plus général d'applications microscopiques.

HAPTIQUE ET SIMULATIONS MOLÉCULAIRES

L'étude du *docking* moléculaire avec retour d'effort suppose que soient connues les spécificités des contraintes imposées par la simulation conformationnelle des protéines pour son couplage bilatéral avec une interface haptique. L'objet de ce chapitre est d'explicitier les différentes propriétés de la simulation des molécules avec, pour objectif final, une interactivité totale entre un bras à retour d'effort et une plate-forme de *docking* moléculaire.

Il est nécessaire de comprendre la structure des protéines pour envisager leur simulation. Ainsi, dans une première partie sont exposées succinctement leurs principales propriétés structurales. Cette description met en lumière les différentes propriétés des protéines et la complexité de modéliser de façon précise leur comportement.

Dans un deuxième temps sont abordées les méthodes de modélisation de l'énergie interatomique. Par modélisation moléculaire, il faut entendre la recherche de positions appelées conformations, dans lesquelles l'énergie moléculaire est minimum. De nombreuses méthodes sont proposées comme la mécanique quantique, les méthodes empiriques ou hybrides. L'ensemble des champs de force est présenté. Les gradients de ces champs aux échelles moléculaires engendrent une complexité forte pour le couplage homothétique bilatéral. A partir de l'exploitation des champs, des méthodes classiques de minimisation peuvent être exploitées pour déterminer la conformation. La simulation moléculaire exploite ainsi de nombreuses méthodes classiques, avec des temps de calculs longs et problématiques pour le couplage. De même, pour des raisons de complexité algorithmique et de temps de calcul, des modèles robotiques robustes sont adaptés à l'exploration conformationnelle. La protéine est considérée comme étant une chaîne polyarticulée. Nous justifierons ensuite le choix des méthodes empiriques pour notre approche.

La troisième partie présente les travaux existants portant sur l'ajout de retour d'effort dans des simulations moléculaires. Issu d'un réel besoin d'interactivité pour accélérer les calculs de dynamique moléculaire, le retour d'effort est abordé avec une réduction de la complexité des systèmes moléculaires. Le calcul des forces à ressentir est en général issu d'un simple couplage

bilatéral. L'ensemble des travaux sera présenté en montrant la nécessité d'une approche générale sans fixer de propriétés fortes pour simplifier le calcul de la simulation moléculaire.

De l'étude des paragraphes précédents découle la plate-forme de téléopération mise en place au cours de ces travaux. En fonction des besoins exprimés par le partenaire industriel du projet, une interface haptique a été sélectionnée, ainsi que les environnements pour la simulation interactive. Cette plate-forme est présentée dans une dernière partie.

1 Généralités

1.1 Du virus au médicament

Les virus sont de petites particules biologiques de 20 à 300 *nm* de diamètre dont la structure résulte de l'assemblage de deux ou trois éléments (figure 1.1(a) page ci-contre) : le génome (un acide nucléique, ARN ou ADN), une capsid (assemblage de petites protéines virales identiques dont le rôle est la protection du génome viral en milieu extérieur) et parfois une enveloppe (celle-ci porte les déterminants viraux qui se lient spécifiquement à des récepteurs cellulaires). Après fixation aux récepteurs de la membrane cellulaire, l'enveloppe, de même nature lipidique, peut fusionner avec cette membrane, introduisant la nucléocapsid dans la cellule.

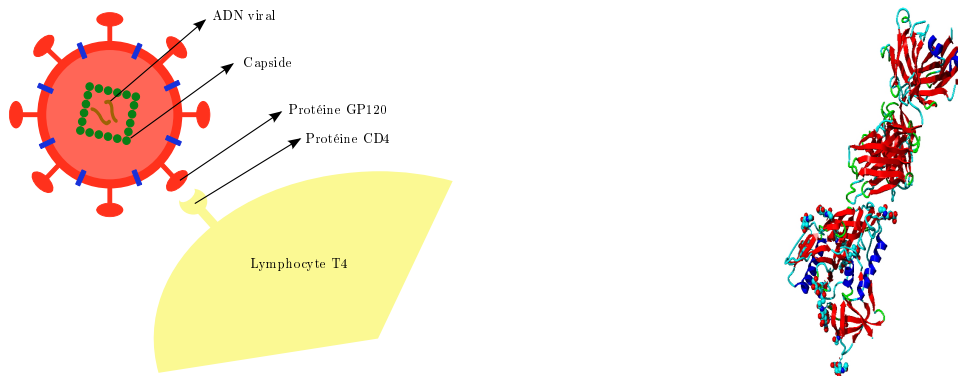
1.1.1 Exemple simplifié du virus du sida (VIH)

Le virus du sida présent dans le sang est capable de se fixer sur des cellules particulières du système immunitaire, par exemple les lymphocytes T4. Ces lymphocytes sont ainsi nommés car ils sont porteurs de la protéine transmembraire CD4. La fixation du virus à ces cellules fait intervenir le complexe CD4 (reconnu par la protéine GP120 du virus), ainsi que d'autres protéines membranaires (les corécepteurs). A partir de cette fixation, le matériel génétique du virus peut pénétrer dans le lymphocyte (figures 1.1(a) et 1.1(b) page suivante).

Une fois dans le cytoplasme, l'ARN du virus est rétrotranscrit en ADN double brin. Cet ADN pénètre dans le noyau et s'intègre au génome de la cellule hôte. L'expression des gènes du virus provoque alors la fabrication de ses protéines. Assemblées, elles forment de nouveaux virions qui bourgeonnent autour de la cellule, s'entourant d'une membrane héritée de celle qui est infectée. Ce mécanisme libère ainsi de nouveaux virus dans le sang de l'organisme infecté.

1.1.2 Vers un médicament

Le médicament est responsable de la désactivation des fonctions d'intégration du virus à la cellule. L'objectif est de concevoir une petite protéine (ligand) dont les propriétés structurales sont proches du récepteur membranaire complémentaire (le complexe CD4 pour l'exemple étudié). Une fois présent dans l'organisme, le ligand se fixera sur les récepteurs du virus, celui-ci ne pouvant plus, par conséquent, se lier à une protéine membranaire. Le virus est toujours présent mais considéré comme inhibé.



(a) Fixation du virus du sida sur un récepteur membranaire d'un lymphocyte T_4 grâce à sa protéine membranaire GP120 compatible avec le complexe CD_4 de la cellule cible.

(b) Représentation moléculaire des complexes GP120 et CD_4 agissant lors de l'interaction du virus du sida avec un lymphocyte T_4 .

Figure 1.1 - Représentation schématique et moléculaire d'une interaction du virus du sida avec un lymphocyte T_4 .

Le processus de recherche d'un ligand dont les caractéristiques géométriques et physiques sont compatibles avec un site actif de protéine est appelé *screening* virtuel s'il s'agit de tester plusieurs centaines de formes géométriques différentes, sinon *docking* moléculaire. Ces méthodes, actuellement entièrement automatisées, consistent à identifier la position et l'orientation d'un ligand maximisant ses interactions avec la protéine. Une autre étape est d'approcher le ligand récemment créé du site actif d'une protéine, afin de s'assurer que celui-ci est attiré par le complexe protéique grâce aux forces intermoléculaires.

Pour garantir une bonne modélisation du *docking* moléculaire, il faut connaître les propriétés des molécules et les outils qui les modélisent et les simulent. Les paragraphes suivants font l'objet de cette étude.

1.2 Biologie structurale

1.2.1 Les protéines

Les cellules sont en grande partie constituées de protéines qui représentent plus de la moitié de leur poids sec. Les protéines déterminent la forme et la structure de la cellule et servent aussi d'instruments principaux de reconnaissance moléculaire et de catalyse. Ces longues chaînes d'acides aminés sont appelées chaînes polypeptidiques. Une chaîne de moins de dix acides aminés est appelée peptide, une chaîne comprenant entre dix et cent acides, polypeptide et enfin une chaîne de plus de cent acides aminés, protéine [Mitchell 99] [Alberts 94].

1.2.2 Les acides aminés et peptides

Les acides aminés sont chimiquement variés mais ils contiennent tous un groupement acide carboxylique et un groupement amine portés par le même atome de carbone appelé « carbone- α » (figure 1.2) [Branden 91].

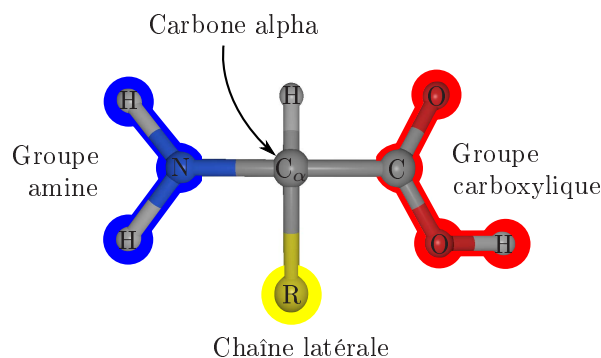


Figure 1.2 - Propriétés structurales d'un acide aminé. R représente une chaîne latérale qui est spécifique à chaque acide aminé.

Ils servent de sous-unités dans la synthèse des protéines, longs polymères linéaires d'acides aminés attachés les uns aux autres par une liaison peptidique entre le groupement carboxyle d'un acide aminé et la fonction amine du suivant. Bien qu'il existe de nombreux acides aminés différents, seuls vingt sont retrouvés dans les protéines, chacun possédant une chaîne latérale différente liée au carbone- α . Les propriétés des chaînes latérales, dans leur ensemble, déterminent les propriétés des protéines et sont à l'origine de toutes leurs fonctions sophistiquées.

1.2.3 Liaison peptidique

La liaison peptidique est la réaction d'accrochage d'un premier acide aminé avec un deuxième. Il s'agit d'une condensation avec élimination d'une molécule d'eau entre le groupe carboxylique du premier acide aminé et la fonction amine de la deuxième molécule (figure 1.3).

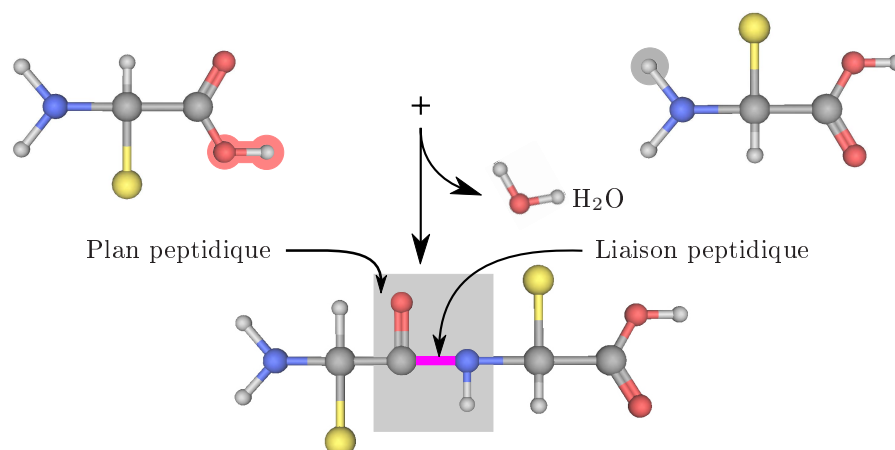


Figure 1.3 - Formation de la liaison peptidique.

Cette liaison est caractérisée par le fait que le centre des quatre atomes CONH et les deux atomes carbone- α des acides aminés sont situés dans un même plan, le plan de la liaison peptidique. En effet, aucune rotation autour de l'axe C-N n'est possible du fait de la présence de l'oxygène, fortement électro-attracteur. Les électrons de la liaison peptidique vont être attirés du côté de l'oxygène en polarisant ainsi la liaison C-N. L'oxygène prend une charge négative partielle alors que l'azote donne son doublet d'électrons libres et prend une charge positive partielle. De ce fait, les électrons entre l'oxygène et l'azote sont délocalisés, entraînant la rigidité du plan peptidique.

En général, les deux atomes carbone- α se placent de part et d'autre du pont C-N dans la configuration *trans* (figure 1.4(a)), la plus favorable thermodynamiquement. En outre, elle présente moins de gêne stérique pour les deux groupes *R* adjacents. Par opposition à la configuration *trans*, la protéine se trouve dans la configuration *cis* lorsque l'atome d'oxygène de la double liaison C=O est du même côté de la liaison peptidique.

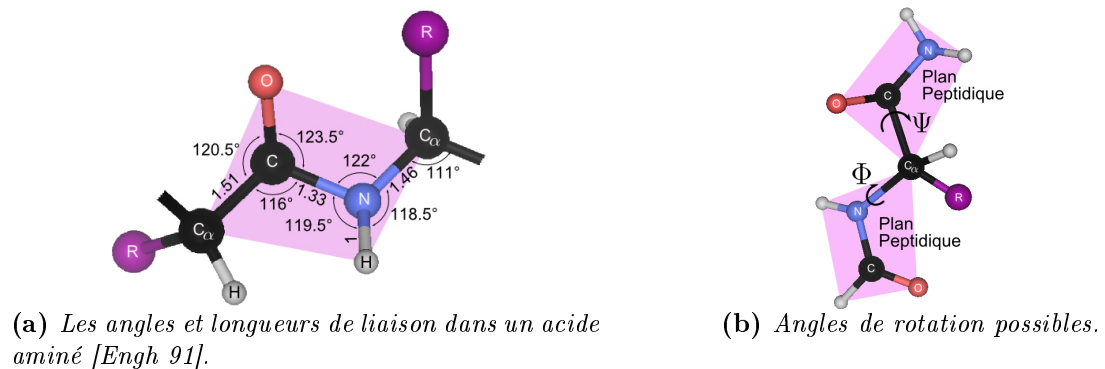


Figure 1.4 - Caractéristiques d'un dipeptide.

La position spatiale d'un plan amide par rapport à son voisin est caractérisée en définissant deux angles de rotation (figure 1.4). L'angle psi (ψ) qui est l'angle de rotation autour de la liaison C $_{\alpha}$ -C (sur le groupement C=O) d'un premier plan, et l'angle phi (ϕ) qui est l'angle de rotation autour de la liaison C $_{\alpha}$ -N (sur le groupement N-H) du deuxième plan. Toutes les valeurs d'angles ne sont pas permises. Le diagramme de Ramachandran montre les valeurs possibles d'un angle en fonction de l'autre (pour les structures tridimensionnelles connues) [Richardson 81].

Les deux configurations *cis* et *trans* sont déterminées par l'angle ω formé par les deux plans (C $_{\alpha}$ -C-N) et (C-N-C $_{\alpha}$) autour de la liaison peptidique. Celui-ci prend uniquement les valeurs de 0° pour la première et 180° pour la seconde. La figure 1.4(a) représente donc la conformation *trans*.

1.2.4 Conclusion

Les protéines sont des chaînes polypeptidiques possédant des caractéristiques géométriques bien définies. Elles possèdent un grand nombre de propriétés. La simulation de leur comportement doit donc les intégrer et les préserver afin de garantir la validité de la simulation.

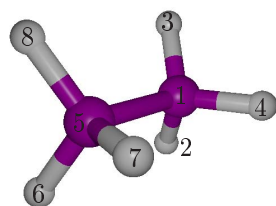
L'ensemble des paramètres, déterminés empiriquement par cristallographie ou résonance magnétique nucléaire (RMN), est regroupé au sein de fichiers *Protein Data Bank* (PDB) afin de

pouvoir représenter et simuler ces systèmes.

1.2.5 Représentation des protéines en mécanique moléculaire

Une molécule peut être représentée de différentes façons. La plus détaillée est le modèle *tout-atome* où tous les atomes de la molécule sont explicitement décrits. Une première simplification de ce modèle consiste à intégrer les hydrogènes non polaires aux atomes auxquels ils sont liés, c'est le modèle *atomes-unis* [Weiner 84]. La chaîne latérale peut être approchée comme une seule sphère possédant un rayon et des propriétés chimiques qui varient. C'est le *modèle-sucette* [Levitt 76]. Enfin, la structure secondaire peut être représentée par des vecteurs [Singh 97].

1.2.5.1 Représentation en coordonnées cartésiennes



Type d'atome	N°	Nom	Molécule	x	y	z
HETATM	1	C	ETH	0.452	-0.007	0.005
HETATM	2	H	ETH	0.805	-1.005	0.005
HETATM	3	H	ETH	0.805	0.492	-0.860
HETATM	4	H	ETH	0.805	0.492	0.869
HETATM	5	C	ETH	-1.056	-0.007	0.005
HETATM	6	H	ETH	-1.420	-1.037	0.005
HETATM	7	H	ETH	-1.420	0.508	0.897
HETATM	8	H	ETH	-1.420	0.508	-0.888
CONNECT	1	4	3	2	5	
CONNECT	5	8	7	6		

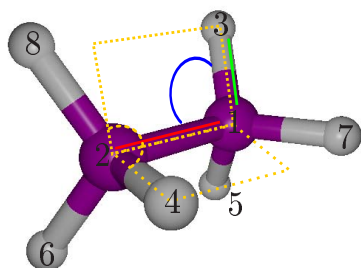
Figure 1.5 - Représentation d'une molécule d'éthane en coordonnées cartésiennes (extrait d'un fichier PDB). HETATM représente un atome d'un groupe non standard, ETH est le nom de la molécule. Les paramètres CONNECT relient les atomes entre eux.

La structure est décrite par la position tridimensionnelle de tous ses atomes. Chaque atome possède un jeu de trois coordonnées. La protéine est alors regardée comme une superposition de positions sans lien explicite entre elles. En effet, la structure covalente n'est pas définie. Lors des simulations de dynamique moléculaire, des précisions sur ces liaisons doivent être apportées, notamment par l'introduction de ressorts qui maintiennent les longueurs de valence proches de leur valeur d'équilibre. Cette représentation est assez physique car les liaisons peuvent se dissocier lors des simulations à température élevée. De plus, les interactions de van der Waals, basées sur la distance entre atomes non liés, peuvent être facilement estimées. Cependant, la majorité des simulations ne concerne pas la dissociation des atomes. Il peut donc être intéressant d'intégrer directement dans le modèle des données sur la structure protéique et ainsi diminuer le nombre de paramètres.

1.2.5.2 Représentation en coordonnées internes

L'analyse des structures des protéines, obtenues expérimentalement et répertoriées dans la *Protein Data Bank* [Berman 00], montre que les longueurs des liaisons varient très peu. La flexibilité d'une protéine n'est donc due qu'en très grande majorité à la rotation autour de la

liaison covalente. Un résidu (acide aminé) possède trois angles de rotation et donc trois degrés de liberté. Ces angles sont les angles ϕ (autour de N-C α), ψ (autour de C α -C) et ω autour de la liaison peptidique (vaut 0 ou 180°, cette dernière valeur étant plus stable énergétiquement, elle est souvent préférée et fixée). Deux degrés de liberté suffisent donc pour décrire assez précisément le squelette du résidu. Dans cette représentation, la chaîne secondaire n'a qu'entre zéro (glycine et alanine) et quatre (arginine) degrés de liberté en torsion. Un résidu est donc décrit par, au plus, 7 degrés de liberté.



1	C							
2	C	1.54	1					
3	H	1.0	1	109.5	2			
4	H	1.0	2	109.5	1	180.0	3	
5	H	1.0	1	109.5	2	60	4	
6	H	1.0	2	109.5	1	-60	5	
7	H	1.0	1	109.5	2	180.0	6	
8	H	1.0	2	109.5	1	60	7	

Figure 1.6 - Représentation d'une molécule d'éthane en coordonnées internes. L'atome de carbone 1 est fixé, toute la molécule est décrite par rapport à lui en termes de longueur de liaison (rouge, vert) et d'angles dièdres (bleu) ou peptidiques (jaune).

Cette représentation décrit la position de chaque atome en fonction des trois qui lui sont liés par covalence. Ainsi, les noms des atomes liés sont connus et trois paramètres complètent le modèle : une longueur de liaison, un angle de valence et un angle dièdre. Ainsi, l'atome 4 est décrit à partir de sa distance à l'atome 2 (la longueur de la liaison 4-2), son angle formé par les liaisons 1-2 et 2-4 et par son angle dièdre entre les plans définis par 1,2 et 3 et celui défini par 1,2 et 4. Cette représentation n'utilise que $3N - 6$ paramètres puisque la position du premier atome est fixée arbitrairement, le second atome est spécifié simplement avec sa distance de liaison et l'angle dièdre n'est pas spécifié pour le troisième. Des contraintes peuvent facilement être imposées dans les liaisons et toute l'information topologique est présente. Cependant, il est très coûteux en temps de calcul de trouver les positions relatives des atomes non liés pour le calcul des forces à longue portée.

1.2.5.3 Conclusion

Beaucoup de travaux concernent l'optimisation de la représentation des atomes dans un système de coordonnées afin de pouvoir décrire très précisément leur position lors d'un changement de conformation. La représentation la plus couramment utilisée, lors des simulations de changements de conformations, décrit la position des atomes les uns par rapport aux autres, en termes d'angles de torsion. Cette méthode sera utilisée dans la suite de ce manuscrit.

2 Simulation moléculaire

La simulation moléculaire a pour but de comprendre et de prévoir les propriétés physico-chimiques macroscopiques d'une substance à partir de la simulation de ses constituants à l'échelle microscopique. Une simulation d'un comportement protéique comporte une phase de **modélisation des interactions** afin d'en déduire l'énergie du système. Cette phase est déduite de l'équation de Schrödinger ou de la description empirique du champ de force. Le deuxième niveau consiste, à partir de cette énergie, à simuler le mouvement des atomes pour obtenir la position d'énergie minimum garantissant une stabilité énergétique. Ces positions atomiques peuvent être obtenues, par exemple, grâce à l'emploi d'algorithmes de minimisation de l'énergie.

Deux orientations différentes sont à distinguer pour simuler l'évolution d'un système moléculaire. La première orientation concerne les méthodes issues du *génie biologique* et utilise des méthodes adaptées de minimisation de l'énergie. La deuxième est une nouvelle application des théories issues de la robotique. Ces dernières assimilent un système polyarticulé à une molécule et diminuent alors les temps de simulation.

L'objet de ce paragraphe est d'apporter des éléments de réponse quant à la sélection d'une méthode pour l'évaluation de l'énergie interatomique et pour la simulation conformationnelle. Les contraintes majeures pour la discrimination de ces méthodes concernent le temps de calcul et l'analogie des résultats avec la biologie moléculaire. Outre le choix d'un modèle d'interaction et d'une méthode de simulation conformationnelle, l'attention devra être portée sur la rapidité des solutions conformationnelles fournies en cours de manipulation.

Les méthodes issues du *génie biologique* modélisent l'énergie du système grâce à trois classes de modèles qui coexistent [Bowen 91], [Allinger 76] :

1. les modèles dits *ab initio*, basés uniquement sur la mise en équation des principes de la mécanique quantique ;
2. les modèles dits *semi-empiriques*, simplifications des précédents. Ils négligent certains termes jugés « secondaires » (comme des interactions à longue distance par exemple) et utilisent des données ajustées sur des résultats expérimentaux afin de simplifier les calculs ;
3. les modèles dits *empiriques*, dans lesquels la mécanique quantique est abandonnée pour être remplacée par la mécanique classique newtonienne.

Les modèles *robotiques* dont la particularité est d'identifier une protéine à un manipulateur polyarticulé utilisent des méthodes d'optimisations robustes. Les contraintes internes des protéines (liaisons peptidiques, plans peptidiques, angles, etc.) sont ainsi traduites en contraintes géométriques qu'il faut résoudre en vue d'obtenir un système stable.

L'applicabilité de ces méthodes varie de plusieurs ordres de grandeur : un modèle *ab initio* est aujourd'hui limité à la simulation de systèmes d'une centaine d'atomes alors que par dynamique moléculaire et en utilisant un modèle des interactions empirique, des systèmes de plusieurs milliers d'atomes peuvent être convenablement simulés. Les systèmes informatiques devront toutefois largement évoluer afin de pouvoir les simuler avec une composante interactive. Une alternative consiste à mettre en place des approximations des énergies d'interaction afin de limiter le temps de calcul.

De même, la résolution numérique d'un système polyarticulé contenant des milliers de degrés

de liberté ne peut être réalisée en temps réel avec les méthodes issues de la robotique.

La simulation de *docking* moléculaire interactive avec retour d'effort nécessite un calcul rapide des forces d'interaction, afin de satisfaire la contrainte temps réel nécessaire à la stabilité d'une commande à retour d'effort. Le choix entre les différentes méthodes de modélisation de l'énergie intermoléculaire doit donc intégrer cette spécificité. Une fois l'énergie modélisée, une méthode de simulation du comportement protéique appropriée au *docking* doit être appliquée au système, afin d'observer l'évolution spatiale des atomes de la molécule et garantir que les propriétés physicochimiques de la protéine soient préservées.

Les paragraphes suivant présentent ces différentes méthodes de modélisation et de simulation des protéines issues du génie biologique ou de la robotique. Elles diffèrent toutes par leur précision acquise au détriment du temps de calcul.

2.1 Méthodes issues du génie biologique

Ce paragraphe décrivant les méthodes de simulation issues de la biologie, s'attache à montrer, de manière succincte, la complexité du problème, tant dans sa formulation que dans sa résolution. Le choix final de la méthode dépend de la rapidité de calcul, de la précision du modèle et de son indépendance vis-à-vis des systèmes moléculaires.

La modélisation de l'énergie d'interaction peut être décrite par la mécanique quantique, des méthodes hybrides ou des méthodes empiriques. La simulation du comportement consiste à calculer, à partir de l'énergie précédemment décrite, la conformation d'énergie minimum, par exemple, grâce à des algorithmes de minimisation, de dynamique moléculaire ou de Monte-Carlo.

2.1.1 Modélisation des interactions

Une méthode de simulation moléculaire est basée sur l'évaluation préalable des énergies d'interaction atomiques. Une fois cette énergie évaluée, il devient alors possible de calculer les changements de conformation de la protéine.

2.1.1.1 Mécanique quantique

La mécanique quantique se ramène très souvent à l'étude de l'équation de Schrödinger. Celle-ci a été proposée en 1926 et s'est développée d'abord dans le but de décrire les atomes constitués d'une seule particule située dans un certain champ de force (l'électron au sein de l'atome d'hydrogène par exemple). L'objet central de la théorie de Schrödinger est une fonction $\Psi(\vec{X}, t)$ à valeurs complexes, appelée fonction d'onde. Cette fonction satisfait [Leach 01] :

$$i\hbar\frac{\partial}{\partial t}\Psi(\vec{X}, t) = -\frac{\hbar^2}{2m}\Delta\Psi(\vec{X}, t) + E_p(\vec{X})\Psi(\vec{X}, t) \quad (1.1)$$

où \vec{X} est le rayon-vecteur repérant la particule dans l'espace ($\vec{X} = x\vec{x} + y\vec{y} + z\vec{z}$), $E_p(\vec{X})$ est l'énergie potentielle de la particule étudiée, Δ est l'opérateur de Laplace, i désigne le nombre imaginaire pur, m représente la masse de la particule, \hbar la constante de Planck [Head-Gordon 96].

Cette équation dépend du temps t . Si l'approximation de Born-Oppenheimer, qui stipule que les électrons s'adaptent instantanément aux mouvements des noyaux, est supposée (leurs masses sont suffisamment négligeables par rapport à celle du noyau), la fonction d'onde s'écrit comme le produit d'une fonction spatiale et temporelle : $\Psi(\vec{X}, t) = \psi(\vec{X})\mathcal{T}(t)$. Dans ce type de modèle, ne sont considérées que les situations où le potentiel est indépendant du temps. L'équation s'écrit alors :

$$E\Psi(\vec{X}) = -\frac{\hbar^2}{2m}\Delta\Psi(\vec{X}) + E_p(\vec{X})\Psi(\vec{X}) \quad (1.2)$$

où E est l'énergie de la particule.

Cette équation aux dérivées partielles ne peut être résolue précisément que dans certains cas simples et ne peut donc pas être utilisée pour décrire un phénomène de l'ordre de la protéine. Son calcul reste très lourd, même dans les cas triviaux, car elle reste fonction des $3n$ coordonnées des n électrons.

Certaines études portent sur la recherche d'approximations des solutions de l'équation car elles rendent compte précisément de l'énergie totale du système. La plus significative fut proposée par Hartree-Fock. Elle a pour but de remplacer les fonctions tests Ψ par des fonctions qui sont un déterminant de N fonctions $L^2(\mathbb{R}^3)$. La forme condensée de l'équation s'écrit [Leach 01] :

$$H_\phi\phi_i = E_i\phi_i \quad (1.3)$$

ici E_i est l'énergie de l'orbitale i , H_ϕ est l'opérateur de Fock, et ϕ_i est la fonction d'orbitale spatiale.

Cette équation est fortement non linéaire et entraîne une complexité algorithmique en N^4 , N étant le nombre d'orbitales. Cette complexité est une des plus faibles pour ce type de méthode [Bris 02]. Par exemple, les méthodes *ab initio* ont des complexités du type kN^p , où $p = 3, 4, \dots, 8$ et k est le nombre de noyaux. Cette propriété pose d'énormes problèmes pratiques au regard de la longueur des simulations : la simulation d'une nanoseconde d'un système moléculaire requiert un million de pas de temps sur la dynamique des noyaux (le temps caractéristique de leur vibration est de 10^{-15} s), donc la résolution d'un million de problèmes de type (1.3). L'objet de cette thèse étant d'interagir en temps réel avec une molécule, ces méthodes quantiques ne peuvent simplement pas être utilisées.

2.1.1.2 Méthodes hybrides

De par la complexité de la résolution des équations précédentes, la communauté scientifique s'est tournée vers des méthodes alliant la précision de l'équation de Schrödinger et l'approximation des méthodes empiriques. Elles sont appelées hybrides ou semi-empiriques [Kollman 00].

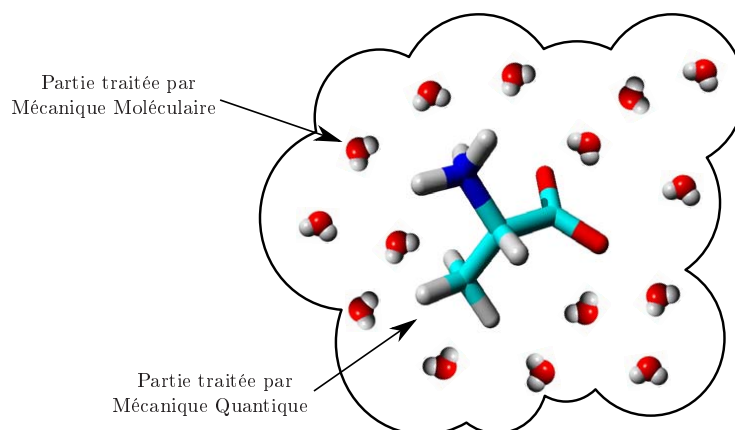


Figure 1.7 - Représentation schématique du procédé QM/MM.

Une méthode semi-empirique est une méthode dans laquelle une partie des calculs nécessaires au modèle de Hartree-Fock est remplacée par des paramètres ajustés sur des valeurs expérimentales. En général, toutes ces méthodes sont très précises pour des familles de produits données voisines de celles utilisées pour la paramétrisation [Fields 00], [Liu 98].

Le but de la méthode est de séparer le problème en deux parties : la première est traitée selon des méthodes de mécanique quantique, la deuxième par la mécanique moléculaire. Sur la figure 1.7, la partie moléculaire est résolue par mécanique quantique, alors que le solvant est résolu par la mécanique moléculaire car il est beaucoup plus important et peut développer beaucoup d'interactions avec la molécule. Le traitement sera donc plus rapide.

L'énergie E_{Total} du système s'écrit :

$$E_{Total} = E_{QM} + E_{MM} + E_{QM/MM} \quad (1.4)$$

où E_{QM} est l'énergie de la partie du système traitée par mécanique quantique et E_{MM} est l'énergie de la partie décrite par mécanique moléculaire. $E_{QM/MM}$ n'est due qu'à des interactions non liantes entre la partie QM (mécanique quantique) et MM (mécanique moléculaire).

Ces méthodes calculent les interactions intraprotéiques par mécanique quantique. Lors de la simulation du comportement du ligand au sein de la protéine, le solvant n'est pas considéré. La protéine est donc simulée dans le vide. Les équations qui régissent l'énergie de la protéine sont donc les équations de la mécanique quantique dont le temps de calcul n'est pas compatible avec une application temps réel. Dans le cadre des travaux présentés dans ce manuscrit, cette méthode de modélisation des interactions n'est donc pas retenue au profit d'une méthode plus rapide, empirique.

2.1.1.3 Mécanique moléculaire ou méthodes empiriques

La mécanique moléculaire est basée sur la mécanique newtonienne. Elle postule que la molécule est représentée par un ensemble d'atomes ponctuels, de masse et de charge électrique

définies, reliés entre eux par des liaisons covalentes déformables de type ressort. La mécanique moléculaire est régie par trois postulats [Lomas 86] :

1. L'énergie potentielle E correspondant à la géométrie du système est décrite comme une somme discrète de termes dépendants des coordonnées atomiques. D'une façon générale, la fonction énergie possède une forme dérivant d'un développement en série de Taylor, limité le plus souvent au second ordre, par rapport aux coordonnées internes (équation 1.5). Pour de petits déplacements des $3N-6$ coordonnées X_i par rapport aux coordonnées de référence (X^0, E_0) , l'énergie potentielle E se présente sous la forme suivante :

$$E = E_0 + \sum_i^{3N-6} \left(\frac{\partial E}{\partial X_i} \right)_0 (X_i - X_i^0) + \frac{1}{2} \sum_i^{3N-6} \sum_j^{3N-6} \left(\frac{\partial^2 E}{\partial X_i \partial X_j} \right)_0 (X_i - X_i^0)(X_j - X_j^0) \quad (1.5)$$

2. La notion de transférabilité est très importante en mécanique moléculaire (MM). L'objectif est de décrire un modèle mathématique pour un ensemble de molécules et non pas une seule. Cette exigence impose d'introduire des termes qui ont des formes analytiques spécifiques. C'est pourquoi, l'énergie potentielle totale peut être divisée en termes indépendants, trouvant leur origine dans une description géométrique de la molécule, ou mettant en jeu des interactions à longue distance.

$$E = \underbrace{E_{\text{Elongation}} + E_{\text{Angle}} + E_{\text{Torsion}}}_{\text{Coordonnées internes}} + \underbrace{E_{\text{van der Waals}} + E_{\text{Electrostatique}}}_{\text{Atomes non liés}} + E_{\text{Supp.}} \quad (1.6)$$

Les termes $E_{\text{Elongation}}$, E_{Angle} et E_{Torsion} représentent la fonction E de l'équation 1.5. Des contributions E_{Supp} spécifiques supplémentaires peuvent apparaître suivant le modèle mathématique utilisé. Ce terme est à l'origine d'un nombre assez important de programmes différents modélisant chacun leur propre effet.

3. Le système est d'autant plus stable que son énergie totale est faible.

Les avantages d'une telle approche empirique résident dans la rapidité accrue des calculs et la maîtrise des termes énergétiques employés dans l'expression de l'énergie totale. Le traitement de molécules de grande taille (de plusieurs centaines d'atomes à des systèmes complexes) est rendu possible par la simplicité de la forme analytique et la transférabilité des termes énergétiques. Cependant, le formalisme de la mécanique moléculaire impose que les liaisons covalentes soient élastiques. Il n'est donc pas possible de modéliser parfaitement, et de façon répétitive, des ruptures de liaisons, même s'il est possible de l'appréhender grâce au résultat de la simulation. Les atomes sont définis comme une sphère et une charge calculée empiriquement et indépendamment de la conformation. Il est, par conséquent, impossible de modéliser des phénomènes de variations importantes de densités électroniques. Pour ce genre de modélisation, il faut avoir recours aux méthodes semi-empiriques ou *ab initio* décrites précédemment.

Cependant, ressentir sur un bras à retour d'effort des interactions moléculaires impose que la simulation des molécules soit réalisée en temps réel, ou dans le même ordre de grandeur, soit 1 *KHz*. De plus, une interaction ligand-protéine ne nécessite pas la création ni la destruction de liaison. Cette approche, certes moins précise que les méthodes quantiques, est donc à privilégier dans le cadre de ces travaux afin de garantir la stabilité haptique recherchée.

Le champ de force. Celui-ci contient la totalité des fonctions énergétiques nécessaires pour calculer l'énergie potentielle E du système et ses dérivées partielles $\frac{\partial E}{\partial X_i}$ et $\frac{\partial^2 E}{\partial X_i \partial X_j}$. Il inclut

les valeurs d'équilibre X_j^0 de ces fonctions, ainsi que les constantes de forces associées. Les paramètres du champ de force sont issus de **mesures expérimentales** (RMN, cristallographie aux rayons X ...).

Les multiples champs de force existants (une dizaine) diffèrent plus par leur jeu de paramètres et leur degré de réalisme pour modéliser des espèces chimiques particulières (protéines par exemple) que par la forme analytique des contributions énergétiques. Lorsque Allinger propose le premier champ de force en 1967 (MM1) [Cooper 03], la précision obtenue était telle (0.01 Å de longueur de liaison et 0.5° d'erreur dans les angles) que, très vite, il devint une référence de comparaison pour tous les nouveaux champs émergents. Certains problèmes mis en évidence, notamment pour l'énergie de l'hydrogène, ont été résolus par le champ MM2 [Allinger 77] puis MM3 [Allinger 89b] et MM4 [Allinger 96]. A l'heure actuelle, les champs MM4 (1996) et MMFF94 (1994) [Halgren 96b] sont considérés comme les plus évolués dans le domaine. Le champ MM4 qui est certainement le plus utilisé puisque ses particularités sont reprises par tous les autres champs, est décrit dans la suite de ce chapitre.

2.1.1.3.1 La liaison covalente

La figure 1.8(a) montre la forme typique de l'énergie potentielle pour ce type de liaison, de la position de plus forte énergie à la dissociation.

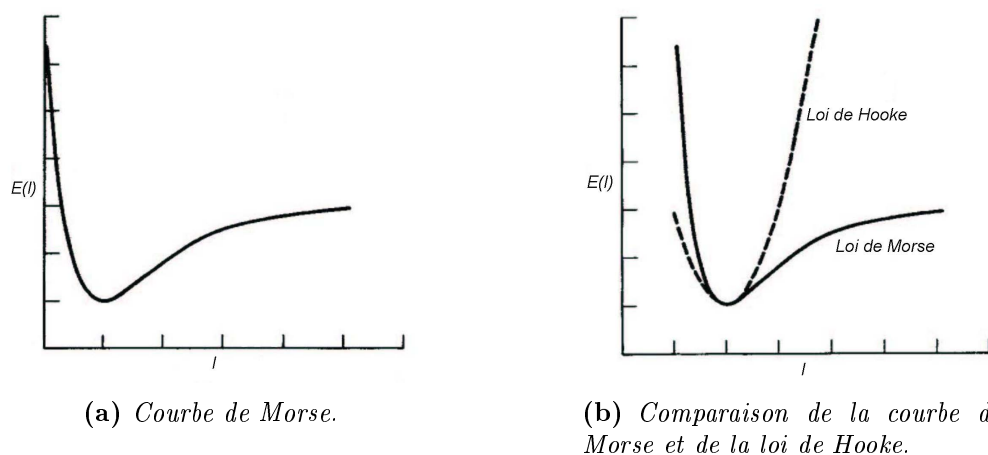


Figure 1.8 - Variations de l'énergie potentielle d'une liaison covalente selon deux modèles [Leach 01]. (a) Le potentiel de Morse modélise l'interaction, (b) la liaison est modélisée par un potentiel harmonique et comparée au potentiel de Morse.

Morse l'a décrite de la manière suivante :

$$E(l) = D_e (1 - e^{-a(l-l_0)})^2 \quad (1.7)$$

où D_e est la valeur minimum de l'énergie et $a = \omega_v \sqrt{m/(2D_e)}$, où m est la masse et ω_v est la fréquence de vibration de la liaison. ω_v est à relier à la constante de raideur K de la liaison par $\omega_v = \sqrt{K/m}$. l_0 est la valeur de référence de la liaison. Cette valeur est atteinte lorsque tous les autres termes du champ sont annulés (c'est la longueur naturelle de liaison et non la valeur

d'équilibre qui est atteinte pour une valeur d'énergie minimum). Cependant, le potentiel de Morse n'est pas utilisé en mécanique moléculaire car il est trop coûteux en temps de calcul. En effet, il requiert la détermination de trois paramètres pour chaque liaison. De plus, il est rare, lors de calculs de mécanique moléculaire, que la liaison dévie significativement de sa position d'équilibre. Par conséquent, une autre expression plus simple, tirée de la loi de Hooke, est utilisée. Celle-ci vaut alors à partir de la position de référence l_0 (généralement déterminée par cristallographie) :

$$E(l) = \frac{K_l}{2}(l - l_0)^2 \quad (1.8)$$

L'équation 1.8 est de forme harmonique et de petites oscillations autour de l_0 sont possibles. Cette approximation offre une certitude de l'ordre du centième d'Angström. Les forces entre atomes sont très fortes et nécessitent une énergie considérable pour que les liaisons se déplacent de manière significative de leur position d'équilibre. La figure 1.8(b) page précédente montre une comparaison du champ harmonique par rapport au potentiel de Morse. Dans les champs MM3 et MM4, la fonction harmonique est remplacée par une fonction cubique voire quadratique afin de mieux épouser la forme de la courbe de Morse.

2.1.1.3.2 L'angle de valence

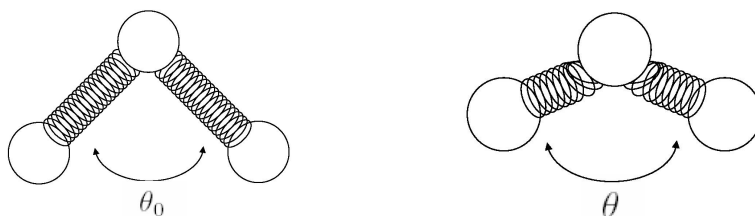


Figure 1.9 - Variation de l'angle de valence.

Comme pour la situation précédente, la variation de l'angle de valence est décrite par la loi de Hooke (ou un potentiel harmonique, figure 1.9) :

$$E(\theta) = \frac{K_\theta}{2}(\theta - \theta_0)^2 \quad (1.9)$$

A nouveau, pour accroître la précision du modèle, il suffit d'augmenter l'ordre de l'équation 1.9.

2.1.1.3.3 Termes de torsion

Les énergies mises en jeu lors des élongations des liaisons ou lors de la déformation des angles de valence doivent être importantes, c'est pourquoi la plupart des changements de conformation des molécules sont dus aux angles de torsion et aux interactions non liées.

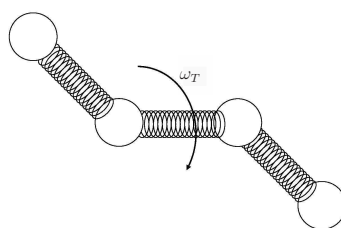


Figure 1.10 - Angle de torsion suivant les deux plans définis par les quatre atomes.

Contrairement aux termes précédents qui ne possèdent qu'une valeur d'équilibre, l'énergie de torsion est caractérisée, dans certains cas, par plusieurs minimums. Une forme fonctionnelle retrouvée dans le champs de force AMBER est :

$$E(\omega_T) = \sum_{n=1}^3 \frac{E_n}{2} (1 + \cos(n\omega_T - \gamma_p)) \quad (1.10)$$

où ω_T est l'angle de torsion, E_n est la hauteur de la barrière d'énergie (due aux gênes stériques), n est la multiplicité. Cette valeur donne le nombre de minimums atteints par la fonction durant l'unique rotation de la liaison sur 360° . γ_p représente la phase et détermine l'endroit où l'angle de torsion atteint son minimum.

Par exemple, le profil de l'énergie pour une rotation autour du carbone de l'éthane peut être représenté par un simple terme avec $n = 3$ et $\gamma_p = 0$. Ce qui donne trois minimums atteints pour $+60^\circ$, -60° et 180° (figure 1.11).

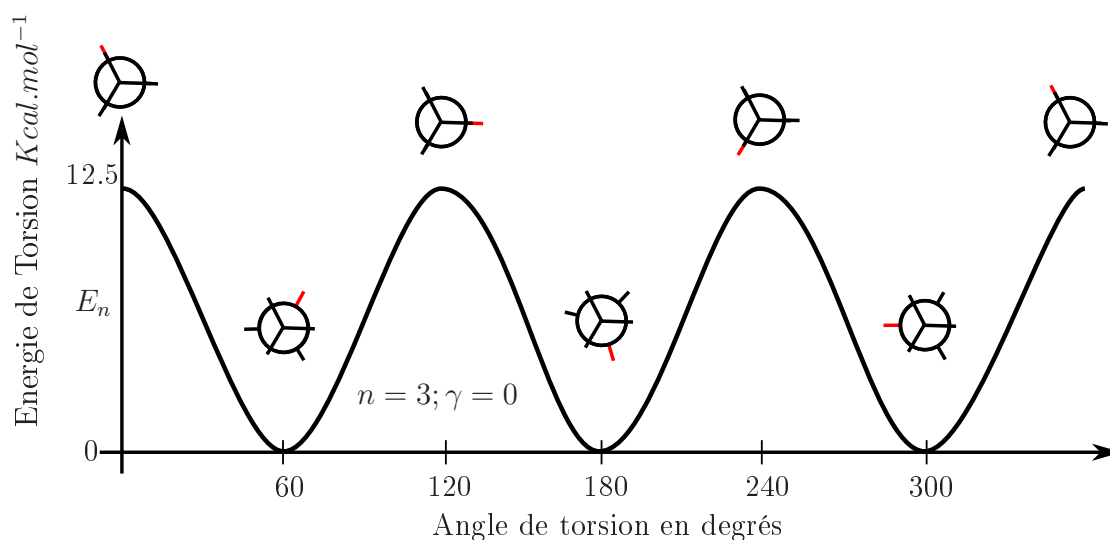


Figure 1.11 - Variation de l'énergie de torsion au cours d'une rotation.

Beaucoup des termes de torsion du champ de force AMBER ne contiennent qu'un cosinus dans l'expression de l'énergie. Mais il peut être nécessaire d'en faire apparaître un autre, notamment pour mettre en évidence les conformations *gauches* (il existe plusieurs minimums différents au cours de la rotation comme pour le butane : $\text{R}-\text{CH}_2-\text{CH}_2-\text{R}$). Le profil de l'énergie pour la rotation ne dépend seulement que des types atomiques de la liaison centrale et non

pas des types atomiques des autres atomes. Par exemple, tous les angles de torsion autour d'un atome de carbone sp^3 auront les mêmes paramètres. Une expression plus précise de l'énergie de rotation est proposée dans le champ de force MM2 [Allinger 77] :

$$E(\omega_T) = \frac{E_1}{2}(1 + \cos\omega_T) + \frac{E_2}{2}(1 - \cos2\omega_T) + \frac{E_3}{2}(1 + \cos3\omega_T) \quad (1.11)$$

Chacun des termes de cette série traduit une stabilisation particulière de l'angle dièdre :

- le premier terme agit sur la stabilisation des formes de conformation *cis* (E_1 est minimum) ou *trans* (E_1 est maximum) ;
- le deuxième terme agit sur la stabilisation planaire des doubles liaisons entre atomes de type sp^2 : si la liaison n'est pas coplanaire, E_2 est minimum, sinon il est maximum ;
- le troisième terme agit sur la stabilisation des formes de conformation décalée (E_3 est minimum) ou éclipse (E_3 est maximum).

2.1.1.3.4 Termes de torsion impropres

Si une molécule est modélisée en ne prenant en compte, en termes de rotation, que l'expression vue précédemment, il se peut que certaines conformations soient éloignées de la réalité. Certains atomes prendraient des valeurs d'équilibre qu'ils n'ont pas en réalité. Pour ramener dans le plan des liaisons des carbones sp^2 les atomes qui sont en dehors, il est nécessaire d'introduire un terme dit de « torsion impropre ».

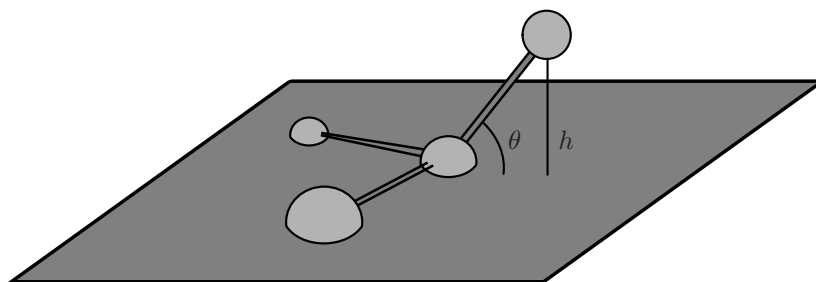


Figure 1.12 - Deux possibilités pour la modélisation de l'angle impropre.

La plupart du temps, l'énergie est introduite sous forme harmonique afin de ressembler aux équations précédentes et pouvoir ainsi facilement être intégrée. L'approche consiste à calculer la hauteur de l'atome hors du plan formé par les trois autres atomes liés au carbone (figure 1.12). L'équation s'écrit alors :

$$E(\theta) = \frac{k_\theta}{2}\theta^2; \quad E(h) = \frac{k_l}{2}h^2 \quad (1.12)$$

Il n'est pas toujours nécessaire d'introduire ce terme qui provoque des altérations dans le calcul du champ de force.

2.1.1.3.5 Les interactions longue distance

Les interactions entre atomes non liés n'impliquent que les atomes séparés par au moins trois liaisons ; elles caractérisent les forces de van der Waals et les forces électrostatiques. Les interactions 1-4 sont cependant traitées à part à cause de leur interférence avec les termes de torsion.

van der Waals. La courbe de la figure 1.13(a) représente la forme classique des forces de van der Waals [Lii 89], [Halgren 96b], [Halgren 96c], [Brooks 83]. L'énergie d'interaction vaut zéro à l'infini. A mesure que les deux molécules s'approchent, l'énergie décroît jusqu'à un minimum puis croît très vite.

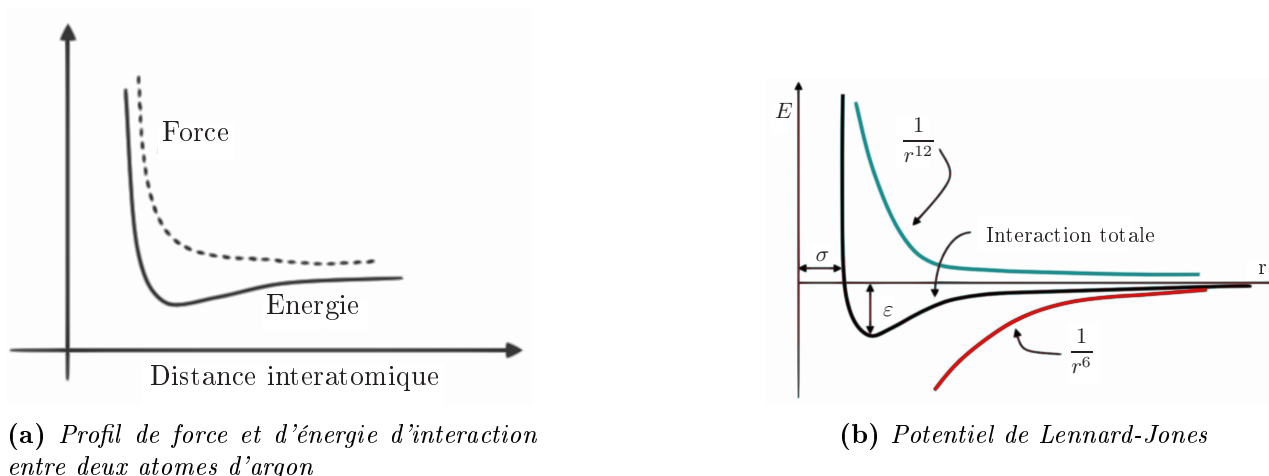


Figure 1.13 - Forces de van der Waals dans un champ de force de type AMBER et CHARMM.

Les forces d'attraction sont des forces à longue distance alors que les forces répulsives agissent à de petites distances. La contribution attractive des forces est due aux forces de dispersion expliquées pour la première fois par mécanique quantique par London.

Les interactions dispersives et attractives entre atomes peuvent être décrites par mécanique quantique mais cela requiert de solides bases théoriques et un temps de calcul conséquent. Pour un champ de force, il faut utiliser une fonction qui approche au mieux la courbe idéale par l'introduction de paramètres empiriques et qui peut être calculée rapidement. L'importance d'une fonction facilement calculable est compréhensible du fait d'un nombre très important d'interactions de type van der Waals dans une molécule. La fonction la plus connue représentant le potentiel de van der Waals est la *fonction 6-12 de Lennard-Jones* qui prend la forme pour une interaction entre deux atomes :

$$E(r) = 4\epsilon_p \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (1.13)$$

Cette fonction ne contient que deux paramètres ajustables : le diamètre de collision σ (la distance d'interaction pour laquelle l'énergie vaut zéro) et la profondeur du puits ϵ_p (figure 1.13(b)).

Le point d'énergie minimum est appelé point de séparation. A ce point ($r = r_m$), la dérivée première est nulle et entraîne donc que $r_m = 2^{1/6}\sigma$. Le potentiel s'écrit alors :

$$E(r) = \varepsilon \left[\left(\frac{r_m}{r} \right)^{12} - 2 \left(\frac{r_m}{r} \right)^6 \right] \quad (1.14)$$

Le potentiel de Lennard-Jones est caractérisé par une forme attractive en r^{-6} et une forme répulsive en r^{-12} (figure 1.13(b) page précédente). Le terme en r^{-6} est retrouvé en mécanique quantique (modèle de Drude le plus souvent utilisé). En revanche, la partie en r^{-12} ne trouve pas de réelle justification [Leach 01]. Ce terme est correct pour les gaz rares mais limité pour les systèmes complexes. Il est utilisé car facilement déductible en effectuant le carré du terme en r^{-6} . Il est possible de trouver, dans des champs de force, l'expression de Lennard-Jones avec des puissances neuf et dix.

Halgren propose une alternative à cette fonction en 1992 [Halgren 96c]. Elle se veut plus fiable et capable de reproduire des données expérimentales. Elle évite d'introduire des potentiels difficiles à calculer en réalité. Pour deux atomes i et j en interaction, ce potentiel est de la forme :

$$E(r) = \varepsilon_{p_{ij}} \left(\frac{1 + \alpha}{\rho_{ij} + \alpha} \right)^{(n-m)} \left(\frac{1 + \beta}{\rho_{ij}^m + \beta} - 2 \right) \quad (1.15)$$

Dans cette équation, $\rho_{ij} = r_{ij}/r_{ij}^*$, α et γ sont des constantes. L'étoile correspond à l'état d'énergie minimum. r_{ij}^* correspond donc à la valeur de la distance correspondant au minimum d'énergie entre les atomes i et j . La forme de Lennard-Jones est retrouvée si $n = 12$, $m = 6$, $\alpha = \beta = 0$.

Halgren propose une forme dite « buffered 14-7 » pour laquelle $n = 14$, $m = 7$, $\alpha = 0.007$ et $\beta = 0.12$:

$$E(r) = \varepsilon_{ij} \left(\frac{1.07r_{ij}^*}{r_{ij} + 0.07r_{ij}^*} \right)^7 \left(\frac{1.12r_{ij}^{*7}}{r_{ij}^7 + 0.12r_{ij}^{*7}} - 2 \right) \quad (1.16)$$

Un potentiel fini est obtenu par cette version lorsque la distance interatomique tend vers zéro (contrairement à Lennard-Jones qui devient infini). Une meilleure approximation de la série représentant la dispersion (modèle de Drude) est ainsi obtenue.

Interactions électrostatiques. Dans le terme électrostatique (ou coulombien) entrent en ligne de compte non seulement les interactions ions-ions où les atomes sont porteurs de charges, mais également toutes les autres interactions dites polaires. Dans ce dernier cas, les atomes possèdent des charges partielles ou ponctuelles. Ces charges sont généralement déterminées par des calculs *ab initio* ou semi-empiriques. L'expression utilisée est l'équation de Coulomb :

$$E(r) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon r_{ij}} \quad (1.17)$$

avec q_i et q_j les charges partielles portées par les deux atomes en interaction ; r_{ij} la distance de séparation des deux atomes ; ε_0 la permittivité diélectrique du vide et ε la permittivité effective.

La liaison hydrogène. Il est possible de tenir compte de l'existence de la liaison hydrogène de deux façons : soit en l'englobant dans les interactions électrostatiques et de van der Waals sans lui accorder de traitement particulier, soit en introduisant des fonctions spécifiques. Ce terme corrige surtout la non-prise en compte de la polarisation dans le modèle de charge ponctuelle utilisé pour l'interaction électrostatique.

Le calcul du potentiel intermoléculaire impose beaucoup plus de temps *CPU* que le calcul du potentiel intramoléculaire. La complexité pour ce type de calcul est évaluée à N^2 . Il est donc impératif d'introduire des procédures limitant ce temps de calcul. La méthode la plus généralement utilisée est l'introduction d'une coupure du potentiel non liant entre 8 et 10 Å pour les interactions de van der Waals. Il existe plusieurs méthodes de troncatures, telle la méthode *Shift* qui ramène brutalement le potentiel à zéro à partir de la distance définie. Elle a pour conséquence de modifier le profil de force. La méthode *Switch* introduit deux distances de coupure. Avant la première distance, le profil n'est pas affecté, entre les deux, il est ramené à zéro selon une certaine fonction puis, après la deuxième distance, est laissé nul. Cette méthode évite des instabilités introduites par la méthode *Shift* lors de la dynamique moléculaire. Les interactions non liées ne sont plus calculées entre toutes les paires d'atomes mais simplement avec leurs voisins (compris dans la sphère de 8 à 10 Å), ce qui nécessite l'introduction d'une table de voisins qui sera mise à jour régulièrement lors des étapes de dynamique et de minimisation.

Si la coupure est économique en temps de calcul, sa justification physique n'est pas évidente : pour les interactions de van der Waals, qui sont effectivement de courte portée (décroissance en $1/r^6$), la coupure à 8 ou 10 Å est valable. Pour l'électrostatique (décroissance en $1/r$), interaction de longue portée, la coupure est discutable.

2.1.1.3.6 Paramétrisation

Ayant choisi la forme mathématique des fonctions énergétiques à employer dans le champ de force, il convient ensuite de rechercher l'ensemble des paramètres qui offriront la meilleure représentation des données expérimentales retenues.

Les données géométriques sont importantes à deux égards : d'abord, tout champ de force doit respecter la géométrie moléculaire et, ensuite, certaines fonctions font appel à des valeurs géométriques de référence. La source d'information structurale la plus abondante est la cristallographie aux rayons X.

Les paramètres de référence (l_0, θ_0) introduits dans les équations de comportement ne doivent pas être confondus avec les paramètres d'équilibre réellement observés. Les paramètres de référence utilisés dans un champ de force n'ont pas de signification précise, ce sont de simples valeurs qui permettent de retrouver par le calcul la géométrie des structures de l'ensemble expérimental. Dans le cas de l'éthane, par exemple, la longueur de la liaison C-C calculée vaut 1.532 Å (valeur expérimentale : 1.534 Å), tandis que la valeur de référence pour une liaison C-C est de 1.523 Å. Cette « élongation » est due aux interactions 1-4 entre atomes d'hydrogène : afin de mieux répartir les différentes contributions énergétiques, les longueurs de liaison ainsi que les angles de valence sont modifiés par rapport aux valeurs de référence.

Les paramètres non géométriques (hauteur des barrières d'énergie, constantes apparaissant

dans l'expression de l'énergie de van der Waals ...) sont nettement plus difficiles à obtenir : c'est à ce niveau que l'aspect « empirique » de la méthode se manifeste le plus. Ainsi, alors que les paramètres de $E(r)$ sont approximativement égaux aux constantes de force spectroscopique, il est nécessaire pour paramétrer $E(\theta)$, de réduire de 20 à 50% les constantes de force de déformation des angles de valence.

Une dernière considération sur les champs de force concerne la non-transférabilité. En effet, un champ de force est paramétré pour un type de molécule précis et adapte ses paramètres afin de retrouver des valeurs expérimentales. Un champ étudié pour les micromolécules n'est, par conséquent, pas utilisable pour étudier le comportement d'un liquide. Le choix du modèle est donc très important pour la validité du résultat.

2.1.2 Simulation du comportement protéique

Une fois l'énergie totale de la molécule évaluée, plusieurs méthodes aux aboutissements différents peuvent simuler le comportement de la protéine. Elles sont toutes basées sur une évaluation préalable de l'énergie et diffèrent par leur applicabilité.

2.1.2.1 Méthodes de minimisation du potentiel

Une des premières utilisations des champs de force dans le domaine de la biologie est de minimiser l'énergie interne des modèles tridimensionnels des macromolécules. La minimisation des énergies moléculaires vient du fait que la forme biologique naturelle des molécules est dans un minimum de potentiel (état stable). La structure correspondant à un minimum de l'énergie interne peut être considérée comme la structure moyenne autour de laquelle évolue le système dans son état d'équilibre.

Le problème posé par la minimisation énergétique des molécules est lié à la recherche de minimums pour des fonctions d'un grand nombre de variables ($3N$, où N est le nombre d'atomes). Il faut donc choisir entre les différentes méthodes qui diffèrent par leur efficacité, leur rayon de convergence, leur convergence vers le minimum et le nombre d'opérations induites.

2.1.2.1.1 Principe

La fonction d'énergie interne à minimiser est une fonction à environ $3N$ variables. Une telle fonction présente, dans le cas le plus général, un minimum global et un grand nombre de minimums locaux. D'une manière très schématique, la figure 1.14 page suivante montre, selon un axe donné (direction ou suite de conformation), le minimum global et des minimums locaux.

La recherche du minimum global est donc difficile, très coûteuse en temps de calcul et n'est absolument pas certaine pour les algorithmes qui ne recherchent qu'un minimum à partir de la position initiale.

D'un point de vue énergétique, la molécule se positionne au fond du puits de potentiel énergétiquement le plus bas, défini comme étant le minimum global. Si la température augmente, la molécule vibre autour de son minimum. Si l'énergie de vibration est assez élevée

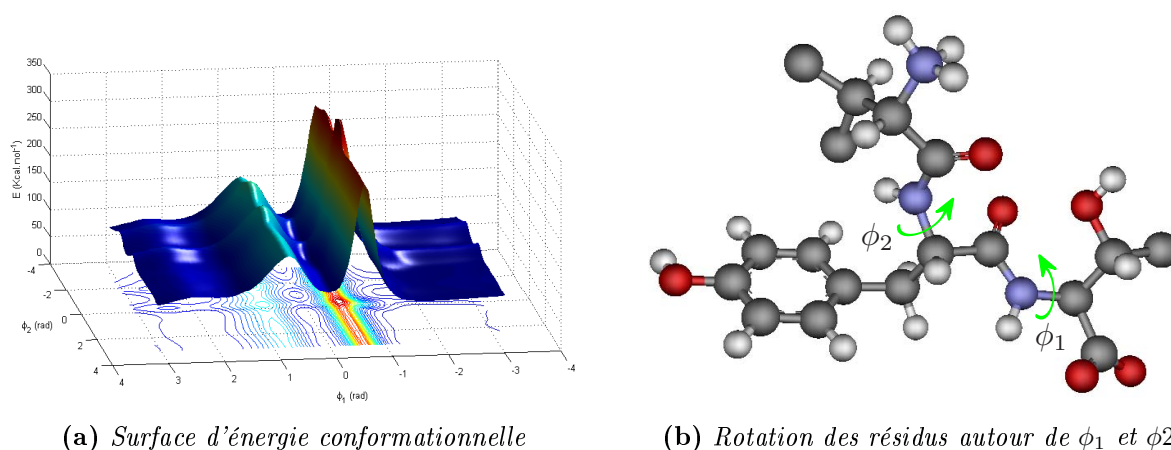


Figure 1.14 - Surface d'énergie conformationnelle obtenue pour la rotation de la valine autour de la thyrosine (ϕ_2) en fonction de la rotation de la thréonine autour de la thyrosine (ϕ_1).

pour atteindre une certaine énergie d'activation, un saut entre deux conformations va pouvoir s'effectuer. D'autres minimums qui ne correspondent pas au minimum global vont ainsi pouvoir être atteints.

2.1.2.1.2 Optimiseurs classiques

Les optimiseurs classiques, à partir des coordonnées initiales des atomes, calculent le minimum énergétique le plus proche de la position de départ. L'accessibilité du minimum global, sauf si la géométrie de départ se trouve proche de ce puits, n'est pas possible avec de telles méthodes. Ces optimiseurs étudient la nullité de la dérivée du potentiel, et la pente de sa dérivée seconde, qui doit être positive.

Trois méthodes aux dérivées partielles sont couramment utilisées pour atteindre un minimum local : la méthode *Steepest Descent*, celle du gradient conjugué et celle aux dérivées secondes, de Newton-Raphson ou Newton tronquée. Elles sont généralement employées ensemble et, selon l'ordre précédemment décrit, de la plus simple à la plus compliquée. La première converge rapidement car chaque pas nécessite peu de temps de calcul, la deuxième calcule les dérivées secondes en se basant sur la première méthode et la troisième méthode affine le calcul.

2.1.2.1.3 Monte-Carlo

La méthode stochastique de Monte-Carlo est basée sur l'utilisation de nombres aléatoires et de probabilités statistiques d'échantillonnage de l'espace conformationnel. L'idée est de fournir, à une liaison donnée, une nouvelle valeur de l'angle dièdre associé. La nouvelle conformation ainsi engendrée est acceptée ou refusée selon le calcul de la probabilité p_{MC} , en utilisant la procédure de Métropolis [Leach 01] :

$$p_{MC} = \min \left[1, e^{\left(-\frac{\Delta E}{k_B T} \right)} \right] \quad (1.18)$$

avec p_{MC} la probabilité d'un processus, k_B la constante de Boltzmann, T la température absolue et ΔE la variation d'énergie entre la nouvelle configuration et l'ancienne.

Si la nouvelle conformation présente une énergie plus basse que l'ancienne ($\Delta E < 0$), celle-ci est toujours acceptée. Si, au contraire, la nouvelle conformation présente une énergie supérieure ($\Delta E > 0$), il existe une certaine probabilité, proportionnelle au facteur de Boltzmann, qu'elle soit acceptée (équation 1.18). Si cette probabilité est supérieure à un nombre aléatoire compris entre 0 et 1, alors la nouvelle configuration est acceptée, sinon elle est refusée.

La méthode de Monte-Carlo explore la surface d'énergie conformationnelle en limitant toutefois l'acceptation des nouvelles conformations à celles d'énergie inférieure aux précédentes. Puisque c'est une méthode d'exploration conformationnelle, elle ne minimise pas une conformation vers le minimum le plus proche mais cherche sur la surface d'énergie engendrée par la nouvelle conformation l'état d'énergie minimum optimum. Des barrières d'énergie peuvent ainsi être franchies. L'état initial de la protéine à simuler étant d'énergie minimum lors des simulations interactives, la méthode de Monte-Carlo n'est donc pas à privilégier, une simple minimisation garantit ainsi le chemin de plus basse énergie.

2.1.2.2 La dynamique moléculaire

La dynamique moléculaire consiste à étudier la trajectoire d'une molécule en appliquant les lois de la mécanique classique newtonienne c'est-à-dire à simuler les mouvements atomiques au cours du temps (exploration conformationnelle). Ces mouvements correspondent à des vibrations autour d'un minimum ou au passage d'un minimum à un autre. Les barrières d'énergie les séparant pouvant être franchies par chauffage du système.

En pratique, les équations classiques du mouvement de Newton sont résolues étant donné une fonction énergie potentielle et son champ de force associé. La dynamique dépend donc du temps, contrairement à la minimisation. Afin de simuler ces mouvements, la loi de Newton est utilisée :

$$m_i \frac{d^2 \vec{X}_i}{dt^2} = \vec{F}_i = -\vec{\nabla}_i(E(\vec{r})) \quad (1.19)$$

\vec{X}_i étant la position de l'atome i , m_i sa masse, F_i la force agissant sur l'atome i et E l'énergie associée au champ de force précédemment calculé.

La température intervient lors de l'écriture de l'énergie cinétique :

$$E_{cinétique} = \frac{3}{2} N k_B T \quad (1.20)$$

avec N le nombre de particules, k_B la constante de Boltzmann et T la température.

Le système est porté à une température élevée, de l'ordre de 1000 K. A une température si haute, l'énergie cinétique fournie à la molécule garantit le franchissement de barrières d'énergie par le mouvement qu'elle induit. Le facteur déterminant dans la procédure ainsi appelée *recuit simulé* est d'effectuer un refroidissement lent afin que les atomes puissent se positionner de la meilleure façon.

La vitesse et le sens du mouvement atomique dépendent des forces qui s'exercent entre les atomes. En pratique, les atomes possèdent une vitesse initiale propre (choisie aléatoirement selon la distribution de vitesse de Boltzmann) qui se conforme à l'énergie cinétique totale du système.

Par suite d'un défaut de solution analytique, les trajectoires ne sont pas directement déterminées par résolution de l'équation 1.19. Une approche différente est utilisée et consiste, par exemple, à calculer :

1. les accélérations des atomes à partir des masses atomiques et des forces s'exerçant sur chacun d'eux.
2. les vitesses de chaque atome à partir de la connaissance des accélérations atomiques.
3. les positions des atomes à partir des vitesses atomiques.

Les méthodes de dynamique moléculaire explorent la surface d'énergie conformationnelle et dépendent du temps au travers du pas d'intégration et du temps de simulation du comportement désiré (de l'ordre de la picoseconde). La simulation interactive de *docking* moléculaire recherchée n'est pas le parcours de la surface d'énergie conformationnelle, mais la recherche de la meilleure conformation pour le ligand et la protéine, quel que soit le temps, pour une position donnée.

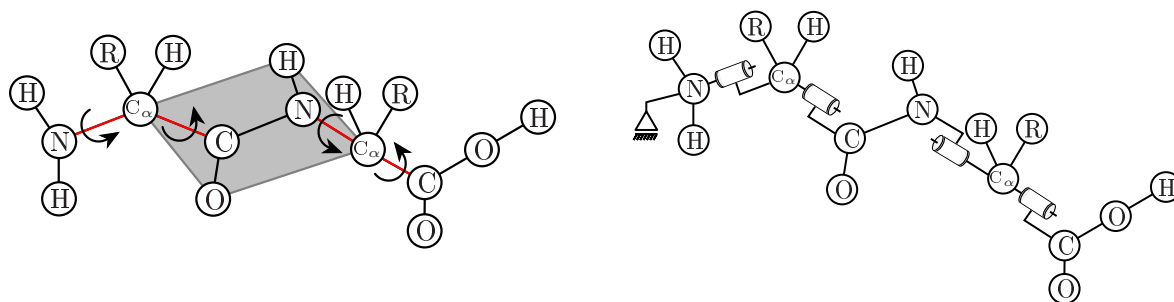
La minimisation du champ de force répond à cette exigence, pas la dynamique moléculaire. C'est donc la méthode de minimisation de l'énergie engendrée par un champ de force qui sera privilégiée.

Cependant, minimiser l'énergie calculée à partir d'un champ de force pour obtenir la conformation d'énergie minimum à partir d'une position donnée, à moins de disposer d'une puissance de calcul certaine, n'est pas temps réel. Au cours de la simulation, les positions des atomes sont actualisées sur le chemin de plus basse énergie, mais la solution optimale n'est pas calculée en temps réel. C'est pourquoi d'autres méthodes ont été privilégiées, assimilant un système polyatomique à un système polyarticulé et identifiant le problème de minimisation à un problème d'optimisation sous contrainte géométrique. Ce sont des méthodes issues de la robotique. Elles consistent à modéliser les chaînes polypeptidiques en chaînes cinématiques puis à résoudre un problème de cinématique inverse. Elles consistent en outre à assimiler un problème de minimisation à un problème de planification de trajectoire dans un espace de contraintes géométriques accessibles.

2.2 Méthodes issues de la robotique

2.2.1 Modélisation par chaîne cinématique

Du point de vue conceptuel, une protéine peut être représentée dans son état dénaturé par une chaîne de $N + 1$ solides connectés par N liaisons pivots car l'énergie mise en jeu pour une liaison covalente est bien supérieure aux autres énergies. Les liaisons covalentes sont alors assimilées à des liaisons rigides et seules les angles de torsion sont modélisés par des liaisons pivots. La chaîne secondaire, plus petite, est représentée de la même manière par des liaisons pivots (de zéro à quatre) connectés à la chaîne principale (figure 1.15 page suivante).



(a) Représentation biologique d'une protéine. Toutes les liaisons sont élastiques (modélisées sous forme de ressorts.)

(b) Représentation robotique de la même protéine. Toutes les liaisons sont rigides ou des pivots.

Figure 1.15 - Une même molécule est représentée différemment selon le point de vue biologique ou robotique. Les groupes *R* représentent des chaînes latérales.

D'après [Kazerounian 04a], le temps de calcul de la géométrie de la chaîne polyarticulée par cette méthode serait divisé par 225. Une des premières études est menée par Manocha [Manocha 95] qui utilise la solution de Ragavan et Roth d'un manipulateur $6R$ pour le problème de cinématique inverse, permettant de réaliser un *docking* de petite molécule. Cette modélisation tend à réduire la dimension de l'espace des configurations, d'où un gain en temps de calcul.

Un des premiers avantages de la modélisation en chaîne cinématique vient de sa description. En effet, les méthodes classiques utilisent le modèle tridimensionnel des molécules. La modélisation des chaînes cinématiques introduit la représentation de Denavit-Hartenberg, la méthode des tenseurs, la méthode duale, les quaternions, etc. Ces notations et ces concepts font que le traitement est plus stable. De petits changements dans la géométrie n'induisent pas de grandes variations dans les paramètres cinématiques [Kazerounian 04a], le temps de calcul est limité et le passage entre chaînes fermées et chaînes ouvertes est simplifié.

Le modèle géométrique donne la position de tous les atomes en termes d'angles de torsion. Les valeurs des liaisons covalentes sont fixées pour les raisons précédemment énoncées. Plusieurs représentations sont alors possibles : Lavalley décrit le changement de conformation à l'aide des paramètres de Denavit-Hartenberg, les valeurs des liaisons, des angles de valence et les angles de torsion étant les paramètres [Lavalley 00]. Il introduit une fonction d'erreur cinématique égale à la somme des erreurs de position entre chaque atome par rapport à une position donnée (par homologie). Il procède ensuite par changement de conformation aléatoire suivant les angles de torsion et garde la meilleure conformation si l'erreur cinématique est minimum et si l'énergie (mécanique) est minimisée. Afin de pouvoir traiter le problème par le système de coordonnées de Denavit-Hartenberg, un atome est choisi comme « ancre » et tout le reste de la molécule est considéré comme un arbre issu du sommet. Les chaînes fermées sont considérées comme étant rigides et sont remplacées par des atomes « spéciaux ». L'avantage de cette méthode est qu'il n'est pas nécessaire de se souvenir de la position précédente des atomes, limitant ainsi les retenues mémoires [Zhang 02]. De plus, les multiples rotations autour d'une liaison peuvent être combinées en une seule. En d'autres termes, lorsque la matrice de rotation d'un atome est construite, la position finale est représentée par la dernière colonne de celle-ci. En outre, l'erreur numérique sur les anciennes positions des atomes ne se propage pas aux nouvelles lorsqu'une rotation est effectuée, la nouvelle position ne dépend donc pas de l'ancienne. L'inconvénient majeur repose sur la multiplicité des repères. Lorsqu'un atome possède deux liaisons, il est

nécessaire d'introduire deux repères, ce qui augmente la multiplicité des matrices de rotation pour les fils. Un autre inconvénient intervient lorsqu'un lien est fixe. Il possède tout de même un repère et alourdit inutilement les calculs. Une autre méthode, appelée *atomgroup local frames*, élimine cette redondance : un repère est attaché à un groupe, celui-ci étant un ensemble d'atome privés de rotation entre eux. La simulation est alors bien plus rapide [Zhang 02].

[Kazerounian 04b] implémente la méthode de la « position zéro » [Gupta 86] [Alvarado 03] afin de décrire le mouvement de la chaîne (figure 1.15 page précédente). Il n'utilise alors que des matrices de rotation. Cette méthode présente l'avantage, par rapport à Denavit-Hartenberg, de ne pas requérir de repères intermédiaires car elle travaille dans le repère global. Tous les vecteurs de description de la méthode (vecteurs d'axes de rotation, d'axes de liaison) sont décrits à partir de la position de référence (état dénaturé) dans le repère global. La chaîne secondaire est traitée de la même manière et raccordée au modèle du squelette. Le champ de force utilisé pour ce modèle ne considère que les interactions longue distance car les longueurs de liaison sont fixes et les contraintes sont intégrées au niveau de la géométrie [Kazerounian 04c]. Ce choix offre une meilleure efficacité au niveau du temps de calcul pour relativement peu d'erreurs. La suite du modèle consiste à intégrer toutes les contraintes dans un Jacobien. Le mouvement est obtenu grâce au modèle géométrique inverse en considérant que les atomes sont soumis à des couples articulaires. Les variations d'angles deviennent un rapport d'un couple articulaire sur un couple articulaire maximal multiplié par un facteur de stabilité (modèle compliant).

[Morin 07] utilise un algorithme adaptatif pour la simulation dynamique de chaînes polyarticulées. Toutes les liaisons autorisant une variation d'angle de torsion sont assimilées à des liaisons de type pivot. L'algorithme, adapté de l'algorithme *Divide-and-Conquer* de Featherstone [Featherstone 99a], [Featherstone 99b], [Redon 05], implémente un modèle dynamique pour la simulation de chaînes polyarticulées dont la particularité est sa rapidité d'exécution comparée aux autres méthodes dynamiques (sa complexité est en $O(N_A) + f \times [O(N_A) + O(N_R^3)]$, où N_A est le nombre de liaisons actives, f est la fréquence d'actualisation de la région active et N_R le nombre de nœuds qui doivent être rigidifiés). Cet algorithme simplifie automatiquement le modèle dynamique, l'utilisateur spécifiant le nombre de degrés de liberté qui doivent être actifs. Toute la dynamique de la chaîne n'est alors pas réactualisée.

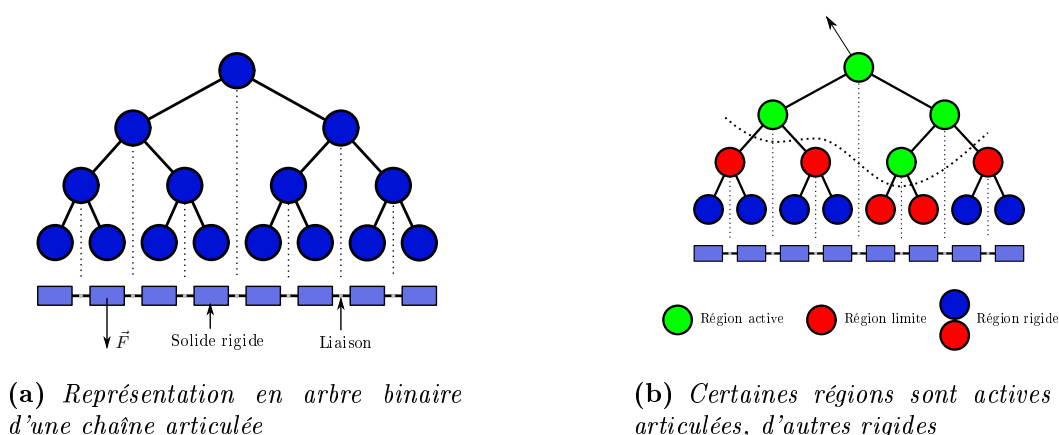


Figure 1.16 - Représentation en arbre binaire d'une chaîne articulée et son exploitation par l'algorithme adaptatif [Redon 05].

Featherstone définit, de manière récursive, une chaîne articulée comme une paire d'objets articulés reliés par une liaison. La chaîne peut alors être représentée par un arbre binaire pour

lequel chaque nœud représente une liaison et chaque feuille, un corps rigide (figure 1.16 page précédente). Chaque corps articulé possède un jeu de « poignées » qui sont le lieu d'application des forces.

Chaque liaison de la chaîne est active (simulée), ou rigide (non simulée). Chaque nœud de l'arbre est rigide si tous les fils du nœud sont rigides ou sont une feuille, ou hybride (active) si la liaison principale du nœud est active, mais certains fils sont rigides. La région rigide correspond à l'ensemble des nœuds ne nécessitant pas d'actualisation, et la région passive, à l'ensemble des nœuds dont le parent est actif. L'algorithme adaptatif combine alors deux algorithmes à chaque pas de temps :

1. un algorithme d'actualisation de région active : il rend active une liaison selon certains critères.
2. un algorithme de propagation de la dynamique : il propage la dynamique lorsque une liaison a été sélectionnée.

Cette méthode, prometteuse pour la faible complexité qu'elle engendre, utilise les contraintes cinématiques de la chaîne polyarticulée comme forces internes. La dynamique de la chaîne n'est calculée que sur certaines zones pouvant être définies par l'utilisateur, réduisant d'autant la complexité et le temps de calcul de la nouvelle conformation. Cette application n'a cependant pas été implémentée pour le *docking* de molécule.

2.2.2 Planification de la trajectoire moléculaire

D'une façon générale, il est possible de modéliser une protéine comme un système mécanique dont les corps rigides sont les éléments de structures secondaires (hélices et feuilletts) et dont les liaisons sont constituées de sous-chaînes articulées d'acides aminés (boucles), les acides aminés pouvant, eux aussi, avoir plusieurs conformations liées à la mobilité de leur chaîne latérale. La protéine apparaît ainsi comme une chaîne cinématique complexe du même type que celles étudiées en robotique. Dès lors, les problèmes d'étude de l'accessibilité d'un ligand vers le site actif d'une protéine se ramènent aux études classiques de planification de mouvements sous contraintes à la fois endogènes et exogènes. Avec cette modélisation, toutes les études sont conduites dans l'espace des configurations de la structure ligand-protéine. Sont ainsi représentées de manière unifiées sous forme d'« obstacles à éviter » dans l'espace des configurations aussi bien des contraintes angulaires (déformation d'angles dièdres) sur le repliement d'une protéine autour d'un axe (contraintes endogènes) que des contraintes issues des interactions moléculaires (contraintes exogènes).

La formulation générale de la planification de mouvement consiste à transformer le problème du déplacement d'un système mécanique dans un environnement tridimensionnel en celui du déplacement d'un point dans un espace caractéristique des positions du système, appelé espace des configurations. Avec cette formulation, sont traitées sous forme d'obstacles à éviter dans cet espace les contraintes endogènes (entre les corps du système mobile) ou exogènes (avec les autres corps présents dans l'environnement) qui caractérisent les mouvements possibles du système. Pour cela, il convient de transformer les contraintes spatiales imposées par les obstacles de l'environnement en un ensemble de contraintes qui définissent l'espace des configurations admissibles. Planifier un mouvement sans collision revient alors à construire et à explorer les composantes connexes de cet espace (les ensembles de positions connectables par un chemin sans collision).

L'algorithmique du mouvement a pour but le calcul de trajectoire sans collision par une recherche de chemins dans un graphe. Alors que la planification de mouvement s'est longtemps heurtée à la complexité combinatoire inhérente au problème (en particulier lorsqu'il s'agit d'explorer des espaces de configuration hautement dimensionnés), de nouvelles approches de type probabiliste développées récemment dans plusieurs équipes de recherche [Cortés 03], [Cortés 07], [Singh 99], [Lotay 04], conduisent aujourd'hui à traiter efficacement une large classe de problèmes : systèmes articulés décrits par des chaînes cinématiques complexes évoluant dans des environnements fortement contraints (nombreuses contraintes exogènes) tout en rendant compte des contraintes endogènes propres aux mouvements internes du système mécanique.

2.2.3 Conclusion

La modélisation moléculaire et la robotique partagent l'approche réductrice du filtrage des mouvements selon des degrés de liberté supposés efficaces pour la fonction étudiée. Mais la différence essentielle entre une protéine et un robot réside dans la complexité combinatoire des conformations. Alors que le nombre de degrés de liberté d'un robot dépasse rarement une dizaine, une protéine peut en posséder plusieurs milliers. C'est dans cette différence quantitative que réside l'enjeu technologique : pour que l'application des techniques robotiques à la modélisation moléculaire soit efficace, il est indispensable qu'un filtrage des degrés de liberté ait d'abord lieu et que les algorithmes résistent à la complexité des structures mécaniques considérées.

L'inconvénient majeur de ces méthodes robotiques, bien que performantes, est qu'elles nécessitent un travail en amont très important. En effet, le filtrage des degrés de liberté pour une molécule est aisé lorsqu'un ligand est considéré mais devient plus ardu s'il faut modéliser le mouvement entier d'une protéine. De plus, les approches de type planification de trajectoire ne sont pas temps réel bien qu'une fois l'espace des configurations et des trajectoires calculé, il soit possible de rejouer la scène en temps réel. Mais le chemin précalculé ne correspond pas forcément à l'attente d'un biologiste qui voudrait s'en écarter pour rechercher des solutions nouvelles. Bien que le chemin calculé soit effectivement optimal du point de vue de l'interaction, s'en écarter peut être intéressant, notamment pour comprendre le champ d'énergie extérieur à la molécule. Cette approche robotique n'est donc pas retenue pour la modélisation et la simulation des interactions ligand-protéine. Une approche empirique avec une minimisation du champ de force est alors privilégiée.

2.3 Bilan

Ce paragraphe s'est attaché à résumer succinctement l'état de l'art des simulations moléculaires. Par l'expression « simulation moléculaire », il faut comprendre le fait d'évaluer l'énergie interatomique d'un système polypeptidique puis le fait d'optimiser la conformation de la protéine afin d'obtenir son minimum assurant une conformation stable proche de la réalité.

Les méthodes qui évaluent l'énergie d'interaction, telles que la mécanique quantique, les méthodes semi-empirique ou empiriques, sont dépendantes du système considéré. Elles dépendent notamment de leurs tailles et de leurs propriétés physicochimiques. La différenciation de ces méthodes s'effectue principalement au niveau de la précision de la modélisation des interactions

mais aussi au niveau du temps de calcul nécessaire pour la modélisation des interactions.

Ainsi, la mécanique quantique est privilégiée pour une meilleure modélisation des interactions. Cependant, si le temps de calcul est un élément important dans le choix d'un modèle, alors un champ de force adapté au système à simuler sera une réponse satisfaisante.

Les méthodes qui simulent le changement de conformation d'un polypeptide sont, dans ce manuscrit, classées en deux parties. L'une considère les méthodes usuellement utilisées en biologie, l'autre, une nouvelle application des algorithmes d'optimisation des configurations des systèmes robotiques. La dynamique moléculaire explore la surface d'énergie conformationnelle d'une protéine sur un temps précis de simulation. La méthode de Monte-Carlo explore aléatoirement la surface d'énergie conformationnelle en vue d'obtenir le minimum optimal selon un critère précis. Enfin, la minimisation de l'énergie engendrée par un champ de force, par des optimiseurs classiques, fait tendre le système vers son minimum d'énergie le plus proche. Les méthodes issues de la robotique optimisent la géométrie du système pour un ensemble de contraintes externes qui lui sont appliquées. Elles jouent donc le rôle d'un minimiseur classique.

A nouveau, plusieurs éléments sont à prendre en compte pour le choix d'une de ces méthodes. L'objectif de ce manuscrit est de réaliser un *docking* moléculaire en interactivité avec un opérateur extérieur. L'idée n'est donc pas d'explorer les conformations des molécules présentes mais de chercher, à tout instant, la conformation de plus basse énergie pour une position et une orientation de ligand donnée. La minimisation par des optimiseurs classiques de l'énergie totale du système est donc la méthode à privilégier et non celle de Monte-Carlo. Les méthodes *robotiques* ne sont pas retenues du fait de leur faible lien avec la biologie.

Il ressort donc de cette étude qu'un champ de force, adapté au type de protéine à simuler, est à privilégier sur les autres méthodes pour des raisons de temps de calcul. Enfin, une minimisation de cette énergie, par une optimisation classique, répond au critère du *docking* moléculaire interactif.

3 Haptique moléculaire

Avant de proposer une orientation pour mes travaux de recherche en haptique moléculaire, il convient d'étudier les travaux existants. Ceux-ci sont regroupés dans la suite de ce paragraphe en deux catégories. La première rassemble les approches qui considèrent un ligand rigide aux abords d'une protéine rigide. Il faut entendre par *rigide* la simplification conformationnelle qui consiste à négliger la variation de position des atomes d'un système. Ainsi, l'énergie engendrée par l'interaction des deux entités, le ligand et la protéine, n'est pas minimisée au cours de la manipulation. Il n'existe donc pas de changement de conformation au cours de la manipulation. La seconde catégorie considère la protéine comme rigide, et le ligand comme flexible. Contrairement à la rigidité, la flexibilité d'une protéine fait intervenir les contraintes géométriques étudiées précédemment. Ainsi, l'énergie totale calculée est cette fois optimisée afin de changer la conformation de la protéine considérée. Les travaux autour du ligand rigide ne considèrent que les interactions moléculaires en termes de forces à ressentir. En revanche, pour l'approche flexible, les équipes de recherche ont cette fois intégré les moments calculés lors de l'interaction.

3.1 Approche rigide

L'haptique moléculaire est née dans les années 2000 de la nécessaire interactivité des simulations de dynamique moléculaire [Stone 01]. En effet, à l'initialisation des simulations de dynamique moléculaire, une force doit être appliquée aux atomes. Si cette force est trop faible, le comportement de la molécule peut ne pas être significatif. [Stone 01] a eu l'idée d'utiliser une interface haptique pour appliquer cette force en cours de simulation afin d'imposer à un ion le passage d'une protéine membranaire, simulation difficilement réalisable en raison de fortes barrières d'énergies présentes à son entrée. Le schéma de couplage entre un phantom 3D, le simulateur (NAMD) et l'interface graphique (VMD) est simple. Trois facteurs sont considérés. Le premier, α , met en correspondance le temps réel et le temps total de la simulation de dynamique moléculaire et vaut typiquement 10^{12} . Le deuxième, β , adapte les forces entre le micromonde et le macromonde. Le dernier, γ , est un facteur de déplacement. La force d'une particule appliquée sur l'interface $\beta \vec{F}_{haptique}$ vaut alors :

$$m \frac{d^2 \vec{X}_s}{dt_s^2} = \beta \vec{F}_{haptique} = m \frac{\alpha^2}{\gamma} \frac{d^2 \vec{X}_u}{dt_u^2} \quad (1.21)$$

avec \vec{X}_s et t_s les coordonnées et temps exprimés dans l'environnement de la simulation, et \vec{X}_u et t_u ceux exprimés dans l'environnement de l'utilisateur, m la masse de la particule à manipuler. La force à ressentir vaut donc :

$$\vec{F}_{haptique} = m \frac{\alpha^2}{\beta \gamma} \frac{d^2 \vec{X}}{dt_u^2} \quad (1.22)$$

En considérant l'équation 1.22, la force appliquée à l'interface dépend étroitement de la rapidité de la simulation. En réalité virtuelle, une force appliquée *via* une interface haptique vaut $\vec{F}_{haptique} = -K \vec{X}_u$, où K est une constante à déterminer et représente la raideur du couplage. La force extérieure ainsi appliquée vaut $\beta K \vec{X}_u = K \beta \gamma \vec{X}_s$. Si la particule est assujettie à un champ d'énergie $E(\vec{X}_s)$, la dynamique de la particule est alors représentée par :

$$m \frac{d^2 \vec{X}_s}{dt_s^2} = -K \beta \gamma \vec{X}_s - \frac{dE(\vec{X}_s)}{d\vec{X}_s} \quad (1.23)$$

La force à ressentir, dépendante du temps de simulation qui doit par conséquent être rapide, implique que la dynamique moléculaire n'est pas une méthode directement applicable à un système. L'influence des facteurs d'échelles peut être prédominante sur l'énergie d'interaction si ces derniers sont trop élevés, altérant ainsi le ressenti des interactions (équation 1.23).

[Nagata 02] propose de ressentir uniquement les forces électrostatiques entre une protéine et un ligand, les deux systèmes étant rigides. La position du ligand considéré comme une sphère est la donnée du système. La force électrostatique actualisée est retournée à l'interface haptique selon un facteur d'échelle fixé arbitrairement. Si les forces ainsi calculées dépassent la valeur maximale des forces admissibles par les moteurs de l'interface, celles-ci sont arbitrairement bornées. Devant la complexité des calculs, seulement vingt atomes sont considérés afin de calculer en temps réel les forces d'interaction.

[Wollacott 07], afin d'éviter le calcul temps réel des forces, propose de précalculer selon une grille tridimensionnelle précédemment établie, les forces électrostatiques et de van der Waals

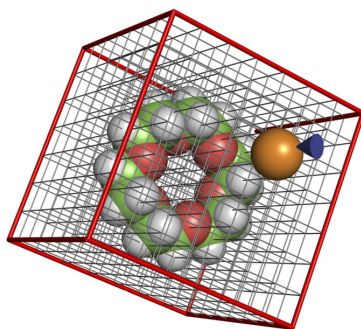


Figure 1.17 - Schéma de principe de calcul. Une grille volumique est construite autour de la molécule 18-crown-6 et densifiée aux alentours du site actif. Le pointeur bleu est utilisé pour sélectionner l'ion potassium (orange) afin de le manipuler au travers de la grille [Wollacott 07].

pour l'interaction d'un ligand avec une protéine (figure 1.17). Cette méthode est très largement utilisée dans les simulations de *docking* moléculaire [Meng 92], [Pattabiraman 86], [Tomioka 94].

Le choix de la taille de la grille est important afin de limiter les retenues mémoires qui ont une complexité en N^3 , où N est le nombre de points pour chaque grille. Précalculer des forces induit une instabilité lorsque l'objet n'est plus situé sur un point de la grille. Une interpolation de la force entre chaque point de la grille doit donc être effectuée. Afin de ne pas altérer l'interface haptique, toutes les forces interprétées par celle-ci ne dépassent pas $1.5 N$, celles-ci étant calculées par un coefficient homothétique. [Wollacott 07] rend compte de la forte instabilité des forces durant un *docking* moléculaire (un ion potassium au travers d'un site actif de protéine 18-crown-6), rendant difficile une manipulation précise. Plusieurs méthodes sont alors mises en place, comme l'ajout de viscosité. L'inconvénient majeur réside dans le fait que les forces de van der Waals sont non linéaires, rendant une interpolation linéaire peu efficace et induisant une discontinuité des forces. Le problème de l'instabilité est notamment causé par la raideur importante des forces de van der Waals dont l'ordre de grandeur est en $1/r^6$.

[Lee 04] considère le *docking* d'un ligand rigide (les distances interatomiques sont fixes) dans une protéine rigide. Son approche consiste, de façon similaire, à précalculer des forces d'interaction en un nombre fini de points d'une grille prédéfinie. Le récepteur est décomposé en une grille régulière. Augmenter le nombre de points rend la précision plus importante mais cette contrainte est limitée par la taille de la mémoire. Pour chaque type atomique du ligand, celui-ci est positionné en un point du maillage et la force exercée par le récepteur sur le ligand est calculée et sauvegardée. Les forces intramoléculaires sont négligées. En chaque point, l'énergie totale calculée vaut alors :

$$E = \sum_{i=1}^{\text{ligand}} \sum_{j=1}^{\text{récepteur}} \left[\frac{A_{ij}}{r_{ik}^{12}} - \frac{B_{ij}}{r_{jk}^6} + 1389.1 \frac{q_i q_j}{\epsilon_0 r_{ij}} \right] \quad (1.24)$$

où A et B sont les paramètres d'attraction et de répulsion de van der Waals, r_{ij} est la distance entre l'atome i et l'atome j , q_i et q_j sont les charges des atomes i et j , ϵ_0 est la permittivité diélectrique du vide.

Physiquement, deux atomes ne peuvent s'approcher à moins de leur distance définie par leur rayon de van der Waals. Le schéma de couplage proposé diffère des travaux précédents. Il considère loin d'une surface de contact les forces précalculées (équation 1.24) puis lors d'un contact, une force proportionnelle à la pénétration d'un rayon de van der Waals dans un autre.

Cette force de contact peut être assimilée à une force de contact entre un objet et un mur. Le modèle assimile le contact à un ressort. La force ainsi calculée en chaque point de la grille, au contact, pendant la manipulation vaut :

$$\vec{F} = \vec{F}_{maillage} + k(\vec{X} - \vec{p}) \quad (1.25)$$

où \vec{p} représente la projection du centre de l'objet \vec{X} au point de contact. La complexité consistant à trouver le point de projection sur le « mur » est proportionnelle au nombre d'atomes présents dans le récepteur.

Afin de réduire le temps de calcul, une autre grille volumique est utilisée. L'espace est alors divisé en cellules de taille unité. Pour chaque cellule, chaque atome du récepteur susceptible d'interagir avec un atome de ligand est précalculé et sauvegardé. Un test d'interaction entre les atomes du ligand et le récepteur est alors effectué. De la collection d'intersections avec le récepteur, les projections peuvent être calculées. Le temps de calcul de la détection de collision est très coûteux. Afin de l'alléger, chaque atome du récepteur possède une liste d'atomes du ligand potentiellement capables d'interagir avec eux.

Cependant, au contact, les forces divergent très rapidement, provoquant des instabilités dans le retour d'effort. Après application d'un simple facteur de coupure, des oscillations apparaissent, rendant difficile un contact lisse car la force ressentie possède un gradient très élevé. Le profil de force adopté autorise alors une faible pénétration et la réponse en force est graduelle. Le facteur de force est défini après une première manipulation. La force maximale calculée au cours de la simulation est gardée comme force maximale admissible par les moteurs de l'interface. Toutes les forces alors calculées sont proportionnelles à cette valeur. Cependant, lors d'une collision, le nombre d'atomes en contact avec le récepteur agit sur la résultante des forces. Plus le nombre d'atomes en contact est élevé, plus la résultante de la force l'est aussi. Le coefficient de force ne peut donc rendre fidèlement toutes les situations. Par exemple, s'il est calculé pour un atome en collision, lors de la collision simultanée de plusieurs atomes, la force calculée provoquera une saturation de l'interface. A l'inverse, si ce coefficient est calculé pour plusieurs collisions, un faible nombre de collisions ne sera pas fidèlement ressenti.

[Stone 01], [Nagata 02], [Wollacott 07] et [Lee 04] décrivent l'intérêt de l'apport de l'haptique dans la compréhension des phénomènes moléculaires. En effet, pour la perception des forces électrostatiques ou, tout simplement, pour scanner une surface moléculaire, les sites actifs des molécules ont été positionnés avec précision grâce au ressenti des barrières d'énergie, des zones attractives et des zones répulsives. Loin du site actif, les forces semblent répulsives. A l'approche de celui-ci, les forces sont attractives si l'objet approché est simplement sphérique et muni d'une charge ou dépendant de l'orientation de la petite molécule manipulée. Auparavant, les sites actifs de certaines molécules n'avaient pu être repérés par les techniques de simulations moléculaires propres. L'haptique simplifie alors à partir du ressenti des propriétés de la molécule la problématique de la localisation de certaines zones.

Cependant, le temps de calcul des forces et leur grande variation sont deux composantes qui rendent difficilement réalisables un *docking* moléculaire en comparaison d'un simple déplacement d'ion sur une surface moléculaire. C'est pourquoi les travaux précédemment cités se sont basés sur un précalcul des forces ainsi que sur des systèmes rigides.

Ce type d'approche nécessite un stockage important de données. La rigidité des modèles amène à appréhender la difficulté de l'échelle moléculaire mais pas d'effectuer un *docking* qui

nécessite une adaptation de la part du site actif. Les schémas de commande proposés sont classiques et basés sur un modèle de ressort entre l'interface haptique et la simulation ne gérant pas les retards dus aux temps de simulation. La stabilité de la commande n'est alors pas garantie et le rendu haptique assuré par un simple couplage homothétique. Il n'est pas fait mention d'un réel travail sur la perception des échelles moléculaires. La grande variation des forces n'est pas abordée dans la plupart des cas ou simplement sur un cas rigide/rigide, par l'introduction d'un facteur de coupure, prévenant la saturation des moteurs de l'interface. De plus, la flexibilité du ligand augmente considérablement cette dynamique. Cette recherche de flexibilité entre le ligand et le site actif de la protéine est au centre des travaux développés au cours de la thèse. Les approches proposées ci-dessus sont difficilement adaptables à un système flexible.

3.2 Approche flexible/rigide

Ce paragraphe recense les études conduites autour des simulations interactives de *docking* avec retour d'effort. Cependant, au cours de la manipulation du ligand, celui-ci est rendu flexible grâce à l'introduction de la minimisation de l'énergie totale. En effet, il a été démontré qu'un ligand rigide est moins apte à se lier à son site actif [Bayazit 01]. Le site actif reste, lui, rigide. Les forces et les moments sont calculés et l'utilisateur ressent ces efforts *via* un couplage bilatéral.

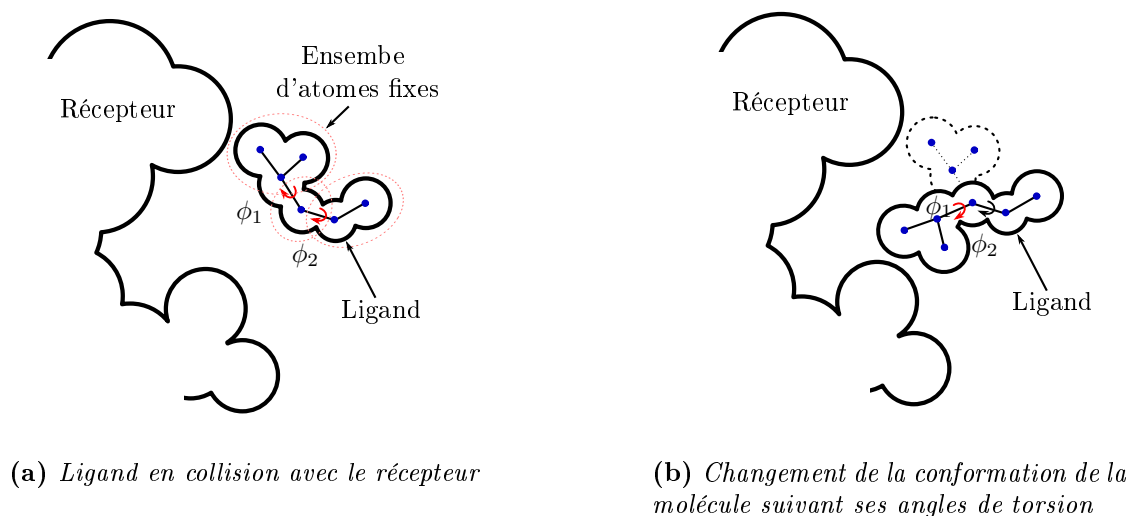


Figure 1.18 - *Changement de conformation d'un ligand au contact d'un récepteur avec minimisation de l'énergie. Les cercles rouges représentent un ensemble d'atomes fixes englobés dans une enveloppe afin d'optimiser le traitement des collisions.*

[Lai-Yen 05] et [Lai-Yen 06] considèrent un *docking* de ligand flexible avec une molécule rigide et un ressenti des forces et des moments, les nouvelles positions du ligand étant déterminées par minimisation du champ de force (figure 1.18). Afin de simplifier les calculs et de fournir les efforts nécessaires au couplage en temps réel, les forces électrostatiques sont négligées devant les forces de van der Waals. La contribution de van der Waals entre le ligand et les atomes du récepteur est choisie sous sa forme 6-12 de Lennard-Jones. A nouveau, deux maillages volumétriques de l'espace sont utilisés. Le premier dont les nœuds sont espacés de 0.5 Å est utilisé pour le précalcul de la force d'interaction. Le deuxième est utilisé pour le précalcul de la liste des collisions possibles entre le ligand et les atomes du site actif afin d'accélérer la détection

de collision.

En reprenant les travaux précédents, la force appliquée sur le ligand s'exprime par :

$$\vec{F}_{\text{Total}} = \vec{F}_{\text{Potentiel}} + \vec{F}_{\text{Collision}} \quad (1.26)$$

$$\text{soit } F_{\text{Total}} = -\frac{d}{dr_{ij}} \left(\sum_{i=1}^{N_{lig}} \sum_{j=1}^{N_{rec}} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right] \right) + \sum_{i=1}^{n_{collision}} \delta K (p_i^{mur} - p_i) \quad (1.27)$$

où δ est le symbole de Kronecker et vaut un si le ligand p_i entre en collision avec un autre atome du maillage et zéro sinon. p_i est le centre de l'atome i du ligand en collision, p_i^{mur} est la projection de p_i sur le mur virtuel, K une constante de raideur prédéfinie et $n_{collision}$ le nombre de collisions entre le ligand et le récepteur.

Les moments sont calculés à partir du centre de masse du ligand p_{cdm} . La force précédemment calculée est considérée appliquée au centre de masse de la molécule. Le couple $\vec{\tau}$ induit en ce point par une collision en p_i vaut :

$$\vec{\tau} = \sum_{i=1}^{N_{lig}} \left[\sum_{j=1}^{N_{rec}} \overrightarrow{p_{cdm} p_i} \times \vec{F}_{ij}^{collision} \right] \quad (1.28)$$

L'opérateur peut alors manipuler le ligand en translation et en rotation. Un processus de minimisation de l'énergie ainsi que d'évitement d'obstacles agit sur les degrés de liberté internes au ligand, ses angles de torsion. Le processus de minimisation ne modifie que les angles $(\phi_k)_{k=1\dots n}$ afin de trouver la conformation de plus basse énergie ne provoquant pas de collision. Si une nouvelle conformation respectant les conditions précédentes est trouvée, cette nouvelle conformation est gardée. Sinon la force d'interaction est renvoyée ainsi que la conformation non actualisée. Les résultats montrent que les zones pouvant potentiellement accepter un ligand sont rapidement identifiées grâce à l'haptique. Celui-ci étant déformable, il pénètre le site actif contrairement au cas rigide.

Ce modèle d'interaction négligeant les forces électrostatiques est discutable. En effet, les forces de van der Waals sont des forces de courte portée comparées aux forces électrostatiques. De plus, le modèle de rigidité de la protéine n'autorise pas l'accès de sites actifs en profondeur nécessitant l'ouverture de la protéine. Enfin, bien que les simulations semblent être temps réel, un accroissement du temps de calcul rendrait la commande instable.

Après avoir modélisé un système atomique comme une chaîne polyarticulée, [Morin 07] définit la force appliquée par l'utilisateur comme étant la différence de la position de l'interface X_h et d'une position de consigne X_c :

$$\vec{F}_{\text{utilisateur}} = k_s \left(\overrightarrow{X_h - X_c} \right) \quad (1.29)$$

Basée sur les équations spatiales de Featherstone décrivant le mouvement d'objets rigides, la force appliquée au solide en question implique que son équation de mouvement devient :

$$\mathbf{I}\vec{\gamma} = \vec{F}_{\text{utilisateur}} - \vec{X} \times \mathbf{I}\vec{X} \quad (1.30)$$

où \mathbf{I} est la matrice d'inertie du solide, $\vec{\gamma}$ son accélération spatiale et \vec{X} la vitesse spatiale. L'objet sélectionné est sujet à d'autres forces \vec{F}_{ext} (comme les forces appliquées par d'autres objets), et des forces de contraintes internes \vec{F}_c dues à la cinématique du solide. L'équation du mouvement s'écrit alors :

$$\mathbf{I}\vec{\gamma} = \vec{F}_{\text{utilisateur}} + \vec{F}_{\text{ext}} + \vec{F}_c - \vec{X} \times \mathbf{I}\vec{X} \quad (1.31)$$

La force $\vec{F}_{\text{haptique}}$ à renvoyer est définie comme étant la différence des deux accélérations des équations 1.30 et 1.31 :

$$\vec{F}_{\text{haptique}} = \vec{F}_{\text{ext}} + \sum \vec{f}_i \quad (1.32)$$

Afin de calculer les forces externes \vec{F}_{ext} et internes \vec{f}_i , le principe de l'algorithme adaptatif expérimenté par [Redon 05] est utilisé. Le calcul des forces est effectué pour une région présélectionnée, et les positions des objets sont modifiées en conséquence. Le retour haptique, tel que spécifié dans les travaux ci-dessus ne font pas état d'une recherche spécifique sur la perception des échelles moléculaires. Les efforts sont de l'ordre du nano-Newton, il n'est pourtant pas fait mention de l'utilisation de coefficients pour l'interprétation de ces forces dans un monde macroscopique et l'adéquation entre les déplacements de l'interface et ceux du monde moléculaire. De plus, le couplage utilisé est un ressort dont le gain variable autorise de jouer sur la rigidité du couplage. Il reste donc à concevoir une méthode pour la perception de ces échelles.

3.3 Bilan

De l'étude des approches proposées depuis les débuts de l'interfaçage des systèmes moléculaires avec un système à retour d'effort, il ressort qu'un des principaux problèmes provient de la grande raideur des forces électrostatiques et de van der Waals. Ces raideurs sont à l'origine de la grande instabilité des profils des forces d'interaction. L'interactivité n'est que relative car dans la plupart des cas une grille volumétrique sert de base au précalcul des forces. Cette approche est proposée pour éliminer le temps de calcul nécessaire pour l'évaluation des forces. Cette méthode ne peut être appliquée que si l'objet manipulé est rigide. En effet, il n'est pas possible d'envisager, pour une position donnée, de précalculer un ensemble de conformations du ligand en interaction avec une protéine.

Des approches moins réductrices envisageant la flexibilité du ligand utilisent à nouveau un maillage volumétrique pour le précalcul des forces de van der Waals. Elles précalculent à nouveau les forces d'interaction en certains points et laissent l'optimiseur d'énergie actualiser les positions du ligand. Cependant, devant le temps de calcul nécessaire pour le changement de conformation, le modèle d'interaction, très réducteur, ne prend en compte que les forces de van der Waals.

Les instabilités liées à la grande raideur des forces de van der Waals ont été résolues par l'application de forces bornées pour le couplage bilatéral. Cependant, malgré la limitation de l'amplitude des forces, l'oscillation résultante n'est atténuée que par l'ajout volontaire de viscosité, au détriment du ressenti. Enfin, les seuls travaux relatant une manipulation tridimensionnelle d'un ligand flexible avec une protéine rigide, avec un retour d'effort en translation et en rotation, décrivent celle-ci selon une grille prédéfinie. Le retour d'effort proposé lors d'un contact est un ressort linéaire dépendant de la profondeur de pénétration du ligand dans la protéine engendrant une forte instabilité. Les moments ne sont ressentis que lors des contacts du ligand avec les protéines et non pas au cours de la manipulation quelle que soit la position du ligand.

Les travaux présentés dans ce paragraphe ne font pas état d'un réel travail sur les métaphores haptiques pour garantir à l'opérateur la compréhension du phénomène moléculaire. Seuls les travaux de [Stone 01] font état d'un couplage bilatéral avec des homothéties en effort et en déplacement. Les autres travaux considèrent un ressort virtuel, correspondant au contrôleur haptique, qui calcule la force à ressentir comme étant proportionnelle à la profondeur de pénétration du ligand au point de contact considéré. Mais en dehors des contacts, il n'est pas fait mention de coefficients en effort et en déplacement pour l'adéquation des grandeurs physiques du monde macroscopique et moléculaire.

Tous les travaux cités cherchent à optimiser le temps de réponse du simulateur moléculaire pour garantir une réponse de celui-ci à la fréquence du kilohertz. Les couplages proposés dépendent donc des champs de force utilisés et des simulateurs moléculaires mis en place. Un couplage haptique indépendant de la simulation ne peut donc être envisagé.

4 Approche proposée

L'objectif recherché par cette thèse est d'insérer le projet de retour d'effort pour des applications moléculaires interactives dans les étapes de la conception des ligands. Une fois les opérations de *screening* virtuel réalisées, il reste à optimiser géométriquement une dizaine de ligands candidats. Le test de l'affinité des candidats pour le récepteur considéré, doit être réalisable à partir des logiciels effectuant le *screening* pour des raisons de continuité des études physico-chimiques. Se pose ainsi la contrainte de la compatibilité de la plate-forme de téléopération avec les logiciels de simulation moléculaire existants. La question de la réutilisation des méthodes de simulation existantes pour la discrimination des géométries restantes est ainsi cruciale. Ces méthodes de simulation ainsi que le logiciel réutilisé ont leurs contraintes propres pour le couplage bilatéral mais semblent, du point de vue de nos partenaires, en corrélation avec leurs travaux.

L'approche proposée dans ce paragraphe tient compte des bilans effectués aux deux paragraphes précédents portant sur la simulation moléculaire et sur l'état de l'art en haptique moléculaire. Des contraintes industrielles concernant le choix de la simulation moléculaire, et des contraintes imposées par les systèmes robotiques doit naître la plate-forme de téléopération moléculaire avec retour d'effort. Son objectif est de pouvoir aider à la discrimination des géométries de ligands candidats dans les étapes de la conception des médicaments.

Cette partie reprend les différents éléments et contraintes pour proposer une solution originale au retour haptique du *docking* moléculaire. Ainsi, le premier paragraphe concerne la flexibilité des protéines. En effet, une approche rigide-rigide engendrerait un calcul rapide des interactions. Cependant le comportement ainsi simulé ne reflète pas la réalité. Une étude est donc conduite afin de déterminer l'intérêt de la flexibilité pour le *docking* moléculaire.

Le deuxième paragraphe concerne la réutilisation d'un logiciel de simulation. En effet, sont envisagées deux possibilités : la première est de concevoir un nouvel algorithme, robuste, pour l'évaluation de l'énergie d'interaction et son optimisation ; la deuxième tend à préférer l'utilisation d'un logiciel complet impliquant des contraintes temps réel à résoudre pour la stabilité du couplage.

Enfin le troisième paragraphe répond aux attentes du retour d'effort et pose le cadre de ma recherche.

4.1 Ligand flexible et protéine flexible

Dans le but de prédire les interactions entre éléments biologiques, la plupart des méthodes de *docking* traitent les protéines comme des corps rigides. Ces méthodes échouent généralement si au cours de l'association les protéines subissent des changements conformationnels internes, par exemple au niveau de leurs boucles de surface.

La flexibilité des protéines est primordiale dans une très grande majorité des mécanismes cellulaires. Les protéines se déforment pour s'adapter à leur partenaire, pour former des interactions spécifiques, pour éviter des gênes stériques ou pour améliorer la complémentarité de surface autorisant la formation de liaisons hydrogènes. Les protéines peuvent encore changer leur conformation au cours de réactions enzymatiques.

De manière générale il est possible de distinguer deux types de mouvements. Le premier correspond à des mouvements de grande amplitude pour lesquels deux ou plusieurs parties de protéine bougent les unes par rapport aux autres. Ce type de mouvement est souvent observé au cours des réactions catalytiques [McCammon 87], [Kumar 99]. Le deuxième type de mouvement implique des mouvements de faible amplitude qui reflètent l'équilibre de la protéine autour de son état natif dans le minimum de son potentiel énergétique.

La simulation du *docking* ligand-protéine doit donc faire intervenir ces contraintes de flexibilité du ligand et de la protéine. Les outils pour simuler une telle flexibilité sont, entre autres, la dynamique moléculaire et la minimisation de l'énergie. Le choix de la méthode de simulation du changement conformationnel s'est orienté vers une minimisation classique de l'énergie du système étudié. En effet, l'objectif n'est pas l'exploration conformationnelle du ligand et de la protéine mais bien la recherche de leur conformation d'énergie minimum au cours de la manipulation. La dynamique moléculaire ainsi que les autres méthodes ne répondent pas exactement à cet impératif. Dès lors, la modélisation de l'énergie d'interaction peut être conduite par mécanique quantique, par des méthodes semi-empiriques ou même empiriques.

La contrainte temps réel imposée par l'utilisation d'une interface à retour d'effort oriente le choix de la modélisation des interactions vers une méthode empirique. En effet, ces modélisations sont plus rapides à calculer bien que moins précises. Les interactions seront donc modélisées par un champ de force qui pourra être choisi en fonction du système moléculaire à simuler.

Le choix de la plate-forme pour un retour haptique intégrera donc un processus de minimisation de l'énergie évaluée par un champ de force empirique, garantissant ainsi que le système biologique est dans un état d'énergie minimum au cours de la manipulation.

Le choix d'une modélisation des interactions fixé - un champ de force - et la méthode d'optimisation arrêtée - minimisation - la question de la création d'un logiciel se pose.

4.2 Logiciel de simulation du *docking* moléculaire

Devant la grande variété de logiciels capables de modéliser et de simuler le comportement protéique, il a été décidé, afin de ne pas orienter spécifiquement ces travaux de thèse vers la conception d'un logiciel, de réutiliser un logiciel existant. Il est vrai que le besoin d'une réponse temps réel pour des applications haptiques aurait pu nécessiter la conception d'un nouvel algorithme. Mais cette thèse, exploratrice par son application, ne pouvait à la fois engendrer la création d'un logiciel dans une unité robotique où la connaissance en bio-informatique est limitée, et une recherche active sur les lois de commande pour l'interfaçage d'un tel système avec un bras à retour d'effort.

La collaboration avec Sanofi-Aventis portant sur le choix des molécules à simuler a orienté le besoin vers une brique logicielle haptique à intégrer à l'environnement de simulation « *Molecular Operating Environment* ». En effet, MOE est souvent utilisé dans les services de conception de nouvelles molécules. L'idée d'intégrer un module haptique capable, après une étude quelconque de fournir des informations supplémentaires sur les forces d'interaction, oriente le choix du simulateur moléculaire vers ce logiciel préexistant.

La partie de la plate-forme de manipulation concernant la modélisation et la simulation moléculaire sera donc réutilisée. L'environnement de simulation sera un logiciel évolutif, certes pas temps réel, mais complet pour la conception de médicaments.

4.3 Retour d'effort

Les travaux sur le retour haptique dans les simulations de *docking* moléculaire, décrits précédemment, tendent à affirmer que l'ajout de l'haptique dans les simulations moléculaires est utile. Il apporte des solutions nouvelles pour la découverte de sites actifs grâce à la détection des zones attractives et répulsives. Il a aussi fourni des informations sur les propriétés physiques des zones externes aux protéines grâce à un parcours de sa surface extérieure.

Ces travaux autorisent à penser qu'un retour d'effort lors d'un *docking* moléculaire pourrait fournir des informations discriminantes, autres que visuelles, à l'opérateur. A partir d'une base de données réduite de formes géométriques potentiellement complémentaires du site actif considéré, l'opérateur peut être en mesure de distinguer grâce au retour d'effort la différence d'affinité des ligands potentiels pour le site actif considéré. Ainsi, en plus des informations fournies par les simulations moléculaires automatisées, l'opérateur peut lui-même agir sur le processus de *docking* et déterminer la forme géométrique la plus adaptée.

Le choix d'un retour haptique dans les trois directions de l'espace en translation et en



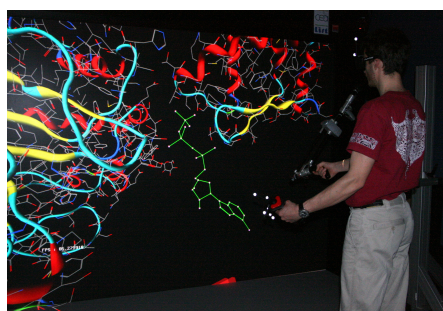
(a) Molécule de glucose dans une protéine membranaire

(b) Molécule de glucose dans une protéine membranaire vue de face

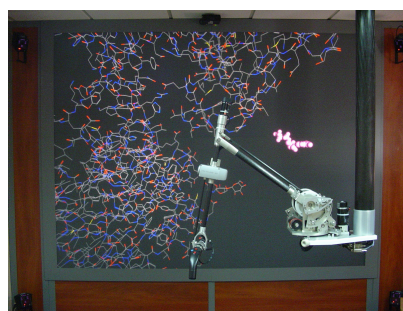
Figure 1.19 - Le passage d'une molécule au travers d'une protéine membranaire est un des cas où les informations sur les moments sont utiles.

rotation peut sembler inutile pour un *docking* de ligand dans une protéine. En effet, lors de l'actualisation des positions du ligand dans le site actif, la variation de position due au calcul d'un moment selon un axe n'est pas aisément quantifiable. Ainsi, dans l'état de l'art en haptique moléculaire, les systèmes étant pour la plupart rigides, le calcul de ces moments est aisé puisqu'il s'agit d'une rotation d'un corps rigide autour de son centre de masse. Cependant, plusieurs applications nécessitent de ressentir les moments lors de la manipulation d'une protéine et il convient d'inclure cette spécificité dans notre cahier des charges.

La figure 1.19 montre, par exemple, une des applications possibles du retour d'effort dans les six degrés de libertés autorisés par l'interface haptique. En effet, lorsque la protéine (le glucose) est entrée dans le canal membranaire, il peut être intéressant d'en étudier les propriétés mécaniques et notamment de déformation le long de son grand axe. Les moments peuvent donc donner une information supplémentaire.



(a) Manipulation d'un ligand avec le *Virtuose* dans une scène en trois dimensions.



(b) Le *Virtuose* dans la salle haute définition.

Figure 1.20 - Interface à retour d'effort *Virtuose* couplée au logiciel *MOE* en salle de réalité virtuelle du *CEA-LIST*.

Devant le besoin d'un retour d'effort selon six degrés de liberté, le choix de l'interface s'est alors orienté vers un dispositif capable d'assurer un tel retour. Le *Virtuose 6D¹* (figure 1.20), bras maître existant au Commissariat à l'Énergie Atomique (*CEA*), peut développer un effort maximal de 35 *N* pour une course verticale de 90 *cm* et horizontale de 1 *m*. Il se commande

¹Haption - www.haption.fr

à partir d'un PC industriel. Il est cependant dédié pour des applications macroscopiques telles que le prototypage virtuel de voitures par exemple. Sa conception n'est donc pas dédiée à des applications de manipulation moléculaire.

Au regard des informations données dans ce chapitre, plusieurs problèmes directement induits par l'utilisation d'une interface haptique possédant ses propres spécificités apparaissent :

- le temps de calcul induit par la minimisation et l'évaluation de l'énergie intermoléculaire semble trop long pour garantir la stabilité de la commande ;
- les interactions moléculaires sont décrites par le champ de force en termes d'énergie et non pas en termes de forces ni de moments ;
- un *docking* moléculaire doit être réalisé avec un ligand flexible et une protéine flexible tendant à augmenter le temps de calcul.

La plate-forme doit en outre être générique par rapport aux modèles d'interaction. En effet, concevoir une plate-forme pour un type particulier de champ de force la rend dépendante des propriétés physico-chimiques à simuler. Les protéines responsables des maladies n'ayant pas toutes les mêmes spécificités, concevoir un outil adaptable à leurs propriétés apparaît être primordial pour une utilisation scientifique active. La méthode de simulation, en l'occurrence une minimisation de l'énergie, quant à elle, n'a pas besoin d'être modifiée. La plate-forme à concevoir devra donc être indépendante des modèles d'interaction et utilisera des algorithmes interchangeables de minimisation de l'énergie.

5 Conclusion

La problématique du *docking* moléculaire, objet de l'étude menée tout au long de ce chapitre, est mieux appréhendée. Il s'agit d'optimiser, à la fois, la position et l'orientation des ligands dans le site actif d'une molécule et la géométrie des candidats pour une affinité optimale entre les deux molécules. Cette affinité est censée modifier la structure tridimensionnelle de la protéine en vue d'inhiber ses effets macroscopiques indésirables fonctions de sa géométrie. Ressentir sur un bras à retour d'effort l'affinité de plusieurs candidats envers un site catalytique conduirait à choisir la meilleure géométrie possible et ainsi garantir l'efficacité du médicament.

Les études conduites sur le retour d'effort, lors de simulations de *docking*, l'ont principalement été avec un simulateur de dynamique moléculaire. Elles considèrent la plupart du temps un ligand rigide ainsi qu'un site actif rigide. Le retour d'effort étudié est effectif uniquement pour la translation négligeant l'importance des moments pour certaines interactions. Enfin, les recherches sur la perception des échelles moléculaires sont inexistantes. La question de l'adéquation des grandeurs physiques entre le macromonde et l'environnement moléculaire n'est abordée qu'au travers de coefficients homothétiques, en effort et en déplacement, fixes.

Devant cet état de fait et compte tenu des besoins de nos partenaires, nous nous sommes orientés vers une plate-forme générique, réutilisant leur logiciel de simulation et proposant un retour d'effort avec six degrés de liberté. Dans la suite des travaux présentés dans ce manuscrit, le retour d'effort est considéré pour la translation mais aussi pour la rotation. En effet, il doit

impérativement intégrer les moments pour tenir compte du cas d'une protéine testée dans un canal membranaire. Le retour d'effort doit pouvoir rendre possible une discrimination géométrique des ligands par le seul ressenti du champ d'énergie intermoléculaire. Cette condition forte impose un ressenti stable des interactions moléculaires et une étude du transfert des propriétés microscopiques vers le macromonde.

La simulation moléculaire sera réalisée par un logiciel dont les caractéristiques rendent son évolution possible. Cependant, les simulations font intervenir l'énergie d'interaction. L'introduction d'une interface haptique nécessite la conversion de l'énergie en forces d'interaction. Enfin, le logiciel MOE, bien que performant dans le domaine moléculaire n'est pas orienté temps réel. Les simulations, du fait de leur complexité, ne fournissent pas d'information d'énergie dans des temps compatibles avec une interface à retour d'effort.

L'approche ligand flexible-protéine flexible n'a pas encore été testée pour une application à retour d'effort faisant intervenir une interface à six degrés de liberté. Les chapitres suivants s'attachent à rendre possible une telle plate-forme de téléopération. Sont ainsi étudiées des lois de commande spécifiques, tenant compte des particularités du *docking* telles que le retard du calcul des forces ou la forte variabilité de celles-ci.

En particulier, le chapitre suivant est un premier contact avec l'interfaçage d'un organe à retour d'effort avec deux molécules en interaction. Il met en place la commande d'un ligand flexible aux abords d'un site actif de protéine et met en place un premier parallèle entre deux mondes aux propriétés différentes.

ETUDE D'UN COUPLAGE SIMPLE POUR LE DOCKING

La plate-forme, telle que proposée au chapitre précédent, se compose de plusieurs parties. La première est l'interface haptique qui est un système de la classe *Virtuose* capable d'assurer un retour de forces et de moments. La deuxième partie est constituée du logiciel de simulation moléculaire MOE. De par ses fonctionnalités étendues, il est capable de modéliser les interactions moléculaires et de simuler le changement de conformation de la protéine pour garantir une stabilité énergétique. Les interactions sont modélisées par un champ de force quelconque et la recherche du minimum énergétique est assurée par la minimisation de l'énergie totale du système. Le principal désavantage de ce logiciel est qu'il n'est pas conçu dans l'optique d'assurer une minimisation temps réel. La troisième partie concerne la commande bilatérale du ligand qui doit assurer la communication entre l'interface à retour d'effort et le logiciel de simulation. L'étude de cette commande est l'objet de ce chapitre.

Ce chapitre propose après un rappel de généralités concernant le retour d'effort l'étude d'un premier couplage homothétique appliqué au *docking* moléculaire. Dans un premier temps, afin d'accélérer la simulation moléculaire, le ligand est considéré rigide puis une fois la commande validée, il est considéré flexible.

La plate-forme de téléopération est ensuite décrite précisément en suivant l'évolution d'une consigne envoyée par l'interface haptique au simulateur moléculaire qui l'évalue puis calcule la force à appliquer, en retour, à l'interface haptique.

Dans un troisième temps, seront étudiées l'influence des coefficients homothétiques sur le ressenti des efforts moléculaires dans le monde macroscopique puis l'influence du retard des consignes causé par une minimisation de l'énergie qui n'est pas temps réel, sur la stabilité de la commande.

La conclusion dresse un bilan des difficultés abordées pour proposer une évolution de la commande afin de garantir sa stabilité.

1 Couplage haptique : généralités

1.1 Introduction

Les débuts de la téléopération sont apparus avec des systèmes maître-esclave mécaniquement liés. La transmission entre les articulations du bras maître et du bras esclave est effectuée par des systèmes à câbles passant par de nombreuses poulies. Cependant, la liaison mécanique entre les parties maître et esclave possède un inconvénient car ces ensembles ne peuvent être éloignés de plus de quelques mètres. Les liaisons mécaniques sont alors remplacées par des liaisons électriques. Chaque articulation est entraînée par un moteur. Ces bras conservent le retour d'effort et, au travers de leurs liaisons électriques, autorisent une distance quelconque entre les parties maître et esclave. L'essor de la téléopération voit réellement le jour avec l'introduction de l'ordinateur dans la chaîne de commande. Sa principale qualité est de transformer toute information d'entrée en information de sortie mais adaptée aux objectifs désirés (la télé-micromanipulation devient par exemple possible) et aux moyens utilisés pour atteindre ces buts. Les transformations mécaniques deviennent des transformations effectuées par l'ordinateur. Selon le cas, l'ordinateur peut être un assistant, un coopérant ou peut même se substituer à l'opérateur.

L'avènement de la réalité virtuelle rend possible un gain de temps conséquent en ne créant plus de maquettes identiques au réel. Ainsi, de nombreux secteurs, telles la chirurgie, l'industrie automobile ou l'aéronautique, sont intéressés par le retour d'effort sur maquette numérique. Avec des simulations interactives, l'idée est d'avancer vers une maquette numérique des produits pour que les mêmes tests soient effectués dans des délais plus restreints et à moindre coût. Cette approche est appelée prototypage virtuel. A l'aide de l'interface haptique, l'utilisateur manipule les objets conçus qu'il visualise à l'écran.

Dans le cas de la pharmacologie, il peut vérifier qu'une molécule nouvellement créée se fixera bien sur le récepteur membranaire d'une cellule grâce à la sensation du retour d'effort. Celui-ci est ainsi utilisé afin que l'opérateur puisse discriminer des géométries de ligands en intervenant directement dans le processus de criblage virtuel actuellement entièrement automatisé. Les simulations ne proposant qu'une information visuelle sous forme de score correspondant à l'accointance du ligand avec son site actif, l'adjonction du retour d'effort comme paramètre décisionnel supplémentaire peut *a priori* fournir un élément supplémentaire de comparaison.

Cependant, l'ordinateur est source de beaucoup d'instabilités dans le schéma de commande notamment en raison de la discontinuité des signaux, du temps de transfert des données dans les lignes de transmission et de la non-continuité des schémas physiques classiques.

Les critères d'évaluation d'un système haptique sont sa stabilité et sa performance. La passivité garantit la stabilité du système, et la transparence, sa performance.

1.2 Passivité

Une première notion très importante concerne la passivité. Un système est dit passif s'il a la capacité de stocker, dissiper et restituer de l'énergie sans pouvoir en créer. La passivité est une condition suffisante pour garantir la stabilité [Adams 98]. Il est donc nécessaire de montrer que

tous les éléments constitutifs de la chaîne de commande sont passifs. Selon leur agencement, la stabilité sera garantie. L'avantage de garantir que chaque élément de la chaîne de commande est passif réside dans le fait qu'ils pourront interagir séparément avec n'importe quel autre environnement passif.

L'information qui circule au sein d'un dipôle constituant une chaîne de commande peut être imaginée en termes de puissance. Intuitivement, un système est passif s'il absorbe plus d'énergie qu'il n'en produit. Soit $P_{entrant}$ la puissance entrante dans le système, qui est par convention positive, représentant le produit de l'entrée (vecteur x) par la sortie (vecteur y) du système :

$$P_{entrant} = x^t y \quad (2.1)$$

La puissance calculée ne peut être que stockée ou dissipée afin de garantir la passivité. Une puissance de stockage $P_{stockée}$ doit donc être définie ainsi qu'une puissance de dissipation $P_{dissipée}$. Ces deux puissances sont définies positives et ne représentent à nouveau pas nécessairement une puissance au sens mécanique du terme. Un système satisfait la condition de passivité si :

$$P_{entrant} = P_{stockée} + P_{dissipée} \quad (2.2)$$

Soit, par rapport à l'énergie stockée :

$$P_{entrant} = \frac{d}{dt} E_{stockée} + P_{dissipée} \quad (2.3)$$

Au cours du temps, le système ne pouvant engendrer d'énergie et ne pouvant restituer que l'énergie acquise ou stockée initialement, l'équation précédente s'écrit :

$$\int_0^t P_{entrant} dt = E_{stockée}(t) - E_{stockée}(0) + \int_0^t P_{dissipée} dt \quad \forall t \geq 0 \quad (2.4)$$

Si la variation d'énergie entre l'entrée et la sortie du système est positive, alors il est dit dissipatif. A l'inverse, si cette variation d'énergie est négative, le système crée de l'énergie. Il n'est donc plus passif et peut devenir potentiellement instable.

Cette définition est valable pour toutes les entrées en forces et en flux. Elle impose la passivité du dispositif (les ports et le dispositif lui-même). Un théorème moins général consiste à ne regarder la passivité qu'aux ports du dipôle. C'est-à-dire de ne considérer la passivité du système de téléopération que du point de vue de l'opérateur si l'environnement distant est passif et vice versa. Cette condition de stabilité inconditionnelle est en effet suffisante dans la mesure où l'opérateur et l'environnement n'interagissent pas directement mais par le biais du système [Hannaford 03]. En d'autres termes, l'équation précédente revient à écrire que la matrice immitance \mathbf{P} (régissant le couplage) est réelle positive. elle vérifie donc que \mathbf{P} n'a pas de pôle dans la partie droite du plan de Laplace, ou simplement des pôles sur l'axe imaginaire pur. Un schéma de couplage en admittance peut s'écrire par exemple sous la forme quadripolaire suivante :

$$\begin{bmatrix} F_{maître} \\ F_{environnement} \end{bmatrix} = \mathbf{P} \begin{bmatrix} V_{maître} \\ V_{environnement} \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} V_{maître} \\ V_{environnement} \end{bmatrix} \quad (2.5)$$

Le système ainsi décrit est inconditionnellement stable en mode couplé si et seulement si le critère de Llewelyn, composé de trois inégalités scalaires, est vérifié :

$$Re(P_{11}) \geq 0, \quad Re(P_{22}) \geq 0$$

$$Re(P_{11})Re(P_{22}) - \left| \frac{(P_{21} + P_{12})^2}{2} \right| \geq 0 \quad \forall \omega \geq 0$$

Ces propriétés doivent être garanties pour chaque système :

- le *Virtuose* : le système haptique n'est constitué que de poulies et de câbles. Il est donc passif car il ne peut pas créer d'énergie ;
- le couplage : en continu, il est passif. Pratiquement, il est impossible de garantir une liaison rigide pour un système échantillonné. Il faut avoir recours à un autre type de couplage étudié *infra*. De plus, les retards dans la transmission influent sur la passivité ;
- l'environnement : l'environnement est une molécule. Elle ne crée pas d'énergie sans excitation. C'est *a priori* un système passif. Cette partie nécessitera une étude approfondie afin de garantir la passivité. En effet, l'environnement étant simulé, il peut être facteur d'instabilité à cause de l'échantillonnage du temps.

1.3 Couplage idéal sans retard

1.3.1 Couplage idéal

Un couplage idéal reproduirait l'action d'une barre rigide entre le déplacement de l'interface haptique et le déplacement de l'esclave sur la scène virtuelle. Cependant, pour des problèmes d'instabilité, ce couplage virtuel n'est pas envisageable. Une première explication serait de montrer qu'un gain infini représentant la raideur de la barre n'est pas réalisable par un système informatique échantillonné. Cette idée n'étant pas réalisable, le couplage mécanique virtuel mis en place correspond à l'utilisation d'un ressort-amortisseur entre le mouvement mesuré sur le dispositif et le mouvement simulé dans l'environnement virtuel. Si la simulation impose une contrainte extrêmement raide, celle perçue par l'opérateur ne sera pas infinie mais correspondra au couplage virtuel.

1.3.2 Commande du ligand

Considérons dans un premier temps un couplage en translation. La commande décrite à la figure 2.1 page suivante utilise les vitesses et les positions. Son équation est :

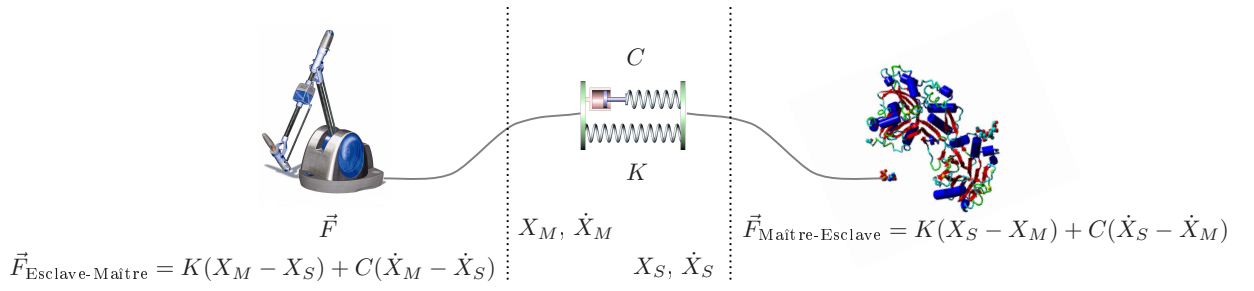


Figure 2.1 - Principe du couplage virtuel mis en place. Le ligand, considéré comme l'esclave est asservi au travers d'un ressort/amortisseur par le Virtuose.

$$F = KX + C\dot{X} \text{ soit } F = Z\dot{X} \text{ en termes d'impédance} \quad (2.6)$$

où F représente le torseur des efforts appliqués (sur l'esclave par exemple : figure 2.1), K la raideur du couplage, C l'amortissement, X et \dot{X} les position et vitesse du maître ou de l'esclave dans la simulation.

Lors du déplacement de l'organe haptique (le maître dans notre cas), une position est mesurée : X_M . Pareillement, à chaque instant, la position de la molécule est mesurée : X_S .

1.3.3 Principe

Dans ce mode de contrôle, le contrôleur (couplage Ressort/Amortisseur) lit en entrée une variation de position et de vitesse qui s'exprime à la sortie du *Virtuose*. L'opérateur impose une force au *Virtuose* qui l'interprète en termes de position et de vitesse par l'intermédiaire de capteurs. Cette variation (de position et de vitesse) est transmise à l'esclave par l'intermédiaire du contrôleur. La force appliquée vaut alors :

$$F_{\text{Maître-Esclave}} = K(X_S - X_M) + C(\dot{X}_S - \dot{X}_M)$$

Cette force est confrontée à la dynamique de l'environnement qui renvoie une nouvelle position de l'esclave notée X_S ainsi qu'une nouvelle vitesse \dot{X}_S . Le contrôleur réalise, cette fois en sens inverse, la commande et impose au maître une force qui vaut alors :

$$F_{\text{Esclave-Maître}} = K(X_M - X_S) + C(\dot{X}_M - \dot{X}_S)$$

La valeur des gains K et C dépend de la période d'échantillonnage, de la raideur et de la masse de l'objet que l'utilisateur manipule dans la simulation. Ces gains vont garantir le ressenti des contacts. S'ils sont trop faibles, l'utilisateur aura la sensation d'un contact mou. A l'inverse, une trop grande valeur des gains rendrait le système instable dans certaines configurations.

Lorsque les valeurs des forces $F_{\text{Esclave-Maître}}$ sont calculées, les valeurs des couples (Γ) à appliquer sur les moteurs sont obtenues à partir de la jacobienne (\mathbf{J}) transposée du dispositif haptique :

$$\Gamma = \mathbf{J}^t F_{\text{Esclave-Maitre}}$$

Le contrôleur prenant en entrée un couple position/vitesse et exprimant à sa sortie une force, l'asservissement est dit admittant.

Le couplage décrit ci-dessus est valable dans un sens comme dans l'autre. Si l'esclave est soumis à une excitation, l'organe haptique répondra de manière adaptée au signal fourni.

1.3.4 Facteurs d'échelle

Le *Virtuose* est conçu pour être manipulé par un opérateur humain sur une course de centaines de radians (de l'ordre du mètre) pour des efforts allant jusqu'à 35 N. En revanche, la taille d'un site actif de molécule est de l'ordre d'une dizaine d'Angström. Le ligand ne peut donc pas être manipulé sur des courses dépassant cette grandeur. Les forces mises en jeu sont, elles, de l'ordre du nano-Newton.

Quelle que soit la nature du couplage, il est nécessaire d'établir des facteurs d'échelle convertissant les efforts et les vitesses macroscopiques en grandeurs microscopiques et vice versa.

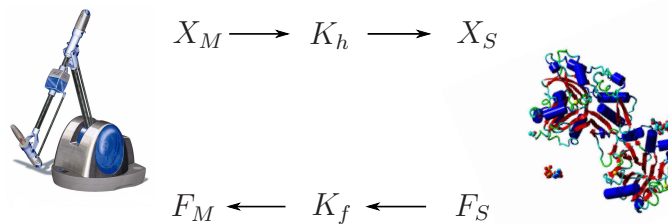


Figure 2.2 - Principe du couplage homothétique.

Pratiquement (figure 2.2), pour l'homothétie en effort K_f , il suffit de considérer le rapport de la force maximum applicable à la simulation sur la force maximum que l'opérateur veut imposer au dispositif haptique. Cette dernière est donnée par les caractéristiques du système. Le facteur d'homothétie en position K_h se calcule de la même manière, c'est le rapport entre la course maximum du ligand et la course maximum du dispositif.

1.4 L'environnement

Tout comme les lois de commande, les environnements peuvent se répartir en deux catégories :

- l'environnement peut se comporter en impédance, en acceptant des vitesses et/ou des positions, et en engendrant des forces selon un certain modèle physique. Le ligand est asservi en position, une force d'interaction ligand-protéine est déduite, l'environnement est donc dit impédant ;

- l'autre possibilité est un comportement semblable à une admittance en acceptant des forces et en retournant des mouvements (positions et/ou vitesses). C'est le cas, par exemple, des approches qui autorisent une gestion des contacts par contraintes.

La simulation de *docking* réalisée avec MOE fonctionne en impédance : Elle utilise en entrée une position pour retourner une force d'interaction.

1.5 Transparence

Un dernier critère de caractérisation de l'interface, après la stabilité, concerne la transparence.

En admettant qu'un dispositif simule le poids d'un objet, l'environnement virtuel calcule ce poids et envoie cette force au dispositif haptique. Le dispositif transmet cette force à l'utilisateur qui sentira non seulement cette force mais aussi le propre poids du dispositif. L'utilisateur se rend alors compte qu'il soulève quelque chose de plus lourd que l'objet virtuel qu'il voulait saisir. Il s'agit donc de compenser cette force parasite. La généralisation de cette technique s'appelle la transparence. Elle consiste à recenser toutes les perturbations que peut apporter un dispositif à retour haptique. La boucle de commande se charge alors de modifier les forces calculées afin de compenser ces perturbations.

Le principe de transparence d'un système de téléopération fait appel à la notion d'impédance mécanique par analogie avec l'électronique. Lorsque le ligand est en contact avec son environnement, il existe une relation qui lie la force F_h exercée par la molécule sur le ligand à la vitesse \dot{X}_S du ligand,

$$F_h = Z_e \dot{X}_S \quad (2.7)$$

où Z_e représente l'impédance transmise par la molécule au ligand.

De la même manière, il existe une relation entre la force F_M exercée par l'opérateur sur le *Virtuose* et sa vitesse \dot{X}_M , mettant en jeu l'impédance Z_M ,

$$F_M = Z_M \dot{X}_M \quad (2.8)$$

La condition pour que l'opérateur ressente exactement les interactions du ligand avec la protéine est que ces deux impédances soient identiques, aux facteurs d'homothétie près,

$$Z_M = -\frac{K_f}{K_h} Z_e \quad (2.9)$$

Cependant, l'impédance transmise par la molécule au ligand est difficilement quantifiable, elle est représentée par l'évaluation du champ de force. Celui-ci pouvant varier selon la protéine à simuler, cette relation générale ne peut donc fournir d'information analytique pour le réglage des gains de l'interface haptique.

2 Système mis en place

2.1 Architecture de la plate-forme de téléopération

La plate-forme de téléopération mise en place se compose de trois parties, toutes reliées entre elles par un protocole de communication de type UDP (figure 2.3).

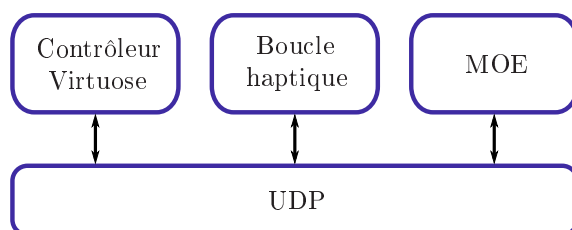


Figure 2.3 - Architecture générale de la plate-forme de téléopération.

La première partie est le contrôleur de l'interface haptique. Celui-ci, contenant les décodeurs responsables de la lecture des positions et l'interprétation des forces, fonctionne à une cadence nécessaire au ressenti haptique, soit une milliseconde.

La deuxième partie est la boucle haptique qui actualise les données entre le *Virtuose* et la simulation. Elle est donc responsable de la communication entre les deux entités. Cette boucle peut ne pas fonctionner à la cadence haptique. En effet, si le contrôleur ne reçoit pas de données actualisées à la cadence désirée, il applique la même donnée que reçue précédemment, cette donnée pouvant être une force, une position ou une vitesse.

Enfin, la troisième partie se compose de la simulation moléculaire. Celle-ci ne fonctionne pas à la cadence haptique. Dès qu'une force d'interaction est calculée, elle est envoyée à la boucle haptique. Celle-ci, après traitement, l'applique au contrôleur.

Ce schéma s'affranchit du traitement temps réel de la simulation en maintenant constantes les consignes en effort lorsque des retards de calcul ne fournissent pas les informations en effort à la cadence définie par le contrôleur. Cependant, afin de garantir que les forces ressenties reflètent bien la simulation en cours, il convient d'optimiser le simulateur moléculaire, représenté par l'outil logiciel MOE. Il s'agit d'optimiser ses performances afin que la réponse en force et en position n'excède pas quelques millisecondes. Cette contrainte pour un système comprenant de grosses molécules est très difficile à obtenir.

Le temps de réponse de la minimisation de l'énergie dépend de la taille du système et de la précision exigée. La figure 2.4 page ci-contre montre le temps de calcul nécessaire pour effectuer la minimisation de l'énergie d'une molécule composée de 250 atomes. Les pics correspondent à une optimisation des positions du ligand et du site actif. Ceux-ci présentent une moyenne de calcul de l'ordre de 200 *ms*. L'évaluation permanente de l'énergie totale du système, hors minimisation, est de l'ordre de 12 *ms*.

En considérant une application haptique, les forces doivent être calculées dans un délai ne dépassant pas quelques millisecondes pour que la stabilité de l'interface ne soit pas affectée.

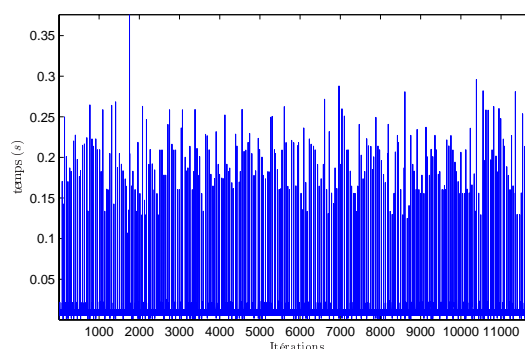


Figure 2.4 - Temps de calcul de la partie moléculaire intégrant la minimisation de 250 atomes au cours d'une simulation de docking moléculaire.

2.2 Optimisation du simulateur moléculaire

La minimisation de l'énergie d'un système moléculaire, induisant la variation des positions des atomes de la molécule est composée de trois méthodes utilisées successivement. Il s'agit, en premier lieu, de la méthode de la plus grande pente (SD) puis, après son critère d'arrêt, de la méthode du gradient conjugué (CG). En dernier lieu, afin d'optimiser les positions des atomes pour garantir un état stable, la méthode de Newton tronquée (TN) est appliquée. Les critères d'arrêt de ces méthodes sont un nombre d'itération limite ou une limite de gradient d'énergie à atteindre entre deux itérations.

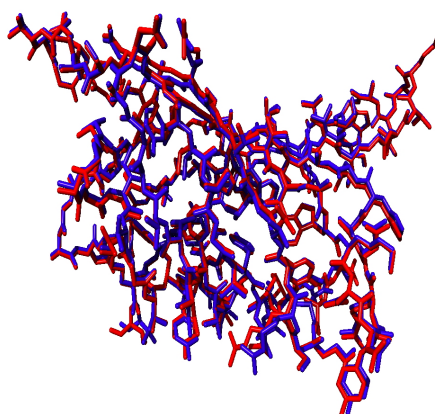


Figure 2.5 - Résultat de deux minimisations superposées pour la molécule de streptavidine. En bleu, $E_{Tot} = 9,09.10^2$ kcal/mol, $RMSG=4,43.10^{-2}$ Å, en rouge, $E_{Tot} = 9,007.10^2$ kcal/mol, $RMSG=3,94.10^{-2}$ Å. La superposition engendre un RMS de 0,41 Å.

La figure 2.5 montre clairement que le résultat d'une minimisation n'est pas une solution unique. Pour les mêmes caractéristiques de simulation, deux conformations distantes de 0,41 Å sont obtenues. Ces résultats ont été calculés au bout de 90 itérations pour 10 secondes de simulation concernant 1000 atomes.

Les temps de convergence des processus de minimisation énoncés ci-dessus dépendent de la taille du système considéré. Au mieux, en considérant la molécule décrite en coordonnées internes et donc par $3N - 6$ paramètres, les algorithmes convergent en $3N - 6$ pas. Les ligands

peuvent contenir jusqu'à une centaine d'atomes, les protéines, des milliers. En considérant simplement la protéine GP120 du virus du sida contenant quelque 3000 atomes, le système converge donc, au mieux, en 9000 itérations. A ce temps, incompressible, s'ajoute celui de l'évaluation de l'énergie interatomique. Les énergies de van der Waals sont calculées pour tous les atomes séparés de plus de trois liaisons et proches d'un certain rayon d'action, les autres énergies se calculent de proche en proche. Le temps de calcul varie donc en fonction de la taille, des contraintes internes de la protéine et ne se fait pas en temps réel.

Plusieurs approximations sont donc à mener afin de rendre le système plus interactif. La fiabilité de la solution n'en est donc que relative mais dans l'état actuel des systèmes informatiques considérer une interactivité totale entre une protéine et un opérateur semble très délicat.

2.2.1 La méthode de minimisation

Une des premières approximations consiste à diminuer les temps de convergence des algorithmes lorsque l'utilisateur manipule le ligand dont la conformation est éloignée de la solution attendue. En d'autres termes, le ligand n'est pas dans le site actif et il n'est pas proche de son état minimisé. Au cours de la manipulation, les erreurs de convergence n'ont que peu d'intérêt. Une minimisation grossière est donc effectuée et lorsque l'opérateur estime être proche de la solution, il peut à sa guise laisser converger l'optimisation. Ainsi, les effets principaux comme les ouvertures des sites actifs de la protéine ou les mouvements des boucles sont conservés sans trop consommer les ressources du système. En revanche, une fois le site actif intégré, le système est modifiable afin d'en approfondir la géométrie. Un aperçu du principe est représenté sur la figure 2.6.

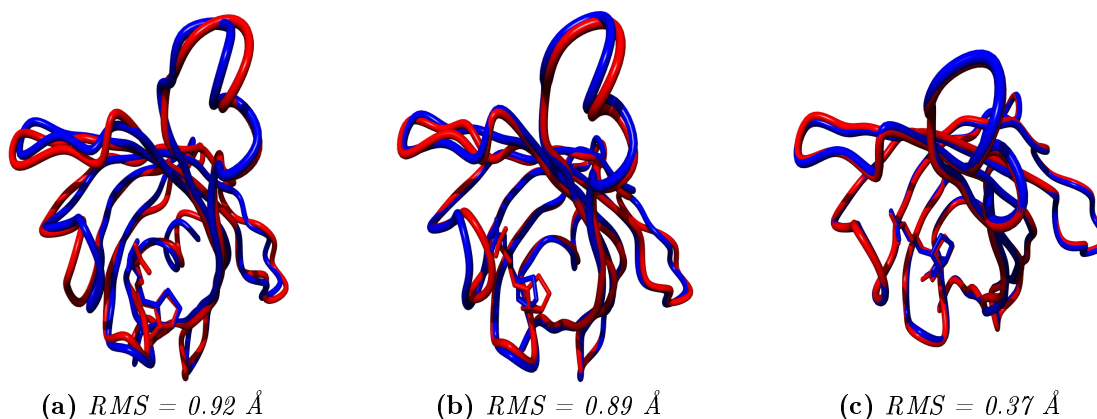


Figure 2.6 - Comparaison de la méthode adoptée, en bleu, pour la minimisation de la protéine streptavidine et de son ligand, la biotine, par rapport à une minimisation idéale, en rouge. Les critères d'arrêt sont énoncés au tableau 2.1.

L'objectif de la comparaison de la figure 2.6 est de montrer que la convergence de la minimisation peut être obtenue au cours de la manipulation même si elle n'est pas exécutée jusqu'à son terme. Le temps de calcul est ainsi limité. Sur chaque figure, la même molécule est simulée avec des critères de minimisation différents (tableau 2.1). La molécule rouge est optimisée, la bleue suit le schéma de simulation voulu, c'est-à-dire peu de convergence au cours de la manipulation, mais une optimisation finale affinée. Au cours de la manipulation du ligand, les

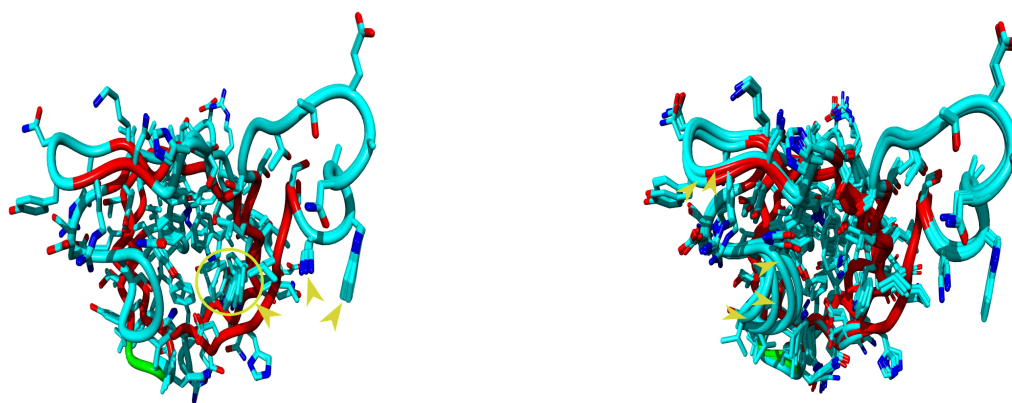
erreurs de déviation ne dépassent pas 1 Å par rapport à l'optimisation idéale au cours de la manipulation et l'optimisation finale est quasiment identique (la solution idéale ne pouvant être obtenue, figure 2.5 page 53) pour un gain de temps considérable (tableau 2.1).

Molécule rouge		Molécule bleue	
Condition	Critère	Condition	Critère
SD : Nombre d'itérations	100	SD : Nombre d'itérations	20
SD : Gradient test	1000	SD : Gradient test	1500
CG : Nombre d'itérations	100	CG : Nombre d'itérations	20
CG : Gradient test	100	CG : Gradient test	1000
TN : Nombre d'itérations	500	TN : Nombre d'itérations	20
TN : Gradient test	0.05	TN : Gradient test	0.05

Tableau 2.1 - Conditions d'arrêt des méthodes utilisées au cours de la minimisation de l'énergie des molécules représentées à la figure 2.6 page ci-contre.

En effet, la convergence optimale nécessite plus d'une centaine d'itérations alors qu'une vingtaine suffit dans le deuxième cas. Cette approche de modification des paramètres de convergence de la minimisation au cours de la manipulation trouve alors tout son sens. Il est à noter que la bonne convergence des algorithmes est due à l'utilisation des méthodes SD et CG qui approchent la solution optimale en peu d'itérations.

2.2.2 Contraintes internes



(a) Chaînes latérales mobiles, squelette fixe

(b) Chaînes latérales mobiles, squelette mobile

Figure 2.7 - L'espace des conformations à explorer dans le cas d'un squelette fixe est moins important que s'il était mobile. Les flèches indiquent les zones flexibles.

Lors de l'interaction d'un ligand avec une protéine, dans la majorité des cas, l'interaction se situe au niveau de la surface protéique. Toute la protéine, du moins sa chaîne carbonée, n'a pas à se mouvoir afin de laisser le ligand s'arrimer. La structure tridimensionnelle des protéines est principalement stabilisée grâce à des interactions entre les atomes de la chaîne principale. En général, la chaîne principale des protéines est beaucoup plus rigide que les chaînes latérales des acides aminés, en particulier ceux exposés au solvant.

Une simplification de la minimisation est alors effectuée en contraignant le mouvement de la protéine. Sa chaîne carbonée principale, le squelette, est considérée fixe. Au contraire, le mouvement des chaînes latérales est autorisé. Le ligand est donc flexible ainsi que les chaînes latérales de la protéine. La chaîne principale étant fixe, une partie de la protéine ne sera pas affectée par la perturbation géométrique induite par le ligand. L'effet des forces électrostatiques et de van der Waals étant limité (pour un changement de géométrie conséquent) à 10 \AA pour une simulation dans le vide, tous les atomes distants de 10 \AA du site actif sont considérés fixes. L'exploration conformationnelle est donc réduite, les variables internes diminuant considérablement (figure 2.7 page précédente). Le fait d'être fixe n'annihile pas la contribution des atomes à l'évaluation de l'énergie mais fait diminuer le nombre de paramètres de l'optimisation conformationnelle.

2.2.3 Conclusion

Afin d'optimiser le temps de calcul des processus successifs de minimisation, le système mis en place propose de limiter les effets des forces électrostatiques et de van der Waals à 10 \AA , de fixer les chaînes principales et d'effectuer les simulations dans le vide. Ces approximations réduisent considérablement le temps de calcul d'une conformation. L'algorithme de minimisation est lui-même restreint à un certain nombre d'itérations au cours de la manipulation.

Mais deux problèmes majeurs demeurent. Bien que ces approximations diminuent le temps de réponse du simulateur moléculaire, celui-ci n'est toujours pas dans les bornes acceptables du retour haptique (1 milliseconde à 50 millisecondes). De plus, toutes les interactions sont décrites en termes d'énergie, les problèmes du ressenti et de la stabilité de la commande restent donc entiers.

2.3 Description d'une manipulation

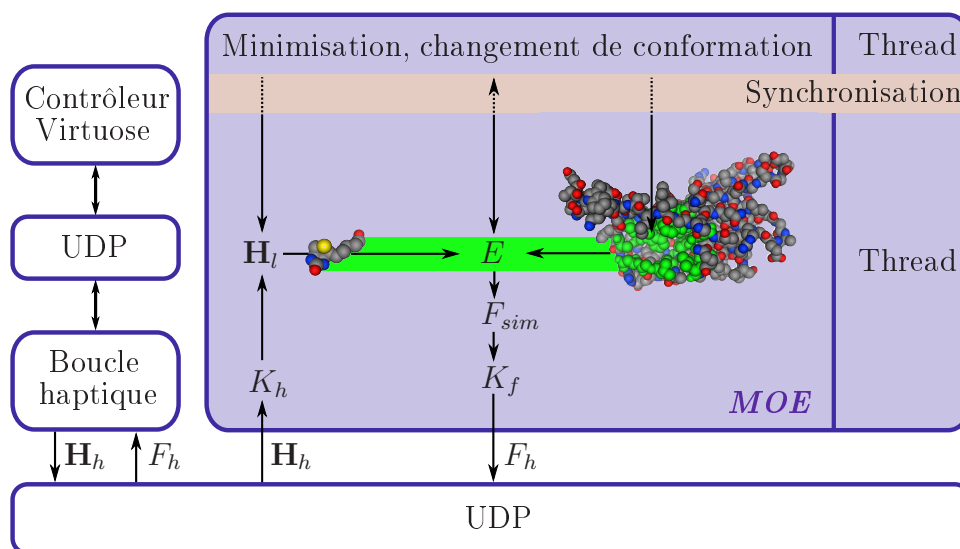


Figure 2.8 - Schéma de couplage.

La figure 2.8 page ci-contre représente schématiquement le fonctionnement d'une simulation. La boucle haptique acquiert les données de positions et d'orientations du *Virtuose* regroupées dans une matrice homogène $[4 \times 4]$, \mathbf{H}_h . Celle-ci les transmet par protocole UDP au simulateur moléculaire. Dans un premier temps, ces données sont traitées en vue d'un affichage simultané des nouvelles consignes en position.

L'énergie d'interaction ligand-protéine est ensuite évaluée, convertie en une force admissible et compréhensible par le *Virtuose* puis envoyée à l'interface haptique. Pendant cette opération, le minimiseur actualise les positions des atomes de la protéine et du ligand.

La simulation décrite à la figure 2.8 page précédente est présentée dans la suite de ce paragraphe dont le plan consiste à suivre en détail l'évolution des consignes, du *Virtuose* vers la simulation, puis de la simulation vers le *Virtuose*. L'évolution de cette consigne est représentée par un synopsis qui résume une opération de manipulation d'un ligand dans le site actif d'une molécule (figure 2.9).

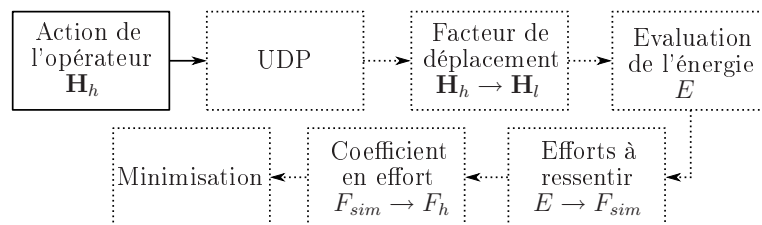


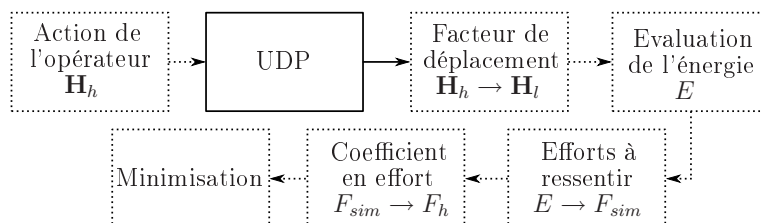
Figure 2.9 - Synopsis d'une opération de docking moléculaire.

Dans le plan du synopsis, la minimisation du système sera traitée au dernier paragraphe. Il faut se souvenir qu'elle agit en parallèle du processus itératif défini par les blocs précédents.

2.3.1 Action de l'opérateur

L'opérateur agit sur l'interface haptique (figure 2.9). La position et l'orientation de l'effecteur lues par les codeurs optiques du *Virtuose* sont alors actualisées. Un quaternion les modélise.

2.3.2 Protocole UDP



User Datagram Protocol (UDP) est un des principaux protocoles de télécommunication utilisés par Internet. Il fait partie de la couche transport du protocole TCP/IP. Son rôle est de transmettre très simplement des paquets entre deux entités, chacune étant définie par une adresse IP et un numéro de port. Contrairement au protocole TCP, il travaille en mode non

connecté : il n'existe pas de moyen de vérifier si tous les paquets envoyés sont arrivés à destination, ni dans quel ordre. Il est souvent décrit, pour cette raison, comme étant un protocole non fiable. En revanche, pour un paquet UDP donné, l'exactitude du contenu des données est assurée grâce à une somme de contrôle. Ce protocole est aussi utilisé lorsqu'il est nécessaire de transmettre soit des données très rapidement (et où la perte d'une partie de ces données n'a pas grande importance), soit des petites quantités de données (où la connexion TCP serait trop lourde).

Dans le cas présent, une interface haptique impose une communication rapide entre le contrôleur et la simulation physique. Le protocole UDP paraît alors être la meilleure solution.

Le quaternion envoyé par protocole UDP représentant la position et l'orientation de l'interface est normalisé en sortie de contrôleur. Afin de garantir que les données reçues soient effectivement un quaternion sans perte de données lors de la transmission, à la réception des paquets du côté simulation, ceux-ci sont normalisés. Une erreur de normalisation sur un paquet provoque son élimination, la simulation utilisant alors l'ancien quaternion.

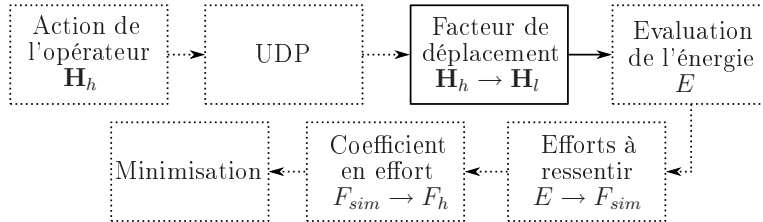
Afin d'être interprété par la simulation, le quaternion est traduit en une matrice homogène notée \mathbf{H}_h .

$$\mathbf{H}_h = \begin{bmatrix} \mathbf{R} & \mathbf{X}_e \\ 0 & 1 \end{bmatrix} \text{ avec } \begin{cases} \mathbf{R} & \text{Matrice de rotation de l'effecteur} \\ \mathbf{X}_e & \text{Position de l'effecteur} \end{cases} \quad (2.10)$$

En considérant un quaternion Q composé du scalaire a et ayant pour partie vectorielle le triplet b, c, d ($Q = a + ib + jc + kd$), le passage en une matrice de rotation s'effectue de la manière suivante [Murray 94] :

$$\mathbf{R} = \begin{pmatrix} 1 - 2c^2 - 2d^2 & 2bc - 2da & 2bd + 2ca \\ 2bc + 2da & 1 - 2b^2 - 2d^2 & 2cd - 2ba \\ 2bd - 2ca & 2cd + 2ba & 1 - 2b^2 - 2c^2 \end{pmatrix} \quad (2.11)$$

2.3.3 Facteur de déplacement



\mathbf{H}_h représente un déplacement macroscopique. Un coefficient responsable de l'adaptation de ces déplacements au monde microscopique doit être mis en place afin de faire correspondre le déplacement du *Virtuose* (macroscopique) au déplacement du ligand (microscopique). \mathbf{H}_h est alors exprimée dans le repère de la molécule puis multipliée par un coefficient de déplacement noté \mathbf{K}_H .

\mathbf{K}_H fait donc correspondre les translations et les rotations dans les grandeurs caractérisant les deux mondes. Cependant, si les translations doivent nécessairement être adaptées par un coefficient, les rotations, elles, sont identiques entre les deux mondes. En effet, la faible course

de la rotation de la poignée du *Virtuose* ne nécessite pas d'adaptation du monde macroscopique vers le monde microscopique. La matrice de rotation pouvant ainsi caractériser le changement d'échelle est fixée à l'identité.

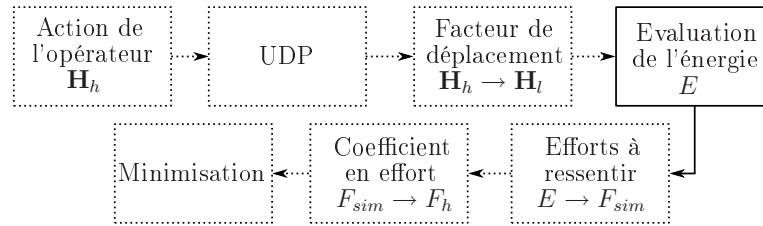
\mathbf{K}_D est la matrice des coefficients de translation pour chaque axe. Pour que l'utilisateur ne soit pas perturbé par une différence de déplacement selon les axes, tous les coefficients de \mathbf{K}_D sont identiques et égaux à K_h .

Le résultat, noté \mathbf{H}_l , représente le déplacement microscopique du ligand associé à l'effecteur.

$$\mathbf{H}_l = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{K}_D \\ \mathbf{0} & 1 \end{bmatrix}}_{\mathbf{K}_H} \mathbf{H}_h \quad (2.12)$$

Il représente le rapport entre la course maximum admissible par l'interface et le déplacement maximal désiré dans le repère microscopique pour un coefficient en rotation identitaire.

2.3.4 Evaluation de l'énergie



Une modification des positions des atomes du ligand ou de la molécule - dans le cas présent, la position du ligand est modifiée en fonction des déplacements de l'interface haptique - entraîne une modification du champ d'énergie modélisé entre le ligand et la protéine. En effet, toutes les composantes de l'énergie totale dépendent d'une distance d'interaction. Non seulement l'énergie d'interaction varie mais l'énergie interne à la molécule est aussi modifiée en conséquence. Une évaluation des différentes énergies est alors effectuée.

Plusieurs régions sont à considérer lors d'une interaction ligand-protéine (figure 2.10 page suivante). La première concerne l'énergie totale qui est la somme des régions A, B et C :

$$E_T = E_A + E_B + E_C \quad (2.13)$$

La région A correspond à la protéine, la région B au ligand et la région C à l'interaction entre ces deux entités. La région B se trouve donc être à l'intérieur du site actif de la protéine. Les contacts ne peuvent avoir lieu à cause des forces d'interaction qui modifieraient soit la géométrie de la protéine, soit celle du ligand. Dès lors, l'énergie d'interaction s'obtient en différenciant les énergies des régions A et B de l'énergie totale :

$$E_{A \cap B} = E_T - E_A - E_B \quad (2.14)$$

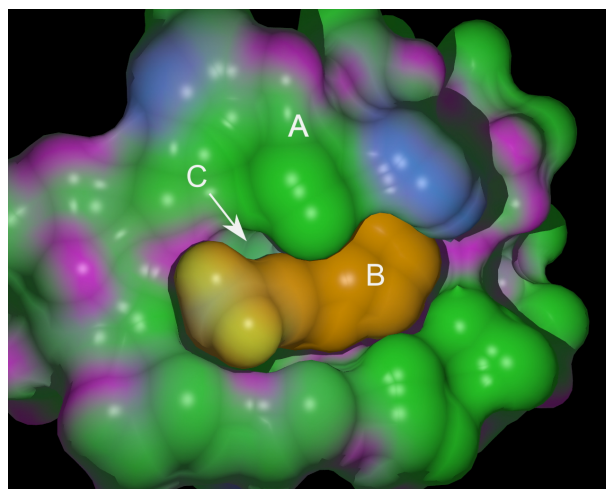
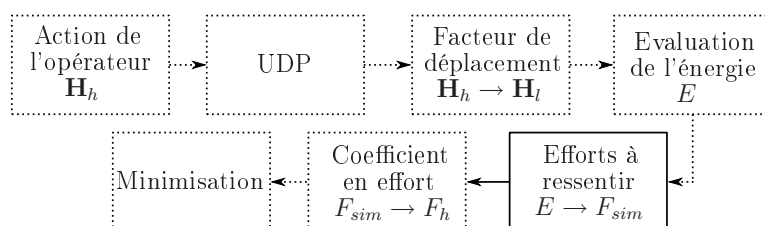


Figure 2.10 - L'évaluation de l'énergie d'interaction fait intervenir trois régions distinctes.

Un système en interaction avec un autre nécessite alors, pour calculer son énergie, trois évaluations préalables des énergies des sous-systèmes. La conséquence immédiate est donc qu'un simple calcul d'interaction peut prendre rapidement beaucoup de temps d'autant plus que le système est grand (le nombre d'atomes est élevé). Ce temps peut être modifié selon la finesse du calcul et la complexité du champ de force.

Le choix d'un tel procédé nécessaire à l'obtention de l'énergie peut surprendre. En effet, n'aurait-il pas été préférable de ne considérer qu'une distance d'interaction entre deux systèmes, comprenant des régions d'interactions internes aux deux molécules? La question est légitime et trouve sa réponse dans le temps de calcul de l'énergie. Considérer des régions d'interaction nécessite de faire un calcul de détection de proximité pour chaque atome considéré dans un champ d'interaction. Il faut calculer l'énergie d'interaction par rapport aux atomes distants. Cette contrainte implique, avant chaque évaluation de l'énergie, un calcul de proximité. Dans le cadre d'un projet dans lequel intervient une interface haptique, et pour ce type de détection, le temps de réponse de cette solution serait trop important. De plus, l'énergie de chaque entité doit être calculée afin de pouvoir minimiser le champ de force. L'interaction ainsi calculée n'impose donc pas de calcul ou d'évaluation supplémentaire.

2.3.5 Efforts à ressentir



Un des problèmes majeurs posé par l'utilisation d'un logiciel existant spécialisé dans la modélisation moléculaire est que toutes les interactions sont décrites en termes d'énergie. Cette propriété est fondamentale pour les utilisateurs de cette classe de logiciels dans les domaines comme la chimie ou la biologie.

Par le biais d'une interface à retour d'effort, une énergie ne peut être ressentie. Sa conversion en effort doit donc être envisagée. Le travail W des efforts exercés sur le ligand s'exprime en fonction de l'énergie d'interaction E :

$$\delta W = -\delta E \quad (2.15)$$

Le travail des efforts exercés sur le ligand s'exprime en fonction de la force \vec{F} extérieure qui lui est appliquée sur un déplacement $\vec{\delta l}$ ainsi que du moment résultant \vec{M} par rapport à la rotation $\vec{\delta \theta}$ de l'interface :

$$\delta W = \vec{F} \vec{\delta l} + \vec{M} \vec{\delta \theta} \quad (2.16)$$

Ainsi la force à ressentir s'exprime par rapport à la variation de position de l'interface haptique (équation 2.17).

$$\vec{F}_{\text{Haptique}} = -\vec{\text{grad}}(E) \quad (2.17)$$

En considérant un temps échantillonné et une trajectoire représentée par un axe quelconque \vec{x}_k , la force ainsi calculée vaut :

$$\vec{F}_{\text{Haptique}} \cdot \vec{x}_k = \frac{E_i - E_{i-1}}{x_{k_i} - x_{k_{i-1}}} \quad (2.18)$$

E_i et x_{k_i} représentent respectivement l'énergie évaluée à l'itération i et le déplacement (selon une direction \vec{x}_k de l'espace) de l'interface haptique - exprimé dans le repère de la molécule après application du facteur de déplacement - à l'itération i .

Les moments à ressentir sont calculés en considérant, cette fois, une variation d'angle.

$$\vec{M}_{\text{Haptique}} \cdot \vec{x}_k = \frac{E_i - E_{i-1}}{\theta_{x_{k_i}} - \theta_{x_{k_{i-1}}}} \quad (2.19)$$

θ_i représentant l'orientation de l'interface haptique à l'instant i .

Le torseur à ressentir est alors composé des forces et des moments calculés par les équations 2.17 et 2.19.

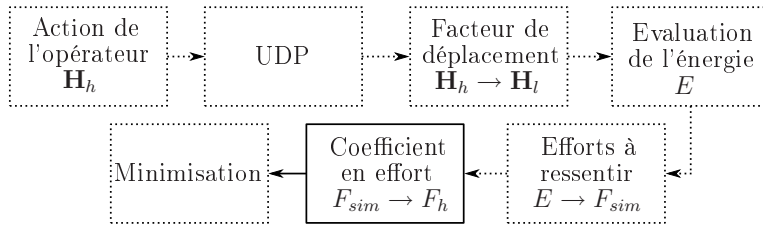
Cependant, une première observation quant à la validité d'une telle approche peut être conduite. Dériver l'énergie suppose d'obtenir ses composantes selon les trois directions de l'espace. Le champ de force intégrant les distances d'interaction dans sa formulation, calculer les énergies selon les trois directions est par conséquent possible. Cependant, le choix d'utiliser pleinement le logiciel MOE comme simulateur moléculaire, impose de n'utiliser que le résultat fourni par celui-ci. L'énergie calculée n'est donc pas décomposable selon les directions de déplacement de l'interface, en translation et en orientation. Utiliser les formulations ci-dessus

définies (équations 2.17, 2.19) revient à ne pas fournir la force exacte, engendrée par le champ de force, pour une position de ligand et de protéine donnée.

Un autre inconvénient de cette méthode réside dans les instabilités qu'elle engendre. En effet, cette approche n'est valable que si la variation de position de l'interface haptique est constamment définie. Or lorsque l'interface est fixe, la variation de position est nulle. Les équations 2.17 et 2.19 ne sont alors pas définies.

Lorsque l'interface haptique est détectée immobile, une première approximation consiste à appliquer les efforts du pas de temps précédent. Cette solution peut être envisagée en considérant que le champ d'énergie du changement de conformation ne présente pas d'instabilité ou de singularité. Le champ d'énergie est défini quelle que soit la position des atomes au sein de la molécule : cette approche est donc possible.

2.3.6 Coefficient en effort



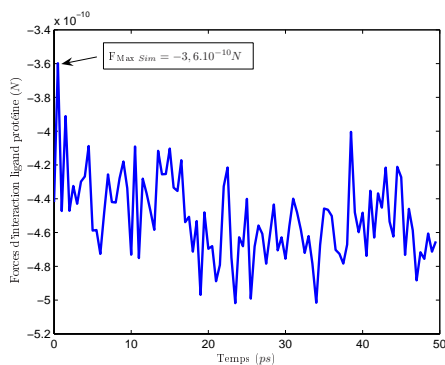
Les efforts calculés sont exprimés dans l'espace moléculaire. Afin de pouvoir les ressentir, un facteur d'effort K_f , responsable de l'adaptation des efforts micro/macro monde, est appliqué. Celui-ci est considéré isotrope pour les forces et isotrope pour les moments. Ce coefficient met en relation les efforts maximaux admissibles par l'interface haptique et les efforts maximaux de la simulation. Celui-ci reste à déterminer et dépend des déplacements imposés par l'utilisateur.

Le résultat est envoyé par protocole UDP à l'interface afin que l'utilisateur puisse ressentir la conséquence de son action. Cette relation s'écrit :

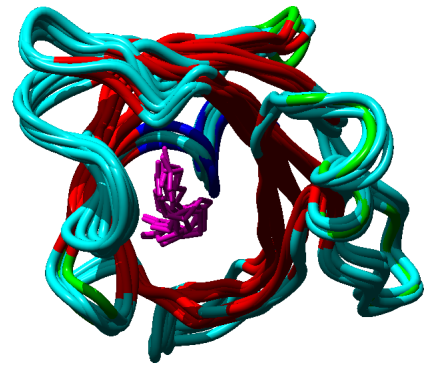
$$K_f = \frac{F_h}{F_{sim}} = \frac{F_{Max\ Virtuose}}{F_{Max\ Sim}} \quad (2.20)$$

où F_h est la force/couple à ressentir, F_{sim} est la force/moment calculée, à partir de l'énergie d'interaction dans le repère de la molécule. $F_{Max\ Virtuose}$ est la force/couple maximum admissible par l'interface haptique et est déterminée par ses caractéristiques, et $F_{Max\ Sim}$ la force maximum atteinte au cours de la manipulation. Le couple maximum est déterminé à partir de la force maximum en considérant qu'elle est appliquée à une extrémité du ligand.

Détermination de $F_{Max\ Sim}$. La force maximum estimée au cours d'une manipulation peut être précalculée soit manuellement par une succession de minimisations au cours d'une manipulation, soit par dynamique moléculaire. Dans ce cas, la force d'interaction moyennée au cours d'une simulation de quelques picosecondes est plus fidèle à la réalité. En effet, la dynamique moléculaire explore la surface d'énergie conformationnelle du ligand et de la protéine. Les conformations du ligand ainsi que son énergie sont estimées (figure 2.11 page suivante).



(a) Valeurs moyennes des forces d'interaction



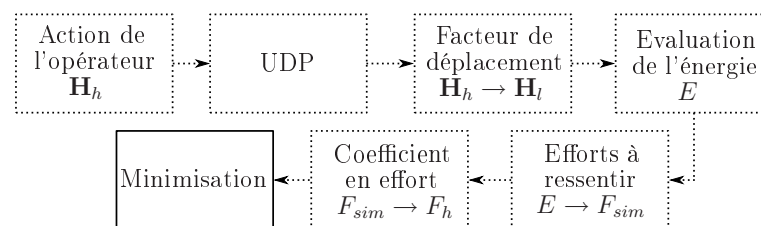
(b) Changements de conformation observés

Figure 2.11 - Expérience de dynamique moléculaire effectuée sur la molécule streptavidine et son ligand (violet), la biotine. Au cours de l'exploration conformationnelle de la protéine, les énergies d'interaction sont calculées, les forces engendrées sont estimées.

Cette expérience reproduit fidèlement le comportement que peut avoir le ligand au sein du site actif de la protéine car toute la surface d'énergie est explorée, les conformations possibles du ligand au cours de la manipulation sont donc *a priori* déjà estimées. Selon la figure 2.11 et si les conditions de minimisation du champ de force sont respectées (les rayons de van der Waals de tous les atomes ne s'interpénètrent pas au cours de la manipulation), la force maximum estimée de la simulation interactive serait $-3,5976.10^{-10} N$. Il convient cependant, afin de ne pas saturer l'interface haptique, d'appliquer un coefficient de sécurité de l'ordre d'une puissance de dix.

2.3.7 Minimisation

A ce stade de la simulation, l'énergie d'interaction est évaluée et les efforts correspondants sont calculés et transmis à l'interface haptique par protocole UDP.



Durant cette phase de calcul des efforts basée sur l'énergie d'interaction, le processus de minimisation est activé. Si cette énergie est différente de l'évaluation précédente, les atomes de la molécule et du ligand doivent se mettre en mouvement afin de garantir un minimum de potentiel induit par cette variation. Le processus de minimisation du champ de force s'active afin d'assurer que le système composé des deux entités considérées se trouve dans un puits de potentiel. C'est certainement la phase la plus longue de la simulation. Chaque groupe d'atomes doit être repositionné en fonction du mouvement des atomes de son champ d'influence.

L'énergie d'interaction est donc calculée avant la minimisation de l'énergie. Le retour d'effort

s'effectue sur un système dont les positions des atomes ne sont pas encore actualisées. Ce postulat peut être admis car dans le cas d'une manipulation microscopique les déplacements sont comme considérés faibles. La variation de position des atomes du ligand et de la protéine est ainsi supposée minime. Le puits de potentiel dans lequel se trouve à chaque instant la molécule assure la validité du postulat. *A priori*, sauf mouvement brusque de l'opérateur, la molécule, de proche en proche, évolue lentement sur sa surface d'énergie conformationnelle. Il peut donc être admis qu'un déplacement suivant un autre n'a qu'un effet limité sur la variation de position des atomes du système.

Il est évident que ce postulat impose, comme condition initiale, un système dans son état minimisé. En effet, faire interagir deux molécules qui ne sont pas à l'avance minimisées aurait pour conséquence possible une grande variation de position des atomes, rendant incompréhensible un retour d'effort. Une deuxième condition concerne l'opérateur qui doit, au cours de la manipulation, imposer de faibles déplacements au ligand afin d'éviter un saut de barrière d'énergie.

Ce schéma de simulation (envoi d'une position - évaluation de l'énergie d'interaction - calcul de la force - envoi de la force - minimisation) est le fil directeur des travaux présentés ici.

2.4 Commande d'un ligand flexible

La commande en admittance est le schéma utilisé, car le simulateur moléculaire utilise et actualise des positions. Il est donc naturel de lui présenter comme information, à son port d'entrée, un changement de position plutôt qu'une force. La figure 2.1 page 49 présente le principe du couplage utilisé.

2.4.1 Repères de travail

Plusieurs environnements sont définis pour garantir, à tout instant, la cohérence entre les mouvements de l'utilisateur et la prise en compte par la simulation.

Trois changements de repères apparaissent (figure 2.12 page ci-contre) :

- Le passage du repère de base au repère de l'écran définit la position du socle du *Virtuose* par rapport à l'écran. Cette transformation est fixe au cours de la simulation.
- Le passage du repère de l'écran au repère caméra définit la position de la caméra par rapport à la scène virtuelle. Cette transformation peut être modifiée en dynamique. Lors du mouvement du repère caméra, les coordonnées des atomes dans $\mathbf{R}_{\text{Ecran}}$ sont fixes. Elles varient alors dans le repère caméra.
- Le passage du repère de la molécule au repère de la caméra. Cette transformation estime les coordonnées de la molécule dans le repère caméra.

En supposant le ligand rigide, lui imposer un mouvement revient à imposer un déplacement identique à chaque atome du système. La rotation peut être représentée par un centre fixe qui est le centre de masse du ligand et un bras de levier qui correspond à l'éloignement des atomes par rapport à ce centre. L'implémentation de ce schéma est décrit ci-dessous.

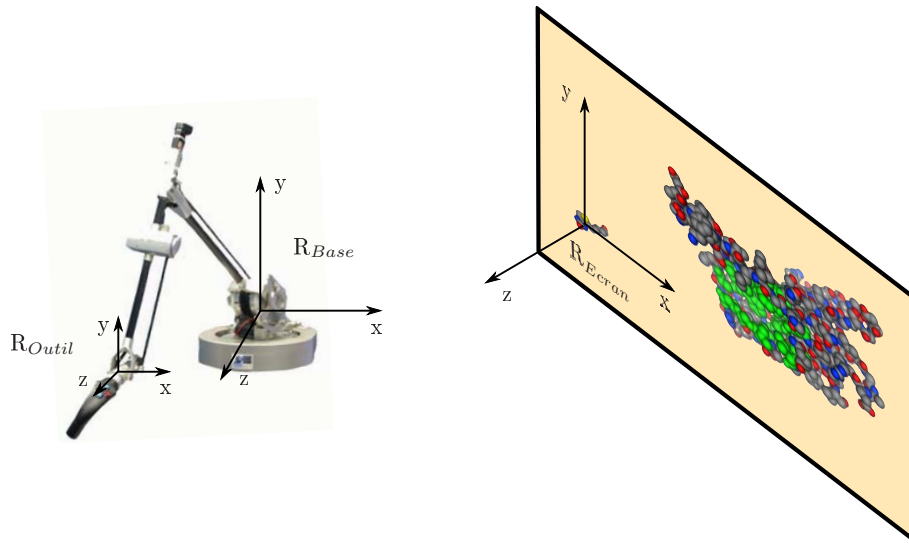


Figure 2.12 - Repères associés aux différentes entités du système.

2.4.1.1 Repères associés au *Virtuose*

${}^{\text{Base}}\mathbf{R}_{\text{Outil}}^{\text{Virtuose}}$ est la matrice de rotation du repère associé à l'effecteur du *Virtuose*, effectuant une rotation de l'outil par rapport à la base, exprimée dans le repère de base.

2.4.1.2 Simulateur moléculaire

Parallèlement à la définition des repères du *Virtuose*, il convient d'associer des repères au simulateur moléculaire. Pour un instant initial noté t_0 , le repère de la molécule est noté $\mathbf{R}_0^{\text{MOE}}$. Ce repère est identifié à l'instant initial au repère du *Virtuose*.

$$\mathbf{R}_0^{\text{MOE}} = {}^{\text{Base}}\mathbf{R}_{\text{Outil}}^{\text{Virtuose}}(t_0) \quad (2.21)$$

A l'instant initial, les positions des atomes notées X_0 valent :

$$\mathbf{X}_0 = \mathbf{X}^{\text{MOE}}(t_0) = \underbrace{\begin{bmatrix} X_{0_1}^t \\ \vdots \\ X_{0_N}^t \end{bmatrix}}_{\text{Positions}} \left. \vphantom{\begin{bmatrix} X_{0_1}^t \\ \vdots \\ X_{0_N}^t \end{bmatrix}} \right\} \text{Nombre d'atomes} \quad (2.22)$$

Le centre de masse \bar{X} du ligand est défini de la façon suivante :

$$\bar{X} = \sum \frac{1}{m_i} \mathbf{X}_i(t_0) \quad (2.23)$$

où les m_i représentent les masses associées aux atomes i .

A tout instant t , la matrice de rotation ${}^{\text{Base}}\mathbf{R}_{\text{Outil}}^{\text{Virtuose}}(t)$ est calculée à partir de l'information des codeurs optiques du *Virtuose*.

La position des atomes du ligand, exprimée dans le repère du *Virtuose*, est égale quel que soit t :

$$\mathbf{X}_{\text{Rotation}_i}^{\text{Ligand}}(t) = \bar{\mathbf{X}} + {}^{\text{Base}}\mathbf{R}_{\text{Outil}}^{\text{Virtuose}}(t) {}^{\text{Base}}\mathbf{R}_{\text{Outil}}^{\text{Virtuose}}(t_0) (\mathbf{X}_{i_0} - \bar{\mathbf{X}}) \quad (2.24)$$

Les nouvelles positions des atomes dans le repère écran ne peuvent être obtenues à partir de l'expression 2.24. Une dernière composition de mouvement est nécessaire. La transformation à obtenir définit les coordonnées de la base vers l'écran soit la matrice ${}^{\text{Base}}\mathbf{R}_{\text{Ecran}}$

Afin d'obtenir cette dernière transformation, une mise en correspondance est effectuée entre les mouvements du *Virtuose* et ceux observés à l'écran, en translation. De cette mise en correspondance, les matrices de passage de l'écran vers la base notée ${}^{\text{Base}}\mathbf{R}_{\text{Ecran}}$ et de la base vers l'écran notée ${}^{\text{Base}}\mathbf{R}_{\text{Ecran}}^t$ sont déduites :

$${}^{\text{Base}}\mathbf{R}_{\text{Ecran}} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

L'expression des coordonnées $\mathbf{X}(t)$ des atomes en fonction du mouvement du *Virtuose*, exprimée dans le repère écran, s'écrit alors :

$$\mathbf{X}_{\text{Rotation}_i}^{\text{Ligand}}(t) = \bar{\mathbf{X}} + {}^{\text{Base}}\mathbf{R}_{\text{Ecran}} {}^{\text{Base}}\mathbf{R}_{\text{Outil}}^{\text{Virtuose}}(t) {}^{\text{Base}}\mathbf{R}_{\text{Outil}}^{\text{Virtuose}}(t_0) {}^{\text{Ecran}}\mathbf{R}_{\text{Base}} (\mathbf{X}_{i_0} - \bar{\mathbf{X}}) \quad (2.25)$$

Le déplacement du ligand rigide représenté par l'équation 2.25 n'effectue que la rotation du ligand autour de son centre de masse. La translation fait apparaître le changement de repère entre la base et l'écran, ${}^{\text{Ecran}}\mathbf{R}_{\text{Base}}$. En conséquence, une translation du ligand s'écrit :

$$\mathbf{X}_{\text{Translation}_i}^{\text{Ligand}}(t) = \mathbf{X}_i^{\text{Ligand}}(t-1) + {}^{\text{Ecran}}\mathbf{R}_{\text{Base}} (X^{\text{Virtuose}}(t) - X^{\text{Virtuose}}(t-1)) \quad (2.26)$$

Le déplacement total du ligand, résultat de la composition du mouvement de translation et de rotation ainsi calculé, s'écrit :

$$\mathbf{X}_{\text{Total}}^{\text{Ligand}}(t) = \mathbf{X}_{\text{Translation}}^{\text{Ligand}}(t) + \mathbf{X}_{\text{Rotation}}^{\text{Ligand}}(t) \quad (2.27)$$

Lors de l'interaction du ligand avec la protéine, le site actif de celle-ci reste sous l'influence du processus de minimisation. L'énergie d'interaction est calculée entre les atomes du site actif et ceux de la molécule, mais seul le site actif est flexible.

Cette première étape de commande du ligand établit le schéma de simulation. Celui-ci consiste à détecter une variation de position du *Virtuose*, à envoyer cette variation *via* un protocole UDP à la simulation en vue de son traitement. L'énergie d'interaction est calculée et la force d'interaction déduite et envoyée par le même protocole au *Virtuose*. La minimisation du site actif est ensuite réalisée.

L'évaluation de l'énergie nécessaire au calcul de la force s'effectue avant l'actualisation des positions des atomes du site actif de la protéine. Cette énergie d'interaction représente l'ensemble des forces appliquées à chaque atome. Elle peut donc être considérée comme l'action du ligand rigide sur le site actif. La dérivation de l'énergie correspond donc à une force globale supposée appliquée au centre de masse du solide. Le système ainsi manipulé correspond à un déplacement du centre de masse du solide qui provoque une réaction en son centre de masse.

Une amélioration de ce système peut être réalisée. Plutôt que de considérer la manipulation du ligand par son centre de masse, l'atome le plus proche de celui-ci peut être considéré comme étant le point d'attache du solide. Cette approche a pour effet une meilleure compréhension visuelle du phénomène. En effet, imposer une rotation du ligand par rapport à son centre de masse a pour effet une rotation autour d'un centre invisible. Suivant les configurations que peut avoir le ligand, l'utilisateur peut être troublé par les effets provoqués. Une visualisation du centre de rotation ou de déplacement par rapport à un atome a pour effet une meilleure compréhension du déplacement.

2.4.2 Cas du ligand flexible

Le processus de minimisation associé à MOE calcule les nouvelles positions des atomes dans le repère de l'écran. A chaque pas de minimisation, un vecteur de l'ensemble des coordonnées du système est calculé. Ce vecteur $\mathbf{X}_m(\mathbf{X}, k)$ contient donc les informations de position des atomes du ligand ainsi que des atomes de la molécule pour une itération k et une position initiale \mathbf{X} . \mathbf{X} est la position des atomes au début de la minimisation. Ce vecteur s'écrit donc :

$$\mathbf{X}_m(\mathbf{X}, k) = [\mathbf{X}_m^{molécule}(k), \mathbf{X}_m^{ligand}(k)] \quad (2.28)$$

$\mathbf{X}_m(\mathbf{X}, k)$ dépend donc de la précision de la minimisation, k étant le nombre d'itérations du processus. C'est-à-dire que pour une position initiale donnée \mathbf{X} , les k vecteurs d'actualisation des positions sont renvoyés pour affichage. Tant que ces k vecteurs ne sont pas calculés, le processus de minimisation n'actualise pas son vecteur de position \mathbf{X} à minimiser. Le processus de minimisation s'interrompt si les critères d'arrêt sont satisfaits. Ces critères d'arrêt sont soit un gradient d'énergie faible, soit un nombre défini d'itérations. Afin de rendre le ligand flexible au cours de la manipulation, une composition des positions de consigne et de minimisation doit être effectuée pour le ligand, le récepteur n'étant pas sujet aux variations de positions de l'interface haptique.

Les positions des atomes du ligand correspondant à l'énergie minimisée sont, dans un premier temps, exprimées dans un repère local propre au ligand. Soit $\mathbf{X}_{m_{init}}^{ligand}(t)$, la position initiale des atomes soumis à la minimisation de l'énergie. $\mathbf{X}_h^{ligand}(t)$ est la position du ligand, après manipulation par l'interface haptique, et $\mathbf{X}_m^{ligand}(k, t)$ la position des atomes du ligand dans

son repère écran, à l'itération k du processus de minimisation. $\mathbf{X}_m^{ligand}(k, t)$ s'écrit donc, en reprenant l'expression de l'équation 2.28,

$$\mathbf{X}_m^{ligand}(\mathbf{X}_{m_{init}}^{ligand}(t), k)|_{\text{Ecran}} = [\mathbf{0}, \mathbf{X}_m^{ligand}(\mathbf{X}_{m_{init}}^{ligand}(t), k)] \quad (2.29)$$

La position du ligand dans son repère local s'exprime ainsi :

$$\mathbf{X}_m^{ligand}(\mathbf{X}_{m_{init}}^{ligand}(t), k)|_{\text{Local}} = \mathbf{X}_{m_{init}}^{ligand}(t)|_{\text{Ecran}} - \mathbf{X}_m^{ligand}(\mathbf{X}_{m_{init}}^{ligand}(t), k)|_{\text{Ecran}} \quad (2.30)$$

En considérant la manipulation du ligand par l'interface, la position du ligand $\mathbf{X}^{ligand}(t)$ s'écrit alors dans le repère écran :

$$\mathbf{X}^{ligand}(t)|_{\text{Ecran}} = \mathbf{X}_h^{ligand}(t)|_{\text{Ecran}} + \mathbf{X}_m^{ligand}(\mathbf{X}_{m_{init}}^{ligand}(t), k)|_{\text{Local}} \quad (2.31)$$

Le processus de minimisation n'est pas implémenté dans la partie logicielle responsable de l'affichage et de l'acquisition des données provenant de l'interface haptique. Ce choix s'explique par la différence de temps de calcul entre les deux entités. La différence de temps de calcul provoque un décalage entre l'acquisition des données et la réception du vecteur des positions issues de la minimisation. Lorsqu'un tel cas se présente, l'actualisation des positions d'affichage ne tient compte que de l'acquisition des données. Ainsi, la nouvelle position à afficher correspond à l'équation 2.31 :

$$\mathbf{X}^{ligand}(t)|_{\text{Ecran}} = \mathbf{X}_h^{ligand}(t)|_{\text{Ecran}} \quad (2.32)$$

Ce choix de ne pas bloquer le processus d'affichage en attendant les données du minimiseur évite ainsi une interruption de l'affichage et du calcul des forces qui se fait dans le même thread.

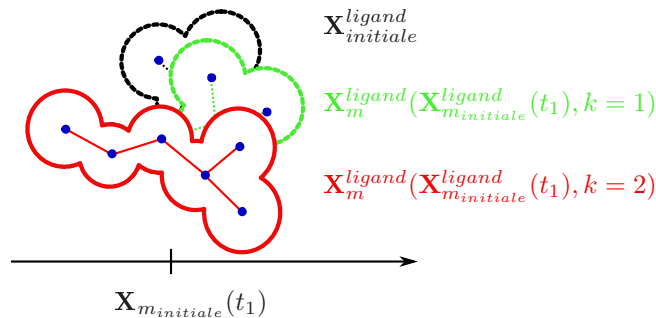


Figure 2.13 - Changement de conformation du ligand pour une consigne en position nulle.

La position des atomes, \mathbf{X}_m^{ligand} , pour un ligand dont l'énergie est en cours de minimisation dépend de la position initiale utilisée par le minimiseur, du nombre d'itérations du processus ou d'un gradient d'énergie minimum. Un premier effet non négligeable peut donc être facteur d'instabilité. Si le temps de calcul de la minimisation est élevé ou que le nombre d'itérations nécessaires pour l'achever est important, l'opérateur peut effectuer de grands déplacements

avec le ligand, mais avec pour positions en cours d'actualisation, celles ayant pour condition initiale des positions éloignées des actuelles. Cette imprécision nécessite un apprentissage de l'opérateur, ou l'ajout de viscosité dans la commande, au détriment d'un bon ressenti des forces.

La figure 2.13 page précédente représente schématiquement le processus de minimisation. Pour une consigne nulle, le ligand ne dépend que des positions actualisées par le minimiseur. Celles-ci sont fonction de la position initiale du ligand, et sur la figure, du nombre d'itérations fixé à deux. En deux itérations, la position idéale est atteinte.

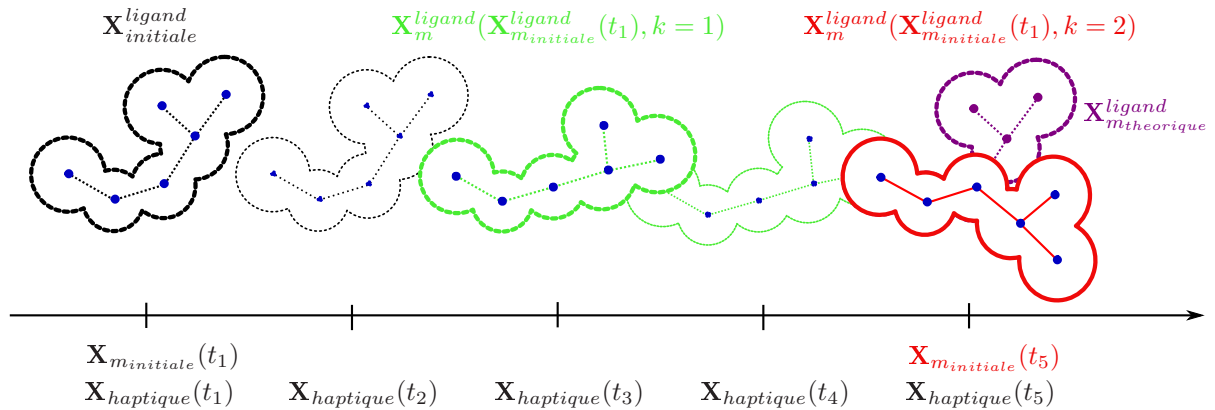


Figure 2.14 - *Changement de conformation du ligand au cours de sa manipulation.*

La figure 2.14 schématise l'évolution d'un ligand au cours d'une manipulation.

A l'instant initial t_1 , pour une consigne en position $\mathbf{X}_{haptique}$, le ligand est dans un état minimisé correspondant à un ensemble de positions atomiques $\mathbf{X}_{m_{initiale}}^{ligand}$.

Entre les instants t_1 et t_2 , le ligand doit être minimisé, le processus de minimisation est alors activé. Il prend comme condition initiale, la position des atomes du ligand, $\mathbf{X}_{m_{initiale}}^{ligand}(t_1)$, et comme critère d'arrêt, deux itérations. Le processus de minimisation étant indépendant de l'acquisition des positions de l'effecteur du *Virtuose*, à l'instant t_2 celle-ci est acquise mais le minimiseur n'a pas procédé à l'actualisation des positions du ligand. L'ensemble à afficher s'affranchit donc de la réponse du minimiseur et la position du ligand correspond au déplacement voulu de l'interface haptique. C'est un déplacement de corps rigide.

A l'instant t_3 , MOE acquiert la position de l'effecteur du *Virtuose* et les positions des atomes correspondants à la minimisation du ligand, renvoyées par le minimiseur. Les positions ainsi affichées tiennent compte, cette fois, de la consigne haptique et du retour du minimiseur.

A l'instant t_4 , comme pour l'instant t_2 , les seules positions à afficher sont celles correspondant au déplacement de l'interface.

Enfin, à l'instant t_5 , les positions affichées sont une différence des positions de consigne et de celles calculées par le minimiseur. Le critère d'arrêt est atteint, le minimiseur s'arrête et prend, comme condition initiale pour le prochain processus, cette dernière position, si l'énergie doit à nouveau être minimisée.

A l'instant t_5 , la position finale du ligand est donc un compromis entre le déplacement de l'interface et la minimisation de l'instant t_1 . A cet instant et pour les forces agissant sur le ligand la conformation d'énergie minimum idéale aurait été $\mathbf{X}_{m_{theorique}}^{ligand}$. Le résultat de la prochaine minimisation, sans déplacement de l'interface, sera exactement cette conformation. Si l'interface se déplace, la prochaine conformation du ligand correspondra à $\mathbf{X}_{m_{theorique}}^{ligand}$, mais appliquée au déplacement de l'interface.

Par cette méthode « asynchrone », la téléopération du ligand est fluide. L'opérateur doit cependant optimiser le minimiseur afin d'obtenir le plus rapidement possible, et le plus précisément possible, les positions des atomes du ligand dans sa conformation minimisée.

L'actualisation des positions du site actif procède de la même méthode. En revanche, le site actif n'étant pas lié à l'interface, les positions des atomes restent dans le repère global de l'écran.

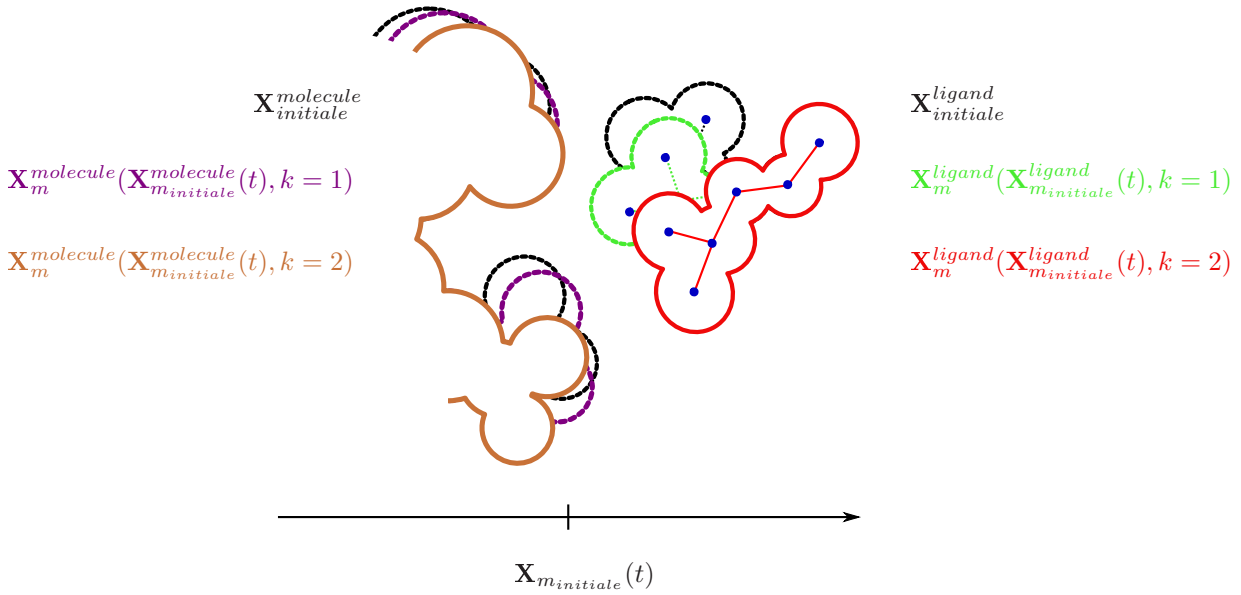


Figure 2.15 - Récepteur flexible au cours de la manipulation.

Soit $\mathbf{X}_{m_{init}}^{molecule}(t)$, la position initiale des atomes de la molécule dont l'énergie est à minimiser. $\mathbf{X}_m^{molecule}(k, t)$ est la position des atomes de la conformation minimisée dans son repère écran à l'itération k du processus de minimisation (figure 2.15). $\mathbf{X}_m^{molecule}(k, t)$ s'écrit donc, en reprenant l'expression de l'équation 2.28 :

$$\mathbf{X}_m^{molecule}(\mathbf{X}_{m_{init}}^{molecule}(t), k)|_{\text{Ecran}} = [\mathbf{X}_m^{molecule}(\mathbf{X}_{m_{init}}^{molecule}(t), k), \mathbf{0}] \quad (2.33)$$

Au cours de la manipulation du ligand par l'interface, la position $\mathbf{X}_{m_{init}}^{molecule}(t)$ des atomes de la molécule, dans l'écran, soumis à la variation de l'énergie, s'écrit :

$$\mathbf{X}_{m_{init}}^{molecule}(t)|_{\text{Ecran}} = \mathbf{X}_m^{molecule}(\mathbf{X}_{m_{init}}^{molecule}(t), k)|_{\text{Ecran}} \quad (2.34)$$

Cette implémentation, indépendante, présente l'avantage de la rapidité de calcul. La boucle lente, constituée de la minimisation du système, n'agit pas sur la fréquence d'exécution de la boucle rapide qui acquiert la position du *Virtuose*, évalue l'énergie d'interaction et calcule la force à appliquer. L'inconvénient réside dans le fait qu'il est possible, après un déplacement de l'interface, pour une consigne en position nulle, que les positions des atomes continuent d'être actualisées, perturbant ainsi la manipulation.

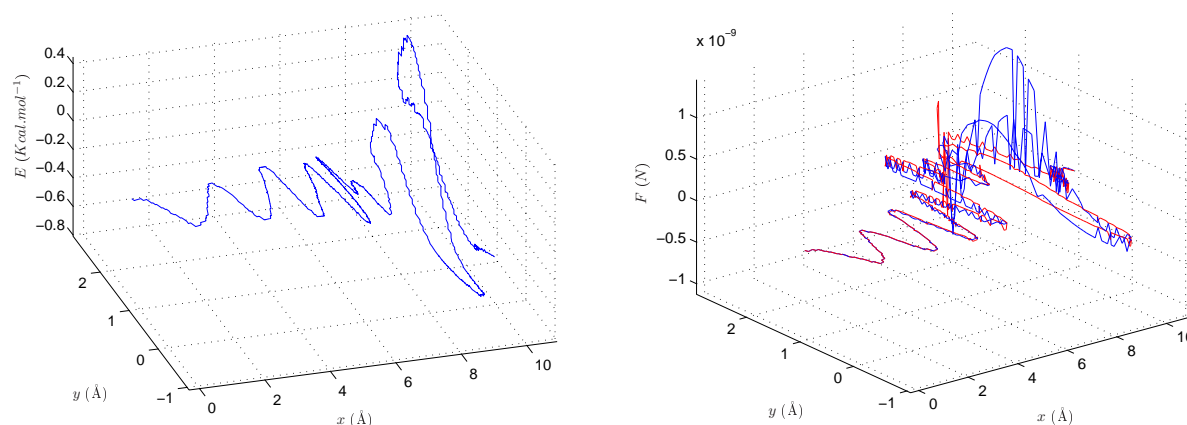
3 Retour d'effort

3.1 Influence du facteur d'effort

3.1.1 Système rigide

Ce paragraphe décrit les premières expérimentations effectuées afin d'appréhender le retour d'effort à l'échelle moléculaire. Une première étude consiste à examiner les forces d'interaction ligand-protéine au cours du *docking*, afin de pouvoir déterminer la valeur des gains mis en jeu notamment le facteur de déplacement, le coefficient en effort, les raideurs et amortissements de l'asservissement.

Ce paragraphe s'attache à montrer l'importance de ces facteurs pour la stabilité du retour d'effort.



(a) Énergie calculée au cours de la manipulation de la biotine dans le site actif de la streptavidine.

(b) Forces calculées au cours de la manipulation de la biotine dans le site actif de la streptavidine. En bleu, est représentée la composante de la force selon x , et en rouge, selon y .

Figure 2.16 - Manipulation de la biotine dans le site actif de la streptavidine à z constant. Au cours de la manipulation, la minimisation n'est pas activée.

La figure 2.16 illustre toute la complexité du ressenti des forces à l'échelle des molécules. Lors de la simulation du *docking* de la biotine au sein du site actif de la streptavidine, l'énergie semble ne varier que très peu (figure 2.16(a)). Lorsque le même parcours est effectué en évaluant les forces, les fortes variations de celles-ci apparaissent clairement.

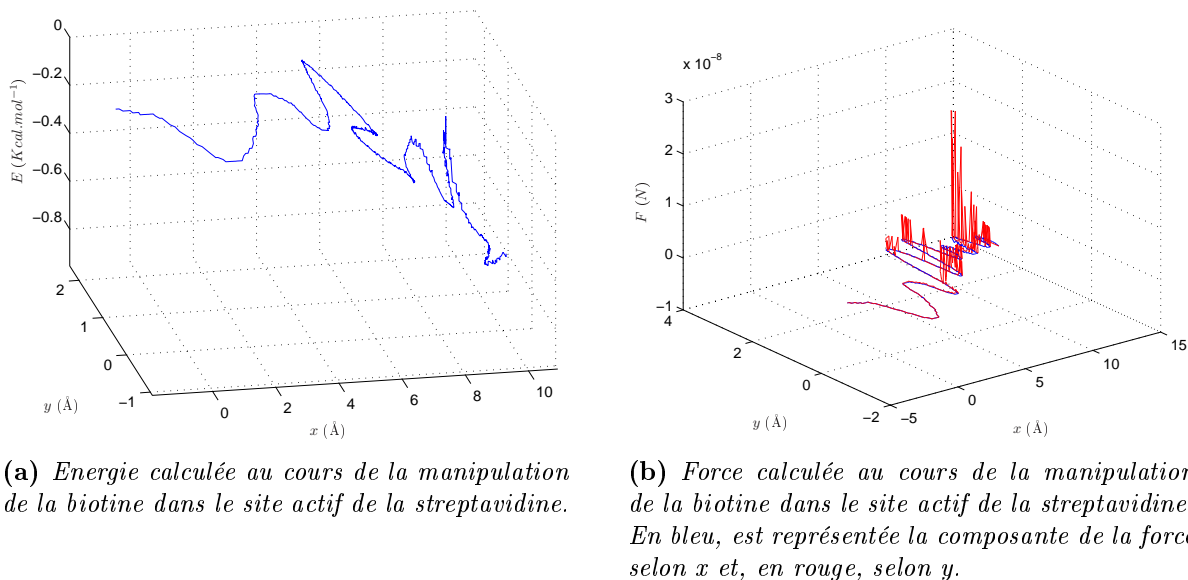
D'après le paragraphe 2.3.6, la force maximum atteinte au cours de cette manipulation dépasse les prévisions. De plus, le profil de force semble difficile à appréhender car il est très discontinu.

3.1.2 Système flexible

Les forces ressenties pour un système rigide semblent discontinues. Qu'en est-il d'un système dont les positions des atomes sont soumises à une minimisation de l'énergie ? Ce paragraphe

montre que la forte variation des forces d'interaction, présente pour un système rigide, est amplifiée lorsqu'il s'agit de manipuler un ligand flexible dans un site actif flexible.

En effet, au cours de la manipulation, les positions des atomes varient. Le temps de calcul nécessaire à l'évaluation de l'énergie d'interaction, et au calcul de la force, fait que certains atomes peuvent entrer en collision, provoquant dès lors un effort infini.



(a) Energie calculée au cours de la manipulation de la biotine dans le site actif de la streptavidine.

(b) Force calculée au cours de la manipulation de la biotine dans le site actif de la streptavidine. En bleu, est représentée la composante de la force selon x et, en rouge, selon y .

Figure 2.17 - Manipulation de la biotine dans le site actif de la streptavidine à \vec{z} constant. Au cours de la manipulation, la minimisation est activée.

La figure 2.17 montre la variation d'énergie et le profil de force obtenus lors de la manipulation de la biotine dans le site actif de la streptavidine alors qu'un processus de minimisation de l'énergie totale est activé.

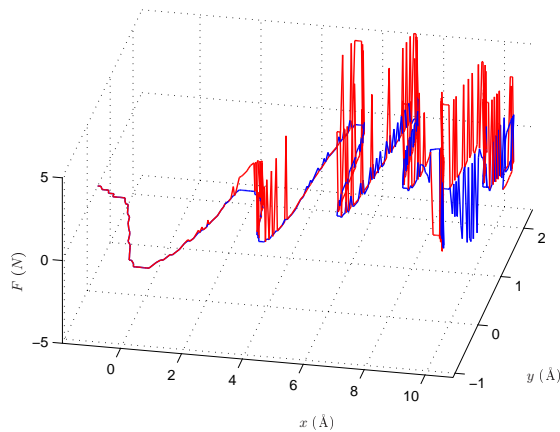
Le profil de force dérivé de l'énergie apparaît très instable. Alors que les forces exercées sur l'axe \vec{y} du *Virtuose* dépassent l'ordre de grandeur initialement prévu, les forces sur l'axe \vec{x} restent de l'ordre de 10^{-10} N. Le coefficient en effort, initialement calculé pour une force maximum de $-3,5 \cdot 10^{-10}$ N, rend délicat le ressenti de telles forces, dont les ordres de grandeurs varient autant. En effet, prévoir un coefficient en effort pour ressentir des efforts de l'ordre de 10^{-8} N fait que l'ensemble des forces inférieures à 10^{-10} N sont difficilement interprétables. A l'inverse, prévoir un retour d'effort avec un coefficient en effort maximal de l'ordre de 10^{-10} N saturerait les moteurs de l'interface si un effort supérieur était calculé.

Afin de garantir la sécurité de l'interface, une force de coupure est introduite. Les forces supérieures à celles désirées seront donc ressenties comme un contact de raideur infinie simulant un mur. Ce procédé trouve une explication physique simple. Si, lors de la dynamique moléculaire, les forces d'interaction possèdent un ordre de grandeur de 10^{-10} N, alors les atomes mis en mouvement au cours de la simulation ont comme contrainte interne de ne pas pénétrer les rayons de van der Waals. La position du ligand ainsi simulée reste théoriquement celle qui minimise cette énergie. Lors de la manipulation du ligand et sans contrainte sur le déplacement des molécules, certains rayons de van der Waals peuvent s'interpénétrer et produire un effort infini. Afin de garantir cette séparation, plutôt que de renvoyer un effort infini qui aurait pour effet de déstabiliser l'interface, une raideur infinie est renvoyée à l'interface.

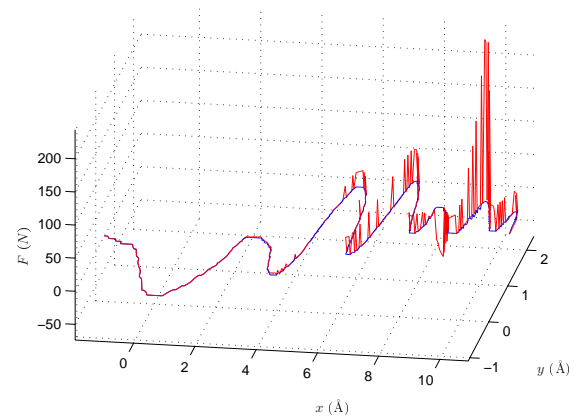
La figure 2.18 montre le profil de force ressenti lors d'une manipulation de *docking* ligand-protéine. La force maximum admissible par le *Virtuose* est 5 N. La force d'interaction biotine-streptavidine maximum estimée avant la manipulation par dynamique moléculaire vaut 5.10^{-10} N. Le coefficient en effort alors appliqué aux forces calculées vaut :

$$K_f = \frac{5}{5.10^{-10}} = 1.10^{10} \quad (2.35)$$

La figure 2.18(b) montre le profil de force théoriquement renvoyé à l'interface haptique pour un tel coefficient. Sans l'application d'une limite en effort, le *Virtuose* aurait à appliquer des forces avoisinant les 200 N. Cette limitation est donc nécessaire mais altère le ressenti du *docking* (figure 2.18(a)).



(a) Force renvoyées à l'interface haptique au cours de la manipulation de la biotine dans le site actif de la streptavidine. En bleu, est représentée la composante de la force selon x et, en rouge, selon y .



(b) Simulation des forces potentiellement envoyées à l'interface haptique lors d'une manipulation de *docking* biotine-streptavidine. En bleu, est représentée la composante de la force selon x et, en rouge, selon y .

Figure 2.18 - Manipulation de la biotine dans le site actif de la streptavidine à z constant. Au cours de la manipulation, la minimisation est activée. (a) Avec limitation des efforts à 5 N, (b) Sans limitation (simulation).

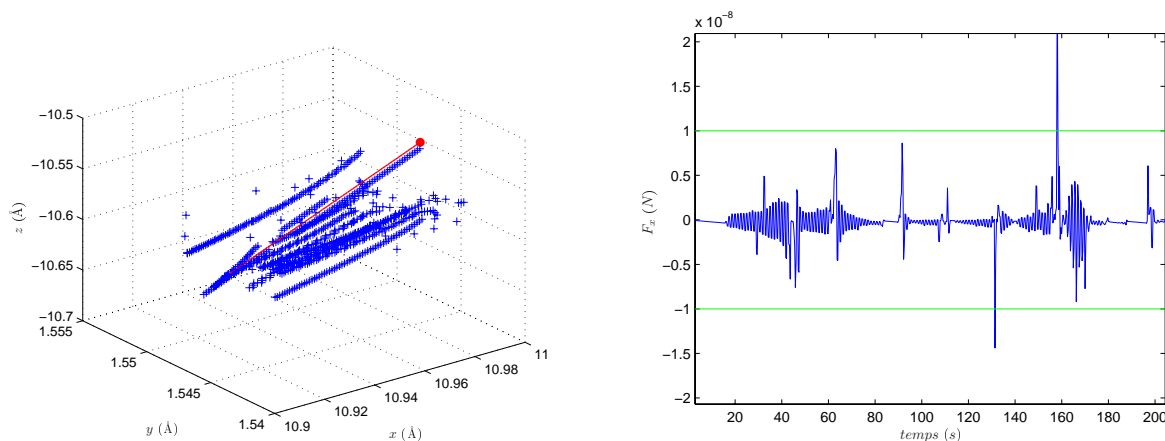
Le problème majeur du ressenti n'est pas tant le blocage de l'interface lorsqu'une limite en effort est détectée que l'instabilité du profil. En effet, d'après la figure 2.18(a), une fois la limite en effort atteinte, les forces tendent vers zéro. Cette grande variation de force engendre une instabilité de l'interface. Après l'application d'un effort important, le ligand peut être attiré vers une zone répulsive qui le renverrait alors en collision vers la zone opposée. Cette oscillation ne pourrait pas être maîtrisée par l'opérateur.

3.2 Influence du facteur de déplacement

3.2.1 Facteur de déplacement faible

La figure 2.19 page suivante montre le résultat obtenu lors de la manipulation d'un ligand dans le site actif d'une protéine réceptrice avec un faible coefficient de déplacement. Un débat-

tement du *Virtuose* selon l'axe \vec{x} correspond ainsi à un déplacement de 0.001 \AA selon le même axe dans le site actif. Le processus de minimisation du système est activé et la puissance du *Virtuose* coupée (le retour d'effort n'est pas assuré).



(a) Evolution du centre de masse au cours de la minimisation, + Positions du centre de masse, • position initiale, — consigne en position.

(b) Forces calculées au cours de la manipulation sur l'axe x , — élément de comparaison du profil avec celui de la figure 2.20(b) page suivante.

Figure 2.19 - Manipulation de la biotine dans le site actif de la streptavidine suivant une consigne en position linéaire.

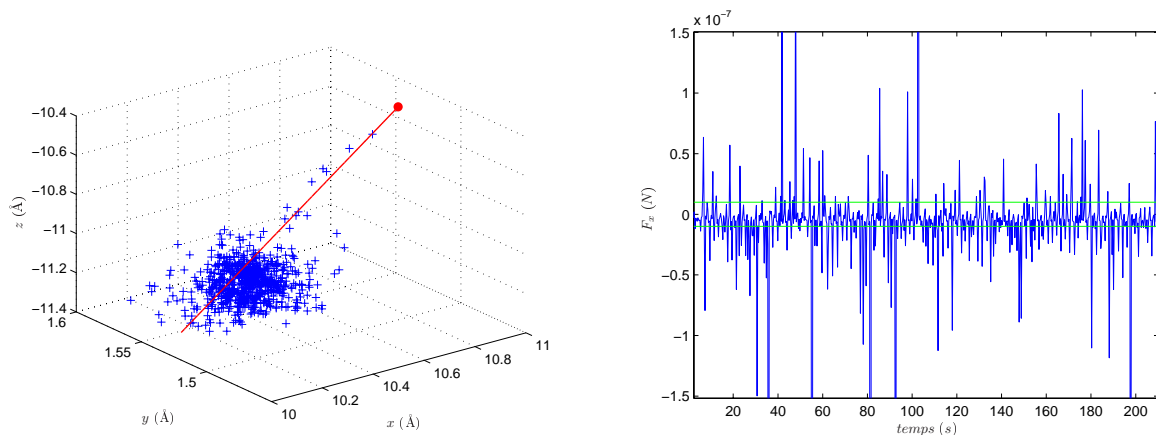
La courbe 2.19(a) retrace l'évolution du centre de masse du ligand flexible au cours d'une consigne linéaire en position. Celui-ci se déplace, dans un premier temps, linéairement, tel un corps rigide, selon la droite de consigne, tant que l'énergie d'interaction ne varie pas. Lorsque le déplacement total du ligand atteint 0.03 \AA , l'énergie d'interaction est modifiée. La minimisation impose un nouveau jeu de coordonnées pour le ligand, la position du centre de masse est modifiée en conséquence. Une fois sa conformation minimisée, le ligand est à nouveau soumis au déplacement linéaire. Le profil de force ainsi calculé est visible sur la figure 2.19(b).

Il est possible de distinguer une oscillation permanente du ligand autour de sa position d'énergie minimum. Cette oscillation permanente empêche d'interpréter de façon transparente ce profil de force. Bien que l'ordre de grandeur des forces d'interaction représente 10^{-8} N , ce qui autorise la mise en place d'un coefficient en effort sans perdre d'information en force, l'oscillation engendrée rendrait instable la manipulation.

L'objectif du *docking* avec retour d'effort est de choisir la meilleure géométrie du ligand pour le site actif considéré. Une telle oscillation ne peut pas fournir d'argument à l'opérateur quant à la meilleure force d'interaction ligand-protéine engendrée par des géométries différentes.

3.2.2 Facteur de déplacement élevé

Pour un coefficient en déplacement élevé, un débatement du *Virtuose* correspondant à un déplacement de 0.1 \AA , le profil de force semble varier avec de plus grandes amplitudes. L'ordre de grandeur, représenté par la droite verte ($F_x = 10^{-8} \text{ N}$), compare les deux profils des figures 2.19(b) et 2.20(b) page suivante. En effet, un facteur de force limitant le ressenti à cette valeur aurait pour conséquence, dans ce cas, une saturation de l'interface, les efforts



(a) Evolution du centre de masse au cours de la minimisation, + Positions du centre de masse, • position initiale, — consigne en position.

(b) Forces calculées au cours de la manipulation

Figure 2.20 - Manipulation de la biotine dans le site actif de la streptavidine suivant une consigne en position linéaire.

étant quasiment supérieurs à cette force tout au long de la manipulation. Ressentir un tel profil mêlant attractivité et répulsion même si les coefficients en efforts le rendent possible, ne peut conduire à une bonne interprétation du champ d'énergie interne à la protéine.

3.3 Conclusion

La grande amplitude des forces peut provenir de plusieurs facteurs. Le premier fait intervenir la rapidité de calcul de la minimisation et le rayon d'action des forces de van der Waals. En effet, si lors d'une manipulation avec un coefficient élevé, l'opérateur interpénètre plusieurs rayons de van der Waals avant que la minimisation ait pu empêcher cette configuration, les forces ainsi produites peuvent être importantes et repousser le ligand en conséquence. Cette force étant supérieure à celle qui est admissible par le *Virtuose*, elle est renvoyée à l'opérateur comme la force maximum admissible par le *Virtuose*. L'opérateur ainsi repoussé ne peut contrôler la déviation du ligand qui interpénétrera d'autres rayons de van der Waals.

Un faible facteur de déplacement permet tout de même, après filtrage et ajout de viscosité, de réaliser une opération de ressenti des forces à l'intérieur du site actif mais celui-ci est altéré. L'inconvénient majeur réside dans le temps de réalisation de la tâche. En effet, imposer un déplacement de 0.001 \AA pour un débattement de l'interface haptique empêche toute réalisation d'un *docking*, car il faudrait constamment repositionner l'interface pour faire avancer le ligand.

A l'inverse, un grand coefficient en déplacement provoque une ample variation des forces, autorisant la tâche de *docking* au détriment d'un bon ressenti de la scène.

4 Stabilité du système

Afin d'illustrer toute la difficulté que pose le retard vis-à-vis de la stabilité d'une commande, il convient d'étudier un exemple particulier qu'il sera possible de généraliser (figure 2.21).

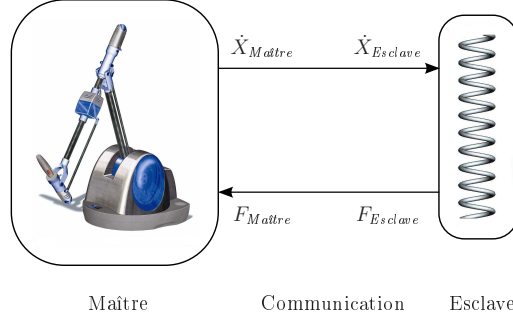


Figure 2.21 - Couplage sans retard du Virtuose sur un ressort virtuel.

Le couplage standard consiste à envoyer une consigne de vitesse à un ressort qui, en retour, envoie une force à l'interface haptique. Deux cas sont à envisager. Soit la communication est instantanée (figure 2.21) et, dans ce cas, le retard ne perturbe pas la liaison. Soit au contraire, le retard n'est plus négligeable, et il agit alors sur la stabilité de cette liaison (figure 2.22(a) page ci-contre).

La démonstration de la passivité de la communication, pour un tel dispositif, revient à écrire la différence de puissance entre le maître et l'esclave. La puissance entrante dans le système vaut alors :

$$P_{\text{Maître}} = \dot{X}_{\text{Maître}} F_{\text{Maître}} \quad (2.36)$$

La puissance retournée par l'esclave, quant à elle, s'écrit :

$$P_{\text{Esclave}} = \dot{X}_{\text{Esclave}} F_{\text{Esclave}} \quad (2.37)$$

La différence des deux définitions précédentes 2.36 et 2.37 vaut zéro car il est clair que $\dot{X}_{\text{Maître}} = \dot{X}_{\text{Esclave}}$ et que $F_{\text{Esclave}} = F_{\text{Maître}}$. La communication est donc passive, elle ne crée et ne dissipe pas d'énergie.

Le système précédent n'est pas réaliste car il existe un retard entre l'instant où la commande est envoyée à l'esclave et donc son interprétation en retour. C'est ce retard qui engendre une instabilité dans les mécanismes virtuels. La démonstration peut se faire en réutilisant l'exemple précédent et en introduisant un retard dans la communication.

En s'appuyant sur la figure 2.22(b) page suivante représentant à l'instant t l'envoi d'une information de vitesse à l'esclave, à l'instant $t + t_0$ sa réception par l'esclave, et à l'instant $t + 2t_0$ son interprétation par l'organe maître, il est possible de montrer que cette commande peut engendrer des instabilités. La puissance mise en jeu doit être étudiée :

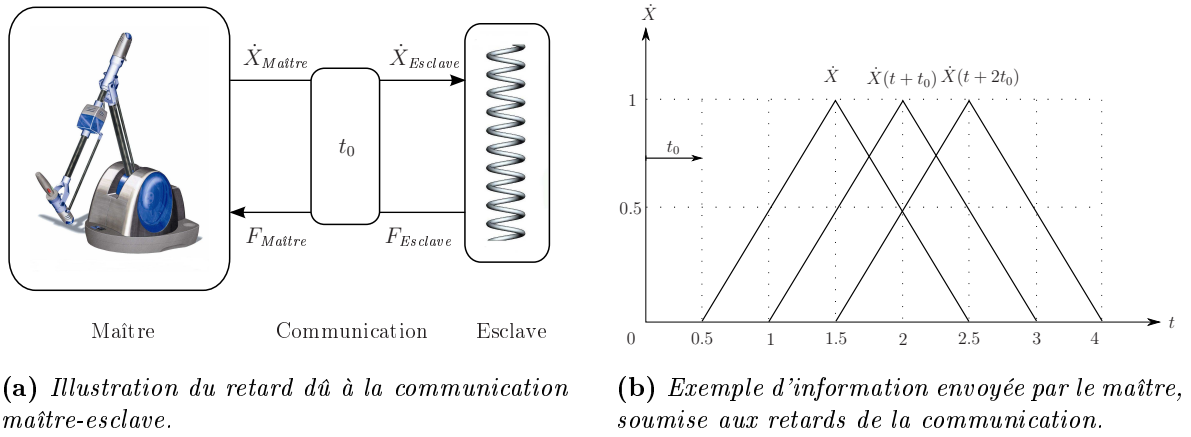


Figure 2.22 - Couplage avec retard du Virtuose sur un ressort virtuel.

$$P_{\text{échangée}} = P_{\text{Maître}} - P_{\text{Esclave}} \quad (2.38)$$

Cette équation peut s'écrire :

$$\begin{aligned} P_{\text{échangée}} &= \dot{X}_{\text{Maître}} F_{\text{Maître}} - \dot{X}_{\text{Esclave}} F_{\text{Esclave}} \\ &= \dot{X}(t) F(t + t_0) - \dot{X}(t + t_0) F(t) \end{aligned} \quad (2.39)$$

où $\dot{X}(t)$ est la vitesse à l'instant t et $F(t)$, la force créée par le ressort (figure 2.22(a)) de raideur K .

A l'instant t , le ressort engendre une force $F(t) = K\dot{X}(t + t_0)$. L'information de force ainsi reçue par le maître vaut alors $F(t + t_0) = K\dot{X}(t + 2t_0)$. La puissance échangée vaut donc, en tenant compte de l'équation 2.39 :

$$\begin{aligned} P_{\text{échangée}} &= \dot{X}(t) K \dot{X}(t + 2t_0) - \dot{X}(t + t_0) K \dot{X}(t + t_0) \\ &= K \dot{X}(t) \dot{X}(t + 2t_0) - K \dot{X}^2(t + t_0) \end{aligned} \quad (2.40)$$

En intégrant sur le temps de manipulation, la puissance échangée vaut :

$$\begin{aligned} P_{\text{échangée}} &= \int_{0.5}^{3.5} \left(K \dot{X}(t) \dot{X}(t + 2t_0) - K \dot{X}^2(t + t_0) \right) dt \\ &\leq 0 \end{aligned} \quad (2.41)$$

La puissance échangée entre le maître et l'esclave étant négative, l'esclave a donc créé de l'énergie au cours de la manipulation. Le couplage n'est donc pas passif. Le retard de communication peut être assimilé à un retard du simulateur moléculaire à cause du temps de calcul

nécessaire.

Un important temps de calcul agit donc directement sur la stabilité de la commande. Le système proposé, qui est basé sur un envoi et une réception simple de consignes sans gestion de retard, ne peut donc satisfaire l'exigence d'une simulation stable.

5 Conclusion

Ce chapitre avait pour but de montrer, à travers une première approche, la complexité du problème à résoudre. L'interface haptique nécessite une actualisation des données de l'ordre du kilohertz, la loi de commande doit être stable et tenir compte des retards de temps de calcul.

La simulation mise en place, bien qu'optimisée pour le processus de minimisation, ne fournit pas les forces à la cadence haptique désirée. Les retards des consignes sont donc à considérer dans la loi de commande, afin d'assurer la passivité de celle-ci. Ce problème de passivité, avec des retards de communication, engendre l'utilisation d'une nouvelle approche autour des variables d'ondes. Elles consistent à ne plus considérer les couples flux-efforts, mais directement les variables en flux d'énergie. Une onde porteuse de l'information se propageant dans les deux directions, et indépendante du retard, est ainsi créée.

La grande amplitude des forces ne peut être ressentie par un simple facteur d'homothétie en déplacement et en effort. Bien qu'un amortissement de cette dynamique nuise à la transparence, la méthode mise en place dans le chapitre suivant propose un filtrage de ces forces par l'intermédiaire des variables d'onde, afin d'intégrer le retard dans la loi de commande et rendre stable le couplage.

COUPLAGE D'UN SYSTÈME RETARDÉ À GRANDE DYNAMIQUE

Le chapitre précédent a fait état de plusieurs résultats dont la conséquence principale est l'instabilité du système à retour d'effort. Le temps de calcul engendré par la minimisation de l'énergie des molécules est élevé et ne garantit pas de réponse à une cadence haptique idéale, de l'ordre de la milliseconde. De plus, les efforts calculés par MOE ont une forte variation d'amplitude. Ces deux caractéristiques, efforts de forte amplitude et retards de calcul, font que le système à retour d'effort est instable.

Le présent chapitre tente d'apporter des solutions à ces deux problèmes. D'une part, la forte amplitude des efforts doit être atténuée afin que l'interface n'oscille pas entre des positions de raideur infinie ; d'autre part la commande du système haptique doit être stable. Elle doit donc tenir compte du fait que les consignes, envoyées par MOE vers l'interface, sont retardées par rapport à celles qui proviennent du *Virtuose*.

Dans un premier temps, sont présentés des opérateurs capables d'assurer la stabilité d'une communication retardée. Il s'agit des variables d'onde qui expriment des consignes de vitesses et d'efforts dans un espace de flux. Il s'agit alors d'en assurer la conservation entre l'interface haptique et la simulation moléculaire. Les consignes sont codées en termes de flux puis décodées en vue de leur interprétation.

Dans un deuxième temps, une solution qui consiste à ne considérer les variables d'onde que comme un amortisseur virtuel entre la simulation moléculaire et l'interface haptique est étudiée. Puisqu'elles agissent sur les vitesses, l'idée est donc d'introduire dans la commande un facteur d'amortissement qui, couplé à cette vitesse, atténue l'amplitude des efforts.

Bien qu'assurant la stabilité d'une communication retardée, les variables d'onde engendrent une raideur de couplage inversement proportionnelle au retard. La transparence s'en trouve donc atténuée.

Ce manque de transparence peut être une solution pour atténuer la forte variation d'amplitude des efforts. Elle est étudiée dans un dernier paragraphe. Ainsi, quelles que soient leurs variations, une modification de la transparence de la commande assure que l'interface ne considère pas les contacts comme une raideur infinie. Les efforts maximaux ne sont alors pas ressentis comme des contacts, la manipulation est, par conséquent, plus stable. Les résultats obtenus par cette deuxième commande sont analysés et comparés à ceux de la commande précédemment décrite.

1 Variables d'onde

1.1 Introduction

La téléopération fait interagir un opérateur dans un environnement quelconque grâce à un bras maître couplé à un organe esclave. Les manipulations à distance sont alors possibles, ainsi que les changements d'échelle par l'introduction de coefficients en effort et en déplacement dans les commandes. Ces systèmes de téléopération rencontrent plusieurs difficultés dont la principale concerne l'existence de délais dans la communication entre les interfaces maître et esclave. Si la téléopération se fait avec un retour d'effort, le système est alors très dépendant des retards de communication et sa stabilité n'est pas assurée. Le besoin d'interagir à distance a été prouvé, par exemple, par les travaux conduits dans l'espace [Sheridan 93] ou les opérations chirurgicales à distance [Guthart 00]. Ces communications, utilisant les ondes radio ou des liaisons TCP/IP sont donc soumises aux retards dus à l'éloignement ou à l'encombrement du réseau [Munir 03], [Niemeyer 98]. Les principales approches qui s'adaptent à ce problème utilisent la prédiction des mouvements de l'esclave [Sheridan 92]. Cependant, à cause du retour d'effort présent dans la boucle de contrôle, un faible décalage de consigne peut conduire à une forte instabilité de celui-ci. Les retards observés lors d'une téléopération à partir d'un réseau rendent impossible un retour d'effort sans compensation de ceux-ci.

Les variables d'onde sont apparues comme des outils capables d'assurer un retour d'effort en présence de retards arbitraires [Niemeyer 04]. Au début, cantonnées à des retards de communication constants, les variables d'onde s'étendent aux problèmes plus généraux des retards variables. Dans [Niemeyer 98], la transformation en variables d'onde compense les retards constants ou variables de la communication. Celle-ci est alors démontrée comme étant stable.

Les variables d'onde ont été principalement utilisées dans les travaux concernant des manipulations macroscopiques. Utilisées en chirurgie à distance, elles autorisent un chirurgien à manipuler ses outils intégrés à des robots. Il n'est mentionné nulle part l'utilisation des variables d'onde pour la commande de systèmes submicroniques à forte variations d'efforts. Certains travaux mentionnent l'utilisation des variables d'ondes pour la commande d'un système microscopique, mais le domaine d'application diffère dans le sens où la réponse du système microscopique est stable [Boukhnifer 04b], [Boukhnifer 04a], [Boukhnifer 06].

Garantissant une stabilité quel que soit le retard dans la communication, elles apparaissent comme un outil intéressant pour la compensation des retards de consigne du simulateur moléculaire MOE. Leurs propriétés, raideur et inertie variables devant l'importance des retards [Niemeyer 96] suggèrent qu'elles pourraient, par la faible transparence qu'elles engendrent, filtrer

les fortes amplitudes des efforts provenant du système moléculaire.

1.2 Présentation

Alors que les systèmes de téléopération traditionnels fonctionnent bien sans retard, ils sont soumis à des problèmes de stabilité, même pour de faibles délais. C'est pourquoi le problème du retard de consigne est abordé par l'introduction de variables d'onde qui sont des combinaisons linéaires de forces et de vitesses. Elles sont notamment applicables à des systèmes non-linéaires [Niemeyer 96]. En encodant les informations en un flux unique, elles sont robustes face aux retards arbitraires [Niemeyer 91]. Niemeyer propose la formulation de l'équation 3.1, représentée graphiquement sur la figure 3.1 :

$$\mathbf{U} = \frac{b\dot{\mathbf{X}} + \mathbf{F}}{\sqrt{2b}} \quad \mathbf{V} = \frac{b\dot{\mathbf{X}} - \mathbf{F}}{\sqrt{2b}} \quad (3.1)$$

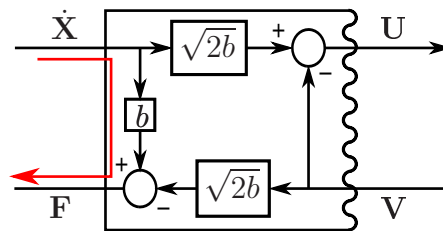


Figure 3.1 - La transformation classique des variables d'onde fait correspondre les vitesses et les forces en des flux entrant et sortant, convertissant les couples $(\dot{\mathbf{X}}, \mathbf{V})$ en (\mathbf{F}, \mathbf{U}) . La ligne brisée correspond au port d'interaction de la variable d'onde, et le chemin rouge représente le chemin de retour direct qui fait ressentir l'impédance apparente b .

Les variables \mathbf{U} et \mathbf{V} (figures 3.1 et 3.2 page suivante) sont respectivement le flux sortant et le flux entrant (pour un quadripôle quelconque). Par convention, le flux sortant \mathbf{U} conduit à une valeur positive de la puissance, alors que le flux entrant introduit un terme négatif. b représente l'impédance caractéristique de l'onde. Son introduction uniformise les unités physiques de l'équation et modifie le comportement de l'interface haptique en introduisant de l'amortissement ou, au contraire, une raideur plus élevée. En effet, si un système quelconque est décrit dans le domaine des variables d'onde, la variable \mathbf{U} est définie. Si l'impédance de la variable d'onde augmente, le poids de la vitesse $\dot{\mathbf{X}}$ sera plus important que celui de la force \mathbf{F} dans la définition. Une même valeur de \mathbf{U} entraîne donc une plus faible valeur pour la vitesse $\dot{\mathbf{X}}$ et une plus forte pour la force \mathbf{F} . En d'autres termes, une augmentation de la valeur de b va conduire à une augmentation de la force \mathbf{F} pour une réduction du mouvement. Le système apparaîtra donc plus amorti.

Le contraire est avéré lorsque l'impédance b est diminuée. Mais le système associe dès lors un poids plus important à la force, qui est donc atténuée comparativement à une augmentation de la vitesse. Le système apparaît, de ce fait, moins amorti.

De plus, en examinant la figure 3.1 page précédente, la transformation en variable d'onde introduit un chemin de retour direct entre la vitesse et la force, produisant un amortissement

b apparent. Cependant, l'amortissement induit n'est qu'apparent. L'énergie n'est pas dissipée mais transmise au reste du système. En effet, si l'énergie transmise par le système téléopéré est nulle, ou si elle n'est retournée qu'après un retard conséquent, cet amortissement apparent sera clairement ressenti. Mais si l'énergie est transmise très rapidement, cet amortissement apparent ne sera pas perceptible. Le facteur b devient donc important au regard du ressenti haptique si le délai augmente.

Lors du choix de b , il faudra envisager une téléopération avec un mouvement lent mais avec une information en force accrue, ou un mouvement rapide avec un retour d'efforts altéré. La transparence de la commande est donc fortement assujettie à la valeur de b .

D'après le chapitre précédent, une chaîne de commande soumise à un retard de communication ou de calcul produit de l'énergie. Elle est par conséquent instable. Or le principe des variables d'onde est de rendre passive une communication perturbée par des retards. Considérer cette communication comme un flux d'énergie dont il suffit d'assurer la conservation permet alors de rendre stable une telle commande. Les variables d'entrée et de sortie ne sont plus une force, une vitesse ou une position, mais une énergie. Le principe des variables d'onde est donc de projeter le problème dans un espace de flux conservatif (équation 3.1).

Cette transformation est bijective et garantit l'obtention de toutes les relations concernant les vitesses et les forces. Le passage du domaine physique au domaine des variables d'onde est alors possible. Le couplage s'effectue par encodage/décodage successifs des variables concernées en vue de leur interprétation tant au niveau du contrôleur que de la simulation.

1.3 Principe

En partant du principe qu'il faut encoder les variables de sortie en flux d'énergie, puis les décoder en vue de leur interprétation par la simulation, un schéma de couplage classique peut être mis en place. Il est présenté à la figure 3.2.

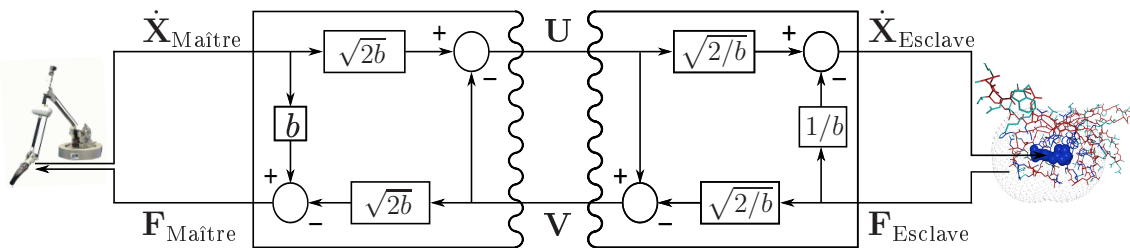


Figure 3.2 - Schéma de couplage entre le Virtuose et la simulation. Les vitesses et forces sont successivement codées puis décodées.

De l'expression du codage en variables d'onde (équation 3.1) des grandeurs de vitesse et de force, découle leur décodage :

$$\mathbf{F} = \sqrt{\frac{b}{2}}(\mathbf{U} - \mathbf{V}) \quad (3.2)$$

$$\dot{\mathbf{X}} = \frac{1}{\sqrt{2b}} (\mathbf{U} + \mathbf{V}) \quad (3.3)$$

En trois dimensions, et en considérant six degrés de liberté, \mathbf{F} et $\dot{\mathbf{X}}$ s'écrivent respectivement en fonction des forces et des moments d'interaction ligand-protéine et en fonction de la vitesse et de la vitesse angulaire du *Virtuose* :

$$\begin{aligned} \mathbf{F}_{[6 \times 1]} &= [F_x, F_y, F_z, M_x, M_y, M_z]^t \\ \dot{\mathbf{X}}_{[6 \times 1]} &= [\dot{X}_x, \dot{X}_y, \dot{X}_z, \Omega_x, \Omega_y, \Omega_z]^t \end{aligned} \quad (3.4)$$

1.4 Passivité

Afin d'établir la passivité des variables d'onde, il convient d'étudier l'exemple précédent et représenté à la figure 3.2 page précédente. Celui-ci peut se représenter sous la forme de la figure 3.3 en tenant compte du codage en variables d'onde.

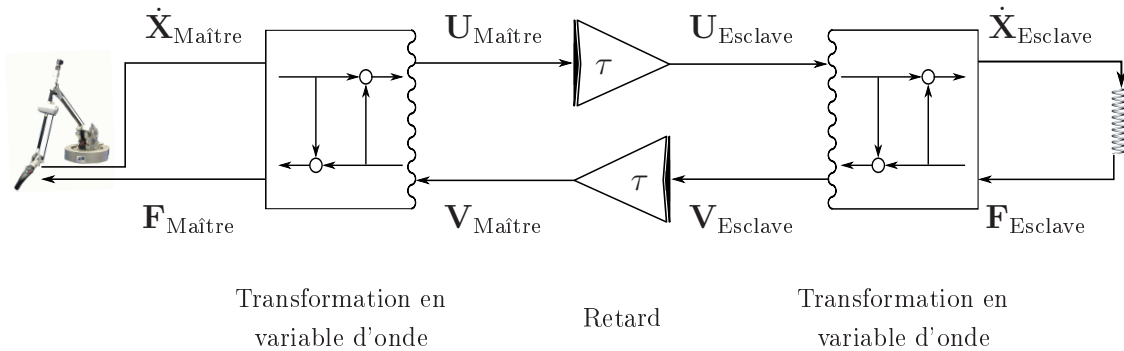


Figure 3.3 - La communication en variables d'onde encode l'information envoyée par le maître puis la décode pour la simulation. L'opération inverse est effectuée pour le retour de l'information.

Dans la suite de la démonstration, par convention, \mathbf{U} représentera l'information envoyée par le maître vers l'esclave. \mathbf{V} est l'information retour. De plus, un scalaire noté \mathbf{U}^2 correspond à l'opération de multiplication matricielle $\mathbf{U}^t \mathbf{U}$.

Les équations qui régissent la communication peuvent s'écrire en fonction du retard τ sous la forme :

$$\mathbf{U}_{\text{Esclave}}(t) = \mathbf{U}_{\text{Maître}}(t - \tau) \quad (3.5)$$

$$\mathbf{V}_{\text{Maître}}(t) = \mathbf{V}_{\text{Esclave}}(t - \tau) \quad (3.6)$$

En considérant l'expression des variables d'onde, les entrées ci-dessus décrites peuvent s'écrire :

$$\mathbf{U}_{\text{Maître}}(t) = \frac{b\dot{\mathbf{X}}_{\text{Maître}}(t) + \mathbf{F}_{\text{Maître}}(t)}{\sqrt{2b}} \quad (3.7)$$

$$\mathbf{V}_{\text{Esclave}}(t) = \frac{b\dot{\mathbf{X}}_{\text{Esclave}}(t) - \mathbf{F}_{\text{Esclave}}(t)}{\sqrt{2b}} \quad (3.8)$$

Afin de montrer la passivité de la communication, il convient d'étudier la puissance échangée entre le maître et l'esclave. Celle-ci doit être positive.

$$P_{\text{échangée}} = \dot{\mathbf{X}}_{\text{Maître}}^t \mathbf{F}_{\text{Maître}} - \dot{\mathbf{X}}_{\text{Esclave}}^t \mathbf{F}_{\text{Esclave}} \quad (3.9)$$

En exprimant les vitesses et les forces à l'aide des équations 3.7 et 3.8, l'équation 3.9 peut s'écrire :

$$P_{\text{échangée}} = \frac{1}{2} \mathbf{U}_{\text{Maître}}^2 - \frac{1}{2} \mathbf{V}_{\text{Maître}}^2 - \frac{1}{2} \mathbf{U}_{\text{Esclave}}^2 + \frac{1}{2} \mathbf{V}_{\text{Esclave}}^2 \quad (3.10)$$

En substituant l'expression des variables retardées (3.5, 3.6) dans celle de la puissance échangée, celle-ci s'exprime de la manière suivante :

$$P_{\text{échangée}} = \frac{1}{2} \mathbf{U}_{\text{Maître}}^2(t) - \frac{1}{2} \mathbf{U}_{\text{Maître}}^2(t - \tau) + \frac{1}{2} \mathbf{V}_{\text{Esclave}}^2(t) - \frac{1}{2} \mathbf{V}_{\text{Esclave}}^2(t - \tau) \quad (3.11)$$

L'intégration de l'équation 3.11 conduit à l'expression ci-dessous :

$$\int_0^t P_{\text{échangée}} dt = \frac{1}{2} \int_0^t [\mathbf{U}_{\text{Maître}}^2(t) - \mathbf{U}_{\text{Maître}}^2(t - \tau) + \mathbf{V}_{\text{Esclave}}^2(t) - \mathbf{V}_{\text{Esclave}}^2(t - \tau)] dt \quad (3.12)$$

La dernière étape consiste à montrer que l'équation 3.12 est positive, garantissant ainsi que l'esclave ne crée pas d'énergie au travers de la communication. Il faut revenir à l'expression de la passivité. Celle-ci stipule dans ce cas précis :

$$\begin{aligned} \int_0^t P_{\text{échangée}} dt &= \frac{1}{2} \int_0^t [\mathbf{U}_{\text{Maître}}^2(t) - \mathbf{U}_{\text{Maître}}^2(t - \tau) + \mathbf{V}_{\text{Esclave}}^2(t) - \mathbf{V}_{\text{Esclave}}^2(t - \tau)] dt \\ &= E_{\text{stockée}}(t) - E_{\text{stockée}}(0) + \int_0^t P_{\text{dissipée}} dt \end{aligned} \quad (3.13)$$

En réécrivant l'égalité précédente, et en considérant que l'énergie stockée à l'instant t est positive tout comme la puissance dissipée au cours du temps, il est possible d'obtenir l'inégalité montrant qu'un système est passif uniquement si l'énergie de la sortie est inférieure à l'énergie de l'entrée, soit :

$$\begin{aligned} \frac{1}{2} \int_0^t [\mathbf{V}_{\text{Esclave}}^2(t-\tau) - \mathbf{V}_{\text{Esclave}}^2(t)] dt &\leq \frac{1}{2} \int_0^t [\mathbf{U}_{\text{Maître}}^2(t) - \mathbf{U}_{\text{Maître}}^2(t-\tau)] dt \\ &+ E_{\text{stockée}}(0) \quad \forall t \geq 0 \end{aligned} \quad (3.14)$$

Cette équation s'écrit, en effectuant le changement de variable (3.5 et 3.6), et en considérant que l'énergie initialement stockée est nulle (hypothèse vérifiée par le fait que $\mathbf{U} \equiv \mathbf{0}$ pour tout $t < 0$) :

$$\int_0^{t-\tau} \frac{1}{2} \mathbf{V}_{\text{Esclave}}^2(t) dt - \int_0^t \frac{1}{2} \mathbf{V}_{\text{Esclave}}^2(t) dt \leq \int_0^t \frac{1}{2} \mathbf{U}_{\text{Maître}}^2(t) dt - \int_0^{t-\tau} \frac{1}{2} \mathbf{U}_{\text{Maître}}^2(t) dt \quad \forall t \geq 0 \quad (3.15)$$

Finalement, le résultat ainsi obtenu s'écrit :

$$\int_{t-\tau}^t \frac{1}{2} \mathbf{V}_{\text{Esclave}}^2(t) dt + \int_{t-\tau}^t \frac{1}{2} \mathbf{U}_{\text{Maître}}^2(t) dt \geq 0 \quad (3.16)$$

Alors, pour un système dont la dissipation est nulle et pour lequel l'énergie initialement stockée est nulle, ce qui constitue le cas le plus contraignant, l'énergie stockée est égale à la puissance échangée :

$$\begin{aligned} \int_0^t P_{\text{échangée}} dt &= E_{\text{stockée}}(t) \\ &= \int_{t-T}^t \frac{1}{2} \mathbf{V}_{\text{Esclave}}^2(t) dt + \int_{t-T}^t \frac{1}{2} \mathbf{U}_{\text{Maître}}^2(t) dt \\ &\geq 0 \end{aligned} \quad (3.17)$$

La puissance échangée est positive ou nulle. Cette propriété garantit la passivité de la communication et donc la stabilité du système. En outre, l'énergie stockée correspond à l'énergie du retard. Le retard est alors considéré comme un élément passif dans une chaîne de transmission passive. Un système exprimé sous forme de variables d'onde devient, en théorie, complètement robuste et indépendant du retard.

1.5 Conclusion

De cette première étude des variables d'onde, plusieurs propriétés fondamentales pour les systèmes à communication retardée doivent être retenues. D'une part, elles assurent la stabilité de la commande grâce à la conservation des flux entre le maître et l'esclave. D'autre part l'impédance caractéristique b est une constante qui influe sur l'amortissement du couplage. Cette impédance laisse à penser que la grande amplitude des forces étudiées au chapitre précédent

pourrait être filtrée par cet amortisseur. De plus, la commande retardée par le simulateur, codée en variables d'onde, serait stable. Il convient dès lors d'intégrer les variables d'onde à la commande du ligand. Cette nouvelle application des variables d'onde nécessite une étude approfondie.

2 Etude d'une commande en variables d'onde

Les variables d'onde ont des propriétés intéressantes sur la stabilité des communications retardées et un filtrage des informations est possible par le biais de l'impédance caractéristique b . Le codage en variables d'onde est donc implémenté entre le *Virtuose* et MOE.

L'objet de ce paragraphe est de montrer l'intégration de la commande dans la plate-forme de téléopération. Par la forme de l'implémentation des variables d'onde, des chemins de réflexion apparaissent. Des solutions sont apportées afin de les éliminer sans perturber la communication. Enfin, la transparence du système est étudiée pour que la performance de l'interface haptique soit évaluée.

2.1 Commande du *Virtuose* dans le cas idéal

Le chapitre précédent propose une commande homothétique pour la manipulation d'un ligand flexible. Cette commande est ici détaillée pour observer son évolution lors d'un codage en variables d'onde. Le couplage possède une inertie et une raideur propre qu'il est possible de modéliser (figure 3.4).

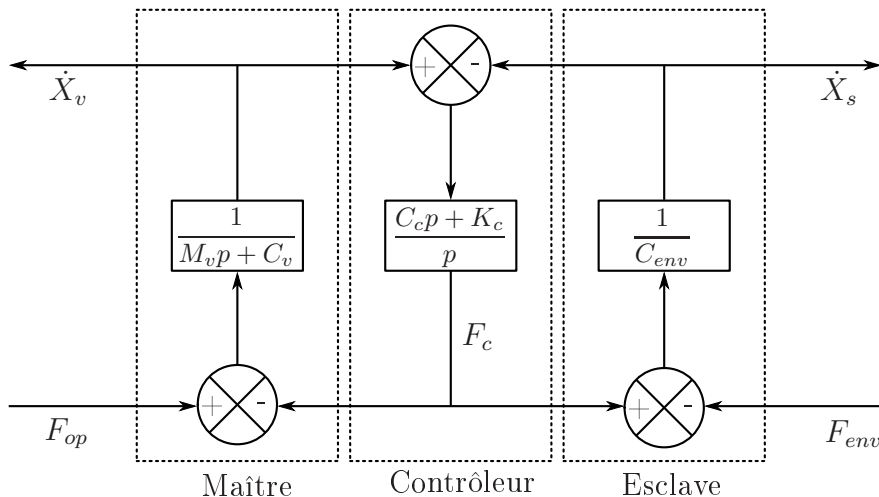


Figure 3.4 - Schéma de commande de la simulation dans le cas idéal sans retard. Le principe est exposé en une dimension. Une généralisation est possible en considérant les éléments comme des matrices. Celles-ci sont énoncées aux équations 3.18, 3.19 et 3.20.

Le *Virtuose*, modélisé par une inertie \mathbf{M}_v et un amortissement \mathbf{C}_v , peut être représenté sous forme matricielle à l'aide des variables de Laplace notées p , par l'expression 3.18 :

$$\begin{bmatrix} \dot{\mathbf{X}}_v \\ \mathbf{F}_{op} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{[3 \times 3]} & \mathbf{0}_{[3 \times 3]} \\ \mathbf{M}_v p + \mathbf{C}_v & \mathbf{1}_{[3 \times 3]} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{X}}_v \\ \mathbf{F}_c \end{bmatrix} \quad (3.18)$$

où \mathbf{F}_{op} représente les efforts exercés par l'opérateur sur le bras maître, et est de dimension $[6 \times 1]$, $\dot{\mathbf{X}}_v$ la vitesse du *Virtuose*, de dimension $[6 \times 1]$, et \mathbf{F}_c l'entrée du correcteur, de dimension $[6 \times 1]$. \mathbf{M}_v est une matrice diagonale $[3 \times 3]$ qui représente l'inertie de l'interface selon chaque axe. \mathbf{C}_v est une matrice d'amortissement selon chaque axe et est de dimension $[3 \times 3]$.

Le contrôleur, qui est l'ensemble ressort/amortisseur présent entre la simulation et le *Virtuose*, s'écrit en considérant que le ressort a une raideur \mathbf{K}_c , de dimension $[3 \times 3]$, et un amortissement \mathbf{C}_c de dimension $[3 \times 3]$:

$$\begin{bmatrix} \dot{\mathbf{X}}_v \\ \mathbf{F}_c \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{[3 \times 3]} & p(\mathbf{C}_c p + \mathbf{K}_c)^{-1} \\ \mathbf{0}_{[3 \times 3]} & \mathbf{1}_{[3 \times 3]} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{X}}_s \\ \mathbf{F}_c \end{bmatrix} \quad (3.19)$$

où $\dot{\mathbf{X}}_s$ représente la vitesse de l'objet asservi. Le système moléculaire peut être globalement représenté par un amortisseur représenté par C_{env} . Son expression s'écrit alors :

$$\begin{bmatrix} \dot{\mathbf{X}}_s \\ \mathbf{F}_c \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{[3 \times 3]} & \mathbf{0}_{[3 \times 3]} \\ C_{env} \cdot \mathbf{1}_{[3 \times 3]} & \mathbf{1}_{[3 \times 3]} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{X}}_s \\ \mathbf{F}_{env} \end{bmatrix} \quad (3.20)$$

L'ensemble de ces définitions peut être représenté par le schéma bloc de la figure 3.4 page ci-contre, décrit en une dimension.

2.2 Utilisation des variables d'onde

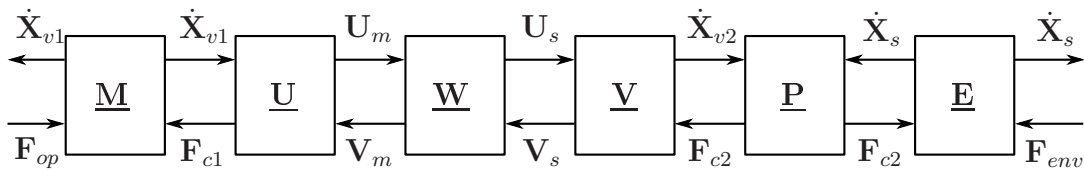


Figure 3.5 - Modélisation de la transformation en variables d'onde à l'aide de blocs élémentaires.

La figure 3.5 présente, sous forme de blocs, la commande appliquée entre le simulateur moléculaire et le *Virtuose*.

- Le bloc $\underline{\mathbf{M}}_{[6 \times 6]}$ représente le *Virtuose*, organe maître,
- le bloc $\underline{\mathbf{U}}_{[6 \times 6]}$ assure le passage du domaine physique à celui des variables d'onde,
- le bloc $\underline{\mathbf{W}}_{[6 \times 6]}$ représente le délai existant dans la communication,
- le bloc $\underline{\mathbf{V}}_{[6 \times 6]}$ assure le décodage de la variable d'onde pour son interprétation dans le domaine physique,
- $\underline{\mathbf{P}}_{[6 \times 6]}$ est le correcteur proportionnel dérivé,
- $\underline{\mathbf{E}}_{[6 \times 6]}$ représente l'environnement au travers du ligand, considéré comme l'esclave.

Le délai ainsi introduit est décrit par la matrice 3.21, dans laquelle τ représente le retard.

$$\underline{\mathbf{W}} = \begin{bmatrix} e^{\tau p} \cdot \mathbf{1}_{[3 \times 3]} & \mathbf{0}_{[3 \times 3]} \\ \mathbf{0}_{[3 \times 3]} & e^{-\tau p} \cdot \mathbf{1}_{[3 \times 3]} \end{bmatrix} \quad (3.21)$$

Les blocs $\underline{\mathbf{U}}$ et $\underline{\mathbf{V}}$ étant régis par les équations 3.22 et 3.23, où b représente l'impédance caractéristique de la variable d'onde, ils s'écrivent :

$$\underline{\mathbf{U}} = \frac{1}{\sqrt{2b}} \begin{bmatrix} \mathbf{1}_{[3 \times 3]} & \mathbf{1}_{[3 \times 3]} \\ b \cdot \mathbf{1}_{[3 \times 3]} & -b \cdot \mathbf{1}_{[3 \times 3]} \end{bmatrix} \quad (3.22)$$

$$\underline{\mathbf{V}} = \frac{1}{\sqrt{2b}} \begin{bmatrix} b \cdot \mathbf{1}_{[3 \times 3]} & \mathbf{1}_{[3 \times 3]} \\ b \cdot \mathbf{1}_{[3 \times 3]} & -\mathbf{1}_{[3 \times 3]} \end{bmatrix} \quad (3.23)$$

La commande mise en place tient compte du retard de communication dans la chaîne de commande de l'environnement. Celle-ci calcule un effort après avoir reçu, à son entrée, une position ou une vitesse, puis l'envoie au *Virtuose*. Cependant cette commande agit directement sur la transparence haptique. D'une part à cause des retards, et d'autre part à cause de l'impédance caractéristique qui est ressentie comme un amortissement.

2.3 Filtrage

Un des inconvénients majeurs des variables d'onde, de par la présence de nombreuses boucles dans la chaîne de commande, est l'apparition de phénomènes de réflexion. Une consigne peut en effet suivre plusieurs chemins sans forcément revenir au *Virtuose*. La perturbation agit donc directement sur la stabilité.

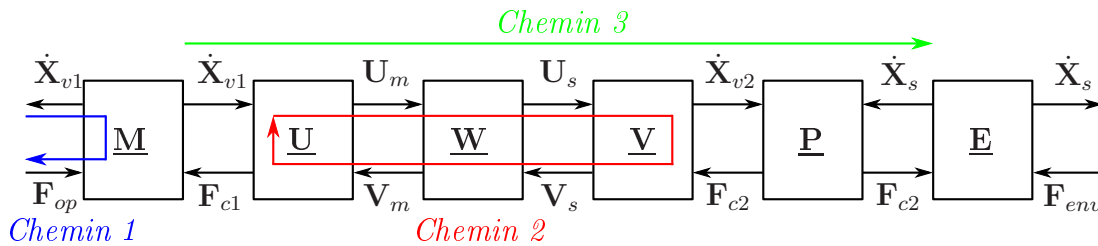


Figure 3.6 - Chemins de rétroactions possibles lors d'un couplage en variables d'onde.

La figure 3.6 présente les chemins possibles que peut emprunter la consigne.

Le premier chemin apparaît comme un amortissement pour l'utilisateur. Il représente l'inertie et la raideur du *Virtuose* modélisées par la figure 3.4 page 86. Ce chemin est indépendant du retard et agit directement sur la transparence du système. Cependant, il n'est pas lié au codage en variables d'onde mais à sa conception mécanique directement.

Le second chemin, directement lié au codage en variables d'onde, perturbe le signal envoyé par le *Virtuose*. Il peut être assimilé à une réflexion d'onde. Lorsque la consigne \mathbf{U}_s , consigne du maître transcrite en variables d'onde et soumise au retard, est présentée au bloc responsable du décodage, une partie du signal est renvoyée par l'onde retour \mathbf{V}_s (figures 3.6 et 3.2 page 82). Cette consigne de retour est alors directement retournée au *Virtuose*. Dans ce cas, cette consigne de retour est erronée puisqu'elle n'est pas transmise par la simulation. Elle peut être réintroduite dans la boucle de contrôle au travers de \mathbf{U}_m . Cette particularité provoque l'effet de réflexion et, par conséquent, agit sur la stabilité du système.

Le troisième chemin représente l'information utile et c'est lui qui doit être conservé.

L'information erronée, découlant de la conception des variables d'onde et représentée par le deuxième chemin, peut être éliminée par un filtrage de la variable. Cependant ce filtre ne doit pas agir sur les consignes découlant d'un mouvement de l'opérateur. Il est courant de considérer que les mouvements humains, et notamment ceux de l'opérateur, agissant sur une interface haptique sont limités à une fréquence de quelques hertz. En effet, la raideur et la conception de l'interface ne permettent pas un mouvement classique sans que l'opérateur ne se fatigue musculairement pour des mouvements rapides. La fréquence peut donc être envisagée inférieure à dix hertz. Dès lors, il est possible de distinguer une information se propageant à une cadence haptique (1 ms) d'un mouvement de l'opérateur par un filtre ne laissant transparaître que les basses fréquences. La variable d'onde \mathbf{U} est donc soumise à un filtre passe-bas de pulsation de coupure λ , qui varie selon les axes considérés. Les hautes fréquences seront ainsi filtrées.

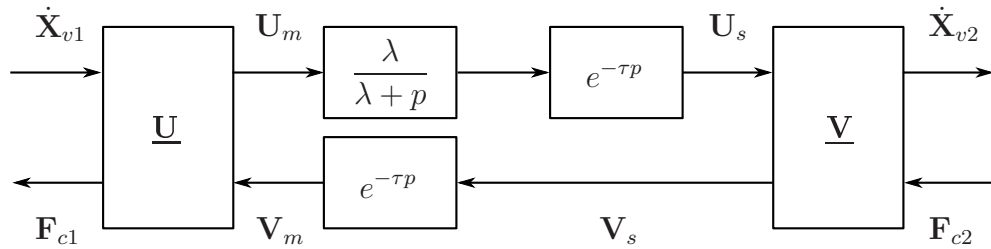


Figure 3.7 - Codage et décodage d'une information en variables d'onde soumise à un retard au travers d'un filtre passe-bas de pulsation de coupure λ .

La figure 3.7 représente le codage et le décodage d'une information en variables d'onde soumise à un retard au travers d'un filtre passe-bas.

La matrice $\underline{\mathbf{W}}$ réunissant, cette fois, le retard et le filtrage de la réflexion peut alors s'écrire :

$$\underline{\mathbf{W}} = \begin{bmatrix} \frac{\lambda+p}{\lambda} e^{\tau p} \mathbf{1}_{[3 \times 3]} & \mathbf{0}_{[3 \times 3]} \\ \mathbf{0}_{[3 \times 3]} & e^{-\tau p} \mathbf{1}_{[3 \times 3]} \end{bmatrix} \quad (3.24)$$

2.4 Transparence

Les variables d'onde agissent comme un ressort entre le maître et l'esclave. [Niemeyer 96] montre que pour un couplage en variables d'onde sans filtrage, la chaîne de commande peut être modélisée par une masse et une raideur équivalentes, les caractéristiques dépendent donc de l'impédance de couplage b ainsi que du délai τ (équation 3.25 et 3.26).

$$K_{\text{Communication}} = \frac{b}{\tau} \quad (3.25)$$

$$M_{\text{Communication}} = b\tau \quad (3.26)$$

Les deux équations précédentes dépendent du retard. Celui-ci dégrade la raideur du couplage. Autrement dit, il est d'autant plus difficile de ressentir un contact dur, tel un mur, que le délai de communication est important. De plus, l'inertie du système augmente proportionnellement. En revanche, pour un retard nul, la raideur du couplage devient infinie, garantissant ainsi la possibilité de ressentir des contacts de façon transparente. L'inertie, quant à elle, tend vers zéro. Cette situation idéale ne peut être atteinte en présence de retard.

En présence d'un filtrage des variables d'onde, la raideur et l'inertie équivalentes dépendent non seulement des constantes b et τ citées précédemment, mais aussi du facteur de coupure λ .

La transparence se calcule pour un régime établi. L'inertie doit être la plus faible possible, pour une consigne nulle retournée par le correcteur, et la raideur être maximale en mode bloqué. Ce mode représente une consigne en vitesse nulle, c'est-à-dire un mur. Ces conditions s'écrivent par l'équation 3.27.

$$\begin{bmatrix} \dot{\mathbf{X}}_{v1} \\ \mathbf{F}_{c1} \end{bmatrix} = \underline{\mathbf{U}} \underline{\mathbf{W}} \underline{\mathbf{V}} \begin{bmatrix} \dot{\mathbf{X}}_{v2} \\ \mathbf{F}_{c2} \end{bmatrix} \quad (3.27)$$

En utilisant les définitions de $\underline{\mathbf{U}}$, $\underline{\mathbf{V}}$ et $\underline{\mathbf{W}}$ ci-dessus décrites aux équations 3.22, 3.23 et 3.24, le système 3.27 s'écrit :

$$\begin{bmatrix} \dot{\mathbf{X}}_{v1} \\ \mathbf{F}_{c1} \end{bmatrix} = \frac{1}{2b} \begin{bmatrix} b \left(\frac{\lambda + p}{\lambda} e^{\tau p} + e^{-\tau p} \right) \cdot \mathbf{1}_{[3 \times 3]} & \frac{\lambda + p}{\lambda} e^{\tau p} + e^{-\tau p} \cdot \mathbf{1}_{[3 \times 3]} \\ b^2 \left(\frac{\lambda + p}{\lambda} e^{\tau p} + e^{-\tau p} \right) \cdot \mathbf{1}_{[3 \times 3]} & b \left(\frac{\lambda + p}{\lambda} e^{\tau p} + e^{-\tau p} \right) \cdot \mathbf{1}_{[3 \times 3]} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{X}}_{v2} \\ \mathbf{F}_{c2} \end{bmatrix} \quad (3.28)$$

Cette dernière équation permet de conclure que la valeur de l'inertie et de la raideur équivalente d'une telle loi de commande [Hart 07] vaut :

$$M_{\text{Communication}} = \lim_{p \rightarrow 0} \left. \frac{\mathbf{F}_{c1}}{\dot{\mathbf{X}}_{v1}} \right|_{\mathbf{F}_{c2}=0} = b\tau + \frac{b}{2\lambda} \quad (3.29)$$

$$K_{\text{Communication}} = \lim_{p \rightarrow 0} \left. \frac{\mathbf{F}_{c1}}{\dot{\mathbf{X}}_{v1}} \right|_{\dot{\mathbf{x}}_{c2}=0} = \frac{2b\lambda}{1 + 2\tau\lambda} \quad (3.30)$$

D'après les expressions 3.29 et 3.30, le ressenti dépend toujours de b . Plus λ augmente, plus l'inertie diminue, tendant vers un système idéal, d'inertie nulle si le retard τ tend vers zéro. Le phénomène inverse se présente pour la raideur de la communication. En effet, si λ augmente et que le retard tend vers zéro, la raideur tend à son tour vers l'infini, rendant le couplage rigide et donc idéal.

L'application de ce filtrage nécessite donc un réglage de la pulsation de coupure. Mais le ressenti haptique, vu de l'opérateur, dépend lui aussi de cette pulsation de coupure qui agit directement sur la transparence de la commande. Accroître l'inertie de l'interface par une diminution de la pulsation de coupure pourrait avantageusement diminuer l'influence des fortes amplitudes des efforts d'interaction ligand-protéine. La raideur, quant à elle, diminuerait. Les fortes amplitudes ne seraient alors plus ressenties par des raideurs infinies, mais par l'illusion d'un contact mou.

Cependant, une commande se doit, à l'origine, d'être transparente. La forte amplitude des forces doit donc être diminuée par l'apport d'un amortisseur extérieur. Celui-ci, multiplié aux vitesses du *Virtuose*, ajouterait de la viscosité à la simulation, limitant les déplacements du ligand dans le site actif. Cette limitation en position agit directement sur le profil des efforts.

2.5 Conclusion

Les variables d'onde sont une solution pour rendre passive une communication intégrant des retards. Cependant, la transparence en est altérée de par la présence d'une impédance interne b , responsable de la stabilisation de la variable d'onde, et d'une pulsation de coupure λ qui filtre les hautes fréquences provenant de la simulation. Un des problèmes du ressenti haptique lors du *docking*, comme l'a montré le chapitre 2, est la présence de hautes fréquences rendant impossible un ressenti compréhensible des interactions moléculaires. De par le manque de transparence des variables d'onde et le filtrage intégré, ces hautes fréquences peuvent être diminuées.

Cependant, il est préférable de concevoir une commande transparente. Les hautes fréquences, ainsi que les grandes amplitudes des efforts d'interaction, peuvent être atténuées par un amortisseur entre l'organe maître et le ligand. Afin de conserver le bénéfice des variables d'onde, celles-ci peuvent être appliquées à un amortisseur.

3 Variables d'onde et amortisseur virtuel

La simulation de *docking* moléculaire utilise, comme entrée la variation de position du *Virtuose*. La première loi de commande utilisant les variables d'onde doit donc envoyer une position au simulateur afin de ne pas le modifier. L'avantage d'utiliser une position plutôt qu'une vitesse est que celle-ci est très bruitée et provoquerait une dérive en position.

Le problème majeur qui doit, entre autres, être géré par la loi de commande est la forte variabilité des forces. L'idée d'envoyer la position de l'interface directement à la simulation et d'utiliser la variable d'onde comme un amortisseur virtuel est donc fondée. La vitesse décodée de la variable d'onde servira alors, *via* une constante d'amortissement, à atténuer la grande variabilité des forces.

3.1 Schéma de commande

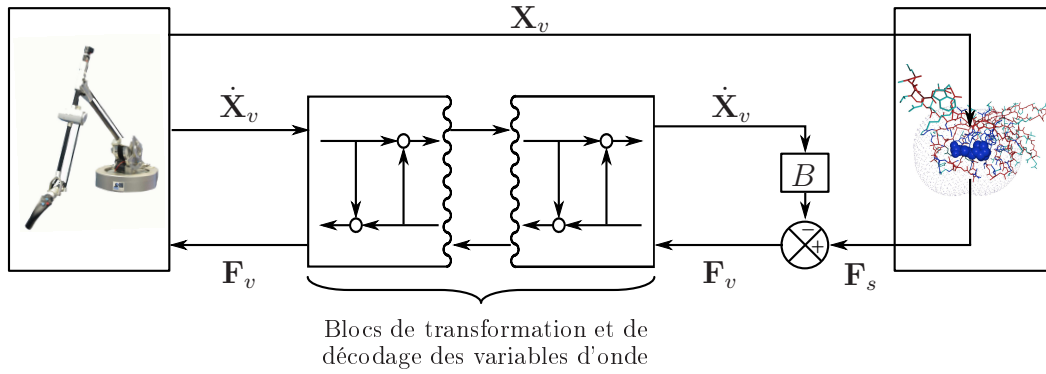


Figure 3.8 - Schéma de couplage utilisant les variables d'onde comme un amortisseur virtuel.

La figure 3.8 présente le schéma de commande utilisant les variables d'onde comme amortisseur virtuel. La position et l'orientation de l'interface haptique \mathbf{X}_v sont envoyées au simulateur moléculaire par protocole UDP. Après traitement de cette information, un effort, noté \mathbf{F}_s , exprimé dans le monde macro, après application du coefficient en effort adéquat, est exprimé en sortie du simulateur. Le facteur d'amortissement B , présent dans la boucle de rétroaction, atténue les efforts ainsi engendrés. La relation 3.31 calcule les efforts à ressentir :

$$\mathbf{F}_v = \mathbf{F}_s - B\dot{\mathbf{X}}_v \quad (3.31)$$

où \mathbf{F}_v représente le total des efforts à retourner à l'utilisateur *via* les variables d'onde.

3.2 Codage et décodage des variables

La figure 3.9 page suivante représente le couplage réalisé entre le *Virtuose* et le simulateur moléculaire. Une modification de la définition des variables d'onde est effectuée afin de conserver une réalité physique (équations 3.32 et 3.33).

$$\mathbf{U} = \frac{b\dot{\mathbf{X}}_v - \mathbf{F}_v}{\sqrt{2b}} \quad (3.32)$$

$$\mathbf{V} = \frac{b\dot{\mathbf{X}}_v + \mathbf{F}_v}{\sqrt{2b}} \quad (3.33)$$

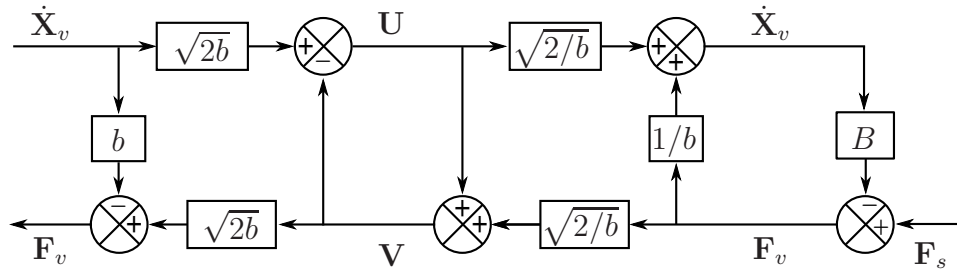


Figure 3.9 - Codage et décodage des variables d'onde.

La boucle en admittance, qui encode les efforts provenant de la simulation en un flux d'énergie et qui décode le flux d'énergie provenant de l'interface en une vitesse est modifiée en conséquence. En effet, lorsque la simulation renvoie un effort, la vitesse du *Virtuose* doit avoir le même sens (équation 3.33). A l'inverse, si la vitesse du *Virtuose* a un sens qui lui est propre, la force calculée n'est pas nécessairement de même sens (équation 3.32). Autrement dit, un effort doit engendrer un mouvement dans le même sens, le mouvement ne pouvant pas être contraire à cette force.

A partir de cette modification, les différentes valeurs physiques de l'interface et de la simulation peuvent être encodées ou décodées selon la figure 3.9. Ainsi la variable d'onde retour V peut s'exprimer en fonction de la force à renvoyer au *Virtuose*, F_v :

$$V = U + \sqrt{\frac{2}{b}} F_v \quad (3.34)$$

Afin d'éviter des instabilités internes, la boucle algébrique qui calcule la vitesse du *Virtuose* à partir des variables d'onde est calculée analytiquement. La vitesse du *Virtuose* doit être exprimée en fonction de l'effort retourné par la simulation. Il est notamment nécessaire de connaître son encodage en variable d'onde pour le retour de la consigne. La vitesse du *Virtuose*, \dot{X}_v , exprimée en fonction de la force à renvoyer, $F_v = F_s - B\dot{X}_v$, vaut d'après 3.32 :

$$\dot{X}_v = \sqrt{\frac{2}{b}} U + \frac{F_v}{b} \quad (3.35)$$

En remplaçant F_v par sa valeur, l'équation précédente devient :

$$\dot{X}_v = \sqrt{\frac{2}{b}} U + \frac{F_s}{b} - \frac{B}{b} \dot{X}_v \quad (3.36)$$

L'expression finale de la vitesse décodée vaut donc :

$$\dot{X}_v = \frac{1}{B+b} \left(\sqrt{2b} U + F_s \right) \quad (3.37)$$

Le codage de la force, à retourner à l'utilisateur, provient de la définition de \mathbf{V} . Ainsi, d'après 3.34, l'onde retour s'écrit :

$$\mathbf{V} = \mathbf{U} + \sqrt{\frac{2}{b}}(\mathbf{F}_s - B\dot{\mathbf{X}}_v) \quad (3.38)$$

Soit, en remplaçant l'expression de la vitesse du *Virtuose* par celle décodée et exprimée à l'équation 3.37, la variable retour s'écrit :

$$\mathbf{V} = \mathbf{U} + \sqrt{\frac{2}{b}} - \sqrt{\frac{2}{b} \frac{B}{b+B}} \left(\sqrt{2b}\mathbf{U} + \mathbf{F}_s \right) \quad (3.39)$$

L'expression finale de la variable qui encode l'effort de la simulation en un flux énergétique s'écrit alors :

$$\mathbf{V} = \left(1 - \frac{2B}{b+B} \right) \mathbf{U} + \left(\sqrt{\frac{2}{b}} - \sqrt{\frac{2}{b} \frac{B}{b+B}} \right) \mathbf{F}_s \quad (3.40)$$

Vérification. La validité de la commande peut être prouvée en mode libre et bloqué, c'est-à-dire pour des valeurs de l'amortisseur B nulles et infinies. En supposant B nulle, l'expression de l'onde retour s'écrit alors :

$$\mathbf{V}|_{B=0} = \mathbf{U} + \sqrt{\frac{2}{b}}\mathbf{F}_s \quad (3.41)$$

Cette condition est vérifiée. En supposant que B tende vers l'infini, le flux \mathbf{V} de retour doit tendre vers un mode bloqué.

$$\begin{aligned} \mathbf{V}|_{B \rightarrow \infty} &= (1-2)\mathbf{U} + \left(\sqrt{\frac{2}{b}} - \sqrt{\frac{2}{b}} \right) \mathbf{F}_s \\ &= -\mathbf{U} \end{aligned} \quad (3.42)$$

La variable retour \mathbf{V} est alors égale à l'opposé de la variable entrante, bloquant par conséquent le système. La rigidité du blocage dépend des constantes du système, représentées par b et, dans une moindre mesure, par B ainsi que du délai τ présent dans la communication entre le codage de \mathbf{V} et son décodage.

3.3 Résultats

Les résultats présentés dans ce paragraphe montrent l'influence du facteur d'amortissement B sur la stabilité de la commande mise en place.

3.3.1 Faible coefficient d'amortissement

La figure 3.10 montre la réponse de l'interface haptique pour un coefficient d'amortissement B nul et pour des coefficients en effort et en déplacement fixés, $K_D = 2.10^{-9}$ et $K_f = 5.10^7$, lors d'une simulation de *docking* biotine-streptavidine. L'énergie provenant de la simulation n'est pas dissipée tout au long de la chaîne de commande, et le seul facteur d'amortissement existant est b .

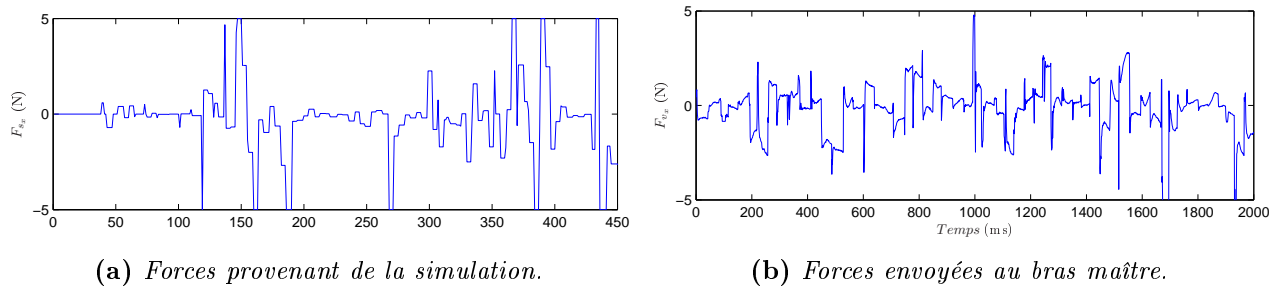


Figure 3.10 - Comportement de la force d'interaction biotine-streptavidine selon l'axe \vec{x} . Les paramètres de simulation utilisés sont un facteur d'amortissement B nul, un facteur de déplacement $K_D = 2.10^{-9}$ et un facteur d'effort $K_f = 5.10^7$.

L'amortissement faible, visible entre les deux figures 3.10(a) et 3.10(b), provient donc de la constante de stabilisation b et du manque de transparence présent à cause du retard. La simulation de *docking* est réalisable, la biotine étant à la fin de celle-ci introduite dans le site actif de la molécule de streptavidine. Cependant, le profil de force engendré et à ressentir (figure 3.10(b)) varie toujours trop pour garantir une information pertinente. Il n'est pas possible, à partir de ces coefficients, de discriminer les géométries possédant la meilleure affinité pour le site actif.

3.3.2 Coefficient d'amortissement élevé

Avec l'augmentation du facteur B , les forces envoyées à l'interface haptique paraissent plus stables que dans le cas précédent. Cependant, il convient d'analyser avec attention les données de la figure 3.11 page suivante.

Le facteur d'amortissement B étant élevé ($B = 50$), la variation des forces induites par la simulation n'a que peu d'effets sur le ressenti, celui-ci étant majoritairement contraint par la viscosité introduite. Cet ajout de viscosité impose à l'utilisateur de faibles déplacements. Le simulateur moléculaire possède donc le temps d'évaluer l'énergie d'interaction et de calculer les efforts à retourner, sans que les rayons de van der Waals des atomes ne s'interpénètrent. Les faibles déplacements engendrent, de fait, une faible variation des efforts.

Les deux graphiques des figures 3.11(a) et 3.11(b) page suivante ne paraissent pas présenter le même résultat. Cependant, une forte viscosité ambiante empêche tout mouvement brusque de l'opérateur, limitant ainsi la forte variation des efforts. Une forte viscosité revient alors à imposer un faible déplacement. Cependant, les efforts ressentis en retour correspondent à cet amortissement et peu à l'action des forces intermoléculaires. Les forces représentées sur la figure 3.11(b) page suivante représentent donc majoritairement l'effet de la viscosité.

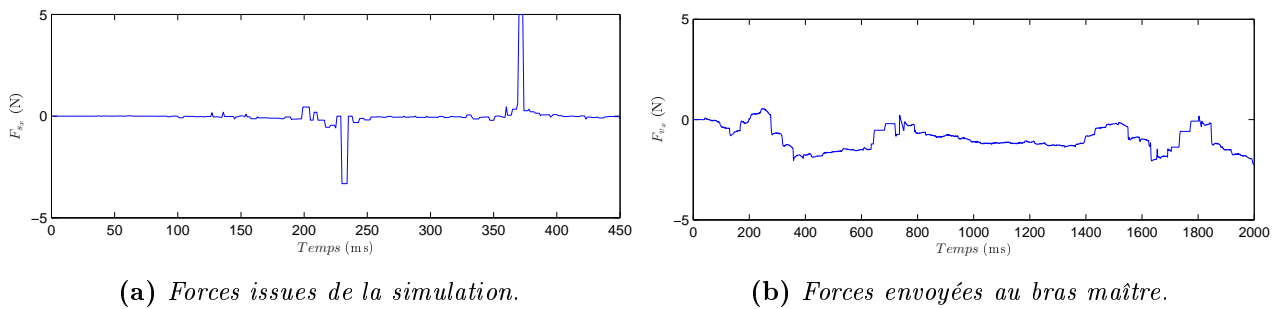


Figure 3.11 - Comportement de la force d'interaction biotine-streptavidine selon l'axe \vec{x} . Les paramètres de simulation utilisés sont un facteur d'amortissement $B = 50$, un facteur de déplacement $K_D = 2.10^{-9}$ et un facteur d'effort $K_f = 5.10^7$.

Avec l'ajout de viscosité dans la commande, le *docking* moléculaire est réalisable. Les efforts d'interaction varient peu, mais le ressenti haptique n'est pas transparent. Il reflète à la fois la viscosité et les efforts qu'il n'est alors plus possible de différencier.

3.4 Conclusion

La commande, assortie d'un amortisseur virtuel couplé aux variables d'onde, gère les retards de communication et amortit la forte variabilité des efforts d'interaction moléculaire. L'ajout de viscosité dans la commande impose de faibles déplacements au ligand. Le profil des efforts ainsi engendré est moins oscillant que dans le cas où une viscosité est nulle. Cependant, cette adaptation se fait au détriment de la transparence. Le ressenti de l'opérateur est l'addition de la viscosité et de la force engendrée par le déplacement du ligand. Si l'objectif est de réaliser un *docking* en utilisant l'haptique comme accélérateur de simulation, alors le schéma proposé ci-dessus garantit la faisabilité de l'opération. En revanche, s'il s'agit de discriminer l'affinité d'une géométrie d'une molécule avec une autre, la faible précision apportée par la viscosité n'aide pas à résoudre le problème posé.

Les deux cas extrêmes ont été testés, l'un sans viscosité, l'autre avec une viscosité importante. Il conviendrait d'effectuer des tests plus poussés, notamment sur un éventail de valeurs pouvant être attribuées à B . Un réglage de ce gain en cours de manipulation pourrait alors être utile. Il pourrait, en effet, être augmenté si le changement conformationnel de la protéine induit une forte variabilité des efforts. Au contraire, si les changements de conformation sont faibles, et les efforts calculés plus facilement interprétables, il pourrait être diminué afin d'accroître la sensation du retour d'effort.

Cependant, si l'objectif de la transparence est recherché, afin de réellement comprendre la surface d'énergie du site actif, la commande ainsi conçue doit donc évoluer. Elle doit à nouveau gérer les retards et favoriser un réel travail sur le ressenti haptique, c'est-à-dire sans l'altérer. L'évolution proposée consiste à modifier le simulateur moléculaire afin que les variables d'onde lui soient compréhensibles. Ce n'est plus, par conséquent, la position de l'interface qui est envoyée, mais sa vitesse. Le schéma classique des variables d'onde est alors applicable.

4 Variables d'onde sans amortissement

Le fait d'appliquer le schéma classique des variables d'onde nécessite une modification du simulateur moléculaire. En effet, ce n'est plus la position qui est directement transmise à la simulation, mais la vitesse de l'interface. Son intégration doit d'abord être réalisée puis le suivi de la trajectoire du *Virtuose* implémenté. Dans une dernière étape, le retour d'effort sera présenté.

4.1 Simulateur moléculaire

Le simulateur moléculaire présenté au chapitre précédent, utilise la variation de position du *Virtuose* afin de transmettre, *via* un coefficient de déplacement, cette variation pour l'affichage et le calcul de l'énergie d'interaction. Cependant, les variables d'onde sont basées sur un calcul de vitesse. Les codeurs optiques du *Virtuose* permettent l'envoi de ces données. Celles-ci doivent être interprétées du côté de la simulation.

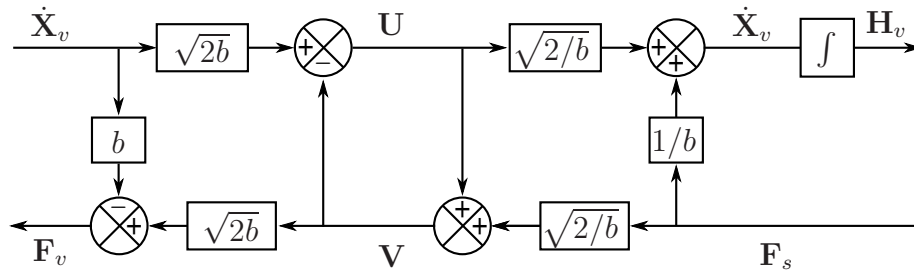


Figure 3.12 - Schéma bloc des variables d'onde intégrées après leur décodage.

Le schéma de commande est donc modifié, et devient celui qui est présenté à la figure 3.12.

4.1.1 Intégration de la vitesse

La vitesse $\dot{\mathbf{X}}_v$ (de dimension 6) calculée au niveau de la poignée du *Virtuose* peut s'écrire sous la forme d'une matrice homogène \mathbf{T}_v . Son expression matricielle fait intervenir la vitesse linéaire du *Virtuose* $\dot{\mathbf{X}}$ et la vitesse angulaire $\boldsymbol{\Omega}$ de l'interface. Elle s'écrit :

$$\mathbf{T}_v = \begin{bmatrix} \boldsymbol{\Omega} & \dot{\mathbf{X}} \\ \mathbf{0} & 1 \end{bmatrix} \quad (3.43)$$

La matrice $[\mathbf{T}_v]$ définit la matrice homogène en vitesse avec sa matrice $\boldsymbol{\Omega}$ antisymétrique, notée $[\boldsymbol{\Omega}]$ (équation 3.44).

$$[\mathbf{T}_v] = \begin{bmatrix} [\Omega] & \dot{\mathbf{X}} \\ \mathbf{0} & 1 \end{bmatrix}, \quad \text{avec } [\Omega] = \begin{bmatrix} 0 & -\Omega_3 & \Omega_2 \\ \Omega_3 & 0 & -\Omega_1 \\ -\Omega_2 & \Omega_1 & 0 \end{bmatrix}, \quad \text{et } \dot{\mathbf{X}} = \begin{bmatrix} \dot{X}_1 \\ \dot{X}_2 \\ \dot{X}_3 \end{bmatrix} \quad (3.44)$$

\mathbf{T}_v peut s'exprimer en fonction des matrices homogènes en position \mathbf{H}_v du *Virtuose* [Murray 94] :

$$[\mathbf{T}_v] \approx \dot{\mathbf{H}}_v \mathbf{H}_v^{-1} \quad (3.45)$$

La discrétisation de l'équation 3.45 peut s'écrire sous la forme :

$$[\mathbf{T}_v] \mathbf{H}_{v_k} = \frac{1}{t} [\mathbf{H}_{v_{k+1}} - \mathbf{H}_{v_k}] \quad (3.46)$$

où t représente la période d'échantillonnage et k l'itération du processus. La position de l'interface à l'itération $k + 1$ s'écrit, d'après l'équation 3.46 :

$$\mathbf{H}_{v_{k+1}} = [\mathbf{1} + t[\mathbf{T}_v]] \mathbf{H}_{v_k} \quad (3.47)$$

Cependant, lors de la mesure de la vitesse, celle-ci apparaît très bruitée. Une dérive en position est observée. De plus, l'intégration provoque une dérive sur la position calculée. Au cours de la simulation, le *Virtuose* doit, par conséquent, être recalé régulièrement afin d'annihiler cet effet.

4.1.2 Expression de la position du ligand

La vitesse de la poignée (outil O) du *Virtuose*, exprimée dans sa base B , s'écrit ${}^B\dot{\mathbf{X}}_O$. C'est la mesure enregistrée par les codeurs optiques. Enfin, la vitesse du ligand L , exprimée dans la fenêtre graphique F du simulateur, s'exprime, quant à elle, comme ${}^F\dot{\mathbf{X}}_L$.

De l'équation 3.47, il est possible d'écrire, en fonction des repères :

$${}^F(\mathbf{H}_{v_{k+1}})_L = [\mathbf{1} + t {}^F(\mathbf{T}_v)_L] {}^F(\mathbf{H}_{v_k})_L \quad (3.48)$$

Or, l'expression de la vitesse du ligand dans la base, en fonction de la vitesse du *Virtuose*, s'écrit :

$${}^B\mathbf{T}_O = {}^B \mathbf{N}_F {}^F\dot{\mathbf{X}}_L \quad (3.49)$$

$$\text{avec } {}^B\mathbf{N}_F = \begin{bmatrix} {}^B\mathbf{R}_F & {}^B[\mathbf{X}]_F & {}^B\mathbf{R}_F \\ \mathbf{0} & {}^B\mathbf{R}_F & \end{bmatrix} \text{ et } {}^B[\mathbf{X}]_F = \begin{bmatrix} 0 & -X_3 & X_2 \\ X_3 & 0 & -X_1 \\ -X_2 & X_1 & 0 \end{bmatrix} \quad (3.50)$$

${}^B[\mathbf{X}]_F$ étant la position du repère de la base du *Virtuose* par rapport à la fenêtre de simulation, dont les composantes selon les trois axes (\vec{x} , \vec{y} , \vec{z}) sont décrites respectivement par ${}^B[\mathbf{X}]_F = [X_1 \ X_2 \ X_3]^t$.

En considérant que le *Virtuose* est confondu avec le centre de la fenêtre de simulation (${}^B[\mathbf{X}]_F = \mathbf{0}$), ${}^B\mathbf{N}_F$ s'exprime de la manière suivante :

$${}^B\mathbf{N}_F = \begin{bmatrix} {}^B\mathbf{R}_F & \mathbf{0} \\ \mathbf{0} & {}^B\mathbf{R}_F \end{bmatrix} \quad (3.51)$$

Ainsi, la vitesse du ligand exprimée dans la fenêtre s'écrit :

$${}^F\dot{\mathbf{X}}_L = {}^F\mathbf{N}_B {}^B\dot{\mathbf{X}}_O \quad (3.52)$$

${}^F\dot{\mathbf{X}}_L$ est exprimée en une matrice homogène ${}^F\mathbf{T}_L$. La nouvelle position du ligand, à l'instant $k+1$, est donc calculée en reportant sa vitesse, exprimée à l'équation 3.52 sous forme de matrice homogène, dans l'équation 3.48, en sachant qu'à l'instant initial :

$${}^F(\mathbf{H}_{v_0})_L = \begin{bmatrix} \mathbf{1} & \mathbf{X}_{cdm} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \quad (3.53)$$

où \mathbf{X}_{cdm} est la position du centre de masse du ligand, exprimée dans la fenêtre du simulateur. Le déplacement de corps rigide correspondant s'écrit :

$${}^F(\mathbf{H}_{v_{k+1}})_L = [\mathbf{1} + t[{}^F\mathbf{T}_L]]^F(\mathbf{H}_{v_k})_L \quad (3.54)$$

La consigne interprétée par la simulation est donc ${}^F(\mathbf{H}_{v_{k+1}})_L$. Cependant, lors de la simulation, le ligand est flexible. Le déplacement de corps solide, précédemment imposé, doit donc être modifié par la variation de position des atomes induite par la minimisation de l'énergie. En supposant que le simulateur calcule une matrice homogène en position ${}^F(\mathbf{H}_{MOE_{k+1}})_L$, le déplacement total du ligand dans la fenêtre de simulation vaut donc :

$${}^F(\mathbf{H}_{Total_{k+1}})_L = {}^F(\mathbf{H}_{v_{k+1}})_L {}^F(\mathbf{H}_{MOE_{k+1}})_L \quad (3.55)$$

Cette dernière matrice est appliquée au ligand. Ainsi, le déplacement de l'interface engendre un déplacement de corps rigide et les nouvelles positions du ligand sont actualisées une fois ce déplacement effectué. La méthode de minimisation « asynchrone » énoncée au chapitre précédent est ici retrouvée. La dérive en position observée lors de la manipulation peut être amoindrie par des recalages successifs de l'interface haptique.

4.2 Analyse des résultats

La figure 3.13 présente la réponse du bras maître (figure 3.13(b)) au regard de l'excitation engendrée par la simulation de *docking* (figure 3.13(a)).

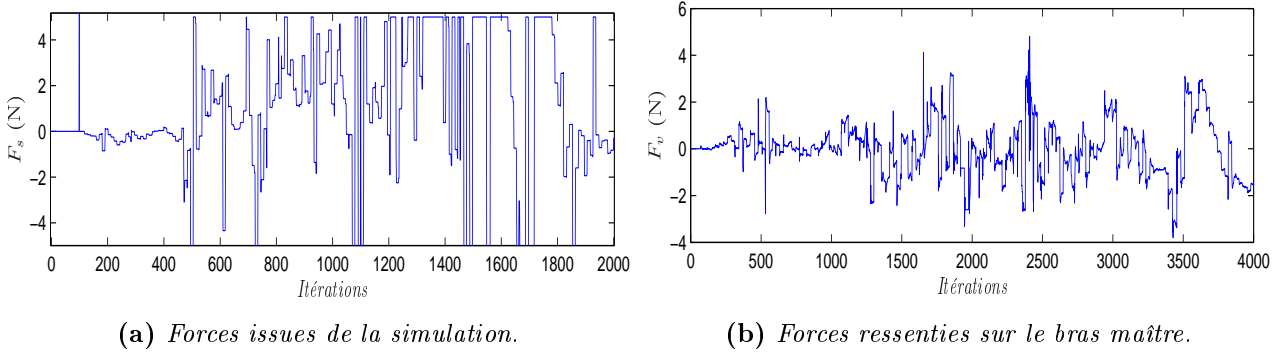


Figure 3.13 - Comportement de la force d'interaction biotine-streptavidine selon l'axe \vec{x} . Les paramètres de simulation utilisés sont un facteur de déplacement $K_D = 1.10^{-9}$ et un facteur d'effort $K_f = 5.10^7$.

Les forces issues de la simulation et interprétées dans le monde macroscopique avec le coefficient K_f sont saturées à 5 N afin de protéger l'interface haptique d'efforts trop violents (figure 3.13(a)). Le coefficient en effort ainsi appliqué ne fait pas correspondre la force maximale atteinte au cours de la manipulation à l'effort maximal admissible par l'interface haptique. Une meilleure résolution des forces est obtenue lors du déplacement du ligand au sein du site actif de la protéine. Mais ce choix présente l'inconvénient majeur de provoquer une saturation de l'interface.

Les forces provenant de la simulation sont atténuées. Le premier filtrage provient de celui des mouvements de l'opérateur afin que l'onde retour ne se propage pas à nouveau vers la simulation. La deuxième atténuation provient de la raideur équivalente du couplage, qui n'est pas idéale mais dépend du retard et des fréquences de coupure du filtre passe-bas. Ces deux particularités provoquent une atténuation des valeurs de consigne renvoyées par la simulation et permettent un travail sur le rendu haptique.

L'avantage de cette méthode est qu'elle autorise un travail futur sur le ressenti macroscopique des efforts d'interaction ligand-protéine. En effet, les efforts appliqués à l'interface ne sont pas sujets à la viscosité introduite. Dès lors, les efforts ressentis sont réellement ceux qui sont calculés. Il est alors possible d'envisager une discrimination de la géométrie des ligands en se basant sur leur affinité. Cependant, il reste à concevoir une méthode pour le ressenti des efforts moléculaires. De simples facteurs d'homothétie, fixés en début de simulation, n'autorisent pas à effectuer, par exemple, des déplacements faibles et précis dans le site actif pour un retour d'effort stable, puis d'effectuer de grands déplacements pour sortir de celui-ci.

Les forces envoyées à l'interface haptique par la méthode classique des variables d'onde ne présentent pas la même variation que celles de la simulation. Il est possible d'envisager, en aval de ce traitement, un travail sur les coefficients en déplacement et en effort. Leur effet étant atténué par l'amortisseur virtuel de la loi de commande précédente.

5 Conclusion

Les variables d'onde, étudiées dans ce chapitre, rendent stable une commande dans laquelle interviennent des retards. Les retards de calculs sont alors assimilés à des retards de communication, rendant stable le couplage. L'avantage majeur des variables d'onde, dans le cas présent, provient de leur faible transparence. Celle-ci est atténuée en présence d'un filtrage des données, elle est fonction du retard de communication et de la valeur donnée à son impédance caractéristique. Cette faible transparence atténue la forte variation des efforts d'interaction.

Le premier couplage étudié propose ainsi de n'utiliser les variables d'onde que comme un amortisseur virtuel. Celui-ci atténue l'amplitude des efforts. Le résultat ainsi obtenu réalise un *docking* d'un ligand flexible dans un site actif de protéine flexible. Les efforts interatomiques renvoyés à l'interface haptique sont atténués proportionnellement à un coefficient B , réglable par l'utilisateur. Celui-ci engendre de faibles déplacements s'il est élevé et un retour d'effort stable. Mais la discrimination géométrique des ligands par la recherche d'affinité n'est pas un objectif atteignable. Le *docking* est possible mais les efforts ressentis ne rendent pas compte de la réalité des interactions.

Le deuxième couplage a pour vocation d'éliminer ce problème. En effet, il filtre et atténue les efforts par les seuls coefficients présents dans sa définition, c'est-à-dire l'impédance caractéristique, le temps de communication et la valeur de la pulsation de coupure du filtre passe-bas. Les consignes envoyées depuis la simulation vers l'interface ne sont pas altérées. Le ressenti des efforts reflète donc dans ce cas la réelle interaction entre le ligand et le site actif. La recherche des zones attractives ou répulsives peut s'effectuer. Cependant, les efforts interprétés par l'interface haptique sont encore très bruités. L'ajout d'un autre filtre passe-bas en sortie de simulateur pourrait être une solution pour atténuer cette variation.

La forte amplitude des efforts, ainsi que leur haute fréquence, peuvent être dues à leur définition jusqu'alors utilisée. La dérivation spatiale de l'énergie n'est qu'une approximation grossière des efforts, d'où pourraient provenir les instabilités. Une autre méthode pour la conversion de l'énergie en efforts doit donc être envisagée.

L'approche finale étudiée dans le dernier chapitre propose de modéliser l'énergie du site actif, afin de déterminer analytiquement les efforts agissant sur le ligand. Ce modèle interagirait alors directement avec l'interface haptique et s'affranchirait des retards de simulation.

L'approche consiste à créer un modèle d'énergie dont les paramètres dépendent de la position, de l'orientation, de l'interface haptique et de l'énergie calculée. Les forces et les moments sont calculés par dérivation analytique. Ce champ interagira, dès lors, directement avec l'interface haptique, garantissant une réponse en temps réel. Ce nouveau modèle, explicité dans le chapitre suivant, rend indépendant le couplage du choix du champ de force.

ESTIMATION DES EFFORTS À PARTIR D'UN MODÈLE D'ÉNERGIE PARAMÉTRABLE

Des chapitres précédents, il ressort notamment qu'en règle générale, les systèmes de simulation moléculaire sont basés sur l'énergie. Le terme « champ de force » est abusivement utilisé et désigne en fait un ensemble d'énergies d'interaction. Le système mis en place utilise un champ de force pour la modélisation des interactions et un processus de minimisation pour garantir la stabilité énergétique des conformations. Se pose alors la question du ressenti des forces d'interaction moléculaire. L'énergie calculée par le simulateur est la somme des composantes de l'énergie selon chaque direction. Le scalaire ainsi calculé ne peut être décomposé selon des axes précis ; les diverses composantes de l'énergie ne sont alors plus accessibles.

Le présent chapitre a pour objet de pallier le manque d'information dans les données, fournies par le simulateur. La méthode mise en place est capable de fournir forces et moments pour tous les types d'interface haptique et tous les couplages à ces échelles. Elle s'affranchit du type de champ de force utilisé. Il est alors possible d'utiliser la méthode décrite ci-dessous pour tout type de champ de force et tout type d'interface haptique. Elle présente notamment l'avantage de se substituer temporairement au simulateur, rendant possible un couplage en temps réel entre l'interface et le modèle créé pendant que le simulateur, fonctionnant à une cadence moindre, enrichit le modèle d'interaction.

1 Estimation directe de la force par un observateur

1.1 Introduction

La simple dérivation échantillonnée n'est pas une bonne approximation de la dérivation analytique. Le problème alors posé est de savoir comment une information de direction peut être

gardée lors du calcul des forces. En effet, le fait d'utiliser l'énergie fournie par le simulateur fait perdre l'information de direction présente dans le champ de force. L'objectif, par la construction d'un observateur, est d'essayer de retrouver cette information.

A l'itération k , sont supposées connues \mathbf{F}_k , la force extérieure appliquée au centre de masse du ligand, E_k l'énergie d'interaction, et \mathbf{X}_k la position du ligand. A l'itération $k + 1$, E_{k+1} et \mathbf{X}_{k+1} sont connus. L'objectif de la méthode décrite au paragraphe suivant est donc d'estimer \mathbf{F}_{k+1} connaissant \mathbf{F}_k par une méthode d'optimisation à un pas. La méthode présentée ici ne considère que le calcul des forces. Les moments peuvent être obtenus en remplaçant les variations de position par des variations angulaires.

La contrainte de l'optimisation est de minimiser l'écart entre la variation d'énergie et la variation des forces estimées correspondante.

1.2 Méthode

La manipulation du ligand se fait avec des déplacements faibles afin d'éviter un gradient d'énergie trop élevé. L'hypothèse initiale de la méthode consiste à considérer un faible gradient en force. Cette hypothèse est en contradiction avec les études menées jusqu'alors. L'observateur peut donc ne pas estimer correctement ces gradients. L'hypothèse des faibles gradients de forces s'écrit :

$$\mathbf{F}_{k+1} = \mathbf{F}_k + \delta\mathbf{F} \quad (4.1)$$

où $\delta\mathbf{F}$ est minimal.

Considérer un faible gradient de force revient à minimiser $\delta\mathbf{F}$ dans l'équation 4.1, tel que :

$$\begin{aligned} E_{k+1} - E_k &= \delta E \\ &= \delta\mathbf{X}^t \mathbf{F}_{k+1} \end{aligned} \quad (4.2)$$

Le Lagrangien d'une optimisation sous contraintes s'écrit comme la somme d'une fonction de coût par une pondération de la contrainte. Le facteur de pondération est un multiplicateur de Lagrange :

$$\mathcal{L} = \text{coût} + \lambda \cdot \text{contrainte}$$

L'optimisation d'une fonction minimise le Lagrangien par rapport à ses paramètres. Dans le cas présent, le paramètre à estimer est \mathbf{F}_{k+1} . Minimiser le Lagrangien par rapport à ce paramètre implique d'annuler sa dérivée. Dans un premier temps, il convient d'écrire le problème d'optimisation. Celui-ci est représenté par l'équation 4.3 :

$$\mathcal{L} = \frac{1}{2} \|\mathbf{F}_{k+1} - \mathbf{F}_k\|^2 + \lambda (\delta\mathbf{X}^t \mathbf{F}_{k+1} - \delta E) \quad (4.3)$$

avec λ un scalaire.

La dérivée du Lagrangien par rapport au paramètre à optimiser s'écrit :

$$\frac{\partial \mathcal{L}}{\partial \mathbf{F}_{k+1}} = 0 \quad (4.4)$$

Soit :

$$(\mathbf{F}_{k+1} - \mathbf{F}_k) + \lambda \delta \mathbf{X} = 0 \quad (4.5)$$

Cette dernière équation se met sous la forme :

$$\mathbf{F}_{k+1} = \mathbf{F}_k - \lambda \delta \mathbf{X} \quad (4.6)$$

En combinant l'équation 4.6 avec l'équation 4.2, il advient que :

$$\delta \mathbf{X}^t \mathbf{F}_k - \lambda \|\delta \mathbf{X}\|^2 = \delta E \quad (4.7)$$

Finalement, l'expression de λ est obtenue :

$$\lambda = - \frac{\delta \mathbf{X}^t \mathbf{F}_k - \delta E}{\|\delta \mathbf{X}\|^2 + \varepsilon_{obs}} \quad (4.8)$$

où ε_{obs} est une constante positive, introduite afin d'éviter une singularité si l'interface haptique est immobile.

En connaissant l'expression de λ (équation 4.8), et d'après l'expression 4.5, l'estimation de la force à l'instant $k + 1$ s'écrit :

$$\mathbf{F}_{k+1} = \mathbf{F}_k - \left(\frac{\delta \mathbf{X}^t \mathbf{F}_k - \delta E}{\|\delta \mathbf{X}\|^2 + \varepsilon_{obs}} \right) \delta \mathbf{X} \quad (4.9)$$

Développer l'expression 4.9 conduit à l'équation suivante :

$$\mathbf{F}_{k+1} = \left(1 - \frac{\delta \mathbf{X} \delta \mathbf{X}^t}{\|\delta \mathbf{X}\|^2 + \varepsilon_{obs}} \right) \mathbf{F}_k + \frac{\delta E}{\|\delta \mathbf{X}\|^2 + \varepsilon_{obs}} \delta \mathbf{X} \quad (4.10)$$

Par cette méthode d'optimisation, la force d'interaction entre le ligand et la protéine est estimée. Estimer les moments suppose de remplacer la variation de position par une variation de rotation.

1.3 Simulation

La figure 4.1 présente les résultats obtenus pour l'estimation des forces en tout point d'un volume d'énergie. Le déplacement au sein du volume est aléatoire mais respecte l'hypothèse des faibles déplacements afin d'engendrer un faible gradient d'énergie.

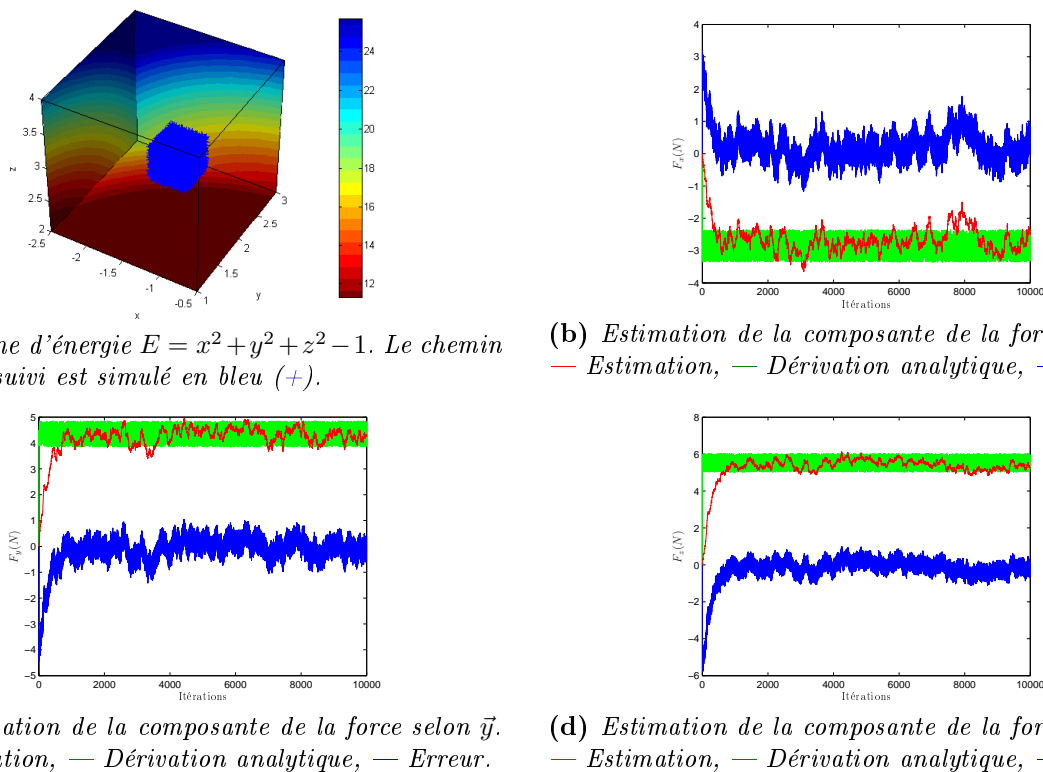


Figure 4.1 - Comparaison de la dérivation analytique selon les axes \vec{x} , \vec{y} et \vec{z} d'un potentiel donné, $E = x^2 + y^2 + z^2 - 1$, avec l'estimation de la dérivée par l'utilisation de projecteurs.

La force estimée ne présente pas les instabilités observées par la méthode utilisée jusqu'alors, qui est une dérivation simple de l'énergie selon le déplacement de l'interface. Mais l'observateur n'estime pas une solution comparable à la dérivée analytique. L'estimation reste une approximation de celle-ci et est très dépendante du chemin suivi ainsi que du gain ε_{obs} (figure 4.2 page ci-contre).

S'il est faible, celui-ci ajoute à chaque pas de temps la totalité de l'estimation. En revanche, s'il est élevé, la contribution de la force précédemment calculée s'en trouve amoindrie. Ce paramètre influence donc sur la convergence de la méthode. L'erreur d'estimation de la force est importante au regard des valeurs de la dérivée analytique, d'autant plus que l'estimation ne semble pas tendre vers une valeur de convergence.

Bien que plus fiable que la dérivation numérique, ce calcul n'est pas précis et sa convergence n'est pas toujours assurée. L'observateur calculé ne peut donc être retenu pour le calcul des forces de l'interaction ligand-protéine, car une divergence de l'algorithme provoquerait l'explosion du système de téléopération. Si l'observateur est très dépendant des gains ainsi que du chemin suivi, la divergence de la méthode peut être rapide lorsque la protéine est soumise à la minimisation de son énergie. Dans ce cas, l'hypothèse des faibles gradients d'énergie n'est pas assurée, le changement de conformation pouvant engendrer de fortes variations de ceux-ci.

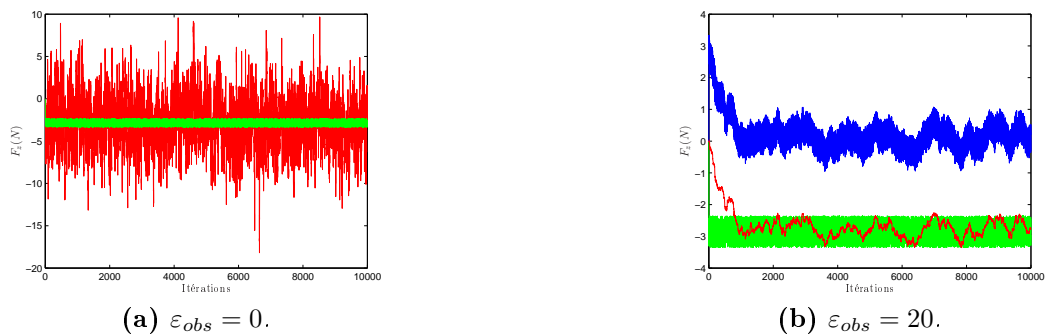


Figure 4.2 - Effet de la variation du gain ε sur la précision de l'estimation. — Estimation, — Dérivation analytique, — Erreur.

Cependant, une évolution non testée dans le cadre de ces travaux consisterait à faire évoluer la méthode d'optimisation vers des méthodes plus précises, par exemple à plusieurs pas, pour la détermination de la force. De plus, il faudrait envisager de calculer les moments pour un retour d'effort avec six degrés de liberté. Cette étude n'a pas été faite car la faible convergence de l'estimation laisse supposer que l'estimation de la rotation ne peut pas non plus converger.

1.4 Conclusion

La méthode qui consiste à optimiser une fonction de coût sous une contrainte convertit l'énergie fournie par le simulateur en une force. Cependant, devant sa faible convergence au regard du chemin suivi ou de la variation du gain ε_{obs} , elle ne garantit pas une estimation fiable des forces dans le cadre de la méthode présentée.

La première alternative serait de concevoir un observateur plus précis et d'étudier sa convergence.

La deuxième alternative consisterait à modéliser une fonction paramétrable qui reflète localement l'énergie réelle. Sa dérivée analytique, connue, fournit l'expression des forces et des moments. L'actualisation des paramètres du modèle peut se faire grâce à différentes méthodes d'interpolations comme par exemple celle des moindres carrés ou celle des moindres carrés récursifs.

Une dernière évolution serait de concevoir un maillage volumique précis aux abords du site actif de la protéine et de calculer, en tout point de ce maillage, la dérivée analytique d'une fonction connue possédant des paramètres à estimer qui assurent que celle-ci approche en tout point l'énergie réelle. Cependant, l'état de l'art illustre des problèmes de non-linéarité lorsque le ligand se déplace entre deux points. L'interpolation des forces de van der Waals n'est pas stable. De plus, les capacités mémoires doivent être étendues, et la manipulation reste dépendante de la taille du maillage.

Le paragraphe suivant propose la modélisation de l'énergie par un modèle paramétrable, actualisé grâce à la méthode des moindres carrés.

2 Estimation des efforts à partir d'un modèle d'énergie paramétrable

Le problème que pose l'utilisation d'un simulateur tel que MOE pour le calcul des énergies d'interaction réside dans le fait que seul un scalaire, correspondant à l'énergie d'interaction, est calculé. Le but est donc de modéliser l'énergie, pour une position donnée, par une autre plus simple, dont les paramètres et la dérivée analytique sont connus. En effet, si la dérivée analytique de l'énergie fournie par le simulateur n'est pas accessible, approcher l'énergie, ainsi que sa dérivée par un modèle connu rend possible le calcul des efforts.

La précision du gradient obtenu dépend de la précision du modèle qui approche l'énergie à une position donnée. Plus le modèle sera complexe, plus grande sera la précision du gradient. Cependant, le modèle doit être actualisé dès que la position est modifiée. La méthode d'optimisation choisie doit donc être suffisamment rapide pour ne pas provoquer de divergence de la commande, problématique pour les couplages haptiques, comme cela a été vu *supra*.

2.1 Principe en trois dimensions

2.1.1 Evaluation de l'énergie

A partir d'une surface d'énergie dont les caractéristiques ne sont pas connues, le principe de la méthode consiste à approcher, aux alentours de la position du ligand, cette surface par une fonction paramétrable. Les paramètres qui la constituent sont à estimer. La figure 4.3 page suivante illustre le propos de ce paragraphe.

Le ligand est supposé être à la position d'énergie minimum (★). La figure 4.3(a) page ci-contre représente, par exemple, la surface d'énergie d'interaction entre le ligand et la protéine. La position d'énergie minimum représente alors la position idéale dans laquelle devrait se trouver le ligand à la fin de la manipulation. Les seules valeurs connues sont celles fournies par cette surface, notées $E_{int}(x, y)$.

2.1.1.1 Détermination d'une fonction approchant localement l'énergie

Dans un premier temps, seules les forces d'interaction sont estimées afin de bien comprendre le principe de la méthode. Par la suite, les rotations et les moments seront introduits.

La détermination analytique de la force impose de connaître la fonction qui approche la surface d'énergie calculée. Le puits de potentiel représenté à la figure 4.3 page suivante peut être approché par une fonction de type quadratique. Cependant, cette fonction n'étant pas centrée à l'origine, les paramètres de son minimum $(\hat{x}, \hat{y}, \gamma)$ doivent par conséquent être estimés, comme ceux qui déterminent sa forme (α, β) :

$$\hat{E}(x, y, \hat{x}, \hat{y}, \alpha, \beta, \gamma) = \alpha(x + \hat{x})^2 + \beta(y + \hat{y})^2 + \gamma \quad (4.11)$$

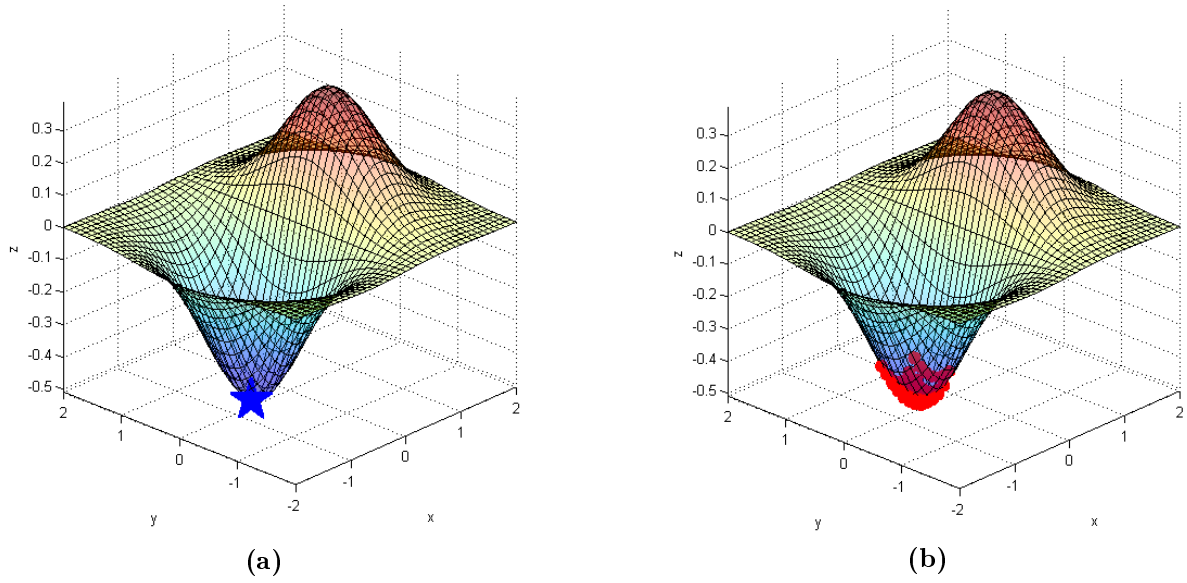


Figure 4.3 - Représentation d'une surface de paramètres inconnus, assimilée à une énergie d'interaction entre un ligand et une protéine, en trois dimensions. **(a)** Les valeurs de l'énergie représentent une surface dont les paramètres ne sont pas connus. ★ Position du ligand. **(b)** • Déplacements aléatoires (autour de la position initiale) du ligand dans le site actif. Les unités en déplacement selon les axes \hat{x} et \hat{y} peuvent être assimilés à des Å. Selon \hat{z} est représentée l'énergie, exprimée en Kcal/mol.

Le choix de la fonction d'estimation $\hat{E}(x, y, \hat{x}, \hat{y}, \alpha, \beta, \gamma)$ doit avoir des caractéristiques proches de la fonction réelle afin d'assurer la convergence de la méthode.

Une fois la fonction d'estimation choisie, une méthode d'optimisation minimise l'erreur entre l'estimation et la fonction réelle. Pour des raisons de simplicité de mise en œuvre, la méthode des moindres carrés a été utilisée. Cependant, elle présente l'inconvénient d'inverser une matrice qui peut être singulière ou de taille importante, influant donc sur la stabilité du système ou le temps de calcul.

2.1.1.2 Méthode des moindres carrés

Les paramètres $\alpha, \beta, \gamma, \hat{x}$ et \hat{y} , qui approchent la surface d'énergie réelle, doivent être estimés. Cette dernière étape, dans le cadre de ce manuscrit, utilise la méthode des moindres carrés. Le but est de minimiser la somme des erreurs ζ suivante :

$$\min_{\Theta} \sum_k \|E_{int}(x, y) - \hat{E}(x, y, \Theta_k)\|^2 = \sum_k \zeta \quad (4.12)$$

où k est l'itération, Θ un vecteur qui représente l'ensemble des paramètres à estimer $\Theta = [\alpha, \beta, \hat{x}, \hat{y}, \gamma]^t$ et ζ est l'erreur d'estimation à chaque itération.

Cependant, l'énergie estimée $\hat{E}(x, y, \Theta_k)$ ne dépend pas linéairement de ses paramètres Θ . Or la minimisation d'un critère des moindres carrés impose que la fonction d'estimation dépende linéairement de ses paramètres.

La linéarisation de la fonction d'estimation par rapport à ses paramètres s'écrit :

$$\widehat{E}(x, y, \Theta) \Big|_{\substack{x=x_k, \\ y=y_k, \\ \Theta=\Theta_{k-1}}} = \widehat{E}(x_k, y_k, \Theta_{k-1}) + \frac{\partial \widehat{E}(x_k, y_k, \Theta_{k-1})}{\partial \Theta} \delta \Theta \quad (4.13)$$

Dans le cas idéal :

$$E_{int}(x, y) = \widehat{E}(x, y, \Theta) \quad (4.14)$$

Le problème 4.12 se met sous la forme :

$$E_{int}(x_k, y_k) = \widehat{E}(x_k, y_k, \Theta_{k-1}) + \frac{\partial \widehat{E}(x_k, y_k, \Theta_{k-1})}{\partial \Theta} \delta \Theta \quad (4.15)$$

Dans l'équation 4.15, l'estimation \widehat{E} est calculée à l'itération k avec les paramètres de l'instant $k - 1$. La résolution du système 4.15 conduit à l'obtention des paramètres $\delta \Theta$ qui minimisent l'équation 4.12.

Finalement, le système 4.15 s'écrit :

$$\frac{\partial \widehat{E}(x_k, y_k, \Theta_{k-1})}{\partial \Theta} \delta \Theta = E_{int}(x_k, y_k) - \widehat{E}(x_k, y_k, \Theta_{k-1}) \quad (4.16)$$

Ce dernier s'écrit alors sous forme matricielle :

$$\begin{bmatrix} \nabla_{\Theta} \widehat{E}(k, \mathbf{X}_k, \Theta_{k-1}) \\ \nabla_{\Theta} \widehat{E}(k-1, \mathbf{X}_{k-1}, \Theta_{k-2}) \\ \vdots \\ \nabla_{\Theta} \widehat{E}(k-m, \mathbf{X}_{k-m}, \Theta_{k-m-1}) \end{bmatrix} \begin{bmatrix} \delta \widehat{x} \\ \delta \widehat{y} \\ \delta \alpha \\ \delta \beta \\ \delta \gamma \end{bmatrix} = \begin{bmatrix} E_{int}(k, \mathbf{X}_k) - \widehat{E}(k, \mathbf{X}_k, \Theta_{k-1}) \\ E_{int}(k-1, \mathbf{X}_{k-1}) - \widehat{E}(k-1, \mathbf{X}_{k-1}, \Theta_{k-2}) \\ \vdots \\ E_{int}(k-m, \mathbf{X}_{k-m}) - \widehat{E}(k-m, \mathbf{X}_{k-m}, \Theta_{k-m-1}) \end{bmatrix} \quad (4.17)$$

où \mathbf{X}_k simplifie l'écriture et correspond au couple de coordonnées (x_k, y_k) .

Plusieurs variantes existent pour la construction de la matrice des gradients par rapport aux paramètres. Dans la méthode présentée ici, les paramètres estimés sont différents sur toutes les lignes. Une variante consiste à modifier la matrice en entier avec les paramètres calculés à l'étape k . Ainsi, toutes les énergies estimées sont actualisées avec le vecteur de paramètres courant. Ce n'est pas le cas pour la suite du manuscrit.

Afin de simplifier l'écriture, les matrices suivantes sont introduites :

$$\mathbf{A}\mathbf{Y} = \mathbf{B} \quad (4.18)$$

avec \mathbf{A} la matrice correspondant au gradient de l'estimée \hat{E} par rapport à ses paramètres Θ , \mathbf{Y} représente le vecteur colonne des paramètres à estimer et \mathbf{B} la matrice de la différence entre le champ réel E_{int} et l'estimation \hat{E} .

A la différence d'une méthode classique des moindres carrés, la linéarisation du problème 4.12 représenté par le système 4.18 conduit à obtenir la matrice \mathbf{Y} représentant non pas les paramètres Θ mais la variation infinitésimale de ses paramètres, soit $\delta\Theta$. Obtenir l'ensemble des paramètres revient alors à inverser la matrice \mathbf{A} du problème 4.18.

2.1.1.2.1 Etude des propriétés de la matrice \mathbf{A}

Inversibilité et déplacements aléatoires. La condition qui assure que le problème a une solution est que \mathbf{A} soit de rang plein. Cette contrainte forte impose que la variation de position du ligand se fasse aléatoirement autour de la trajectoire de consigne (figure 4.3(b) page 109). Lors de la manipulation du ligand selon la trajectoire imposée par l'opérateur et afin d'assurer la convergence de la méthode, l'algorithme qui approche l'énergie réelle par une énergie estimée doit imposer autour de cette trajectoire un mouvement aléatoire du ligand. Celui-ci n'a pas d'effet sur le changement de conformation mais enrichit la matrice \mathbf{A} en évaluant aléatoirement l'énergie.

L'inversion de \mathbf{A} , en considérant qu'elle est bien conditionnée, se fait en calculant sa pseudo-inverse. Les paramètres sont obtenus selon l'équation 4.19 :

$$\mathbf{Y} = \mathbf{A}^+\mathbf{B} \quad \text{avec } \mathbf{A}^+ = (\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t \quad (4.19)$$

Influence de la taille du système. La matrice \mathbf{A} est de dimension $[m \times n]$, où m est un nombre de ligne modifiable et n correspond au nombre de paramètres. Cette dernière dimension est donc fixe. Faire varier m revient à augmenter la taille de la matrice à inverser. Bien que plus précise, la solution n'est obtenue que pour un temps croissant avec la dimension du système. La précision du modèle dépend de la grande taille \mathbf{A} mais les contraintes liées au temps de calcul en imposent une faible, afin de fournir une modélisation de l'énergie en temps réel.

Un avantage non négligeable de choisir un nombre m important réside dans le fait que le résultat, tenant compte d'une « mémoire » plus ample, engendre une plus faible variation d'énergie au cours du temps. Une plus faible valeur provoque une tendance à l'oscillation autour d'une solution idéale. m agit donc comme un filtre moyenneur dont la taille de la fenêtre est la valeur de cette constante.

Un compromis entre précision du modèle et temps de calcul doit donc être trouvé, celui-ci tendant vers une optimisation du temps de calcul.

2.1.1.2.2 Détermination des paramètres \mathbf{Y}

Grâce au calcul de la pseudo-inverse de \mathbf{A} , l'inversion du système admet pour solution la matrice \mathbf{Y} des paramètres pour l'étape k :

$$\mathbf{Y}|_{\Theta_k} = \mathbf{A}|_{x_k, y_k, \Theta_{k-1}}^+ \mathbf{B}|_{x_k, y_k, \Theta_{k-1}} \quad (4.20)$$

Cependant, d'après l'équation 4.16, la matrice \mathbf{Y} représente la variation infinitésimale des paramètres et non pas les paramètres eux-mêmes :

$$\mathbf{Y} = \delta\Theta \quad (4.21)$$

L'excitation étant aléatoire, additionner les paramètres de $\delta\Theta_k$ à ceux de Θ_{k-1} aurait pour conséquence une forte oscillation des valeurs de Θ_k . Afin de remédier à cette oscillation, les paramètres sont actualisés par rapport à un gain G définissable par l'utilisateur :

$$\Theta_k = \Theta_{k-1} + \frac{\mathbf{Y}}{G} \quad (4.22)$$

L'influence du gain G est étudiée au paragraphe suivant.

2.1.1.2.3 Résultats expérimentaux

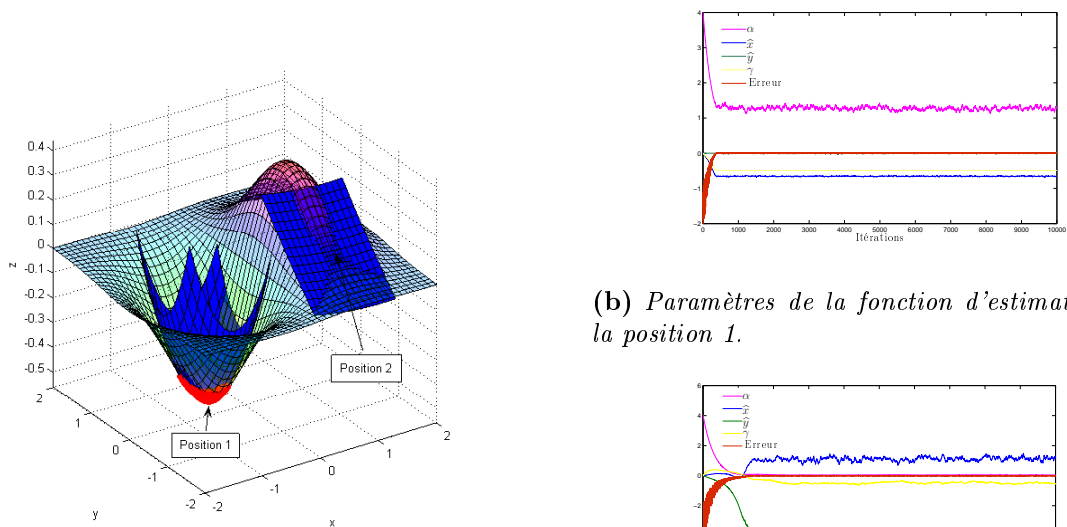
La figure 4.4 page suivante montre les résultats d'une simulation qui approche une surface d'énergie inconnue, en deux positions différentes, par la fonction d'estimation 4.23 :

$$\widehat{E}(x, y, \theta) = \alpha [(x + \widehat{x})^2 + (y + \widehat{y})^2] + \gamma \quad (4.23)$$

A partir de la connaissance des deux positions, l'objectif est d'approcher la surface d'énergie réelle par celle décrite ci-dessus (équation 4.23).

Afin de procéder à l'identification des paramètres, le ligand est soumis à des déplacements aléatoires autour de la position désirée, soit successivement la position 1 et 2 de la figure 4.4 page ci-contre. A la fin de l'estimation, aux deux points considérés, la fonction inconnue peut être approchée par la fonction d'estimation dont les paramètres sont présentés respectivement pour la position 1 et 2 à la figure 4.4(a) et 4.4(b) page suivante.

Position 1. La forme du puits de potentiel présent à la position 1 semble correspondre à la forme d'une surface quadratique. Cette corrélation est visible sur l'identification des paramètres. En effet, ceux-ci ne semblent pas osciller au cours de l'estimation.



(a) Surface d'énergie inconnue, approchée aux positions 1 et 2 par une fonction dont les paramètres sont connus. Les paramètres sont obtenus grâce aux déplacements aléatoires (+)

(b) Paramètres de la fonction d'estimation pour la position 1.

(c) Paramètres de la fonction d'estimation pour la position 2.

Figure 4.4 - Obtention des paramètres de la fonction d'estimation $E = \alpha [(x + \hat{x})^2 + (y + \hat{y})^2] + \gamma$ afin d'approcher une surface d'énergie inconnue en deux positions.

Position 2. A cette position, la forme de la surface inconnue présente un arrondi qu'une surface quadratique ne peut correctement modéliser. Ce manque de précision est visible sur les paramètres d'estimation de la figure 4.4(c). En effet, l'évolution du paramètre \hat{y} est instable et ne présente pas de tendance particulière, bien qu'il oscille autour de la valeur $\hat{y} = -6$. Regarder la figure 4.4(a) à la position 2 confirme ce manque de précision. La surface quadratique coupe la surface inconnue sans réellement l'approcher à cette position. Cependant, l'erreur est tout de même minimum, c'est-à-dire que les paramètres obtenus en fin de simulation, pour la surface quadratique d'estimation, sont ceux qui approchent au plus près de la surface inconnue.

Cette observation confirme le fait que la forme de la fonction d'estimation initiale doit correspondre à l'évolution de la surface d'énergie à approcher au cours du déplacement. Dans le cas de la position 2, pour la figure 4.4(a), une fonction cubique aurait permis de représenter la surface d'énergie de façon plus précise.

Influence de la taille de \mathbf{A} . La figure 4.5 page suivante compare la modélisation de l'énergie pour différentes tailles de la matrice \mathbf{A} . Plus le nombre de ligne m est élevé, plus le profil aléatoire de l'énergie réelle est gardé en mémoire. La conséquence immédiate est donc la restitution d'un profil qui oscille après l'inversion de \mathbf{A} . De plus, s'il existe des erreurs d'initialisation, elles risquent de rendre la convergence très lente voire infructueuse. A l'inverse, un faible nombre de lignes ne garde que très peu d'informations sur la variation du profil d'énergie, rendant ainsi la solution moins oscillante mais la convergence est plus lente.

Une faible valeur de m a pour conséquence non seulement un lissage de la solution mais aussi un gain de temps pour l'inversion de \mathbf{A} . Cette solution est donc à privilégier pour la modélisation de l'énergie réelle.

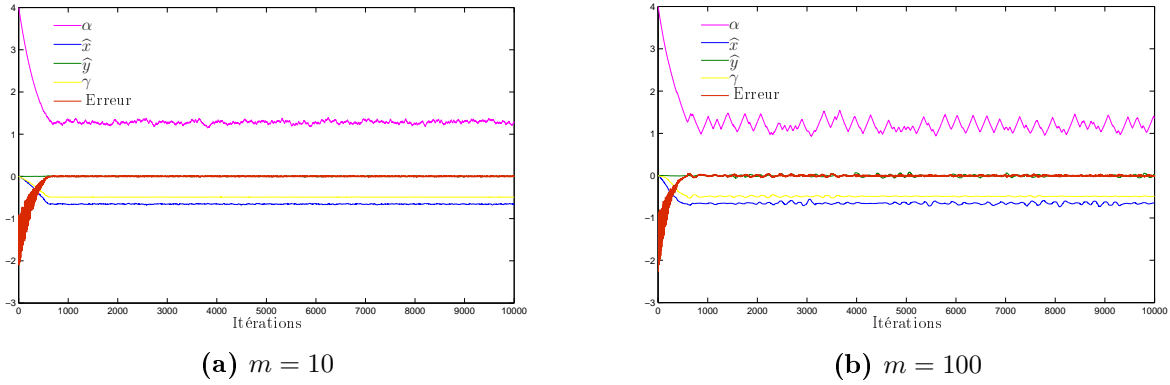


Figure 4.5 - Comparaison de la précision de la modélisation pour différentes tailles de la matrice \mathbf{A} .

Influence du gain G . La figure 4.6 compare les profils des paramètres obtenus pour différentes valeurs du gain G . L'évaluation de l'énergie s'effectuant aléatoirement, la diminution du gain tend à faire osciller les paramètres autour d'une valeur qui est celle de convergence. Cependant, l'algorithme converge rapidement vers la solution. Un gain élevé n'ajoute que très peu d'informations à chaque itération sur la variation des paramètres, assurant ainsi une meilleure convergence mais au prix de beaucoup d'itérations.

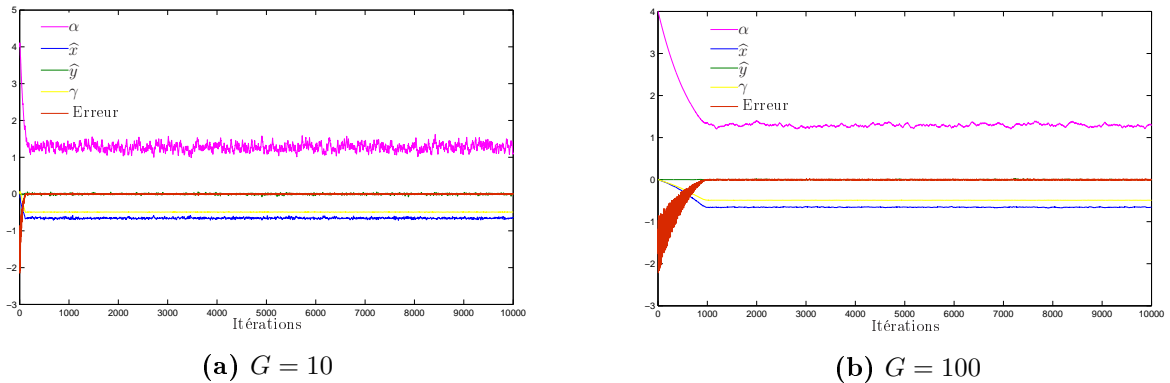


Figure 4.6 - Comparaison de la précision de la modélisation pour différentes valeurs du gain G .

Le besoin d'une information en temps réel pour la modélisation de l'énergie impose le choix d'un faible coefficient G .

Conclusion. De cet exemple simple pour appréhender la modélisation d'une énergie, il ressort que la forme de la fonction d'estimation initiale est importante. Il est possible de choisir comme paramètre d'estimation la puissance de la fonction. Cette approche étant tout de même fortement non linéaire, il serait préférable de mettre en place d'autres techniques avec des paramètres de nature différente. Mais un compromis entre temps de calcul et précision du modèle doit être trouvé. En effet, la modélisation de la surface d'énergie doit se faire suffisamment rapidement pour fournir, en temps réel, une force dérivant de celle-ci. Un nombre minimal de paramètres à estimer doit donc être préféré.

2.1.1.2.4 Conclusion

La méthode des moindres carrés décrite ci-dessus modélise une surface d'énergie, dont les paramètres ne sont pas connus, par une autre énergie à estimer, dont les paramètres à l'étape k sont connus. La démonstration ci-dessus est fondée sur un exemple tridimensionnel simple et fait apparaître l'évaluation de trois matrices :

- \mathbf{A} : matrice $[m \times n]$, qui regroupe à l'étape k pour un ensemble de paramètres connus de l'étape $k - 1$, les n composantes du gradient de l'estimée par rapport à ses paramètres. m est le nombre de lignes de la matrice et doit être déterminé selon la précision désirée.
- \mathbf{Y} : matrice $[n \times 1]$, qui regroupe l'ensemble des variations infinitésimales des paramètres recherchés à l'étape k . Les paramètres de l'étape k sont obtenus par :

$$\Theta_k = \Theta_{k-1} + \mathbf{Y} \quad (4.24)$$

Les paramètres ainsi obtenus estiment la surface pour les nouvelles valeurs de Θ .

- \mathbf{B} : matrice $[m \times 1]$, qui regroupe, à l'étape k , la différence sur m lignes entre le champ réel et le champ estimé déterminé à partir des paramètres Θ_{k-m} .

A partir du modèle d'énergie, il est possible de connaître la valeur de la force au point considéré en dérivant analytiquement le champ estimé. Les valeurs obtenues sont alors une approximation de la valeur des forces pour le champ réel au même point.

2.1.2 De l'énergie modélisée à la force d'interaction

La figure 4.7 page suivante présente le résultat d'une simulation qui calcule les forces associées à l'approximation d'une surface d'énergie inconnue.

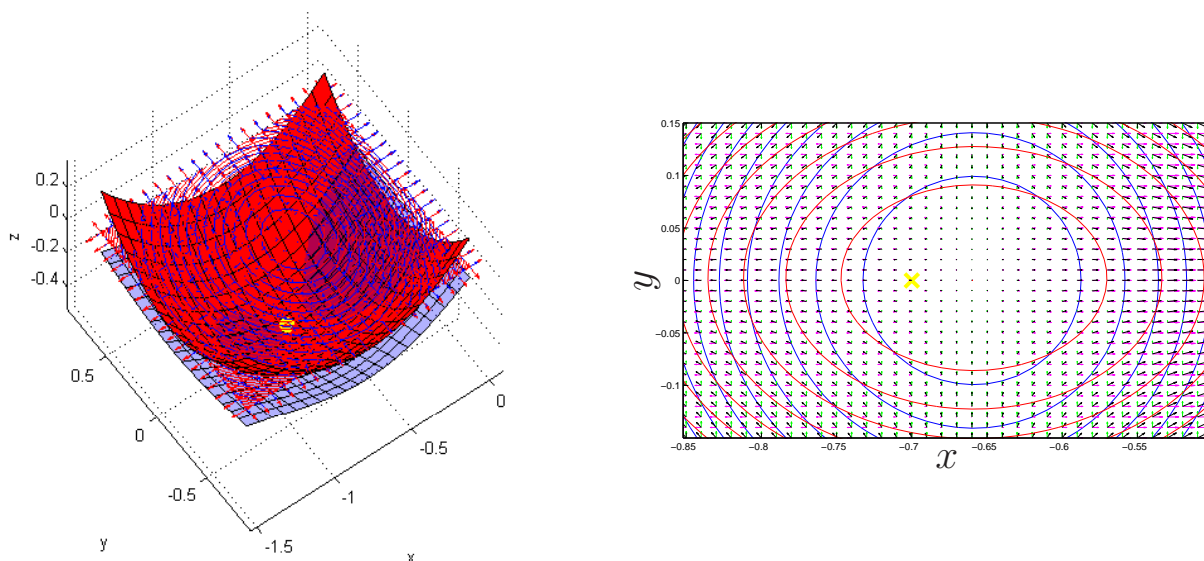
La figure 4.7(a) page suivante représente l'approximation de la surface d'énergie inconnue, en bleu, par la surface dont les paramètres sont obtenus après l'application de la méthode des moindres carrés. En reprenant l'exemple précédent, cette surface est calculée grâce aux paramètres de fin de simulation (étape k) reportés dans l'équation 4.25 :

$$\widehat{E}(x, y, \Theta_k) = \alpha(x + \widehat{x})^2 + \beta(y + \widehat{y})^2 + \gamma \quad (4.25)$$

Les gradients sont calculés analytiquement à partir de l'équation 4.25 en tout point du maillage. L'expression des forces associées à la surface modélisée est obtenue en dérivant l'énergie par la variation de position. Ainsi l'expression des forces, pour la surface modélisée et pour l'exemple considéré, s'écrit :

$$\vec{\nabla}_{(x,y)} \widehat{E} = \begin{cases} \frac{\partial \widehat{E}(x, y, \Theta)}{\partial x} = 2\alpha(x + \widehat{x}) \\ \frac{\partial \widehat{E}(x, y, \Theta)}{\partial y} = 2\beta(y + \widehat{y}) \end{cases} \quad (4.26)$$

Au point de calcul de l'estimation de paramètres de modèle, représenté par une croix (\times) sur la figure 4.7(b) page suivante, l'énergie calculée ainsi que ses gradients approchent la surface à



(a) Approximation de la surface d'énergie représentée en bleu par la surface rouge. Projection de leurs gradients respectifs.

(b) Représentation plane de la différence des gradients des surfaces de la figure (a) selon les axes \vec{x} , \vec{y} et \vec{z} . (\rightarrow) erreur des gradients selon \vec{x} , (\rightarrow) erreur des gradients selon \vec{y} , (\rightarrow) erreur des gradients selon \vec{z} , (\times) position d'estimation des forces.

Figure 4.7 - Approximation de l'énergie par la méthode des moindres carrés et calcul des forces associées.

estimer et ses gradients de façon fiable. En effet, les coefficients calculés le sont pour un ensemble de points aléatoires proches de la position centrale, assurant que sur de faibles déplacements, les forces calculées sont proches de celles qui l'auraient été si l'équation de la surface était connue.

La forme de l'équation de l'estimation de l'énergie doit tout de même être proche de celle qui doit être modélisée. Ainsi, pour quelques déplacements autour d'une position centrale, la surface d'énergie calculée peut approcher une zone plus grande que celle qui est limitée par les déplacements locaux inhérents à la méthode.

2.2 Conclusion

La méthode des moindres carrés décrite dans ce paragraphe approche une surface d'énergie par une fonction dépendant de paramètres qu'elle estime. De cette fonction connue sont déduites les dérivées spatiales correspondant aux forces.

Les gradients étant des dérivées analytiques d'un polynôme, leur définition est donc assurée sur l'ensemble des déplacements du ligand.

Cependant, par la forme de l'estimation choisie, les gradients ne représentent que des forces, car les paramètres de l'estimation ne dépendent que de la position du ligand. Vouloir ressentir des forces et des moments implique qu'il faut modéliser la fonctionnelle inconnue par une énergie dont les paramètres représentent à la fois les déplacements linéaires du ligand et son orientation. Le gradient de l'énergie par rapport à la variation de déplacement représentera alors les forces d'interaction et le gradient de l'énergie par rapport à l'orientation représentera les moments. Cette étude fait l'objet du paragraphe suivant.

3 Modélisation de l'énergie d'interaction ligand-protéine

L'objet de ce paragraphe est la modélisation de l'énergie d'interaction ligand-protéine par une énergie dépendant de paramètres à estimer, afin d'en déduire les forces et les moments d'interaction. La fonction d'estimation, base de la dérivation analytique pour le calcul des efforts, doit donc représenter à la fois les déplacements du ligand et son orientation. Les étapes de la modélisation sont la recherche d'une fonctionnelle pour la translation et la rotation.

3.1 Détermination d'une forme de l'estimation

Les interactions ligand-protéine, simulées dans le vide, sont régies par les énergies électrostatiques et de van der Waals. La forme de ces fonctions, différentes par les paramètres utilisés, ne varie que très peu selon le champ de force considéré. Ainsi, les interactions électrostatiques ont une forme générale en $1/r^2$, r étant la distance d'interaction, et les forces de van der Waals sont modélisées par une forme de Lennard-Jones dont la puissance varie selon le champ. Elles sont, en général, de la forme en $1/r^6$ ou $1/r^{12}$.

Dans un premier temps, afin de ne pas altérer l'optimisation des paramètres lors de leur identification initiale, le ligand ainsi que le site actif de la protéine sont considérés rigides. Ensuite, au cours de la simulation, les gradients de l'énergie d'interaction étant faibles, les paramètres de l'énergie modélisée peuvent évoluer plus lentement que la simulation de *docking*, sans provoquer une divergence de l'algorithme. Les paramètres initiaux doivent cependant modéliser correctement l'énergie d'interaction initiale.

La figure 4.8 page suivante présente le profil de l'énergie d'interaction entre la biotine et la streptavidine pour différents champs de forces. Les deux molécules sont dans leur état minimisé ; l'énergie pour l'état initial est donc minimum.

Des données ainsi calculées, une forme d'énergie à estimer doit être arrêtée. Celle-ci doit faire apparaître deux termes. La dérivée spatiale du premier terme doit approcher les forces d'interaction tandis que le second terme doit fournir une formulation qui, à partir de sa dérivée par rapport à l'orientation du ligand, calcule les moments de l'interaction.

L'énergie à estimer se décompose donc en deux termes :

$$\hat{E}(\mathbf{X}, \mathbf{R}, \Theta) = \hat{E}_{\text{Trans}}(\mathbf{X}, \Theta_{\text{Trans}}) + \hat{E}_{\text{Rot}}(\mathbf{R}, \Theta_{\text{Rot}}) \quad (4.27)$$

où \mathbf{X} est la position du centre de masse du ligand à l'instant considéré, \mathbf{R} la matrice de rotation du ligand et Θ le vecteur des paramètres de l'estimation.

3.1.1 Forme de l'estimation pour la translation

La comparaison des différents profils amène naturellement à orienter le choix de la forme de l'énergie à estimer vers une fonction quadratique. Les paramètres doivent donc positionner

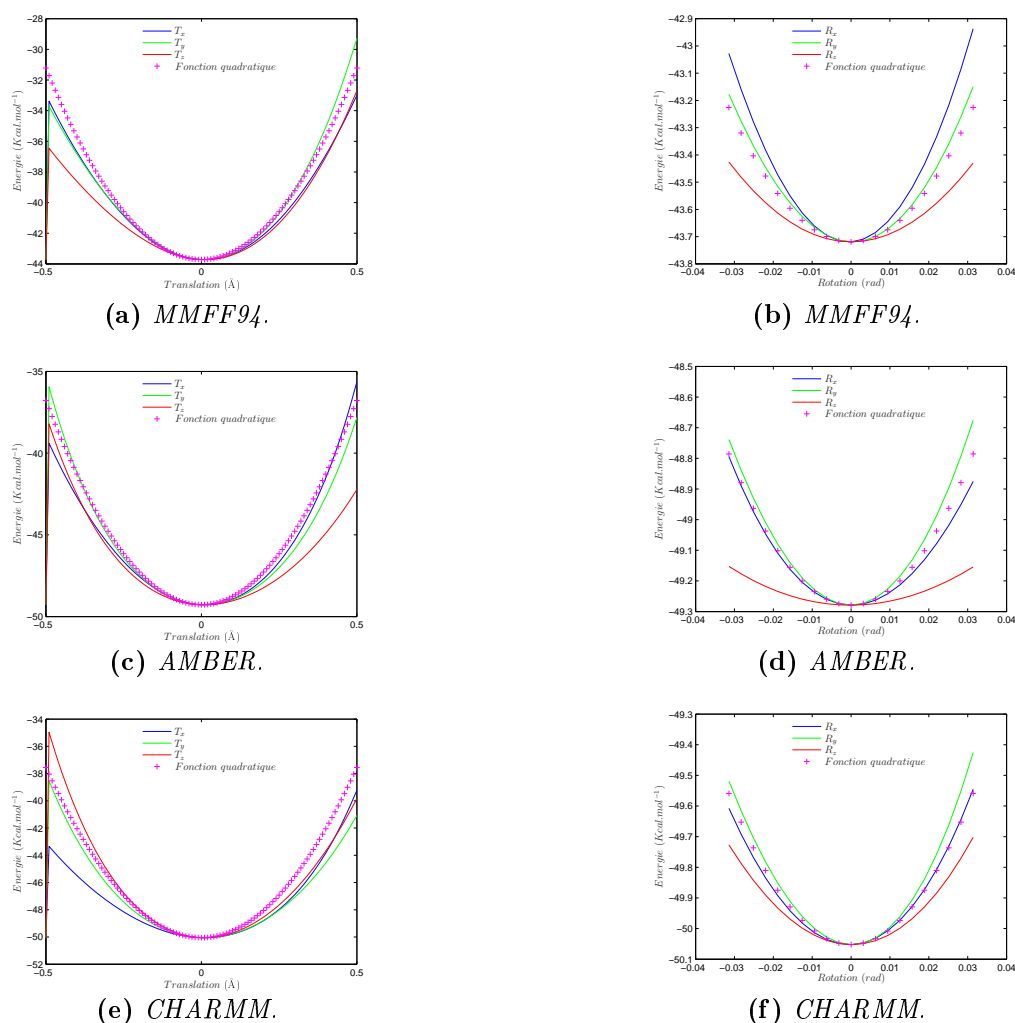


Figure 4.8 - Profils d'énergies obtenus pour des rotations et des translations du ligand dans le site actif d'une protéine selon les trois axes de coordonnées. Ces profils sont comparés à une fonction quadratique. Trois champs de forces différents sont comparés.

le minimum de la fonction, noté \mathbf{X}_e , sur le minimum de l'énergie d'interaction. \mathbf{X}_e représente la matrice colonne des coordonnées du minimum de la fonction d'estimation. En reprenant l'exemple précédent, ce paramètre correspondrait dans le cas tridimensionnel à $(\hat{x}, \hat{y}, \hat{z})$.

La puissance de la fonction d'estimation peut être utilisée comme paramètre. Cependant, la forme générale pour la translation et la rotation doit rester identique afin que les contributions respectives des forces et des moments le restent également. Imposer une puissance différente entre l'énergie modélisant la translation et l'énergie modélisant la rotation modifierait par conséquent les contributions associées des forces et des moments. Dans le cadre des travaux présentés dans ce manuscrit, les puissances utilisées sont considérées comme identiques et constantes.

L'approximation du profil d'énergie doit intégrer le fait que la méthode des moindres carrés nécessite une inversion matricielle qui peut être conséquente si le paramètre m est élevé. Une application en temps réel tend à réduire la taille du système ainsi que le nombre de paramètres.

Ainsi, une forme pour l'énergie modélisant la translation peut être :

$$\widehat{E}_{\text{Trans}}(\mathbf{X}, \Theta_{\text{Trans}}) = \frac{1}{2}k_t \|\mathbf{X}|_{\mathfrak{R}_m} - \mathbf{X}_e|_{\mathfrak{R}_m}\|^2 + \alpha \quad (4.28)$$

avec k_t une constante positive et α une constante. $\mathbf{X}|_{\mathfrak{R}_m}$ représente la position du ligand à l'instant considéré et est projetée dans le repère fixe de la molécule : \mathfrak{R}_m . La position du minimum de l'énergie estimée en translation, $\mathbf{X}_e|_{\mathfrak{R}_m}$, est par conséquent exprimée dans le repère fixe de la molécule.

L'ensemble Θ_{Trans} des paramètres d'estimation de l'énergie de translation est donc représenté par :

$$\Theta_{\text{Trans}} = [k_t, \mathbf{X}_e, \alpha]^t \quad (4.29)$$

La représentation de l'estimation de la translation dans la matrice \mathbf{A} représentant le gradient de la prédiction s'écrit à partir de la dérivée de l'estimation par rapport à ses paramètres.

Elle s'écrit autour de la position \mathbf{X} :

$$\nabla_{\Theta_{\text{Trans}}} \widehat{E}_{\text{Trans}}(\mathbf{X}, \Theta_{\text{Trans}}) = \frac{1}{2} \|\mathbf{X}|_{\mathfrak{R}_m} - \mathbf{X}_e|_{\mathfrak{R}_m}\|^2 \delta k_t - k_t (\mathbf{X}|_{\mathfrak{R}_m} - \mathbf{X}_e|_{\mathfrak{R}_m}) \delta \mathbf{X}_e|_{\mathfrak{R}_m} + \delta \alpha \quad (4.30)$$

Il est alors possible de déduire de l'équation 4.30 la matrice \mathbf{A} à inverser :

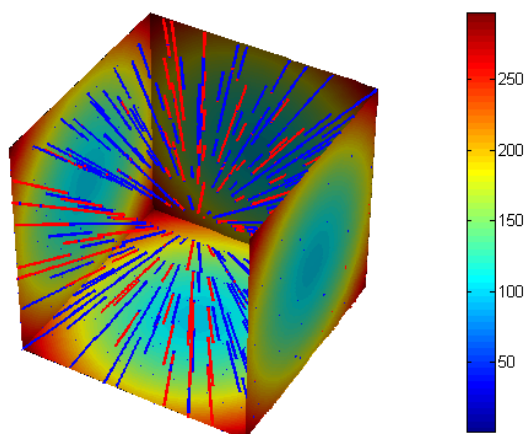
$$\mathbf{A}|_{[(k-m) \times 5]} = \left[\begin{array}{ccc} \frac{1}{2} \|\mathbf{X}|_{\mathfrak{R}_m} - \mathbf{X}_e|_{\mathfrak{R}_m}\|^2, & -k_t (\mathbf{X}|_{\mathfrak{R}_m} - \mathbf{X}_e|_{\mathfrak{R}_m}), & \mathbf{1} \end{array} \right]_{[(k-m) \times 5]} \quad (4.31)$$

La matrice colonne \mathbf{Y} contient l'ensemble des paramètres Θ_{Trans} de 4.29.

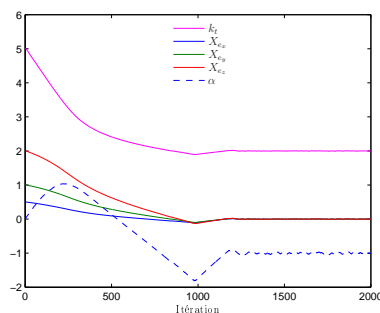
3.1.1.1 Validation en simulation

Afin de valider la méthode, la fonction d'estimation décrite au paragraphe précédent est comparée à la fonction $E = x^2 + y^2 + z^2 - 1$. Le processus de simulation est alors le suivant :

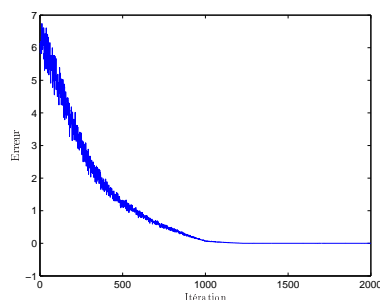
- Détermination d'un chemin aléatoire autour d'une position initiale,
- initialisation des paramètres de la fonction d'estimation,
- évaluation du gradient de l'estimation, étape k ,
- évaluation de la fonction d'estimation, étape k , Θ_{k-1} ,
- construction de la matrice $\mathbf{A}(k, \Theta_{k-1})$, $\mathbf{Y}(\Theta_{k-1})$, $\mathbf{B}(k, \Theta_{k-1})$,
- $\mathbf{Y}_k = \mathbf{A}^+(k, \Theta_{k-1})\mathbf{B}(k, \Theta_{k-1})$,
- $\Theta_k = \Theta_{k-1} + \mathbf{Y}_k$.



(a) Surface d'énergie $E = x^2 + y^2 + z^2 - 1$ comparée à l'estimation. Les lignes rouges et bleues indiquent respectivement la direction de plus grande pente pour l'énergie réelle et estimée.



(b) Estimation des paramètres pour la translation.



(c) Erreur entre l'estimation et l'énergie réelle.

Figure 4.9 - Estimation des paramètres du modèle d'énergie pour la translation.

La figure 4.9 présente les résultats de l'optimisation des paramètres pour la translation. La figure 4.9(a) représente la superposition de l'énergie réelle E et de l'énergie estimée \hat{E} avec ses paramètres optimaux obtenus en fin de simulation (figure 4.9(b)). La superposition des gradients, représentés par des lignes bleues et rouges, confirme que le volume réel est bien approché par la fonction d'estimation. La figure 4.9(c) montre que l'erreur entre l'estimation et la mesure tend vers zéro en fin de simulation. Les paramètres modélisent donc un volume d'énergie simple avec une erreur proche de zéro.

3.1.1.2 Validation expérimentale

La méthode des moindres carrés converge dans un cas d'énergie volumique proche de la fonction d'estimation. Cependant, cette fonction d'estimation doit être testée en situation expérimentale afin d'assurer qu'elle modélise l'énergie d'interaction ligand-protéine dont les termes ne sont pas quadratiques.

La figure 4.10 page suivante présente les résultats de la modélisation de l'énergie d'interaction entre la biotine et la streptavidine pour le champ de force MMFF94. L'objet de cette simulation est de parvenir à identifier les paramètres du modèle d'estimation décrit ci-dessus. La position initiale du ligand est la position d'énergie minimum obtenue après une minimisation du système. Cette position est représentée en pointillé sur la figure 4.10(c) page ci-contre.

D'après les deux graphiques 4.10(c) et 4.10(f), l'algorithme pour la fonction de forme choisie n'a pas convergé vers une valeur optimum. En effet, la position d'énergie minimum, $\mathbf{X}_{e_{opt}}$ en pointillé sur la figure 4.10(c) page suivante, aurait dû être la valeur de convergence pour le paramètre \mathbf{X}_e qui représente l'évolution du minimum de la fonction d'estimation au cours des

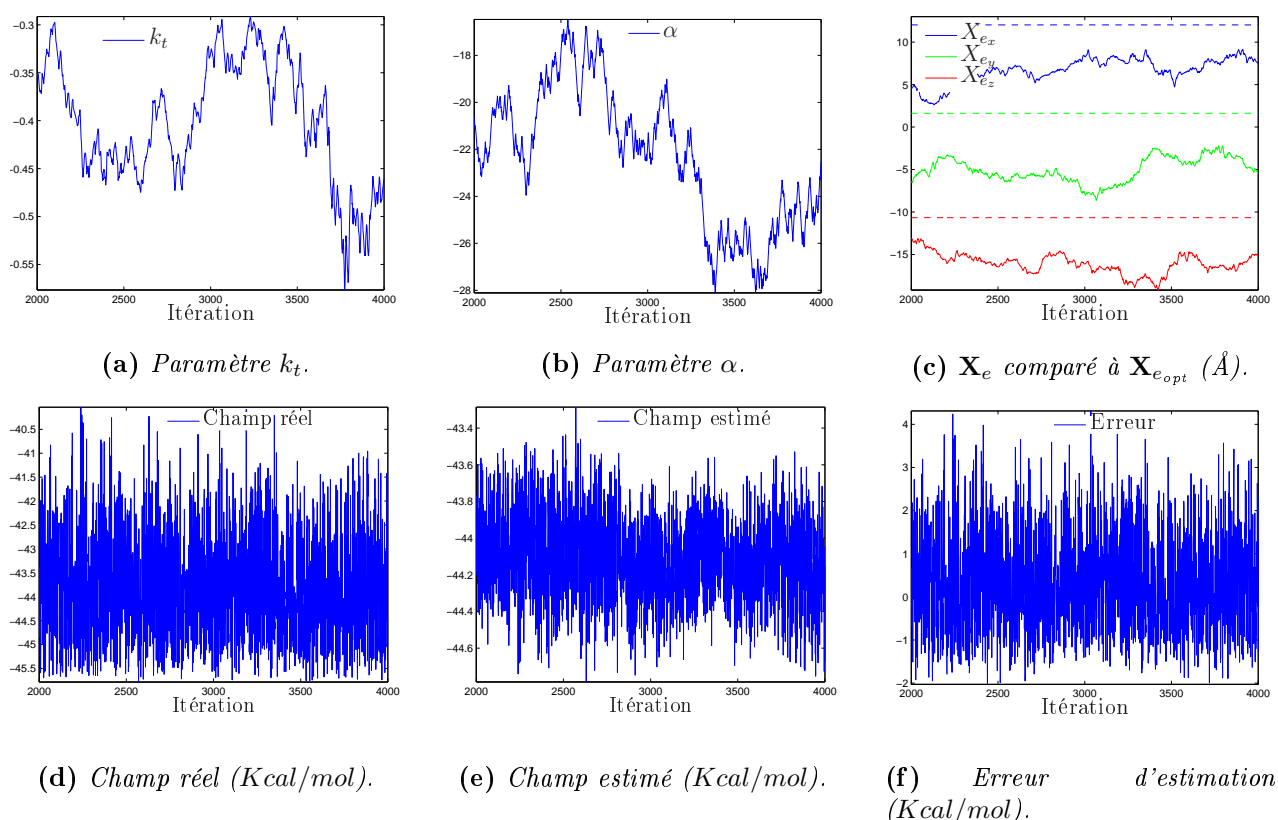


Figure 4.10 - Estimation des paramètres du modèle d'énergie pour la translation à partir de la position d'énergie minimum $\mathbf{X}_{e_{opt}}$.

étapes de calcul. La convergence n'est pas optimum.

Malgré la convergence relative, il ressort de l'étude des graphiques 4.10(d) et 4.10(e) que les valeurs du champ estimé, en moyenne, ne sont pas éloignées de la solution réelle. Les grandes variations n'ont que peu d'effets sur l'évolution des paramètres. Le champ réel, du fait de l'excitation aléatoire, présente une variation de l'ordre de 5 Kcal.mol^{-1} , tandis que l'estimation ne varie que de $1.5 \text{ Kcal.mol}^{-1}$. Cette faible variation s'explique par un gain G élevé (équation 4.22).

L'excitation étant aléatoire, une partie des variations des paramètres peut être expliquée par ce phénomène. Les paramètres k_t et α ne sont pas non plus dans un état de convergence optimum mais leur variation reste bornée.

3.1.1.3 Conclusion

Bien que faiblement convergents, les résultats expérimentaux montrent tout de même qu'une forme quadratique pour l'estimation des paramètres approche l'énergie d'interaction ligand-protéine. La convergence n'est pas optimum mais une partie de l'erreur provient de l'excitation aléatoire du système. Si l'opérateur désire une transparence parfaite, la solution proposée ne l'est pas. Cependant, l'avantage non négligeable d'utiliser une méthode dont l'estimation varie lentement au cours d'une manipulation est le filtrage qui lui est inhérent. En effet, une minimisation du système provoque des variations brusques de déplacement et par conséquent

des variations brusques d'énergie. Cette variation n'est pas immédiatement répercutée par la fonction d'estimation. Par conséquent, la manipulation du ligand peut être plus stable.

3.1.2 Forme de l'estimation pour la rotation

3.1.2.1 Formulation

Toute la difficulté de la modélisation d'une énergie par une rotation tient dans son estimation en termes de rotation. La forme de la fonction associée pour la translation est quadratique : il convient de la garder pour la modélisation utilisant une rotation.

Ainsi, un champ quadratique peut s'exprimer sous la forme de la trace d'une matrice de rotation :

$$\widehat{E}_{\text{Rot}}(\mathbf{R}, \Theta_{\text{Rot}}) = -g_0 (\text{tr}(\mathbf{R}^t \mathbf{R}_e) - 3) \quad (4.32)$$

où g_0 est une constante positive. \mathbf{R} représente la rotation du ligand à l'instant considéré et \mathbf{R}_e la rotation du minimum de la fonction d'estimation.

La forme quadratique se révèle par une décomposition en éléments simples de la trace de la matrice de rotation \mathbf{R} [Murray 94] :

$$-\text{tr}(\mathbf{R}) = -(1 + 2 \cos \chi) \approx -2 \left(1 - \frac{\chi^2}{2}\right) \approx \chi^2 \quad (4.33)$$

χ représentant un angle quelconque, formé par la rotation du ligand, dans le repère global fixe de la molécule \mathfrak{R}_m .

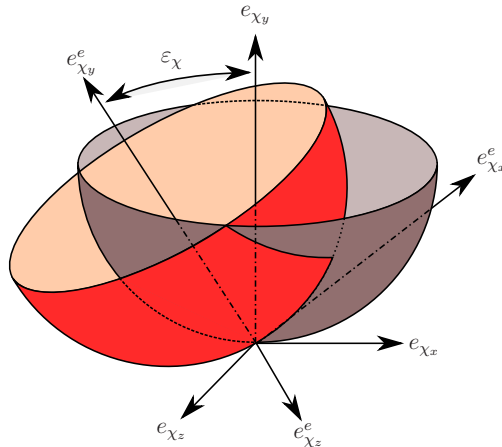


Figure 4.11 - Représentation schématique du principe de minimisation de la fonction modèle dans le repère $(e_{\chi_x}^e, e_{\chi_y}^e, e_{\chi_z}^e)$ par rapport à la fonction réelle dans son repère $(e_{\chi_x}, e_{\chi_y}, e_{\chi_z})$.

En prenant pour exemple la figure 4.11, chercher à minimiser $\widehat{E}_{\text{Rot}}(\mathbf{R}, \Theta_{\text{Rot}})$ revient à minimiser l'erreur d'orientation ε_χ du minimum de l'estimation dans le repère $(e_{\chi_x}^e, e_{\chi_y}^e, e_{\chi_z}^e)$, par

rapport au minimum réel de la fonction qui doit être modélisée, dans le repère $(e_{\chi_x}, e_{\chi_y}, e_{\chi_z})$ de la figure 4.11 page précédente. Alors que la minimisation du modèle pour la translation tend à minimiser l'écart en translation entre la fonction modèle et la fonction réelle, la minimisation en termes de rotation minimise, à partir d'un minimum optimum l'angle entre les deux repères ainsi engendré.

Lorsque le modèle représente la fonction réelle avec une erreur proche de zéro, le terme $\mathbf{R}^t \mathbf{R}_e$ tend vers la matrice identité. Le minimum de la minimisation de l'erreur doit donc être atteint pour cette valeur. La trace de l'identité valant trois, l'annulation de l'équation 4.32 nécessite de différencier la trace de l'identité à l'estimation $\text{tr}(\mathbf{R}^t \mathbf{R}_e)$.

3.1.2.2 Estimation des paramètres

Les paramètres Θ_{Trans} de la matrice \mathbf{Y} du modèle pour la translation s'obtiennent en inversant le système $\mathbf{A}\mathbf{Y} = \mathbf{B}$. Les paramètres actualisés sont alors l'addition des termes précédents et des termes calculés pour \mathbf{Y} . Les paramètres de la rotation ne peuvent pas s'obtenir par addition des paramètres de l'étape k et de l'étape $k - 1$; en effet, le résultat ne serait plus une rotation.

Construction de la matrice \mathbf{A} . Le modèle de l'équation 4.32 est fondé sur la matrice de rotation \mathbf{R} du ligand, calculée dans le repère fixe de la molécule \mathfrak{R}_m , ainsi que sur la matrice de rotation du minimum de l'estimation \mathbf{R}_e , exprimée dans le même repère \mathfrak{R}_m . La matrice \mathbf{A} contient les termes de la dérivation de l'estimation par rapport à ses paramètres.

Les paramètres Θ_{Rot} du modèle pour la rotation sont g_0 et \mathbf{R}_e :

$$\Theta_{\text{Rot}} = [g_0, \mathbf{R}_e]^t \quad (4.34)$$

Il convient dès lors de calculer $\nabla_{\Theta_{\text{Rot}}} \widehat{E}_{\text{Rot}}(\mathbf{R}, \Theta_{\text{Rot}})$

Dérivation de $\text{tr}(\mathbf{R}^t \mathbf{R}_e)$. Pour cette dérivation non triviale, il faut reprendre un cas général. La dérivation de la trace d'une matrice de rotation quelconque \mathbf{R} s'écrit :

$$\delta \text{tr}(\mathbf{R}) = \text{tr}(\delta \mathbf{R}) \quad (4.35)$$

Une des propriétés fondamentales des matrices de rotation assure que $\mathbf{R}\mathbf{R}^t = \mathbf{1}$. L'équation 4.35 se met donc sous la forme :

$$\text{tr}(\delta \mathbf{R}) = \text{tr}(\mathbf{R}\mathbf{R}^t \delta \mathbf{R}) \quad (4.36)$$

$\mathbf{R}^t \delta \mathbf{R}$ est une matrice antisymétrique notée $[\delta \mathbf{w}]$. Soit :

$$\text{tr}(\mathbf{R}\mathbf{R}^t\delta\mathbf{R}) = \text{tr}(\mathbf{R}[\delta\mathbf{w}]) \quad (4.37)$$

La trace d'une matrice de rotation quelconque \mathbf{R} , multipliée par une matrice antisymétrique, ne fait donc apparaître que la contribution des termes qui ne sont pas sur la diagonale de \mathbf{R} . Autrement dit, les éléments de \mathbf{R} qui contribuent à la trace du produit $\mathbf{R}[\delta\mathbf{w}]$ sont ceux de sa matrice antisymétrique notée $[\mathbf{R}]$. L'équation 4.37 se réécrit donc :

$$\text{tr}(\mathbf{R}\mathbf{R}^t\delta\mathbf{R}) = 2 \left([\mathbf{R}]^V \right)^t \delta\mathbf{w} \quad (4.38)$$

$[\mathbf{R}]^V$ est l'écriture vectorielle de la matrice antisymétrique $[\mathbf{R}]$ et présente donc trois composantes. L'équation 4.38 calcule l'expression de la dérivée de la fonction d'estimation par rapport à sa matrice de rotation \mathbf{R}_e :

$$\delta\text{tr}(\mathbf{R}^t\mathbf{R}_e) = 2 \left([\mathbf{R}^t\mathbf{R}_e]^V \right)^t \delta\mathbf{w}_e. \quad (4.39)$$

avec $\delta\mathbf{w}_e = [\mathbf{R}^t\delta\mathbf{R}_e]^V$.

L'expression de la dérivée de la fonction, par rapport à ses paramètres, s'écrit finalement :

$$\delta\widehat{E}_{\text{Rot}}(\mathbf{R}, \boldsymbol{\Theta}_{\text{Rot}}) = - \left(\text{tr}(\mathbf{R}^t\mathbf{R}_e) - 3 \right) \delta g_0 + 2g_0 \left([\mathbf{R}^t\mathbf{R}_e]^V \right)^t \delta\mathbf{w}_e. \quad (4.40)$$

La matrice \mathbf{A} à inverser se déduit de l'équation 4.40 et s'écrit :

$$\mathbf{A}|_{[(k-m)\times 4]} = \left[\begin{array}{cc} - \left(\text{tr}(\mathbf{R}^t\mathbf{R}_e) - 3 \right), & 2g_0 \left([\mathbf{R}^t\mathbf{R}_e]^V \right)^t \end{array} \right]_{[(k-m)\times 4]} \quad (4.41)$$

Obtention des paramètres du modèle. La matrice \mathbf{Y} contient, en une colonne, l'ensemble des paramètres $\boldsymbol{\Theta}_{\text{Rot}}$ de l'équation 4.34. A l'étape k , pour un ensemble de paramètres calculés à l'étape $k-1$, l'actualisation des paramètres de l'étape k nécessite l'inversion de \mathbf{A} . La résolution du système conduit à l'obtention de δg_0 à l'étape k . Une fois cette étape réalisée, le paramètre g_0 s'actualise grâce à l'équation 4.42 :

$$g_0(k) = g_0(k-1) + \frac{\delta g_0(k)}{G} \quad (4.42)$$

La rotation calculée par l'inversion de \mathbf{A} calcule l'estimation du paramètre $\delta\mathbf{R}_e$. $\delta\mathbf{R}_e$ représente alors la variation de rotation tendant à minimiser l'erreur $\varepsilon_{\boldsymbol{\Theta}}$ exprimée en un vecteur. La quantification de \mathbf{R}_e s'établit grâce à l'équation 4.43 :

$$\mathbf{R}_e(k) = \mathbf{R}_e(k-1)e^{[\delta\mathbf{w}_e]} \quad (4.43)$$

3.1.2.3 Validation en simulation

Afin de valider la méthode, la fonction d'estimation décrite au paragraphe précédent est comparée à une fonction décrite en termes de rotation : $E = g_{0_{\text{nom}}} (\text{tr}(\mathbf{R}^t \mathbf{R}_{e_{\text{nom}}}) - 3)$. $g_{0_{\text{nom}}}$ représentant la valeur vers laquelle doit tendre g_0 à la fin de l'optimisation, \mathbf{R} un ensemble de matrices de rotation aléatoires et $\mathbf{R}_{e_{\text{nom}}}$ la rotation de l'équilibre vers laquelle \mathbf{R}_e doit tendre en fin de simulation.

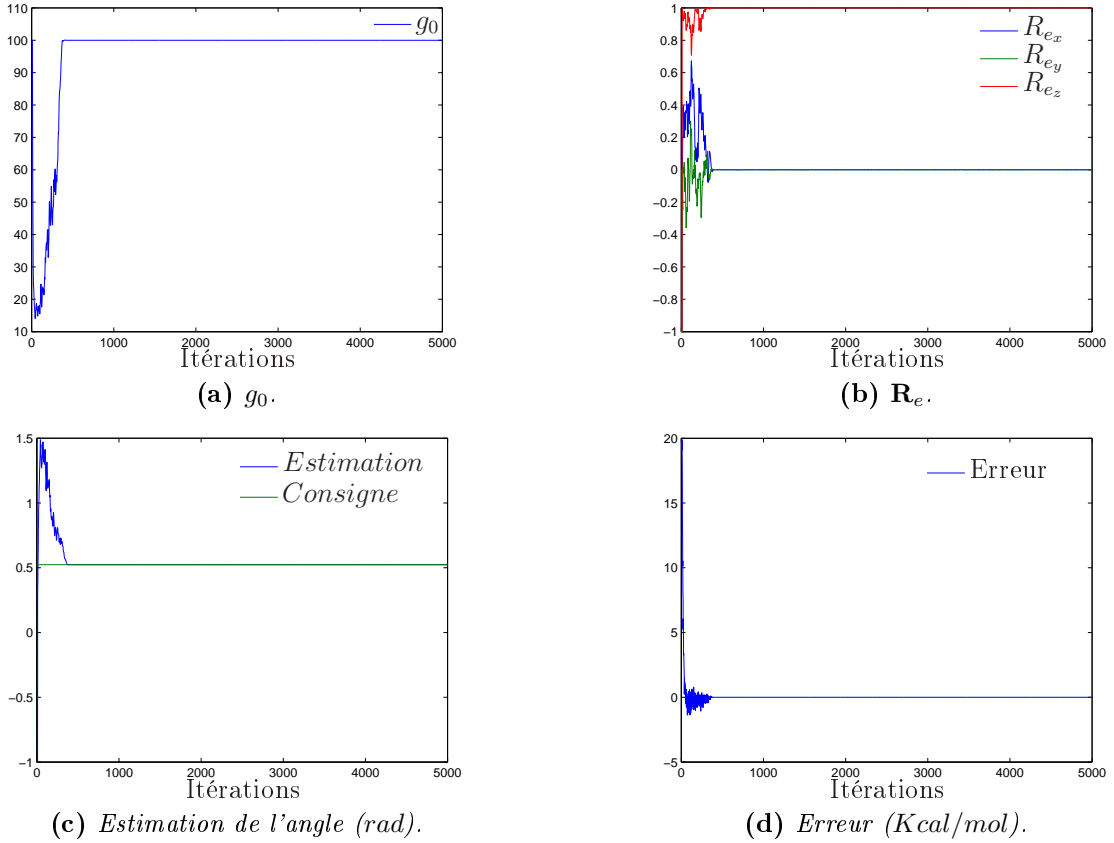


Figure 4.12 - Estimation des paramètres du modèle d'énergie pour la rotation à partir de l'énergie consigne $E = g_0(\text{tr}(\mathbf{R}^t \mathbf{R}_{e_{\text{nom}}}) - 3)$, $g_{0_{\text{nom}}} = 100$, $\mathbf{R}_{e_{\text{nom}}}$ représente une rotation d'axe \vec{z} et d'angle $\pi/6$.

Le processus de simulation est le suivant :

- Détermination d'un chemin aléatoire composé de matrices de rotation,
- initialisation des paramètres de la fonction d'estimation,
- évaluation du gradient de l'estimation, étape k ,
- évaluation de la fonction d'estimation, étape k , Θ_{k-1} ,
- construction de la matrice $\mathbf{A}(k, \Theta_{k-1})$, $\mathbf{Y}(\Theta_{k-1})$, $\mathbf{B}(k, \Theta_{k-1})$,
- $\mathbf{Y}(\Theta_k) = \mathbf{A}^+(k, \Theta_{k-1})\mathbf{B}(k, \Theta_{k-1})$,
- $(g_0, \beta)(k) = (g_0, \beta)(k-1) + \mathbf{Y}(g_0, \beta)(k)$,
- $\mathbf{R}_e(k) = \mathbf{R}_e(k-1)e^{[\mathbf{Y}(\delta \mathbf{w}_e)(k)]}$.

Pour une excitation initiale aléatoire et au cours de la simulation, l'algorithme converge très rapidement vers les valeurs de consigne, c'est-à-dire, $g_{0_{\text{nom}}} = 100$ et une matrice de rotation

correspondant à une rotation d'axe \vec{z} et d'angle $\pi/6$. En effet, le graphique 4.12(b), représentant l'évolution du paramètre de rotation normalisé, présente une rotation d'axe \vec{z} . Ces paramètres représentent la rotation de l'estimation par rapport à une rotation réelle (d'axe \vec{z} et d'angle $\pi/6$). Cette évolution peut être traduite, par exemple en trois dimensions, grâce à la figure 4.11 page 122. La rotation réelle de la fonction représentant l'énergie serait alors représentée par la demi-sphère grise, celle-ci étant supposée d'axe \vec{z} et d'angle $\pi/6$. L'estimation correspondrait alors à la demi-sphère rouge, qui évoluerait en rotation au cours des itérations successives jusqu'à juxtaposer la demi-sphère grise. Les paramètres d'estimation tendraient alors vers les caractéristiques $g_{0_{\text{nom}}}$ et $\mathbf{R}_{e_{\text{nom}}}$ de celle-ci (figures 4.12(b) et 4.12(c) page précédente).

Cette convergence parfaite laisse supposer que les expérimentations présenteront un taux de convergence, sinon identique, du moins proche.

3.1.2.4 Validation expérimentale

Pour l'estimation d'une énergie d'interaction biotine-streptavidine, la méthode des moindres carrés est prouvée convergente d'après la figure 4.13. L'énergie de la biotine et du site actif de la streptavidine est minimisée avant l'approximation de l'énergie d'interaction. La biotine est animée d'une rotation aléatoire, d'un angle faible, autour de sa position d'énergie minimum. La convergence des paramètres est donc assurée.

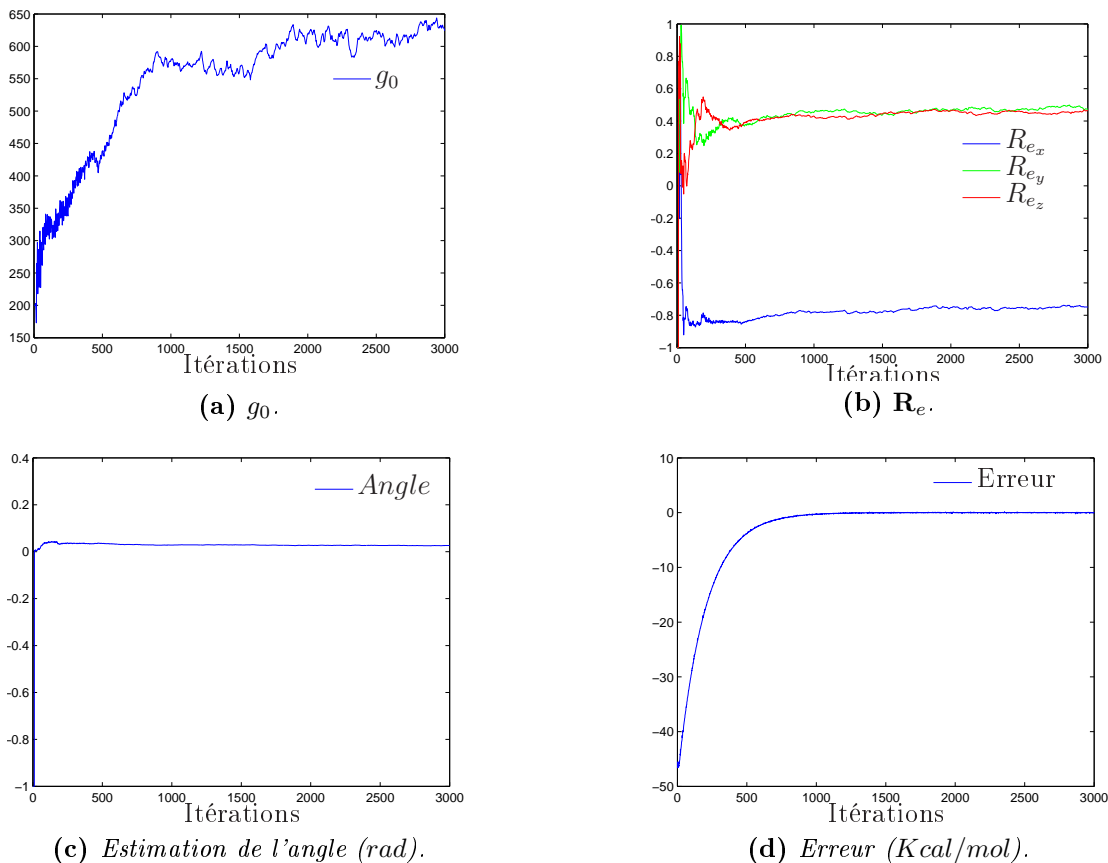


Figure 4.13 - Estimation des paramètres du modèle d'énergie pour la rotation à partir de l'énergie d'interaction biotine/streptavidine.

Les paramètres obtenus en fin de simulation n'approchent qu'une zone située près de la position d'énergie minimum et délimitée par l'ensemble des déplacements aléatoires imposés au ligand. Cependant, au cours de la manipulation du ligand dans le site actif, en considérant que les déplacements imposés par l'utilisateur sont faibles et que l'énergie d'interaction varie peu, l'estimation des paramètres sera une bonne approximation de l'énergie à des positions éloignées de la position d'énergie minimum.

3.1.2.5 Conclusion

La convergence de la méthode d'approximation de l'énergie d'interaction ligand-protéine par une énergie dont certains termes sont à estimer est prouvée pour deux cas. Le premier considère un déplacement en translation : l'objectif est donc la modélisation de l'interaction par une fonction d'estimation des translations. Le deuxième considère l'approximation de l'énergie d'interaction par une énergie représentant une rotation. L'étape suivante consiste à prouver qu'une énergie d'interaction peut être approchée par un terme en translation additionné à un terme en rotation. Cette démonstration est l'objet du paragraphe suivant.

3.2 Formulation générale

L'interaction ligand-protéine peut être approchée par une énergie modèle dont les termes traduisent successivement un déplacement et une rotation. Cependant, l'association des deux n'est peut-être pas convergente pour autant.

La formulation générale du modèle peut être décrite par l'équation 4.44.

$$\widehat{E}(\mathbf{X}, \mathbf{R}, \Theta) = \widehat{E}(\mathbf{X}, \Theta_{\text{Trans}}) + \widehat{E}(\mathbf{R}, \Theta_{\text{Rot}}) \quad (4.44)$$

Cette dernière formulation étant la concaténation des approximations précédentes, le modèle d'énergie s'écrit :

$$\widehat{E}(\mathbf{X}, \mathbf{R}, \Theta) = \frac{1}{2}k_t \|\mathbf{X}|_{\mathfrak{R}_m} - \mathbf{X}_e|_{\mathfrak{R}_m}\|^2 - g_0 (\text{tr}(\mathbf{R}^t \mathbf{R}_e) - 3) + \alpha \quad (4.45)$$

3.3 Validation expérimentale

La figure 4.14 page suivante présente la convergence de la méthode pour un système dont l'énergie initiale est minimisée et qui est soumis à une minimisation au cours de l'estimation des paramètres. La biotine et la streptavidine sont dans leur état de plus faible énergie, le centre de masse du ligand étant à la position ($x_{\min} = 11.66 \text{ \AA}$, $y_{\min} = 2.33 \text{ \AA}$, $z_{\min} = -10.22 \text{ \AA}$).

La convergence est assurée d'une part par la faible erreur entre l'estimation et le champ réel en fin de simulation, et d'autre part par le fait que la position d'énergie minimum (figure 4.14(a) page suivante) estimée par la fonction de l'équation 4.28, traduit la position d'équilibre dans

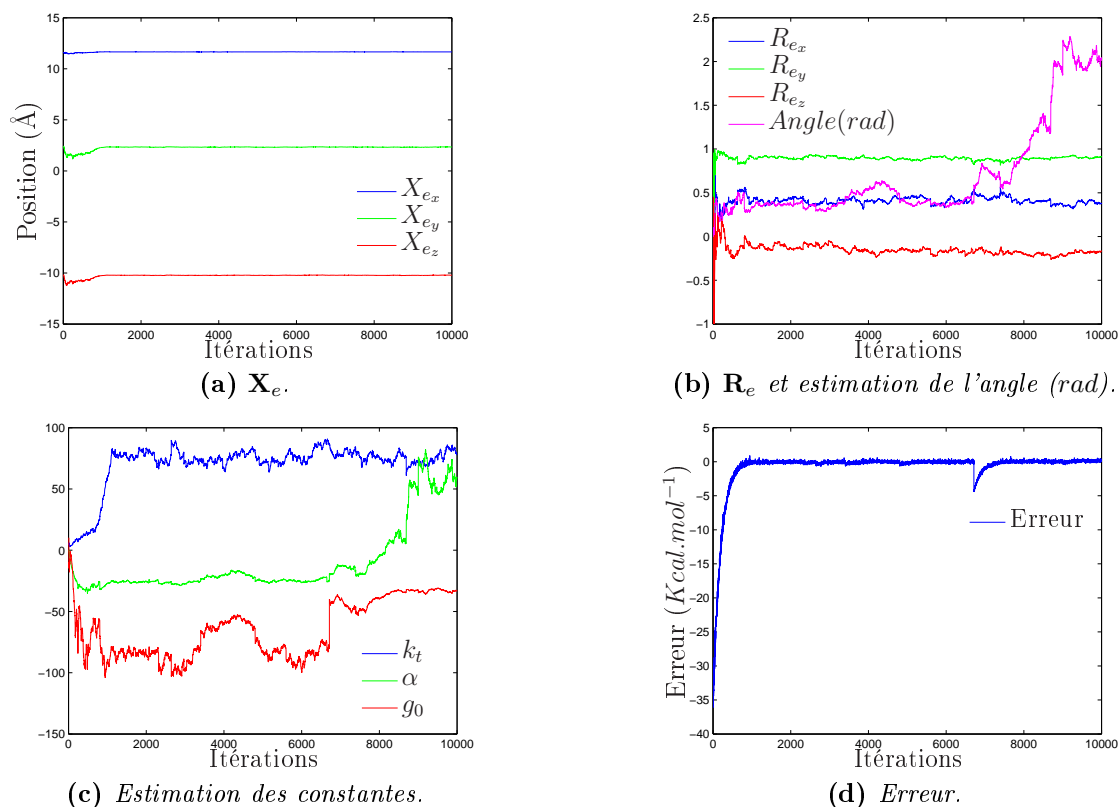


Figure 4.14 - Estimation des paramètres du modèle d'énergie faisant intervenir la rotation et la translation.

laquelle se trouve le ligand. Le puits de potentiel dans lequel est situé le ligand peut être modélisé par les paramètres calculés en fin de simulation.

Les fortes oscillations des paramètres sont notamment dues au faible gain ($G = 20$) utilisé lors de l'approximation du champ ambiant. Un gain plus élevé produit une oscillation moindre et des valeurs de convergence plus stables. Cependant, l'adaptation à l'évolution de l'énergie d'interaction au cours du déplacement est plus faible car les paramètres n'évoluent que très peu, reproduisant ainsi moins précisément les interactions ligand-protéine.

L'adaptation des paramètres doit se faire assez rapidement, en quelques itérations, afin de garantir qu'à tout instant l'énergie d'interaction est correctement modélisée. Une faible adaptation aurait pour conséquence une divergence temporaire de l'algorithme. L'erreur entre l'estimation et le champ réel ne peut pas être réduite en quelques itérations. Le gain G est ainsi un paramètre essentiel qui, selon qu'il est élevé ou faible, engendre une faible évolution des paramètres au cours du temps et une manipulation stable, ou une bonne adaptation aux variations énergétiques au détriment d'une manipulation sans oscillation. Le choix de G dépend donc de la tâche à accomplir. S'il s'agit de ressentir précisément les interactions, ce gain devra être élevé. En revanche, s'il s'agit de tester les ouvertures de sites, le gain peut être faible.

Bien que les paramètres de modélisation de l'énergie varient, les forces et les moments ne suivent pas exactement cette variation puisqu'ils dépendent de la dérivée par rapport aux paramètres de la fonction modèle. Cette propriété est vérifiée dans le paragraphe suivant, qui présente la méthode de calcul des efforts en fonction du modèle estimé.

3.4 Conclusion

Ce paragraphe était destiné à montrer qu'une énergie d'interaction, même fortement non linéaire, peut être approchée par une modélisation quadratique. Cette modélisation est très locale et utilise des paramètres qui évoluent au cours de la manipulation. Le modèle étudié est composé de deux termes représentant respectivement un déplacement du ligand ainsi qu'une rotation. L'optimisation de la fonction modèle est obtenue grâce à une méthode des moindres carrés.

La méthode converge pour une forme quadratique de l'estimation. Cette convergence doit être assurée quelle que soit la position du ligand dans le site actif. Or le conditionnement de la matrice \mathbf{A} impose un déplacement aléatoire du ligand autour de la position d'estimation des paramètres. Cette contrainte va à l'encontre de la nécessité d'un calcul rapide pour la stabilité de l'interface haptique mais elle est une condition nécessaire pour la stabilité du modèle. Cette variation de position devra donc être réalisée soit par l'utilisateur, soit par l'algorithme qui, pour chaque nouvelle position, imposera un ensemble de mouvements aléatoires transparents pour l'utilisateur.

Ce paragraphe montre notamment l'importance des gains G et m . Ceux-ci doivent être réglés pour une bonne précision mais pour un temps de calcul accru. Un compromis doit donc être réalisé entre précision et temps de calcul acceptable.

L'énergie d'interaction ligand-protéine est approchée par une fonction dont les paramètres sont connus à tout instant. Sa dérivée analytique fournit l'expression des forces et des moments à ressentir.

4 De la modélisation d'une énergie au calcul des efforts

Les paragraphes précédents ont montré qu'il est possible de modéliser une énergie d'interaction ligand-protéine par une énergie dont les paramètres sont à estimer et traduisent intrinsèquement une translation et une rotation. Le terme de translation affine le positionnement du minimum de la fonction d'estimation par rapport à celui de l'énergie d'interaction. Le terme de rotation oriente la forme du minimum vers celle de l'énergie réelle.

Une fois l'énergie modélisée, les paramètres décrivent un champ proche de celui qui représente l'interaction ligand-protéine. Obtenir les forces et les moments d'un tel champ consiste à dériver cette prédiction, non pas par rapport aux paramètres, mais par rapport à la variation de position et de rotation.

4.1 Calcul des forces d'interaction

Les forces d'interaction \mathbf{F}_{int} sont obtenues par dérivation du modèle d'énergie par rapport à la variation de position :

$$\mathbf{F}_{\text{int}} = \nabla_{\mathbf{X}} \widehat{E}(\mathbf{X}, \mathbf{R}, \Theta) \quad (4.46)$$

En reprenant l'expression du modèle d'énergie de l'équation 4.45, les forces d'interaction s'écrivent :

$$\mathbf{F}_{\text{int}} = \nabla_{\mathbf{X}} \frac{1}{2} k_t (\mathbf{X} - \mathbf{X}_e)^2 + \alpha = k_t (\mathbf{X} - \mathbf{X}_e) \quad (4.47)$$

4.1.1 Validation expérimentale

La figure 4.15 présente les caractéristiques de l'énergie estimée au cours d'une manipulation d'un ligand avec minimisation de l'énergie totale. La manipulation consiste à déplacer le ligand autour de sa position d'énergie minimum en lui imprimant une translation pure selon les axes \vec{x} puis \vec{y} et enfin \vec{z} . Celui-ci étant soumis à une minimisation, sa position d'équilibre tend vers la position d'énergie minimum 4.15(e), celle du début de la simulation.

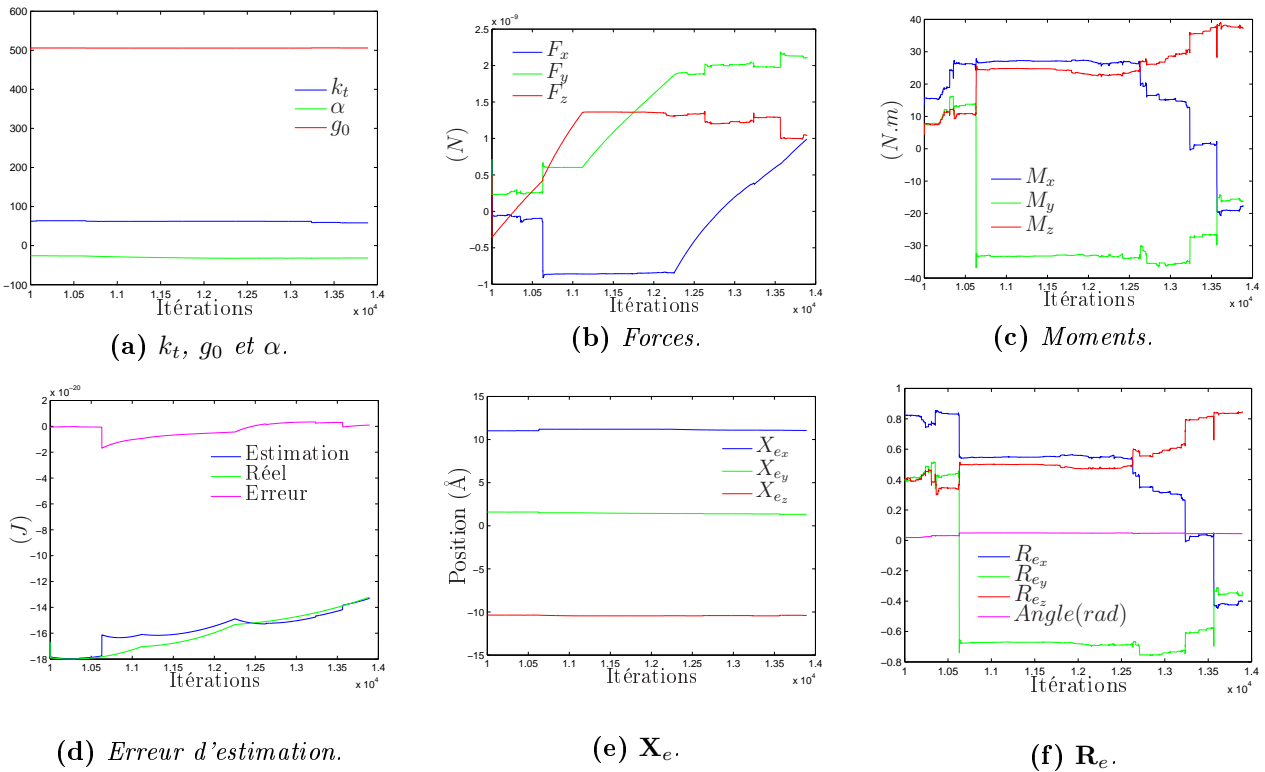


Figure 4.15 - Estimation des paramètres du modèle d'énergie pour la translation et la rotation à partir de la position d'énergie minimum. Une translation selon les axes \vec{x} (10000 \rightarrow 11000), \vec{y} (11000 \rightarrow 13000) et \vec{z} (13000 \rightarrow 14000) est effectuée. L'itération 10000 correspond à l'itération de convergence optimum de l'algorithme.

Au cours de la manipulation, la fonction d'estimation actualise ses paramètres. L'erreur entre l'énergie réelle et l'énergie estimée est minimisée. La figure 4.15(d) montre une bonne convergence de l'algorithme malgré une adaptation lente à l'évolution de l'énergie réelle. Le

gain utilisé pour l'actualisation des paramètres est élevé, limitant les effets des oscillations potentielles de l'énergie réelle, mais réduisant d'autant la rapidité de convergence.

Les forces obtenues et envoyées à l'interface haptique ont une faible amplitude, autorisant leur interprétation d'un point de vue haptique. En effet, les résultats ainsi présentés diffèrent totalement de ceux qui, dans un premier temps, réalisaient le couplage haptique. Une forte oscillation de l'énergie d'interaction n'influe pas directement sur le calcul des efforts car la fenêtre de moyennage modélisée par la valeur de m filtre les instabilités. Les paramètres évoluant lentement par rapport à la fréquence de variation de l'énergie d'interaction, les efforts calculés présentent un profil stable. Dès lors, pour ressentir précisément les interactions, un travail sur le rendu haptique peut être conduit, en préférant la commande en variables d'onde, étudiée *supra*, qui ne présente pas d'amortissement virtuel. Ainsi, les efforts réels entre le ligand et la protéine peuvent être ressentis sans altération.

Cependant, les déplacements du ligand lors de la simulation correspondaient à un coefficient en déplacement faible. L'estimation converge donc rapidement. Un coefficient élevé aurait pour conséquence un changement brusque de l'énergie d'interaction qui ne pourrait être modélisée convenablement.

4.1.2 Conclusion

La formulation généralisée de la fonction d'estimation, telle que présentée à l'équation 4.45, réalise une modélisation des forces au cours d'une manipulation de la biotine au sein du site actif de la streptavidine. L'erreur entre la fonction d'estimation et l'énergie réelle est minimisée au cours de la manipulation. Les paramètres α , k_t et g_0 ne semblent varier que très peu, au détriment du paramètre d'estimation de la rotation \mathbf{R}_e , certainement du fait du mauvais conditionnement de \mathbf{A} . Cette remarque amène à penser que les moments divergent.

Une phase de test de la convergence de la rotation doit être conduite pour l'estimation des moments.

4.2 Calcul des moments d'interaction

Les moments d'interaction \mathbf{M}_{int} sont calculés par dérivation du modèle d'énergie par rapport à la variation de rotation réelle :

$$\mathbf{M}_{\text{int}} = \nabla_{\mathbf{R}} \widehat{E}(\mathbf{X}, \mathbf{R}, \Theta) \quad (4.48)$$

En reprenant l'expression du modèle d'énergie de l'équation 4.45, les moments d'interaction s'écrivent :

$$\mathbf{M}_{\text{int}} = \nabla_{\mathbf{R}} (-g_0 (\text{tr}(\mathbf{R}^t \mathbf{R}_e) - 3)) = 2g_0 \left([\mathbf{R}^t \mathbf{R}_e]^V \right)^t \quad (4.49)$$

4.2.1 Validation expérimentale

La figure 4.16 montre l'estimation des paramètres du modèle général au cours d'une manipulation de la biotine dans le site actif minimisé de la streptavidine. Au cours de la manipulation, la minimisation des deux molécules est activée. La manipulation consiste cette fois à réaliser trois rotations successives du ligand, par rapport à son centre de masse, selon les axes \vec{x} , \vec{y} puis \vec{z} .

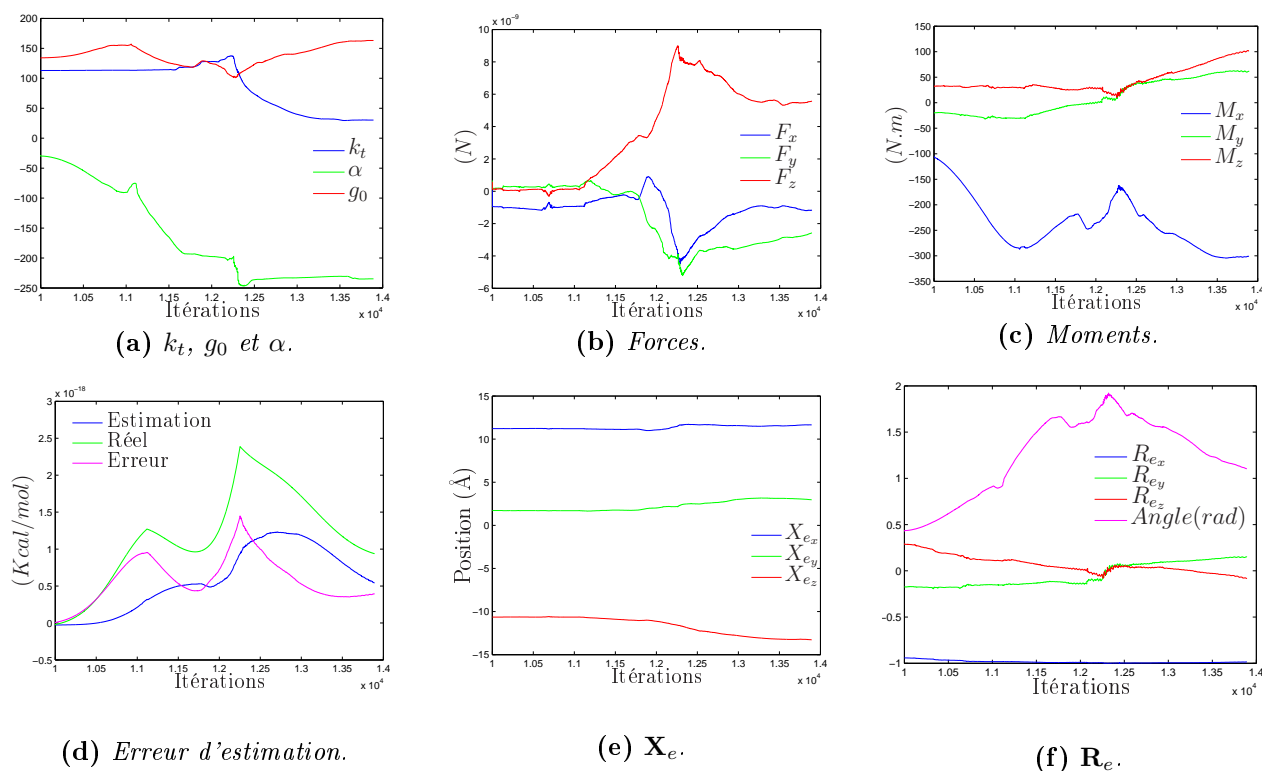


Figure 4.16 - Estimation des paramètres du modèle d'énergie pour la translation et la rotation à partir de la position d'énergie minimum. Une rotation selon les axes \vec{x} ($10000 \rightarrow 11000$), \vec{y} ($11000 \rightarrow 13000$) et \vec{z} ($13000 \rightarrow 14000$) est effectuée. L'itération 10000 correspond à l'itération de convergence optimum de l'algorithme. Au cours de la manipulation, le ligand est soumis à une minimisation du champ de force.

Au cours du temps, les paramètres du modèle sont estimés (figures 4.16(a), 4.16(e) et 4.16(f)). Les paramètres du modèle ne modélisent pas exactement la variation d'énergie engendrée par la rotation (figure 4.16(d)), mais le profil de l'estimation représente le comportement général de l'évolution de l'énergie d'interaction. Cette faible adéquation entre l'estimation et l'énergie réelle peut être imputée au gain G . Celui-ci étant élevé, la faible adaptation à l'évolution du système est une conséquence. Ce modèle et notamment la méthode d'optimisation peuvent donc être améliorés.

Au cours du mouvement, la matrice \mathbf{A} n'est constituée que des gradients de l'estimation pour les positions effectives du ligand. Aucun déplacement aléatoire fictif n'est effectué autour de la position de prédiction des paramètres. Par conséquent, l'adaptation du modèle ne peut qu'être faible puisqu'elle dépend des mouvements imprimés par l'opérateur. Afin d'assurer la convergence optimum de l'algorithme au cours du déplacement, le ligand aurait dû osciller autour de la position de prédiction des paramètres, à un instant t donné, afin d'enrichir pour

cet instant le modèle de prédiction. La prédiction n'en aurait été que plus fiable et l'erreur entre l'estimation et la fonction réelle minimisée.

Concernant cette phase de manipulation qui consiste à n'imprimer qu'une rotation pure au ligand, les moments semblent être moins soumis à une variation arbitraire des paramètres. Ainsi, des zones d'attraction et de répulsion apparaissent clairement. Un coefficient en effort doit être calculé et appliqué afin de rendre fidèlement cette variation de moments et de forces.

4.2.2 Conclusion

La formulation générale, quadratique, assure un calcul des moments à partir de l'estimation des paramètres. Les moments apparaissent stables, contrairement à ceux qui auraient pu être déduits d'une dérivation numérique. Cependant, l'estimation des paramètres semble ne pas autoriser, dans le cas d'une rotation pure, une adaptation rapide de l'estimation à l'énergie réelle. Une solution élégante serait d'enrichir les paramètres par une oscillation aléatoire du ligand, autour des positions de consigne, afin de connaître précisément l'environnement énergétique du point considéré.

5 Conclusion

Au cours de ce chapitre, il a été montré que la conversion des énergies en efforts n'est pas intuitive. Le premier paragraphe proposait une solution consistant à projeter, selon les axes de déplacement, une partie de l'énergie au cours du déplacement. Cette solution ne calcule pas les forces avec précision et dépend pour beaucoup de l'excitation du système. Le deuxième et le troisième paragraphes proposaient, par une fonction dont certains termes sont à estimer, une modélisation locale de l'énergie fournie par le simulateur. Grâce à une méthode des moindres carrés, les paramètres estimés font s'approcher le modèle d'énergie et le champ réel. Cette énergie modèle, dont la forme est connue et déterminée en fonction de l'énergie à modéliser, fournit une solution analytique pour les efforts. La méthode est convergente si les matrices du modèle sont bien conditionnées.

La méthode de modélisation locale de l'énergie d'interaction présente l'avantage non négligeable d'être indépendante du choix du champ de force. En effet, seule l'énergie fournie par le simulateur moléculaire sert de base au calcul des paramètres de la fonction d'estimation. La forme de cette estimation doit tout de même présenter des similitudes avec l'énergie à modéliser. Une forme trop éloignée de la solution provoquerait une divergence de l'algorithme. Cette forme doit donc être déterminée, *a priori*, en fonction d'une estimation de la forme de l'énergie réelle. Une solution serait de n'imposer aucune forme particulière à l'énergie d'estimation, et d'utiliser la puissance comme paramètre. Mais les contraintes imposées par le retour haptique limitent le nombre de paramètres pour une réponse rapide. En outre, cette solution est non linéaire et les risques de divergence sont plus importants.

Le choix de la méthode des moindres carrés s'explique par sa simplicité de mise en œuvre et par sa précision. Cependant, elle n'est adaptée à des contraintes temps réel que si le nombre de paramètres et la taille du système sont limités. L'alternative serait une méthode des moindres

carrés récursifs, avec l'introduction d'un coefficient d'oubli qui tiendrait plus ou moins compte de la contribution de l'énergie réelle.

La précision de la méthode dépend essentiellement de l'enrichissement de la matrice constituée par le gradient de l'estimation par rapport aux paramètres. Cet enrichissement est effectué par une oscillation aléatoire du ligand autour de ses positions successives. Ces déplacements aléatoires permettent de connaître l'environnement énergétique en translation et en rotation autour des positions successives du ligand. Cependant, ce schéma ne peut être envisagé car l'évaluation de l'énergie d'interaction est très peu compatible avec une téléopération en temps réel.

Grâce à ce modèle d'interaction, des applications de *docking* peuvent être réalisées. Les énergies utilisées pour la minimisation des systèmes moléculaires sont approchées par un modèle qui fournit une solution exacte pour les efforts. L'algorithme d'optimisation doit cependant garantir une réponse rapide pour ne pas rendre instable le couplage. L'avantage réside dans le fait que, par l'introduction d'un facteur d'oubli ou en jouant sur la taille de \mathbf{A} , les fortes variations de l'énergie sont filtrées lors des changements de conformations. La manipulation apparaît alors stable.

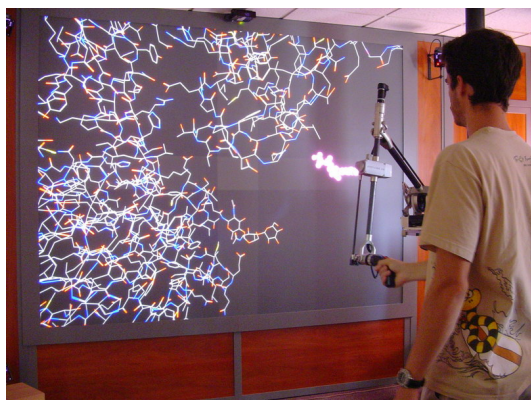


Figure 4.17 - Manipulation de la biotine aux abords du site actif de la streptavidine. Une commande en variables d'onde assure la stabilité du système. Les efforts sont calculés grâce à un modèle paramétrable.

La méthode présentée dans ce chapitre a été implémentée sur MOE, avec une commande en variables d'onde sans amortissement, et testée avec une interface haptique du type *Virtuose 6D*. Les efforts sont ressentis dans les trois directions de l'espace, en translation et en rotation. Le profil des efforts est stable et la commande bilatérale assure l'adéquation des grandeurs physiques entre les deux mondes, macroscopique et moléculaire.

Cependant, il reste au partenaire industriel à confirmer que l'objectif est atteint. Cet objectif est la discrimination de formes géométriques différentes en fonction de leur affinité avec le site actif de la protéine. Si, pour un utilisateur quelconque, le ressenti des efforts illustre les zones attractives et répulsives, il reste à confirmer son utilité comme élément de décision.

Une série de tests doit donc être réalisée afin de préciser le besoin en termes de transparence de la commande et de précision du modèle d'interaction. Des expérimentations plus poussées pourraient permettre d'identifier les gains du couplage en fonction des besoins de l'utilisateur.

Il reste notamment à vérifier, par des manipulations successives, la stabilité de la méthode qui calcule les efforts. Si l'utilisateur n'assure pas au cours de la manipulation, un échantillonnage

spatial régulier autour d'une position, la matrice à inverser pourrait être singulière. Ces limites d'utilisation doivent être définies.

Enfin, il reste à tester la fiabilité de la plate-forme quant au positionnement des molécules dans le site actif. Les résultats de simulations de *docking* automatisés doivent être comparés à ceux qui sont obtenus avec un retour d'effort.

La validité des expérimentations corroborées par des données cristallographiques rendrait alors la plate-forme de téléopération utile pour l'exploration des environnements submicro-niques. Elle s'inscrit donc dans un contexte où la réalité virtuelle est considérée comme l'outil autorisant cette exploration, les données expérimentales étant très difficilement accessibles [Hamdi 04], [Hamdi 07].

CONCLUSION ET PERSPECTIVES

Abordant le problème de la mise en œuvre d'un retour d'effort pour le *docking* moléculaire, une plate-forme de téléopération a été proposée. L'objectif est de manipuler, avec six degrés de liberté en effort, un ligand flexible dans un site actif de protéine flexible. La modélisation moléculaire utilise un logiciel de simulation et l'interface haptique est un *Virtuose*. Une commande bilatérale entre le ligand et l'interface haptique assure le lien entre ces deux composantes. Cependant, plusieurs difficultés sont apparues :

- Il existe une large variété de simulateurs de conformations des protéines et une multitude de modèles pour la détermination de l'énergie du système. Concevoir une plate-forme de téléopération pour un modèle d'interaction associé à un simulateur moléculaire la rend peu modulaire devant la diversité des propriétés physico-chimiques à simuler.
- Faire le choix d'utiliser un logiciel standard pour la simulation moléculaire impose de s'adapter à ses spécificités. En l'occurrence, le logiciel MOE est certes évolutif grâce à son API mais n'est pas prévu pour une réponse temps réel à une requête quelconque. De plus, la simulation conformationnelle impose un nombre important de calculs. La stabilité de la commande bilatérale n'est alors pas garantie. Le choix initial de concevoir la téléopération d'un ligand flexible dans le site actif d'une protéine flexible ne permet pas d'envisager une simulation conformationnelle en temps réel.
- Les interactions sont décrites par des modèles d'énergies classiques. Les changements de conformations sont obtenus grâce à une optimisation de l'énergie globale du système. Il est donc complexe de calculer les forces et les moments directement à partir de l'expression de l'énergie. De plus, les composantes de cette énergie selon les axes de déplacement du *Virtuose* ne sont pas connues.
- Lors d'un premier couplage bilatéral entre le monde moléculaire et le macromonde sont apparus des forces et des moments de grande amplitude. Cette dernière est engendrée par les forces à distance à fort gradient, les forces électrostatiques et de van der Waals. La nécessité d'imposer des facteurs d'échelle, en déplacement et en effort, pour ressentir les efforts dans le macromonde et imposer des déplacements nanométriques, se heurte alors à cette grande variabilité. La vibration de l'organe à retour d'effort provoque ainsi

une instabilité de la commande.

Pour résoudre ces verrous scientifiques et devant les solutions trop simples apportées actuellement, plusieurs réponses adaptées sont proposées.

Des coefficients en effort et en déplacement, fixés en début de simulation, l'un par dynamique moléculaire, l'autre arbitrairement pour le couplage bilatéral de deux mondes aux échelles différentes, sont mis en place. L'amplitude et la variabilité des efforts d'interaction sont amplifiées dès lors que les systèmes en interaction sont flexibles. La grande variabilité des efforts est atténuée par les variables d'onde qui sont, dans un premier temps, utilisées comme un amortisseur virtuel. La vitesse décodée est pondérée par un coefficient d'amortissement réglable puis additionnée à la force provenant de la simulation. Bien qu'atténuant l'instabilité engendrée par les efforts d'interaction, ce couplage présente l'inconvénient majeur de rendre difficilement dissociable la viscosité introduite et les efforts d'interaction à ressentir.

Dans un deuxième temps, les variables d'onde sont utilisées comme simple canal de communication entre l'organe haptique et la simulation moléculaire. Les forces provenant de la simulation sont atténuées par le manque de transparence de la communication. En effet, sa raideur dépend du temps de latence de la communication et d'un coefficient d'amortissement et de stabilisation de la variable d'onde. La transparence n'est pas assurée mais les contacts, auparavant ressentis comme une raideur infinie, présentent une raideur moins importante. L'interface n'oscille donc plus entre des modes de raideur importante. Ce couplage présente, en outre, l'avantage d'être stable au regard des retards de simulation.

La spécificité du champ de force et la grande variété des méthodes d'optimisation de l'énergie d'un système rendent dépendante la plate-forme de téléopération. De plus, les interactions sont modélisées par des énergies alors que l'interface haptique utilise le torseur des efforts d'interaction pour sa commande. Le calcul de ce torseur d'interaction doit être effectué à partir de cette énergie, sans connaissance de ses projections spatiales. L'idée d'identifier un modèle d'énergie, dont la dérivation analytique fournit les forces et les moments de l'interaction garantit à la fois une solution pour le calcul du torseur et l'indépendance de la plate-forme vis-à-vis du modèle d'interaction et de la méthode d'optimisation utilisée. Le modèle d'énergie intègre des paramètres à estimer. Ces derniers s'adaptent à la variation de l'énergie calculée par l'optimiseur. L'interface haptique est alors couplée au nouveau modèle d'énergie. Celui-ci fournit les informations en temps réel, pendant que la réactualisation de ses paramètres s'exécute en temps différé. Ce procédé évite ainsi de devoir optimiser le temps de réponse du simulateur moléculaire. De plus, la lente convergence du nouveau modèle devant les variations rapides de l'énergie, garantit une réponse en effort de plus faible amplitude. Le profil des forces et des moments apparaît en effet plus lisse, ouvrant la voie à un réel travail sur le ressenti haptique des échelles moléculaires.

Les travaux conduits dans le cadre de ce projet sont un premier pas dans le monde de la simulation moléculaire. En effet, cette application apporte ses spécificités propres, pas toujours compatibles avec les exigences des systèmes à retour d'effort. S'il est désormais possible de ressentir les forces et les moments d'une interaction moléculaire grâce à une interface à retour d'effort, il reste à pouvoir réellement les comprendre et les interpréter. Dans l'état actuel des recherches, l'objectif de discriminer plusieurs géométries différentes par leur affinité au site actif n'est pas atteint. En effet, s'il est possible de ressentir les interactions, le manque de transparence et les nombreux changements de conformations sont autant de facteurs altérant la perception de l'interaction ligand-protéine. Si la question concerne la possibilité, pour un

ligand, d'intégrer un site actif de protéine, alors la plate-forme mise en place peut y répondre. En effet, le simulateur se basant sur la minimisation d'un champ de force, l'état énergétiquement stable est constamment garanti et les efforts présents au cours de la simulation peuvent être ressentis. Mais la comparaison de deux manipulations successives, en l'état actuel, ne fournit pas d'informations discriminantes.

Le fait que la protéine soit flexible est certainement le problème majeur pour la comparaison des procédures d'insertion. Une réponse claire, comme un rejet dû aux forces électrostatiques, est compréhensible. Mais si tous les ligands sont compatibles avec le site actif, la plate-forme ne semble pas pouvoir répondre à la recherche de la meilleure affinité.

Pour que la plate-forme soit fonctionnelle, dans le sens d'un réel ajout d'éléments de comparaison, plusieurs travaux doivent être conduits. Ces recherches forment un ensemble de perspectives pour cet outil futur.

Si le parallèle entre le nanomonde et le macromonde est réalisé par un couplage homothétique, le travail consistant à concevoir des métaphores haptiques, pour que l'opérateur comprenne l'interaction, est un problème ouvert. Le fait que la commande ne soit pas transparente n'est pas forcément un désavantage si ce manque de transparence est constructif, dans le sens où il n'est pas utile à l'opérateur de ressentir le profil des efforts à grande raideur dans son intégralité. Ainsi, la grande raideur des forces de van der Waals, par exemple, pourrait être modélisée par une fonction moins instable à l'approche d'un contact. L'ajout de métaphores visuelles ou d'outils de la réalité augmentée peut être une solution adaptée. La plate-forme peut être très facilement améliorée, en intégrant les simulations moléculaires dans un milieu immersif tridimensionnel, dans lequel l'opérateur se trouverait « immergé » au sein de la protéine. Dès lors, en étant très proche de la frontière ligand-protéine qui représente l'interaction, l'opérateur, en manipulant le ligand, peut se rendre compte de la réelle distance d'interaction. Ainsi, au stimulus du toucher s'en ajouterait un, visuel, correspondant à notre monde tridimensionnel. Cette mise en correspondance entre le nanomonde et le macromonde impliquerait, pour l'utilisateur, une meilleure interprétation des informations renvoyées par simple homothétie. Dans ce cas, il semblerait possible d'apporter une réponse quant à la discrimination de la géométrie des ligands.

La plate-forme de téléopération utilise un organe à retour d'effort de la classe *Virtuose*. Ces interfaces sont conçues pour les besoins de la conception assistée par ordinateur. Si elles autorisent la manipulation de portières de voiture, en réalité virtuelle, leur application au nanomonde est délicate. En effet, ressentir une interaction nanométrique sur une poignée de taille conséquente altère forcément le jugement. Une interface de raideur et d'inertie moindre, avec des dimensions réduites, semble être une meilleure solution. Ainsi, il pourrait être pertinent d'envisager la manipulation du ligand avec un seul doigt, dont les terminaisons nerveuses sont bien plus nombreuses. Le réel besoin d'une interface dédiée pour les applications microscopiques a été mis en valeur par l'application proposée.

Les coefficients en déplacement et en effort sont fixes au cours de la manipulation. Cependant, réaliser un *docking* suppose que le ligand parcourt de grandes distances, comparativement à celles qu'il peut parcourir au sein du site actif. Imaginer un coefficient en déplacement variable pourrait être une solution évitant à l'opérateur une succession de recalages avec le *Virtuose*. Intégrer un site actif suppose de franchir des barrières d'énergie engendrées par les chaînes secondaires repliées. En revanche, une fois à l'intérieur du site catalytique, il n'existe plus de barrières d'énergie à franchir. Imposer un coefficient en effort fixe pour toute la manipulation ne semble pas être la meilleure approche. Envisager un coefficient en effort variable selon le type de tâche à réaliser semble être une méthode originale avec de nouveaux problèmes en termes

de stabilité.

L'idée de modéliser l'énergie fournie par le simulateur par une énergie dépendant de paramètres à estimer, et dont la dérivée analytique fournit les efforts d'interaction, permet d'envisager d'autres applications pour la plate-forme. Ainsi, tous les systèmes nanométriques utilisant un modèle d'énergie pour la description de ses interactions peuvent être intégrés à la plate-forme. Il devient dès lors possible de ressentir, par exemple, les forces d'interaction entre deux brins d'ADN. L'outil ainsi conçu ouvre la possibilité de la caractérisation d'entités biologiques du nanomonde dès lors qu'une énergie les caractérise. Il reste cependant à confirmer expérimentalement la validité des efforts d'interaction calculés par la méthode d'optimisation mise en place. La portabilité d'une interface à retour d'effort, dédiée pour des applications moléculaires ou microscopiques, est une nouvelle approche pour l'étude du monde microscopique, ouvrant la voie à une exploration approfondie et plus rapide de ces environnements complexes.

LISTE DES PUBLICATIONS

Daunay, B. and Micaelli, A. and Régnier, S.

6 DOF Haptic Feedback for Molecular Docking Using Wave Variables.

Actes de ICRA'07 IEEE International Conference on Robotics and Automation. Rome, Italie, 2007.

Daunay, B. and Micaelli, A. and Régnier, S.

Energy-Field Reconstruction for Haptics-Based Molecular Docking Using Energy Minimization Processes.

Actes de IROS'07 : IEEE International Conference on Intelligent Robots and Systems. San Diego, USA, 2007.

Daunay, B. and Haliyo, S. and Régnier, S.

6D haptic feedback for molecular docking.

Actes de EUROMECH Colloquium 476, Real-time Simulation and Virtual Reality Applications of Multibody Systems. Ferrol, Espagne, 2006.

Daunay, B. and Abbaci, A. and Micaelli, A. and Régnier, S.

The wave variables, a solution for stable haptic feedback in molecular docking simulations.

IEEE CISSE / IETA 2006, International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering. Pages 67-73, 2006.

BIBLIOGRAPHIE

- [Adams 98] **R. Adams et B. Hannaford.** *A two-port framework for the design of unconditionally stable haptic interfaces.* IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 1254–1259, 1998.
- [Alberts 94] **B. Alberts, D. Bray, J. Lewis, M. Raff et al.** *Biologie moléculaire de la cellule.* Médecine-Sciences, troisième édition, 1994.
- [Allinger 76] **N.L. Allinger.** *Calculation of Molecular Structure and Energy by Force-Field Methods.* Advan. Phys. Organ. Chem., vol. 13, pages 1–82, 1976.
- [Allinger 77] **N.L. Allinger.** *Conformational analysis. 130. MM2. Hydrocarbon force field utilizing V1 and V2 torsional terms.* American Chem. Soc., vol. 99, n° 25, pages 8127–8134, 1977.
- [Allinger 89a] **N.L. Allinger et J.H. Lii.** *Molecular Mechanics. The MM3 Force Field for hydrocarbons. (2). Vibrational frequencies and Thermodynamics.* American Chem. Soc., vol. 111, pages 8566–8575, 1989.
- [Allinger 89b] **N.L. Allinger, Y.H. Yuh et J.H. Lii.** *Molecular Mechanics. The MM3 Force Field for hydrocarbons. (1).* American Chem. Soc., vol. 111, pages 8551–8565, 1989.
- [Allinger 96] **N.L. Allinger.** *An improved force field (MM4) for saturated Hydrocarbons.* J. of Comp. Chem, vol. 17, n° 5-6, pages 642–668, 1996.
- [Alvarado 03] **C. Alvarado et K. Kazerounian.** *On the rotational operators in protein structure simulations.* Journal of protein engineering, vol. 16, n° 10, pages 717–722, 2003.
- [Bayazit 01] **O. B. Bayazit, G. Song et N. M. Amato.** *Ligand binding with OBPRM and user input.* International Conference on Robotic and Automation, pages 954–959, 2001.

- [Berman 00] **H. Berman, J. Westbrook et al.** *The protein Data Bank*. Nucl. Acids Res., vol. 28, pages 235–242, 2000.
- [Boukhnifer 04a] **M. Boukhnifer et A. Ferreira.** *H₂ Optimal Controller Design for Micro-Teleoperation with Delay*. International Conference on Intelligent Robots and Systems, pages 224–229, 2004.
- [Boukhnifer 04b] **M. Boukhnifer, A. Ferreira et J.G. Fontaine.** *Scaled Teleoperation Controller Design for Micromanipulation over Internet*. International Conference on Robotics and Automation, pages 481–487, 2004.
- [Boukhnifer 06] **M. Boukhnifer et A. Ferreira.** *Stability and Transparency for Scaled Teleoperation System*. International Conference on Intelligent Robots and Systems, pages 4217–4222, 2006.
- [Bowen 91] **J.P. Bowen et N.L. Allinger.** *Molecular Mechanics : The Art and Science of Parametrization*. Reviews in Comp. Chem, vol. 2, pages 81–97, 1991.
- [Branden 91] **C. Branden et J. Tooze.** Introduction to protein structure. Garland, New-York, 1991.
- [Bris 02] **C. Le Bris.** *Problématiques numériques pour la simulation moléculaire*. Actes du 32e congrès national d'analyse numérique, vol. 11, pages 127–140, 2002.
- [Brooks 83] **B.R. Brooks et R.E. Bruccoleri.** *CHARMM : A program for macromolecular energy, minimization, and dynamics calculations*. J. of Comp. Chem, vol. 4, n° 2, pages 187–217, 1983.
- [Cooper 03] **C.F. Cooper.** *Molecular mechanics - modeling structure and energetics : the 2002 Benjamin Franklin Medal in chemistry presented to Norman Allinger*. Journal of the Franklin Institute, vol. 340, pages 191–203, 2003.
- [Cortés 03] **J. Cortés.** *Motion Planning Algorithms for General Closed-Chain Mechanisms*. Thèse de doctorat, Institut National Polytechnique de Toulouse, Décembre 2003.
- [Cortés 07] **Juan Cortés, Léonard Jaillet et Thierry Siméon.** *Molecular Disassembly with RRT-like Algorithms*. International Conference on Robotic and Automation, 2007.
- [Coutsias 03] **E.A. Coutsias, C. Seok et al.** *A kinematic View of Loop Closure*. Journal of Computer Chemistry, vol. 25, pages 510–528, 2003.
- [Engh 91] **A. Engh et R. Huber.** *Accurate bond and angle parameters for X-Ray protein structure refinement*. Acta Cryst., vol. A47, pages 392–400, 1991.
- [Featherstone 99a] **R. Featherstone.** *A divide-and-conquer articulated body algorithm for parallel $o(\log(n))$ calculation of rigid body dynamics. Part 1 : Basic algorithm*. International Journal of Robotics Research, vol. 18, n° 9, pages 867–875, 1999.
- [Featherstone 99b] **R. Featherstone.** *A divide-and-conquer articulated body algorithm for parallel $o(\log(n))$ calculation of rigid body dynamics. Part 2 : Trees, loops*

- and accuracy.* International Journal of Robotics Research, vol. 18, n° 9, pages 876–892, 1999.
- [Fields 00] **M.J. Fields, M. Albe et C. Bret et al.** *The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials.* Journal of Computational Chemistry, vol. 21, pages 1088–1100, 2000.
- [Gupta 86] **K.C. Gupta.** *Kinematic analysis of manipulators using the zero reference position description.* International Journal of Robotic Research, vol. 5, n° 2, pages 5–13, 1986.
- [Guthart 00] **G.S. Guthart et J. Salisbury.** *The intuitive telesurgery system : Overview and application.* International Conference on Robotics and Automation, vol. 1, pages 618–621, 2000.
- [Halgren 96a] **T.A. Halgren.** *Merck molecular force field (3). Molecular geometries and vibrational frequencies for MMFF94.* J. of Comp. Chem, vol. 17, n° 5-6, pages 553–586, 1996.
- [Halgren 96b] **T.A. Halgren et et al.** *Merck molecular force field (1). Basis, Form, Scope, parametrization, and performance of MMFF94.* J. of Comp. Chem., vol. 17, n° 5-6, pages 490–519, 1996.
- [Halgren 96c] **T.A. Halgren et et al.** *Merck molecular force field (2). MMFF94 van der waals and Electrostatic Parameters for intermolecular interactions.* J. of Comp. Chem., vol. 17, n° 5-6, pages 520–552, 1996.
- [Halgren 96d] **T.A. Halgren et et al.** *Merck molecular force field (4). Conformational Energies and geometries for MMFF94.* J. of Comp. Chem., vol. 17, n° 5-6, pages 587–615, 1996.
- [Hamdi 04] **M. Hamdi, A. Ferreira et J.G. Fontaine.** *Virtualized Reality Interface for Tele-Micromanipulation.* International Conference on Robotics and Automation, pages 2776–2781, 2004.
- [Hamdi 07] **M. Hamdi et A. Ferreira.** *Multiscale Design and Modeling of Nanorobots.* International Conference on Intelligent Robots ans Systems, pages 3821–3827, 2007.
- [Hannaford 03] **B. Hannaford, G. Hirzinger, C. Preusch et J. Ryu.** *Time domain passivity control with reference energy behavior.* : IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 2932–2937, 2003.
- [Hart 07] **J. S. Hart et G. Niemeyer.** *Design guidelines for wave variable controllers in time delayed telerobotics.* Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07), pages 182–187, 2007.
- [Head-Gordon 96] **M. Head-Gordon.** *Quantum chemistry and molecular processes.* J. Phys. Chem., vol. 100, pages 13213–13225, 1996.
- [Kazerounian 04a] **K. Kazerounian.** *From Mechanisms and robotics to protein conformation and drug design.* Journal of Mechanical Design, vol. 126, pages 40–45, 2004.

- [Kazerounian 04b] **K. Kazerounian, K. Latif et al.** *PrototfoldPart (Part 1) : Nanokinematics for analysis of protein molecules*. ASME, 2004.
- [Kazerounian 04c] **K. Kazerounian, K. Latif et al.** *PrototfoldPart (Part 2) : A successive kineto-static compliance method for protein conformation prediction*. ASME, 2004.
- [Kollman 00] **P.A. Kollman et I. Massova et al.** *Calculating structures and free energies of complex molecules : Combining molecular mechanics and continuum models*. American Chemical Society, vol. 33, pages 889–897, 2000.
- [Kumar 99] **S. Kumar, B. Ma, C.J. Tsai, H. Wolfson et R. Nussinov.** *Folding funnels and conformational transitions via hinge-bending motions*. Cell Biochem Biophys, vol. 31, n° 2, pages 141–164, 1999.
- [Lai-Yen 05] **S. K. Lai-Yen et Y. S. Lee.** *Computer-Aided molecular design (CAMD) with force-torque feedback*. International conference on computer aided design and computer graphics, pages 199–204, 2005.
- [Lai-Yen 06] **S. K. Lai-Yen et Y. S. Lee.** *Energy-field optimization and haptic-based molecular docking and assembly search system for computer-aided molecular design (CAMD)*. Symposium on haptic interfaces for virtual environment and teleoperator systems, pages 233–240, 2006.
- [Lavalle 00] **S.M. Lavalle, P.W. Finn, L. Kavraki et J.C. Latombe.** *A Randomized kinematics based approach to pharmacophore-constrained conformational search and database screening*. Journal of computational chemistry, vol. 21, n° 9, pages 731–747, 2000.
- [Leach 01] **A.R. Leach.** *Molecular modelling principles and applications*. Prentice Hall, second édition, 2001.
- [Lee 04] **Y. G. Lee et K. W. Lyons.** *Smoothing haptic interaction using molecular force calculations*. Computer-aided design, vol. 36, pages 75–90, 2004.
- [Levitt 76] **M. Levitt.** *A simplified representation of protein conformations for rapid simulation of protein folding*. Journal of Molecular Biology, vol. 104, pages 49–107, 1976.
- [Lii 89] **N.L. Allinger and J.H. Lii.** *Molecular Mechanics. The MM3 Force Field for hydrocarbons. (3). The van der Waals potential and Crystal data for aliphatic and aromatic hydrocarbons*. American Chem. Soc., vol. 111, pages 8576–8582, 1989.
- [Liu 98] **Y.P. Liu, K. Kim et B.J. Berne et al.** *Constructing ab initio forcefield for molecular dynamics simulations*. Journal of Chemical Physics, vol. 108, pages 4739–4755, 1998.
- [Lomas 86] **J.S. Lomas.** *La mécanique moléculaire, une méthode non quantique pour le calcul de la structure et de l'énergie d'entités moléculaires*. L'actualité chimique, Mai 1986.
- [Lotay 04] **I. Lotay.** *Algorithms exploiting the chain structure of proteins*. Thèse de doctorat, Stanford University, August 2004.

- [Manocha 95] **D. Manocha et Y. Zhu.** *Conformation analysis of molecular chains using Nano-Kinematics.* Computer applications of biological sciences, vol. 11, n° 1, pages 71–86, 1995.
- [McCammon 87] **J.A. McCammon et S.C. Harvey.** Dynamics of proteins and nucleic acids. Cambridge university press édition, 1987.
- [Meng 92] **E. C. Meng, B. K. Shoichet et I. D. Kuntz.** *Automated docking with grid-based energy calculation.* Journal of Computational Chemistry, vol. 13, n° 4, pages 505–524, 1992.
- [Mitchell 99] **L.G. Mitchell, N.A. Campbell et J.B. Reece.** Biology. Addison Wesley, fifth édition, 1999.
- [Morin 07] **S. Morin et S. Redon.** *A force-feedback algorithm for adaptive articulated-body dynamics simulation.* International Conference on Robotics and Automation, pages 3245–3250, 2007.
- [Munir 03] **S. Munir et W.J Book.** *Control techniques and programming issues for time delayed internet based teleoperation.* Journal of Dynamic Systems Measurement and Control, vol. 125, n° 2, pages 205–214, 2003.
- [Murray 94] **R. M. Murray, Z. Li et S. S. Sastry.** A mathematical introduction to robotic manipulation. CRC Press, 1994.
- [Nagata 02] **H. Nagata, H. Mizushima et H. Tanaka.** *Concept and prototype of protein-ligand docking simulator with force feedback technology.* Bioinformatics, vol. 18, pages 140–146, 2002.
- [Niemeyer 91] **G. Niemeyer et J. J. E. Slotine.** *Stable adaptive teleoperation.* IEEE Journal of oceanic engineering, vol. 16, pages 152–162, 1991.
- [Niemeyer 96] **G. Niemeyer.** *Using Wave Variables in Time Delayed Force Reflecting Teleoperation.* Thèse de doctorat, Massachusetts Institute of Technology, September 1996.
- [Niemeyer 98] **G. Niemeyer et J.J.E. Slotine.** *Towards force reflecting teleoperation over the internet.* International Conference on Robotics and Atomation, vol. 3, n° 3, pages 1909–1915, 1998.
- [Niemeyer 04] **G. Niemeyer et J.J.E. Slotine.** *Telemanipulation with time delays.* International Journal of Robotics Research, vol. 23, n° 9, pages 873–890, 2004.
- [Pattabiraman 86] **N. Pattabiraman, M. Levitt, T. E. Ferrin et R. Langridge.** *Computer graphics in real-time docking with energy calculation and minimization.* Journal of computational chemistry, vol. 6, n° 5, pages 432–436, 1986.
- [Redon 05] **S. Redon, N. Galoppo et M. C. Lin.** *Adaptive Dynamics of Articulated Bodies.* ACM Transactions on Graphics, vol. 24, n° 3, 2005.
- [Richardson 81] **J.S. Richardson.** *The anatomy and taxonomy of protein structure.* Advanced Protein Chem., vol. 34, pages 167–339, 1981.
- [Sheridan 92] **T.B. Sheridan.** Telerobotics, automation, and human supervisory control. The MIT Press, Cambridge, 1992.

- [Sheridan 93] **T. B. Sheridan.** *Space teleoperation through time delay : Review and prognosis.* IEEE Transactions on Robotics and Automation, vol. 9, n° 5, pages 592–606, 1993.
- [Singh 97] **A.P. Singh et D.L. Brutlag.** *Hierarchical protein structure superposition using both secondary and atomic representation.* In Proc. Int. Conf. Intell. Syst. Mol. Biol., pages 284–293, volume 5, 1997.
- [Singh 99] **A.P. Singh, J.C. Latombe et D.L. Brutlag.** *A Motion Planning Approach to Flexible Ligand Binding.* Int. Conf. on Intelligent Systems for Molecular Biology (ISMB), pages 252–261, 1999.
- [Stone 01] **J. Stone, J. Gullingsrud, P. Grayson et K. Schulten.** *A system for interactive molecular dynamics simulation.* ACM Symposium on Interactive 3D Graphics, pages 191–194, 2001.
- [Tomioka 94] **N. Tomioka et A. Itai.** *Green : a program package for docking studies in rational drug design.* Journal of computer-aided molecular design, vol. 8, n° 4, pages 347–366, 1994.
- [Weiner 84] **S.J. Weiner, P.A. Kollman et et al.** *A new force field for molecular mechanical simulation of nucleic acids and proteins.* Journal of American Chemical Society, vol. 106, n° 3, pages 765–784, 1984.
- [Wollacott 07] **A. M. Wollacott et K. M. Merz Jr.** *Haptic applications for molecular structure manipulation.* Journal of Molecular Graphics and Modelling, vol. 25, n° 6, pages 801–805, 2007.
- [Zhang 02] **M. Zhang et L. Kavradi.** *A new method for fast and accurate derivation of molecular conformations.* Journal of American Chemical Society, vol. 42, pages 64–70, 2002.