



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Université de Nancy 1

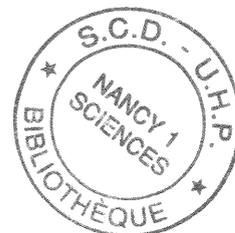
U.F.R Sciences et Techniques de la Matière et des Procédés
G.F.D Chimie et Physico-Chimie Moléculaire et théorique



**Thèse de Doctorat de l'Université de Nancy 1
en Chimie Informatique et Théorique**

Présentée par

Ahmed MOKRANE



**MODELISATION DE LA REACTIVITE ENZYMATIQUE
PAR LE FORMALISME *LSCF* ET ETUDE DES EFFETS ELECTROSTATIQUES
A LONGUE DISTANCE. APPLICATION A L'ACTIVATION DU RECEPTEUR
DE LA THROMBINE**

Soutenue publiquement devant la commission d'examen le 12 novembre 1997

Membres du jury :

Président : J. L. Rivail Professeur, Université Henri Poincaré, Nancy 1

Rapporteurs : G. Vergoten Professeur, U.S.T.L, Lille

G Folkers Professeur, E.T.H, Zurich

Examineurs :

J. Smith Directeur de recherche, CEA Saclay

B. Maigret Directeur de recherche, CNRS, Nancy

A. Cartier Chargé de recherche, CNRS, Nancy

Ce travail a été réalisé au sein du Laboratoire de Chimie théorique de l'Université Henri Poincaré-Nancy 1, dirigé par Monsieur le Professeur J.L. Rivail. Je tiens à lui exprimer toute ma gratitude pour son accueil et pour son précieux soutien scientifique.

J'exprime toute ma reconnaissance au docteur B. Maigret, directeur de l'U.R.A. 510 pour m'avoir permis de réaliser cette thèse. L'intérêt constant qu'il a manifesté à ce travail m'a beaucoup encouragé.

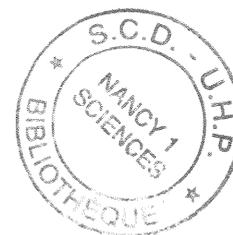
J'adresse mes plus vifs remerciements à Monsieur A. Cartier pour avoir accepté de diriger cette thèse, pour son soutien jusque dans les moments difficiles. Qu'il trouve ici l'expression de mon amitié indéfectible.

Je tiens à remercier Messieurs G. Vergoten, Professeur à l'U.S.T.L, Lille et G. Folkers, Professeur à l'U.T.H, Zurich, qui ont bien voulu me faire l'honneur de juger ce travail en tant que rapporteurs.

Je remercie également Monsieur le Dr. J. Smith, Directeur de recherche au C.E.A, Saclay, pour avoir bien voulu participer aux travaux de ce jury.

J'associe enfin à mes remerciements tous mes collègues du laboratoire, en particulier D. Rinaldi et M. Ruiz-López, pour leur aide amicale, leur soutien et leurs précieux conseils.

Madame J. Grillot a bien voulu relire ce manuscrit, je l'en remercie bien sincèrement.



SOMMAIRE

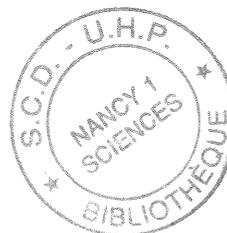
INTRODUCTION

4

CHAPITRE 1 : LES MÉTHODES DE LA CHIMIE QUANTIQUE ET DE LA MODÉLISATION MOLÉCULAIRE

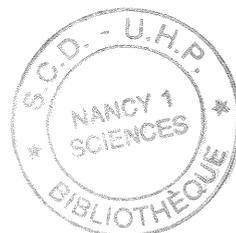
1. La Mécanique Quantique en Chimie Théorique	7
1.1 Principes	7
1.2. Approximation de Born-Oppenheimer	8
1.3 Théorie du champ auto-cohérent	9
1.4 Les calculs <i>abinitio</i>	11
1.5 Les méthodes semi-empiriques	12
2. Méthodes de la Mécanique Moléculaire	14
3. Minimisation de l'énergie	16
4. Méthodes Hybrides : Mécanique Quantique-Mécanique Moléculaire (MQ-MM)	18
4.1 Aperçu des différentes méthodes mixtes MQ-MM	18
4.2 La méthode du champ auto-cohérent local (LSCF)	20
4.2.1 Principe de la méthode	20
4.2.2 Construction des orbitales hybrides	21
4.2.3 Expression de l'hamiltonien LSCF	24
4.2.4 Expression de l'opérateur de Hartree-Fock LSCF	25
4.2.5 Expression de l'énergie dans le formalisme LSCF	26

4.2.6 Dérivées de l'énergie dans le formalisme LSCF	28
5. Chemin réactionnel	30
6. Le modèle du continuum	32
6.1 Introduction	32
6.2 Modèle de Pise (Tomasi et al.)	33
6.3 Modèle à cavité de Nancy	34
Bibliographie	37
CHAPITRE 2 : EFFET DES INTERACTIONS À LONGUE DISTANCE	
1. Champ électrostatique moyen	40
1.1 Introduction	40
1.2 Perturbation électrostatique	41
1.2.1 Répartition du système	41
1.2.2 Expression du champ électrostatique moyen	42
1.3 Matrice de Hartree-Fock et expression de l'énergie	45
1.4 Expression de la dérivée première de l'énergie.	46
1.5 Validation	48
1.5.1 Choix du cutoff	49
1.5.2 Application	50
2. Prise en compte de la polarisation de l'enzyme (Modèle du continuum)	51
2.1 Introduction	51
2.2 Modèle à cavité	52
Bibliographie	55

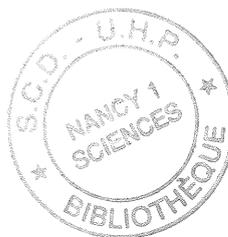


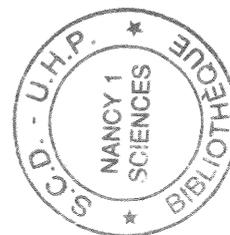
CHAPITRE 3 : APPLICATION À LA CATALYSE ENZYMATIQUE

1. Les enzymes	56
1.1 Composition et structure	57
1.2 Fonction.	58
1.3 La cinétique enzymatique	60
1.4 Stabilité des enzymes et les facteurs catalytiques	63
1.4.1 Les contraintes stériques	63
1.4.2 Les effets entropiques	64
1.4.3 Les effets électrostatiques	65
1.5 La modélisation de la catalyse enzymatique par LSCF	65
2. Protéases à sérine : la thrombine	67
2.1 Fonction et description	67
2.2 Triade catalytique des protéases à sérine	70
2.3 Mécanisme catalytique des protéases à sérine	71
3. Le récepteur de la thrombine	76
3.1 Description	76
3.2 Mécanisme d'activation du récepteur	77
4. Modélisation du complexe thrombine/récepteur par la méthode de Mécanique Moléculaire (MM)	79
5. Modélisation de l'activation du complexe thrombine/RS par la méthode LSCF	82
5.1 Le complexe de Michaelis	86
5.2 Chemins réactionnels vers l'intermédiaire tétraédrique	88
5.3 Etape intermédiaire tétraédrique-acylenzyme	92



5.4 Etape acylenzyme-intermédiaire tétraédrique	94
5.5 Etape intermédiaire tétraédrique-enzyme/produits(EP)	97
6. Etudes des effets électrostatiques et stériques de la protéine	100
6.1 Etude des profils énergétiques	100
6.2 Effets électrostatiques et stériques de l'environnement.	102
6.3 Effets du champ électrostatique moyen	105
7. Effets électrostatiques dus à la polarisation : Modèle à Cavité	110
Bibliographie	118
CONCLUSION	121





Introduction

Les catalyseurs du monde vivant sont indispensables aux mécanismes les plus complexes qui composent celui-ci et figurent au plus bas niveau du fonctionnement de la vie. Ces catalyseurs sont les enzymes qui séduisent par leur efficacité et sont l'objet d'une grande curiosité scientifique. Ils sont devenus en même temps les outils de nombreuses applications à caractère thérapeutique. Bien que les enzymes soient soumises aux mêmes lois de la nature qui régissent le comportement des autres substances, ils diffèrent des catalyseurs chimiques ordinaires par un grand degré de spécificité envers les substrats et une vitesse de réaction élevée.

L'une des premières familles d'enzymes à avoir été étudiée de manière intensive a été celle des protéases à sérine, qui se caractérisent par la présence d'un pôle sérine essentiel dans leur site catalytique. Parmi cette famille d'enzyme, la thrombine, clé de voûte de l'hémostase, possède des effets physiologiques complexes en agissant comme agent coagulant, anticoagulant (activation de la protéine C) ou stimulant du mécanisme inflammatoire. La thrombine, par son rôle dans les interactions sang-vaisseaux, représente une cible pharmacologique multifonctionnelle de première importance. La reconnaissance de la thrombine se fait par un mécanisme particulier qui implique une protéolyse du récepteur et l'élimination des 41 premiers résidus de la séquence. L'activation du récepteur est causée par l'apparition d'un nouveau N-terminal après hydrolyse, et le mécanisme ressemble à celui des protéases à sérine que nous avons modélisé dans ce mémoire. Mieux comprendre le mécanisme d'interaction du complexe thrombine/récepteur permet la mise au point de molécules pharmacologiques actives qui pourraient, soit bloquer le récepteur, soit la thrombine.

Dans ce type de mécanisme réactionnel, un réseau de liaisons hydrogène, dans le site actif, assure la stabilité des intermédiaires de la réaction, et pour étudier d'éventuels transferts de proton, nous devons utiliser une méthode de calcul quantique. En effet, les approches empiriques sont fondées sur les techniques de mécanique moléculaire (MM) et de dynamique moléculaire (DM), utilisant des fonctions empiriques pour le calcul de l'énergie du système. Ceci implique que, si ces méthodes offrent la possibilité d'étudier des systèmes très larges, elles ne peuvent en revanche rendre compte de phénomènes

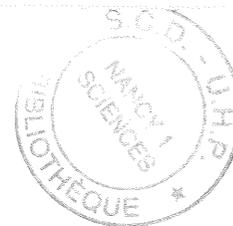


essentiels tels que la création ou la rupture des liaisons chimiques. De plus, elles sont généralement paramétrisées avec des données expérimentales fournissant peu d'informations sur les parties des surfaces d'énergie potentielle éloignées des configurations stables ou d'équilibre. Malheureusement, une étude quantique de tout le système thrombine-récepteur n'est pas réalisable en raison de la taille du système. On peut néanmoins remédier aux inconvénients de ces deux méthodes, quantique et empirique, en combinant à la fois les propriétés de la mécanique moléculaire pour représenter l'essentiel du système, et de la mécanique quantique pour traiter la partie du système qui subit l'essentiel des modifications électroniques au cours de la réaction. Ce sont les méthodes *mixtes ou hybrides*.

La méthode LSCF (Local Self Consistent Field), développée dans notre laboratoire permet de combiner le traitement quantique d'un petit sous-système, ici le site catalytique et son substrat, avec un traitement classique sur le reste du système. Elle est fondée sur la représentation de chaque liaison séparant le fragment quantique du fragment classique au moyen d'une orbitale strictement localisée supposée connue. Dans cette méthode, la contribution électrostatique de l'environnement protéique est représentée par un jeu de charges ponctuelles localisées sur les atomes.

Le calcul de la perturbation, dans le formalisme LSCF, due à ces charges sur les fonctions de base de la partie quantique prend un temps de calcul considérable. Afin de réduire ce temps de calcul tout en représentant correctement le potentiel électrostatique de l'environnement (choix imposé par la taille des systèmes auxquels on s'intéresse, et qui dans l'ensemble agit d'une manière assez globale), nous avons développé une méthode qui nous permet de représenter la perturbation de la partie lointaine de l'environnement par un potentiel en une expansion multipolaire de $1/r$ au centre de la partie quantique. De cette façon le champ créé ne sera calculé qu'une seule fois au cours de l'optimisation de géométrie, plutôt qu'à chaque fois comme dans l'ancienne version du programme. Cette amélioration du programme a permis de gagner un temps de calcul considérable (accélération d'un facteur moyen de 3), au prix d'une perte de précision infime. Nous nous attacherons à étudier l'influence de ce champ électrostatique moyen sur le mécanisme réactionnel de la thrombine.

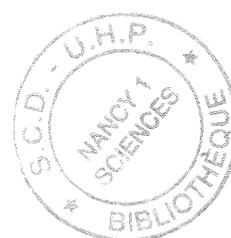
Le procédé de prise en compte de l'environnement protéique est perfectible et une représentation du potentiel électrostatique moléculaire par un jeu de charges fait

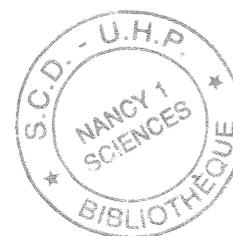


encore l'objet de recherches. Cet environnement protéique dans son ensemble intervient de façon non négligeable au niveau du mécanisme réactionnel en stabilisant ou en déstabilisant les réactants, l'intermédiaire réactionnel ou les produits. En général, dans les modèles classiques, la protéine est représentée comme un solvant non polaire tant que ses moments permanents et le solvant sont pris en compte explicitement et non par l'intermédiaire de la constante diélectrique. Ainsi, si toutes les contributions polaires étaient considérées explicitement, en incluant les dipôles permanents et induits de la protéine et leurs relaxations par la réorganisation des charges, la constante diélectrique devrait être réduite à l'unité. Il est clair que la polarisation du système protéique joue un rôle important dans la détermination des structures et des fonctions des protéines. Ce phénomène peut être modélisé en supposant que l'enzyme joue le rôle du solvant de constante diélectrique faible dans une région interne et localisé en l'occurrence le site actif de l'enzyme. De cette façon on superpose à l'effet des charges classique de l'environnement protéique, un champ de réaction dû à la polarisation du système.

Le présent mémoire est donc organisé de la façon suivante. Dans un premier chapitre, nous nous attacherons à rappeler les bases théoriques de la méthodologie de calcul. Le chapitre II est consacré aux améliorations méthodologiques que nous avons mises au point. Le dernier chapitre traite de la modélisation d'une réaction enzymatique d'actualité, l'activation du récepteur de la thrombine, par hydrolyse d'une liaison peptidique. L'étude sera réalisée par la méthode LSCF en prenant en compte l'environnement enzymatique et, de ce fait, par comparaison avec le chemin réactionnel dans le vide (sans environnement protéique), on dégagera certains aspects de la catalyse enzymatique (effets électrostatique et stérique).

Nous nous arrêterons ensuite sur les effets à longue distance, l'effet du champ électrostatique moyen et l'effet du solvant. Dans ce modèle de solvant le fragment quantique est entouré par un continuum, l'environnement protéique, de constante diélectrique faible, dans le but de tenir compte de la polarisation du système classique. Le mémoire se termine par une conclusion qui envisage d'autres développements de la méthodologie.





Chapitre 1

Les Méthodes de la Chimie Quantique et de la Modélisation Moléculaire

1. La Mécanique Quantique en Chimie Théorique

1.1. Principes

Etudier la structure ou la réactivité d'un système en chimie théorique consiste essentiellement à déterminer le comportement des électrons de ce système. Le seul outil capable de fournir des informations précises sur la densité électronique ainsi que sa variation au cours d'une réaction est la mécanique quantique. Celle-ci consiste à résoudre pour le système considéré, que nous supposons non relativiste et n'évoluant pas dans le temps, l'équation de Schrödinger :

$$\hat{H}\Psi = E\Psi \quad 1.1$$

où \hat{H} est l'opérateur hamiltonien du système ; E son énergie ; et Ψ une fonction d'onde représentant l'état du système d'énergie E .

En nous plaçant dans le cas où les électrons sont considérés comme non-relativistes, nous pouvons scinder l'hamiltonien en plusieurs termes, relatifs aux noyaux et aux électrons :

$$\hat{H} = \hat{T}_N + \hat{T}_e + \hat{V}_{eN} + \hat{V}_{ee} + \hat{V}_{NN} \quad 1.2$$

où \hat{T}_N et \hat{T}_e représentent les opérateurs d'énergie cinétique respectivement des noyaux et des électrons et \hat{V}_{eN} , \hat{V}_{ee} et \hat{V}_{NN} sont les termes d'énergie potentielle respectivement d'attraction électrons-noyaux, d'interaction entre les électrons et de répulsion entre les noyaux.

1.2. Approximation de Born-Oppenheimer

L'une des premières approximations possibles¹ consiste à séparer le mouvement des noyaux de celui des électrons. Comme les noyaux sont beaucoup plus lourds que les électrons, ils se déplacent plus lentement. On peut donc, dans une bonne approximation, considérer que les électrons d'un système se déplacent dans un champ de noyaux fixes dans l'espace. Cela revient alors à négliger l'énergie cinétique des noyaux et à considérer le terme V_{NN} comme n'ayant qu'un rôle de paramètre dans la définition de l'hamiltonien et donc dans le calcul de l'énergie. On définit alors un opérateur \hat{H}' par :

$$\hat{H}' = \hat{H}_e + V_{NN} \quad 1.3$$

Avec

$$\hat{H}_e = T_e + V_{eN} + V_{ee} \quad 1.4$$

Nous obtenons un hamiltonien électronique \hat{H}_e , ayant pour fonction propre la fonction Ψ_e , fonction propre de \hat{H}' , qui décrit le comportement des électrons dans une molécule. Dans l'équation (1.1), E décrit alors la contribution électronique à l'énergie. En unités atomiques, le développement des différents termes de l'hamiltonien électronique donne l'équation suivante :

$$\hat{H}_e = -\frac{1}{2} \sum_i^N \Delta_i - \sum_K^M \sum_j^N \frac{Z_K}{R_{jk}} + \sum_i^N \sum_{j<i}^N \frac{1}{r_{ij}} \quad 1.5$$

où les indices i et j se rapportent aux N électrons du système, alors que K est relatif aux M noyaux.

1.3 Théorie du champ auto-cohérent

Les méthodes MO-LCAO expriment la fonction d'onde totale de la molécule comme le produit antisymétrisé des orbitales moléculaires:

$$\Psi = \left| \prod_i \psi_i \right| \quad 1.6$$

La fonction propre doit respecter le principe d'exclusion de Pauli selon lequel la fonction d'onde d'un système multiélectronique doit changer de signe lors de la permutation des coordonnées de deux électrons quelconques. Dans le cadre de l'approximation orbitale, la solution approchée de la fonction d'onde évoquée ci-dessus se met sous la forme d'un déterminant de Slater² :

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_a(1) & \chi_b(1) & \cdots & \chi_p(1) \\ \cdots & \cdots & \cdots & \cdots \\ \chi_a(i) & \cdots & \cdots & \chi_p(i) \\ \cdots & \cdots & \cdots & \cdots \\ \chi_a(N) & \cdots & \cdots & \chi_p(N) \end{vmatrix} \quad 1.7$$

où $\chi_a(i)$ est une spinorbitale, produit d'une orbitale et de sa fonction de spin associée.

L'orbitale moléculaire ψ_i est généralement représentée par une combinaison linéaire des orbitales atomiques φ_j constituant la base de fonctions choisie,

$$\psi_i = \sum_j c_{ij} \varphi_j \quad 1.8$$

d'où la désignation LCAO (*Linear Combination of Atomic Orbitals*).

On détermine les coefficients c_{ij} pour la construction des orbitales moléculaires à l'aide de la méthode du champ auto-cohérent (SCF).

Ce problème revient à résoudre les équations de Hartree-Fock³, qui prennent la forme :

$$F\chi_k = e_k\chi_k \quad 1.9$$

définissant un ensemble de fonctions parmi lesquelles se trouvent des spinorbitales permettant de construire un déterminant de Slater qui approche le mieux la fonction d'onde multiélectronique du système étudié. En raison de la nature variationnelle de la méthode, la valeur calculée de l'énergie s'approche de la valeur exacte par valeur supérieure, et tend, avec l'augmentation de la qualité de la base de fonctions utilisée, vers une limite, appelée "limite de Hartree-Fock", due à l'approximation orbitales (MO) qui ne prend pas en compte la corrélation existant entre le mouvement des divers électrons.

L'opérateur monoélectronique de Hartree-Fock est défini comme :

$$F(1) = \hat{H}^c(1) + \sum_i [J_i(1) - K_i(1)] \quad 1.10$$

où \hat{H}^c est l'hamiltonien de cœur associé à un seul électron et défini par :

$$\hat{H}^c = \left[-\frac{1}{2} \Delta_i + \sum_K \frac{Z_K}{r_{iK}} \right] \quad 1.11$$

J et K sont les opérateurs biélectroniques coulombien et d'échange, définis par :

$$J_i(1) = \int_{(2)} \chi_i^*(2) \frac{1}{r_{12}} \chi_i(2) d\tau_2 \quad 1.12$$

$$K_i(1)\chi_t(1) = \chi_i(1) \int_{(2)} \chi_i^*(2) \frac{1}{r_{12}} \chi_t(2) d\tau_2 \quad 1.13$$

L'énergie calculée avec l'hamiltonien électronique \hat{H} de ce système a pour expression :

$$E = \sum_k \langle \chi_k(1) | \hat{H}^c(1) | \chi_k(1) \rangle + \sum_k \sum_{l>k} \left[\langle \chi_k(1)\chi_l(2) | \frac{1}{r_{12}} | \chi_k(1)\chi_l(2) \rangle - \langle \chi_k(1)\chi_l(2) | \frac{1}{r_{12}} | \chi_k(2)\chi_l(1) \rangle \right] \quad 1.14$$

On retrouve les coefficients c_{ij} selon une procédure itérative : à partir de valeurs initiales de ces coefficients, les éléments F_{ij} de la matrice de Hartree-Fock sont calculés, les équations sont résolues et les nouveaux coefficients obtenus servent de point de départ pour le cycle de calcul suivant; cette procédure est répétée jusqu'à ce que le système atteigne une cohérence interne (variation de la matrice densité inférieure ou égale à 10^{-6}), d'où le nom auto-cohérent.

1.4 Les calculs abinitio

Deux formalismes sont alors possibles :

(1) les orbitales ou les spinorbitales utilisées sont orthogonales et on peut les multiplier par les fonctions de spin α ou β . On se place dans le cadre Hartree-Fock restreint (**RHF**) qui, pour les systèmes à couches fermées, impose à chaque orbitale de représenter deux électrons. Nous obtenons ainsi les équations de Roothaan⁴ :

$$\sum_{\eta} C_{\eta\mu} [F_{\eta\mu} - e_i S_{\eta\mu}] = 0 \quad 1.15$$

(2) il est en revanche à noter, que lorsque les distances interatomiques deviennent importantes, la fonction d'onde du système tend vers le produit de deux fonctions d'espace différentes. Il faut donc utiliser deux orbitales moléculaires différentes à toutes les distances interatomiques. On est alors dans le cadre du formalisme Hartree-Fock non restreint (**UHF**). Une autre imprécision est apportée du fait du développement des orbitales moléculaires sur une base de gaussiennes. En effet, la précision des calculs dépend de l'étendue de la base. L'énergie ainsi calculée approche de la valeur exacte par valeur supérieure et tend vers une limite appelée limite Hartree-Fock.

Malheureusement, l'approche Hartree-Fock nécessite le calcul d'un nombre important d'intégrales biélectroniques multicentriques qui sont difficiles à obtenir avec des orbitales de Slater. On développe alors les exponentielles $e^{-\zeta r}$ sur une base de fonctions gaussiennes de la forme $e^{-\omega^2 r^2}$ qui présentent la propriété suivante : le produit de deux gaussiennes centrées en deux points différents est équivalent à une gaussienne unique. Ainsi toutes les intégrales biélectroniques multicentriques se ramènent à des sommes d'intégrales monocentriques, calculables aisément. L'utilisation de ces bases de fonctions gaussiennes permet de calculer toutes les intégrales intervenant dans les calculs SCF sans autres approximations que celles provenant de la méthode elle-même. C'est pourquoi ces calculs sont appelés calculs **abinitio**. Si les fonctions peuvent bien représenter les orbitales pour des distances interatomiques correspondant à des phénomènes de liaisons chimiques, il faut, pour des distances interatomiques élevées, utiliser des bases très étendues et multi-dzéta. On peut ainsi corriger la tendance des gaussiennes à s'annuler plus rapidement que les orbitales de Slater. L'autre problème se situe au niveau des noyaux. En effet, les fonctions gaussiennes ont la particularité de posséder une dérivée nulle pour $r = 0$, ce qui ne permet pas de représenter l'énergie potentielle électrons-noyaux et la répulsion électronique au voisinage des noyaux. Cependant, les fonctions gaussiennes représentent de façon satisfaisante les orbitales de Slater dans la région où a lieu la plupart des phénomènes chimiques et sont donc à même de rendre compte de la réactivité chimique.

1.5 Les méthodes semi-empiriques

L'un des problèmes majeurs des calculs ab-initio est le nombre très élevé des intégrales biélectroniques qu'il est nécessaire de calculer, limitant ainsi les études à des systèmes comportant un nombre réduit d'atomes (actuellement quelques dizaines tout au plus). En réponse à ce problème ont été développées des méthodes semi-empiriques qui réduisent de manière importante le nombre des intégrales à calculer, tout en utilisant, en compensation, des données tirées de l'expérience. Ces méthodes sont fondées sur les approximations suivantes :

- (1) La base d'orbitales utilisée est constituée par les orbitales de Slater de valence.
- (2) Les intégrales de recouvrement sont négligées dans la résolution des équations SCF.

- (3) Toutes les intégrales biélectroniques à trois et à quatre centres sont supposées nulles.
- (4) Les termes non-diagonaux de la matrice de l'hamiltonien de coeur sont estimés au moyen de relations empiriques reposant sur l'hypothèse que ces intégrales sont proportionnelles à l'intégrale de recouvrement des orbitales atomiques concernées.
- (5) La plupart des intégrales mono ou biélectroniques à un centre sont estimées à partir de données tirées des spectres électroniques des atomes des éléments considérés, entre autres les paramètres de Slater-Condon. Ces paramètres sont ajustables de façon à obtenir la meilleure concordance entre les calculs et l'expérience, et ceci, en fonction de la grandeur que l'on veut calculer.

Parmi les méthodes semi-empiriques, nous pouvons citer les méthodes du type :

- (1) *CNDO*⁵ (Complete Neglect of Differential Overlap) dans laquelle les recouvrements différentiels sont complètement négligés. Néanmoins, les valeurs obtenues pour la géométrie et les distributions de charges sont acceptables, par contre, les valeurs de l'énergie restent éloignées des valeurs expérimentales. Cette méthode est inapte à prendre en compte les différences d'énergie d'interaction entre électrons de même spin et électrons de spins contraires, puisque les intégrales d'échange sont négligées. Par conséquent, elle ne permet pas le calcul de certaines grandeurs liées aux spins électroniques.
- (2) *INDO*⁶ (Intermediate Neglect of Differential Overlap) qui partage un grand nombre de points communs avec la méthode CNDO, mis à part l'introduction des intégrales d'échange monocentriques qui sont évaluées à partir des paramètres de Slater-Condon.
- (3) Les méthodes de type *NDDO*⁷ (Neglect of Diatomic Differential Overlap) dans lesquelles on néglige toutes les intégrales faisant intervenir le recouvrement différentiel de deux orbitales centrées sur deux atomes différents. Une version modifiée par Dewar, la méthode *MNDO*⁸ tire toutes les intégrales de valeurs expérimentales à l'aide de la méthode d'Oleari. Cette méthode est très bien adaptée au calcul de la plupart des grandeurs (énergie de formation, géométrie moléculaire, moments dipolaires), en revanche, elle ne représente pas bien les liaisons hydrogène. Pour palier cette difficulté, d'autres versions révisées ont vu le jour, telle que *AM1*⁹ (Austin Model) ou *PM3*¹⁰ (Parametric Method 3) qui diffèrent de la première par les valeurs de certains paramètres, et l'ajout de paramètres supplémentaires.

Les méthodes de la mécanique quantique traitent les molécules au niveau de leurs représentations électroniques. Elles permettent d'étudier précisément les propriétés électroniques des systèmes considérés mais également leurs variations lors du mécanisme réactionnel, cependant ces calculs deviennent très coûteux en temps de calcul lorsque le nombre d'atomes du système étudié est très important. Les méthodes semi-empiriques sont de plus en plus utilisées dans le cadre de la réactivité biologique, puisqu'elles donnent la possibilité d'étudier des systèmes plus importants tels que le site actif d'un enzyme, mais ne permettent pas de modéliser l'enzyme entièrement.

2. Méthodes de la Mécanique Moléculaire

Du fait du nombre élevé d'atomes qui constituent une macromolécule, les seules techniques de modélisation utilisables sont celles dans lesquelles l'énergie d'un système est calculée à l'aide d'une représentation empirique fondée sur la transférabilité des propriétés de certains groupements d'atomes d'une molécule à l'autre. Ces méthodes de la mécanique moléculaire sont fondées sur une toute autre méthodologie que les méthodes de la mécanique quantique. En effet, ici, les molécules seront traitées de manière classique. L'énergie du système, ne comportant que des termes d'énergie potentielle, s'écrit :

$$E = E_C + E_{vdw} + E_S + E_H + E_m \quad 1.14$$

où E_C représente l'énergie coulombienne, E_{vdw} l'énergie de Van der Waals, E_S l'énergie stérique, E_H l'énergie des liaisons hydrogène et E_m un terme énergétique fonction de la méthode.

L'énergie coulombienne représente les interactions qui existent entre les différentes charges réparties dans la molécule : charges atomiques partielles, ionisation de certains groupements.

Le terme représentant l'énergie de van der Waals, fonction de la distance entre atomes non liés, regroupe deux types d'interactions :

(1) les phénomènes de répulsion à courte distance,

(2) les phénomènes d'attraction à des distances plus importantes qui correspondent à un terme permettant de calculer l'énergie de dispersion.

Deux types de potentiels permettent de représenter ces interactions, le potentiel de Lennard-Jones et le potentiel de Buckingham, qui utilisent un terme en $1/r^6$ pour représenter l'énergie de dispersion. Le potentiel de Lennard-Jones prend en compte l'énergie de répulsion grâce à un terme en $1/r^{12}$, alors que celui de Buckingham utilise une exponentielle.

Pour toute molécule, on peut déterminer une conformation idéale où les paramètres géométriques (longueurs de liaison, angles de valence) garderaient une valeur théorique. Or, lors de notre étude, cette molécule est soumise à des champs de force complexes qui conduisent à des déformations géométriques de notre système. Le terme d'énergie stérique a pour but de calculer la différence énergétique entre ces deux états. Pour cela, on utilise la loi de Hooke :

$$E_s = E_{st} + E_b + E_{sb} + E_t \quad 1.15$$

E_{st} représente l'énergie d'étirement (*stretching*) qui tient compte des modifications des longueurs des liaisons. E_b , énergie de déformation angulaire (*bending*), correspond aux modifications apportées dans les angles de liaison. E_{sb} est un terme croisé (*stretching-bending*), qui traduit l'interaction entre les deux termes précédents. Enfin E_t (*torsion*) correspond aux modifications d'énergie qu'il est nécessaire d'apporter pour corriger les barrières de rotation. Pour que l'énergie des liaisons hydrogène soit prise en compte dans les interactions de van der Waals, on ajoute un terme électrostatique à un potentiel de type Lennard-Jones ou Buckingham.

Les méthodes de la mécanique moléculaire sont bien adaptées à l'étude de systèmes complexes, comportant un nombre de molécules ou d'atomes important, mais leur utilisation reste néanmoins limitée à des problèmes essentiellement structuraux. En plus ces méthodes supposent que les liaisons chimiques sont bien établies, ceci rend difficile la représentation des phénomènes chimiques qui accompagnent les changements dans la structure des système étudiés (formation ou coupure de liaisons).

3. Minimisation de l'énergie

Les méthodes de minimisation d'une fonction à N variables sont toutes basées sur le même principe : à chaque étape du calcul, le minimum de la fonction est déterminé suivant une direction donnée. Les principaux algorithmes disponibles diffèrent par la méthode qu'ils utilisent pour définir les directions de recherche successives. Parmi ces méthodes on peut citer :

(1) *Steepest descent*^{11,12}

La direction de recherche est déterminée par le gradient de la fonction. C'est une technique très simple, dans laquelle le vecteur déplacement est choisi dans la même direction mais dans le sens contraire du gradient. Cette méthode est surtout efficace dans le cas d'une minimisation préliminaire, visant l'élimination des mauvais contacts de van-der Waals entre les atomes de la molécule. Par contre, à proximité du minimum du potentiel, là où la pente de la surface d'énergie est beaucoup moins grande, cette méthode a tendance à trouver le minimum le plus proche, sans pouvoir franchir les barrières de potentiel peu élevées. La convergence de cette méthode au voisinage du minimum n'est pas bonne du fait que chaque nouveau vecteur déplacement est orthogonal au vecteur calculé dans le pas précédent. C'est pourquoi elle est très souvent utilisée dans des procédures hybrides, en préliminaire à un autre algorithme plus rapide au voisinage de l'extremum.

(2) *Gradients conjugués*¹³

Cet algorithme présente la particularité d'affiner la direction de recherche à chaque étape de la minimisation, en tenant compte des directions antérieures. Le protocole de minimisation prend en compte les résultats de deux itérations successives (déplacement ou gradient) qui seront cumulés afin de déterminer le vecteur du déplacement dans l'étape suivante. Par cette méthode, il est possible de franchir, avec certaines limites, les barrières de potentiel. Sa convergence vers un minimum d'énergie est meilleure que celle de la méthode précédente.

(3) *Newton-Raphson*¹⁴

Ces méthodes utilisent les dérivées secondes de la fonction à minimiser pour déterminer le point de changement de direction. Le hessien n'est pas estimé mais calculé intégralement à chaque itération. Ces méthodes sont très répandues et sont notamment très souvent couplées à une procédure de type " steepest descent ", car elles ont l'avantage de converger rapidement au voisinage du minimum.

(4) *Quasi Newton-Raphson*¹⁵

Cette dernière catégorie de méthodes utilise le principe de Newton-Raphson que nous venons de décrire, en calculant le hessien numériquement, à partir des dérivées premières.

(5) Méthode de Schlegel (Berny)¹⁶

A chaque cycle d'optimisation le hessien est modifié selon les résultats collectés au long de plusieurs itérations précédentes. Il est estimé selon la relation :

$$G_{ij} = \frac{\langle (g_i - g_k) | r_j \rangle - \sum_{m=i+1}^{k-1} G_{mj} \langle (x_i - x_k) | r_m \rangle}{\langle (x_i - x_k) | r_i \rangle} \quad 1.16$$

où $|g_i\rangle$ est le gradient d'une fonction à un point $|x_i\rangle$ de l'espace avec $j \leq i$; $i = k-1, k+2$.

Les vecteurs unitaires du sous-espace sont :

$$\bar{r}_i = |x_i - x_k\rangle - \sum_{m=i+1}^{k-1} |r_m\rangle \langle (x_i - x_k) | r_m \rangle \quad 1.17$$

$$r_i = \frac{\bar{r}_i}{|\bar{r}_i|} \quad 1.18$$

La correction finale apportée au hessien est donc :

$$G_k = G_{k-1} + \sum_{i \leq j} [G_{ij} - \langle r_i | G_{k-1} | r_j \rangle] [|r_i\rangle \langle r_j| - (1 - \delta_{ij}) |r_j\rangle \langle r_i|] \quad 1.19$$

4. Méthodes Hybrides : Mécanique Quantique-Mécanique Moléculaire (MQ-MM)

4.1 Aperçu des différentes méthodes mixtes MQ-MM

Ces méthodes, que l'on nomme mixtes ou couplées, ont vu le jour récemment et allient le côté positif de la mécanique moléculaire et de la chimie quantique pour permettre l'étude des gros systèmes moléculaires. Elles sont basées sur une partition du système à étudier en deux sous-systèmes. L'un, de fort intérêt réactionnel par exemple, sera traité quantiquement, l'autre sera traité classiquement. Dans le cas des réactions enzymatiques, quelques résidus ont un rôle catalytique important et doivent être traités quantiquement alors que le reste est traité de façon classique en tant que partie exerçant une influence sur le site quantique.

Le problème qui se pose dans une telle approche est celui du traitement de la zone frontière entre les parties classique et quantique. Puisque les électrons de la région classique ne sont pas traités explicitement, il est nécessaire de concevoir une méthodologie dans laquelle la densité électronique le long de la liaison quantique-classique soit déterminée de façon satisfaisante. Les effets de l'interaction des atomes "classiques" frontière avec les atomes "quantiques" et leurs électrons doivent reproduire, le plus exactement possible, les résultats d'un calcul où le système complet est traité par la mécanique quantique. On remarquera que ce problème est évité si les régions quantique et classique constituent deux molécules différentes. Plusieurs schémas ont été proposés pour aborder ce problème. On se limitera ici à une rapide présentation de quelques-unes des méthodes mixtes classiques-quantiques :

- **QCFF/ALL** (*Quantum Consistent Force Field*)

Une des premières descriptions de la représentation d'une liaison covalente reliant les régions quantique et classique a été donnée par Warshel et Levitt¹⁷. Ils utilisent une seule orbitale hybride sur l'atome quantique dans la région de coupure. Les liaisons covalentes entre les atomes quantiques et classiques sont traitées de manière habituelle. Dans cette étude, la méthode semi-empirique QCFF/ALL est utilisée en incluant tous les électrons de valence. Elle constitue une extension d'une formulation similaire pour

les systèmes conjugués utilisant le modèle PPP de Pariser-Parr-Pople^{18,19}. Dans QCFF/ALL, les intégrales sont paramétrées par des résultats numériques et expérimentaux. Cette méthode a été employée pour étudier la réaction de la lysosyme. Il n'apparaît cependant pas clairement comment a été réalisé le partage des liaisons covalentes dans la région MQ-MM.

- **EVB/MM** (Empirical Valence Bond/Molecular Mechanics)

Dans des recherches plus récentes, Warshel et Weis²⁰ ont porté leur attention sur les méthodes mixtes EVB/MM. Dans cette représentation, l'effet majeur du changement de traitement de l'environnement réside dans la modification de l'interaction électrostatique des formes résonantes ioniques avec le reste du système. La construction de l'hamiltonien EVB/MM comporte l'évaluation de deux types d'éléments de matrice :

- (1) les termes relatifs aux états covalents qui sont considérés égaux en solution (ou dans un environnement enzymatique) et en phase gazeuse,
- (2) les éléments relatifs aux états ioniques qui sont exprimés comme une somme d'une énergie de la molécule isolée et d'une énergie "enzymatique" ou de solution.

La stabilisation des formes de résonance ioniques dans les protéines est calculée par l'évaluation de l'interaction entre les charges atomiques dans ces formes avec les charges partielles et les dipôles induits des atomes de l'enzyme. La protéine est considérée fixe.

- **VB/MM** (Valence Bond/Molecular Mechanic)

Bernardi et al²¹ ont proposé un autre schéma fondé sur un modèle de liaison de valence paramétrisée combiné avec un champ de force standard de mécanique moléculaire (VB/MM). Dans cette approche, le concept de base est la séparation du système en orbitales *{actives}* qui sont les orbitales impliquées dans la création et/ou la coupure des liaisons entre les atomes, les orbitales restantes étant *inactives* et représentant la suite du système (*framework*). Sur la base de ces orbitales, on définit les éléments de l'hamiltonien qui sont les intégrales de Coulomb, Q , et celles d'échanges, K . Ces termes sont considérés comme des paramètres du système. L'énergie de Coulomb a la forme suivante :

$$Q = V_{MM} + \sum_{ab} Q_{ab} \quad 1.20$$

où V_{MM} est un potentiel de mécanique moléculaire représentant les orbitales inactives et Q_{ab} le potentiel de Coulomb des orbitales actives. Le terme d'échange, K_{ab} , ne concerne que les orbitales actives. Le terme classique V_{MM} est un potentiel standard pour les atomes qui ne contiennent pas d'orbitales actives et un potentiel modifié pour les atomes ayant une orbitale active. Un autre aspect important de cette méthode est la modification apportée aux termes K_{ab} et Q_{ab} initialement calculés sur un système modèle, de façon à inclure l'influence de l'environnement.

- **MQ/CHARMM**

Karplus et al ²² ont proposé un schéma simple où des atomes "fantômes" ou "de terminaison" sont introduits le long de la liaison frontière pour satisfaire la valence des atomes quantiques qui sont liés aux atomes classiques. Ces atomes sont traités comme des atomes quantiques d'hydrogène et ils sont invisibles des atomes "classiques", c'est-à-dire que l'interaction entre ces atomes et les atomes classiques n'est pas calculée. Ils n'ont ni charge, ni paramètres de van der Waals. De même, ces atomes ne participent pas aux termes énergétiques des atomes classiques.

4.2 La méthode du champ auto-cohérent local (LSCF)

4.2.1 Principe de la méthode

La méthode LSCF (Local Self Consistent Field) ²³, mise au point au laboratoire, reprend le concept des orbitales hybrides développé par Warshel et Levitt ¹⁶. L'idée de base de cette méthode est que la liaison chimique qui reliera le sous-système quantique au reste de la macromolécule possède des propriétés constantes et bien définies. Elle est exprimée comme une liaison simple décrite par une orbitale moléculaire strictement localisée sur les deux atomes, considéré en première approximation comme une combinaison linéaire de deux orbitales hybrides centrées l'une sur l'atome classique et l'autre sur l'atome quantique. L'atome de la liaison qui appartient au système quantique est appelé atome frontière et la liaison de jonction des deux sous-systèmes liaison frontière. Cette méthode est fondée sur l'hypothèse que l'orbitale hybride correspondant à l'atome quantique est, d'une part, gelée et, d'autre part, possède la propriété de

transférabilité à partir d'un système modèle. Aussi doit-elle satisfaire à la condition d'orthogonalité avec les orbitales moléculaires du sous-système quantique.

Dans le cadre des méthodes semi-empiriques, il suffit, pour remplir cette condition, que l'orbitale hybride de la liaison frontière soit orthogonale aux autres orbitales atomiques de l'atome frontière. L'interaction des électrons du sous-système quantique avec la population électronique de l'atome classique de la liaison frontière est exprimée par une intégrale de Coulomb faisant intervenir la charge nette de cet atome.

4.2.2 Construction des orbitales hybrides

Considérons une liaison σ entre l'atome frontière **X** et l'atome classique de la liaison frontière **Y** (figure 1.1). Si nous désignons par $|s\rangle$, $|x\rangle$, $|y\rangle$ et $|z\rangle$, les orbitales atomiques de valence de l'atome X, l'orbitale hybride gelée peut s'écrire :

$$|l\rangle = a_{11}|s\rangle + a_{12}|x\rangle + a_{13}|y\rangle + a_{14}|z\rangle \quad 1.21$$

Les paramètres a_{11} , a_{12} , a_{13} et a_{14} doivent remplir les conditions suivantes :

- (1) l'orbitale $|l\rangle$ doit être dirigée vers l'atome Y : a_{12} , a_{13} et a_{14} sont donc proportionnels aux cosinus directeurs du vecteur représentant la liaison X-Y.
- (2) la contribution s à l'hybride normalisée est supposée être une propriété transférable de la liaison X-Y, si bien que le coefficient a_{11} est un paramètre donné du système.
- (3) les quatre coefficients sont reliés par la condition de normalisation de $|l\rangle$.

L'orbitale hybride $|l\rangle$ est définie complètement par sa direction, la contribution a_{11} de l'orbitale atomique $|s\rangle$ et sa densité électronique P_{ll} .

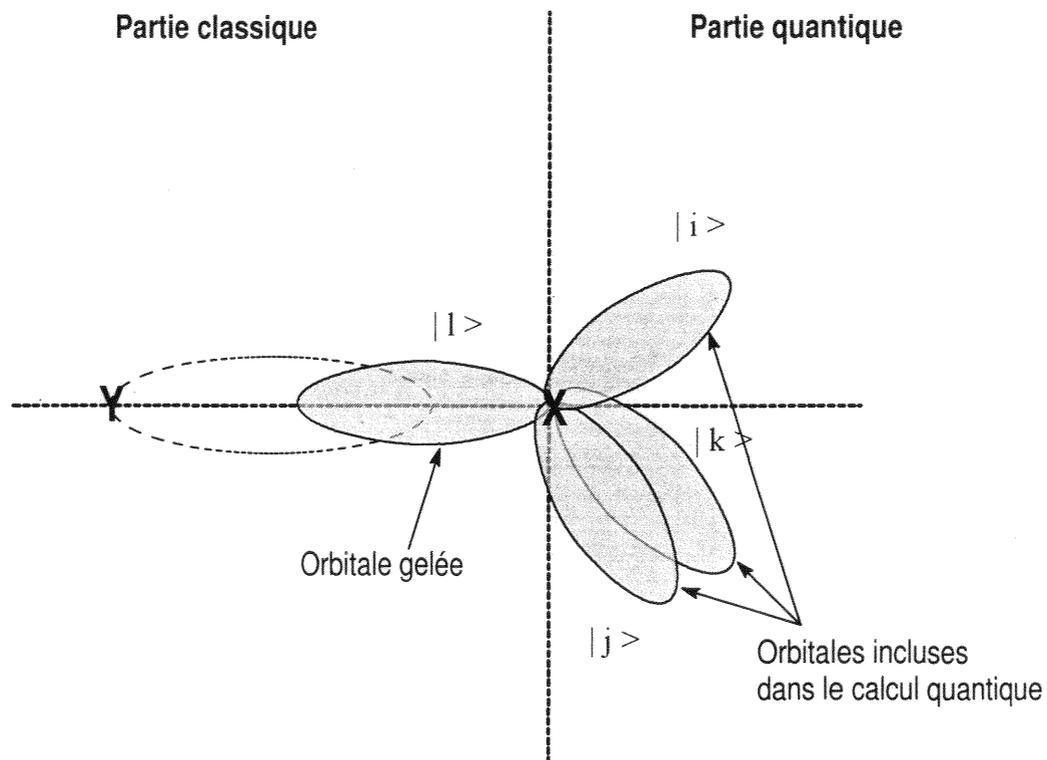


Figure 1.1 : Représentation schématique de la frontière entre les parties classique et quantique.

Les orbitales hybrides figées appartenant aux différents atomes frontières considérés seront alors enlevées du jeu d'orbitales atomiques traités quantiquement. Toutes les autres orbitales appartenant à ce sous-système subiront une combinaison linéaire de façon à former les orbitales moléculaires selon le formalisme de Hartree-Fock. La transformation orthogonale des orbitales atomiques de l'atome **X**, pour donner les quatre orbitales hybrides se réalise en deux étapes :

- 1) la construction d'une matrice d'hybridation **A** pour définir l'ensemble des orbitales hybrides orthogonales et normées $|i\rangle$, $|j\rangle$, $|k\rangle$ et $|l\rangle$,
- 2) l'application de la rotation **R** sur cet ensemble de façon à ce que l'orbitale $|l\rangle$ pointe vers la direction de la liaison X-Y. On obtient ainsi :

$$\begin{bmatrix} |i\rangle \\ |j\rangle \\ |k\rangle \\ |l\rangle \end{bmatrix} = \mathbf{B} \begin{bmatrix} |s\rangle \\ |x\rangle \\ |y\rangle \\ |z\rangle \end{bmatrix} = \mathbf{R}\mathbf{A} \begin{bmatrix} |s\rangle \\ |x\rangle \\ |y\rangle \\ |z\rangle \end{bmatrix} \quad 1.22$$

La matrice \mathbf{A} a la forme suivante :

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sqrt{1-a_n^2} & 0 & 0 & -a_n \\ a_n & 0 & 0 & \sqrt{1-a_n^2} \end{bmatrix} \quad 1.23$$

Ainsi, l'orbitale $|l\rangle$ est définie par combinaison linéaire de l'orbitale $|s\rangle$ et de l'orbitale $|z\rangle$ colinéaire à la direction $\mathbf{X-Y}$, l'orbitale $|k\rangle$ est construite avec la combinaison orthogonale des même orbitales. Les deux autres orbitales $|i\rangle$ et $|j\rangle$ sont simplement identifiées respectivement à $|x\rangle$ et $|y\rangle$

Les paramètres électroniques a_n et $P_{||}$, qui définissent respectivement le caractère s et la population électronique de l'orbitale figée, sont calculés à partir des coefficients des orbitales atomiques, bases de l'orbitale moléculaire localisée entre les deux atomes de la liaison frontière. Ces paramètres sont définis par :

$$a_n = \left(\frac{C_{X_1}^2}{\sum_{i=1}^4 C_{X_i}^2} \right)^{1/2} \quad 1.24$$

et

$$P_{||} = 2 \frac{\sum_{i=1}^4 C_{X_i}^2}{\left(\sum_{i=1}^4 C_{X_i}^2 + \sum_{i=1}^4 C_{Y_i}^2 \right)} \quad 1.25$$

α_{11} a une valeur proche de 0.5 pour une hybridation de type sp^3 et P_{11} une valeur proche de 1 pour une liaison σ non polarisée. Dans la pratique, ces valeurs sont tirées d'un calcul d'orbitales localisées sur des molécules modèles. Par exemple, dans le cas des enzymes, la localisation des orbitales est réalisée sur un tripeptide Gly-X-Gly pris comme modèle, dans lequel X est l'acide aminé qui nous intéresse.

4.2.3 Expression de l'hamiltonien dans le formalisme LSCF

Dans le modèle LSCF, toujours dans l'approximation de Born-Oppenheimer et en admettant qu'il n'y a pas de transfert de charge entre les différentes régions, l'opérateur hamiltonien du système peut être séparé en différents termes :

$$\hat{H} = \hat{H}_S + \hat{H}_{ES} + \hat{H}_E \quad 1.26$$

avec \hat{H}_S , l'hamiltonien de la partie quantique, qui s'écrit sous la forme :

$$\hat{H}_S = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_{K \in S} \sum_i \frac{Z_K}{r_{ik}} + \sum_{i < j} \frac{1}{r_{ij}} + \sum_{\substack{L < K \\ L \in S}} \frac{Z_K Z_L}{R_{KL}} \quad 1.27$$

où i, j, K et L sont les indices des électrons et des noyaux respectivement, r est la distance entre deux électrons ou un électron et un noyau, R la distance noyau-noyau, K le nombre total des atomes quantiques et Z la charge nucléaire effective.

L'hamiltonien \hat{H}_{ES} décrit l'interaction des atomes quantiques S et classiques E. Les atomes classiques, dans un champ de force normal, sont représentés par des charges ponctuelles et des paramètres de Leonard Jones. Ainsi, cet hamiltonien peut se mettre sous la forme :

$$\hat{H}_{ES} = -\sum_i \sum_{A \in E} \frac{Q_A}{r_{iA}} + \sum_{K \in S} \sum_{A \in E} \frac{Z_K Q_A}{R_{KA}} + \sum_{K \in S} \sum_{A \in E} \left[\frac{A_{KA}}{R_{KA}^{12}} - \frac{B_{KA}}{R_{KA}^6} \right] \quad 1.28$$

où les indices i et K correspondent aux électrons et noyaux quantiques et A aux atomes classiques.

Dans cette équation, les deux premiers termes sont des termes électrostatiques représentant l'interaction des atomes classiques avec les électrons et les noyaux quantiques respectivement. Le dernier est le terme de Leonard Jones.

Le dernier terme de (1.26) \hat{H}_E (calculé seulement dans la version CQFF qui fera l'objet d'une thèse prochainement dans notre laboratoire, extension de la méthode LSCF après ajout d'un champ de force classique), représente une interaction de mécanique moléculaire entre les atomes classiques E. Il est fonction de la position des atomes classiques et ne dépend pas des coordonnées électroniques et nucléaires des atomes quantiques.

4.2.4 Expression de l'opérateur de Hartree-Fock dans la formalisme LSCF

Après les transformations des intégrales qui nous permettent de les exprimer dans la base des orbitales hybrides, on peut considérer les éléments de la base de construction des orbitales moléculaires et adopter la notation générale des orbitales atomiques $|\mu\rangle$, $|\nu\rangle$, $|\lambda\rangle$ et $|\sigma\rangle$. Dans la méthode LSCF, on admet que l'interaction électrostatique de l'environnement est représentée par les charges partielles Q_A localisées sur les atomes A de l'environnement E. Ainsi l'élément de l'opérateur de Hartree-Fock local, que l'on notera $F_{\mu\nu}$, peut se mettre sous la forme générale suivante:

$$F_{\mu\nu} = F_{\mu\nu}^S + F_{\mu\nu}^{GS} + F_{\mu\nu}^{ES} \quad 1.29$$

Dans cette équation, le premier terme $F_{\mu\nu}^S$ est la forme habituelle de l'élément de l'opérateur de Hartree-Fock pour un système quantique, à cette modification près que cette expression est exprimée dans la base des orbitales. La forme de ce terme dépendra de la méthode quantique choisie pour le calcul.

Le deuxième terme a pour expression, dans sa forme générale :

$$F_{\mu\nu}^{GS} = \sum_{\ell \in X} P_{\ell\ell} \left[(\mu\nu|\ell\ell) - \frac{1}{2}(\mu\ell|\nu\ell) \right] \quad 1.30$$

où GS désigne l'interaction entre l'orbitale gelée et les électrons du sous-système quantique.

Enfin le dernier terme, $F_{\mu\nu}^{ES}$, est celui de la perturbation de la distribution électronique du sous-système quantique S par les charges atomiques de l'environnement E. Dans le formalisme NDDO, il se met sous la forme suivante :

$$F_{\mu\nu}^{ES} = \sum_{A \in E} Q_A (\mu\nu | s_A s_A) \quad 1.31$$

Ce terme est une correction apportée par rapport à la méthode de Warshell-Levitt¹⁷ qui traite cette interaction comme une interaction Coulombienne pure. Ce terme est important pour les atomes de l'environnement proches du sous-système quantique et, plus encore, pour l'atome classique de la liaison frontière. Il prend en compte les effets de pénétration des nuages électroniques.

Dans la base des orbitales hybrides, un élément de la matrice de Hartree-Fock , regroupant tous les termes développés précédemment s'écrit :

$$F'_{\mu\nu} = F_{\mu\nu} + \sum_{\lambda\eta} P_{\lambda\eta} \left[(\mu\nu | \lambda\eta) - \frac{1}{2} (\mu\eta | \lambda\nu) + \sum_I P_{II} (\mu\nu | \ell\ell) - \frac{1}{2} (\mu\ell | \lambda\ell) \right] + \sum_{A \in E} Q_A (\mu\nu | s_A s_A) \quad 1.32$$

4.2.5 Expression de l'énergie dans le formalisme LSCF

L'énergie totale du système mixte classique-quantique peut se mettre sous la forme suivante :

$$E = E_S + E_{GS} + E_{GE} + E_{SE} \quad 1.33$$

dans laquelle E_S représente l'énergie du sous-système quantique, E_{GS} et E_{GE} représentent respectivement l'énergie d'interaction de la population électronique des orbitales hybrides gelées avec le sous-système quantique et l'environnement classique et enfin E_{SE} , l'énergie de l'interaction entre le sous-système quantique et l'environnement E. Alors que le premier terme E_S de l'équation (1.33) est un terme habituel des calculs

quantiques, sans oublier toutefois d'y inclure les interactions des orbitales hybrides gelées, on écrira le second E_{GS} :

$$E_{GS} = \sum_{\ell} P_{\ell\ell} \left\{ \sum_{\mu} \sum_{\nu} P_{\mu\nu} \left[(\mu|\nu|\ell\ell) - \frac{1}{2}(\mu\ell|\nu\ell) \right] + \frac{1}{2} \sum_{\ell'} P_{\ell'\ell'} (\ell\ell|\ell'\ell') + \sum_{Z \in S} Z_K (\ell\ell|s_K s_K) \right\} \quad 1.34$$

où μ, ν représentent les orbitales atomiques et hybrides et ℓ et ℓ' les orbitales gelées. $P_{\ell\ell}$ est l'élément diagonal de la matrice densité correspondant à l'orbitale gelée et Z_K est le nombre d'électrons de valence de l'atome K.

Le premier terme de l'équation 1.34 correspond à l'interaction de la population électronique de l'orbitale gelée avec le reste de la distribution électronique, le second terme représente l'interaction électronique entre orbitales gelées et le troisième l'interaction des électrons "gelés" avec les noyaux du système quantique.

Le terme E_{GE} qui est l'énergie d'interaction de la population électronique des orbitales gelées avec l'environnement classique s'écrit :

$$E_{GE} = \sum_{\ell} P_{\ell\ell} \sum_{A \in E} Q_A (\ell\ell|s_A s_A) \quad 1.35$$

où Q_A est la charge atomique nette de l'atome A dans l'environnement classique.

Le terme E_{ES} , qui correspond à l'interaction entre les parties quantique (en excluant l'orbitale gelée) et classique du système, a la forme suivante :

$$E_{ES} = \sum_{A \in E} \left\{ Q_A \sum_{\mu} \sum_{\nu} P_{\mu\nu} (\mu\nu|s_A s_A) + \sum_{K \in S} Z_K \left[Z_A f_{KA} - P_A (s_K s_K|s_A s_A) \right] \right\} + V'(E, S) \quad 1.36$$

Le premier terme représente l'interaction électron-électron, le deuxième est l'énergie de répulsion entre noyaux quantiques et classiques où f_{KA} est une fonction semi-empirique de répulsion entre noyaux de charge $Z_A Z_K$. Le troisième terme est l'énergie d'interaction entre les noyaux quantiques et les populations électroniques de valence des atomes classiques. En accord avec le formalisme NDDO, cette interaction est proportionnelle à l'interaction Coulombienne entre les fonctions s , s_A et s_K .

L'équivalent de la population électronique de valence pour un atome classique A est défini par :

$$P_A = Z_A - Q_A \quad 1.37$$

où Z_A est le nombre d'électrons de valence et Q_A la charge atomique nette. Enfin, le dernier terme $V(E,S)$ représente toutes les interactions entre les deux sous-systèmes, à l'exception des interactions électrostatiques.

L'énergie totale regroupant tous les termes est représentée par l'équation 1.38 écrite sous le forme :

$$\begin{aligned}
 E = & \underbrace{\sum_i^{\text{occ}} \sum_{\mu} \sum_{\nu} \left\{ H_{\mu\nu} + \frac{1}{2} \sum_{\lambda} \sum_{\eta} P_{\lambda\eta} \left[(\mu\nu|\lambda\eta) - \frac{1}{2} (\mu\eta|\lambda\nu) \right] \right\} + \frac{1}{2} \sum_{\substack{K \in S \\ L \in S \\ L \neq K}} Z_K Z_L f_{KL}}_{\text{énergie SCF de S}} \\
 & + \underbrace{\sum_{\ell} P_{\ell\ell} \left\{ \sum_{\mu} \sum_{\nu} P_{\mu\nu} \left[(\mu\nu|\ell\ell) - \frac{1}{2} (\mu\ell|\nu\ell) \right] + \sum_{\ell'} \frac{1}{2} P_{\ell'\ell'} (\ell\ell|\ell'\ell') + \sum_{A \in E} Q_A (\ell\ell|S_A S_A) + \sum_{K \in S} Z_K (\ell\ell|S_K S_K) \right\}}_{\text{contribution des électrons des orbitales frontières gelées}} \\
 & + \underbrace{\sum_{AE} \left\{ Q_A \sum_{\mu} \sum_{\nu} P_{\mu\nu} (\mu\nu|S_A S_A) + \sum_{K \in S} Z_K \left[Z_A f_{KA} - P_A (S_K S_K|S_A S_A) \right] \right\}}_{\text{interaction électrostatique E-S}} + \underbrace{V(E,S)}_{\text{interaction non électrostatique E-S}}
 \end{aligned}$$

4.2.6 Dérivées de l'énergie dans le formalisme LSCF

Il s'agit d'un point important dans l'étude de la surface de potentiel engendrée par le macrosystème, tant pour la localisation des structures minimisées que pour celles des états de transition. L'étape supplémentaire dont on doit s'affranchir dans le cadre de la méthode LSCF, par rapport à une dérivation habituelle de l'énergie, est évidemment la présence des orbitales hybrides gelées centrées sur chaque atome frontière. Les propriétés de ces orbitales doivent rester inchangées lors de chaque simulation. Ainsi, dans le processus de dérivation, nous devons tenir compte non seulement des dérivées

des intégrales électroniques, mais aussi des dérivées de la matrice \mathbf{B} , définie dans le paragraphe 4.2.2, qui génère les orbitales hybrides sur chaque atome frontière.

En se reportant à l'équation (1.33), on écrira la dérivée première de l'énergie par rapport à la coordonnée q :

$$\partial_q E = \partial_q E_S + \partial_q E_{GS} + \partial_q E_{GE} + \partial_q E_{ES} \quad 1.39$$

Le développement des différents termes introduits par le formalisme LSCF est le suivant :

$$\partial_q E_{GS} = \sum_{\ell} P_{\ell\ell} \left\{ \sum_{\mu} \sum_{\nu} P_{\mu\nu} \partial_q \left[(\mu\nu|\ell\ell) - \frac{1}{2} (\mu\ell|\nu\ell) \right] + \frac{1}{2} \sum_{\ell'} P_{\ell\ell'} \partial_q (\ell\ell|\ell'\ell') + \sum_{Z \in \mathcal{S}} Z_K \partial_q (\ell\ell|s_K s_K) \right\} \quad 1.40$$

$$\partial_q E_{GE} = \sum_{\ell} P_{\ell\ell} \sum_{A \in E} Q_A \partial_q (\ell\ell|s_A s_A) \quad 1.41$$

$$\begin{aligned} \partial_q E_{ES} = & \sum_{A \in E} \left\{ Q_A \sum_{\mu} \sum_{\nu} P_{\mu\nu} \partial_q (\mu\nu|s_A s_A) + \sum_{K \in \mathcal{S}} Z_K \left[Z_A \partial_q f_{KA} - P_A \partial_q (s_K s_K|s_A s_A) \right] \right\} \\ & + \partial_q V'(E, S) \end{aligned} \quad 1.42$$

Dans toutes les équations de ce paragraphe, les dérivées premières de la matrice densité $[P]$ n'interviennent pas²⁵ et les intégrales de recouvrement, ainsi que leurs dérivées, sont négligées au niveau NDDO. En regroupant tous les termes, la dérivée première de l'énergie s'écrit, (équation 1.43) :

$$\begin{aligned}
& \partial_q E = \sum_i \sum_{\mu} \sum_{\nu} \left\{ \partial_q H_{\mu\nu} + \frac{1}{2} \sum_{\lambda} \sum_{\eta} P_{\lambda\eta} \left[\partial_q (\mu|\lambda\eta) - \frac{1}{2} \partial_q (\mu\eta|\lambda\nu) \right] \right\} + \frac{1}{2} \sum_{\substack{K \in \mathcal{L} \in \mathcal{S} \\ L \neq K}} Z_K Z_L \partial_q f_{KL} \\
& + \sum_{\ell} P_{\ell\ell} \left\{ \sum_{\mu} \sum_{\nu} P_{\mu\nu} \left[\partial_q (\mu\nu|\ell\ell) - \frac{1}{2} \partial_q (\mu\ell|\nu\ell) \right] + \sum_{\ell'} \frac{1}{2} P_{\ell'\ell'} \partial_q (\ell\ell|\ell'\ell') + \sum_{A \in \mathcal{E}} Q_A \partial_q (\ell\ell|S_A S_A) + \sum_{K \in \mathcal{S}} Z_K \partial_q (\ell\ell|S_K S_K) \right\} \\
& + \sum_{A \in \mathcal{E}} \left\{ Q_A \sum_{\mu} \sum_{\nu} P_{\mu\nu} \partial_q (\mu\nu|S_A S_A) + \sum_{K \in \mathcal{S}} \left[Z_A \partial_q f_{KA} - \partial_q P_A(S_K S_K|S_A S_A) \right] \right\} + \partial_q V(E, \mathcal{S})
\end{aligned}$$

5. Chemin réactionnel

Dans l'approximation de Born-Oppenheimer l'énergie d'une molécule est fonction, entre autres choses, des coordonnées des atomes constituant cette molécule. Nous avons vu, au paragraphe précédent, comment calculer cette énergie au moyen des méthodes quantiques élaborées. Nous pouvons définir une hypersurface d'énergie potentielle de dimension $3N-6$ où N est le nombre d'atomes du système étudié. Cette hypersurface est caractérisée par des points critiques (points où toutes les dérivées premières de l'énergie par rapport aux coordonnées atomiques sont nulles) qui se répartissent en quatre catégories suivant les valeurs des dérivées secondes de l'énergie.

- 1) Si toutes les dérivées secondes possèdent des valeurs positives, alors le point critique correspondant sera un *minimum*.
- 2) Si toutes, sauf une, prennent des valeurs positives et la dernière une valeur négative, alors le point critique sera un *état de transition* (ou point selle). Le vecteur propre correspondant à la seule valeur propre négative du hessien (matrice dérivées secondes) au point de la structure de transition est appelé *vecteur de transition*.
- 3) Si N dérivées seconde possèdent des valeurs négatives, alors le point critique sera un point *selle d'ordre N* (ce point critique n'a pas de signification physique).

4) Si toutes les dérivées secondes possèdent des valeurs négatives alors le point critique s'appellera un *maximum*

Les minima de l'hypersurface de potentiel correspondent aux espèces chimiques stables, c'est à dire les réactifs, les produits et les intermédiaires réactionnels. Il est toujours possible de relier deux minima en suivant un chemin qui passe par un ou plusieurs points selle. Le chemin qui lie les réactifs aux produits est appelé *chemin de réaction*.

Les programmes GEOMOS²⁵ et plus récemment GEOMOP²⁶ (qui inclue la méthode LSCF) ont fait la preuve de leur efficacité dans la recherche des minima et des états de transition. Ces programmes utilisent l'algorithme BERNY¹⁷ qui localise bien les minima et les structures de transition grâce à une méthode modifiée du gradient conjugué. En effet, les gradients sont calculés analytiquement par les routines de GEOMOP chaque fois que l'énergie est calculée. La matrice des dérivées secondes est mise à jour selon un procédé propre à BERNY et ceci en employant les gradients : à chaque pas de l'itération, une minimisation unidimensionnelle est réalisée (sauf dans le cas de la recherche d'une structure de transition) en utilisant un polynôme quadratique , suivi d'une recherche n-dimensionnelle en utilisant la matrice des dérivées secondes²⁷.

Le chemin réactionnel dépend du système de coordonnées choisi et il est évident qu'un changement du système de coordonnées ne doit pas altérer la nature des points stationnaires (points où le vecteur gradient est nul). Ainsi on définit un chemin réactionnel indépendant du système de coordonnées, d'où l'idée du chemin *réactionnel intrinsèque*. Un tel chemin peut être trouvé en suivant le chemin de moindre énergie de l'état de transition vers les produits et les réactifs. Cet algorithme du type IRC²⁸, fondé sur les équations classiques du mouvement quasi-statique, et implanté dans le programme GEOMOP, propose un système de coordonnées internes pondérées en masse.

En pratique l'algorithme BERNY, dans le cas d'une recherche d'une structure de transition, force le hessien à avoir une seule valeur propre négative. Si plusieurs valeurs propres négatives sont rencontrées, on choisit l'une d'elles (celle de plus forte valeur absolue) et les autres sont changées de signe. Ainsi, le système est assuré d'évoluer dans une et une seule direction ascendante. La géométrie de départ, pour un calcul type IRC, doit être choisie dans la région quadratique de la structure de transition, c'est-à-dire la

région où le hessien ne possède qu'une seule valeur propre négative. Cette géométrie peut être obtenue par des recherches préliminaires, telles que la localisation d'un point maximum sur un chemin de minimisations successives allant d'un minimum à l'autre, en fixant la coordonnée supposée dominer le vecteur de transition. Si deux vecteurs dominent ce vecteur de transition alors l'étude d'une grille (minimisation pas à pas suivant les deux coordonnées qui dominent ce vecteur propre) peut s'avérer assez efficace, bien que coûteuse en temps de calcul.

6. Le modèle du continuum

6.1 Introduction

En général, la plupart des processus chimiques ont lieu en solution et la nature du solvant joue un rôle très important, notamment dans le cas d'espèces chargées ou présentant une séparation de charge importante. Il est donc important que son effet soit pris en compte pour analyser des résultats provenant de calculs théoriques. De très nombreuses études ont été publiées concernant les techniques de calcul de l'énergie d'interaction soluté-solvant. Parmi ces techniques, nous retrouvons les méthodes de simulation MC^{29,30} (Monte-Carlo), DM³¹⁻³³ (Dynamique Moléculaire) et les méthodes de description quantique du soluté.

Les méthodes de chimie quantique en phase condensée se distinguent des modèles de simulations statistiques (MC et DM) par l'aspect uniforme du solvant. En effet, ces méthodes ne permettent pas de traiter explicitement le transfert de charges entre le solvant et le soluté. Nous pouvons, néanmoins, nous affranchir de cette difficulté en incluant dans le système soluté quelques molécules du solvant, qui seront donc traitées quantiquement. Contrairement aux méthodes de simulation, elles nous permettent de calculer aisément les modifications de la structure de l'état de transition et par conséquent du chemin réactionnel.

Ces méthodes à solvant modèle "continuum", sont basées sur une représentation macroscopique de l'environnement liquide, dans laquelle le solvant est assimilé à un milieu continu et infini caractérisé par ses propriétés diélectriques globales. La molécule

est soumise à un potentiel électrostatique, appelé potentiel du champ de réaction, dû à la polarisation du solvant par la distribution de charges du soluté lui-même. Parmi les multiples approches utilisant ce modèle, bon nombre lui associent le formalisme de la cavité. Celui-ci consiste à placer le soluté à l'intérieur d'une cavité baignant dans le liquide.

Nous allons maintenant détailler deux approches qui utilisent ce modèle et qui considèrent le diélectrique polarisé comme une perturbation de l'hamiltonien du soluté. La première, développée par Tomasi et ses collaborateurs, utilise une distribution de charges fictives sur la surface de la cavité afin de représenter la polarisation du solvant³⁴, tandis que la deuxième, mise au point au laboratoire de Chimie théorique de Nancy s'appuie sur un développement multipolaire de la distribution du soluté³⁵.

6.2 Modèle de Pise (Tomasi et al.)

La polarisation du diélectrique infini est équivalente à la création d'un système de charges virtuelles sur la surface de la cavité. La densité de charges surfacique $\sigma(\vec{s})$ relative à cet ensemble de charges, produit un potentiel électrostatique V_σ qui doit être ajouté à l'hamiltonien du soluté de façon à prendre en compte la polarisation due au solvant :

$$H = H_0 + V_\sigma \quad 1.44$$

La densité de charges surfacique en un point \vec{s} de l'enveloppe de la cavité, s'exprime à partir du potentiel total agissant en ce point, sous la forme :

$$\sigma(\vec{s}) = \frac{\epsilon - 1}{4\pi} \left(\frac{\partial V(\vec{s})}{\partial n} \right)_{s_-} \quad 1.45$$

où ϵ représente la constante diélectrique du solvant et S_- signifie que la dérivée doit être effectuée sur la surface interne de la cavité. Le potentiel total $V(s)$, qui intervient également dans cette expression, est défini à partir du potentiel dû à la distribution de charges du soluté et de V_σ :

$$V(s) = V_p(s) + V_\sigma(s) \quad 1.46$$

La détermination de $\sigma(s)$ implique donc la mise en œuvre d'une procédure itérative décrite en détail dans les références ^{34,36}. A l'origine, la méthode que nous venons de détailler utilisait une cavité définie par la surface de van der Waals de la molécule. Elle a ensuite été adaptée pour prendre en compte une cavité s'appuyant sur une surface moléculaire³⁷, rendant ainsi mieux compte du volume d'exclusion du liquide. Le principal inconvénient de cette méthode est le calcul numérique, à chaque cycle SCF ou d'optimisation, des charges superficielles ce qui limite cette méthode du point de vue temps de calcul.

6.3 Modèle à cavité de Nancy

Ce modèle consiste à utiliser un développement multipolaire pour représenter la distribution de charges du soluté ³⁸. Kirkwood³⁹ a été le premier, en 1934, à utiliser ce type de développement pour calculer l'énergie d'interaction électrostatique d'un ensemble de charges placées dans une cavité sphérique entourée par un milieu diélectrique continu. Le premier calcul de chimie quantique utilisant le modèle à cavité, basé sur le principe de Kirkwood et Onsager⁴⁰, a été effectué par Rivail et Rinaldi³⁸ en 1973. Le champ électrostatique de réaction ainsi que les différentes composantes de ses dérivées s'expriment alors en fonction des composantes M_ℓ^m des moments et des facteurs du champ de réaction $f_{\ell\ell'}^{mm'}$, selon :

$$R_\ell^m = \sum_{\ell'=0}^{\infty} \sum_{m=-\ell'}^{\ell'} f_{\ell\ell'}^{mm'} M_\ell^m \quad 1.47$$

Ces derniers ne dépendent que des propriétés géométriques de la cavité créée au sein du continuum par le soluté et de la constante diélectrique du solvant. Dans le cas d'une cavité de type sphérique ces facteurs prennent la forme suivante :

$$f_{\ell\ell'}^{mm'} = \frac{(\varepsilon + 1)(\varepsilon - 1)}{\varepsilon(\varepsilon + 1) + 1} \frac{1}{a^{2\ell+1}} \delta_{\ell\ell'} \delta_{mm'} \quad 1.48$$

Les $2\ell + 1$ composantes du multipôle de rang ℓ s'écrivent, en utilisant le formalisme des tenseurs sphériques proposé par Gray ⁴², sous la forme :

$$M_\ell^m = \sum_i q_i S_\ell^m(\vec{r}_i) \quad 1.49$$

Dans cette expression apparaissent les polynômes sphériques S_ℓ^m , qui sont définis à partir des harmoniques sphériques :

$$S_\ell^m(\vec{r}) = \sqrt{\frac{4\pi}{2\ell + 1}} r^\ell Y_\ell^m(\theta, \chi) \quad 1.50$$

L'opérateur d'interaction entre la molécule du soluté et le champ de réaction peut s'écrire :

$$H' = - \sum_{l=0}^{\infty} \sum_{m=-l}^l R_\ell^m M_\ell^m \quad 1.51$$

Le champ de réaction représente la réponse de l'environnement à la distribution de charges de la molécule. Les éléments $f_{\ell\ell'}^{mm'}$ du tenseur des facteurs du champ de réaction ne dépendent que de la géométrie de la cavité et de la constante diélectrique ϵ du solvant. Dans cette approche, l'expression de ces facteurs a été développée dans le cas d'une cavité ellipsoïdale par la méthode de Rinaldi ⁴³.

En incluant le terme d'énergie électrostatique du solvant dans le Fockien, les éléments de la matrice associée s'écrivent :

$$F'_{\mu\nu} = F_{\mu\nu} + \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} R_\ell^m \langle \mu | M_\ell^m | \nu \rangle \quad 1.52$$

où $F_{\mu\nu}$ est l'élément de la matrice associée au Fockien non perturbé et μ et ν sont deux fonctions de base.

L'énergie électrostatique est :

$$E_{\text{el}} = - \frac{1}{2} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} R_\ell^m M_\ell^m \quad 1.53$$

Le facteur 1/2 provient du fait que l'énergie nécessaire à la polarisation du milieu est exactement la moitié de l'énergie électrostatique.

La détermination du volume de la cavité et de sa forme joue un rôle très important dans ce modèle. Il a été proposé, en accord avec le critère d'Onsager, que la forme du système moléculaire soit déterminée de façon à ce que son volume soit égal au volume moléculaire partiel. Il est apparu que cette surface, dont la forme analytique n'est pas connue, présenterait beaucoup de similitudes avec la surface de van der Waals pour laquelle les rayons associés aux atomes sont surévalués⁴⁴. Pour palier cet inconvénient, une bonne approximation est de considérer une enveloppe ellipsoïdale, dans laquelle les dérivées secondes pourront être calculées analytiquement. Le volume moléculaire est obtenu à partir du volume de van der Waals, corrigé en multipliant les rayons par un facteur de 1.105, par la formule :

$$V_{\text{mol}} = 1.480 V_{\text{vdW}}^{\text{Cor}} + 0.42 \left(\text{en \AA}^3 \right) \quad 1.54$$

Le moyen le plus simple pour définir les axes de l'ellipsoïde est de considérer les axes principaux et les valeurs principales d'une propriété tensorielle de la molécule. On définit préférentiellement l'ellipsoïde à partir des valeurs et vecteurs propres du tenseur d'inertie correspondant au volume de densité uniforme défini par la surface de van der Waals.

BIBLIOGRAPHIE

1. M. Born, J.R. Oppenheimer Ann. Physik, **84**, 457 (1927)
2. J.C. Slater, Phys. Rev, 1293 (1929)
3. D.R. Hartree, Proc. Cambridge Phil. Soc, **24**, 89 (1928)
4. C.C.J. Roothaan, Rev. Mod. Phys, **23**, 69 (1951)
5. J.A. Pople and G.A. Segal, J. Chem. Phys, **44**, 3298 (1966)
6. J.A. Pople, D.L. Beveridge and P.A. Dobosh, J. Chem. Phys, **47**, 2026 (1967)
7. J.A. Pople, D.L. Beveridge and P.A. Dobosh, J. Chem. Phys, **43**, 5129 (1965)
8. M.J.S. Dewar et W. Thiel. J. Am. Chem. Soc, **99**, 4899 (1977)
9. M.J.S. Dewar, E.G. Zoebisch, E.F. Healy and J.J.P. Stewart, J. Am. Chem. Soc, **107**, 3902 (1985)
10. J.J.P. Stewart, J. Comp. Chem, 209 (1989)
11. R. Courant, Bull. Amer. Math. Soc, **49**, 1 (1943)
12. H.D. Curry. Qu. App. Maths, **2**, 258 (1944)
13. M.R. Hestenes et E. Stiefel, J. Res. N. B. S, **49**, 4.9 (1952)
14. A.S. Householder, Principles of numerical Analysis, New york: McGraw Hill, (1953)
15. R. Fletcher et M.J.D. Powell, The Computer Journal, **6**, 163 (1963)
16. A. Warshel, M. Levitt, J. Mol. Biol, **103**, 227 (1976)
17. H.B. Schlegel, J. Comp. Chem, **3**, n° 2, 214 (1982)
18. R. Pariser, R.G. Parr J. Chem. Phys, **21**, 466 (1953)
19. J.A. Pople, Trans. Faraday Soc, **49**, 1375 (1953)
20. A. Warshel, R.M. Weis, J. Am. Chem. Soc, **102**, 6218 (1980)
21. F. Bernardi, M. Olivucci, M.A. Robb, J. Comp. Chem., **114**, 1606 (1992)

22. M.J. Field, P.A. Bash, M. Karplus, *J. Comp. Chem*, **11**, 700 (1990)
23. V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, G. Ferenzy, *J. Comp. Chem.*, **15**, 269 (1994)
24. S. Bratoz, *Coll. Int. CNRS (Paris)*, **82**, 287 (1958)
25. D. Rinaldi, P.E. Hoggan, A. Cartier, GEOMOS, QCPE 584
26. V. Théry, **Thèse**, Université Nancy1 (1993)
27. K. Baka, **Thèse**, Université Nancy1 (1997)
28. K. Fukui, *Acc. Chem. Res*, **14**, 363 (1981)
29. W. L. Jorgensen, *J. Phys. Chem*, **87**, 5304 (1983)
30. D. L. Beveridge, M. Mezei, G. Ravishanker et B. Jayaram, *Int. J. Quantum Chem*, **29**, 1513 (1986)
31. J. P. Bergsma, B. J. Gertner, K. R. Wilson et J. T. Heynes, *J. Chem. Phys*, **86**, 1356 (1987)
32. B. J. Gertner, J. P. Bergsma, K. R. Wilson, S. Lee et J. T. Heynes, *J. Chem. Phys*, **86**, 1377 (1987)
33. B. J. Gertner, K. R. Wilson, et J. T. Heynes, *J. Chem. Phys*, **90**, 3537 (1989)
34. S. Miertus, E. Scrocco et J. Tomasi, *Chem. Phys*, **55**, 117 (1981)
35. J. L. Rivail et D. Rinaldi, *Chem. Phys*, **18**, 233, (1976)
36. R. Bonaccorsi, R. Cimiraglia et J. Tomasi, *J. Comput. Chem*, **4**, 567 (1983).
37. J. L. Pascual-Arthuir, E. Silla, J. Tomasi et R. Bonaccorsi, *J. Comput. Chem*, **8**, 778 (1987)
38. D. Rinaldi et J. L. Rivail, *Theoret. Chim. Acta*, **32**, 57 (1973)
39. J. G. Kirkwood, *J. Chem. Phys*, **2**, 351 (1920)
40. L. Onsager, *J.A.C.S.*, **58**, 1486 (1936)
41. D. Rinaldi *Computers and Chemistry*, **6**, 155 (1982)
42. C. G. Gray, *Can. J. Phys*, **54**, 505 (1936)

43. D. Rinaldi, *Comput.Chem*, **6**, 155 (1982)

44. J. L. Rivail et B. Terryn *J. Chem. Phys*, **79**, (1982)

Chapitre 2

Effet des interactions à longue distance

1. Champ électrostatique moyen

1.1 Introduction

Comme nous l'avons décrit dans le chapitre précédent, la méthode LSCF a été développée afin de traiter un sous-système (site actif d'une protéine) à l'aide des méthodes de la chimie quantique (semi-empirique), alors que le reste de la macromolécule est traité de façon classique. Dans cette méthode, le potentiel électrostatique de l'environnement protéique est représenté par un jeu de charges ponctuelles. Plusieurs études^{1,2} ont montré l'importance des interactions électrostatiques

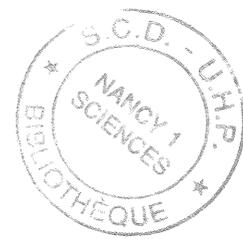
entre le fragment quantique et le reste de la macromolécule, dans l'énergie d'un processus chimique. La définition de la prise en compte du solvant (environnement protéique) est perfectible et une représentation du potentiel électrostatique moléculaire par un jeu de charges fait encore l'objet de recherches³. Le calcul de la perturbation due à ces charges sur la fonction d'onde de la partie quantique prend un temps de calcul énorme, en particulier le calcul des intégrales $\langle ij|S_A S_A \rangle$ (équation 1.38 chapitre 1) entre la partie quantique et l'environnement. Afin de réduire ce temps de calcul tout en représentant correctement le potentiel électrostatique de l'environnement (choix imposé par la taille des systèmes auxquels on s'intéresse), nous avons développé une méthode qui nous permet de représenter la perturbation de la partie lointaine de l'environnement par un potentiel, en une expansion multipolaire de $1/r$ au centre de la partie quantique.

En effet, au lieu de traiter individuellement les interactions à longue distance entre les charges de la partie classique et les électrons de la partie quantique, nous avons mis au point un algorithme permettant de ne considérer, pour l'ensemble des charges classiques situées plus loin qu'une certaine distance (6 Å), que le champ global créé au centre de la partie quantique. De cette façon le champ créé ne sera calculé qu'une seule fois au cours de l'optimisation de la géométrie, au lieu de l'être à chaque pas comme c'était le cas dans l'ancienne version du programme. Cette amélioration du programme a permis de gagner un temps de calcul considérable (accélération d'un facteur moyen de 3), au prix d'une perte de précision infime.

1.2 Perturbation électrostatique

1.2.1 Répartition du système

Dans l'expression de l'énergie E (équation 1.38), on reconnaît l'expression de l'interaction entre la partie quantique S et son environnement E . Cette correction apportée par rapport à celle de Warshel-Levitt⁴, qui traitent cette interaction de manière purement électrostatique est surtout importante pour les atomes de l'environnement proches du système quantique et, en particulier, pour les atomes exclus du calcul liés aux atomes frontières.



L'amélioration que nous proposons de développer dans ce chapitre répartit la macromolécule en trois régions : la partie quantique (région 1), une partie intermédiaire dans laquelle le calcul est effectué, comme auparavant, en tenant compte des charges ponctuelles (région 2), et la partie lointaine de l'environnement où le développement de $1/r$ est utilisé (région 3) (figure 2.1).

Le potentiel électrostatique entre les charges de la région 3 et la partie quantique est calculé, en général, par un développement en série autour d'un point : le centre de la partie quantique. Ceci conduit à une expression faisant intervenir les moments multipolaires du système quantique dont le calcul est très aisé car toutes les quantités sont calculées analytiquement. Dans cette approche, la distribution de charges de la partie lointaine du système classique (distance supérieure à 6 Å) est représentée par un développement multipolaire de $1/r$. Nous utilisons le formalisme des tenseurs sphériques⁵ pour lequel le multipôle d'ordre ℓ a « $2\ell+1$ » composantes indépendantes notées M_ℓ^m .

1.2.2 Expression du champ électrostatique moyen

La perturbation électrostatique engendrée par la partie lointaine de l'environnement sur le sous-système quantique, représente une énergie potentielle de la forme :

$$H^{cut} = \sum_i Q_i V_i \quad 2.1$$

où

$$V_i = \sum_k \frac{Q_k}{r_{ik}} \quad 2.2$$

V_i est le potentiel dû aux charges Q_k de l'environnement à une distance r_{ik} de ce point, en n'importe quel point i du sous-système quantique, et Q_i la charge de la particule (électron ou noyau) située au point i (figure 2.1).

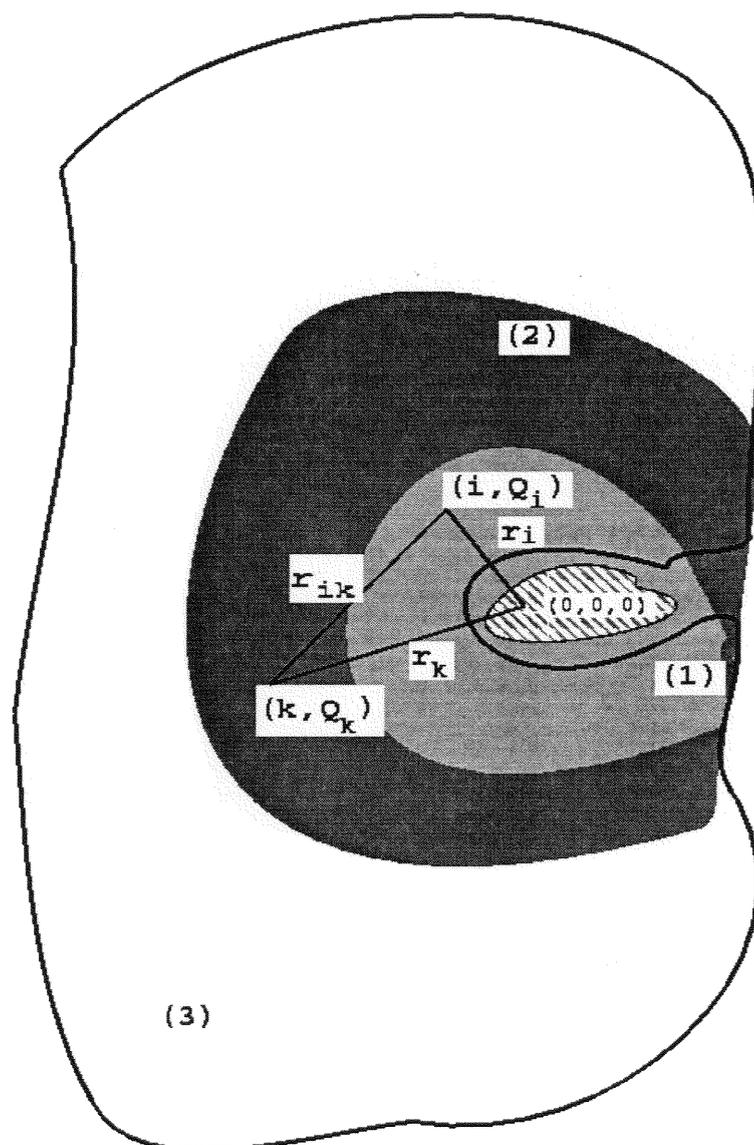


Figure 2.1 : Les trois régions du système protéique dans la méthode QM/MM. La région 1 représente la partie quantique. La région 2 contient les atomes de l'environnement classique situés à une distance $r_{ik} < 8\text{\AA}$. La région 3 contient le reste du système.

Le rapport $1/r_{ik}$ peut être développé en harmoniques sphériques suivant l'expression :

$$\frac{1}{r_{ik}} = \sum_{\ell}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} \frac{r_{<}^{\ell}}{r_{>}^{\ell+1}} Y_{\ell}^m(\theta_i, \varphi_i) Y_{\ell}^{m*}(\theta_k, \varphi_k) \quad 2.3$$

avec $r_{<} = \min(r_i, r_k)$ $r_{>} = \max(r_i, r_k)$

Les fonctions $Y_{\ell}^m(\theta_i, \varphi_i)$ et $Y_{\ell}^m(\theta_k, \varphi_k)$ représentent les harmoniques sphériques exprimées à l'aide des coordonnées angulaire des points i et k. Dans le cas qui nous intéresse, r_k est toujours supérieur a r_i (figure 2.1) puisque l'origine est choisie au barycentre des charges i (sous-système quantique). L'équation 2.2 devient :

$$\frac{1}{r_{ik}} = \sum_{\ell}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} r_i^{\ell} Y_{\ell}^m(\theta_i, \varphi_i) \frac{1}{r_k^{\ell+1}} Y_{\ell}^{m*}(\theta_k, \varphi_k) \quad 2.4$$

La perturbation électrostatique s'écrit :

$$H^{\text{cut}} = \sum_{\ell}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} Q_i r_i^{\ell} Y_{\ell}^m(\theta_i, \varphi_i) \frac{Q_k}{r_k^{\ell+1}} Y_{\ell}^{m*}(\theta_k, \varphi_k) \quad 2.5$$

Les moment électriques multipolaires, selon Gray⁵, dépendent de la charge de la particule et des tenseurs sphériques S_{ℓ}^m , eux-mêmes fonction des harmoniques sphériques Y_{ℓ}^m .

Nous obtenons ainsi une expression de la perturbation électrostatique sous la forme :

$$H^{\text{cut}} = \sum_1^{\ell_{\text{max}}} \sum_{m=-\ell}^{\ell} C_{\ell}^m M_{\ell}^m \quad 2.6$$

où M_{ℓ}^m désigne l'opérateur moment multipolaire défini par :

$$M_i^m = \sum_i Q_i r_i^{\ell} Y_{\ell}^m(\theta_i, \varphi_i) \quad 2.7$$

et C_{ℓ}^m , la composante du champ moyen généré au centre de la partie quantique, définie par :

$$C_{\ell}^m = \sum_k \frac{Q_k}{r_k^{\ell+1}} Y_{\ell}^{m*}(\theta_k, \varphi_k) \quad 2.8$$

1.3 Matrice de Hartree-Fock et expression de l'énergie

Dans la base des orbitales hybrides, un élément de la matrice de Hartree-Fock en fonction du champ moyen crée par la partie lointaine du système s'écrit :

$$F'_{\mu\nu} = F_{\mu\nu} + \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} C_{\ell}^m M_{\ell}^m \quad 2.9$$

où $F_{\mu\nu}$ est l'élément de la matrice associée au Fockien, perturbée par la première couche de charges ponctuelles située à une distance inférieure à un rayon de troncature (appelé par abus de langage "cutoff") choisi, et μ et ν sont deux fonctions de base. Par conséquent, l'effet électrostatique de la partie lointaine de l'environnement protéique est pris en compte de façon auto-cohérente, ce qui permet d'inclure les effets de polarisation au cours du calcul SCF. L'énergie totale du système peut alors s'écrire :

$$\begin{aligned}
E = & \underbrace{\sum_i \sum_{\mu} \sum_{\nu} \left\{ H_{\mu\nu} + \frac{1}{2} \sum_{\lambda} \sum_{\eta} P_{\lambda\eta} \left[(\mu\nu|\lambda\eta) - \frac{1}{2} (\mu\eta|\lambda\nu) \right] \right\} + \frac{1}{2} \sum_{\substack{K \in S, L \in S \\ L \neq K}} Z_K Z_L f_{KL}}_{\text{énergie SCF de S}} \\
& + \underbrace{\sum_{\ell} P_{\ell\ell} \left\{ \sum_{\mu} \sum_{\nu} P_{\mu\nu} \left[(\mu\nu|\ell\ell) - \frac{1}{2} (\mu\ell|\nu\ell) \right] + \sum_{\ell'} \frac{1}{2} P_{\ell'\ell'} (\ell\ell|\ell'\ell') + \sum_{A \in E} Q_A (\ell\ell|_{SASA}) + \sum_{K \in S} Z_K (\ell\ell|_{SKSK}) \right\}}_{\text{contribution des électrons des orbitales frontières gelées}} \\
& + \underbrace{\sum_{A \in E} \left\{ Q_A \sum_{\mu} \sum_{\nu} P_{\mu\nu} (\mu\nu|_{SASA}) + \sum_{K \in S} Z_K \left[Z_A f_{KA} - P_A (SKSK|_{SASA}) \right] \right\}}_{\text{interaction électrostatique E-S avec } r_{E-S} \leq \text{cutoff}} + \quad 2.10 \\
& \underbrace{V_{E-S}}_{\text{interaction non électrostatique E-S avec } r_{E-S} \leq \text{cutoff}} + \underbrace{\sum_{\ell} \sum_m C_{\ell}^m M_{\ell}^m}_{\text{potentiel électrostatique de la partie lointaine avec } r_{E-S} > \text{cutoff}}
\end{aligned}$$

Dans l'expression de E , le dernier terme exprime la perturbation due au potentiel moyen électrostatique créée par les charges partielles de la partie lointaine du système. Les coefficients C_{ℓ}^m ne sont calculés qu'une seule fois au cours de l'optimisation de géométrie, puisque l'environnement (régions 2 et 3) reste fixe.

1.4 Expression des dérivées de l'énergie

La méthode LSCF a été implantée dans le programme de calculs quantiques semi-empiriques GEOMOP⁶ (i.e. version modifiée de GEOMOS⁷). Cette implantation comprend le calcul LSCF ainsi que celui des dérivées de l'énergie par rapport aux coordonnées internes des atomes autres que les atomes frontière du système quantique, l'environnement restant fixe. La connaissance des dérivées secondes de l'énergie par rapport aux coordonnées est indispensable, aussi bien pour localiser les structures de transition que pour calculer les grandeurs thermodynamiques. Nous avons développé un logiciel de calcul analytique des dérivées premières et secondes de la partie de l'énergie

correspondant au champ électrostatique moyen que nous avons incorporé dans la version ancienne du programme⁷.

Dans le cas du champ moyen créé par la partie lointaine de la macromolécule, pour des raisons de clarté, nous définissons les moments électriques multipolaires en fonction de leurs composantes électroniques et nucléaires :

$$M_l^m = \sum_A Z_A M_l^m + \sum_i \sum_\mu \sum_\nu P_{\mu\nu} M_l^m \quad 2.11$$

Afin de simplifier les expressions qui vont suivre, nous définissons les opérateurs dérivées premières des moment M_l^m par rapport à une coordonnée géométrique quelconque q_k :

$$\partial_k M_l^m = \sum_A Z_A \partial_k M_l^m + \sum_i \sum_\mu \sum_\nu P_{\mu\nu} \partial_k M_l^m \quad 2.12$$

Les dérivées premières de l'énergie ne dépendent pas des dérivées de la matrice densité. En outre, dans la méthode NDDO^{8,9,10}, les intégrales de recouvrement et leurs dérivées sont négligées. Les dérivées de l'énergie dans la bases des orbitales hybrides s'écrivent donc :

$$\begin{aligned}
E = & \sum_i \sum_{\mu} \sum_{\nu} P_{\mu\nu} \left\{ \partial_q H_{\mu\nu} + \frac{1}{2} \sum_{\lambda} \sum_{\eta} P_{\lambda\eta} \left[\partial_q (\mu|\lambda\eta) - \frac{1}{2} \partial_q (\mu\eta|\lambda\nu) \right] \right\} + \frac{1}{2} \sum_{\substack{K \in \mathcal{S} \\ L \neq K}} Z_K Z_L \partial_q f_{KL} \\
& + \sum_{\ell} P_{\ell\ell} \left\{ \sum_{\mu} \sum_{\nu} P_{\mu\nu} \left[\partial_q (\mu|\ell\ell) - \frac{1}{2} \partial_q (\mu\ell|\nu\ell) \right] + \sum_{\ell} \frac{1}{2} P_{\ell\ell} \partial_q (\ell\ell|\ell\ell) + \sum_{A \in \mathcal{E}} Q_A \partial_q (\ell\ell|S_A S_A) + \sum_{K \in \mathcal{S}} Z_K \partial_q (\ell\ell|S_K S_K) \right\} \\
& + \sum_{A \in \mathcal{E}} \left\{ Q_A \sum_{\mu} \sum_{\nu} P_{\mu\nu} \partial_q (\mu\nu|S_A S_A) + \sum_{K \in \mathcal{S}} Z_K \left[Z_A \partial_q f_{KA} - \partial_q P_A(S_K S_K|S_A S_A) \right] \right\} + \partial_q V_{E-S} \\
& + C_{\ell}^m \partial_q M_{\ell}^m \qquad \qquad \qquad 213
\end{aligned}$$

La contribution électrostatique dépend des dérivées des intégrales $\langle \mu | S_{\ell}^m | \nu \rangle$ pour le terme électronique et des dérivées des fonctions $S_{\ell}^m(\vec{R}_A)$ pour le terme nucléaire. Les coefficients C_{ℓ}^m sont des constantes puisque l'environnement reste fixe dans l'espace.

1.5 Validation

Afin d'évaluer le gain en temps de calcul offert par cette amélioration méthodologique, nous avons effectué une optimisation de géométrie sur un système de taille assez importante, avec et sans développement multipolaire. Pour cela, nous avons étudié le complexe thrombine/dipeptide (Arg41-Ser42) à l'aide des deux méthodes semi-empiriques AM1 et PM3.

1.5.1 Choix du cutoff

Afin de représenter correctement le champ électrostatique moyen de la partie lointaine du système, nous devons estimer la distance à partir de laquelle les intégrales

$\langle S_K S_K | S_A S_A \rangle$, termes d'interaction entre électrons de la région 1 et charges de l'environnement, coïncident avec un développement en $1/r_{KA}$ (figure 2.2).

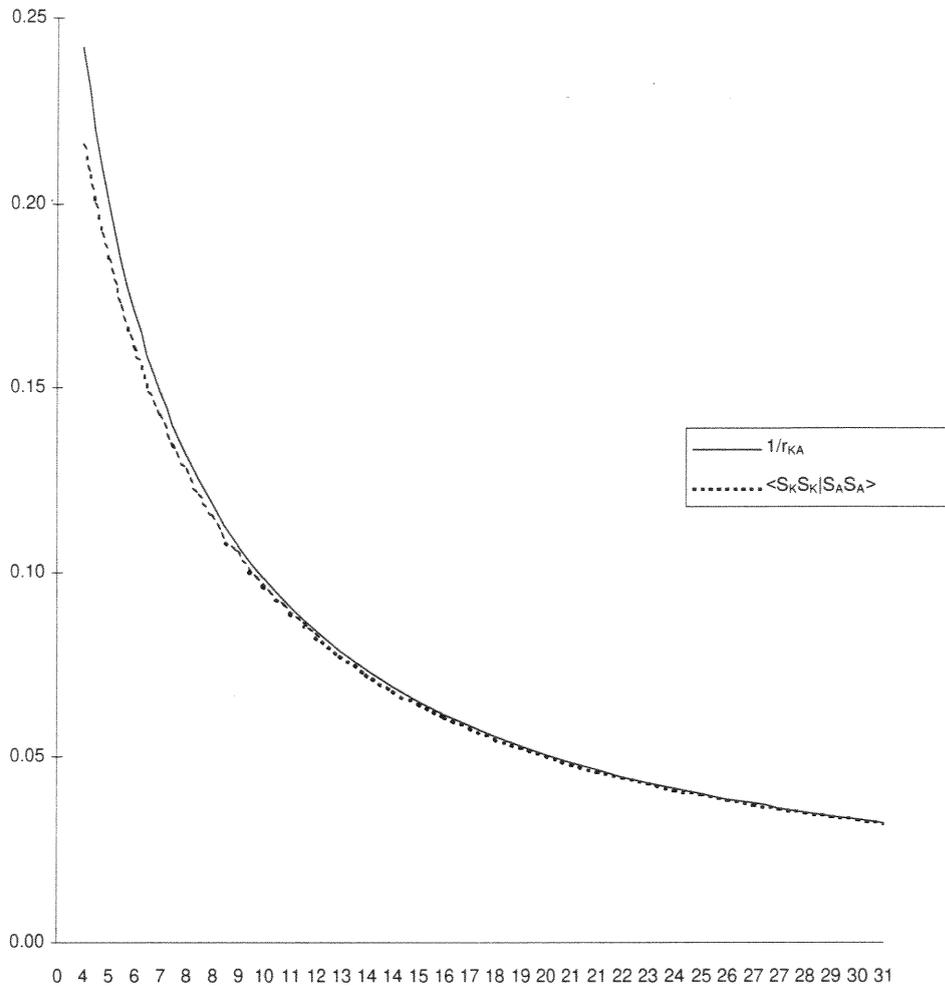


Figure 2.2 : Courbes de comparaison entre les intégrales $\langle S_K S_K | S_A S_A \rangle$ et $1/r_{KA}$

Pour deux atomes quelconques du système, l'un appartenant à la partie quantique et l'autre à l'environnement, nous avons comparé les quantités $\langle S_K S_K | S_A S_A \rangle$ et $1/r_{KA}$ correspondant (r est la distance entre les noyaux des deux atomes). La figure 2.2 montre bien qu'à partir d'une distance de 6 Å l'interaction entre charges de la partie quantique et charges de l'environnement peut être représentée par un potentiel moyen correspondant à $1/r_{KA}$. Dans les études des effets électrostatiques et stériques de la protéine, présentées plus loin, nous avons vérifié, que l'expansion multipolaire au centre de la cavité de $1/r$ convergeait bien à partir de 6 Å également.

1.5.2 Application

On trouve dans le tableau 2.1, les différences en géométrie (sous forme de RMS), le moment dipolaire de la partie quantique optimisée, ainsi que l'énergie totale du système, l'énergie de liaison, le potentiel moyen dû à la partie lointaine du système et le temps de calcul en CPU qui a été nécessaire simultanément avec l'ancienne et la nouvelle version du programme (cutoff=6 Å).

Tableau 2.1

	Thrombine/RS			
	AMI		PM3	
	Sans cutoff	Avec cutoff	Sans cutoff	Avec cutoff
RMS		0.015		0.067
DE (totale)		6.769		2.980
Energie de liaison	-6801.189	-6794.420	-6845.506	-6842.122
Moment dipolaire	31.595	31.462	30.005	30.445
Temps CPU par cycle*	188.0	68.0	197.0	71.0

* temps CPU calculé pour un cycle d'optimisation. Energie en (kcal/mol), moment dipolaire (Debye) et le temps en (seconde)

Cette méthode de calcul des interactions à longue distance entre charges de la partie classique et électrons de la partie quantique nous a permis de gagner un temps de calcul considérable (facteur moyen de 3) sans perte de précision comme le montre la différence en géométrie (RMS) de la partie quantique optimisée.

2. Prise en compte de la polarisation de l'enzyme (Modèle du continuum)

2.1. Introduction

Il est reconnu que l'environnement protéique¹¹ dans son ensemble intervient de façon non négligeable au niveau du mécanisme réactionnel en stabilisant ou en

déstabilisant les réactants, les intermédiaires réactionnels ou les produits. On peut considérer dans ce cas de figure que l'enzyme joue le rôle d'un solvant dans une région interne et localisée, en l'occurrence le site actif de l'enzyme. En effet, l'ensemble des structures secondaires et tertiaires ont un rôle reconnu dans le problème de reconnaissance du substrat et l'importance des perturbations électroniques engendrées est aujourd'hui reconnue et de plus en plus souvent prise en considération¹². Il a été aussi démontré, qu'une paire d'ions est plus efficacement stabilisée par les dipôles d'un enzyme que par l'eau¹¹. Ceci est dû à la faible valeur de la constante diélectrique du milieu protéique¹². En général, dans les modèles classiques, la protéine est représentée comme un solvant non polaire¹³ tant que ses moments permanents et le solvant sont pris en compte explicitement et non par l'intermédiaire de la constante diélectrique. Apparemment, la constante diélectrique représente les effets qui ne sont pas inclus explicitement dans un tel modèle. Ainsi, si toutes les contributions polaires étaient considérées explicitement, en incluant les dipôles permanents et induits de la protéine et leur relaxation par la réorganisation des charges, la constante diélectrique de ce solvant devrait être égale à l'unité¹⁴.

Il est clair que les effets électrostatiques jouent un rôle important dans la détermination des structures et des fonctions des protéines et la prise en compte des énergies électrostatiques est probablement le moyen le plus efficace pour corréler structure et spécificité des protéines¹⁵. Aussi, dans les études théoriques le fait d'apporter une amélioration, aussi minime soit-elle, dans l'évaluation de cette énergie électrostatique peut avoir des conséquences non négligeables dans la compréhension des phénomènes biochimiques.

2.2 Modèle à cavité

Dans le formalisme LSCF, la configuration électronique de l'environnement dans son ensemble est représentée par un champ de force (un jeu de charges ponctuels localisées sur les atomes). La perturbation engendrée par cet ensemble de charges sur les fonctions de base de la partie quantique est représentée uniquement par un terme électrostatique et un terme d'interaction non électrostatique (équation 1.38). Ce modèle, ne prend pas en compte la polarisation de l'environnement électrostatique. Dans le souci

d'amélioration de la prise en compte de la perturbation engendrée par l'environnement protéique, sur une réaction enzymatique, nous devons inclure une contribution de la polarisation dans les interactions entre partie quantique et classique.

Dans les enzymes le site actif est généralement très exposé au solvant (milieu physiologique) de constante diélectrique plus importante. Le fait que le site actif (fragment quantique) se trouve à l'interface de ces deux continuums, va engendrer une polarisation du système non négligeable. D'où l'idée d'une cavité autour de la partie quantique, en considérant la protéine comme un continuum de constante diélectrique faible ($\epsilon = 2-10$)¹⁶. Ceci revient à superposer à l'effet statique de l'environnement classique sur la partie quantique, un effet induit par sa distribution de charge: la polarisation du continuum.

Un des objectifs de ce travail a donc été d'implanter le modèle à cavité de Nancy¹⁷ dans GEOMOP⁶ (qui inclut la méthode LSCF). Ceci, n'a posé de problème que pour la définition du volume et la forme de la cavité. En effet, dans ce modèle à cavité on définit l'ellipsoïde (forme qui s'adapte en général au site actif des protéines) à partir des valeurs et des vecteurs propres du tenseur d'inertie correspondant au volume de densité uniforme défini par la surface de van der Waals¹⁷⁻¹⁸. Ceci est beaucoup plus correct avec les petites molécule, dans lesquelles l'espace entre les sphères de VdW de chaque atome est petit par rapport à ceux existant entre les atomes d'un acide aminé de la partie quantique.

Pour notre part, et dans le but d'adapter une cavité de forme ellipsoïdale à la partie quantique du système, nous avons défini les axes de l'ellipsoïde à partir des valeurs et des vecteurs propres du tenseur d'inertie, et corrigé par un facteur choisi ou calculé automatiquement. En effet, on définit les demi axes de l'ellipsoïde par :

$$a = \frac{1}{2} \sqrt{\sum_{i=1}^N x_i^2} \quad b = \frac{1}{2} \sqrt{\sum_{i=1}^N y_i^2} \quad c = \frac{1}{2} \sqrt{\sum_{i=1}^N z_i^2} \quad 2.14$$

où N est le nombre d'atomes du système quantique, (x_i, y_i, z_i) sont les coordonnées de l'atome i .

Nous avons calculé des facteurs correctifs (k_x, k_y, k_z) respectivement pour les axes $\bar{a}, \bar{b}, \bar{c}$. Ces facteurs sont définis en fonction de la coordonnée maximale T_M^i ($T^i = X^a, Y^b$ ou Z^c), le long de chaque axe d'inertie :

$$k_x = \frac{X_M^a}{a} \quad k_y = \frac{Y_M^b}{b} \quad k_z = \frac{Z_M^c}{c}$$

Les axes de l'ellipsoïde s'écrivent, dans la nouvelle notation :

$$\bar{a}' = k_x \bar{a} \quad \bar{b}' = k_y \bar{b} \quad \bar{c}' = k_z \bar{c}$$

De cette façon, on est sûr que l'ellipsoïde enveloppe toute la partie quantique. Il est clair qu'une cavité définie de cette façon ne représente qu'une approximation d'une cavité bien adaptée à la forme du site catalytique.

En incluant le terme électrostatique du solvant dans le Fockien, l'équation 1.32 devient :

$$F'_{\mu\nu} = F_{\mu\nu} + \sum_{\lambda\eta} P_{\lambda\eta} \left[(\mu\nu|\lambda\eta) - \frac{1}{2}(\mu\eta|\lambda\nu) + \sum_l P_{ll} (\mu\nu|l\ell) - \frac{1}{2}(\mu\ell|\lambda\ell) \right] \quad 2.15$$

$$+ \sum_{A \in E} Q_A (\mu\nu|s_A s_A) + \sum_{l=0}^{\infty} \sum_{m=-l}^l R_l^m \langle \mu | \hat{M}_l^m | \nu \rangle$$

où $F_{\mu\nu}$ est l'élément de la matrice associé au Fockien non perturbé et μ et ν sont deux fonctions de base.

L'énergie totale s'écrit sous la forme , équation 2.16 :

$$\begin{aligned}
E = & \sum_i^{\text{occ}} \sum_{\mu} \sum_{\nu} \left\{ H_{\mu\nu} + \frac{1}{2} \sum_{\lambda} \sum_{\eta} P_{\lambda\eta} \left[(\mu\nu|\lambda\eta) - \frac{1}{2} (\mu\eta|\lambda\nu) \right] \right\} + \frac{1}{2} \sum_{\substack{K \in \mathcal{S} \\ L \in \mathcal{S} \\ L \neq K}} Z_K Z_L f_{KL} \\
& + \sum_{\ell} P_{\ell\ell} \left\{ \sum_{\mu} \sum_{\nu} P_{\mu\nu} \left[(\mu\nu|\ell\ell) - \frac{1}{2} (\mu\ell|\nu\ell) \right] + \sum_{\ell'} \frac{1}{2} P_{\ell'\ell'} (\ell\ell|\ell'\ell') + \sum_{A \in \mathcal{E}} Q_A (\ell\ell|S_A S_A) + \sum_{K \in \mathcal{S}} Z_K (\ell\ell|S_K S_K) \right\} \\
& + \sum_{A \in \mathcal{E}} \left\{ Q_A \sum_{\mu} \sum_{\nu} P_{\mu\nu} (\mu\nu|S_A S_A) + \sum_{K \in \mathcal{S}} Z_K \left[Z_A f_{KA} - P_A (S_K S_K|S_A S_A) \right] \right\} + V'(E, S) \\
& - \frac{1}{2} R_{\ell}^m M_{\ell}^M
\end{aligned}$$

Par conséquent l'effet du solvant est pris en compte de façon auto-cohérente, ce qui permet d'inclure les effets de polarisation au cours du calcul SCF.

BIBLIOGRAPHIE

1. A. Warshel , Proc. Natl. Acad. Sci, **75**, 5250 (1978) ; A. Warshel , Acc. Chem. Res, **14**,284 (1978) ; A. Warshel et J. Aqvist, Annu. Rev. Biophys. Chem, **20**, 267 (1991)
2. D. Bashford et M. Karplus, J. Phys. Chem, **95**, 9556 (1991)
3. C. Chipot, J. G. Ángyán, B. Maigret et H. A. Sheraga, J. Phys. Chem, **97**, 6628 (1993)
4. A. Warshel , et M. Levitt, J. Mol. Biol, **103**, 227 (1976)
5. C. G. Gray, Can. J. Phys, **54**, 505 (1976)

6. V. Théry, Thèse, Université Nancy1 (1993) ; V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, G. Ferency, *J. Comp. Chem.*, **15**, 269 (1994)
7. D. Rinaldi, P.E. Hoggan, A. Cartier, GEOMOS, QCPE 584
8. J.A. Pople, D.L. Beveridge and P.A. Dobosh, *J. Chem. Phys.*, **43**, 5129 (1965)
9. M.J.S. Dewar, E.G. Zoebisch, E.F. Healy and J.J.P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985)
10. J.J.P. Stewart, *J. Comp. Chem.*, 209 (1989)
11. V. Daggett, S. Shroder et J. P. Kollman, *JACS*, **113**, 8926 (1991)
12. C. Mitchison et R. L. Baldwin, *Biopolymers*, **21**, 131 (1975)
13. A. Warshel *Accts. Chem. Res.*, **17**, 284 (1981)
14. J. Aqvist et A. Warshel, *Biochemistry*, **28**, 4680 (1989)
15. A. Warshel et S. T. Russel, *Q. Rev. Biophys.*, **17**, 283 (1984)
16. F. S. Lee, Z. T. Chu et A. Warshel, *J. Comput. Chem.*, **14**, 161 (1993)
17. K. A. Sharp et B. Honig, *Annu. Rev. Biophys. Bio-Phys. Chem.*, **19**, 301 (1990)
18. J. L. Rivail et D. Rinaldi, *Chem. Phys.*, **18**, 233, (1976)
19. J. L. Rivail et B. Terryn *J. Chim. Phys.*, **79**, 1 (1982)
20. D. Rinaldi, J. L. Rivail et N. Riguni, *J. Comput. Chem.*, **13**, 675 (1992)

Chapitre 3

Application

à la

catalyse enzymatique

1. Les enzymes

Les protéines les plus diversifiées et les plus hautement spécialisées sont celles qui possèdent une activité catalytique : les enzymes. Toutes les réactions chimiques des biomolécules organiques dans les cellules sont pratiquement catalysées par les enzymes. Elles ont une grande spécificité pour le substrat, accélèrent des réactions chimiques spécifiques et fonctionnent en solution aqueuse, dans des conditions de température et de pH très douces. Les enzymes agissent en séquences organisées, elles catalysent les centaines d'étapes réactionnelles des voies métaboliques qui permettent aux molécules nutritives d'être dégradées, à l'énergie libre d'être conservée et transformée, aux

macromolécules biologiques d'être synthétisées à partir de précurseurs simples. Elles accélèrent énormément la plupart des réactions chimiques du monde biologique et interviennent dans tous les mécanismes vitaux. Un examen détaillé de la structure des protéines est donc indispensable pour comprendre le mode d'action des enzymes.

1.1 Composition et structure

Les enzymes sont construites à l'aide d'une ou plusieurs chaînes polypeptidiques. Chaque chaîne polypeptidique est un copolymère d'*acides aminés* ($\text{NH}_3^+ \text{-CHR-COO}^-$), qui présentent 20 formules différentes et sont assemblés dans un ordre défini où séquence. Les acides sont reliés entre eux par les *liaisons peptidiques* CO-NH , créées par élimination d'eau entre deux groupements carboxyle (-COOH) et une amine (-NH_2), que l'on retrouve dans la fonction amide. Les différents maillons du polymère sont individualisés par les *chaînes latérales R* des acides aminés successifs. L'établissement de la liaison peptidique fait donc disparaître les charges ioniques potentielles du carboxyle et de l'amine. Cependant la chaîne conserve à une extrémité une fonction aminée libre, et à l'autre une fonction carboxylique libre. Ces fonctions sont ionisées à pH neutre. L'extrémité *N-terminale* marque par convention le début de la chaîne polypeptidique, par opposition à l'extrémité *C-terminale*. Toutes les liaisons peptidiques sont orientées dans le sens de $\text{C} = \text{O}$ vers N-H , dans le même ordre que les lettres C et N dans le mot CARBONE. L'assemblage de la séquence ou *structure primaire* se fait de proche en proche au niveau des ribosomes à partir de l'extrémité N-terminale, au cours d'un mécanisme appelé *traduction* : la succession des nucléotides (monomères dans lequel un groupe phosphate (-O-PO_3) est lié de manière covalente à un sucre, le désoxyribose qui est un ribose) dans l'*ARN messenger* est traduite en séquences d'acides aminés en fonction du code génétique.

Cette séquence d'acides aminés, ou structure primaire, est appelée à se replier dans l'espace, mais de façon non triviale. Un premier repliement, la *structure secondaire*, se fait au niveau du squelette peptidique séquence -[NH-CH-CO]- , pour donner des *hélices*, des *feuilletts*, des *coudes* et des *pelotes statistiques* (structure aléatoire n'entrant pas dans les trois catégories précédentes). Les structures secondaires interagissent entre elles par les chaînes latérales pour former la *structure tertiaire*. C'est souvent ce repliement qui définit en grande partie la forme du *site actif* (l'ensemble des acides

aminés intervenant dans la catalyse enzymatique pour y recevoir un substrat donné). Il existe enfin une *structure quaternaire* qui est le résultat d'une interaction des structures tertiaires des différentes sous unités constitutives de l'enzyme. Cette complexité de la structure spatiale d'une enzyme est due à la diversité des chaînes latérales des acides aminés qui peuvent être acides, basiques, aromatiques, aliphatiques, posséder une fonction alcool, thiol ou amine. La polarité des groupements R varie fortement d'un état totalement non polaire ou *hydrophobe* (insoluble dans l'eau) à un état extrêmement polaire ou hydrophile (soluble dans l'eau). Ainsi, parmi toutes les combinaisons possibles, seules quelques-unes auront une activité catalytique. La longueur de ces groupements R varie également de façon importante (de 1 atome pour la glycine à 18 pour le tryptophane). Cependant, cette structure complexe demeure fragile. Par exemple, une mutation simple (remplacement d'un acide aminé par un autre) peut modifier de façon importante l'activité catalytique.

Certains enzymes sont actifs par eux même, sans autres groupes fonctionnels que ceux de leurs résidus en acides aminés. D'autres, au contraire nécessitent la présence d'un composé chimique supplémentaire appelé *cofacteur*. Le cofacteur peut être soit un ion inorganique, Fe^{2+} , Mg^{2+} , Mn^{2+} ou Zn^{2+} , soit une molécule complexe, organique ou métal-organique appelée *coenzyme*. Un coenzyme ou un ion métallique lié de façon covalente à la protéine enzymatique est appelé *groupement prosthétique*. L'enzyme catalytiquement actif complet avec coenzyme et/ou ions métalliques est appelé *holoenzyme*. La partie protéique d'un tel enzyme est appelé *apoenzyme* ou *apoprotéine*.

1.2 Fonction

Les enzymes sont les catalyseurs biologiques qui permettent aux réactions chimiques nécessaires à la vie et à la multiplication cellulaire de s'effectuer à des vitesses élevées, dans des conditions de température et de pH compatibles avec la vie, et cela avec une spécificité qui élimine la formation de sous-produits indésirables, voire toxiques, pour tout organisme vivant. Les enzymes sont répertoriés en six classes principales, chacune divisée en sous-classes, en fonction des réactions catalysées. Le nom des enzymes a été formé en ajoutant le suffixe "ase" au nom de leur substrat ou à une phrase décrivant leur activité (exemple réductase, oxydase) (tableau 3.1).

Certaines enzymes nécessitent la présence d'un cofacteur de nature non protéique. Dans ce cas, ce n'est pas un acide aminé de la protéine, mais ce cofacteur qui intervient dans la réaction. La séquence protéique n'intervient alors que dans la structure spatiale du site de reconnaissance du substrat et le positionnement adéquat du cofacteur par rapport à ce substrat.

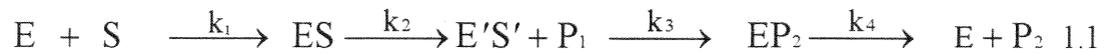
Tableau 3.1: Classification des enzymes

N	Classe	Réactions catalysées
1	<i>les oxydoréductases</i>	modifient simultanément le degré d'oxydoréduction de deux molécules, l'une étant réduite, l'autre oxydée.
2	<i>les transférases</i>	catalysent le transfert d'un groupement monocarbone.
3	<i>les hydrolases</i>	catalysent la rupture de liaisons ester, osidique ou peptidique.
4	<i>les lyases</i>	catalysent l'enlèvement d'un groupe autrement que par hydrolyse ou, au contraire, l'addition d'un groupement.
5	<i>les isomérases</i>	catalysent l'isomérisation.
6	<i>les ligases ou synthétases</i>	permettent l'union de deux molécules, union couplée avec la rupture d'une liaison à haut potentiel énergétique (mono-nucléotide triphosphates A TP, GTP, ...).

Ce cofacteur peut être un métal de transition, un groupement prosthétique (groupement organique non protéique lié de façon covalente à l'enzyme), un coenzyme, très fortement lié à l'enzyme mais de manière non covalente, ou un co-substrat (les vitamines appartiennent à ces deux derniers groupes).

1.3 La cinétique enzymatique

La réaction enzymatique à un seul substrat d'une protéase à sérine peut être décrite par le schéma cinétique suivant :



où S est le substrat (amide ou ester), P₁ et P₂ sont les deux produits issus du substrat (P₁ amine ou alcool et P₂ acide) et E'S' est l'acylenzyme.

L'enzyme peut hydrolyser non seulement des liaisons peptidiques mais aussi des petites molécules déterminées ayant une liaison amide ou ester. Les complexes ES et EP₂ sont non covalents et assimilables à des complexes de Michaelis (EP₂ avec un produit de la réaction). Dans la suite du traitement cinétique du schéma 1.1, on admet que la concentration en enzyme est négligeable par rapport à la concentration en substrat et nous supposons que les constantes k₁ et k₄ sont élevées par rapport à k₂ et k₃. C'est une hypothèse réaliste, étant donné la haute efficacité cinétique des enzymes. La réaction catalytique est divisée en deux processus. D'abord, l'enzyme et le substrat se combinent pour donner le complexe enzyme-substrat (ES). Cette étape est supposée être rapide et réversible (k₂ ≪ k₁ et k₋₁), sans changements chimiques. L'enzyme et le substrat sont réunis par des forces physiques. La réaction chimique a lieu dans un deuxième temps avec une constante de vitesse de premier ordre, k₂. L'approximation de l'état-fixe (la concentration de l'intermédiaire ES demeure constante quand sa vitesse de destruction égale celle de sa formation) donne, pour la vitesse initiale, l'expression suivante :

$$v = \frac{\frac{k_2 K_3}{k_2 + k_3} [E_T] [S]}{\frac{k_3 K_S}{k_2 + k_3} + [S]} \quad 1.2$$

où K_S est la constante de dissociation du complexe enzyme substrat (ES). La constante de Michaelis, K_M, est donnée par

$$K_M = \frac{K_3 K_s}{K_2 + K_3} \quad 1.3$$

La vitesse de la réaction suit la *cinétique de saturation* par rapport à la concentration du substrat [S]. En effet, la quantité de produit formé dépend directement de la concentration en complexe enzyme-substrat, [ES], et cette dernière dépend initialement de la concentration en substrat [S]. Il arrive un moment où, même si [S] continue de croître, la quantité en ES sera limitée par la quantité d'enzyme disponible. La vitesse tend alors vers une valeur maximale (V_{\max}) (figure 3.1). Ainsi,

$$v_{\max} = \frac{k_2 K_3}{k_2 + k_3} [E_T] \quad 1.4$$

Deux cas de figure sont à considérer :

1) La désacylation est très rapide, ce qui veut dire que $k_3 \gg k_2$: le coefficient K_M est sensiblement égal à K_S , la constante de dissociation du complexe ES, et on a

$$v_{\max} = k_2 [E_T] \quad 1.5$$

2) La désacylation est très lente : k_3 est petit devant k_2 et on a :

$$K_M = \frac{k_3}{k_2} K_S, \quad v_{\max} = k_3 [E_T] \quad 1.6$$

On voit tout de suite que le coefficient K_M peut être beaucoup plus petit que la constante de dissociation (l'enzyme est saturé par des faibles valeurs de [S] parce qu'il ne peut pas se débarrasser assez vite du produit P_1). C'est ce qui arrive notamment si la chymotrypsine hydrolyse un *ester*. L'acylation de l'enzyme est très rapide, et la désacylation qui suit n'arrive plus à tenir le rythme. C'est le contraire qui se passe dans le cas des amides.

On peut considérer $\frac{k_2}{K_M}$ comme une constante d'ordre apparent deux. Cette constante est le paramètre le plus important dans la détermination de la spécificité d'un enzyme pour son substrat.

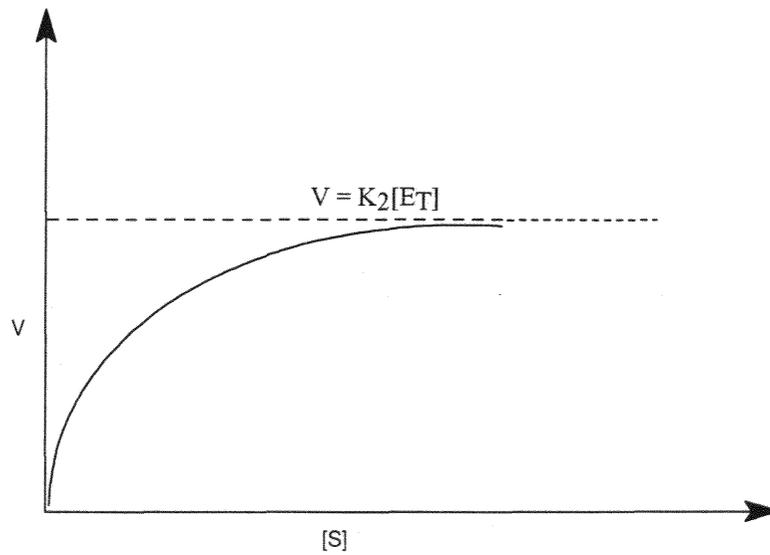


Figure 3.1 : Relation entre la vitesse v et $[S]$ dans une réaction enzymatique type.

La signification des paramètres k_2 , K_M et k_2/K_M peut être mieux comprise en considérant le diagramme d'énergie libre de la figure 3.2.

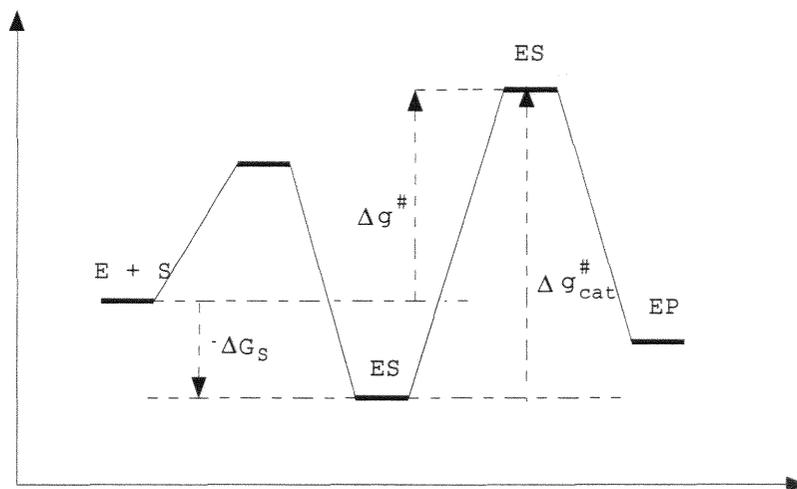


Figure 3.2 : Diagramme schématique d'énergie libre pour une réaction enzymatique type.

Le paramètre k_{cat} est apparenté à la barrière d'activation $\Delta g_{\text{cat}}^{\ddagger}$ et se réfère aux propriétés des complexes enzyme-substrat, enzyme-intermédiaire et enzyme-produit. Le paramètre K_M , apparenté à ΔG_S et sous la condition $K_M \approx K_S$, est une constante de dissociation du substrat de l'enzyme.

1.4 Stabilité des enzymes et facteurs catalytiques

La notion moderne de catalyse enzymatique a été proposée par Haldane en 1930 et élaborée par Linus Pauling en 1946. Pour catalyser une réaction, un enzyme doit être complémentaire de *l'état de transition* de la réaction. Ceci signifie que les conditions optimales d'interaction entre le substrat et l'enzyme sont obtenues seulement à ce stade (par la formation de liaisons faibles). Selon cette théorie de la *complémentarité*^{1,2}, les enzymes catalysent leurs réactions en stabilisant les états de transition correspondants. Depuis, les expériences par diffraction aux rayons X ont montré l'évidence directe de cette complémentarité. La question que se sont posée les chimistes est de savoir comment est accomplie cette stabilisation. Plusieurs propositions ont été formulées pour rationaliser l'énorme pouvoir catalytique des enzymes. Dans cette section, nous étudierons les options importantes.

1.4.1 Les contraintes stériques

On peut considérer comme contraintes stériques les interactions de van der Waals répulsives et les contributions des déformations des liaisons, des angles de valence et des angles dièdres. L'hypothèse des contraintes suggère que les forces stériques des sites actifs des enzymes réduisent l'énergie d'activation par la déstabilisation de l'état fondamental.

Deux éléments cependant semblent porter préjudice à l'hypothèse des contraintes comme facteur clé de la catalyse enzymatique : premièrement, le fait qu'un enzyme est généralement flexible et donc peut accompagner les changements structuraux du substrat sans une grande augmentation de l'énergie libre²; deuxièmement, même si quelques réactions sont associées à des effets stériques significatifs sur Δg^{\ddagger} , une telle déstabilisation du complexe ES (équation 1.1) ne peut provoquer une augmentation de

k_2/K_M qui est (comme c'est illustré sur la figure 3.2) affecté seulement par la différence entre l'énergie de l'état de transition ES^\ddagger et l'énergie de l'état $E+S$.

1.4.2 Les effets entropiques

L'idée que les enzymes catalysent les réactions en utilisant des effets entropiques^{3,4} a été émise fréquemment. Cette idée implique que l'énergie libre de l'état fondamental ES est augmentée par fixation des réactifs et des produits dans une orientation exacte, et ceci est l'effet catalytique majeur. La dépendance entre la barrière d'activation et l'entropie d'activation peut être écrite

$$\begin{aligned}\Delta\Delta G_{s\rightarrow p}^\ddagger &= \Delta G_p^\ddagger - \Delta G_s^\ddagger = \Delta\Delta G_{s\rightarrow p}^\ddagger - T\Delta\Delta S_{s\rightarrow p}^\ddagger \\ \Delta\Delta G_{s\rightarrow p}^\ddagger &= \Delta G_p^\ddagger - \Delta G_s^\ddagger = (S_p^\ddagger - S_p^0) - (S_s^\ddagger - S_s^0)\end{aligned}\quad 1.7$$

avec s et p pour le solvant et la protéine respectivement; S^0 et S^\ddagger désignant l'entropie dans l'état fondamental ES et dans l'état de transition ES^\ddagger .

Au vu de l'équation 1.7, il est clair qu'il est, en principe, possible de réduire $\Delta\Delta G^\ddagger$ en diminuant S_p^0 ou en augmentant S_p^\ddagger par rapport aux termes correspondants dans le solvant. On peut imaginer plusieurs cas pour illustrer cette hypothèse. L'enzyme maintient, en général, les fragments réactifs dans une orientation définie. On peut donc s'attendre à ce que S_p^0 soit réduite par rapport à S_s^0 et donc que $\Delta\Delta G_{s\rightarrow p}^\ddagger$ soit réduite.

Un autre exemple peut être celui d'une réaction non enzymatique qui nécessite d'ordonner les molécules de solvant en réponse à un développement de charges dans l'état de transition, ce qui n'arrive pas dans la réaction enzymatique, puisque les réactifs dans le site actif sont réordonnés. Il se peut donc que ΔS_p^\ddagger soit moins négatif que ΔS_s^\ddagger et donc que $\Delta\Delta G_{s\rightarrow p}^\ddagger$ soit moins important. Cependant, il faut noter que le traitement des effets entropiques reste très difficile à mettre en oeuvre et les méthodes théoriques ne sont pas suffisamment fiables pour traiter un tel problème.

1.4.3 Les effets électrostatiques

Les contributions électrostatiques sont les interactions charge--charge, charge--dipôles induits et les interactions attractives de van der Waals. Ces dernières peuvent être également classées comme contributions stériques. Comme il a été discuté plus haut, les effets catalytiques des différentes classes d'enzymes peuvent être attribués à la stabilisation électrostatique de l'état de transition ES^\ddagger . Les enzymes peuvent créer des micro-environnements qui complètent par leur potentiel électrostatique les changements dans les charges pendant les réactions.

Ces contributions électrostatiques sont considérées comme le facteur le plus important de la catalyse enzymatique^{5,6}, puisqu'elles peuvent "stocker" de grandes contributions énergétiques. On peut expliquer qualitativement ce phénomène par le fait que les dipôles stabilisateurs du site actif sont préorientés vers les charges de l'état de transition (par exemple, le trou "oxyanion" et la "triade catalytique" des protéases à sérine). Dans les réactions non enzymatiques, le système doit dépenser beaucoup d'énergie pour réorienter les molécules du solvant afin de stabiliser l'état de transition, ce qui est déjà fait dans le site actif des enzymes. On peut donc voir le site actif des enzymes comme un "*super-solvant*" qui offre la solvation optimale pour l'état de transition.

1.5 La modélisation de la catalyse enzymatique par LSCF

Les études théoriques peuvent se révéler très efficaces dans la compréhension des contributions spécifiques des différents facteurs catalytiques dans les enzymes. Cependant, vu le nombre élevé d'atomes dans un enzyme, ces études sont limitées le plus souvent à l'étude du système site actif-substrat, voire à une représentation de ces deux entités par des groupements chimiques adéquats.

Le modèle le plus généralement utilisé, encore de nos jours, consiste en une représentation d'une part des chaînes latérales des acides aminés intervenant dans la catalyse enzymatique, et d'autre part du substrat, par des petites molécules ayant un mimétisme de fonction et de structure avec les entités réagissantes. Or, comme on vient de le voir dans la section 1.4, la prise en compte des interactions avec toute la protéine

est très importante pour une meilleure compréhension de la catalyse enzymatique. Les méthodes mixtes MQ-MM sont un moyen élégant pour le traitement d'un tel type de réactions. L'introduction des algorithmes d'optimisation de la partie quantique du système permet l'étude des chemins réactionnels au sein de l'enzyme et la comparaison avec ce qui se passe dans le vide. La méthode LSCF sera appliquée à l'étude du chemin réactionnel de l'activation du récepteur de la thrombine.

Afin de mieux cerner les possibilités et les limites de notre programme, introduisons les équations thermodynamiques suivantes, à température et à pression constantes, l'énergie libre de Gibbs est minimale à l'équilibre et la barrière d'activation correspondante est calculée comme

$$\Delta G^\ddagger = \Delta E^\ddagger + \Delta H^\ddagger(T) - T\Delta S^\ddagger(T) \quad 1.8$$

où ΔE^\ddagger est la barrière d'énergie déterminée par

$$\Delta E^\ddagger = E(TS) - E(MIN) \quad 1.9$$

où TS désigne la structure de transition et MIN le minimum. $\Delta H^\ddagger(T)$ est le coût en enthalpie thermique pour la formation du TS, calculé par

$$\Delta H^\ddagger(T) = \Delta E(T, TS) - \Delta E(T, MIN) + P\Delta V \quad 1.10$$

où P est une constante et $\Delta E(T, X)$ représente la correction d'énergie thermique interne pour l'arrangement de la structure X (X=TS, MIN). ΔS^\ddagger représente le coût entropique de la formation du TS.

Les possibilités du programme LSCF, dans son état actuel, sont limitées au calcul de la barrière d'énergie donnée par la formule 1.9. Notons que la formule 1.9 comprend le terme *ZPE (Zéro Point Energy)* qui n'est pas calculé par notre programme. Ce calcul ne devrait pas cependant poser de problèmes particuliers, étant donné que nous calculons le

hessien à chaque point stationnaire. C'est à partir des résultats obtenus pour la barrière d'énergie que nous étudierons les effets stériques et électrostatiques de la catalyse enzymatique.

2. Protéases à sérine : la thrombine

2.1 Fonction et description

Il n'est pas nécessaire d'expliquer la nature de la *coagulation du sang* et son utilité physiologique. Une lésion dans un vaisseau sanguin amène rapidement la transformation du *fibrinogène* plasmique en *fibrine*, qui précipite pour former un gel capable de combler la brèche. La coagulation du sang est une cascade enzymatique complexe, où des zymogènes (enzymes synthétisés sous forme inactifs, ce qui permet apparemment de protéger la cellule contre tout effet désastreux) sont convertis successivement en protéases actives. Sa réalisation implique la combinaison caractéristique des facteurs suivants : la protéase de l'étape précédente, un cofacteur protéique non enzymatique, une surface de contact (surface étrangère, membrane des plaquettes sanguines, paroi de l'endothélium) où les partenaires s'associent pour former un complexe dont la stabilité nécessite fréquemment des ions calcium.

Un mécanisme d'urgence capable d'accélérer la coagulation à la suite d'un traumatisme est *la voie extrinsèque* fondée sur le facteur VII et un facteur tissulaire. Il converge avec *la voie intrinsèque* (amorcée par le facteur XII) au niveau du facteur X, qui est responsable à son tour de l'activation de la prothrombine (ou facteur II) en thrombine (facteur IIa) (figure 3.3 a).

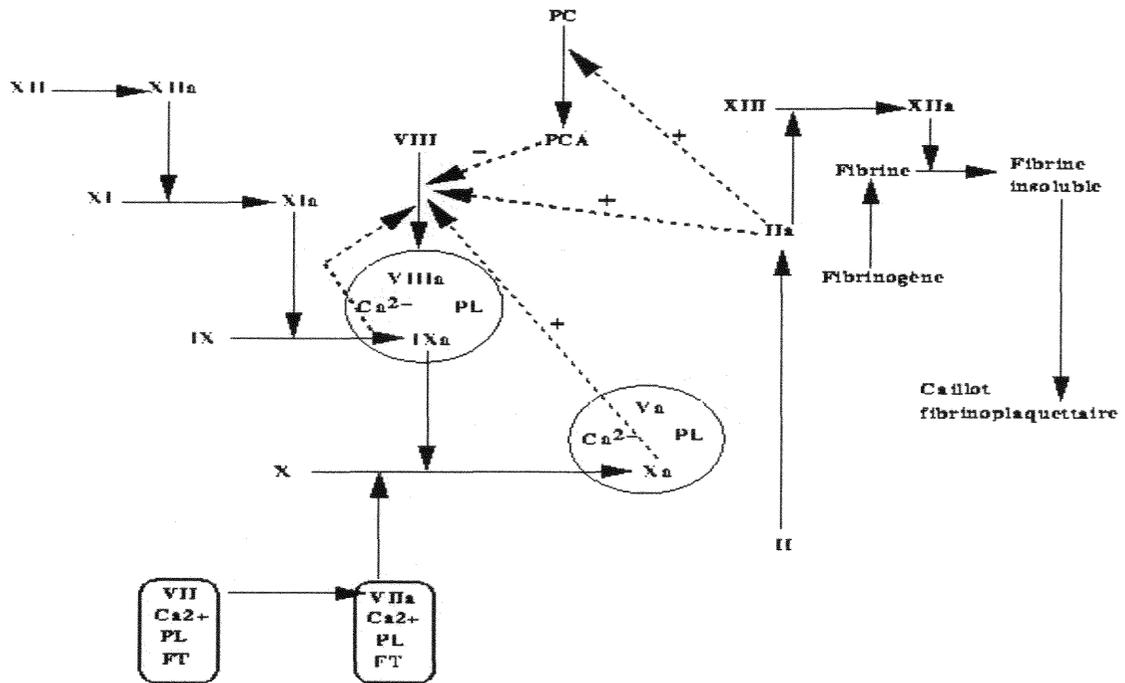


Figure 3.3 a : Schéma simplifié de la coagulation. La coagulation comprend plusieurs étapes de conversion d'un zymogène (facteurs XI, XII, VII, etc.) en enzyme actif (caractérisé par l'indice "a"). Certaines réactions sont accélérées en présence de cofacteurs (facteurs Va ou VIIIa), de phospholipides (PL) et d'ions de calcium (Ca^{2+}). Les mécanismes d'activation (+) et (-) sont représentés ($--\rightarrow$). FT : facteur tissulaire, PCA : protéine C activée.

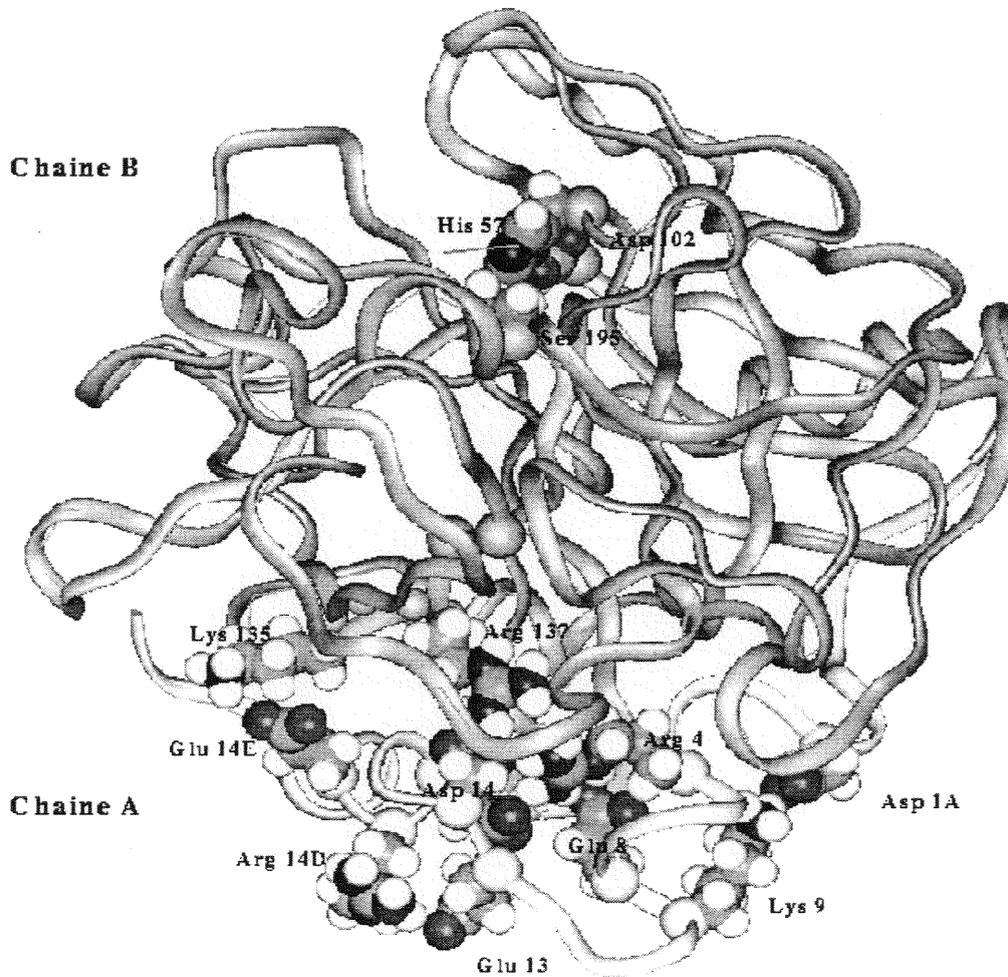


Figure 3.3 b : Vue stéréo des interactions ioniques dans la thrombine entre la chaîne A (en gris clair) et la chaîne B (en gris foncé).

Celle-ci est la pièce maîtresse de la coagulation du sang et va hydrolyser localement le fibrinogène en le faisant précipiter sous forme de fibrine. Le facteur IIa, qui est la thrombine, active lui-même la cascade d'opérations qui conduit à sa formation. Extraordinairement cela veut dire qu'une fois mise en circulation la thrombine est capable de provoquer une formidable auto-accélération de sa propre activation. La thrombine a des effets physiologiques complexes en agissant dans 6 directions : comme agent coagulant (action principale) ; comme anticoagulant en activant la protéine C ; comme stimulant du mécanisme inflammatoire ; comme agent mitogène sur les fibroblastes ; comme agent vasodilatateur (relâchement de l'endothélium) ; comme agent vasoconstricteur et inducteur de la libération des plaquettes^{7,8}. Il est clair que la clé de voûte de l'hémostase est la thrombine, et que cette protéine a des visages multiples. La cristallographie a permis d'accomplir des progrès remarquables et on commence à entrevoir pourquoi cette enzyme possède des propriétés aussi perfectionnées⁹.

Sur le plan structural, elle présente de grandes similitudes avec d'autres protéases telles que la trypsine, la kallikreine ou l'élastase. La thrombine a une masse de 34 kDa et est constituée de deux chaînes. La chaîne A de 49 résidus ne montre pas d'homologie décelable avec les enzymes pancréatiques. La chaîne B, cependant, est très similaire par sa séquence à la trypsine, la chymotrypsine et l'élastase (figure 3.3).

2.2 Triade catalytique des protéases à sérine

La thrombine comme la trypsine contient une triade catalytique, une cavité oxyanionique et un résidu aspartate au fond de la crevasse qui lie le substrat. L'activité catalytique de ces enzymes dépend de l'exceptionnelle réactivité de la sérine 195. Un groupe $-CH_2OH$ est habituellement très inerte dans les conditions physiologiques mais très réactif dans le site actif des protéines à sérine. Une explication plausible a été apportée par des études par rayons X de la structure tridimensionnelle de l'enzyme. L'histidine 57 est voisine de la sérine 195, et le groupe carboxyle de l'aspartate 102, enfoui dans la protéine, est lui aussi proche de l'histidine 57. Ces trois résidus forment une *triade catalytique* (figure 3.4).

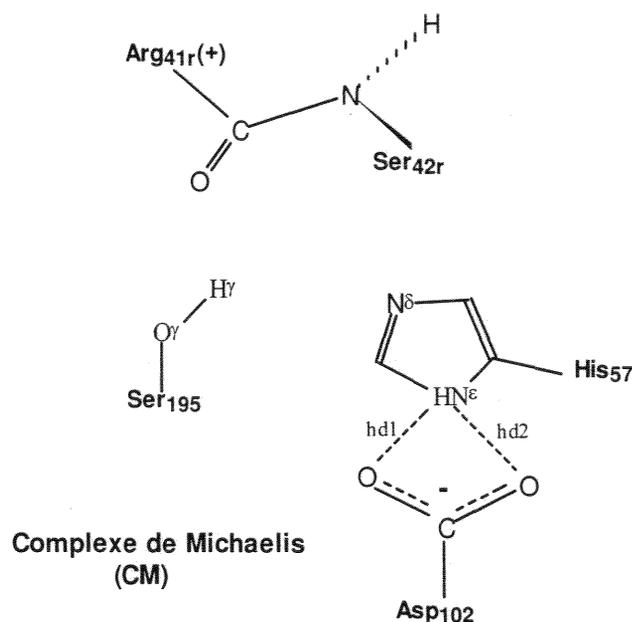


Figure 3.4 : Représentation schématique du site actif de la thrombine et du dipeptide RS.

En l'absence de substrat, l'histidine 57 est déprotonée. Cependant, elle est en équilibre avec l'aspartate 102 pour accepter le proton venant du -OH de la sérine 195 lorsque cet atome d'oxygène effectue une attaque nucléophile sur le substrat. On peut penser que l'aspartate 102 à son tour se protone, et qu'ainsi, cette triade peut être considérée comme un réseau de relais de charge. Cependant, des études par diffraction de neutrons ont montré que le proton de l'azote N^{ϵ} demeure sur l'histidine 57 et que l'aspartate 102 reste chargée négativement. En réalité, le rôle du groupe $-\text{COO}^-$ de l'aspartate 102 est de stabiliser la forme chargée positivement de l'histidine 57 dans l'état de transition. En plus, l'aspartate 102 assure l'orientation de l'histidine 57 dans la forme tautomérique appropriée pour accepter un proton venant de la sérine 195.

2.3 Mécanisme catalytique des protéases à sérine

Un mécanisme catalytique plausible de cette famille d'enzymes a été déduit des nombreuses informations fournies par cristallographie des rayons X et données chimiques. Dans ce mécanisme, l'histidine 57 et la sérine 195 participent directement au clivage de la liaison peptidique sensible du substrat. L'hydrolyse de la liaison peptidique commence par une attaque par l'atome d'oxygène du groupe hydroxyle de la sérine 195 sur l'atome de carbone du carbonyle de la liaison peptidique sensible (figure 3.5).

La liaison carbone-oxygène de ce groupe carbonyle devient une simple liaison et l'atome d'oxygène acquiert une charge négative. Les quatre atomes maintenant liés au carbone du carbonyle sont disposés au sommet d'un tétraèdre. La formation de cet *intermédiaire tétraédrique transitoire* (IT1) à partir d'un groupe amide est rendue possible par les liaisons hydrogène entre l'atome d'oxygène du carbonyle chargé négativement (*oxyanion*) et les deux groupes NH de la Sérine 195 et la Glycine193 (figure 3.5). Ces deux groupes participent à une interaction type liaison hydrogène avec l'atome oxygène du carbonyle en augmentant ainsi son électrophilie avant l'addition du groupe hydroxyle de la Sérine 195. Un autre effet important du trou oxyanion est la stabilisation de l'oxyanion de l'intermédiaire tétraédrique.

Les structures cristallines obtenues par diffraction des rayons X des complexes non covalents de quelques inhibiteurs liés à certaines protéases à sérine, incluant la thrombine¹⁰, ont révélé que l'oxygène du carbonyle du substrat est localisé près du trou oxyanion. De plus, la géométrie normalement trigonale du carbonyle se trouve être plutôt tétraédrique.

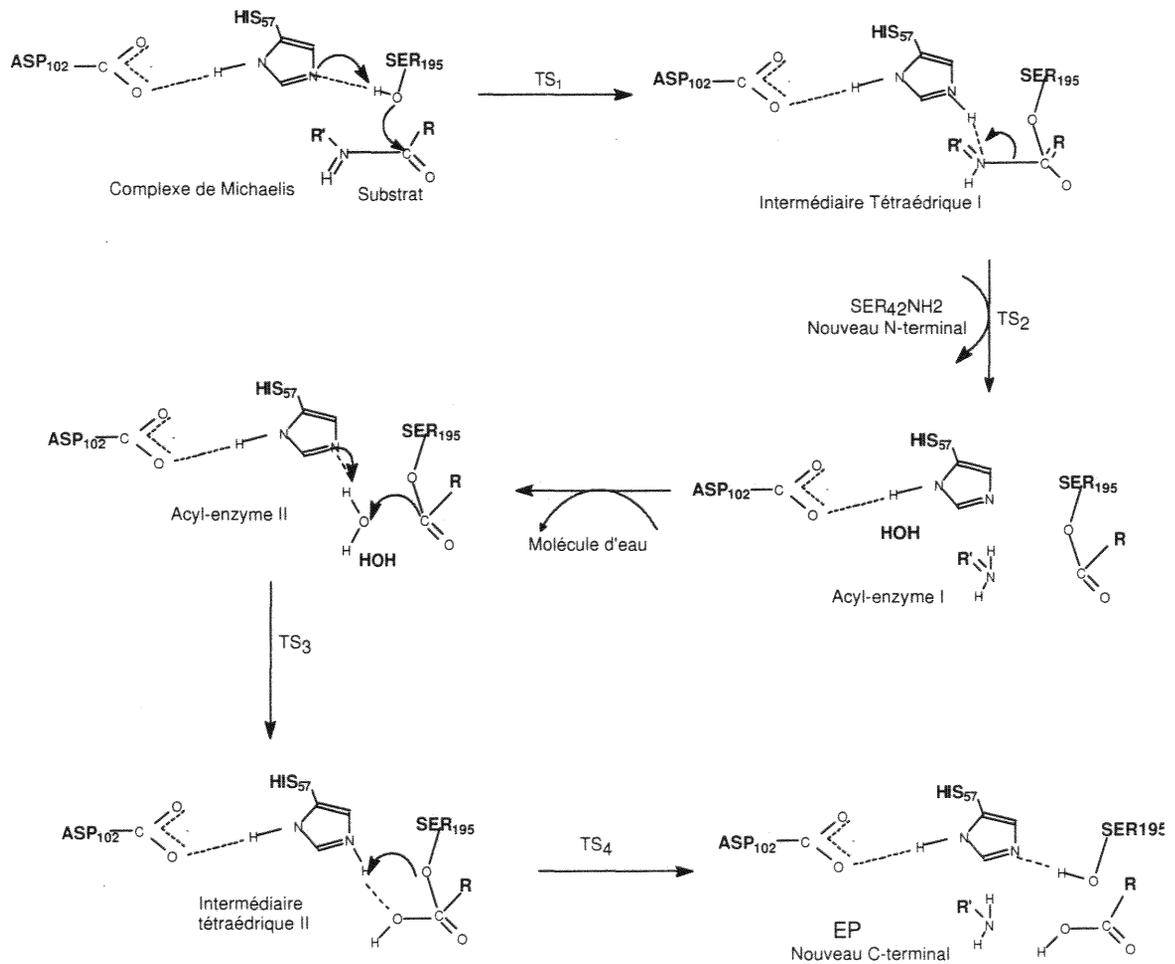


Figure 3.5 : Représentation schématique d'une réaction d'hydrolyse d'une liaison peptidique catalysée par une protéase à sérine.

La pyramidalisation du groupe carbonyle contribue à diminuer l'énergie d'activation requise pour la formation de l'intermédiaire tétraédrique. Aussi, la polarisation de la liaison carbonyle du substrat qui résulte des interactions

électrostatiques avec le trou oxyanion a été postulée être le facteur principal dans la formation de l'acyl-enzyme¹¹.

L'autre événement essentiel dans la formation de cet état de transition tétraédrique est le transfert d'un proton de la sérine 195 sur l'histidine 57. Ce transfert de proton est considérablement facilité par la présence de la triade catalytique. Le proton capté par la forme protonée de l'histidine 57 est ensuite donné à l'atome d'hydrogène de la liaison peptidique sensible qui est alors clivée. A ce stade, le composant amine est uni par liaison hydrogène à l'histidine 57, tandis que le composant acide du substrat est estérifié par la sérine 195. Le composant amine s'élimine, ce qui termine la *phase d'acylation* de la réaction hydrolytique.

La phase suivante, la *désacylation* (figure 3.5), commence lorsque une molécule d'eau prend la place occupée antérieurement par le groupement amine du substrat. En bref, la désacylation est la réaction inverse de l'acylation, où H₂O se substitue au constituant amine. Tout d'abord, le réseau de relais de charge capte un proton du carbonyle du groupe acyle lié à la sérine 195. Comme dans l'acylation, un *intermédiaire tétraédrique transitoire* (IT2) est formé. L'histidine 57 donne ensuite un proton à l'atome d'oxygène de la sérine 195. Ce composant acide s'écarte et l'enzyme est prêt pour un autre cycle de catalyse.

La reconnaissance du substrat par l'enzyme est beaucoup plus complexe que dans le cas de la trypsine, car elle met en jeu une multiplicité de critères qui doivent être respectés pour que le site catalytique, assez profondément enfoncé dans la protéine, puisse fonctionner. Il existe des protéines inhibitrices de ces protéases à sérine qui forment elles aussi une grande famille. L'inhibition provient d'une adhérence très forte entre l'enzyme et l'inhibiteur en des zones de contact déterminées. Par exemple les protéases du type trypsine sont inhibées par plusieurs inhibiteurs caractéristiques appelés inhibiteurs de *Kunitz* : l'inhibiteur de trypsine du soja (STI), l'inhibiteur de trypsine de pancréas de boeuf (BPTI)¹². La thrombine résiste à ces inhibiteurs. On a pu montrer que la cause en est une insertion de séquence qui crée deux boucles supplémentaires dans la structure de la thrombine. L'une d'elles entrave la fixation des inhibiteurs de Kunitz sur la thrombine. Toutefois, si l'insertion d'une boucle au cours de l'évolution qui conduit à la thrombine modifie ses propriétés en la préservant de certains inhibiteurs, en revanche

la séquence supplémentaire lui donne de nouvelles possibilités. Dans le cas de la thrombine, un petit motif de séquence (YPPW)¹³ appartenant à une boucle supplémentaire modifie son action sur le fibrinogène et facilite ses interactions avec un inhibiteur (antithrombine III), la protéine C et la thrombomoduline (figure 3.6).

L'hydrolyse est pratiquée par la triade des trois acides aminés fondamentaux (His₅₇, Asp₁₀₂ et Ser₁₉₅). Les portions de la chaîne du ligand scindé par la thrombine s'encastrent très exactement dans le site catalytique à la faveur de liaisons ioniques multiples. Des interactions fortes sont établies sur le même principe avec les inhibiteurs tels que l'*hirudine* (protéine anticoagulante de la salive des sangsues-*Hirudo medicinalis*). La partie N-terminale s'encastre dans la thrombine, en particulier par des charges anioniques qui interagissent avec les charges positives de l'exosite. Des régions de la thrombine qui participent à différents phénomènes de reconnaissance entre la protéine et la matrice extracellulaire (séquence RGD et domaine chimiotactique) se trouvent près du site. La thrombine est aussi reconnue à la surface des plaquettes par un récepteur qu'elle scinde (voir section suivante). D'autres mécanismes de reconnaissance, également très précis, sont mis en oeuvre avec les facteurs V et VIII, ainsi qu'avec la protéine C et la thrombomoduline (TM cofacteur de la thrombine).

Les interactions électrostatiques jouent un rôle très important pour la reconnaissance des substrats par la thrombine. La thrombine contient une grande proportion d'acides aminés polaires et un grand nombre de résidus chargés. La chaîne A- et la chaîne B de l'espèce humaine contiennent 9 et 31 résidus acides, et 6 et 37 résidus basiques (avec 4 histidines), respectivement, ainsi que deux amines N-terminales, deux carboxyles C-terminales, et deux groupes acides. Plusieurs études théoriques et expérimentales ont montré que la distribution de charges le long du site catalytique peut clairement influencer l'orientation et l'approche des substrats et des inhibiteurs possédant un grand moment électrique¹⁴.

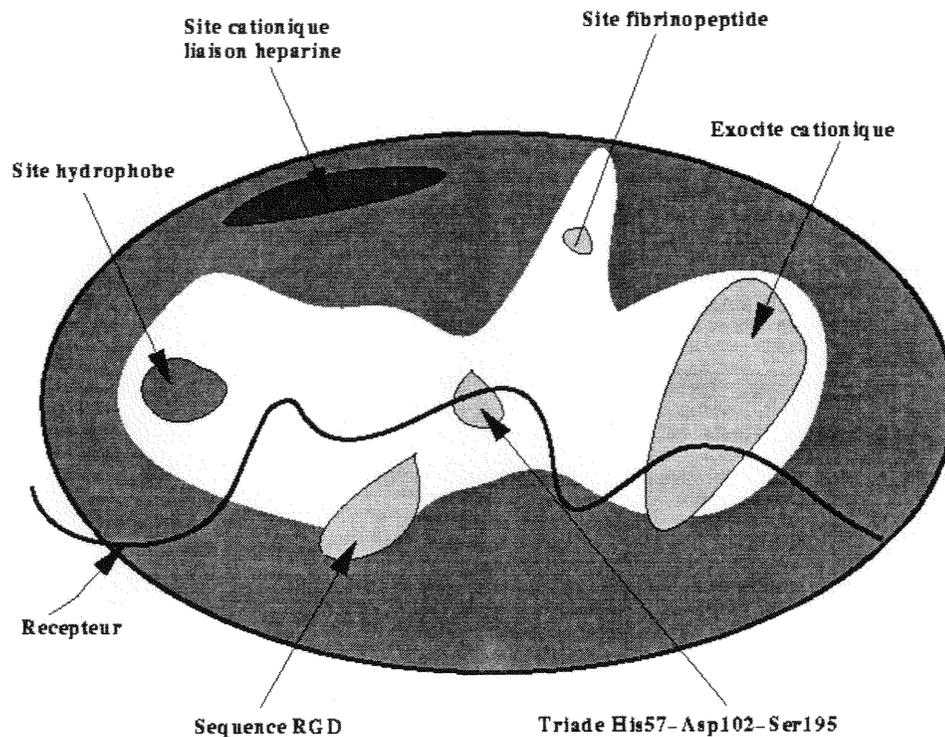


Figure 3.6: Croquis symbolisant la molécule de thrombine avec les sites de reconnaissance. La zone en blanc figure la région où s'installent les polypeptides des substrats et inhibiteurs.

L'analyse électrostatique présentée dans ces travaux distingue trois ensembles de groupes chargés repartis suivant leur rôle dans la stabilité structurale et la fonction de la thrombine¹⁵. Le premier ensemble comprend les groupes engagés dans les ponts salins et les clusters ioniques (Arg4, Glu14, Lys202, Arg137, Asp14...). Ces groupes participent à la stabilité globale de la molécule. Un autre ensemble (Glu164, Asp243, Arg15,..) contribue à la stabilité de la protéine à travers les interactions électrostatiques à longue distance seulement. Ces résidus ne participent à aucun cluster. Le troisième ensemble complète les deux secteurs de charges positives et négatives de la thrombine. Cependant, les groupes chargés de la thrombine ne sont pas uniquement des facteurs stabilisants mais peuvent jouer d'autres rôles essentiels¹⁶ (tel que l'orientation du ligand). La

thrombine est donc une trypsine que l'évolution a complétée et modulée de façon à rendre son action plus spécifique, tout en la dotant de mécanismes de reconnaissances efficaces qui lui confèrent une multiplicité de fonctions physiologiques.

3. Le récepteur de la thrombine

3.1 Description

Le clonage en 1991 du récepteur de la thrombine des mégacaryocytes humains¹⁷ et des fibroblastes de poumons de hamster (CCL39)¹⁸ a constitué une étape capitale pour l'identification moléculaire d'un récepteur, ainsi que pour l'étude des interactions de la thrombine avec les nombreux types cellulaires. Le récepteur de la thrombine appartient à la superfamille des protéines à sept hélices transmembranaires qui caractérisent les récepteurs couplés aux protéines G. Le récepteur est constitué de 427 acides aminés. Dans son domaine N-terminal extracellulaire (séquence NATLDPR/SFLLRNPNDK---KNE), le récepteur possède un site de clivage spécifique pour la thrombine (figures 3.6, 3.7). A côté de ce site, se trouve une séquence qui correspond au vrai ligand du récepteur, suivie par un domaine riche en résidus acides; ce dernier présente une forte homologie de séquence avec un inhibiteur spécifique de la thrombine, l'hirudine. C'est ce domaine acide, très caractéristique des substrats de la thrombine, qui confère la spécificité de l'interaction entre la thrombine et ses substrats physiologiques (fibrinogène par exemple).

Comme l'activité protéolytique de la thrombine est nécessaire pour la quasi-totalité de ses actions biologiques (agrégation plaquettaire, mitogénécité), la présence d'un site de protéolyse suggérait que le récepteur membranaire pouvait être un substrat pour la thrombine elle-même¹⁹.

3.2 Mécanisme d'activation du récepteur

Les études de mutagenèse dirigée du récepteur, où des expériences avec des peptides synthétiques correspondant au site de clivage entrant en compétition avec le récepteur de clivage par la thrombine, ont montré que le clivage après l'arginine-41 est à la fois nécessaire et suffisant pour l'activation du récepteur²⁰. Selon le modèle actuel d'activation du récepteur, la thrombine, en clivant son récepteur entre les résidus arginine-41 et sérine-42, démasque une nouvelle extrémité NH₂ qui joue le rôle de ligand. En faveur de ce mécanisme d'activation par protéolyse, il a pu être montré que des peptides synthétiques, correspondant à la séquence du récepteur à partir du nouveau résidu N-terminal (Sérine-42 ; figure 3.7), sont capables de mimer l'action de la thrombine dans l'agrégation des plaquettes humaines²¹, et l'activation des voies de signalisation couplées aux protéines G dans les cellules en culture²².

Les études menées par plusieurs groupes sur la relation structure-activité des agonistes peptidiques du récepteur de la thrombine ont révélé que les 6 premiers résidus, SFLLRN*(récepteur humain), sont capables d'induire une activation maximale du récepteur (les peptides de 6 résidus sont les plus puissants). De plus, le groupement NH₃⁺ du résidu N-terminal est crucial pour l'activité agoniste du peptide. Après clivage par la thrombine, le ligand engendré se fixe sur un site du récepteur et induit un changement de conformation qui permet l'interaction du récepteur avec les protéines G. La stimulation du récepteur par la thrombine, ou par les agonistes peptidiques, entraîne un état de désensibilisation qui se traduit par l'atténuation de la réponse biologique²³.

Interaction thrombine-récepteur

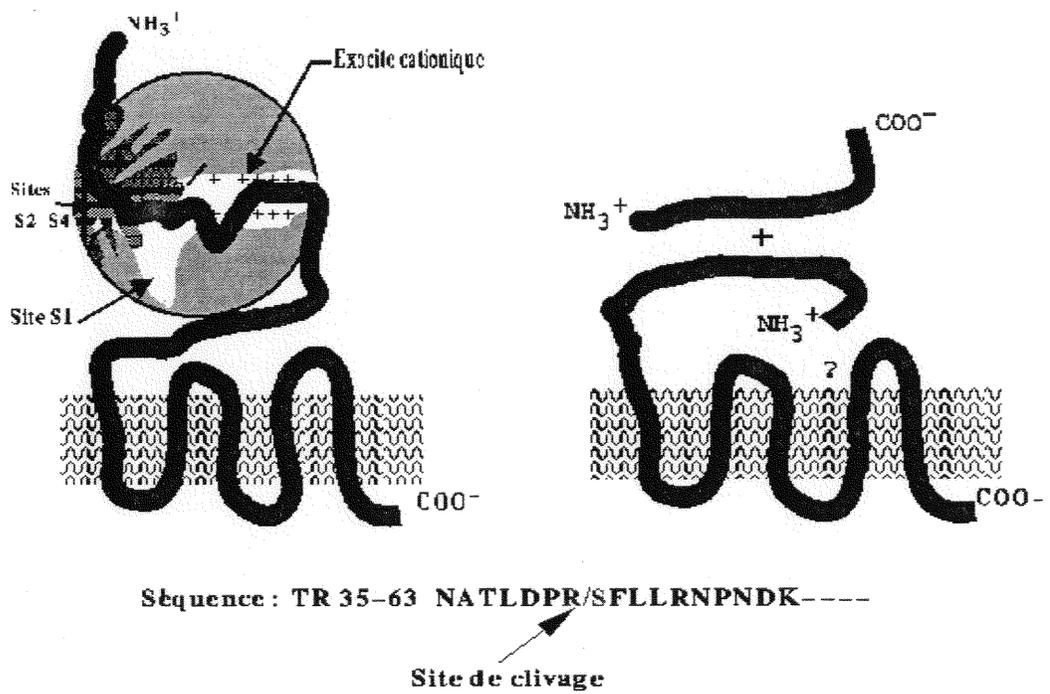


Figure 3.7 : Représentation schématique symbolisant le clivage du N-terminal du récepteur par la thrombine, libérant ainsi, un nouveau N-terminal.

4. Modélisation du complexe thrombine/récepteur par la méthode de Mécanique Moléculaire (MM)

Nous avons utilisé les données cristallographiques du fichier PDB du complexe thrombine /NRS²⁴ (ligand synthétique de séquence NATLDPR/SFLL--) pour créer notre modèle. La structure que nous avons utilisée est donnée pour une résolution de 3.1 Å avec une erreur cristallographique résiduelle de 0.173 Å.

Dans cette structure, la séquence LDPR du peptide NRS n'est pas idéalement positionnée pour contracter un mode de liaison productive dans le site actif. De plus, cette structure cristalline présente quelques distorsions par rapport à la géométrie standard des acides aminés et ne contient pas la séquence complète de la thrombine. Le site de clivage du récepteur, Arg_{41r}-Ser_{42r} se trouve à une distance de 10 Å du site catalytique. Cependant, il est possible que l'approche du peptide NRS soit orientée par l'influence de l'exosite basique et des charges négatives autour du site catalytique (figure 3.8). Pour toutes ces raisons nous avons construit notre modèle à partir de cette structure cristalline en se basant sur le mode de liaison d'un inhibiteur de la thrombine le PPACK²⁵ (D-Phe-Pro-Arg-chlorométhylcétone) (figure 3.9 a).

Nous avons introduit les atomes hydrogène, ceux-ci n'étant pas présents dans les données cristallographiques, en complétant la structure de la thrombine par homologie à une structure bien résolue, le complexe thrombine/PPACK. Une minimisation d'énergie a été effectuée à l'aide du logiciel DISCOVER avec un champ de force CVFF²⁶ sur le système complet (i.e. thrombine/NRS). Elle consiste en la mise en oeuvre de 1000 cycles par la méthode de descente rapide, suivis du nombre de cycles nécessaires pour atteindre la convergence 10^{-3} kcal/mol par la méthode de gradient conjugué. A partir de cette structure, nous avons utilisé une contrainte de distance pour rapprocher la paire d'atomes Arg_{41r}-C et Ser₁₉₅-O , du récepteur et de la thrombine, respectivement, en minimisant l'énergie potentielle du complexe. L'ensemble du système thrombine/récepteur a été soumis à une minimisation d'énergie en fixant la chaîne principale jusqu'à convergence, suivie à nouveau par une minimisation d'énergie en supprimant l'ensemble des contraintes.

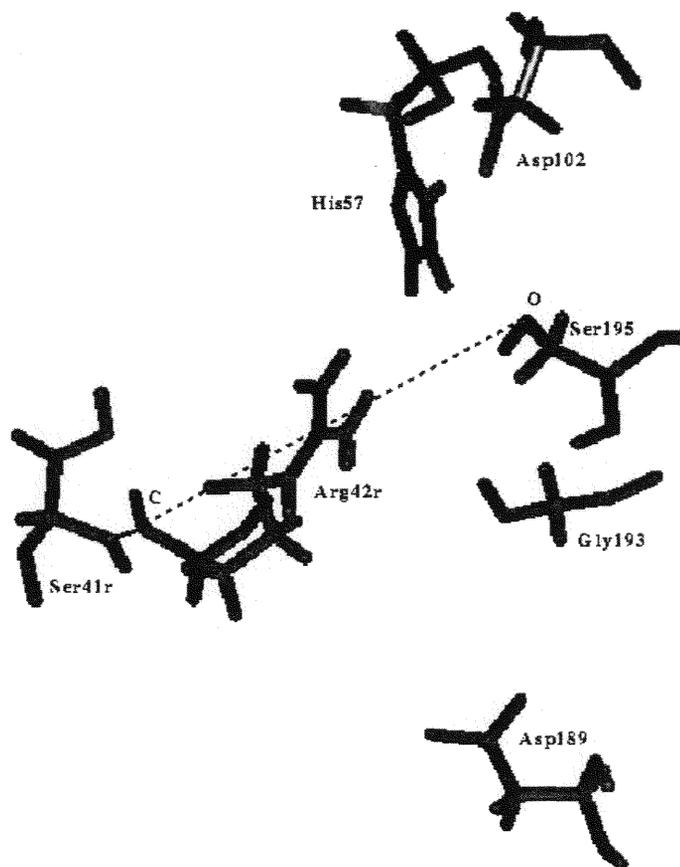


Figure 3.8 : Complexe thrombine/récepteur donné par le fichier PDB.

Dans cette structure minimisée (min_discover) la chaîne latérale de l'Arg_{41r} se trouve dans le site S1 de la thrombine, et la liaison peptidique à hydrolyser se trouve à une distance de moins de 4 Å de l'oxygène de la Ser₁₉₅ (figure 3.9 b).

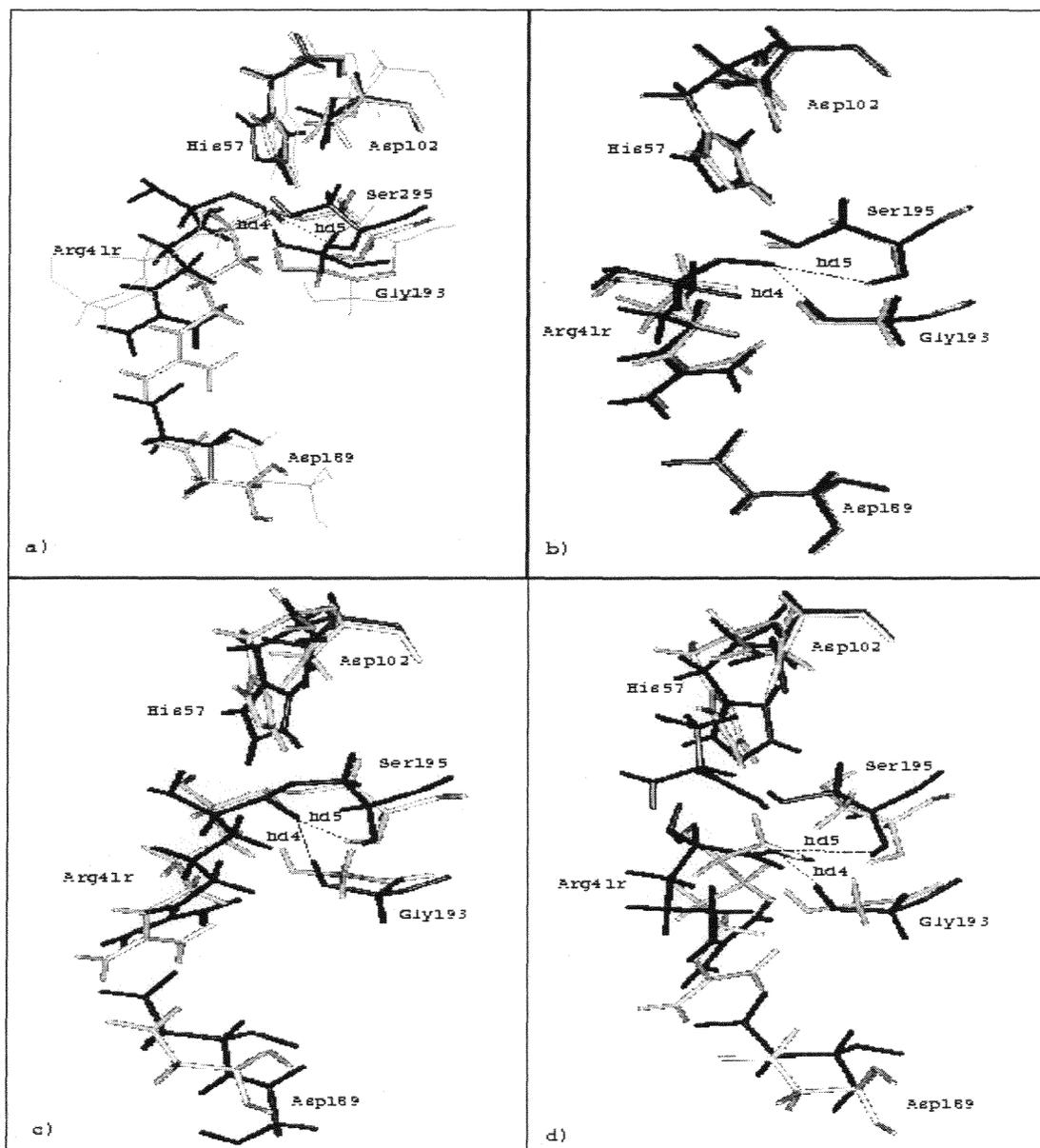


Figure 3.9 : Vue stéréo du site actif de la thrombine et du dipeptide RS du récepteur. superposition de : a, NRS complexe (trait gris fin) et NRP+H complexe (trait gris épais) et min_discover (trait noir); b, min_discover (trait gris épais) et min_geomop (trait noir); c, IT_2(trait noir) et le complexe thrombine/PPACK (trait gris); d, structure EP (trait noir) et NRP+H (trait gris). Les distances hd4 et hd5 sont données en Å et sont indiquées par les pointillés : a, 3.6 et 2.74 (NRS); 2.4 et 2.74 (min_discover); b, 2.38 et 2.71 (min_geomop); c, 2.31 et 2.40 (IT_1); 2.26 et 2.22 (thrombine/PP A CK); d, 3.42 et 2.52 (EP)

Ce mode de liaison de l'Arg_{41r} ressemble à peu près à celui de l'inhibiteur suicide PPACK de la thrombine. Cette chaîne latérale (i.e.Arg_{41r}) forme une paire de liaisons hydrogène avec Asp189 de la thrombine (figure 3.9 b).

5. Modélisation de l'activation du complexe thrombine/RS par la méthode LSCF

Les protéases à sérine constituent, sans doute, l'une des familles les plus étudiées, expérimentalement ou théoriquement (voir, par exemple, les revues²⁷ de Longo et al et Nakagawa et al). Le rôle de la dyade Asp₁₀₂-His₅₇, ainsi que ses formes ioniques ont été longuement discutés. Il est établi que cette dyade augmente le caractère nucléophile de la sérine en facilitant l'attaque sur le carbone du substrat. Cette augmentation du caractère nucléophile résulte du transfert du proton de la sérine vers la dyade. Blow et al²⁸ proposent un double saut de proton conduisant à la triade Asp₁₀₂^o-His₅₇^o-O^r-Ser₁₉₅. Cette hypothèse est soutenue expérimentalement par Hunkapiller et al²⁹, Koeppe et Stroud³⁰, Markeley et Porubcan³¹, et théoriquement par Dewar et Storch³². D'autres études soutiennent l'idée du simple saut de proton, conduisant à la formation de la triade Asp₁₀₂^r-His₅₇⁺-O^r-Ser₁₉₅. Ceci a été suggéré par Wang et al³³, Polgar et Bender³⁴ et confirmé expérimentalement par Robillard et Shulman³⁵, Kossiakoff et Spencer³⁶ et théoriquement par Kollman et Hayes³⁷ et Stamato et Tapia³⁸. L'historique publié par Longo et al²⁶ ainsi que la revue de Kollman et Warshel³⁹, présentent les modèles théoriques proposés dans la littérature et tentent d'expliquer ce dilemme. La plupart de ces études ne suggèrent un chemin réactionnel que pour le processus d'acylation, et soutiennent l'idée du mécanisme concerté du transfert du proton qui passe par un intermédiaire tétraédrique (IT)³⁹⁻⁴².

Dagett et al⁴² proposent un chemin réactionnel pour le processus de désacylation, qui implique aussi un intermédiaire tétraédrique. Aleksandrov et Antonov⁴⁰ suggèrent un mécanisme assisté par une molécule d'eau qui sert comme transmetteur de proton entre la sérine et l'histidine, confirmé théoriquement par Dive et al⁴³.

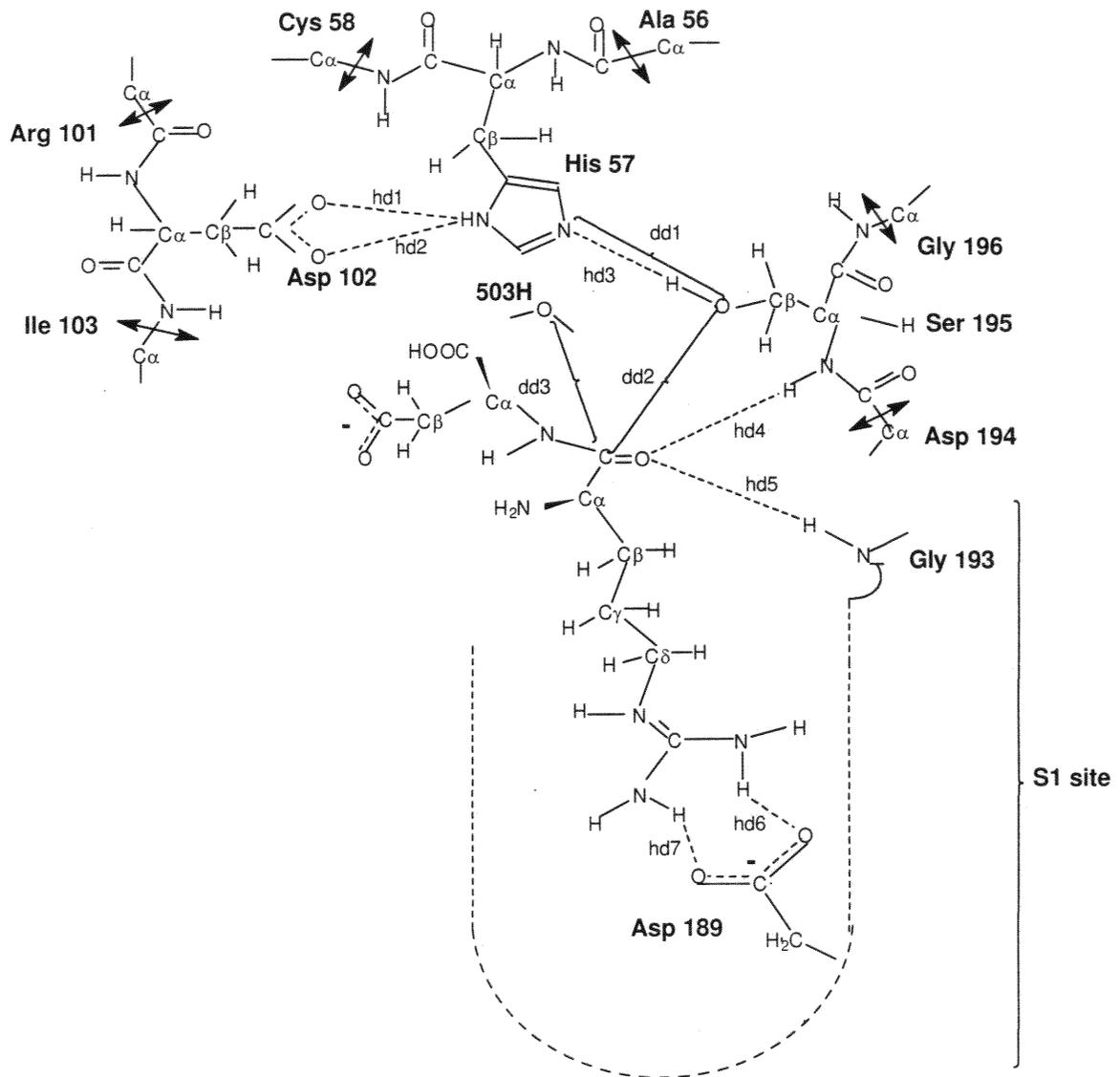
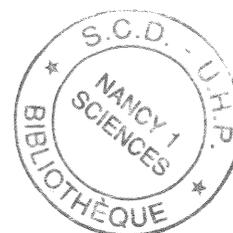


Figure 3.10 : Représentation schématique de la partie quantique du système thrombine/dipeptide RS du récepteur (thrombine : atomes entre double flèches; récepteur : Arg_{41r} et Ser_{42r}). Les liaisons frontières sont les liaisons Arg₁₀₁ C α --CO, Ile₁₀₃ N--C α , Ala₅₆ C α --CO et Cys₅₈ N--C α , Asp₁₉₄ C α --CO et Gly₁₉₆ N--C α .

Komiyama et Bender⁴⁴ proposent une réaction du type S_N2. Dans la première étape du processus, le transfert du proton de la sérine vers l'histidine est accompagné par



l'attaque nucléophile de l'oxygène O^y sur le carbonyle du substrat, rendant ainsi l'azote du substrat plus basique. Dans la deuxième étape, l'hydrogène quitte l'histidine vers l'azote pyramidal du substrat pendant que la liaison peptidique se casse.

Notre objectif est d'essayer de déterminer le chemin réactionnel possible de l'hydrolyse de la liaison sérine⁴¹-arginine⁴² du récepteur dans le site actif de la thrombine. La connaissance du mode de liaison des substrats et du mécanisme de l'hydrolyse des liaisons peptidiques a facilité le développement d'inhibiteurs très puissants des protéases à sérine. Pour créer notre modèle thrombine-récepteur, nous avons remplacé le ligand complet de séquence NRS par le dipeptide RS. L'ensemble a, par la suite, été soumis à une minimisation par un calcul empirique CVFF. Le traitement de notre modèle par LSCF nécessite la définition des sous-systèmes quantique et classique. Notre choix du sous-système quantique est représenté sur la figure 3.10. Le choix d'étendre le fragment quantique jusqu'aux liaisons $C\alpha$ -CO et NH-C des acides aminés entourant les deux groupements du site actif a été dicté par les résultats préliminaires sur l'influence des liaisons frontières sur le sous-système quantique. En effet, ces résultats montrent que la coupure entre les carbones α et β des acides aminés introduit des perturbations non négligeables dans le système^{45a}.

Des études ont montré^{45b} que la séquence KYEPF liée au site fibrinopeptide, bien que prise en compte de façon classique, a une influence sur la déformation de la triade catalytique. Tous les calculs qui suivront dans cette section sont réalisés au niveau semi-empirique AM1. Cette paramétrisation semble bien adaptée à notre étude. Les paramètres a_{11} et P_{11} (voir chapitre 1) utilisés pour les orbitales hybrides gelées sont tirés du calcul sur les acides aminés en jeu pris séparément après localisation des orbitales moléculaires par la méthode de Ruedenberg^{45c}. Les valeurs sont données dans le tableau 3.2.

Orbitale hybride gelée	Paramètres à I1 et PII	
	$a_{\ell 1}$	$p_{\ell \ell}$
A sp-194 (liaison C α --CO) :	0.6988	0.5040
Gly-196 (liaison C α --N) :	0.4892	0.6125
A rg-101 (liaison C α --CO) :	0.6926	0.4919
Ile-103 (liaison C α --N) :	0.4955	0.6255
A la-56 (liaison C α --CO) :	0.6893	0.4885
Cys-58 (liaison C α --N) :	0.4961	0.6212

Tableau 3.2: Paramètres $a_{\ell 1}$ et $p_{\ell \ell}$ calculés au niveau A M1.

Nous savons que le substrat se lie à l'enzyme avec le nombre maximum de liaisons faibles, ou plus exactement que son attachement au site met en jeu des interactions spécifiques suivant le type de substrat utilisé. Les zones du substrat modifiées par l'enzyme doivent être placées avec précision, en orientation et en distance, par rapport aux groupes essentiels appartenant à l'enzyme (i. e. His₅₇, Asp₁₈₉, Arg₁₉₃ et Ser₁₉₅). Il est donc important que la structure du point de départ de la réaction soit bien déterminée. Comme on ne connaît pas de structure cristalline de la thrombine avec un substrat, nous avons préféré construire l'intermédiaire tétraédrique I (IT_1) en se basant sur le mode de liaison d'un inhibiteur de la thrombine : le PPACK⁴⁶ (figure 3.9 c). A partir de cette structure, nous avons modélisé l'étape du passage de l'intermédiaire tétraédrique I au complexe de Michaelis. Cette étape est réalisée par le saut de proton H γ de l'histidine protonée vers l'oxygène de la Ser₁₉₅ avec, simultanément, le clivage de la liaison ester du substrat avec la Ser₁₉₅. Cette étape préliminaire nous a permis de modéliser le complexe de Michaelis, point de départ de notre étude (min_geomop figure 3.9 b).

5.1 Le complexe de Michaelis

Le complexe de Michaelis est le point de départ de notre étude. Il consiste en un complexe non covalent du peptide RS avec la thrombine (figures 3.9 b et 3.11). La caractéristique principale de ce complexe est l'ensemble des deux liaisons hydrogène entre l'oxygène O du carbonyle et les hydrogène HN des acides aminés Gly₁₉₃ et Ser₁₉₅. Nous avons rassemblé dans le tableau 3.3 les paramètres géométriques pour le complexe de Michaelis des deux structures min_discover et min_geomop avec la structure cristallographique NRS.

Dans toute cette étude du chemin réactionnel de l'activation du récepteur, nous avons relaxé en chaque point le système complet par la méthode MM, en fixant la partie quantique, puis minimisé l'énergie de cette dernière dans le formalisme LSCF. Cette relaxation minimise l'influence de l'approximation de l'état figé de l'environnement et permet aux différentes espèces chimiques de mieux se positionner pour une éventuelle interaction.

	Paramètres géométriques sélectionnés		
	Complexe de Michaelis		
	NRS	min_discover	min_geomop
Distances (Å)			
N-C	01.32	01.35	01.37
C-O	10.17	03.62	03.64
O-H	01.03	00.97	00.97
H-N δ	03.07	01.97	02.37
O-HN _{Ser195}	10.26	02.74	02.65
O-HN _{Gly193}	03.61	02.47	02.72

Tableau 3.3 : Paramètres géométriques sélectionnés du complexe Substrat-Site actif de la thrombine dans la protéine. Résultats obtenus au niveau de calcul AM1.

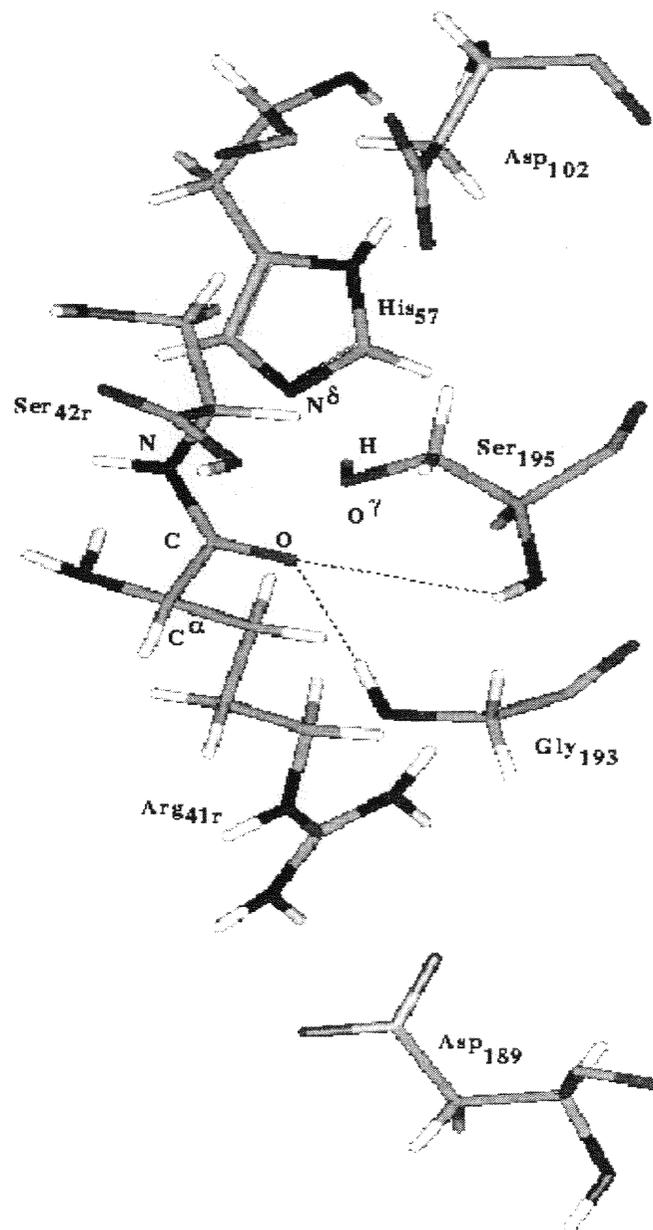


Figure 3.11 : Complexe thrombine/dipeptide RS après minimisation par la méthode LSCF. En pointillés sont représentées les interactions entre le dipeptide et la triade catalytique.

5.2 Chemins réactionnels vers l'intermédiaire tétraédrique

La formation de l'intermédiaire tétraédrique est la première étape de la réaction d'hydrolyse. Elle est caractérisée par l'attaque de l'oxygène O_γ de la sérine 195 sur le carbone C de la liaison carbonyle du dipeptide RS. Le proton H_γ de la sérine saute sur l'azote N_δ de l'histidine 57.

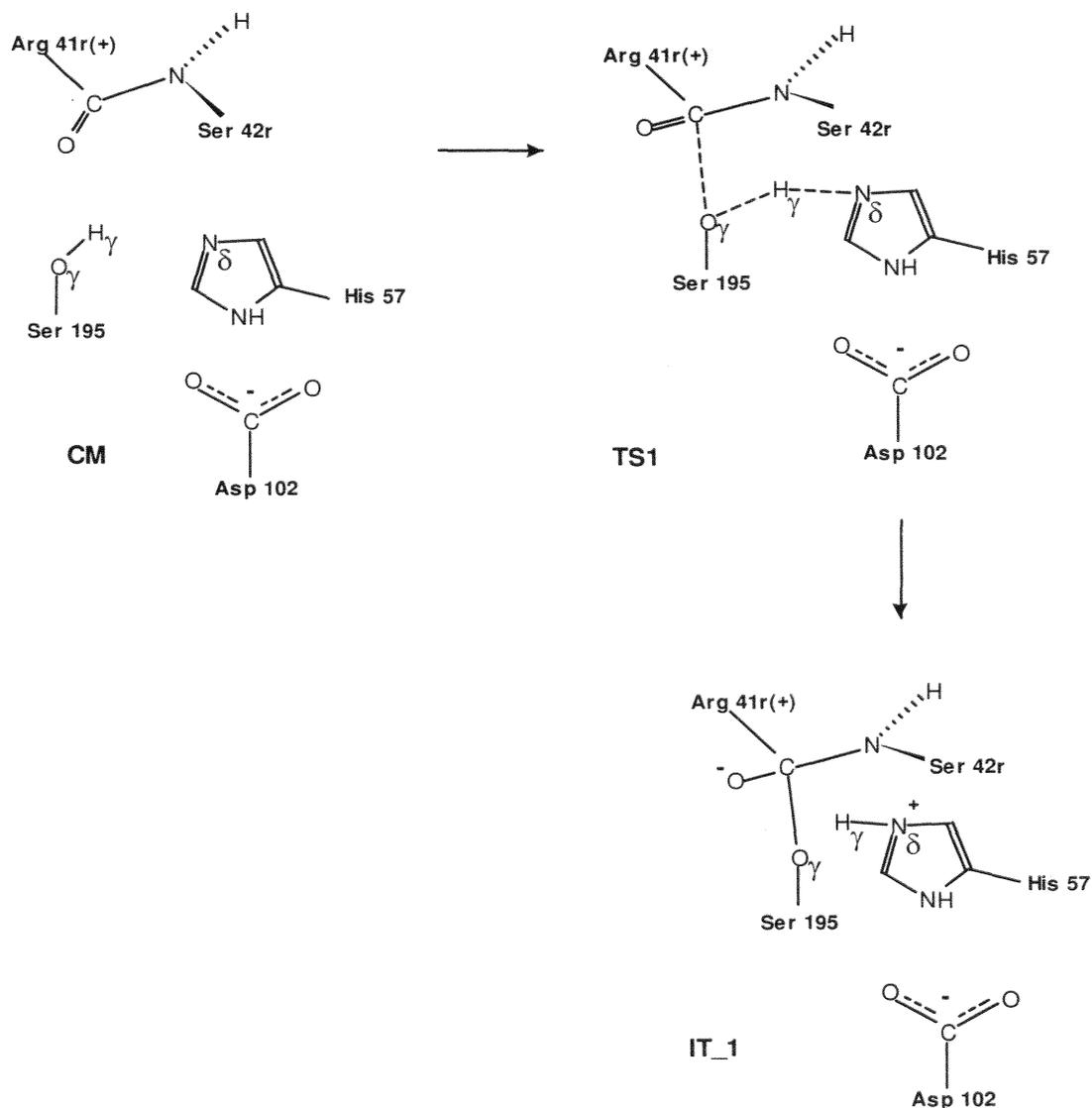


Figure 3.12: Représentation schématique de la première étape de la réaction, passage du complexe de Michaelis à l'intermédiaire tétraédrique selon le mécanisme concerté.

La recherche du chemin réactionnel de cette première étape nous a conduit à l'établissement d'un mécanisme concerté possible vers l'intermédiaire tétraédrique. Ce mécanisme est caractérisé par attaque de l'oxygène O_γ de la sérine 195 sur le carbone C et le saut simultané du proton H_γ sur N_δ de l'histidine 57 (figure 5.2). Les paramètres géométriques sélectionnés sont rapportés dans le tableau 3.4. Pendant le mécanisme concerté, le système passe directement du complexe de Michaelis (CM) à l'intermédiaire tétraédrique (IT_1) via la structure de transition (TS1).

Acylation : Complexe de Michaelis-Intermédiaire Tétraédrique			
	CM	TS1	IT_1
Distances (Å)			
C- O_γ	2.4	1.8	1.5
O_γ - H_γ	1.0	1.8	2.1
H_γ - N_δ	2.0	1.1	1.0
Angles dièdres (degrés)			
C- O_γ ^b	180.0	148.7	126.2
Angles de valence (degrés)			
N- C- O_γ	94.4	102.1	107.4
Energies relatives (kcal/mol)	0.0	21.3	17.0

b : N-C- O_γ -C α

Tableau 3.4: Paramètres géométriques sélectionnés de la première étape de la réaction d'acylation, passage du complexe de Michaelis à l'intermédiaire tétraédrique dans la protéine par un mécanisme concerté. Résultats obtenus au niveau de calcul AM1. a, liaisons hydrogène dans le site actif (figures 3.12 et 3.13 a). b, angle dièdre N- C- O_γ -C α .

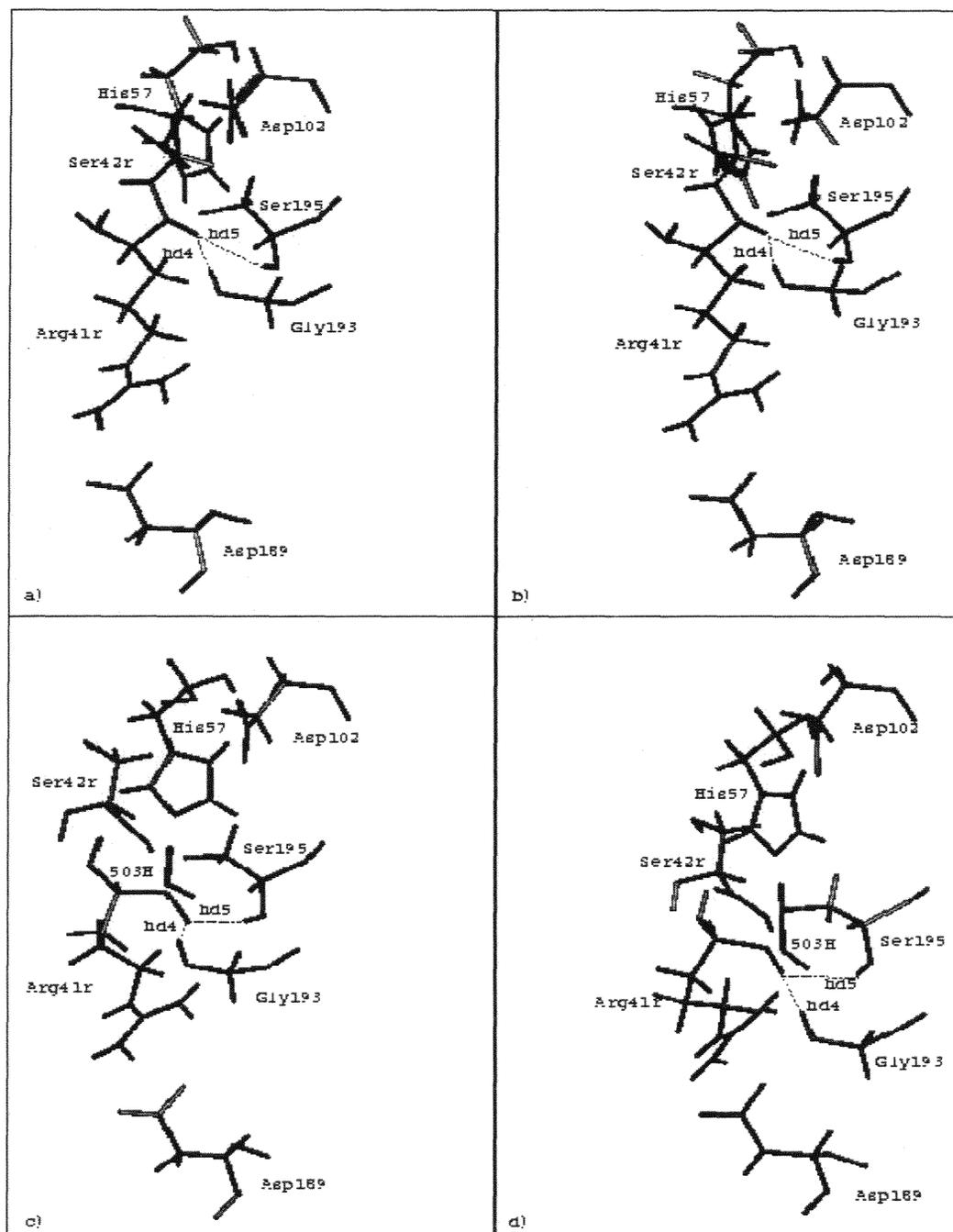


Figure 3.13 : Vue stéréo des états de transition localisés par la méthode LSCF. a : TS1, b: TS2, c: TS3 et d: TS4.

Le vecteur de transition contient deux contributions importantes qui correspondent au saut du proton $H\gamma$ de la sérine 195 sur $N\delta$ de l'histidine 57, et à l'approche de l'oxygène O^γ vers le carbone C du carbonyle.

Il existe également une petite contribution qui correspond à l'angle dièdre C-O (passage à une conformation tétraédrique). Cette étape a une barrière d'activation de 21.31 kcal/mol (10 kcal/mol de moins que celle calculée par Dagett et al⁴²). L'interaction de l'oxygène du carbonyle avec le trou oxyanion est considérablement plus forte dans l'intermédiaire tétraédrique que dans le complexe de Michaelis, et l'atome de carbone du substrat est bien devenu tétraédrique. Dans l'état de transition TS1 (figure 3.13a), la liaison peptidique du substrat s'est sensiblement allongée (ΔR_{C-N} 0.07 Å au niveau AM1) et devient ainsi une simple liaison. L'examen du tableau 3.4 et plus précisément les distances hd1 et hd2 (figure 3.10) montre bien que le proton demeure sur l'histidine contrairement aux conclusions des études théoriques antérieures²⁸ qui soutiennent l'idée du double saut de proton entre les trois résidus Asp¹⁰²-His⁵⁷-Ser¹⁹⁵, où l'His⁵⁷ prend un proton de la Ser¹⁹⁵ et donne un proton à l'Asp¹⁰².

5.3 Etape intermédiaire tétraédrique-acylenzyme

L'étape correspondant au passage de l'intermédiaire tétraédrique à l'acylenzyme est réalisée par le saut du proton H^γ de l'azote N^δ de l'histidine protonée à l'azote N de la liaison peptidique à hydrolyser (figure 3.14).

Acylation : Intermédiaire Tétraédrique-Acylenzyme			
Protéine			
Paramètres géométriques sélectionnés et énergies			
	IT_1	TS2	A CYL_1
Distances (Å)			
N-C	1.50	1.55	2.83
C-O _γ	1.50	1.48	1.36
H _γ -N ^δ	1.01	1.22	4.77
H _γ -N	3.81	1.42	1.00
hd1	2.30	3.67	2.28
hd2	1.89	2.29	2.00
hd4	2.31	2.39	2.43
hd5	2.92	2.39	2.40
Angles dièdres (degrés)			
C-O _γ ^b	126.21	127.47	96.34
Energies relatives (kcal/mol)	16.9	27.7	-19.2

Tableau 3.5 : Paramètres géométriques sélectionnés de la deuxième étape de la réaction, passage de l'intermédiaire tétraédrique à l'acylenzyme dans la protéine. Résultats obtenus au niveau de calcul AM1. b, angle dièdre N- C-O-C^α.

Le phénomène le plus important qui accompagne la création de l'acylenzyme est le clivage de la liaison C-N lors de la protonation de l'azote N. Dans cette deuxième étape de l'acylation, le système passe d'un ensemble à quatre charges (Asp⁻-His⁺-O⁻-Arg_{41r}⁺)

(intermédiaire tétraédrique IT_1) à un système doublement chargé (Asp^- - His^0 - O^0 - Arg_{41r}^+) (l'acylzyme ACYL_1). La barrière d'activation de ce processus est de 27.8 kcal/mol.

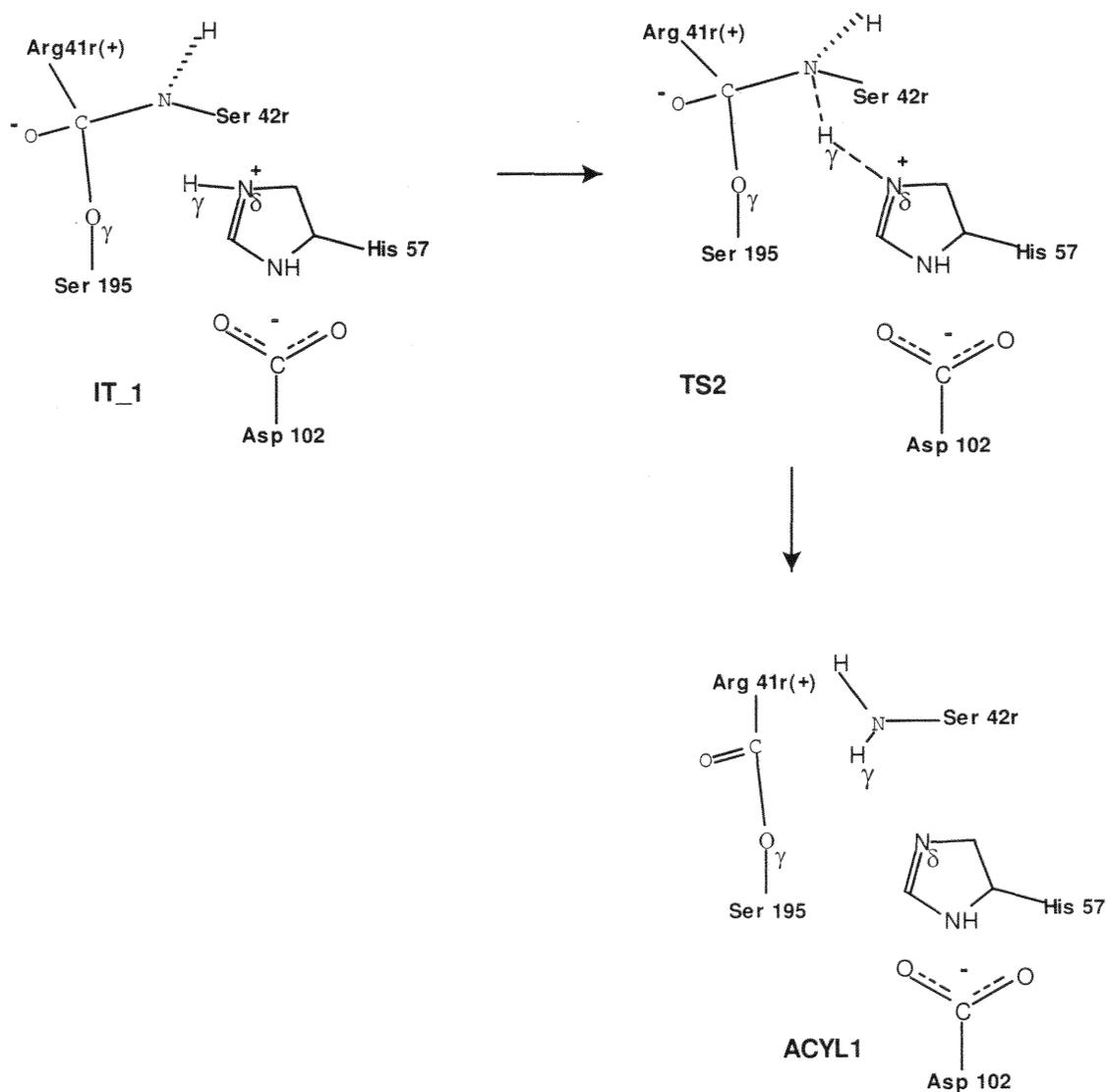


Figure 3.14 :Représentation schématique de la deuxième étape de la réaction, passage de l'intermédiaire tétraédrique à l'acylzyme.

Un point très important de cette étape est l'inversion de l'azote N de la liaison peptidique à cliver, avant que l'hydrogène H' ne le rejoigne. Ce phénomène est le résultat des effets stéréoelectroniques dans les réactions des protéases. Pour l'attaque nucléophile du carbone C du groupe carboxyle par l'oxygène de la sérine, l'état de transition le plus favorable possède une paire libre de l'azote N antiparallèle à la liaison C-O qui se forme. Or, ceci implique que la paire libre de l'azote pointe dans la direction contraire à celle de l'hydrogène qu'il doit recevoir. Ainsi, afin que l'hydrogène H' se dirige vers cet atome, une inversion est nécessaire. Dans la protéine, l'angle dièdre N-C (N-C : C α (Ser42r)-N-H-C) passe de -163.0 degrés dans le complexe de Michaelis (paire libre antiparallèle et qui pointe dans la direction contraire à l'hydrogène) à 131.0 degrés dans l'intermédiaire IT_2 (paire libre prête à recevoir l'hydrogène). Bien que la pyramidalité soit exagérée avec la méthode AM1, on peut remarquer que ces résultats sont en accord avec la théorie du contrôle stéréoelectronique de l'acylation dans les protéases⁴⁶. L'autre point important de cette étape est la diminution sensible de la liaison hydrogène entre l'Asp₁₀₂ et l'His₅₇ (hd1 et hd2 tableau 3.5). Cette interaction faible est le résultat d'une rotation du cycle imidazole pour mieux se positionner pour le transfert de proton. Cette orientation de l'histidine a été montrée récemment dans plusieurs protéases à sérine⁴². Ce mouvement d'ensemble du site actif de la thrombine et du substrat est nécessaire pour le bon déroulement de ce processus chimique.

5.4 Etape acylenzyme-intermédiaire tétraédrique

L'étape de l'activation est réalisée par une attaque nucléophile de l'oxygène O_w (l'indice w caractérise la molécule d'eau) de la molécule d'eau 503H sur le carbone de l'ester formé avec la Ser₁₉₅ (figure. 3.15). Dans la structure ACYL_1, la molécule d'eau 503H est dans une position défavorable pour une attaque nucléophile sur le carbonyle de la liaison ester. En effet, à partir de cette position, la localisation du TS de cette étape par un mécanisme concerté (i. e. attaque nucléophile de l'oxygène O_w sur carbone de la liaison ester et simultanément transfert du proton H1 sur l'azote N^δ de l'histidine 57) ou non concerté (i. e. attaque nucléophile puis le transfert de proton) devenait très difficile.

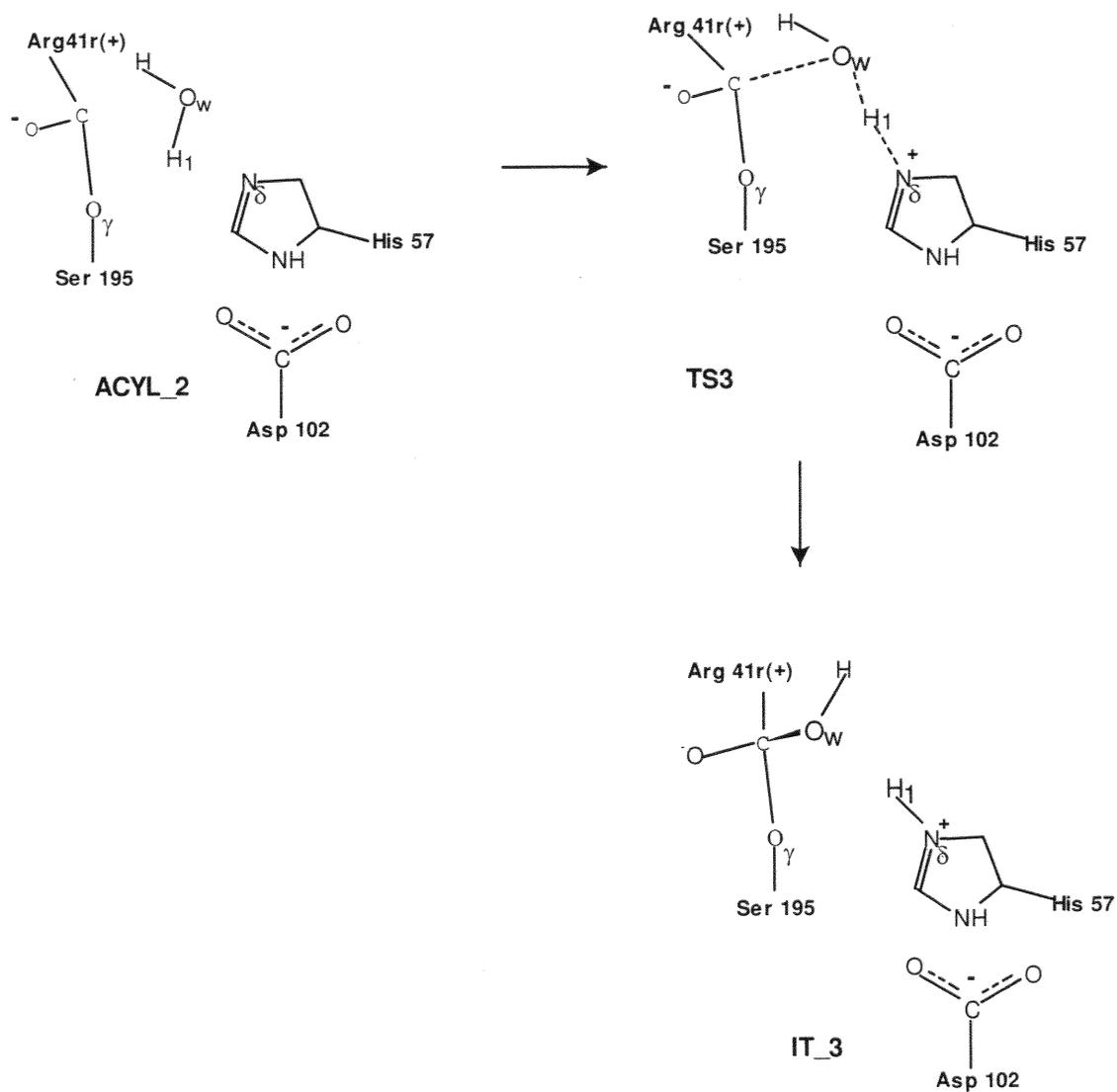


Figure 3.15 : Représentation schématique de la première étape de la réaction de désacylation, passage de l'acylenzyme au complexe enzyme plus produits(EP).

Le vecteur propre négatif du Hessien de cette étape correspond en grande partie et en proportion comparable, à 1) un mouvement d'ensemble de l'amine formé dans la phase d'acylation, 2) au mouvement de l'Arg41r du substrat et 3) d'une orientation de la

molécule d'eau 503H. Il est clair qu'à partir de cette position, la molécule d'eau ne peut pas jouer pleinement son rôle de nucléophile. Pour garder le même système le long du chemin réactionnel, nous n'avons pas voulu pour cette première tentative, remplacer l'amine par une molécule d'eau comme cela a été fait par Kollman et al⁴². Elle doit normalement évoluer de façon à se positionner favorablement pour une éventuelle attaque nucléophile.

Désacylation : Acylenzyme-Intermédiaire Tétraédrique II

	Protéine		
	Paramètres géométriques sélectionnés et énergies		
	A	CYL_2 TS3	IT_2
Distances (Å)			
H ₁ -N δ	2.47	1.21	1.02
O _w -C	3.62	1.53	1.44
C-O γ	1.36	1.43	1.47
hd1	2.34	2.11	1.99
hd2	1.99	2.05	2.03
hd4	2.26	2.67	2.48
hd5	2.41	4.28	3.55
Energies relatives (kcal/mol)	-10.2	29.2	15.21

Tableau 3.6 : Paramètres géométriques sélectionnés de la première étape de la réaction de désacylation, passage de l'acylenzyme au deuxième intermédiaire tétraédrique dans la protéine. Résultats obtenus au niveau de calcul A M1 (figure 3.15).

Pour ce faire, nous avons utilisé la méthode de la *grille* pour placer cette molécule d'eau dans une position favorable au transfert du proton à l'azote N^δ de l'His₅₇. Cette méthode, dite de "grille", consiste en une diminution de la distance N^δ-O_w par pas de 0.05 Å, (figure 3.15) à chaque pas de la formation de la liaison C-O_w (diminution de longueur de cette liaison par le même pas). Un autre minimum (ACYL_2) a été trouvé sans barrière d'activation, dans lequel l'oxygène O_w de la molécule d'eau 503H est à 3.6Å de l'oxygène de l'ester. A partir de cette structure, une attaque nucléophile de l'oxygène O_w sur le carbone de l'ester est réalisée en donnant un deuxième intermédiaire tétraédrique (IT_2). La barrière d'activation correspondant à cette étape est de 29.1 kcal/mol.

L'examen des distances O_w-C et H-N^δ dans le tableau 3.6 montre que, dans cette étape concertée, l'attaque nucléophile sur le carbonyle de l'ester est favorable quand 88% du transfert du proton de l'oxygène O_w de la molécule d'eau vers l'azote de l'His₅₇ est réalisé. Dans la structure de transition TS3, la liaison hydrogène hd4 (figure 3.13c) est affaiblie puisque la distance C-O de l'ester a augmenté. Cette étape de formation du deuxième intermédiaire tétraédrique II est limitante dans ce processus chimique d'hydrolyse de la liaison peptidique. Ces résultats sont en accord avec des études expérimentales réalisées sur des protéases à sérine⁴⁷.

5.5 Etape intermédiaire tétraédrique-enzyme/produits(EP)

Cette dernière étape de la réaction est la régénération de l'enzyme. L'hydrolyse de la liaison ester est le résultat du transfert du proton de l'His₅₇ sur la Ser₁₉₅ (figure 3.16).

Désacylation : Intermédiaire Tétraédrique II-Enzyme/produits(EP)

	Protéine		
	Paramètres géométriques sélectionnés et énergies		
	IT_2	TS4	EP
Distances (Å)			
H ₁ -N _δ	1.02	1.19	1.97
O _w -C	1.44	1.41	1.37
C-O _γ	1.47	1.57	2.42
hd1	1.99	2.08	2.23
hd2	2.03	1.96	2.00
hd4	2.48	2.38	2.52
hd5	3.55	3.19	3.42
Energies relatives (kcal/mol)	15.2	23.8	-8.7

Tableau 3.7 : Paramètres géométriques sélectionnés de la première étape de la réaction de désacylation, passage du deuxième intermédiaire tétraédrique à l'enzyme régénéré dans la protéine. Résultats obtenus au niveau de calcul AM1.

Ce processus est favorable et correspond à une barrière d'activation de 23.7 kcal/mol. Dans cette étape de désacylation la distance C-O de l'ester passe de 1.47 Å (IT_2) à 2.42 Å (EP) (tableau 3.7) (élongation précédée par un mouvement d'atomes de l'Arg_{41r} dans le site S1, site supposé fixe dans l'état actuel du programme).

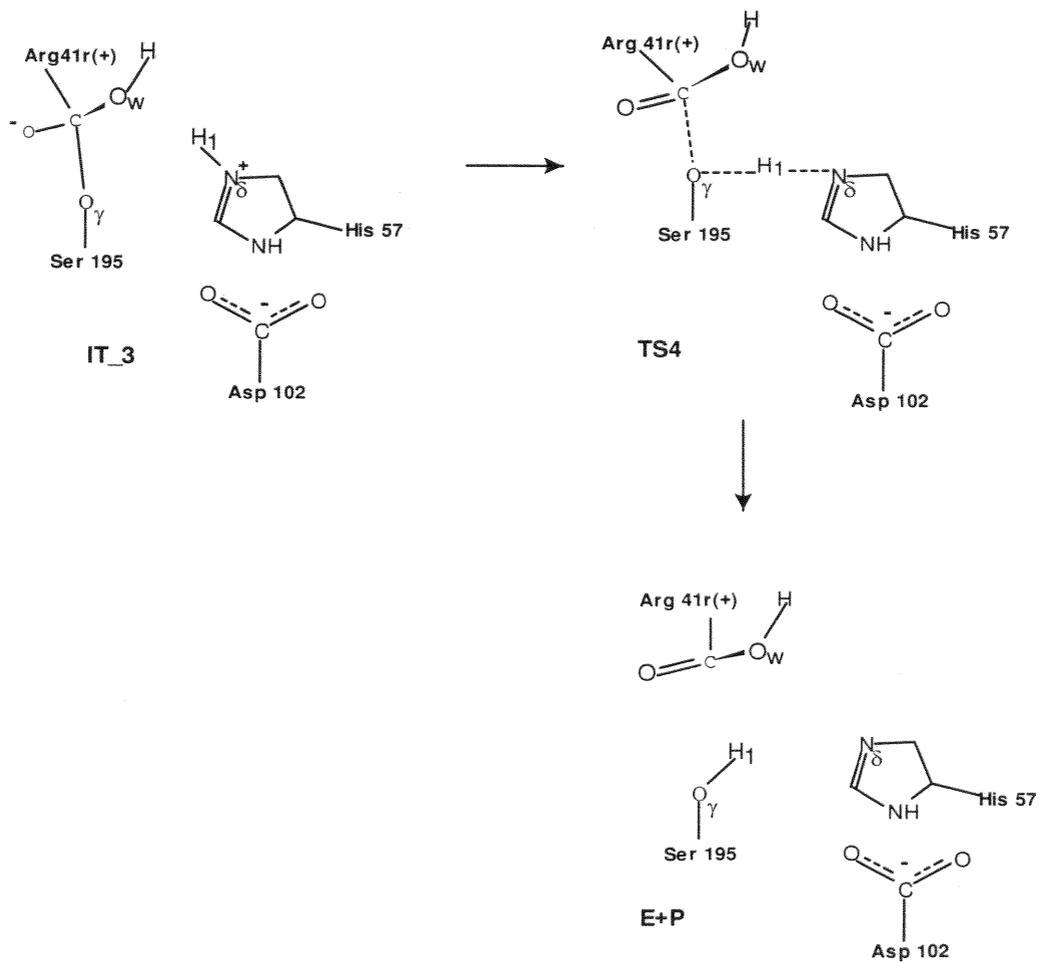


Figure 3.16 : Représentation schématique de la première étape de la réaction de désacylation, passage de l'acylenzyme au complexe enzyme plus produits (EP).

6. Etudes des effets électrostatiques et stériques de la protéine

Comme il a été signalé dans le chapitre 1, l'environnement protéique (i. e. reste du système) est représenté par des charges ponctuelles localisées sur les atomes, et dans le formalisme LSCF, l'interaction de la partie quantique avec le reste du système est représentée par deux termes : un terme électrostatique correspondant à l'interaction des charges de l'environnement avec les électrons et les noyaux de la partie quantique, et un terme stérique correspondant à un potentiel de van der Waals. Afin d'évaluer l'importance relative de ces effets, nous avons dans un premier temps, annulé l'interaction avec la protéine, en considérant le système quantique sans environnement (nous avons appelé par abus de langage cet état "vide"), et nous avons procédé ensuite à une annulation successive de ces effets pour pouvoir estimer leur influence sur les intermédiaires de la réaction.

6.1 Etude des profils énergétiques

Dans cette section, nous procéderons à une étude comparative des profils énergétiques des différents mécanismes dans la protéine et dans le vide. Dans le graphique 3.16, nous avons reporté les profils énergétiques du mécanisme concerté dans la protéine et dans le vide. Les résultats rapportés sur cette figure sont une belle illustration de la catalyse enzymatique. La barrière d'activation de l'étape limitante (la formation de l'intermédiaire tétraédrique II) dans la protéine est de 29.2 kcal/mol alors que dans le vide cette barrière s'élève 48.72 kcal/mol. La détermination de la formation de l'intermédiaire tétraédrique II comme étant étape limitante dans la protéine est en accord avec l'expérience^{48,49}, alors que dans le vide, c'est l'acylation qui est l'étape limitante. Le résultat révélateur de la catalyse enzymatique de la thrombine est la stabilisation très importante de l'intermédiaire tétraédrique dans la protéine par rapport au vide. Cette stabilisation se chiffre en moyenne à 27.5 kcal/mol dans la phase d'acylation et à 4 kcal/mol dans la phase de désacylation. Ce phénomène est dû principalement à l'interaction électrostatique du trou oxyanion. Cependant, il importe de noter que le système étudié dans le vide prend en compte pour moitié l'influence du trou

oxyanion, puisque le groupe NH de la sérine est inclus dans le traitement quantique. Il est possible que le trou oxyanion ne représente qu'une partie de la stabilisation électrostatique par la protéine.

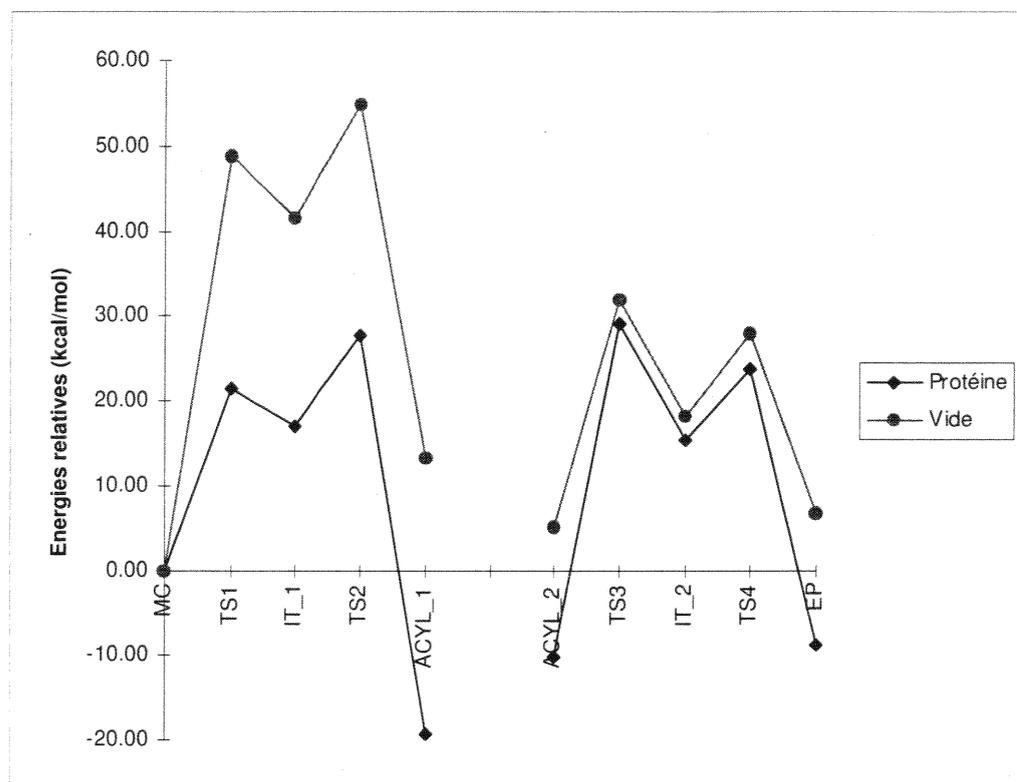


Figure 3.17 : Comparaison des profils énergétiques du mécanisme concerté dans la protéine (trait plein) et dans le vide (pointillés).

La figure 3.17 montre que le fait de ne pas prendre en compte l'influence de la protéine entraîne un comportement différent du système optimisé dans la protéine par rapport au système entièrement étudié dans le vide. La façon dont le système évolue du complexe de Michaelis aux différents intermédiaires reste un sujet de controverses. Différentes études théoriques sur des systèmes mimant les sites actifs des protéases à sérine indiquent que le système évolue de façon non concertée⁵⁰. D'autres soutiennent le mécanisme concerté^{42,51}.

6.2 Effets électrostatiques et stériques de l'environnement.

Nous avons réalisé une série de calculs ponctuels au niveau AM1 sur les structures optimisées dans la protéine en considérant successivement les seuls effets électrostatiques (annulation du terme $V(E,S)$ de l'équation 1.38) ou les seuls effets stériques de la protéine (charges de l'environnement ramenées à zéro). La comparaison des profils énergétiques nous permet de juger de l'importance des effets électrostatiques par rapport aux effets stériques.

	Effets		
	Vide	Electrostatique ^a	Stérique ^b
MC	0.00	0.00	0.00
TS1	48.72	27.40	43.96
IT_1	41.53	22.90	37.16
TS2	54.81	31.49	52.02
ACYL_1	13.26	-14.46	16.76
ACYL_2	5.14	-12.83	-2.63
TS3	31.86	30.57	28.91
IT_2	18.07	18.64	6.60
TS4	27.84	26.55	22.22

Tableau 3.8 : Energies relatives des intermédiaires en fonction de l'environnement. a, les effets stérique ne sont pris en compte ; b, les effets électrostatiques ne sont pas pris en compte.

Le tableau 3.8 regroupe les énergies relatives des différents intermédiaires calculées par rapport au CM, i. e. complexe de Michaelis, choisi comme origine. Ainsi, nous constatons sur la figure 3.18, pour la phase d'acylation, que l'influence principale de la protéine dans la catalyse enzymatique est électrostatique et non stérique. Cet effet est prépondérant dans la phase d'acylation, où la séparation de charge est stabilisée par

un effet électrostatique. Par contre, dans la désacylation, l'effet stérique prédomine. Dans cette réaction, inverse de l'acylation, le nucléophile (en l'occurrence la molécule d'eau 503H), par la méthode de la grille n'a pas pris la place occupée antérieurement par le produit aminé de la phase précédente. On peut penser qu'un mouvement d'ensemble devrait normalement positionner cette molécule d'eau à la place du constituant amine. Or, comme nous l'avons souligné dans la partie méthodologique de ce travail, l'environnement est gardé fixe tout au long de cette étude.

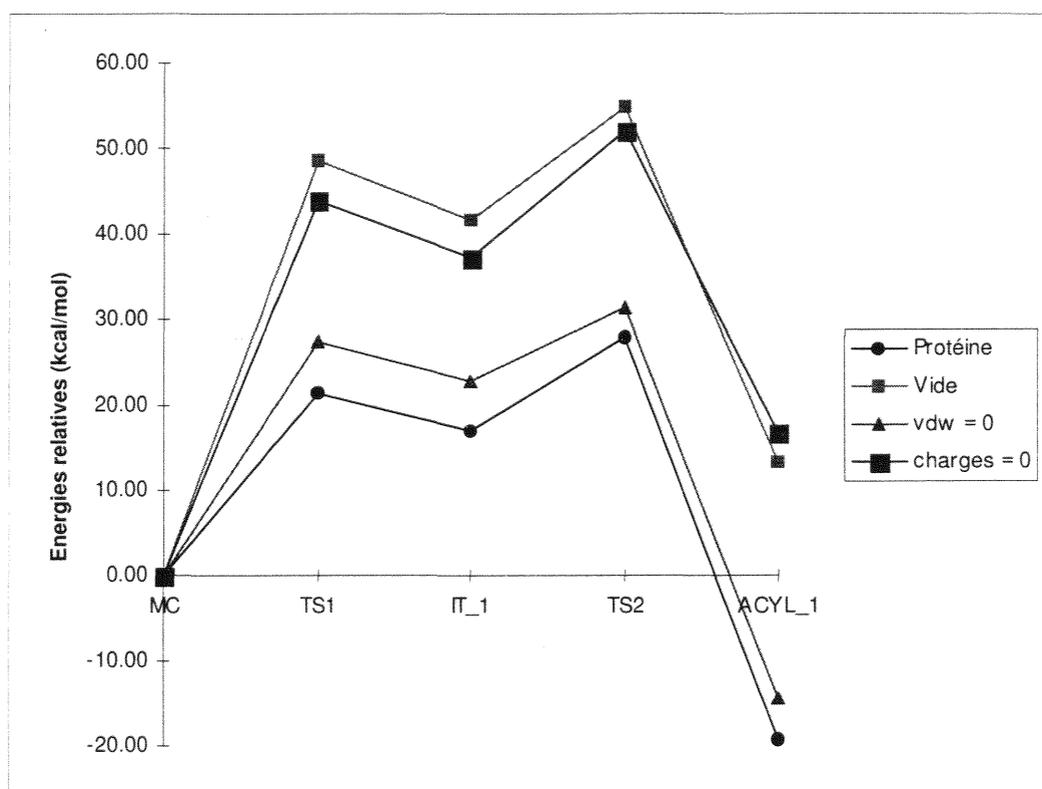


Figure 3.18 : Diagramme énergétique de l'étape d'acylation. Profil énergétique de la réaction d'hydrolyse dans la protéine (cercle), dans le vide (petit carré), effet uniquement électrostatique (en triangle) et effet uniquement stérique (en grand carré).

On peut prévoir alors que, dans le cas où la protéine serait libre de se mouvoir et que, par conséquent, les contraintes stériques seraient mieux représentées le long du chemin réactionnel, l'influence stérique sur nos résultats serait encore plus faible.

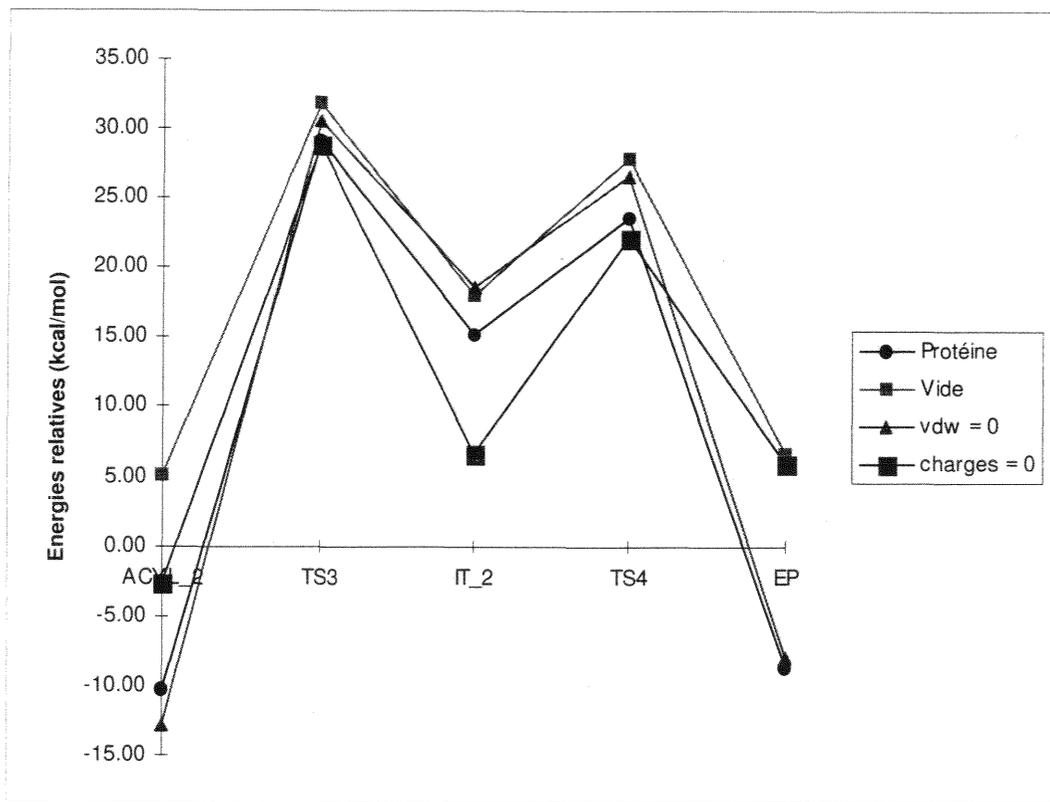


Figure 3.19 : Diagramme énergétique de l'étape de désacylation. Profil énergétique de la réaction d'hydrolyse dans la protéine (cercle), dans le vide (petit carré), effet uniquement électrostatique (en triangle) et effet uniquement stérique (en grand carré).

Dans cette phase de désacylation, contrairement aux effets électrostatiques, les effets stériques déstabilisent les deux intermédiaires ACYL_2 et EP, en revanche, les structures TS3, IT_2 et TS4 sont déstabilisées par ce jeu de charges. Ces résultats sont probablement dûs à des interactions spécifiques qui déstabilisent ce processus de désacylation. En général la conformation du site actif de l'enzyme est déterminée par un équilibre subtil entre plusieurs interactions, et spécialement quand cet enzyme contient plusieurs sites d'interactions complémentaires (figure 3.6). Comme il a été démontré expérimentalement⁵², la catalyse peut être affectée par un ou plusieurs groupes qui ne

sont pas directement impliqués dans la catalyse, mais dont les états ionisés modifient la vitesse de la réaction de la catalyse.

6.3 Effets du champ électrostatique moyen

Pour analyser l'effet de l'enzyme et savoir si cet effet est global ou spécifique sur les différentes étapes de la réaction d'hydrolyse, il serait intéressant de corrélérer la courbe d'étude du chemin réactionnel d'une réaction enzymatique avec celle calculée avec le potentiel électrostatique. Pour cela nous avons entrepris l'étude de la perturbation due à la partie lointaine du système sur la réaction d'hydrolyse du dipeptide RS par la thrombine. Le profil énergétique de cette réaction a été recalculé en utilisant le développement multipolaire, décrit au chapitre 2, pour différents cutoffs (distance à partir de laquelle les charges sont représentées par un développement multipolaire, et au dessous de laquelle les charges sont considérées explicitement).

Afin de tester la validité de nos résultats et de les comparer avec ceux obtenus dans la protéine, nous avons réalisé un calcul ponctuel à l'aide de la méthode AM1 avec différents cutoffs 3, 6 Å. Nous ne présenterons dans cette analyse concernant l'effet du champ moyen que les résultats obtenus pour un cutoff de 3 Å, et pour les autres cutoffs une comparaison est donnée dans l'étude du profil énergétique.

Les tableaux 3.9 et 3.10 montrent que, dans le cas d'un champ électrostatique moyen avec un cutoff de 3 Å, les moments dipolaires des intermédiaires augmentent et, d'après la théorie d'Onsager⁵³, on s'attend à ce que le champ stabilise ces intermédiaires. De plus, l'examen des populations électroniques, de type Mulliken, semble indiquer que la population électronique autour des atomes d'oxygène et d'azote est plus importante avec un champ électrostatique que dans la protéine, tandis qu'autour du carbone et de l'hydrogène elle est plus faible.

	Phase d'acylation				
	MC	TS1	IT_1	TS2	A CYL_1
Moment dipolaire	29.89(26.88)	35.35(31.39)	36.40(32.28)	36.59(32.55)	39.82(35.07)
Charge de Mulliken					
O γ	-0.36(-0.38)	-0.56(-0.56)	-0.44(-0.44)	-0.42(-0.42)	-0.25(-0.26)
N δ	-0.25(-0.29)	-0.16(-0.18)	-0.11(-0.13)	-0.19(-0.20)	-0.20(-0.26)
C	0.26(0.25)	0.37(0.36)	0.35(0.35)	0.32(0.32)	0.27(0.25)
H γ	0.24(0.24)	0.35(0.34)	0.33(0.31)	0.32(0.16)	0.33(0.15)
N-Ser41r	-0.34(-0.36)	-0.36(-0.37)	-0.38(-0.39)	-0.33(-0.32)	-0.33(-0.34)
Energies relatives	0.00	36.10	31.10	39.68	-20.86

Tableau 3.9 : Energies relatives (kcal/mol), moments dipolaires (Debye) et charges de Mulliken (u.a) pour l'étape d'acylation avec un cutoff de 3 Å. Entre parenthèses sont donnés les valeurs des charges de Mulliken et les moments dipolaires dans la protéine.

	Phase de désacylation				
	ACYL_2	TS3	IT_2	TS4	EP
Moment dipolaire	7.70(33.64)	32.11(27.65)	28.03(23.89)	30.54(26.38)	37.80(34.15)
Charge de Mulliken					
O γ	-0.27(-0.26)	-0.40(-0.38)	-0.45(-0.43)	-0.44(-0.43)	-0.35(-0.37)
N δ	-0.21(-0.26)	-0.19(-0.23)	-0.13(0.26)	-0.21(0.24)	-0.25(-0.29)
C	0.27(0.26)	0.34(0.35)	0.35(0.34)	0.36(0.35)	0.27(0.26)
H1	0.21(0.24)	0.37(0.36)	0.35(0.33)	0.37(0.36)	0.23(0.23)
Ow	-0.44(-0.46)	-0.44(-0.45)	-0.50(-0.50)	-0.44(-0.45)	-0.34(-0.35)
Energies relatives	-11.42	39.71	24.95	33.67	1.99

Tableau 3.10 : Energies relatives (kcal/mol), moments dipolaires (Debye) et charges de Mulliken (u.a) pour l'étape de désacylation avec un cutoff de 3 Å. Entre parenthèses sont données les valeurs des charges de Mulliken et les moments dipolaires dans la protéine.

Tout ceci tend à prouver que le champ électrostatique moyen a une grande influence sur la distribution électronique des intermédiaires de la phase d'acylation. Inversement, dans la phase de désacylation, la population électronique autour des atomes

d'oxygène et de carbone diminue par rapport à ceux calculés dans la protéine, tandis qu'autour de l'atome d'azote et d'oxygène O_w elle augmente.

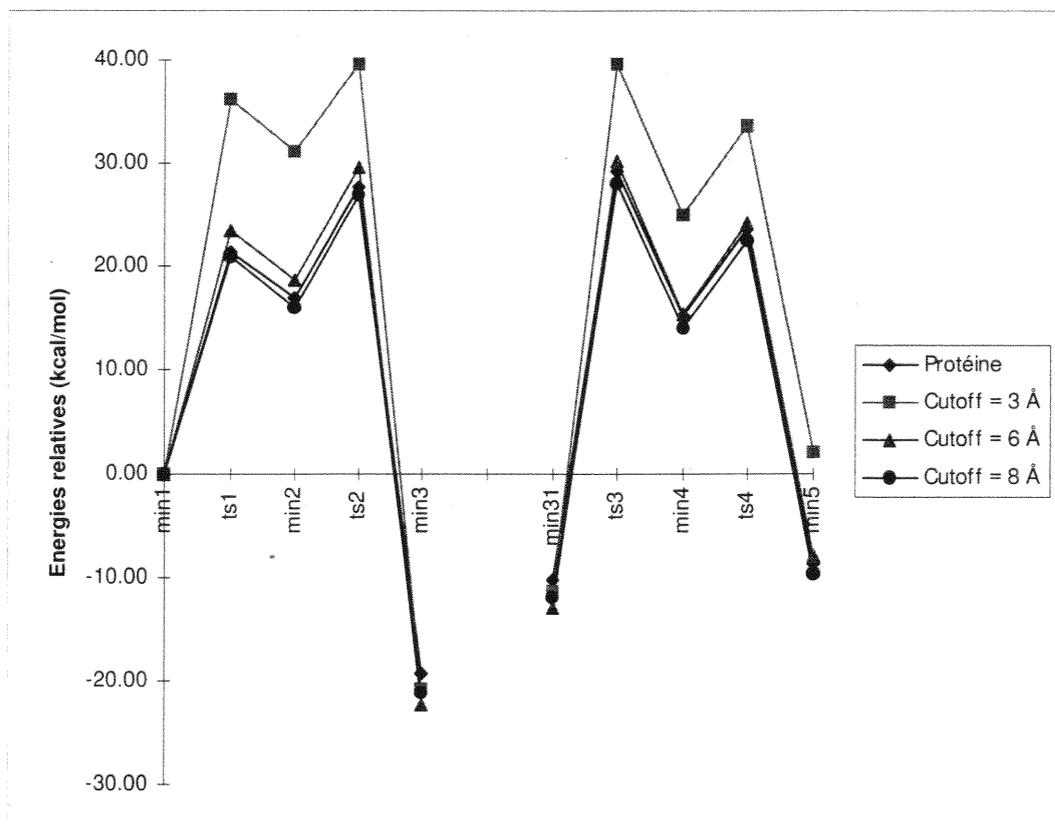


Figure 3.20 : Comparaison des profils énergétiques du mécanisme concerté dans la protéine (en petit losange) avec différents cutoff : de 3 Å courbes en petit carré, 6 Å courbes en triangle et 8 Å courbe en cercle.

Nos résultats prévoient que ces constatations restent valable tant que le cutoff utilisé reste inférieur à 6 Å. A partir de cette distance, le champ électrostatique moyen reproduit les mêmes résultats que celle utilisant une distribution de charges ponctuelles de l'environnement.

Les énergies relatives des différents intermédiaires calculées au niveau AM1 avec un champ électrostatique (figure 3.20) montrent que pour un cutoff de 3 Å, les réactifs et les TS dans la phase d'acylation sont déstabilisés en moyenne par 15 kcal/mol, tandis que dans la phase de désacylation ces intermédiaires sont déstabilisés par 10 kcal/mol. Nous avons corrélé la courbe de l'étude du chemin réactionnel de cette réaction d'hydrolyse avec celle calculée avec le potentiel moyen. La figure 3.20 comporte les profils énergétiques du mécanisme concerté dans la protéine avec et sans cutoff. On constate aisément dans cette figure que le profil énergétique calculé pour un cutoff supérieur ou égal à 6 Å s'apparente bien plus au profil énergétique sans cutoff qu'à celui calculé avec un cutoff de 3 Å. Un autre point important de cette étude est que, l'influence de l'environnement (i. e. partie classique du complexe thrombine/RS) dans la phase d'acylation n'est pas la même que celle de la phase de désacylation. Il est clair que cette première couche de charges située dans un rayon de 6 Å déstabilise l'étape de désacylation (figure 3.19) et que au delà de cette distance, le champ électrostatique moyen reproduit l'effet global de la partie lointaine de l'environnement protéique. Toujours dans le même souci de comparaison, nous avons recalculé le chemin réactionnel du point de vue énergétique, par un calcul ponctuel au niveau AM1, en annulant l'effet de ce champ moyen sur les fonctions de bases de la partie quantique.

L'examen de la figure 3.21 et du tableau 3.11 montre clairement que dans la phase de désacylation, les intermédiaires réactionnels sont déstabilisés par cette première couche de charges située dans un rayon de 3 Å, à l'exception des intermédiaires ACYL_2, IT_2 et EP qui sont au contraire légèrement stabilisés. Ceci vient confirmer la constatation que l'effet de l'environnement protéique sur les fonctions de bases de la partie quantique n'est pas le même tout au long du chemin réactionnel.

	Effets		
	Protéine	C _{1m} ≠ 0.0	C _{1^m} =0.0
MC	0.00	0.00	0.00
TS1	21.31	36.10	47.98
IT_1	16.96	31.10	40.75
TS2	27.79	39.68	52.80
ACYL_1	-19.22	-20.86	14.85
ACYL_2	-10.26	-11.42	1.89
TS3	29.18	39.71	31.37
IT_2	15.21	24.95	13.88
TS4	23.68	33.67	28.00
EP	-8.76	1.99	4.84

Tableau 3.11 : Energies relatives du chemin réactionnel dans la protéine comparées à celles recalculées avec et sans champ électrostatique moyen ($C_1^m=0.0$).

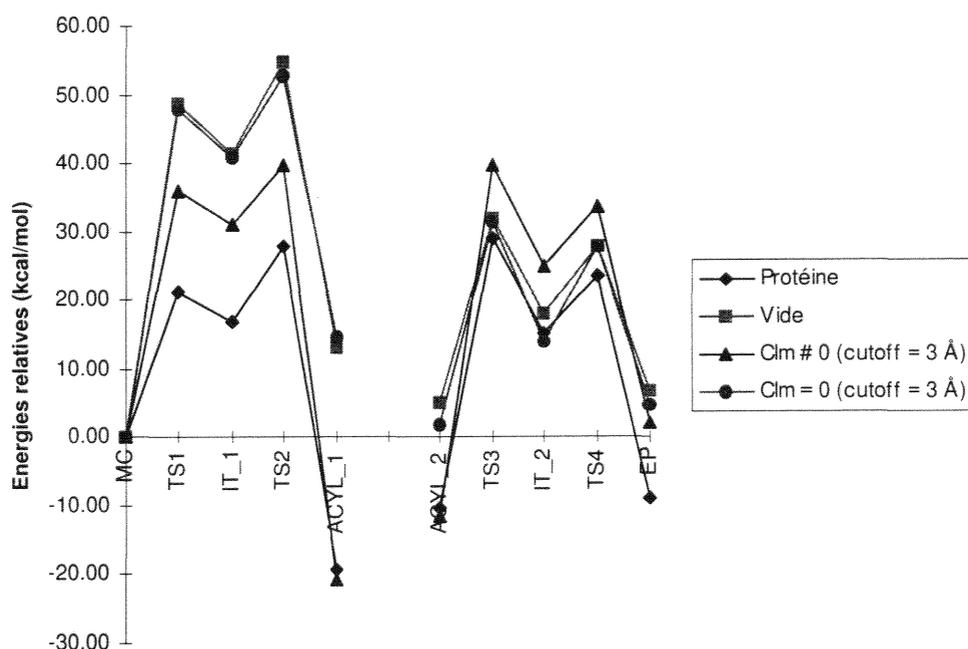


Figure 3.21 : Diagramme énergétique de la réaction d'hydrolyse de la liaison peptidique. Comparaison des profils énergétiques : Réaction dans la protéine (losange), dans le vide (petit carré), réaction avec un champ de réaction pour un cutoff de 3 Å (en triangle) et, en cercle, réaction sans champ moyen (uniquement l'environnement protéique au dessous de 3 Å).

Ces résultats montrent bien que pour cette phase de désacylation, des interactions spécifiques des groupes entourant le centre de la réaction jouent un rôle important dans la stabilité relative des différentes configurations du site actif.

Dans la plupart des cas, et surtout si le système étudié est chargé, les théoriciens utilisent des contre-ions pour neutraliser le système. Warshel et al⁵⁴ ont montré récemment que l'utilisation de ces contre-ions est nécessaire pour éviter l'artefact de la mauvaise représentation des effets à longue distances, et en plus n'influent pas sur les barrières d'activations. Il serait intéressant de tester l'effet de ces contre-ions et de quantifier les charges qui déstabilisent notre système. Une étude est en cours sur un autre système (complexe trypsine/BPTI), qui fait l'objet d'une thèse et pourrait nous donner des informations supplémentaires sur l'utilisation de ces contre-ions.

La question que l'on peut se poser, concernant la définition du potentiel électrostatique, est de savoir si cet ensemble de charges partielles qui représente l'environnement protéique est suffisant pour décrire à lui seul la polarisation du système ?

Il est vrai que même si ces charges sont paramétrées en fonction de la polarisation du système, elles ne suivent pas l'évolution du processus chimique qui passe du complexe de Michaelis (CM) à l'enzyme régénéré, et que leurs relaxations jouent un rôle sans doute important dans ce processus.

7. Effets électrostatiques dûs à la polarisation : Modèle à Cavité

Dans cette partie, nous présenterons quelques résultats qui illustrent l'utilisation du modèle à cavité (décrit dans le chapitre 2) dans l'étude de la réactivité chimique en solution. Nous avons examiné le rôle des interactions électrostatiques avec le solvant sur la réaction décrite précédemment, l'hydrolyse de la liaison peptidique du dipeptide RS du récepteur par la thrombine, en utilisant le modèle à cavité.

	électrostatique ($\epsilon = 2.5$)							Total
	l=6	l=5	l=4	l=3	l=2	l=1	l=0	
MC	-0.17235	-0.6639	-0.3971	-1.5477	-1.0568	-1.7744	-3.2840	-8.8964
TS1	-0.43691	-0.7692	-0.9775	-1.9138	-1.9414	-2.4599	-3.2096	-11.7080
IT_1	-0.43232	-0.7733	-0.9662	-1.9715	-1.9219	-2.5982	-3.2096	-11.8733
TS2	-0.40301	-0.8782	-0.9651	-2.0603	-1.9711	-2.4682	-3.2124	-11.9585
ACYL_1	-0.36175	-0.8378	-0.8918	-2.0515	-1.8674	-2.5960	-3.2346	-11.8411

Tableau 3.12 : Contributions multipolaires à l'énergie électrostatique de perturbation (kcal/mol) pour les réactifs et les TS de la phase d'acylation dans un ellipsoïde avec ($\epsilon = 2.5$).

	électrostatique ($\epsilon = 2.5$)							Total
	l=6	l=5	l=4	l=3	l=2	l=1	l=0	
ACYL_2	-0.3697	-1.0055	-0.8987	-2.2623	-1.8557	-2.4941	-3.2068	-12.0930
TS3	-0.3431	-0.7236	-0.8033	-1.7396	-1.6958	-2.2108	-3.2246	-10.7411
IT_2	-0.3004	-0.5863	-0.7213	-1.4504	-1.6111	-1.9523	-3.2222	-9.8444
TS4	-0.3110	-0.6800	-0.7357	-1.6263	-1.6235	-2.0824	-3.2221	-10.2811
EP	-0.3532	-0.9331	-0.7953	-2.1253	-1.6863	-2.5091	-3.2268	-11.6294

Tableau 3.13 : Contributions multipolaires à l'énergie électrostatique de perturbation (kcal/mol) pour les réactifs et les TS de la phase de désacylation dans un ellipsoïde avec ($\epsilon = 2.5$).

	électrostatique ($\epsilon = 10.0$)							Total
	l=6	l=5	l=4	l=3	l=2	l=1	l=0	
MC	-3.100	-1.1770	-.6948	-2.6515	-1.7500	-2.9654	-4.9260	-14.4753
TS1	-8.061	-1.3940	-1.7745	-3.3824	-3.3780	-4.2760	-4.8144	-19.8257
IT_1	-7.967	-1.4048	-1.7520	-3.4975	-3.3412	-4.5359	-4.8145	-20.1422
TS2	-7.462	-1.5895	-1.7612	-3.6236	-3.4478	-4.2505	-4.8186	-20.2376
ACYL_1	-6.677	-1.5090	-1.6221	-3.5745	-3.2634	-4.3833	-4.8519	-19.8721

Tableau 3.14 : Contributions multipolaires à l'énergie électrostatique de perturbation (kcal/mol) pour les réactifs et les TS de la phase d'acylation dans un ellipsoïde avec ($\epsilon = 10.0$).

	électrostatique ($\epsilon = 10.0$)							Total
	l=6	l=5	l=4	l=3	l=2	l=1	l=0	
ACYL_2	-.6887	-1.8289	-1.6493	-3.9888	-3.2490	-4.2593	-4.8102	-20.4745
TS3	-.6327	-1.3139	-1.4561	-3.0747	-2.9380	-3.8714	-4.8369	-18.1240
IT_2	-.5536	-1.0644	-1.3056	-2.5730	-2.7844	-3.4609	-4.8333	-16.5754
TS4	-.5724	-1.2290	-1.3304	-2.8633	-2.8055	-3.6396	-4.8332	-17.2737
EP	-.6475	-1.6750	-1.4334	-3.6946	-2.9117	-4.2647	-4.8402	-19.4675

Tableau 3.15 : Contributions multipolaires à l'énergie électrostatique de perturbation (kcal/mol) pour les réactifs et les TS de la phase de désacylation dans un ellipsoïde avec ($\epsilon = 10.0$).

		Etape d'acylation		
		Protéine	e = 2.5	e = 10.0
MC		26.88	32.05	32.38
	O γ	-0.26	-0.36	-0.36
	N δ	-0.29	-0.24	-0.24
	C	0.25	0.27	0.27
	H γ	0.24	0.24	0.24
	N-Ser42r	-0.36	-0.34	-0.35
TS1		-31.39	-38.17	38.63
	O γ	-0.56	-0.55	-0.55
	N δ	-0.18	-0.15	-0.16
	C	0.36	0.37	0.37
	H γ	0.34	0.35	0.36
	N-Ser42r	-0.37	-0.38	-0.38
IT_1		32.28	39.18	39.96
	O γ	-0.44	-0.44	-0.44
	N δ	-0.13	-0.11	-0.11
	C	0.35	0.35	0.35
	H γ	0.31	0.33	0.34
	N-Ser42r	-0.38	-0.39	-0.39
TS2		32.55	39.48	39.96
	O γ	-0.42	-0.42	-0.42
	N δ	-0.20	-0.19	-0.19
	C	0.32	0.32	0.32
	H γ	0.16	0.19	0.20
	N-Ser42r	-0.32	-0.33	-0.33
ACYL_1		35.07	41.51	41.95
	O γ	-0.26	-0.28	-0.28
	N δ	-0.24	-0.19	-0.18
	C	0.25	0.26	0.26
	H γ	0.15	0.16	0.16
	N-Ser42r	-0.34	-0.33	-0.33

Tableau 3.16 : Charges de Mulliken (en u.a) et moments dipolaires en gras (en Debye).

	Etape de désacylation		
	Protéine	e = 2.5	e = 10.0
	33.64	40.86	41.36
Og	0.26	-0.27	-0.27
Nd	-0.26	-0.20	-0.19
C	0.26	0.26	0.26
Hl	0.24	0.22	0.22
Ow	-0.46	-0.44	-0.44
	27.65	35.00	35.48
Og	-0.38	-0.41	-0.41
Nd	-0.23	-0.19	-0.19
C	0.35	0.34	0.34
Hl	0.36	0.37	0.37
Ow	-0.45	-0.43	-0.43
	23.89	30.71	31.14
Og	-0.43	-0.45	-0.45
Nd	0.26	0.27	0.27
C	0.35	0.35	0.35
Hl	0.33	0.35	0.35
Ow	-0.50	-0.49	-0.49
	26.38	33.33	33.77
Og	-0.43	-0.44	-0.44
Nd	0.24	0.25	0.25
C	0.35	0.36	0.36
Hl	0.36	0.37	0.37
Ow	-0.45	-0.44	-0.44
	34.15	40.70	41.12
Og	-0.37	-0.35	-0.35
Nd	-0.29	-0.24	-0.24
C	0.27	0.27	0.27

Tableau 3.17 : Charges de Mulliken (en u.a) et moments dipolaires en gras (en Debye).

Les géométries, entièrement optimisées dans la protéine au niveau AM1, ont été réévaluées par des calculs ponctuels en solution. La cavité, dans laquelle la molécule de soluté est placée, est une cavité de forme ellipsoïdale au centre de laquelle le potentiel électrostatique est développé en série multipolaire.

Des calculs théoriques de la constante diélectrique avec un modèle de champ de réaction⁵⁵ dans les protéines ont montré que, dans le cas de la trypsine la valeur de la constante dépend du site d'interaction. Ces valeurs varient de 2 à 10.0 pour le trou oxyanion de la trypsine⁵⁶. Nous avons donné au continuum deux valeurs de constantes diélectriques ($\epsilon = 2.5$ et $\epsilon = 10.0$) pour, d'une part, simuler les contributions qui ne sont pas représentées explicitement (i. e. dipôle permanent et induit de la protéine et leurs relaxations par le réarrangement de charges) et d'autre part représenter correctement les effets électrostatiques dans la protéine.

Pour réaliser nos calculs, nous avons utilisé un calcul SCF sans optimisation de géométrie ("single point" ou calcul ponctuel), à partir de la structure obtenue par minimisation de l'énergie dans la protéine.

Dans le cas d'une constante diélectrique $\epsilon = 10.0$, les composantes électrostatiques sont supérieures pour les réactifs et les TS. Les tableaux 3.13 et 3.14 montrent que, pour ces intermédiaires, la convergence du terme électrostatique est moins bonne que dans le cas d'un milieu avec une constante diélectrique $\epsilon = 2.5$. Comme nous pouvons le constater dans les tableaux 3.12, 3.13, 3.14, 3.15, les réactifs et les TS sont plus solvatés en milieu de forte constante diélectrique, et les moments dipolaires augmentent (tableau 3.16 et 3.17) avec la valeur de la constante diélectrique. On s'attend à ce que le champ de réaction stabilise les différents intermédiaires de réaction. L'analyse de la population électronique (i. e. charges de Mulliken), montre pour les atomes d'oxygène et de carbone, une augmentation sensible des charges correspondantes en fonction de la constante diélectrique. Les valeurs montrent une large polarisation de la molécule sous l'influence du champ de réaction qui peut être représentée grossièrement, par la polarisation des liaisons N δ H γ et CO γ . Il est important de noter que cette étude, réalisée en conservant les coordonnées atomiques optimisées dans la protéine, présente peu de réalité physique, et ne donne qu'une tendance.

	Effets du solvant		
	Protéine	e = 2.5	e = 10.0
MC	0.00	0.00	0.00
TS1	21.31	17.74	15.29
IT_1	16.96	12.58	9.98
TS2	27.79	23.33	20.72
ACYL_1	-19.22	-23.57	-25.95
ACYL_2	-10.26	-14.85	-17.56
TS3	29.18	25.93	24.20
IT_2	15.21	12.84	11.75
TS4	23.68	20.88	19.52
EP	-8.76	-12.90	-15.10

Tableau 3.18 : Energies relatives du chemin réactionnel dans la protéine et dans le solvant (kcal/mol)

En effet, la polarisation du site catalytique par le solvant (continuum superposé à des charges ponctuelles) doit vraisemblablement beaucoup plus stabiliser cette catalyse. En d'autres termes, les géométries d'équilibre présentent certainement des différences importantes avec celles utilisées ici. De même, l'énergie totale calculée en considérant un diélectrique, s'avère toujours inférieure à celle obtenue dans la protéine (tableau 3.18). Par conséquent, les résultats doivent être considérés d'un point de vue qualitatif, c'est-à-dire en ne s'intéressant qu'aux variations et tendances induites par la superposition de ces deux effets électrostatiques sur les intermédiaires de la réaction.

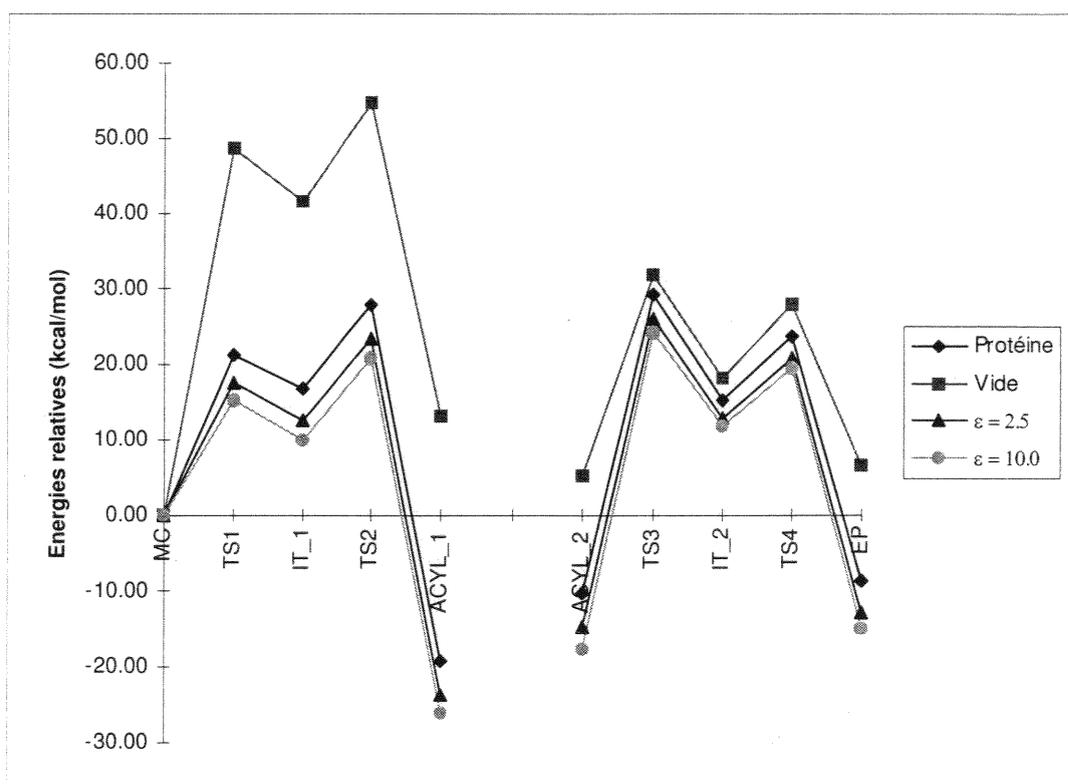
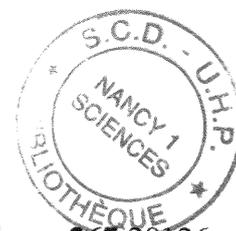


Figure 3.22 : Diagramme des énergies relatives du chemin réactionnel dans la protéine (losange), dans le vide (carré) et dans le solvant avec $\epsilon = 2.5$ (triangle) et $\epsilon = 10.0$ (cercle). Calcul réalisé au niveau AM1.

C'est l'étape d'acylation qui bénéficie le plus de cet abaissement énergétique devant la phase de désacylation. Il en résulte que les barrières énergétiques de tous les intermédiaires augmentent lorsque le pouvoir du solvant augmente (figure 3.22). Ce résultat est en accord avec celui de Dive et al⁵⁷.

BIBLIOGRAPHIE

1. L. Pauling, Chem. Eng. News, **263**, 294 (1946)
2. L. Pauling, Nature London, **161**, 707 (1948)
3. A. Warshel, M. Levitt, J. Mol. Biol, **103**, 227 (1976)
4. D.R. Storm, D. E. Koshland, J.A.C.S, **94**, 5805 (1972)
5. W. P. Jenks, Catalysis in Chemistry and enzymologie. (New York)
6. A. Warshel, Proc. Natl. Acad. Sci, USA **75**, 5250 (1978)
7. C.T. Esmon, Ann. Rev. Cell. Biol, **9**, 1 (1993)
8. T. K. H. Vu & coll, Cell, **64**, 1057 (1991)
9. T. J. Rydel & coll, J. Mol. Biol, **221**, 583 (1991)
10. W. Bode & coll, Protein Sci, **1**, 427 (1992)
11. M. T. Stubbs & coll, Eur. J. Biochem, **206**, 187 (1992)
12. M. Marquart, J. Walter, J. Dosehoger, W. Bode et R. Hubert, Acta Crystallo, Sect. B **39**, 480 (1983).
13. M.T. Milton Stubbs et W. Bode, Thrombosis research, **69**, 1 (1993)
14. A. Karshikov et W. Bode, Seminars in Thrombosis and Hemostasis, **19**, 334, (1993)
15. A. Nicolls et B. Honig, J. Comp. Chem, **12**, 435 (1991)
16. W. Bode, D. Turk et A. Karshikov, Protein Sci, **1**, 426, (1992)
17. T.K.H. Vu, D.T. Hung, W. Wheaton et S.R. Coughlin, Cell, **64**, 1057, (1991)
18. U.B Rasmussen, V. Vouret-Gravianni, S. Jallat, Y. Schesinger, A. Parivani, J.P. Lecoq, J. Pouysségur et E. Vanobberghen, E. Schilling, FEBS Lett, **288**, 123 (1991)
19. D. T. Hung, T.K.H. Vu, V Wheathon, I. F. Charo; N. A Nelken, N. Esmon et S.R. Coughlin, J. Clin, Invest, **89**, 444 (1992)
20. T.K.H. Vu, Wheaton, D.T Hung, I. F. Charo, S. R. Coughlin Nature, **674**, 3553, (1991)



21. D. Troyer, R. Padilla, T. Smith, J. Kreisberg and W. Glass, *J. Biol. Chem.*, **267**, 20126 (1992)
22. V. Vouret-Gravianni, E. Van Obberghien-Shilling, U. B. Rasmussen, A. Pavitani, J. P Lecocq et J. Pouysségur, *Mol. Biol. Cell*, **3**, 95 (1992)
23. S. Paris, I. Magnaldo, J. Bouysségur, *J. Biol. Chem.*, **263**, 11250 (1988)
24. I. I. Mathews, K. P. Padmanathan, V. Ganish et A. Tulinsky, *Biochemistry*, **33**, 3266 (1996)
25. W. Bode, I. Mayer, U. Baumann, R. Huber, S. R. Stone et J. Hofsteenge, *The EMBO J*, **8**, 3467 (1989)
26. Biosym Technologie, 9685 Sranton Road, San Diego CA 9212-2777 USA
27. E.Longo, F. M. L. Stomato, R. Ferreira et O. Tapia, *J. Theor. Biol.*, **112**, 783, (1985)
; S. Nakagawa, H. A. Yu, M. Karplus, H. Umeyama, *Proteins*, **16**, 1772 (1993)
28. D. W. Blow, J. J. Birktoff et B. S. Hartely, *Nature*, **221**, 337 (1969)
29. M. W. Hunkapiller, S. H. Smallcombe, D. R. Whitaker et J. H. Richards, *Biochemistry*, **12**, 4732 (1973)
30. R.E. Koeppe et R. M. Stroud, *Biochemistry*, **15**, 3450(1976)
31. J. L. Markely et M. A. Porubcan, *J. Mol. Biol.*, **102**, 487 (1976)
32. M. J. S. Dewar et D. M. Storch, *Proc. Natl. Acad. Sci USA*, **82**, 2225 (1985)
33. J. H. Wang *Science*, **161**, 328 (1968)
34. L. Polgar et M. L. Bender, *Proc. Natl. Acad. Sci USA*, **64**, 1335 (1969)
35. G. Robillard et R. G. Shulmann, *J. Mol. Biol.*, **71**, 507 (1972)
36. A. A. Kossiakoff et S. A. Spencer, *Nature*, **228**, 414 (1980)
37. P. A. Kollman et D. M. Hayes, *JACS*, **103**, 2955 (1981)
38. F. M. L. Stamato et O. Tapia, *Int. J. Quant. Chem.*, **33**, 187 (1988)
39. A. Warshel, *Curr. Opin. Strut. Biol.*, **2**, 230 (1992)
40. S. L. Aleksandrov et V. K. Antonov, *Mol. Biol (Moscou)*, **21**, 147 (1987)

41. S. Shneider et W. N. Lipscomb, *Proc. Natl. Acad. Sci USA*, **73**, 432 (1976)
42. V. Dagett, S. Shroder et P. Kollman, *JACS*, **113**, 8926 (1991)
43. G. Dive, D. Peeters, G. Levy et G. M. Ghoyesen, *J. Mol. Struct (Theochem)*, **107**, 117 (1984)
44. M. Komiyama et M. Bender, *Proc. Natl. Acad. Sci USA*, **76**, 557 (1979)
45. a. V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, G. Ferenzy, *J. Comp. Chem.*, **15**, 269 (1994)
b. A. D. Wallace, S. Dennis, J. Hofsteenge et S. R. Stone, *Biochemistry* **28**, 10079 (1989)
c. C. Edminston et K. Rudenberg, *Rev. Mod. Phys*, **35**, 457 (1963)
46. K. Taira, D. G. Gorenstein, *Bull. Chem. Soc. Jpn*, **60**, 3625 (1987)
47. A. Fresht *Enzyme Structure and Mechanism*, W. H. Freeman & Co New York (1985)
48. H. Hirohara, M. L. Bender, S. Stark, *Proc. Natl. Acad. Sci USA*, **71**, 1643 (1974)
49. H. Hirohara, M. Philipp et M. L. Bender, *Biochemistry*, **16**, 1573 (1977)
50. S. Sheiner, W. Lipscomb, *Proc. Natl. Acad. Sci USA*, **82**, 2225 (1985)
51. G. Dive et al *J. Mol. Struct*, **107**, 117 (1984)
52. R. Stone, A. Betz et J. Hofsteenge, *Biochemistry*, **30**, 9841 (1991)
53. L. Onsager *JACS*, **58**, 1486 (1936)
54. F. S. Lee et al, *J. Comput. Chem*, **2**, 161 (1992)
55. F. S. Lee, Z. T. Chu et Arie Warshel, *J. Comput. Chem*, **14**, 161 (1993)
56. G. King, F. S. Lee et A. Warshel, *J. Chem. Phys*, **95**, 4366 (1991)
57. G. Dive, D. Dehareng et D. Peeters, *Int. J. Quant. Chem*, **58**, 85 (1996)

Conclusion

L'objectif du travail présenté dans ce mémoire est l'étude du chemin réactionnel de l'activation du récepteur de la thrombine par cette dernière. Le site actif de cet enzyme ainsi que le substrat sont traités de façon quantique en tenant compte de l'effet du reste de l'enzyme qui, lui, est traité par la mécanique moléculaire. En étudiant ce système, nous nous sommes intéressés à certains aspects de la catalyse enzymatique, notamment aux effets électrostatiques et stériques, en nous appuyant sur la comparaison des résultats obtenus dans la protéine au moyen du formalisme LSCF avec ceux obtenus dans le vide. Cette comparaison a montré l'importante stabilisation qu'apporte la prise en compte de l'influence protéique.

Dans la première étape de la réaction d'hydrolyse, (le passage du complexe de Michaelis à l'intermédiaire tétraédrique), nous avons déterminé un mécanisme concerté (le saut du proton de la sérine vers l'histidine et simultanément l'attaque nucléophile de l'oxygène de la sérine sur l'atome de carbone du groupe carbonyle du substrat). Lors de l'étude de l'étape d'acylation, nous avons mis en évidence un phénomène important des réactions dans les protéases à sérine. Il s'agit de l'inversion de l'azote de la liaison peptidique du substrat avant l'approche de l'hydrogène (qui provoque la cassure de cette liaison). Cette inversion est une étape intrinsèque de ce type de réactions. En effet, le chemin le plus favorable de la formation de l'intermédiaire tétraédrique (intermédiaire réactionnel précédant l'acylation) aboutit précisément à une conformation de l'azote telle que la paire libre est à l'opposé de l'atome d'hydrogène qu'elle doit recevoir. Ainsi, l'étape de l'inversion de l'azote avant l'arrivée de l'hydrogène est une étape incontournable du processus de l'acylation dans les protéases à sérine.

L'examen des profils énergétiques indique que l'étape limitante de l'activation du récepteur (dipeptide RS) par la thrombine est la formation de l'intermédiaire tétraédrique II. Un autre résultat important de cette étude est la constatation que l'influence de l'environnement protéique dans la phase d'acylation n'est pas la même que celle dans la phase de désacylation. Ce comportement est probablement dû à des interactions spécifiques de la première couche de charges qui déstabilise ce processus de

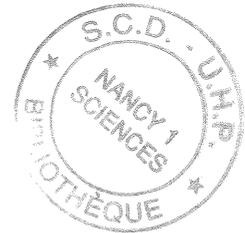
désacylation. Le formalisme LSCF développé au laboratoire nous a permis de comprendre les mécanismes d'interaction du complexe thrombine/récepteur, et par conséquent, nous permet d'étudier des molécules pharmacologiques actives qui pourraient soit bloquer le récepteur, soit bloquer la thrombine.

Le deuxième objectif de ce travail était l'amélioration méthodologique que nous avons apportée au logiciel GEOMOP (logiciel basé sur le formalisme LSCF). En effet, l'utilisation de ce logiciel sur des systèmes biologiques est très gourmand en temps de calcul, et la représentation de la partie lointaine de l'environnement protéique par un potentiel électrostatique moyen a permis de gagner un temps de calcul considérable (accélération d'un facteur moyen de 3), au prix d'une perte de précision infime. Il est clair que cette représentation électrostatique de la partie lointaine de l'environnement protéique ne doit pas se faire aux dépens de la prise en compte des interactions spécifiques. Ces dernières peuvent être particulièrement importantes pour des systèmes présentant à la surface des groupements fortement polarisables.

Un autre point important est l'étude de l'influence de la polarisation de l'environnement protéique sur la stabilité des intermédiaires de la réaction. En effet, le champ de réaction dû au continuum qui représente la contribution de la polarisation du milieu (environnement protéique) dans les interactions entre parties quantique et classique, stabilise les intermédiaires réactionnels de quelques kcal/mol. Il est important de noter, que par ce champ de réaction, nous avons montré que le fait de négliger la polarisation de la protéine intervenant au cours du processus constitue une restriction à la représentation du potentiel électrostatique moléculaire.

La prise en compte de l'environnement protéique et du solvant avec un modèle bien adapté est nécessaire pour rapprocher résultats théoriques et expérimentaux, mais ces résultats initiaux sont encourageants et informatifs. Dans le futur, nous espérons poursuivre ces études avec une méthode combinant le formalisme LSCF et la dynamique moléculaire, pour mieux prendre en compte les effets de l'environnement, et des modifications de structure pouvant apparaître dans ce dernier.

Conclusion



Nom: MOKRANE

Prénom: Ahmed

DOCTORAT de l'UNIVERSITE HENRI POINCARÉ, NANCY-I
en CHIMIE INFORMATIQUE & THEORIQUE

VU, APPROUVÉ ET PERMIS D'IMPRIMER

Nancy, le 24 NOV 1997 n° 115

Le Président de l'Université



L'objectif de ce travail est l'étude du chemin réactionnel de l'activation du récepteur d'un enzyme par ce dernier dans le cadre de la méthode hybride mécanique quantique/mécanique moléculaire LSCF (Local Self Consistent Field).

Pour ce faire, nous avons apporté une amélioration méthodologique au logiciel GEOMOP (logiciel basé sur le formalisme LSCF), permettant de traiter les interactions a longue distance en gagnant un temps de calcul considérable (accélération d'un facteur moyen de 3), au prix d'une perte de précision infime.

Nous avons aussi introduit une méthodologie permettant une étude de l'influence de la polarisation de l'environnement protéique sur la stabilité des intermédiaires de la réaction. En effet, le champ de réaction dû au continuum qui représente la contribution de la polarisation du milieu (environnement protéique) dans les interactions entre partie quantique et classique, stabilise les intermédiaires réactionnels de quelques kcal/mol.

L'application au complexe thrombine/récepteur de la thrombine nous a permis de comprendre la base pour étudier des molécules pharmacologiques actives qui pourraient soit bloquer le récepteur, soit bloquer la thrombine.

Mots clefs : méthode mixte mécanique quantique/mécanique moléculaire (LSCF), structures de transitions, potentiel électrostatique moyen, champ de réaction, thrombine, récepteur