



Modélisation de la réaction d'alkylation du motif zinc-thiolate

Delphine Picot

► To cite this version:

Delphine Picot. Modélisation de la réaction d'alkylation du motif zinc-thiolate. Chimie. Ecole Polytechnique X, 2008. Français. NNT : . pastel-00004269

HAL Id: pastel-00004269

<https://pastel.hal.science/pastel-00004269>

Submitted on 23 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THESE

Présentée pour obtenir le grade de
DOCTEUR DE L'ÉCOLE POLYTECHNIQUE

Spécialité : Chimie

Par

Delphine PICOT

**Modélisation de la réaction d'alkylation
du motif zinc-thiolate**

Directeur de thèse : Gilles Ohanessian

Soutenue le 29 septembre 2008 devant la commission d'Examen composée de :

M. Philippe Maître

Mme Annick Dejaegere

M. Olivier Parisel

M. Gilles Ohanessian

M. Gilles Frison

Président

Rapporteur

Rapporteur

Directeur de thèse

Co-directeur de thèse

Remerciements

Un grand merci à **Gilles Ohanessian**, directeur du laboratoire des mécanismes réactionnels et directeur de ma thèse, pour m'avoir accueillie au sein de son laboratoire et de son équipe, ainsi que pour sa disponibilité et sa patience.

Ces travaux de thèse ont été encadrés par **Gilles Frison**. Je tiens tout particulièrement à le remercier pour son aide précieuse et ses conseils avisés.

Merci aussi à **Carine Clavaguéra** pour l'aide informatique et sur Turbomole, ainsi que les discussions scientifiques et sympathiques.

J'adresse mes plus sincères remerciements à Madame **Annick Dejaegere** et Monsieur **Olivier Parisel** d'avoir accepté d'être les rapporteurs de ce travail. Merci aussi à Monsieur **Philippe Maître** d'avoir accepté de participer à son évaluation.

Merci à **Thérèse**, l'irremplaçable secrétaire.

Je tiens à remercier **les étudiants et le personnel du DCMR** pour le soutien au quotidien et les bons moments partagés.

Merci à **Magalie, Guylaine, Angèle, Emilie, Marie, David...**

Je n'oublierai pas de remercier également,

Mes parents,

Pour m'avoir toujours soutenue pendant toutes ces années d'études.

Anne-Laure, Eric, Zoé et Léa,

Pour les week-ends dépaysants passés avec vous.

Djipe ♥

Mes amis du roller,

Pour les week-ends de défoulement à roulettes dans Paris 😊

Eli, Nadia, Elsa, Philou, Vincent et tous les autres ovsiens

Pour tous les moments mémorables partagés ensemble.

SOMMAIRE

Introduction générale.....	-13-
<u>Chapitre I</u> : INTRODUCTION.....	-19-
I. Le métal Zinc.....	-21-
A. L'élément chimique	-21-
B. Le zinc en biologie	-22-
C. Le zinc et les autres métaux	-22-
II. Le soufre.....	-24-
A. La chimie du soufre	-24-
B. Les différentes formes du soufre	-24-
III. Le motif zinc-thiolate dans les protéines.....	-25-
A. Rôle des métaux et du zinc dans les protéines	-25-
B. Structure du motif zinc-thiolate dans les protéines	-26-
C. Réactivité du motif zinc-thiolate dans les protéines	-28-
C.1. Réactivité des sites catalytiques	-28-
C.2. Réactivité des sites structuraux	-29-
C.2.a. Mécanisme de l'alkylation	-31-
C.2.b. Sélectivité de l'alkylation	-34-
Bibliographie.....	-37-
<u>Chapitre II</u> : METHODES DE CALCULS.....	-41-
I. Généralités.....	-43-
A. Introduction	-43-
B. Stratégie	-43-
II. Bases d'orbitales atomiques.....	-44-
A. Les bases de gaussiennes	-44-
A.1. Les bases tous électrons	-45-
A.1.a. Les bases initiales	-45-

A.1.b.	<i>Les bases étendues</i>	-45-
A.2.	Les potentiels effectifs	-46-
A.2.a.	<i>La base initiale</i>	-46-
A.2.b.	<i>La base étendue</i>	-46-
B.	Les bases utilisées par TURBOMOLE	-46-
B.1.	Les bases initiales	-47-
B.2.	Les bases étendues	-47-
III.	Méthodes de chimie quantique.....	-48-
A.	Introduction	-48-
B.	Méthode Hartree-Fock (HF)	-48-
C.	Méthode Post-HF: MP2	-49-
D.	Les méthodes de la DFT	-51-
D.1.	L'approximation de la Densité Locale (LDA)	-51-
D.2.	L'approximation du Gradient Généralisé (GGA)	-52-
D.3.	Les fonctionnelles hybrides	-52-
IV.	Les approximations de la Résolution de l'Identité.....	-53-
A.	Principe de l'approximation RI	-53-
B.	Principe de l'approximation MARI-J	-54-
V.	Les méthodes de solvation.....	-55-
A.	Introduction	-55-
B.	Généralités thermodynamiques	-56-
C.	Les modèles de Born et d'Onsager	-57-
D.	La méthode PCM	-57-
E.	Les approches COSMO et C-PCM	-59-
E.1.	La théorie COSMO	-59-
E.2.	La théorie C-PCM	-60-
F.	Calculs effectués en solvant	-60-
F.1.	Avec GAUSSIAN	-60-
F.2.	Avec TURBOMOLE	-60-
VI.	Les grandeurs thermodynamiques.....	-61-
A.	En phase gazeuse	-61-
A.1.	Correction au Point-Zero	-61-
A.2.	Variation d'énergie interne à T	-61-
A.3.	Variation d'enthalpie à T	-62-
A.4.	Variation d'énergie libre à T	-62-
B.	En phase aqueuse	-62-

Bibliographie.....	-63-
Annexes.....	-65-

Chapitre III : REACTIVITE ET MECANISMES D'ALKYLATION DES COMPLEXES BIOMIMETIQUES A ZINC..... -69-

I. Introduction.....	-71-
II. Résultats.....	-72-
III. Publication.....	-73-
A. Introduction	-74-
B. Chemical and computational models	-79-
B.1. Chemical models	-79-
B.2. Level of computation	-80-
B.3. Evaluation of chemical model and level of computation	-81-
C. Results	-82-
C.1. Structure of complexes 1-4	-82-
C.2. Reaction mechanisms of complex 1	-84-
C.3. Solvation effects on complex 1	-88-
C.4. Influence of zinc ligands	-89-
C.5. Influence of the hydrogen bond	-90-
D. Discussion	-92-
D.1. On the reactivity of zinc site	-92-
D.2. On the mechanism of zinc site	-94-
E. Conclusion	-96-
F. References	-97-

Bibliographie.....	-101-
--------------------	-------

Chapitre IV : ETUDE DE LA DISSOCIATION DE LIAISONS ZINC-THIOLATE..... -103-

I. Introduction.....	-105-
A. Observation expérimentale des structures zinc-soufre	-105-
A.1. Nombres de coordination du zinc	-105-

A.2.	Structures du zinc à 4 ligands soufrés	-106-
A.3.	Structures du zinc à 3 ligands soufrés	-110-
A.4.	Dimères du zinc	-112-
B.	Méthodes théoriques utilisées dans la littérature	-114-
II.	Mise en évidence de la dissociation de $[\text{Zn}(\text{SPh})_4]^{2-}$.....	-115-
A.	$[\text{Zn}(\text{SPh})_4]^{2-}$ dans la littérature	-115-
B.	Calculs préliminaires sur $[\text{Zn}(\text{SPh})_4]^{2-}$	-116-
C.	Problématiques	-118-
III.	Structures de $[\text{Zn}(\text{SMe})_4]^{2-}$ et $[\text{Zn}(\text{SPh})_4]^{2-}$.....	-119-
A.	Influence de la base	-120-
B.	Influence de la méthode	-120-
C.	Influence de l'environnement	-124-
D.	Conclusions	-124-
IV.	Etude de la dissociation.....	-125-
A.	Influence de la géométrie	-125-
B.	Influence de la solvation	-126-
C.	Barrière d'activation de la dissociation	-129-
C.1.	Complexe $[\text{Zn}(\text{SMe})_4]^{2-}$	-129-
C.2.	Complexe $[\text{Zn}(\text{SPh})_4]^{2-}$	-131-
V.	Discussion sur la stabilité des complexes à zinc.....	-133-
A.	Stabilité en phase gazeuse	-133-
B.	Stabilité en phase aqueuse	-136-
C.	Conséquences pour la stabilité des sites actifs ZnCys_4 des protéines	-138-
	Bibliographie.....	-140-

Chapitre V : INDICES DE NUCLEOPHILIE DU MOTIF ZINC-THIOLATE..... -143-

I.	Introduction.....	-145-
A.	Nucléophilie et basicité	-146-
B.	Les échelles de réactivité	-146-
C.	Les charges	-149-
C.1.	Charges de Mulliken	-149-

C.2.	Charges NBO	-149-
C.3.	Charges de Hirshfeld	-150-
D.	Les indices de réactivité issus des méthodes DFT	-150-
D.1.	Coefficient de Fukui	-150-
D.1.a.	Définition	-150-
D.1.b.	Calcul du coefficient de Fukui	-151-
D.1.c.	Limites du coefficient de Fukui	-152-
D.2.	Dureté, mollesse et indices dérivés	-152-
D.3.	Indice d'électrophilie	-153-
E.	Application aux composés soufrés	-154-
II.	Calculs préliminaires sur les complexes biomimétiques.....	-155-
A.	Charges	-155-
B.	Coefficient de Fukui et mollesse locale	-156-
C.	Affinité protonique	-158-
III.	Nucléophilie des thiolates liés au zinc.....	-159-
A.	Introduction	-159-
B.	Résultats	-160-
C.	Publication	-161-
C.1.	Introduction	-162-
C.2.	Computational methods	-163-
C.3.	Results and discussion	-164-
C.3.a.	Reactivity of zinc-bound phenylthiolate	-165-
C.3.b.	Reactivity of zinc-bound alkylthiolate	-168-
C.3.c.	Influence of hydrogen bonding	-169-
C.3.d.	Nucleophilicity index	-170-
C.4.	Conclusion	-172-
C.5.	References	-173-
IV.	Comparaison avec la réactivité des thiols et des thiolates.....	-176-
	Bibliographie.....	-179-
	Annexe.....	-182-
	 Chapitre VI : MODELISATION D'UN SITE ACTIF A ZINC.....	 -187-
I.	Introduction.....	-189-

A.	Importance de l'environnement du zinc dans les sites actifs	-189-
B.	Les doigts de zinc de la nucléocapside p7 du virus VIH	-190-
B.1.	La famille des doigts de zinc	-190-
B.2.	La protéine nucléocapside p7	-191-
B.2.a.	Généralités	-191-
B.2.b.	Structure	-192-
B.2.c.	Réactivité	-193-
II.	Evaluation du niveau de calcul.....	-196-
A.	Faisabilité des calculs	-196-
B.	Structure tridimensionnelle	-198-
C.	Précision des résultats	-202-
III.	Calcul de l'affinité protonique d'un fragment de protéine.....	-205-
A.	Référence	-205-
B.	Influence du squelette peptidique	-206-
C.	Influence de la chaîne latérale de Lys41	-210-
D.	Influence de la solvation sur la structure du doigt de zinc	-215-
E.	Influence de l'ensemble de l'environnement sur la réactivité du site	-217-
IV.	Conclusion.....	-220-
	Bibliographie.....	-222-
	Conclusion et perspectives.....	-225-

Introduction générale

Le zinc est, après le fer, le métal de transition le plus abondant en biologie. Un corps humain contient en moyenne 2.3 g de zinc. Ce métal est en effet indispensable à la croissance et au développement et présente des effets thérapeutiques et préventifs envers des maladies infectieuses comme la pneumonie et la malaria. L'influence du zinc vient du rôle essentiel qu'il joue dans les protéines. En effet, environ 300 enzymes sont répertoriées dans lesquelles le zinc possède des fonctions catalytiques et/ou structurales.

Le site actif de la plupart de ces métalloenzymes est composé d'un zinc coordonné de façon tétraédrique et lié au squelette protéique par des résidus comme, par exemple, l'histidine et la cystéine. Le zinc est présent à la fois dans des sites catalytiques possédant une molécule d'eau dans la sphère de coordination du métal, et dans des sites structuraux ayant une fonction de stabilisation de la structure protéique. Or, une nouvelle classe de sites actifs a récemment émergé, qui malgré l'absence d'une molécule d'eau, présente une réactivité au niveau d'une liaison zinc-soufre. Le thiolate lié au zinc subit alors une alkylation et le métal a pour rôle d'activer le thiolate pour l'attaque nucléophile (Figure ci-dessous).

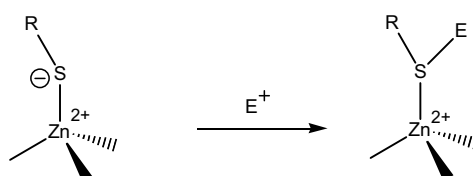


Figure 1 : Alkylation d'un thiolate lié au zinc

L'objectif de cette thèse est la compréhension du mécanisme d'action et de la réactivité d'alkylation du motif zinc-thiolate dans les sites des métalloenzymes. A cette fin, nous étudierons tout d'abord l'alkylation sur des complexes biomimétiques du zinc, choisis comme modèles des métalloenzymes et synthétisés par de nombreux groupes dans la littérature. Une fois cette réactivité bien comprise pour ces modèles chimiques, nous chercherons à extrapoler ces résultats à l'échelle protéique.

Les méthodes quantiques seront utilisées pour l'étude de différents types de composés, depuis de petits systèmes modèles en phase gazeuse contenant quelques dizaines d'atomes jusqu'à des parties entières de protéines contenant plusieurs centaines d'atomes. Pour certains de ces systèmes, comme nous le montrerons, le solvant joue un rôle majeur. Cette diversité de taille et d'environnement nécessite des modélisations adaptées. Les méthodes utilisées, qui sont implémentées dans les logiciels GAUSSIAN03 et TURBOMOLE, seront présentées dans le chapitre II et feront l'objet d'une validation à chaque étape de notre étude.

La première étape de notre recherche (Chapitre III) consistera à déterminer le mécanisme d'alkylation des groupes thiolates liés au zinc dans des complexes biomimétiques. Pour cela, nous avons construit, sur la base du squelette moléculaire d'un complexe biomimétique, plusieurs modèles

représentant la diversité chimique observée dans les protéines. Trois mécanismes ont été considérés : (i) un mécanisme de type S_N2 dissociatif, dans lequel le thiolate lié au zinc se dissocie du métal avant alkylation ; (ii) un mécanisme de type S_N2 associatif, où l'espèce nucléophile est le thiolate lié au zinc ; (iii) un mécanisme de type métathèse σ avec passage par un état de transition incluant un métallacycle à quatre centres. L'étude de l'influence de l'environnement du zinc, notamment la nature de ses autres ligands ou la participation du groupe thiolate dans une liaison hydrogène, sera également réalisée. Nous dégagerons de ces travaux l'importance de la dissociation de la liaison zinc-soufre et de la nucléophilie de l'atome de soufre pour la compréhension de la réactivité de ces espèces.

La dissociation de la liaison zinc-soufre dans des complexes biomimétiques sera appréhendée dans le chapitre IV. Au vu des résultats de la partie précédente, une dissociation aisée du thiolate lié au zinc dans les complexes dianioniques est attendue. Ce phénomène sera plus précisément considéré pour les composés $[Zn(SPh)_4]^{2-}$ et $[Zn(SMe)_4]^{2-}$ et les calculs d'énergies de dissociation seront confrontés aux observations expérimentales disponibles. L'effet stérique et électronique des ligands, ainsi que l'influence de l'environnement (solvant), feront partie des étapes essentielles dans notre compréhension de la dissociation.

Le chapitre V sera consacré à la détermination d'un indice de nucléophilie du motif zinc-thiolate. En effet, le calcul de la barrière énergétique de l'alkylation, tel qu'il a été effectué pour mesurer la nucléophilie des complexes biomimétiques au chapitre III, n'est pas envisageable pour les modèles protéiques de grande taille que nous souhaitons étudier. De plus, nous ne disposons pas à ce jour d'un indice nous permettant de déterminer la nucléophilie relative de différents groupes thiolates. Un critère de réactivité rapide et simple sera recherché pour une large famille de complexes biomimétiques présentant l'avantage, pour certains, de posséder plusieurs thiolates susceptibles d'être alkylés. Ceci permettra non seulement de comparer la réactivité relative de différents sites, mais aussi la réactivité relative de différents ligands thiolates présents dans un même site actif. La barrière énergétique requise pour l'alkylation du soufre nous servira de référence pour rechercher la meilleure corrélation avec les différents indices testés (charges atomiques, coefficients de Fukui, énergie de la HOMO, affinité protonique).

Enfin, l'indice de nucléophilie sélectionné dans la partie précédente nous permettra d'étudier la réactivité d'un site protéique dans le chapitre VI. Nous effectuerons la modélisation progressive d'un fragment protéique, en indiquant les problèmes rencontrés lors de la simplification d'un tel système, flexible et de grande taille. Le site sélectionné, celui de la nucléocapside p7 du virus VIH, permettra d'étudier la sélectivité envers un thiolate par rapport aux autres thiolates d'un même site actif. La réactivité nucléophile des trois cystéines du doigt de zinc du côté C-terminus sera déterminée à partir du calcul de leurs affinités protoniques et comparée aux données de la littérature. L'influence de l'environnement sur la nucléophilie des thiolates sera appréhendée avec la prise en compte des

liaisons hydrogène et l'inclusion d'une charge, portée par un bras ammonium d'une lysine, à proximité de la cystéine réactive.

Chapitre I

Introduction

Le but de cette thèse est l'étude de la réactivité du motif zinc-soufre. Elle s'inscrit dans une thématique du laboratoire portant sur l'influence de l'environnement (squelette peptidique, seconde sphère de coordination, solvation, ...) sur la structure et l'activité biologique de protéines à zinc. La démarche consiste à étudier des composés chimiques modèles (complexes biomimétiques et structures simplifiées de sites actifs) afin de mieux appréhender la chimie des protéines à zinc.

I. Le métal Zinc

A. L'élément chimique

Le zinc est un métal de transition dont le numéro atomique est $Z=30$ et la configuration électronique est $[\text{Ar}] 3d^{10} 4s^2$. Comme les autres éléments de transition, il possède une forte densité (7133 kg/m^3), une température de fusion (692.7 K) et de vaporisation (1180 K) élevées. Le zinc est un métal blanc, légèrement bleuâtre et brillant. Il est cassant à froid mais devient malléable et ductile entre 100 et 150°C . Dans un environnement humide, il se couvre d'une mince couche d'hydrocarbonate qui le patine et le protège contre l'oxydation. Son unique état d'oxydation est $+II$ (Handbook of Chemistry 1990, Merck Index 1996).

Figure I.1 : Le zinc dans le tableau périodique

Le zinc bénéficie d'une réputation de "métal écologique". Il n'est pas considéré comme toxique. Toutefois, lorsque de l'oxyde de zinc à peine formé est inhalé, des désordres physiologiques peuvent être observés. De ce fait, le taux maximum d'oxyde de zinc recommandé est de 5 mg/m^3 d'air dans les lieux de son traitement.

Les travaux sur le zinc sont en nombre plus limité que les autres métaux, du fait d'avoir été longtemps considéré comme « un élément ennuyeux » (Vahrenkamp 2007). En effet, le zinc n'a pas de couleur ou de magnétisme et sa couche d'électrons d pleine est finalement son unique caractéristique chimique intéressante. De plus, les complexes à zinc sont très réactifs et versatiles. Avec des stabilités thermodynamiques généralement faibles, ces propriétés en font un élément difficile à travailler en laboratoire.

Alors que ses propriétés chimiques ne semblent pas exploitables au premier abord, le zinc présente une importante biodisponibilité. Il est second dans l'ordre d'abondance des systèmes biologiques derrière le fer (Christianson 1991) et son abondance naturelle est de 2.3 g pour une personne

moyenne (Sun 1999). Le zinc trouve alors son importance en biologie, et plus particulièrement dans l'étude des métalloenzymes.

B. Le zinc en biologie

Le zinc est considéré comme un élément essentiel des systèmes biologiques, du fait de son abondance naturelle importante dans les protéines (environ 300) (Vallee 1990) et des nombreux processus enzymatiques auxquels il participe.

Les propriétés biochimiques du zinc les plus remarquables sont les suivantes (Parkin 2000 et 2004, Vahrenkamp 2007) :

- L'ion divalent zinc est exceptionnellement stable envers l'oxydation et la réduction. Il ne participe donc pas aux réactions redox, contrairement aux métaux manganèse, fer et cuivre.
- Puisque sa couche d est remplie, les complexes à zinc ne sont pas sujets aux effets de champs des ligands. Leur nombre de coordination et leur géométrie sont donc seulement fonction de la taille et de la charge des ligands.
- La flexibilité dans la géométrie de coordination facilite l'échange de ligands et la capacité du zinc à effectuer un cycle catalytique.
- En solution, le cation métallique est solvaté par six molécules d'eau, alors que dans les enzymes, le zinc montre une grande préférence pour la coordination tétraédrique (Vallee 1993, Patel 2007). L'acidité de Lewis du métal est augmentée lorsqu'il est chélaté aux résidus protéiques, plus polarisables que les molécules d'eau, et le transfert de charge des résidus au zinc est alors plus important qu'en solution (Dudev 2003). Le zinc peut exalter la réactivité de certains ligands (auxquels il est complexé) qui sont alors capables de réagir avec le substrat. C'est le cas des molécules d'eau dont l'acidité de Brønsted augmente de façon significative par coordination à l'ion métallique. Seul l'ion divalent cuivre est un meilleur acide de Lewis.
- Du fait de son petit rayon ionique (0.74 Å) et de son électronégativité relativement élevée ($\chi=1.66$), l'ion divalent zinc se trouve à la frontière entre les cations durs et mous, si bien qu'il peut accommoder des ligands de nature azotés, oxygénés, soufrés et halogénés. Bien qu'il apparaisse aussi coordonné aux acides aspartiques et glutamiques (Dudev 2003), le zinc préfère, dans les enzymes, les ligands plus mous comme la cystéine et l'histidine auxquelles il se lie respectivement par leur soufre ou leur azote.
- Les anions OH^- , OR^- et SR^- gardent un caractère nucléophilie important lorsqu'ils sont liés au zinc. Seuls Mn^{II} , Fe^{III} et Cu^{II} sont "meilleurs" dans ce cas.

C. Le zinc et les autres métaux

Les métaux sont très abondants dans les enzymes puisqu'ils sont présents sous forme ionique dans environ la moitié des protéines (Thomson 1992). Parmi les plus courants, on trouve également les ions fer et magnésium, qui avec le calcium sont les plus étudiés dans la littérature.

Le magnésium, du fait de son rayon atomique proche de celui du zinc, et le calcium ont souvent été considérés comme des analogues. Cependant, il s'agit de métaux durs qui préféreront, contrairement au zinc, les ligands oxygénés aux ligands soufrés. De plus, le magnésium et le calcium possèdent des nombres de coordination variant de 6 à 8, ce qui limite leur comparaison au zinc (Figure I.2) (Dudev 2003).

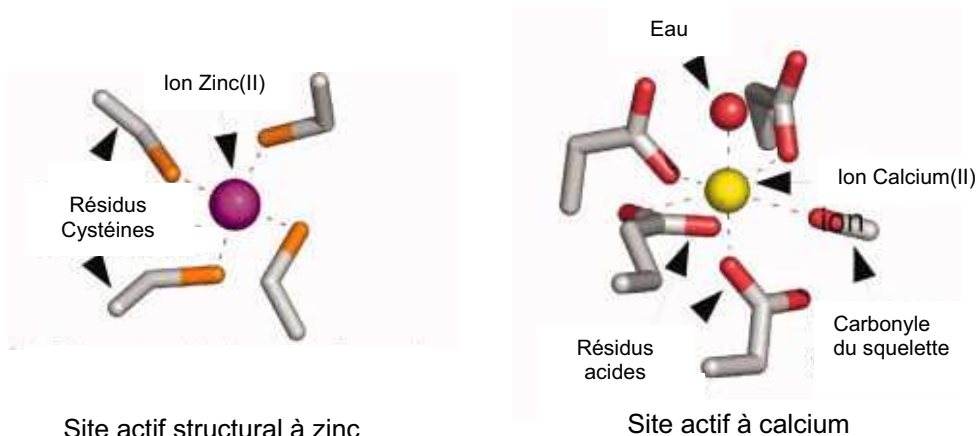


Figure I.2 : Exemples de structures de sites actifs métalliques (Torrance 2007)

Le cuivre, quant à lui, partage avec le zinc la même préférence pour les ligands soufrés. De plus, le cation Cu^{2+} a une capacité prononcée à accepter le transfert de charge des ligands anioniques et pour cette raison, assiste efficacement la déprotonation des thiols cystéiniques des sites métalloactifs, similairement à Zn^{2+} (Dudev 2002). L'étude de complexes hydroxydes du zinc et du cuivre a cependant montré que le cuivre avait une tendance à être coordonné de façon bidentate, ce qui mène à une réactivité différente de ses complexes par rapport à ceux du zinc (Parkin 2004, Hikichi 1994).

Des efforts considérables ont été faits dans la substitution du zinc par d'autres métaux dans les protéines dites « silencieuses » en spectroscopie. En effet, le noyau diamagnétique du zinc (d^{10}) ne permet pas d'effectuer des mesures spectroscopiques (ni ESR, ni ^{67}Zn -NMR) sur les métalloenzymes correspondantes. Une substitution par des ions métalliques divalents tels que le cobalt et le cadmium, permet d'obtenir des informations structurales sur le site actif et sur la nature des intermédiaires (Parkin 2004).

Le cobalt est typiquement utilisé car il a des propriétés électroniques spectroscopiques distinctes et qu'il a une préférence pour former, comme le zinc, des complexes tétraédriques (Horrocks 1980, Bertini 1983, Maret 1993).

Le cadmium permet l'utilisation de la spectroscopie ^{113}Cd -NMR. Cependant, si le cadmium prend correctement la place du zinc dans le site actif, son comportement, en termes de coordination et stabilité, est totalement différent. Le cadmium, qui possède une acidité de Lewis plus petite que le zinc, rend également le thiolate plus nucléophile (Chiou 2000 et 2003). Par ailleurs, l'utilisation du cadmium a également été réduite du fait de sa toxicité.

II. Le soufre

A. La chimie du soufre

Dans la classification périodique des éléments, le soufre ^{16}S se situe juste en dessous de l'oxygène. La fonction thiol (-SH) possède donc une structure électronique et une réactivité souvent comparées à celles de la fonction hydroxyle (-OH).

Ainsi, les groupes thiols présentent des propriétés acides et nucléophiles importantes, similairement aux fonctions alcools (Carey 2006). Malgré la plus importante électronégativité de l'oxygène ($\chi=3.44$) par rapport au soufre ($\chi=2.58$), les thiols sont des composés plus acides que leurs homologues hydroxyles (pK_a de l'ordre de 11 pour les thiols alors qu'il est supérieur à 15 pour les alcools).

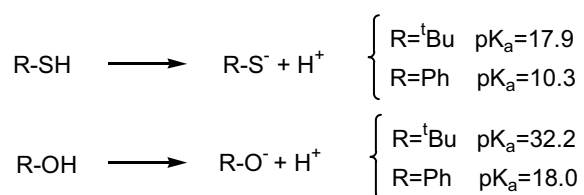


Schéma I.1 (Bordwell 1982)

De plus, les thiols sont de meilleures bases de Lewis. Cette importante réactivité des thiols, acidité de Brønsted et basicité de Lewis, en fait un partenaire chimique privilégié pour le zinc. L'interaction entre les groupes thiols et le métal se retrouve d'ailleurs fréquemment dans les systèmes biologiques.

B. Les différentes formes du soufre

Le soufre est un élément essentiel pour les êtres vivants. Il est nécessaire dans la synthèse des acides aminés cystéine et méthionine, mais aussi homocystéine et taurine (Figure I.3).

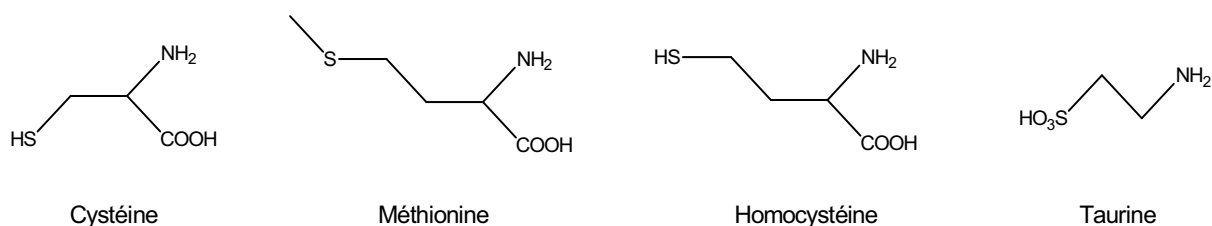


Figure I.3

Le groupe thiol de la cystéine est très nucléophile dans les conditions physiologiques et très facilement oxydable. A cause de cette réactivité élevée, il possède de nombreuses fonctions biologiques et notamment structurales : l'oxydation de deux cystéines est à l'origine des ponts disulfures qui jouent un rôle très important dans l'assemblage et la structure tridimensionnelle des protéines (Figure I.4).

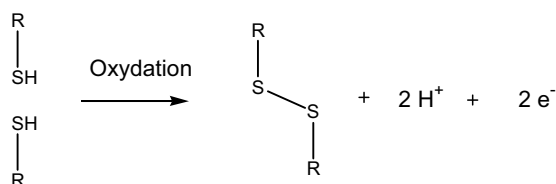


Figure I.4

Le groupe thiol présente une affinité élevée pour les métaux et de ce fait, les cystéines protéiques peuvent se coordonner très fortement à un métal. Dans cette étude, nous avons plus particulièrement considéré le soufre sous forme de cystéinate coordonnée au zinc (Figure I.5). Le pK_a de la fonction thiol de la cystéine étant de 8.0, celle-ci est donc protonée en milieu physiologique à $\text{pH}=7$. Lorsqu'elle est liée au zinc, du fait de la forte acidité de Lewis de ce dernier, le pK_a de la fonction thiol de la cystéine est abaissé, conduisant à la formation du motif zinc-thiolate (Bulaj 1998). Celui-ci se retrouve par exemple dans les doigts de zinc et l'alcool déshydrogénase. Le zinc joue alors un rôle structural ou catalytique essentiel pour la protéine.

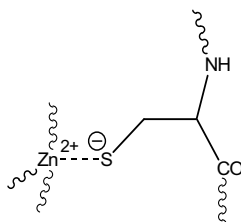


Figure I.5 : Motif zinc-cystéinate

III. Le motif zinc-thiolate dans les protéines

A. Rôle des métaux et du zinc dans les protéines

Les métaux ont un rôle structural et/ou fonctionnel primordial dans les protéines et participent à une grande diversité de processus chimiques. Dans certains cas, tels les ions métalliques divalents Mg^{2+} et Ca^{2+} , ils servent à stabiliser la structure des protéines. Mais ils peuvent également permettre des réactions catalytiques, en particulier les métaux de transition comme le fer, le cuivre et le manganèse qui sont impliqués dans de nombreux processus rédox. Les alcalins et les alcalino-terreux et spécialement les ions Na^+ , K^+ et Ca^{2+} jouent quant à eux un rôle essentiel dans le déclenchement des réponses cellulaires (Dudev 2003).

Parmi tous ces métaux, le zinc est celui qui présente le plus de fonctions dans les organismes vivants. On le retrouve en effet dans les six groupes qui forment la famille des enzymes (oxidoréductases, transférases, hydrolases, lyases, isomérases et ligases) (Vallee 1990). L'ion Zn(II) se retrouve dans des familles d'enzymes très connues telles que l'anhydrase carbonique, enzyme ubiquitaire qui catalyse l'hydratation du dioxyde de carbone, et dans les doigts de zinc qui jouent un rôle important dans la régulation du métabolisme des cellules et dont les séquences ADN qui les

codent occupent 3% du génome humain (Fausto da Silva 1991). Les fonctions biochimiques du zinc dans les enzymes peuvent être classées selon trois types (Prasad 1993, Vallee 1993) :

- Catalytiques : la présence du métal conditionne l'existence d'une activité catalytique (carboxypeptidase, aldolases ...)
- Structurales : le métal stabilise les structures tertiaire et quaternaire, permettant à la protéine de garder une conformation active (protéines de liaison à l'ADN)
- Régulatrices : le zinc a un effet activateur ou inhibiteur sur l'activité catalytique (aminopeptidase bovine)

B. Structure du motif zinc-thiolate dans les protéines

Les sites à zinc adoptent le plus fréquemment une conformation tétraédrique dans les protéines (cf Chapitre IV, Partie I.A.1). L'histidine est le ligand le plus courant, suivie par la cystéine, puis les résidus glutamates et aspartates (Vallee 1990, Alberts 1998, Dudev 2003, Patel 2007, Tamamès 2007). Ainsi, une étude récente établie à partir de 994 protéines à zinc de la PDB a déterminé qu'une fois sur quatre, le ligand coordonné au métal est une cystéine (Figure I.6).

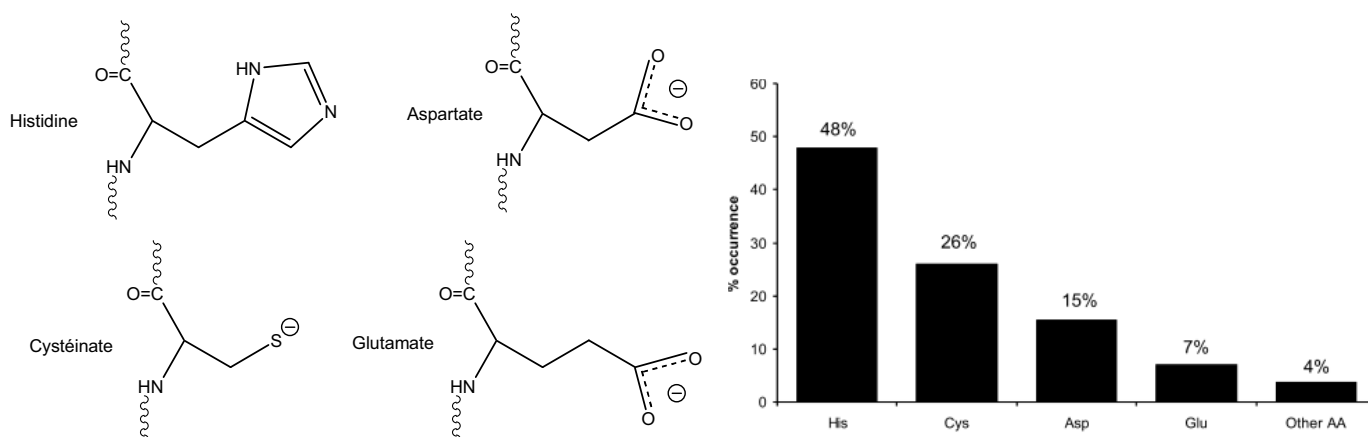


Figure I.6 : Résidus interagissant le plus fréquemment avec le zinc
(Parkin 2000, Tamames 2007)

L'histidine se coordonne via son noyau imidazole et constitue donc un donneur neutre, alors que les résidus glutamate et aspartate se lient par l'intermédiaire de leurs groupes anioniques carboxylates déprotonés.

Du fait de son pK_a déjà proche de la neutralité en milieu physiologique et de son interaction avec le zinc, qui joue le rôle d'acide de Lewis, la cystéine est vue le plus souvent comme un ligand anionique du zinc. Il faut noter cependant que l'état de protonation des résidus cystéines est encore sujet à controverse. Ainsi, alors qu'il est généralement considéré que les sites $Zn(Cys)_2(His)_2$, qu'on trouve par exemple dans les doigts de zinc, sont neutres et que leurs cystéines sont toutes déprotonées, l'état de protonation des cystéines des sites $Zn(Cys)_3(His)$ et $Zn(Cys)_4$ est encore incertain. Une expérience de spectrométrie de masse indique qu'une des cystéines est protonée dans les sites $Zn(Cys)_3(His)$ et que deux le sont dans les sites $Zn(Cys)_4$ de telle sorte que les sites aient une charge totale nulle (Fabris 1999). Cependant, de récents calculs suggèrent que les résidus cystéines

sont toujours déprotonés (Simonson 2002, Dudev 2002). Ceci explique la formation de liaisons hydrogènes intramoléculaires avec les acides aminés lysine et arginine pour stabiliser l'interaction Zn-Cys et est en accord avec les observations expérimentales précédemment effectuées en diffraction des rayons X (Dauter 1996) et EXAFS (Chance 1992, Clark-Baldwin 1998). Le plus grand nombre de liaisons hydrogènes et le plus fort écrantage par des charges résiduelles observé pour les sites $\text{Zn}(\text{Cys})_4$, comparativement aux sites $\text{Zn}(\text{Cys})_3(\text{His})$ et $\text{Zn}(\text{Cys})_2(\text{His})_2$, suggèrent également que les résidus liés au zinc sont sous forme cystéinate (Maynard 2001).

Le motif zinc-cystéinate se trouve donc dans de très nombreuses enzymes (doigts de zinc, facteurs de transcription de l'ADN, ...), à la fois dans des sites actifs ou structuraux. Les sites appelés actifs ou catalytiques possèdent le plus souvent une molécule d'eau, contrairement aux sites structuraux.

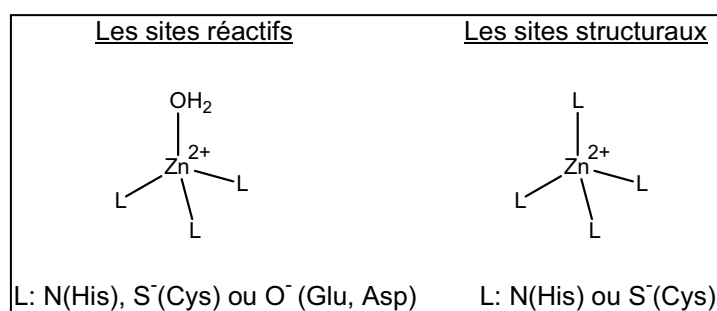


Figure I.7 : Les deux types de sites protéiques à zinc

Les sites structuraux possèdent une proportion plus importante en résidus cystéine et histidine : une étude récente indique que pour 61% des 994 enzymes à zinc extraites de la PDB, les ligands du métal sont exclusivement ces deux acides aminés (Figure I.8).

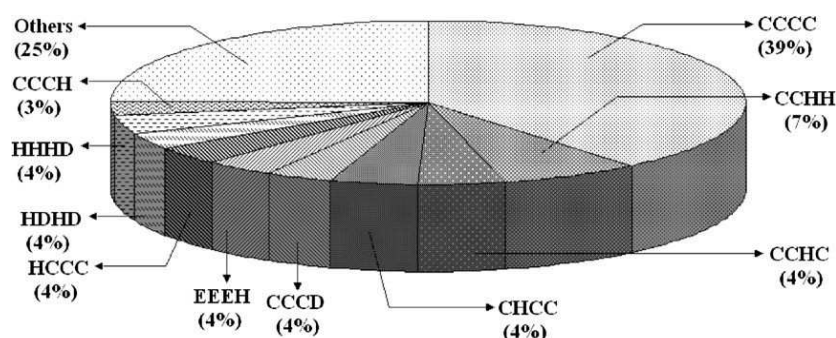


Figure I.8 : Ligands des sites structuraux à zinc
(H=Histidine, C=Cystéine, D=Acide Aspartique, E=Acide Glutamique)
(Patel 2007)

De plus, pour les sites structuraux, une conservation dans la séquence des acides aminés et un arrangement précis de ces séquences sont observés. Ainsi, des séquences courtes et conservées de 2 à 7 résidus relient les deux premiers ligands liés au zinc, ainsi que les deux derniers. Un espace relativement plus long, comprenant entre 20 et 120 résidus, connecte les deuxième et troisième ligands. La présence de ces courtes séquences pour les sites structuraux est cohérente avec leur rôle

de stabilisation conformationnelle, puisque cet arrangement des ligands apporte une certaine rigidité à la protéine. Un exemple est la famille des doigts de zinc, dans laquelle le métal assure la cohésion des sites de liaison à l'ADN (Vallee 1989 et 1993, Patel 2007).

Une conservation des séquences des sites catalytiques peut également être observée, même si elle est moins régulière. Les espaces entre les ligands sont alors composés d'un plus grand nombre de résidus pour garantir une certaine flexibilité conformationnelle à la protéine pendant l'activité enzymatique et pour permettre l'attachement d'un substrat.

C. Réactivité du motif zinc-thiolate dans les protéines

1. Réactivité des sites catalytiques

Les sites catalytiques sont tenus pour responsables de l'activité enzymatique des protéines à zinc. La réaction s'effectue au niveau de la molécule d'eau liée au métal et peut mener à trois mécanismes de réaction différents (Figure I.9).

Le plus fréquemment, le zinc agit comme acide de Lewis et active la molécule d'eau par déprotonation, de façon à engendrer une espèce hydroxyde plus nucléophile à un pH proche de la neutralité (a) (Vallee 1993, Parkin 2004). Ce mécanisme se retrouve, par exemple, pour l'anhydrase carbonique qui possède, outre la molécule d'eau liée au zinc, trois ligands histidines neutres en première sphère de coordination (Lindskog 1983 et 1997, Kumar 1994, Silverman 1991, Christianson 1996 et 1999, Liljas 1994, Tripp 2001).

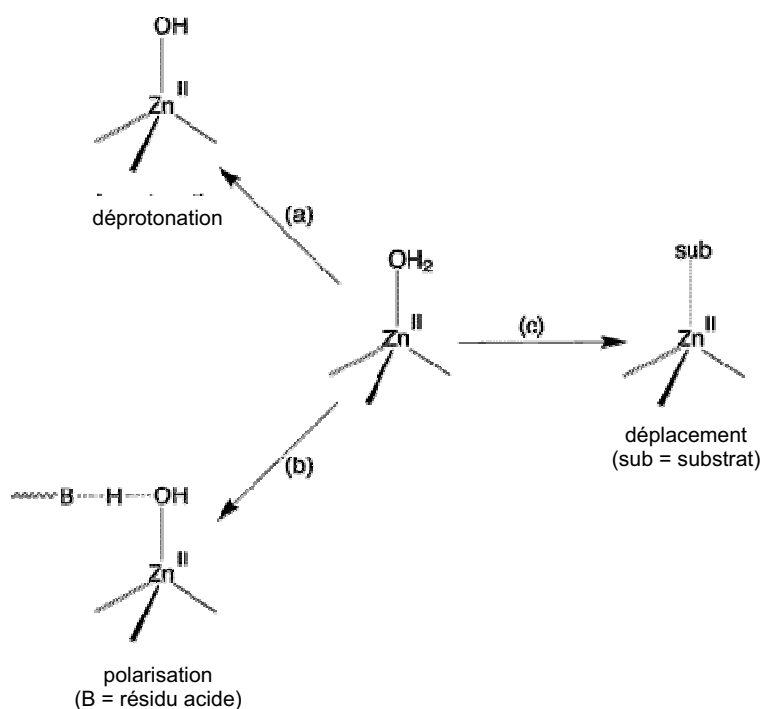


Figure I.9 : Mécanismes d'action des sites catalytiques à zinc

Un deuxième moyen d'activer la molécule d'eau apparaît dans le cas de l'enzyme carboxypeptidase possédant un résidu glutamate dans son site actif. La charge anionique de ce résidu rend la déprotonation à pH neutre de la molécule d'eau plus difficile et incomplète. Dans ce cas, un glutamate sert de base pour activer la molécule d'eau par polarisation et permet la formation d'une espèce hydroxyde (b) (Lipscomb 1970 et 1974, Christianson 1989, Alavarez-Santos 1994).

Le dernier type de participation du métal implique le déplacement complet de la molécule d'eau par un substrat (c) (Vallee 1993). Ceci se produit, par exemple, dans le cas de l'alcool déshydrogénase du foie, pour laquelle l'eau est déplacée par une molécule d'alcool. Deux groupes anioniques cystéines du site empêchent en effet la déprotonation du ligand d'origine, du fait d'une plus faible acidité de Lewis du zinc (Ryde 1994 et 1996).

Le substrat peut également être un groupe thiol. Par exemple, pour la protéine farnésyltransférase, une cystéine peptidique prend la place de la molécule d'eau dans la première sphère de coordination du zinc. Cette cystéine est ensuite alkylée par la farnésylphosphatase : le transfert du groupe farnésyl sur le thiol nucléophile est ainsi facilité par son activation par le zinc (Hightower 1999). Un tel mécanisme d'alkylation du thiol, après déplacement de la molécule d'eau du site catalytique, est aussi présent chez la géranylgeranyltransférase (Matthews 1997).

2. Réactivité des sites structuraux

Les sites structuraux présentent une forte proportion de ligands histidine et surtout cystéine. Alors que les sites catalytiques possédant une coordination avec les atomes donneurs N/O sont inertes envers les réactions d'oxydation et réduction, les ligands soufrés sont capables d'induire une activité redox dans les sites structuraux à zinc. Un tel environnement oxydo-réducteur, où les ligands, et non le métal, sont oxydés et réduits par la libération et la chélation concomitantes du zinc (Figure I.10), est une propriété remarquable des métalloprotéines (Maret 2004).

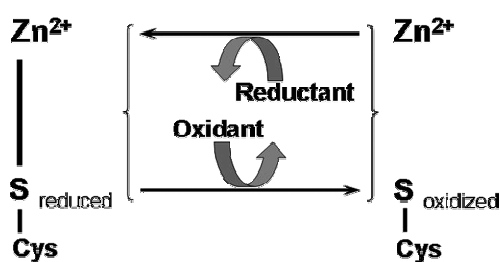


Figure I.10 : Chimie redox de l'interaction zinc-thiolate

La chimie redox de ces motifs zinc-thiolate constitue un moyen de contrôler la quantité de zinc, en stockage dans les protéines ou disponible pour les autres sites. Ce principe est utilisé par la famille des métallothionéines qui régule la distribution du zinc entre les diverses protéines dont les fonctions nécessitent la chélation au métal (Krezel 2007).

La fluctuation des quantités de zinc constitue des signaux intercellulaires modulables et capables d'initier des changements biochimiques ou physiologiques. Ainsi, la protéine Hsp-33 (Heat-Shock protein) de *E. Coli* procède à l'oxydation des ligands thiolates du zinc pour induire un important

changement de conformation qui entraîne l'activité de sa protéine chaperonne (Jacob 1999, Graf 2002). Le motif zinc-thiolate peut également servir de cible au stress oxydant (Figure I.11), comme par exemple chez le facteur *rsrA* anti-sigma de *Spreptomyces coelicor* (Li 2003). La formation d'un pont disulfide entre les cystéines de ses ligands métalliques, induite par stress oxydant, provoque le rejet du métal et stabilise une conformation empêchant l'attachement du facteur sigma.

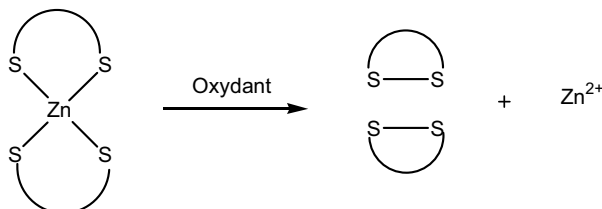


Figure I.11 : Stress oxydant provoquant la formation de ponts disulfides

Un autre exemple est la protéine de réplication A pour laquelle le site redox à zinc est indispensable pour la liaison à l'ADN et la reconnaissance de l'ADN endommagé (Park 1999).

La réactivité redox des motifs zinc-thiolate joue donc un rôle essentiel dans le contrôle de la fonction protéique. Cependant, malgré son importance dans les processus biologiques, nous nous concentrerons, dans cette étude, sur un autre aspect de la réactivité du motif zinc-thiolate, à savoir son pouvoir nucléophile et son alkylation.

Les sites structuraux, outre leur forte proportion en ligands cystéine et histidine, sont caractérisés par leur absence de molécule d'eau liée au zinc. Certaines familles d'enzymes présentent cependant une activité catalytique malgré l'absence de cette molécule d'eau (Parkin 2004). Les résidus protéiques coordonnés au zinc dans une sélection de ces enzymes indiquent que la cystéine est un composant nécessaire (Tableau I.1). La mise en évidence que ces sites, considérés comme structuraux, présentent néanmoins une activité non oxydante au niveau du motif zinc-thiolate est une des découvertes récentes les plus importantes dans la chimie bioinorganique du zinc. Contrairement à leurs homologues possédant un motif Zn-OH_2 , la réactivité particulière de ces enzymes à zinc, bien que d'un intérêt fondamental pour de nombreux processus biologiques, a été jusqu'ici peu étudiée.

Tableau I.1 : Ligands du zinc dans des sites structuraux présentant une activité (nd=Non déterminé) (Hightower 1999, Parkin 2004, Penner-Hahn 2007)

	L ₁	L ₂	L ₃	L ₄
Méthylcobamide :co-enzyme M méthyltransférase	Cys	His	nd	nd
Cobalamine Independent enzyme (Met E)	Cys	Cys	His	N/O
Cobalamine Dependent enzyme (Met H)	Cys	Cys	Cys	N/O
Betaine Homocystéine Transférase	Cys	Cys	Cys	nd
Ada	Cys	Cys	Cys	Cys

Dans les sites structuraux présentant une activité, le thiolate lié au zinc subit une alkylation (Figure I.12). Le zinc joue un rôle primordial en activant le thiol pour l'attaque nucléophile (Matthews 1997 et 2001, Penner-Hahn 2002 et 2007).

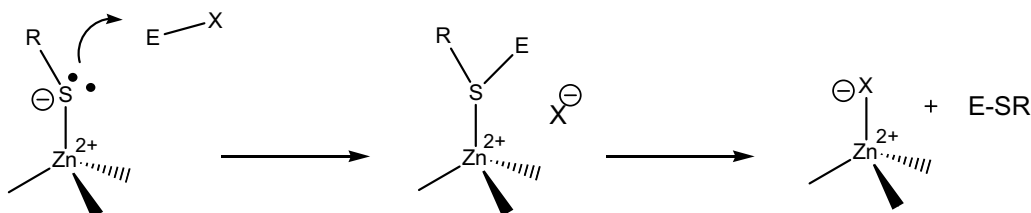


Figure I.12 : Alkylation d'un thiolate lié au zinc

L'alkylation de thiolates cystéiniques est une étape-clé dans le mécanisme de nombreuses protéines, comme par exemple la protéine Ada réparatrice de l'ADN (Myers 1993), les méthionines synthases (Gonzalez 1996, Jarrett 1997, Goulding 1997), les méthyltransférases méthanol:coenzyme M, méthylcobamide:coenzyme M (LeClerc 1996, Sauer 1997), bétaine-homocystéine (Millian 1998), la méthyltransférase de l'O⁶-alkylguanine-ADN (Fang 2005) et l'époxide carboxylase (Allen 1997).

a. Mécanisme de l'alkylation

Le premier site structural présentant une réactivité a été découvert pour la protéine Ada réparatrice de l'ADN. Cette protéine se sacrifie par autométhylation après une alkylation accidentelle de l'ADN (Figure I.13). Au niveau du site réactif, l'atome de zinc est lié à quatre ligands soufrés : Cys38, Cys42, Cys69 et Cys72. La cystéine Cys38 est irréversiblement méthylée par réaction avec les méthylphosphotriesters de l'ADN, alors que les autres cystéines ne sont pas réactives (Sedgwick 1988, Myers 1993 et 1995).

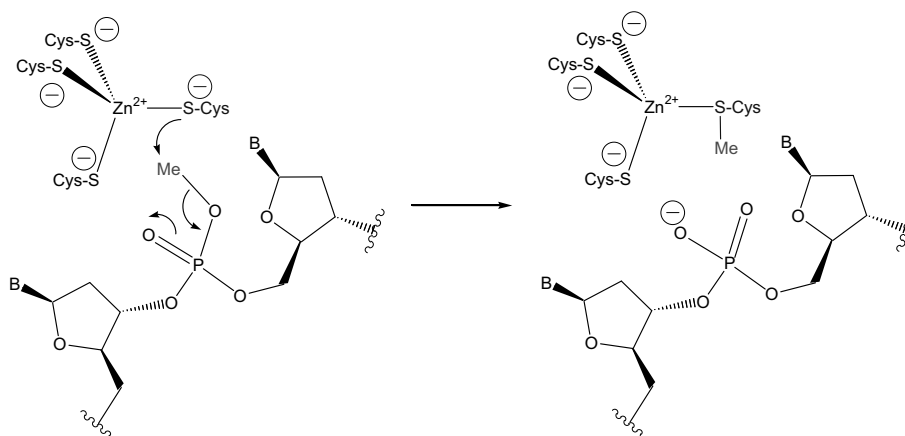


Figure I.13 : Réaction catalysée par la protéine Ada réparatrice de l'ADN

Avec son étude s'est posée la question du mécanisme réactionnel de l'alkylation subie au niveau du thiolate et plus particulièrement de la nature du nucléophile. En effet, plusieurs mécanismes peuvent exister : celui proposé pour la protéine Ada consiste à dissocier le thiolate coordonné au zinc

avant l'alkylation (Wilker 1997). L'espèce thiolate non-liée est alors plus nucléophile et est à l'origine de la réactivité dans le site (Figure I.14).

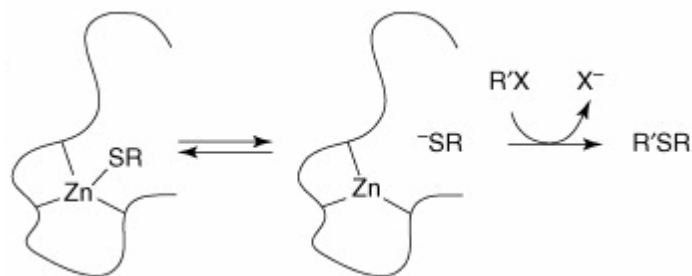


Figure I.14 : Mécanisme dissociatif de type S_N2 (Penner-Hahn 2007)

Un deuxième mécanisme possible fait intervenir le thiolate encore coordonné au métal, comme l'espèce nucléophile. Le mécanisme est alors dit associatif (Figure I.15). Deux possibilités peuvent alors être envisagées : le système zinc-soufre réagit via un mécanisme de type S_N2 , avec un état de transition classique où le groupe partant et le soufre sont sensiblement alignés, ou via un mécanisme concerté avec passage par un état de transition à quatre centres.

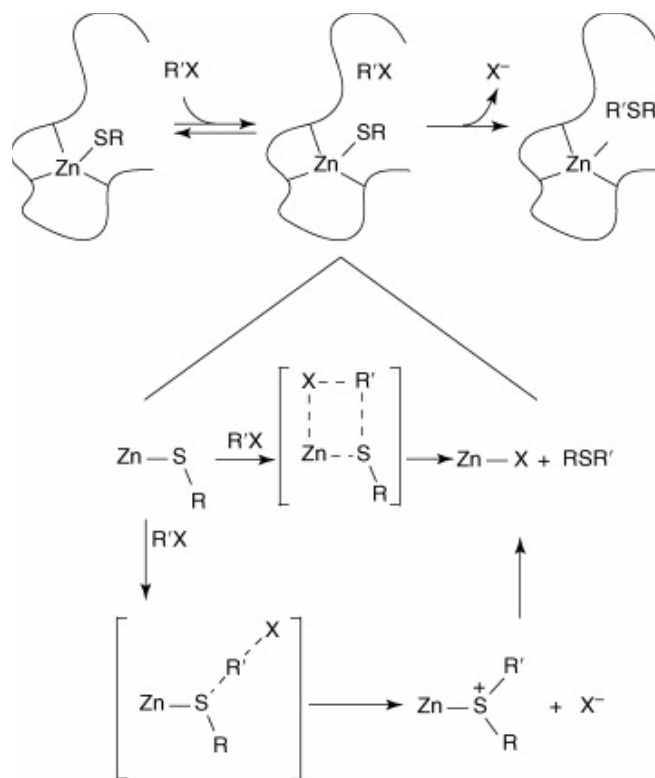


Figure I.15 : Mécanismes associatifs (Penner-Hahn 2007)

La littérature reste cependant contradictoire au sujet du caractère associatif ou dissociatif de la réaction enzymatique : alors que la protéine Ada est supposée réagir via un thiolate dissocié (Wilker 1997), un état de transition associatif est suggéré pour la farnésyl transférase (Huang 2000, Penner-Hahn 2007).

Le mécanisme de l'alkylation du thiolate pour la protéine Ada a été étudié expérimentalement avec le composé biomimétique $[\text{Zn}(\text{SPh})_4]^{2-}$ pour modéliser son site réactif $\text{Zn}(\text{Cys})_4$ localisé du côté N-terminal (Wilker 1995 et 1997). L'étude cinétique de la réaction entre $[\text{Zn}(\text{SPh})_4]^{2-}$ et l'agent alkylant $(\text{MeO})_3\text{PO}$ dans le solvant DMSO consiste à déterminer si le thiolate lié au zinc se dissocie ou non avant l'alkylation, en déduisant le mécanisme à partir des constantes de vitesse. Il a été montré que la réaction suit une cinétique de deuxième ordre et que la réactivité du complexe biomimétique $[\text{Zn}(\text{SPh})_4]^{2-}$ provient de son thiolate dissocié (Wilker 1995 et 1997). Un mécanisme dissociatif, où les thiophénolates libres sont les espèces nucléophiles, est donc suggéré pour le mécanisme de la protéine Ada.

La dissociation des ligands thiolates dans certains complexes à zinc a également été observée par RMN avec la mise en évidence d'échanges de ligands entre divers complexes (Melnick 2006). Cet échange facile de ligands renforce l'hypothèse d'une dissociation des thiolates liés au zinc avant alkylation. Il n'est cependant pas un argument décisif puisqu'il peut procéder via un mécanisme alternatif associatif (Penner-Hahn 2007). Enfin, alors que les études cinétiques ne pouvaient pas conclure quant à la nature du mécanisme d'alkylation de ces composés, l'observation d'un effet isotope normal a été interprétée comme l'indication d'un chemin dissociatif, même si, là encore, un mécanisme dissociatif ne peut être exclu (Morlok 2005).

Alors qu'il est postulé que l'alkylation enzymatique réalisée par la protéine Ada procède via un mécanisme réactionnel dissociatif, une étude portant sur la protéine farnésyltransférase suggère un mécanisme associatif (Hightower 1999). Cette enzyme catalyse le transfert du groupe farnésyl depuis la farnésyl pyrophosphatase (FPP) à un résidu cystéine d'un substrat peptidique. Le site à zinc est constitué des trois acides aminés Asp297, Cys 299 et His 362 et d'une molécule d'eau, qui n'a pas de rôle catalytique (Zhang 1996, Park 1997).

Plusieurs mécanismes ont été proposés pour l'alkylation catalysée par la farnésyltransférase : électrophile ($\text{S}_\text{N}1$ dissociatif) avec formation d'une charge positive sur le carbone terminal de la FPP (Pompliano 1992, Dolence 1995, Cassidy 1996), ou nucléophile ($\text{S}_\text{N}2$ associatif) avec attaque du thiolate d'un résidu cystéinique du substrat (Pompliano 1992, Huang 1997). Ces études sont en faveur d'un mécanisme nucléophile associatif, avec la présence d'un thiolate du substrat lié au zinc de la protéine (Hightower 1999, Huang 2000) (Figure I.16).

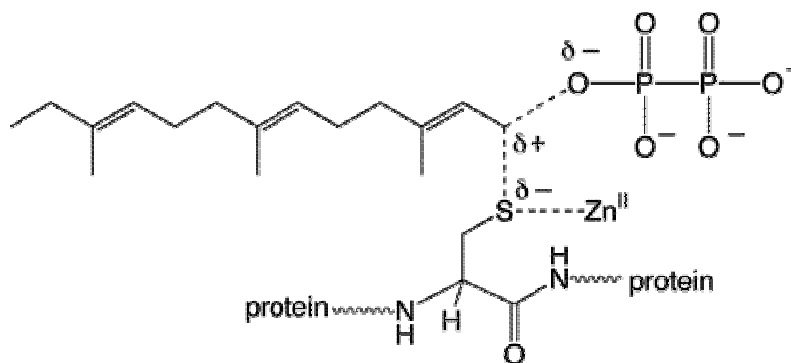


Figure I.16 : Etat de transition associatif proposé pour la farnésyltransférase (Huang 2000)

Cette étude sur l'enzyme farnésyltransférase est l'une des rares à considérer la réaction d'alkylation en milieu biologique. Les autres travaux mécanistiques sur le transfert d'alkyle dans les protéines à zinc portent sur des complexes biomimétiques modélisant uniquement le site métallique. Généralement, ces systèmes montrent peu ou pas d'échange de ligands dans les milieux non polaires (Seebacher 2004) et réagissent, d'après les études cinétiques, via un mécanisme associatif. Ainsi, une série de complexes biomimétiques, neutres ou monocationiques, ont montré qu'ils étaient alkylés sans dissociation préalable de leur thiolate (Grapperhaus 1998, Bridgewater 2000, Chiou 2000 et 2003, Warthen 2001, Brand 2001, Smith 2003 et 2005, Ibrahim 2005, Morlok 2005, Ji 2005, Rombach 2006, Notni 2007). Cependant, l'étude de ces complexes n'apporte pas d'indication certaine quant au mécanisme enzymatique, puisqu'à l'exception du site neutre $\text{Zn}(\text{Cys})_2(\text{His})_2$ de l'alkylguanine transférase de l'ADN, tous les sites protéiques à zinc sont monoanioniques ou dianioniques.

b. Sélectivité de l'alkylation

Un autre aspect de l'alkylation restant sujet à controverse est sa sélectivité envers un unique thiolate au sein d'un site à zinc. Par exemple, dans le site ZnS_4 de la protéine Ada, la cystéine Cys38 est exclusivement alkylée. Il a été remarqué que son soufre est le seul ne possédant pas de liaison hydrogène (He 2005, Takinowaki 2006) et que la présence de ces liaisons sur les autres cystéinates réduisait grandement leur nucléophilie (Habazettl 1996, Myers 1995). Il est alors postulé que l'environnement immédiat des thiolates, et en particulier l'existence ou non de liaisons hydrogène, est responsable de ces réactivités différentes.

Ce phénomène de sélectivité se retrouve pour la nucléocapside p7 (Ncp7) du virus VIH qui contient deux copies de séquences conservées formant les sites structuraux $\text{Zn}(\text{Cys})_3(\text{His})$ (Summers 1992). Un de ces deux doigts de zinc, situé du côté C-terminal, peut être alkylé, menant à l'inhibition du HIV et de nombreux inhibiteurs ont été développés dans le cadre de la recherche d'un traitement anti-sida (Tummino 1996, Loo, 1996, Hathout 1996, Maynard 1998, Huang 1998). Il a été montré que parmi les trois cystéines du doigt de zinc réactif, la cystéine 49 est alkylée préférentiellement. Les causes avancées pour cette sélectivité sont l'accès plus facile au solvant de la cystéine et son plus grand caractère nucléophile mesuré par des coefficients de Fukui (Maynard 1998, Huang 1998). De plus, un examen des liaisons hydrogène indique que la cystéine 49 se situe dans une région plus flexible du doigt de zinc que les deux autres cystéinates, grâce à un réseau moins fourni d'interactions de type liaisons hydrogène. Sa plus faible participation dans les interactions intramoléculaires la rend plus accessible et surtout conduit à une densité électronique plus importante de l'atome de soufre de la cystéine 49 par rapport aux deux autres soufres (Summers 1992, Hathout 1996).

Des études biomimétiques ont également été conduites pour vérifier l'influence des liaisons hydrogène sur la réactivité des soufres et donc la sélectivité de l'alkylation enzymatique (Chiou 2000 et 2003). Cependant, la rationalisation et la quantification de ces effets, en particulier la nucléophilie du motif zinc-thiolate, sont encore à l'heure actuelles insuffisantes (Fox 2004).

Le problème de la sélectivité se retrouve pour un groupe thiolate comparativement à un autre au sein d'un même site, mais aussi entre deux sites. Les réactivités des différents sites protéiques $\text{Zn}(\text{Cys})_4^{2-}$, $\text{Zn}(\text{Cys})_3(\text{His})^-$ et $\text{Zn}(\text{Cys})_2(\text{His})_2$ ont été étudiées à l'aide des composés modèles $[\text{Zn}(\text{SPh})_4]^{2-}$, $[\text{Zn}(\text{SPh})_3(\text{ImMe})]^-$ et $[\text{Zn}(\text{SPh})_2(\text{ImMe})_2]$ respectivement. Les réactions de ces complexes avec l'agent alkylant $(\text{CH}_3\text{O})\text{PO}$ ont été effectuées dans le DMSO et les constantes de vitesse correspondantes mesurées (Wilker 1995). L'ordre de réactivité montre que le complexe dianionique est le plus réactif, suivi du complexe monoanionique, puis du composé neutre. Ceci indique que la nucléophilie intrinsèque des sites à quatre cystéines est supérieure à celle des sites $\text{Zn}(\text{Cys})_3(\text{His})^-$ et $\text{Zn}(\text{Cys})_2(\text{His})_2$. Ces réactivités sont estimées comme variant suivant un rapport 1000 (Cys_4) :100 (Cys_3His) :1 (Cys_2His_2). La première sphère de coordination du zinc a donc un impact important sur la nucléophilie des sites actifs.

Maynard et Covell ont également mis en évidence les différences de réactivité des différents sites à zinc, mais en prenant en compte cette fois l'environnement plus distant que la première sphère de coordination du métal (Maynard 2001). Les effets stériques et électroniques des résidus environnant le motif central ont été évalués par les distributions radiales de la densité protéique et de l'écrantage électrostatique autour du site. Cet écrantage provient des charges résiduelles et des interactions intramoléculaires avec le squelette protéique, du type liaisons hydrogène NH-O et NH-S. Les deux paramètres ainsi calculés indiquent que l'écrantage électrostatique et stérique augmente avec la charge totale du site à zinc. Ainsi, les sites anioniques les plus nucléophiles se trouvent être les plus compacts en densité et ceux possédant le plus de liaisons hydrogènes, alors que les sites structuraux neutres présentent peu d'écrantage. Ceci est cohérent avec l'ordre de réactivité trouvé expérimentalement par Lippard et Wilker et indique que la réactivité des sites est donc aussi à mettre en relation avec l'influence de l'environnement plus distant. Ainsi, le site à quatre cystéines, le plus nucléophile intrinsèquement, peut voir sa réactivité modulée fortement par la présence de résidus basiques ou donneurs de liaisons hydrogènes.

Par exemple, malgré leur similitude structurale (South 1991), les deux doigts de zinc du nucléocapside Ncp7 de la protéine HIV présentent une réactivité différente envers des agents alkylants. Ainsi, le site à zinc localisé du côté C-terminal possède une vitesse de réaction sept fois plus importante que celui situé du côté N-terminal (Tummino 1996, Hathout 1996, Loo 1996). Or, une comparaison des deux doigts de zinc de la nucléocapside indique que celui du C-terminal est plus accessible pour l'approche d'inhibiteurs (Maynard 1998), et également que celui-ci compte un nombre moins important de liaisons hydrogène que son homologue (Hathout 1996). Le moindre écrantage électrostatique du C-terminal permet d'apporter une explication à sa plus grande capacité à l'alkylation, alors que les interactions intramoléculaires et l'environnement compact autour du site à zinc du côté N-terminal diminuent sa réactivité (Maynard 2001).

Ceci est en cohérence avec le fait que le site réactif $\text{Zn}(\text{Cys})_3(\text{His})$ de la nucléocapside est plus vulnérable à une attaque électrophile que le site $\text{Zn}(\text{Cys})_4$ du doigt de zinc GATA-1 (Huang 1998). En effet, l'encombrement stérique et l'écrantage électrostatique du site dianionique sont respectivement 1.5 et 2.8 fois plus importants que ceux du site monoanionique. Ainsi, le site $\text{Zn}(\text{Cys})_4$ de GATA-1, possédant intrinsèquement une nucléophilie élevée, voit sa réactivité fortement réduite

par la présence de résidus basiques et d' un réseau important de liaisons hydrogène dirigées sur ses thiolates. L'environnement d'un site actif a donc une influence toute aussi importante sur sa réactivité que la première sphère de coordination du zinc.

Divers aspects de l'alkylation enzymatique d'un thiolate restent à éclaircir : la nature du mécanisme, dissociatif ou associatif, et par conséquent la nature de l'espèce nucléophile restent incertaines. De plus, l'environnement proche du zinc, à savoir la nature de ses ligands et l'environnement plus distant, comme la présence de liaisons hydrogène sur les thiolates et la proximité de charges dans le site, ont un impact important sur la capacité nucléophile d'un site. La balance des facteurs contrôlant l'alkylation enzymatique est subtile et nous essaierons d'établir les règles de réactivité inhérentes à cette nouvelle classe de sites actifs, grâce notamment à l'étude quantique de complexes biomimétiques.

Bibliographie

- Alberts, I. L.; Nadassy, K.; Wodak, S. *J. Prot. Sci.* **1998**, 7, 1700.
- Allen, J. R.; Ensign, S. A. *J. Biol. Chem.* **1997**, 272, 32121.
- Alvarez-Santos, S.; Gonzalez-Lafont, A.; Lluch, J. M.; Oliva, B.; Aviles, F. X. *Can. J. Chem.* **1994**, 72, 2077.
- Bertini, I.; Luchinat, C. *Acc. Chem. Res.* **1983**, 16, 272.
- Bordwell, G.; Hugues, D. L. *J. Org. Chem.* **1982**, 47, 3224.
- Brand, U.; Rombach, M.; Seebacher, J.; Vahrenkamp, H. *Inorg. Chem.* **2001**, 40, 6151.
- Bridgewater, B. M.; Fillebeen, T.; Friesner, R. A.; Parkin, G. *J. Chem. Soc., Dalton Trans.* **2000**, 4494.
- Bulaj, G.; Kortemme, T.; Goldenberg, D. *Biochemistry* **1998**, 37, 8965.
- Carey, F. A. *Organic Chemistry*, 6th edition, McGraw Hill edition, Chapitre 15, **2006**.
- Cassidy, P. B.; Poulter, C. D. *J. Am. Chem. Soc.* **1996**, 118, 8761.
- Chance, M.; Sagi, I.; Wirt, M.; Frisbie, S.; Scheuring, E.; Chen, E.; Bess, J.; Henderson, L.; Arthur, L.; South, T.; Perez-Alvarado, G.; Summers, M. *Proc. Nat. Acad. Sci. U.S.A.* **1992**, 89, 10041.
- Chiou, S.-J.; Innocent, J.; Riordan, C. G.; Lam, K.-C.; Liable-Sands, L.; Rheingold, A. L. *Inorg. Chem.* **2000**, 39, 4347.
- Chiou, S.-J.; Riordan, C. G.; Rheingold, A. L. *Proc. Nat. Acad. Sci. U.S.A.* **2003**, 100, 3695.
- Christianson, D. W.; Lipscomb, W. N. *Acc. Chem. Res.* **1989**, 22, 62.
- Christianson, D. W. *Adv. Protein Chem.* **1991**, 42, 281.
- Christianson, D. W.; Fierke, C. A. *Acc. Chem. Res.* **1996**, 29, 331.
- Christianson, D. W.; Cox, J. D. *Annu. Rev. Biochem.* **1999**, 68, 33.
- Clark-Baldwin, K.; Tierney, D.; Govindaswamy, N.; Gruff, E.; Kim, C.; Berg, J.; Koch, S.; Penner-Hahn, J. *J. Am. Chem. Soc.* **1998**, 120, 8401.
- Dauter, Z.; Wilson, K.; Sieker, L.; Moulis, J.; Meyer, J. *Proc. Nat. Acad. Sci. U.S.A.* **1996**, 93, 8836.
- Dolence, J. M.; Poulter, C. D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 5008.
- Dudev, T.; Lim, C. *J. Am. Chem. Soc.* **2002**, 124, 6759.
- Dudev, T.; Lim, C. *Chem. Rev.* **2003**, 103, 773.
- Fabris, D.; Hathout, Y.; Fenselau, C. *Inorg. Chem.* **1999**, 38, 1322.
- Fang, Q.; Kanugula, S.; Pegg, A. E. *Biochemistry* **2005**, 44, 15396.
- Fausto da Silva, J. J. R.; Williams, R. J. P. *The Biological Chemistry of Elements*, Clarendon Press, Oxford, **1991**.
- Fox, D. C.; Fiedler, A. T.; Halfen, H. N.; Brunold, T. C.; Halfen, J. A. *J. Am. Chem. Soc.* **2004**, 126, 7627.
- Gonzalez, J. C.; Peariso, K.; Penner-Hahn, J. E.; Matthews, R. G. *Biochemistry* **1996**, 35, 12228.
- Goulding, C. W.; Matthews, R. G. *Biochemistry* **1997**, 36, 15749.
- Graf, P. C.; F.; Jacob, U. *Cell. Mol. Life* **2002**, 59, 1624.
- Grapperhaus, C. A.; Tuntulani, T.; Reibenspies, J. H.; Darensbourg, M. Y. *Inorg. Chem.* **1998**, 37, 4052.
- Habazettl, J.; Myers, L. C.; Yuan, F.; Verdine, G. L.; Wagner, G. *Biochemistry* **1996**, 35, 9335.

- Handbook of Chemistry and Physics*, 71st edition, CRC Press, Ann Arbor, Michigan, **1990**.
- Hathout, Y.; Fabris, D.; Han, M. S.; Sowder, R. C.; Henderson, L. E.; Fenselau, C. *Drug. Metab. Dispos.* **1996**, *24*, 1395.
- He, C.; Hus, J. C.; Sun, L. J.; Zhou, P.; Norman, D. P. G.; Dotsch, V.; Wei, H.; Gross, J. D.; Lane, W. S.; Wagner, G.; Verdine, G. L. *Mol. Cell* **2005**, *20*, 117.
- Hightower, K. E.; Fierke, C. A. *Curr. Opin. Chem. Biol.* **1999**, *3*, 176.
- Hikichi, S.; Tanaka, M.; Moro-oka, Y.; Kitajima, N. *J. Chem. Soc., Chem. Commun.* **1994**, 1737.
- Horrocks, W. D., Jr.; Ishley, J. N.; Holmquist, B.; Thompson, J. S. *J. Inorg. Biochem.* **1980**, *12*, 131.
- Huang, C. C.; Casey, P. J.; Fierke, C. A. *J. Biol. Chem.* **1997**, *272*, 20.
- Huang, M.; Maynard, A.; Turpin, J. A.; Graham, L.; Janini, G. M.; Covell, D. G.; Rice, W. G. *J. Med. Chem.* **1998**, *41*, 1371.
- Huang, C. C.; Hightower, K. E.; Fierke, C. A. *Biochemistry* **2000**, *39*, 2593.
- Ibrahim, M. M.; Shu, M.; Vahrenkamp, H. *Eur. J. Inorg. Chem.* **2005**, *7*, 1388.
- Jacob, U.; Muse, W.; Eser, M.; Bardwell, J. C. A. *Cell* **1999**, *96*, 341.
- Jarrett, J. T.; Choi, C. Y.; Matthews, R. W. *Biochemistry* **1997**, *36*, 15739.
- Ji, M.; Benkmil, B.; Vahrenkamp, H. *Inorg. Chem.* **2005**, *44*, 3518.
- Krezel, A.; Hao, Q.; Maret, W. *Arch. Biochem. Biophys.* **2007**, *463*, 188.
- Kumar, V.; Kannan, K. K. *Indian J. Biochem. Biophys.* **1994**, *31*, 377.
- LeClerc, G. M.; Grahame, D. A. *J. Biol. Chem.* **1996**, *271*, 8725.
- Li, W.; Bottrill, A. R.; Bibb, M. J.; Buttner, M. J.; Paget, M. S. B.; Kleanthous, C. *J. Mol. Biol.* **2003**, *333*, 461.
- Liljas, A.; Håkansson, K.; Jonsson, B. H.; Xue, Y. F. *Eur. J. Biochem.* **1994**, *219*, 1.
- Lindskog, S. *Met. Ions Biol.* **1983**, *5*, 77.
- Lindskog, S. *Pharmacol. Ther.* **1997**, *74*, 1.
- Lipscomb, W. N. *Acc. Chem. Res.* **1970**, *3*, 81.
- Lipscomb, W. N. *Tetrahedron* **1974**, *30*, 1725.
- Loo, J. A.; Holler, T. P.; Sanchez, J.; Gogliotti, R.; Maloney, L.; Reily, M. D. *J. Med. Chem.* **1996**, *39*, 4313.
- Maret, W.; Vallee, B. L. *Methods Enzymol.* **1993**, *226*, 52.
- Maret, W. *Biochemistry* **2004**, *43*, 3301.
- Matthews, R. G.; Goulding, C. W. *Curr. Opinion Chem. Biol.* **1997**, *1*, 332.
- Matthews, R. G. *Acc. Chem. Res.* **2001**, *34*, 681.
- Maynard, A. T.; Huang, M.; Rice, W. G.; Covell, D. G. **1998**, *95*, 11578.
- Maynard, A. T.; Covell, D. G. *J. Am. Chem. Soc.* **2001**, *123*, 1047.
- Melnick, J. G.; Zhu, G.; Buccella, D.; Parkin, G. *J. Inorg. Bioch.* **2006**, *100*, 1147.
- Merck Index*, 12th edition, Ralway, N.J., USA, Merck and co., **1996**.
- Millian, N. S.; Garrow, T. A. *Arch. Biochem. Biophys.* **1998**, *356*, 93.
- Morlok, M. M.; Janak, K. E.; Zhu, G.; Quarless, D. A.; Parkin, G. *J. Am. Chem. Soc.* **2005**, *127*, 14039.
- Myers, L. C.; Terranova, M. P.; Ferentz, A. E.; Wagner, G.; Verdine, G. L. *Science* **1993**, *261*, 1164.
- Myers, L. C.; Verdine, G. L.; Wagner, G. *Biochemistry* **1993**, *32*, 14069.

- Myers, L. C.; Wagner, G.; Verdine, G. L. *J. Am. Chem. Soc.* **1995**, *117*, 10749.
- Notni, J.; Gunther, W.; Anders, E. *Eur. J. Inorg. Chem.* **2007**, 985.
- Park, H. W.; Boduluri, S. W.; Moomaw, J. F.; Casey, P. J.; Beese, L. S. *Science* **1997**, *275*, 1800.
- Park, J. S.; Wang, M.; Park, S. J.; Lee, S. H. *J. Biol. Chem.* **1999**, *274*, 29075.
- Parkin, G. *Chem. Comm.* **2000**, 1971.
- Parkin, G. *Chem. Rev.* **2004**, *104*, 699.
- Patel, K.; Kumar, A.; Durani, S. *Biochim. Biophys. Acta* **2007**, *1774*, 1247.
- Penner-Hahn, J. *Indian J. Chem., Sect A* **2002**, *41*, 13.
- Penner-Hahn, J. *Curr. Opin. Chem. Biol.* **2007**, *11*, 166.
- Pompliano, D. L.; Rands, E.; Schaber, M. D.; Mosser, S. D.; Anthony, N. J.; Gibbs, J. B. *Biochemistry* **1992**, *31*, 3800.
- Prasad, A. S. *Biochemistry of Zinc*, Plenum Press, New-York, **1993**.
- Rombach, M.; Seebacher, J.; Ji, M.; Zhang, G. F.; He, G. S.; Ibrahim, M. M.; Benkmil, B.; Vahrenkamp, H. *Inorg. Chem.* **2006**, *45*, 4571.
- Ryde, U. *Int. J. Quantum Chem.* **1994**, *52*, 1229.
- Ryde, U. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 153.
- Ryde, U. *Eur. Biophys. J.* **1996**, *24*, 213.
- Sauer, K.; Thauer, R. K. *J. Biochem.* **1997**, *249*, 280.
- Sedgwick, B.; Robins, P.; Totty, N.; Lindhal, T. *J. Biol. Chem.* **1988**, *263*, 4430.
- Seebacher, J.; Ji, M.; Vahrenkamp, H. *Eur. J. Inorg. Chem.* **2004**, *2*, 409.
- Silverman, D. N. *Can. J. Bot.* **1991**, *69*, 1070.
- Simonson, T.; Calimet, N. *Proteins* **2002**, *49*, 37.
- Smith, J. N.; Shirin, Z.; Carrano, C. J. *J. Am. Chem. Soc.* **2003**, *125*, 868.
- Smith, J. N.; Hoffman, J. T.; Shirin, Z.; Carrano, C. J. *Inorg. Chem.* **2005**, *44*, 2012.
- South, T. L.; Blake, P. R.; Hare, D. R.; Summers, M. F. *Biochemistry* **1991**, *30*, 6342.
- Summers, M. F.; Henderson, L. E.; Chance, M. R.; Bess, J. W.; South, T. L.; Blake, P. R.; Sagi, I.; Perez-Alvarado, G.; Sowder, R. C.; Hare, D. R.; Arthur, L. O. *Protein Sci.* **1992**, *1*, 563.
- Sun, G.; Budde, R. J. A. *Biochemistry* **1999**, *68*, 5659.
- Takinowaki, H.; Matsuda, Y.; Yoshida, T.; Kobayashi, Y.; Ohkubo, T. *Protein Sci.* **2006**, *15*, 487.
- Tamames, B.; Sousa, S. F.; Tamames, J.; Fernandes, P. A.; Ramos, M. J. *Proteins, Struct. Funct. Bioinf.* **2007**, *69*, 466.
- Thomson, A. J.; Gray, H. E. *Curr. Opin. Struct. Biol.* **1992**, *2*, 155.
- Torrance, J. W.; MacArthur, M. W.; Thornton, J. M. *Proteins* **2007**, *71*, 813.
- Tripp, B. C.; Smith, K.; Ferry, J. G. *J. Biol. Chem.* **2001**, *276*, 48615.
- Tummino, P. J.; Scholten, J. D.; Harvey, P. J.; Holler, T. P.; Maloney, L.; Gogliotti, R.; Domagala, J.; Hupe, D. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 969.
- Vahrenkamp, H. *Dalton Trans.* **2007**, *42*, 4751.
- Vallee, B. L.; Auld, D. S. *FEBS Letters* **1989**, *257*, 138.
- Vallee, B. L.; Auld, D. S. *Biochemistry* **1990**, *29*, 5647.
- Vallee, B. L.; Auld, D. S. *Acc. Chem. Res.* **1993**, *26*, 543.

- Warthen, C. R.; Hammes, B. S.; Carrano, C. J.; Crans, D. C. *J. Biol. Inorg. Chem.* **2001**, 6, 82.
- Wilker, J. J.; Lippard, S. J. *J. Am. Chem. Soc.* **1995**, 117, 8682.
- Wilker, J. J.; Lippard, S. J. *Inorg. Chem.* **1997**, 36, 969.
- Zhang, F. L.; Casey, P. J. *Annu. Rev. Biochem.* **1996**, 65, 241.

Chapitre II

Méthodes de calculs

I. Généralités

A. Introduction

L'essor des méthodes de calculs d'orbitales moléculaires au cours des trente dernières années a été largement lié à celui de l'informatique ainsi qu'à celui des algorithmes. Ces méthodes théoriques, aujourd'hui largement répandues, permettent non seulement de déterminer les conformations des molécules, comme les méthodes de mécanique moléculaire, mais aussi leurs énergies, les chemins de réactions et certaines de leurs propriétés physico-chimiques. Elles sont souvent très coûteuses en temps de calcul et en mémoire et donc associées à la chimie d'espèces de petite taille, c'est-à-dire d'au plus de quelques dizaines à quelques centaines d'atomes, principalement en phase gazeuse. Elles permettent de décrire des composés n'ayant jamais été isolés ou synthétisés et même des entités ayant des durées de vie trop courtes pour être observées comme, par exemple, des états de transition.

De nombreux logiciels ont été développés permettant le calcul quantique des molécules, tels que ADF, NWChem, Molpro, Molcas ou Turbomole, le plus connu et utilisé étant le logiciel Gaussian (Frisch 2004). De nombreuses fonctionnalités sont disponibles dans ces programmes, que ce soit, par exemple, les méthodes de chimie quantique (semi-empirique, Hartree-Fock (HF) et post-Hartree-Fock, multiples fonctionnelles de la densité), les bases d'orbitales moléculaires et les pseudo-potentiels associés, ou encore les possibilités d'optimisation de la géométrie avec contrainte ou l'inclusion d'effet de solvant. Face à un problème chimique donné, il apparaît donc opportun de sélectionner les méthodes de modélisation, et donc le ou les logiciels les plus adaptés.

B. Stratégie

L'objectif tout au long de notre travail est d'expliquer la structure et la réactivité d'alkylation du motif zinc-thiolate dans les sites actifs d'enzymes. Différents types de composés ont été étudiés dans ce but, depuis de petits systèmes modèles en phase gazeuse contenant quelques dizaines d'atomes jusqu'à des parties entières de protéines contenant plusieurs centaines d'atomes. De plus, sur certains de ces systèmes, comme nous le montrerons, le solvant joue un rôle majeur. Cette diversité de taille nécessitait des modélisations adaptées.

Les méthodes DFT et post-HF présentes dans le logiciel GAUSSIAN03 ont été utilisées pour l'étude de la réactivité de complexes biomimétiques (Chapitre III) et l'étude de la structure et de la nucléophilie de divers thiolates coordonnés au zinc (Chapitres IV et V). Le modèle de solvation C-PCM (Barone 1998, Cossi 2003), continuum diélectrique polarisable, a été utilisé pour ces systèmes de petite taille. Dans ce cas, seuls des calculs utilisant la géométrie obtenue en phase gazeuse ont pu être effectués. En effet, il n'a pas été possible d'obtenir la convergence des optimisations de géométrie avec cette méthode de solvation. Cette sophistication de nos modélisations s'est néanmoins révélée nécessaire dans certains cas (Chapitre IV), et dans ce cas, nous avons eu recours au logiciel TURBOMOLE (Ahlrichs 1989) qui permet une optimisation aisée avec le système de solvation COSMO (Klamt 1993).

Le logiciel TURBOMOLE permet également de faire des calculs dans l'approximation de la « Résolution de l'Identité » et dans l'approximation plus récente de la « Résolution de l'Identité accélérée par les Multipoles », qui offre un gain notable en temps de calcul pour les systèmes de grande taille. Ces méthodes seront donc employées pour l'étude de fragments de protéines (chapitre VI).

GAUSSIAN03 et TURBOMOLE utilisant des bases d'orbitales atomiques différentes et les calculs n'étant pas effectués exactement au même niveau, des calibrations sur les géométries et/ou sur les énergies relatives seront nécessaires pour la comparaison des résultats.

II. Bases d'orbitales atomiques

A. Les bases de gaussiennes

Les bases standard utilisées dans les calculs *ab initio* et DFT, par exemple avec le logiciel GAUSSIAN, sont des combinaisons linéaires de fonctions gaussiennes centrées sur les noyaux de la molécule considérée. Un tel choix de représentation des fonctions est très avantageux surtout pour le calcul des intégrales multicentriques, car le produit de deux gaussiennes centrées en deux points différents est une gaussienne centrée en un autre point. Ainsi, toutes les intégrales de répulsion électronique se ramènent à une somme d'intégrales monocentriques qui se calculent aisément. La précision des résultats dépend fortement du nombre de gaussiennes utilisées pour développer chaque orbitale.

Une fonction gaussienne n'est pas a priori bien adaptée pour décrire une densité électronique. Pour éviter ce problème, on utilise des gaussiennes multiples pour représenter une même orbitale. On sait intuitivement que certaines parties de la densité électronique des atomes restent inchangées quelque soit la molécule dont ils font partie. Il s'agit en particulier des régions proches du noyau. On détermine donc les coefficients d'une combinaison linéaire de fonctions gaussiennes, dits coefficients de contraction, une fois pour toutes, par un calcul sur l'atome seul. Seules les gaussiennes d'exposants faibles, destinées à décrire la partie externe de la densité électronique de chaque atome, ne sont pas contractées.

De nombreuses bases de fonctions gaussiennes ont été développées depuis 30 ans environ, l'un des ensembles les plus complets étant dû à l'équipe de J. A. Pople (Hehre 1986).

Différentes bases d'orbitales atomiques, de qualités variées, ont été utilisées au cours de ce travail. Celles-ci sont décrites dans les paragraphes suivants, ainsi que les notations abrégées pour les nommer par la suite.

Les différentes bases et pseudo-potentiels utilisés pour le zinc et l'iode sont explicités en annexe de ce chapitre.

1. Les bases tous électrons

a. Les bases initiales (utilisées pour l'optimisation de géométries)

La grande majorité des optimisations de géométries réalisées avec GAUSSIAN03 utilise la base d'orbitales atomiques notée BS1 dans ce manuscrit. Elle se décompose de la façon suivante :

- Pour tous les atomes à l'exception du zinc et de l'iode, c'est-à-dire les atomes H, B, C, N, O, F, Si, P et S, la base standard de Pople 6-31G(d,p) est utilisée. Cela signifie que l'on décrit les couches de cœur avec six gaussiennes par fonction contractée. Pour la partie interne de chaque orbitale de valence, on utilise une contraction de trois gaussiennes, et pour la partie externe, une seule gaussienne (ce qui permet une bonne adaptation à l'environnement de l'atome). Enfin, pour tous les atomes, on ajoute une fonction gaussienne de nombre quantique azimutal supérieur, appelée *fonction de polarisation*, ce qui donne une meilleure description de la déformation du nuage électronique de l'atome dans la molécule.
- Pour le zinc, la base tous électrons développée par Wachters est employée. Elle est accessible en utilisant le mot-clé 6-311G* dans le logiciel GAUSSIAN03. Elle a été développée par Wachters (Wachters 1970) et modifiée par Hay (Hay 1977), puis par Raghavachari et Trucks (Raghavadi 1989). Cette base contractée [14s9p5d1f/9s5p3d1f] est explicitée en annexe. Nous n'avons effectué aucune modification ni réoptimisation des exposants de gaussiennes.
- Pour l'iode, la base utilisée est associée à un pseudo-potentiel, le tout étant décrit dans le paragraphe suivant (Partie II.A.2) de ce chapitre.

Dans certains cas, en particulier pour les calculs de solvatation, nous avons inclus à BS1 une fonction diffuse pour tous les atomes sauf les hydrogènes. Cette extension, notée BS1+, permet de tenir compte du caractère diffus des thiolates anioniques (Chapitre IV). Cette base apporte donc une meilleure description dans la région de valence (à de grandes distances du noyau). De même, l'ajout d'une seconde fonction de polarisation sur tous les atomes sera notée BS1*.

b. Les bases étendues (utilisées pour les calculs d'énergie finale)

Pour le calcul final des énergies, nous avons utilisé la base standard de Pople 6-311+G(2d,2p) (notée BS2) pour tous les atomes autres que l'iode et le zinc. C'est une base triple- ξ pour les couches de valence, à laquelle sont ajoutées deux fonctions de polarisation sur tous les atomes et des fonctions diffuses s et p pour les atomes lourds.

Pour le zinc, la base notée BS2 correspond à la base étendue de Wachters, contractée [15s11p6d2f/10s7p4d2f] (Wachters 1970). Cette base est explicitée en annexe. Nous n'avons effectué aucune modification ni réoptimisation des exposants de gaussiennes.

2. Les potentiels effectifs ou pseudo-potentiels de cœur

Pour l'atome d'iode, calculer l'énergie en considérant tous les électrons allonge le temps de calcul. De plus, les électrons les plus internes atteignent des vitesses telles qu'ils ne peuvent plus être décrits de façon raisonnable sans prendre en compte les effets de la relativité restreinte.

La description des électrons de cœur est alors remplacée par un potentiel déterminé de façon à ce que les couches de valences soient les plus proches possibles de ce que produirait un calcul incluant tous les électrons, prenant en compte les principaux termes relativistes. Ce potentiel est appelé potentiel effectif ou pseudo-potential. La valence est représentée par une base associée au potentiel de cœur.

a. La base initiale (utilisée pour l'optimisation de géométries)

La base, notée BS1, et le pseudo-potential associé, utilisés pour les optimisations de géométrie et les calculs de fréquence sont ceux développés par Christiansen *et al.* (Hurley 1986). Le potentiel de cœur, appelé CRENBL, ne « gèle » que 36 électrons et est parfois qualifié de « shape-consistent » puisqu'il permet de maintenir la partie radiale externe des orbitales atomiques dans la région de valence et l'énergie monoélectronique associée.

Cette base de valence d'extension (111/111/1111) et le potentiel de cœur de l'iode sont explicités en annexe.

b. La base étendue (utilisée pour les calculs d'énergie finale)

Pour les simples points d'énergie, on utilisera le pseudo-potential ECP28MDF développé récemment par l'équipe de Stuttgart (Peterson 2006). Ce pseudo-potential de « petit cœur » (il ne gèle que 28 électrons) prend en compte l'effet relativiste à un haut niveau de qualité.

La base associée, notée ECP28MDF_AVTZ par l'équipe de Stuttgart, ou encore aug-cc-pVTZ-PP, est contractée [13s12p10d2f/6s4p4d2f] (Peterson 2006). Elle sera notée BS2.

B. Les bases utilisées par TURBOMOLE

TURBOMOLE utilise les fonctions de base de Karlsruhe développées par l'équipe d'Ahlrichs (Schäfer 1992). Elles ont été développées pour la série d'atomes H-Rn en 1992 à partir de fonctions de bases CGTO (*Contracted Gaussian-type Orbitals*) optimisées. Les fonctions de base employées en conjonction avec des pseudopotentiels de cœurs (ECP) sont ainsi généralement de qualité double- ξ (split valence – SV) et triple- ξ (TZV), augmentées de fonctions de polarisation. Des exemples typiques sont les bases LANL2DZ associées aux pseudopotentiels de Hay and Wadt et les bases de Stuttgart.

Les bases SV et TZV de TURBOMOLE ne sont que des versions légèrement modifiées et réoptimisées de ces dernières (Eichkorn 1997, Weigend 1998).

Cette famille de base a ensuite été remodifiée en 2005 pour donner les bases plus performantes def2-SV(P)/SVP/TZVP/TZVPP/QZVP/QZVPP. Ces bases sont les dernières implémentées dans TURBOMOLE et ont été testées sur plus de 300 systèmes représentatifs (Weigend 2005). Les recommandations pour les calculs sont les suivantes :

Tableau II.1 : Recommandations du manuel des utilisateurs de TURBOMOLE pour l'utilisation des bases

QUALITE DU CALCUL	Exploratoire	Qualitative	Quantitative	Limite
DFT		def2-SV(P)	def2-TZVP	def2-QZVP
HF		def2-SVP	def2-TZVPP	def2-QZVPP
MP2	def2-SVP	def2-TZVPP	def2-QZVPP	

Parmi ces bases, celles recommandées pour les calculs DFT ont été utilisées dans nos calculs et sont comparées ci-dessous aux bases Pople utilisées avec GAUSSIAN.

1. Les bases initiales (utilisées pour l'optimisation)

Pour les calculs d'optimisation d'énergie et de fréquences, nous utiliserons la base def2-SVP que nous noterons PB. Elle est composée, pour l'atome de carbone, de cinq primitives pour la zone de cœur, de deux gaussiennes pour l'orbitale 2s, de deux contractions de respectivement trois et une primitives pour les orbitales 2p et d'une orbitale d de polarisation (soit (511/31/1)). On constate que les orbitales s sont décrites par plus de primitives dans la base 6-31G* de Pople que dans la base def2-SVP, cependant les coefficients des orbitales 2s de def2-SVP sont différents de ceux des 2p (alors que ce sont les mêmes coefficients pour 6-31G*). Par ailleurs, les coefficients et exposants des gaussiennes représentant les orbitales 2p des deux bases sont très proches. On peut donc considérer que les bases 6-31G* et def2-SVP sont approximativement de même qualité.

2. Les bases étendues (utilisées pour les calculs d'énergie finale)

La base étendue utilisée est def2-TZVPP, notée GB. Sa comparaison avec la base 6-311+G(2d,2p) utilisée par GAUSSIAN conduit à une conclusion similaire à celle du paragraphe précédent : les bases TZVPP et 6-311+G(2d,2p) se décomposent respectivement en (62111/411/11/1) et (63111/3111/11). La dernière orbitale s contractée de TZVPP correspond à l'orbitale s diffuse de 6-311+G(2d,2p) et la gaussienne de polarisation f est une orbitale diffuse. Par ailleurs, les gaussiennes représentant les orbitales d de polarisation sont rigoureusement identiques. On a donc des bases de qualité voisine pour TURBOMOLE et GAUSSIAN03.

III. Méthodes de chimie quantique

A. Introduction

Il existe une grande diversité de méthodes de chimie quantique. Le but des premières méthodes développées, dites *ab initio*, est de résoudre l'équation de Schrödinger indépendante du temps :

$$\mathbf{H}\Psi_j^{tot} = E_j^{tot}\Psi_j^{tot}$$

où $\mathbf{H}$ est l'opérateur Hamiltonien d'un système composé de n électrons (de masse m_i et de position r_i) et de N noyaux (de masse M_k , de position R_k et de charge Z_k). Les états propres stationnaires du système sont caractérisés par les couples $(E_j^{tot}, \Psi_j^{tot})$. Cependant, la résolution analytique de cette équation est impossible en général et l'on procède donc à différentes approximations. Dans toutes les méthodes considérées ici, on procède d'une part à l'approximation de Born-Oppenheimer qui considère les noyaux fixes. En effet les noyaux étant beaucoup plus lents que les électrons, le principe est de séparer les variables et de rechercher Ψ_j^{tot} comme le produit d'une fonction nucléaire $\Xi(R_k)$ et d'une fonction électronique $\Psi_j^{el}(R_k, r_i)$. La fonction électronique est obtenue en résolvant une équation dite électronique, où la position des noyaux est fixe. D'autre part, pour déterminer la fonction d'onde, il est nécessaire de trouver les OM Φ_i (orbitales moléculaires, cf ci-dessous) sur lesquelles est bâtie $\Psi_j^{el}(R_k, r_i)$. On procède alors à l'approximation CLOA (*Combinaison Linéaire d'Orbitales Atomiques*) où chaque OM est une combinaison linéaire de fonctions atomiques (χ_p). Pour des raisons pratiques, leurs parties radiales sont définies comme des combinaisons de gaussiennes.

On doit alors résoudre l'équation : $\mathbf{H}^{el}\Psi_j^{el} = E_j^{el}\Psi_j^{el}$ où $\mathbf{H}^{el} = \underbrace{\sum_i \left(-\frac{\Delta_i}{2} - \sum_{\mu} \frac{Z_{\mu}}{R_{i,\mu}} \right)}_{\text{Terme mono-électronique}} + \underbrace{\sum_i \sum_{j < i} \frac{1}{r_{i,j}}}_{\text{Terme bi-électronique}}$

$$= \sum_i h_i$$

La partie mono-électronique ne pose pas de grandes difficultés pour être calculée, contrairement au terme bi-électronique. On est alors obligé de faire de nouvelles approximations. Dans la suite, nous n'avons utilisé que des méthodes basées sur une description mono-déterminantale de la fonction d'onde électronique.

B. Méthode Hartree-Fock (HF)

Principe :

Dans le cadre de la méthode Hartree-Fock, la fonction d'onde du système est considérée comme le produit antisymétrisé et orthonormé des fonctions d'onde mono-électroniques Φ_i , donc comme le déterminant de Slater de ces Φ_i . Les Φ_i sont les fonctions propres de l'hamiltonien approché de Fock :

$$F(i) = h(i) + \sum_j^{\text{occupées}} ((ii|jj) - (ij|ij)) = h(i) + \sum_j^{\text{occupées}} J_{ij} - K_{ij}$$

où les intégrales bi-électroniques J_{ij} et K_{ij} sont nommées intégrales coulombiennes et intégrales d'échange, respectivement.

$$(ii|jj) = J_{ij} = \iint \phi_i^*(1) \phi_i(1) \frac{1}{r_{1,2}} \phi_j^*(2) \phi_j(2) d\tau_1 d\tau_2$$

$$(ij|ij) = K_{ij} = \iint \phi_i^*(1) \phi_j(1) \frac{1}{r_{1,2}} \phi_i^*(2) \phi_j(2) d\tau_1 d\tau_2$$

Dans l'approximation CLOA : $\phi_i = \sum_{p=1}^N c_{ip} \chi_p$, la décomposition de $F(i)$ sur les χ_p conduit aux équations de Roothaan : $\sum_p c_{ip} (F_{pq} - \varepsilon_i S_{pq}) = 0$ où $S_{pq} = \langle \chi_p | \chi_q \rangle$ et $F_{pq} = \langle \chi_p | F | \chi_q \rangle$.

Les inconnues du système sont les c_{ip} . Cependant pour résoudre les équations de Roothaan, il est nécessaire de connaître ces c_{ip} qui interviennent dans le terme F_{pq} . La solution est de résoudre ce problème de manière itérative, à partir d'un jeu d'orbitales d'essai $(\Phi_i^1, \varepsilon_i^1)$. La résolution des équations de Roothaan conduit à un nouveau jeu d'orbitales $(\Phi_i^2, \varepsilon_i^2)$. Le processus itératif est stoppé lorsque l'énergie ou les c_{pq}^n sont stables, selon un critère de convergence fixé.

Conséquences :

D'une part, la résolution des équations de Roothaan fait apparaître des termes bi-électroniques à quatre centres : $\langle pq | rs \rangle$, dont le nombre croît comme la puissance 4 du nombre d'orbitales atomiques. Cela pose donc des problèmes techniques de résolution de ces équations.

D'autre part, l'expression mono-déterminantale sous forme d'un déterminant de Slater de la fonction d'onde implique que l'on considère les électrons comme des particules indépendantes, c'est à dire que le mouvement d'un électron n'a qu'un effet moyen sur le mouvement des autres électrons. On néglige donc la corrélation électronique. Il existe plusieurs solutions pour pallier cette approximation, dont la méthode de perturbation de Møller-Plesset à l'ordre 2 (MP2).

C. Méthodes Post-HF: MP2

Dans la méthode de perturbation Møller-Plesset d'ordre 2, la corrélation électronique est traitée en introduisant des configurations électroniques mono et di-excitées en plus de la configuration électronique fondamentale (Hartree-Fock). Pour minimiser la lourdeur des calculs, on traite le problème par perturbation au second ordre, l'hamiltonien non perturbé étant l'hamiltonien HF (H^0).

Il s'agit donc de considérer la différence entre l'hamiltonien exact H et H^0 dont on sait calculer les fonctions propres $|\psi_m^0\rangle$ et les énergies E_m^0 , comme une perturbation λh (le terme λ étant petit devant 1) :

$$H = H^0 + \lambda h$$

h est appelé opérateur de corrélation. On développe les fonctions propres $|\psi_m\rangle$ de H et les énergies E_m correspondantes en série de Taylor en puissance de λ :

$$\begin{aligned} |\psi_m\rangle &= |\psi_m^0\rangle + \lambda |\psi_m^1\rangle + \lambda^2 |\psi_m^2\rangle \\ E_m &= E_m^0 + \lambda E_m^1 + \lambda^2 E_m^2 \end{aligned}$$

En remplaçant ces deux termes par leur développement dans l'équation de Schrödinger, on constate que pour que celle-ci soit vérifiée, il faut que les facteurs des différentes puissances de λ soient nuls. On déduit que :

$$\text{Pour le terme en } \lambda : (H^0 - E_m^0) |\psi_m^1\rangle = (E_m^1 - h) |\psi_m^0\rangle$$

$$\text{Pour le terme en } \lambda^2 : (H^0 - E_m^0) |\psi_m^2\rangle = E_m^2 |\psi_m^0\rangle + (E_m^1 - h) |\psi_m^1\rangle$$

On aboutit ensuite aux corrections à l'énergie aux premier et second ordres, notées respectivement E_m^1 et E_m^2 :

$$E_m^1 = \langle \psi_m^0 | h | \psi_m^0 \rangle \quad E_m^2 = \langle \psi_m^0 | h | \psi_m^1 \rangle$$

La correction de l'énergie au premier ordre s'obtient simplement en calculant la valeur moyenne de l'opérateur de perturbation dans l'état non perturbé. En revanche, pour déterminer la correction au deuxième ordre, il faut connaître la fonction $|\psi_m^1\rangle$. Celle-ci peut se développer dans la base des fonctions non-perturbées $|\psi_n^0\rangle$:

$$|\psi_m^1\rangle = \sum_n C_{nm} |\psi_n^0\rangle$$

On peut alors montrer que dans le cadre de la théorie de Møller-Plesset, la correction d'énergie au second ordre s'exprime finalement comme :

$$E_0^2 = - \sum_{n \neq 0} \frac{(\langle \psi_0^0 | H | \psi_n^0 \rangle)^2}{E_n^0 - E_0^0}$$

La méthode de perturbation MP2 améliore l'énergie Hartree-Fock, mais elle a tendance à légèrement surestimer la correction d'énergie. Elle permet d'obtenir des résultats fiables lors de calculs de grandeurs dépendant peu de la corrélation, mais pour un coût très important, que ce soit en temps de calcul ou en mémoire. Dans notre étude, nous lui préférons le plus souvent les méthodes de fonctionnelle de la densité qui incluent également la corrélation électronique.

D. Méthodes de la DFT

Principe :

La théorie de la fonctionnelle de la densité (DFT) a été développée par P. Hohenberg, W. Kohn et L. J. Sham (Hohenberg 1966, Kohn 1965). Ils ont démontré que toutes les propriétés d'un système dans un état fondamental non dégénéré sont complètement déterminées par sa densité électronique $\rho(\vec{r})$. L'énergie totale du système apparaît alors comme une fonctionnelle de cette densité électronique. De plus ils ont étendu le principe variationnel à la densité : $\frac{\partial E(\rho(\vec{r}))}{\partial \rho(\vec{r})} = 0$, ce qui permet de trouver la

densité électronique, puisque la fonction $\rho(\vec{r})$ exacte correspond au minimum d'énergie. Le problème cependant est que la fonctionnelle exacte de l'énergie est inconnue. On est alors amené à faire des approximations.

L'énergie peut être décomposée en quatre termes : le terme d'énergie cinétique des électrons, un terme d'énergie potentielle d'attraction noyau-électron, un terme de répulsion coulombienne entre électrons et un terme d'énergie d'échange et de corrélation E_{xc} . La fonctionnelle d'énergie cinétique a été résolue par Kohn et Sham par l'introduction de fonctions mono-électroniques. Le problème se ramène alors à la résolution d'une équation mono-électronique très analogue à celle de Hartree-Fock :

$$E = \sum_{i=1}^N \langle \phi_i | -\frac{1}{2} \Delta | \phi_i \rangle + \int \rho(\vec{r}) V_{Ne}(\vec{r}) d\vec{r} + \frac{1}{2} \iint \frac{\rho(\vec{r}_1) \rho(\vec{r}_2)}{r_{1,2}} d\vec{r}_1 d\vec{r}_2 + E_{xc}[\rho]$$

Le terme de répulsion électron-électron développé sur la base des orbitales de Kohn et Sham fait apparaître l'opérateur J précédemment décrit dans le principe de la méthode HF :

$$J = \iint \frac{\rho(\vec{r}_1) \rho(\vec{r}_2)}{r_{1,2}} d\vec{r}_1 d\vec{r}_2 = \sum_i \sum_j (ii|jj) = \sum_i \sum_j J_{ij} = \sum_i \sum_j \iint \phi_i^*(1) \phi_i(1) \frac{1}{r_{1,2}} \phi_j^*(2) \phi_j(2) d\tau_1 d\tau_2$$

Les différentes générations de fonctionnelles :

Il existe trois générations de fonctionnelles qui présentent une sophistication croissante du terme d'échange et de corrélation.

1. L'approximation de la densité locale (LDA)

La première génération de type LDA (*Approximation de la Densité Locale*) considère qu'en un point, la contribution à l'énergie ne dépend que de la densité en ce point. Elle possède une partie d'échange basée sur le calcul du gaz uniforme d'électrons. La forme du potentiel d'échange-corrélation $\left(\frac{\partial E_{xc}[\rho]}{\partial \rho(\vec{r})} \right)$ est alors un opérateur multiplicatif, il est donc plus simple que l'opérateur d'échange K qui est intégral.

Ce type de fonctionnelle a permis d'interpréter de nombreuses propriétés. Cependant, elle a atteint ses limites dans les années 80.

2. L'approximation du gradient généralisé (GGA)

La seconde génération comporte les fonctionnelles GGA (*Approximation du Gradient Généralisé*) qui dépendent de la densité et de son gradient (par exemple : BLYP, BP86...). Les développements actuels incluent également un terme dépendant du laplacien de la densité. On parle alors de fonctionnelles méta-GGA.

L'utilisation de l'approximation GGA permet d'obtenir des géométries et des énergies de liaison plus correctes que LDA. Les énergies de liaison restent toutefois surestimées en GGA.

Dans la méthode BP86 que nous utiliserons avec le logiciel TURBOMOLE, l'énergie d'échange est décrite par les fonctionnelles PW86 de Perdew et Wang (Perdew 1986), VWN de Vosko, Wilk et Nusair (Vosko 1980) et B88 de Becke (Becke 1988) et l'énergie de corrélation par la fonctionnelle de Lee-Yang-Parr (Lee 1988) qui inclut les termes de gradient de densité.

Une troisième génération a été développée pour pallier les défauts des deux autres types de fonctionnelles, notamment pour mieux décrire les liaisons faibles.

3. Les fonctionnelles hybrides

Elles prennent en compte explicitement le terme d'échange Hartree-Fock, calculé au moyen des orbitales de Kohn et Sham. La fonctionnelle hybride la plus utilisée est B3LYP. Dans cet algorithme, la fonctionnelle d'échange-corrélation E_{xc}^{B3LYP} combine la fonctionnelle d'échange exacte Hartree-Fock (E_x^{HF}) à deux autres fonctionnelles d'échange : la fonctionnelle locale 'standard' E_x^{LDA} et la fonctionnelle de correction de gradient ΔE_x^{B88} proposée par Becke (Becke 1988). A ces trois termes s'ajoutent deux fonctionnelles de corrélation : la fonctionnelle de corrélation locale de Vosko E_c^{VWN} (Vosko 1980) et celle proposée par Lee-Yang-Parr E_c^{LYP} (Lee 1988). L'expression finale de E_{xc}^{B3LYP} est donc :

$$E_{xc}^{B3LYP} = (1 - a_0)E_x^{LDA} + a_0E_x^{HF} + a_x\Delta E_x^{B88} + a_cE_c^{LYP} + (1 - a_c)E_c^{VWN}$$

où les coefficients $a_0=0.2$, $a_x=0.72$ et $a_c=0.81$ ont été proposés par Becke.

La fonctionnelle hybride B3LYP donne généralement d'excellents résultats sur les structures électroniques et les propriétés associées. Nous avons donc systématiquement utilisé ce niveau de calcul pour l'optimisation des molécules avec le logiciel GAUSSIAN, parfois en le confrontant à d'autres fonctionnelles ou à la méthode MP2.

Même si les fonctionnelles hybrides ont su prouver qu'elles étaient très efficaces pour les déterminations structurales, les fonctionnelles de type GGA sont aussi de bonnes alternatives, surtout pour de gros systèmes, puisqu'elles permettent un gain non négligeable en temps de calcul. Ainsi, nos fragments de protéine ont été optimisés avec la méthode BP86 à l'aide du logiciel TURBOMOLE.

IV. Les approximations de la résolution de l'identité

Pour les études de fragments protéiques, nous avons choisi, après différents tests décrits dans le chapitre VI, d'utiliser la fonctionnelle BP86, couplée à la méthode d'approximation de « résolution de l'identité accélérée par les multipôles ». Cette double approximation est décrite dans les paragraphes suivants.

A. Principe de l'approximation RI

Dans le cas des fonctionnelles GGA, le terme coulombien est le plus difficile à calculer, car comme dans le cas de la méthode HF, il fait intervenir des termes bi-électroniques à quatre centres. L'approximation RI « résolution de l'identité » permet de réduire notablement le temps de calcul en se limitant à des intégrales à trois centres au maximum. Cette technique a été développée par R. Ahlrichs *et al.* (Eichkorn 1995, Von Armin 1998) et implémentée dans le logiciel TURBOMOLE que nous avons utilisé. Cette approximation consiste à développer la densité électronique ρ sur un jeu de bases auxiliaires α étendues et centrées sur les atomes. L'intérêt de cette méthode est que la dimension de la base auxiliaire N^{aux} est inférieure à N^2 . On a alors :

$$\rho(\vec{r}) = \sum_{k,l}^{N,N} D_{kl} k(\vec{r}) l(\vec{r}) \approx \tilde{\rho}(\vec{r}) = \sum_{\alpha}^{N^{aux}} c_{\alpha} \alpha(\vec{r}), \text{ où } D_{kl} \text{ sont les éléments de la matrice densité. Les}$$

coefficients c_{α} sont déterminés de manière à minimiser le terme $(\rho - \tilde{\rho} | \rho - \tilde{\rho}) = (\rho - \tilde{\rho} | J | \rho - \tilde{\rho})$.

On trouve alors :

$$\frac{\partial (\rho - \tilde{\rho} | \rho - \tilde{\rho})}{\partial c_{\alpha}} = 0 \Leftrightarrow \sum_{\beta}^{N^{aux}} c_{\beta} (\alpha | \beta) = \sum_{k,l}^{N,N} D_{kl} (\alpha | kl)$$

La résolution de ce système d'équations linéaires, qui n'impliquent que des intégrales bi-électroniques à deux et trois centres, donne accès aux c_{α} .

Une fois les coefficients c_{α} déterminés, on peut calculer l'énergie de coulomb J et ses éléments de matrice J_{kl} par les formules :

$$J = \frac{1}{2} (\rho | \rho) \approx \frac{1}{2} (\rho | \tilde{\rho}) = \frac{1}{2} (\tilde{\rho} | \tilde{\rho}) = \sum_{\alpha, \beta}^{N^{aux}, N^{aux}} c_{\alpha} c_{\beta} (\alpha | \beta) \text{ et } J_{kl} = \sum_{\alpha}^{N^{aux}} c_{\alpha} (\alpha | kl).$$

Dans l'approximation RI, il faut donc déterminer les N^{aux} coefficients c_{α} par la méthode ci-dessus et les N coefficients c_{ip} par le processus itératif identique aux méthodes de DFT classiques.

Remarque

Les intégrales bi-électroniques à quatre centres valent : $(ij|kl) \approx \sum_{\alpha, \beta}^{N^{aux}, N^{aux}} (ij|\alpha)(\alpha|\beta)^{-1}(\beta|kl)$. Ceci correspond mathématiquement à l'insertion d'une résolution de l'identité, d'où le nom de cette méthode. Remarquons que cette relation serait strictement vraie si $N^{aux}=N^2$, car c'est une formule de changement de base.

Conséquences

Le calcul des termes bi-électroniques à quatre centres peut être approximé par des calculs de termes bi-électroniques à trois centres et les termes $(\alpha|\beta)^{-1}$. Si la base auxiliaire a une taille raisonnable (les bases auxiliaires utilisées dans le programme TURBOMOLE ont une dimension $N^{aux} \approx 3N$), la matrice $(\alpha|\beta)^{-1} = (\alpha|J^{-1}|\beta)$ peut être stockée en mémoire. Il ne sera donc pas nécessaire de la recalculer à chaque itération. Ahlrichs *et al.* ont montré que pour de grands systèmes la détermination des coefficients c_α varie en N^2 (Sierka 2003).

B. Principe de l'approximation MARI-J

Récemment, une approximation supplémentaire (*Multipôle Accelerated Resolution of Identity J*) a été ajoutée à la résolution de l'identité pour les méthodes DFT du gradient de la densité (Sierka 2003).

Cette approximation s'inspire des méthodes FMM (*Fast Multipôle Method*) (Greengard 1987, White 1994) qui consistent à réduire le temps de calcul en partitionnant les interactions coulombiennes du système en deux parties : Les interactions de courtes portée (Near Field, NF) sont calculés précisément, alors que les interactions à longue portée (Far Field, FF) sont généralement calculées par expansion des multipôles, afin de réduire le temps de calcul. Ces interactions à longue distance constituent la majorité des interactions dans une molécule de grande taille, telle qu'une protéine.

De la même façon que les méthodes basées sur l'expansion des multipôles, l'approche MARI-J partitionne l'interaction coulombienne de la densité électronique en deux portions distinctes. Le calcul de la contribution à courte portée s'appuie sur la technique RI-J et la densité électronique ρ est donc approximée par une combinaison linéaire de fonctions auxiliaires de bases centrées sur les atomes, comme expliqué précédemment (cf Chapitre II, Partie IV.A).

Quant à la portion à longue distance, elle est évaluée par expansions de multipôles centrés sur des atomes. Le terme coulombien est alors basé sur la formulation de l'approximation des multipôles donnée par White et Head-Gordon (White 1994). Celle-ci propose l'expression du terme

$$\frac{1}{|R - (a + b)|} \text{ en polynômes de Legendre.}$$

$\sigma^O(r)$ est une distribution de charge continue centrée sur un point Q et composée de K charges de distributions primitives. L'interaction coulombienne entre deux distributions de charges $\sigma^O(r)$ et $\sigma^R(r)$, une centrée sur O et la deuxième sur un point distant R, peut alors être factorisée de la façon suivante :

$$J_{OR} = \frac{1}{2} \iint \frac{\sigma^O(r_1) \sigma^R(r_2)}{|r_1 - r_2|} = \iint \frac{\sigma^O(a) \sigma^R(R-b)}{|R - (a+b)|} da db = \frac{1}{2} \sum_{l=0}^{\infty} \sum_{m=-l}^l \Omega_{lm}^O \Xi_{lm}^R$$

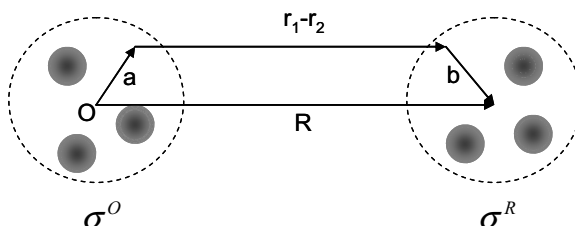


Figure II.1 : Représentation schématique de la factorisation d'une interaction coulombique entre deux distributions de charge continues σ^O et σ^R centrées en O et en un point distant R

Les moments multipolaires Ω_{lm}^O et Ω_{lm}^R centrés sur O et R sont les sommes de moments multipolaires de distributions primitives de charges et Ξ_{lm}^R sont les coefficients d'un développement local en série de Taylor du potentiel dû à une distribution de charge distante de R.

L'interaction coulombienne des distributions de charge est alors exprimée à l'aide de quantités indépendantes que sont les moments multipolaires Ω_{lm} . Une fois calculés, ils peuvent être utilisés pour calculer l'interaction coulombienne avec toutes les autres distributions.

L'avantage de cette factorisation est évident : si l'on considère deux groupes distants de M distributions de charges, la sommation directe des intégrales de Coulomb sur $M^2 K^2$ distributions primitives est remplacée par la sommation des produits des moments multipolaires et des coefficients de Taylor sur les centres de distribution σ . Pour des valeurs suffisamment grandes de K, l'approximation multipolaire est beaucoup plus rapide que l'évaluation directe des charges.

V. Les méthodes de solvation

A. Introduction

La plupart des processus chimiques, notamment les réactions enzymatiques, s'effectuent en solvant et il est donc important de considérer l'influence de la solvation sur le comportement d'un système. Dans certains cas, les molécules de solvant sont directement impliquées et doivent alors être modélisées de façon explicite. Dans d'autres systèmes, le solvant n'agit pas directement sur le réactif

mais fournit un environnement qui va influencer fortement le comportement des solutés. Le solvant est alors généralement qualifié de « bulk medium », c'est-à-dire qu'il agit seulement comme une déviation du système par rapport à son comportement en phase gazeuse. Ses propriétés générales, notamment sa constante diélectrique, sont alors suffisantes pour le décrire et il n'est pas utile dans ce cas d'inclure explicitement chaque molécule de solvant dans le système. C'est le but des modèles de solvation appelés « continuum » (Smith 1994).

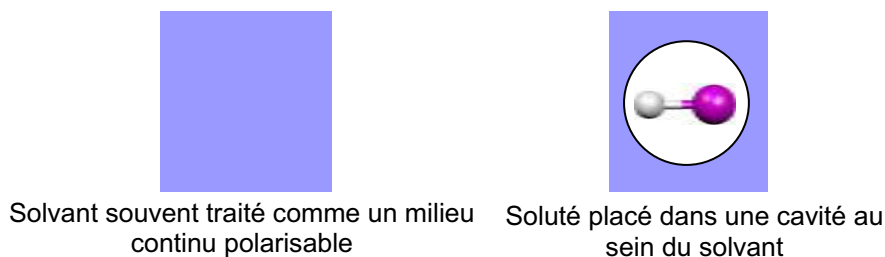


Figure II.II : Méthodes de solvation traditionnelles

B. Généralités thermodynamiques

L'énergie de solvation est dans tous les cas, définie comme la quantité d'énergie libre requise pour transférer une molécule du gaz au solvant. L'énergie libre de solvation ΔG_{solv} peut être décomposée en trois contributions (Leach 2001):

$$\Delta G_{\text{solv}} = \Delta G_{\text{elec}} + \Delta G_{\text{vdw}} + \Delta G_{\text{cav}}$$

ΔG_{elec} mesure l'interaction entre le soluté et le solvant, modélisé comme un milieu uniforme de constante diélectrique ϵ . Cette contribution est particulièrement importante dans le cas de solutés polaires et/ou chargés. ΔG_{vdw} est l'interaction de Van der Waals entre la molécule et le solvant, qui peut être décomposée en un terme répulsif ΔG_{rep} et un terme de dispersion attractif ΔG_{disp} . Enfin, ΔG_{cav} est l'énergie libre requise pour créer la cavité du soluté dans le solvant. Cette dernière contribution est positive et comprend la pénalité entropique associée à la réorganisation des molécules du solvant autour de la cavité, ainsi que le travail exercé contre la pression du solvant pour former la cavité. A ces trois termes peut s'ajouter une contribution ΔG_{hb} tenant compte des liaisons hydrogène existant entre le soluté et le solvant.

Les termes énergétiques ΔG_{vdw} et ΔG_{cav} sont souvent combinés et calculés d'après la formule suivante :

$$\Delta G_{\text{vdw}} + \Delta G_{\text{cav}} = \gamma S + b$$

S est la surface accessible du solvant et γ et b sont des constantes. La dépendance linéaire par rapport à la surface S peut être expliquée de la façon suivante : l'interaction ΔG_{vdw} est proportionnelle au nombre de molécules de la première couche de solvation (l'interaction de Van der Waals diminue rapidement au-delà) et donc approximativement à la surface accessible du solvant. De même, le terme de cavitation ΔG_{cav} implique la pénalité entropique des molécules de solvant entourant le soluté, dont le nombre est proportionnel à S .

Les différents modèles de solvation diffèrent surtout par leur méthode de traitement de la composante électrostatique ΔG_{elec} . Les deux premiers travaux sur la solvation ont été les modèles développés par Born (en 1920) et Onsager (en 1936).

C. Les modèles de Born et d'Onsager

Le modèle le plus simple est celui de Born qui considère une charge ponctuelle placée dans une cavité de solvant de forme sphérique (Born 1920). Onsager l'étend à un dipôle placé dans une cavité sphérique (Onsager 1936). Alors que le modèle de Born n'est approprié que pour les espèces chimiques possédant une charge formelle, celui d'Onsager est applicable à de nombreuses molécules.

Le modèle d'Onsager représente un cas spécial du modèle de Kirkwood, qui considère une distribution arbitraire des charges au sein de la cavité sphérique (Kirkwood 1934).



Figure II.III

Dans le modèle d'Onsager, le moment dipolaire du soluté induit un moment dipolaire du solvant. En retour, la polarisation du milieu induit une polarisation du soluté, c'est-à-dire influe sur les orbitales moléculaires du soluté. Ce champ de réaction créé par le solvant sur le soluté (*Reaction Field* = RF) est donné par :

$$\vec{R}_{RF} = \frac{2(\epsilon - 1)}{(2\epsilon + 1)d^3} \vec{\mu} \text{ où } d \text{ est le rayon de la cavité sphérique du solvant.}$$

La méthode la plus simple pour évaluer ΔG_{elec} est alors la méthode de *Self-Consistent Reaction Field* (SCRF) qui consiste à considérer le champ de réaction comme une perturbation du système :

$$\hat{H} = \hat{H}_{\text{vide}} + \hat{H}_{RF} \text{ avec } \hat{H}_{RF} = -\vec{\mu}^T \vec{R}_{RF} \vec{\mu} \langle \Psi | \vec{\mu} | \Psi \rangle$$

et $\hat{H}_{\text{vide}}$ l'opérateur hamiltonien de la molécule dans le vide.

L'énergie de solvation ΔG_{elec} est alors calculée comme :

$$\Delta G_{\text{elec}} = \langle \Psi | \hat{H} | \Psi \rangle - \langle \Psi_{\text{vide}} | \hat{H}_{\text{vide}} | \Psi_{\text{vide}} \rangle + \frac{1}{2} \frac{2(\epsilon - 1)}{(2\epsilon + 1)d^3} |\vec{\mu}|^2.$$

L'inconvénient d'un tel modèle est l'utilisation d'une cavité sphérique, ce qui est très peu réaliste.

D. La méthode PCM

Principe :

L'approche prise par la méthode PCM (*Polarisable Continuum Method*) tient compte d'une polarisation auto-cohérente entre le soluté et le solvant ainsi que d'une cavité de forme plus réaliste : la cavité du

soluté est définie à partir de sphères de van der Waals centrées sur chaque atome (Miertus 1981, Barone 1996 et 2004, Cossi 2002). La surface de la cavité est alors découpée en un grand nombre de petits éléments de surface appelés *tesserae* (*tessera*=tuile).

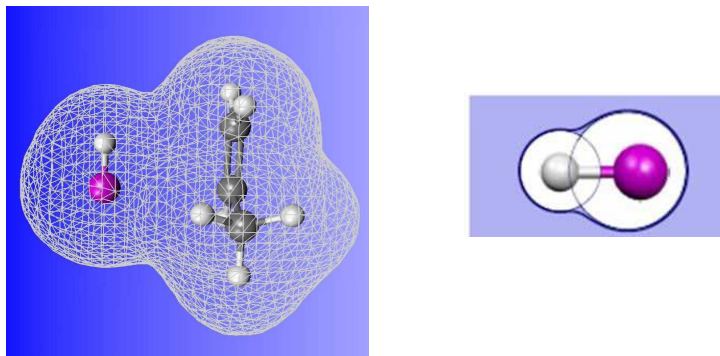


Figure I.IV : Découpe de la surface de la molécule en tesserae

L'énergie ΔG_{elec} est alors calculée en plaçant une charge ponctuelle polarisable au centre de chaque *tessera*. L'ensemble de ces charges ponctuelles représente la polarisation du solvant et l'amplitude de chaque charge surfacique est proportionnelle au gradient du champ électrique en ce point. Le potentiel électrostatique total sur chaque élément de surface est égal à la somme du potentiel dû au soluté ϕ_{solute} et du potentiel dû aux autres charges surfaciques ϕ_{charges} , soit $\phi(r) = \phi_{\text{solute}}(r) + \phi_{\text{charges}}(r)$.

L'algorithme PCM débute par le calcul d'une valeur initiale de chaque charge ponctuelle pour chaque *tessera* à partir du gradient du champ électrique causé par le soluté seul :

$$q_i = - \left[\frac{\epsilon - 1}{4\pi\epsilon} \right] E_i \Delta S$$

où E_i est le gradient du champ électrique et ΔS la surface de la *tessera* considérée

La contribution des autres charges surfaciques peut alors être calculée à partir de la loi de Coulomb. Ces charges sont modifiées de façon itérative jusqu'à leur auto-cohérence. Le potentiel $\phi_{\text{charges}}(r)$ est alors ajouté à l'hamiltonien du soluté ($H = H_0 + \phi_{\text{charges}}(r)$) et le calcul SCF débute. Après chaque cycle SCF, des nouvelles valeurs des charges surfaciques sont obtenues à partir de la nouvelle fonction d'onde pour donner une nouvelle valeur de $\phi_{\text{charges}}(r)$, qui est utilisée dans l'itération suivante jusqu'à ce que la fonction d'onde du soluté et les charges surfaciques convergent.

Pour calculer l'énergie ΔG_{elec} , il faut ensuite tenir compte du travail effectué par la création de la distribution des charges au sein de la cavité dans le milieu diélectrique. Ceci correspond à la moitié de l'énergie de l'interaction électrostatique entre la distribution de charges du soluté et le milieu diélectrique polarisé, soit :

$$\Delta G_{\text{elec}} = \int \Psi H \Psi d\tau - \int \Psi_0 H_0 \Psi_0 d\tau - \frac{1}{2} \int \phi(r) \rho(r) dr$$

Limites du modèle :

Un premier problème apparaît du fait de représenter une distribution de charges continue sur la surface de la cavité par un ensemble de charges ponctuelles. L'utilisation de surfaces (tesserae) de petite taille (typiquement 0.2 Å) permet de limiter l'influence de ce problème.

Un autre inconvénient de ce modèle est l'extension de la fonction d'onde du soluté en-dehors de la cavité. Ainsi la somme des charges sur la surface n'est pas strictement égale et de signe opposé à la charge du soluté.

*E. Les approches COSMO et C-PCM**1. La théorie COSMO**Principe :*

La théorie COSMO (*Conductor-Like Screening Model*) est une variante de l'approche PCM, basée sur le pouvoir d'écrantage des conducteurs (Klamt 1993, 1995 et 1998). La cavité est considérée comme entourée d'un solvant idéal présentant une constante diélectrique infinie, ce qui a l'avantage de mener à des écrantages très forts pour lesquels les calculs sont plus aisés à traiter. Le champ de réaction du solvant est en effet représenté par une distribution de charges de polarisation apparentes sur la surface de la cavité, qui sont déterminées en imposant une condition aux limites très pratique : le potentiel électrostatique surfacique doit être égal à zéro, ce qui revient à dire que le potentiel dû au soluté doit être équivalent à celui dû aux charges surfaciques.

Conséquences :

Cette condition d'annulation permet une simplification des calculs, notamment dans l'expression des gradients de l'énergie et permet donc de traiter des systèmes plus complexes que le modèle PCM. Cependant, le modèle du conducteur est moins fondé que celui du diélectrique et est surtout valable pour des milieux très polaires ($\epsilon \rightarrow \infty$). Pour l'eau dont la constante électrique est élevée ($\epsilon=78,4$), il suffit de corriger légèrement les données obtenues avec le conducteur pour obtenir des résultats fiables. Pour une autre constante diélectrique, les charges surfaciques sont corrigées avec le facteur

$$\text{suivant : } q' = q \frac{\epsilon - 1}{\epsilon + 0,5}.$$

Un autre problème existant avec cette stratégie est l'existence d'une fraction de la densité électronique du soluté à l'extérieur de la cavité pré-définie. Il faudra alors effectuer une correction sur l'énergie pour tenir compte de cette « charge extérieure ».

Implémentations :

La méthode COSMO est utilisée dans le logiciel TURBOMOLE (Klamt 1993). Afin d'obtenir une meilleure construction de la cavité, COSMO construit des segments hexagonaux, pentagonaux et triangulaires, à partir d'une union de sphères construites autour de chaque atome et dont le rayon est déterminé par l'utilisateur. Celui-ci peut utiliser des valeurs optimisées pour chaque atome et prédéfinies dans le

logiciel à partir de données expérimentales. Un facteur supplémentaire concerne un rayon externe pour la construction d'une seconde cavité permettant d'inclure le nuage électronique du soluté sortant de la première cavité.

2. La théorie C-PCM

La méthode C-PCM (*Conductor-like PCM*) est une extension du modèle PCM où l'approche COSMO a été implémentée dans GAUSSIAN94 (Barone 1998, Cossi 2003). Par défaut, le programme construit la cavité en utilisant le modèle UA0 (*United Atom*) qui place une sphère autour de chaque atome lourd du soluté (les hydrogènes sont inclus dans la sphère de l'atome auquel ils sont liés). La cavité UA0 emploie le modèle topologique UATM (*United Atom Topological Model*) appliqué sur les rayons atomiques du champ de force UFF. La méthode a ensuite été améliorée dans GAUSSIAN03 avec le choix d'une série de modèles de construction de la cavité. Basé sur le modèle A0, les cavités UAHF et UAKS utilisent des rayons optimisés avec les niveaux de théories HF/6-31G* et PBE0/6-31G* respectivement (Barone 1997). De plus, dans la plus récente version de PCM implémentée dans GAUSSIAN, le logiciel tient compte implicitement du problème des charges extérieures à la cavité (Cossi 2002). Par contre, même si celle-ci est théoriquement possible, l'optimisation de la géométrie s'est révélée problématique avec le modèle C-PCM et n'a donc pas été effectuée.

F. Calculs effectués en solvant

1. Avec GAUSSIAN

Pour le calcul des énergies de solvation ΔG_{solv} de chaque composé, nous avons utilisé la méthodologie recommandée par l'équipe de Houk (Takano 2005). Les valeurs de ΔG_{solv} sont ainsi déterminées avec un simple point d'énergie au niveau B3LYP/BS1+ sur des structures préalablement optimisées au niveau B3LYP/BS1.

La cavité utilisée est UAKS et la surface moyenne par *tessera* est la valeur par défaut du logiciel, soit $0,2 \text{ \AA}^2$.

2. Avec TURBOMOLE

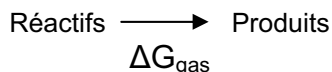
L'énergie des molécules solvatées peut être calculée avec un simple point d'énergie B3LYP/GB sur une molécule préalablement optimisée en phase gazeuse ou optimisée directement en solvant au niveau BP86/PB (ou dans certains cas B3LYP/GB) avec l'approximation MARI-J.

Le paramètre de solvation le plus important pour TURBOMOLE est le rayon élément utilisé pour chaque élément lors de la construction de la cavité. Ceux-ci sont déterminés interactivement par le script *cosmoprep*, et comme préconisé par le logiciel, les rayons optimisés ont été choisis pour tous les atomes, à l'exception de l'atome de zinc. Cet élément n'ayant pas de rayon optimisé, le rayon non-optimisé correspondant a été utilisé.

VI. Les grandeurs thermodynamiques

A. En phase gazeuse

Considérons la réaction :



L'énergie électronique de la réaction est la grandeur que l'on obtient directement à partir des calculs *ab initio*. Il suffit pour l'obtenir de faire la différence entre l'énergie des produits et l'énergie des réactifs, au niveau de calcul considéré : $E_{\text{él}} = E_{\text{él}}(\text{prod}) - E_{\text{él}}(\text{réact})$.

1. *Correction au point-zéro*

L'énergie de vibration d'une molécule à 0K, phénomène quantique, n'est pas nulle. Par conséquent, une fois le point stationnaire localisé (minimum ou état de transition), son énergie est inférieure à l'énergie réelle de la molécule. L'énergie doit être corrigée en ajoutant à la valeur obtenue l'énergie du niveau vibrationnel le plus bas ou *ZPVE* (« Zero Point Vibrational Energy ») définie comme :

$$ZPVE = \sum_{i=1}^n \frac{1}{2} h \nu_i$$

où h est la constante de Planck et ν_i les n fréquences des modes normaux de vibration. La valeur de *ZPVE* est obtenue par un calcul des fréquences dans l'approximation harmonique. Elle est ensuite corrigée par un terme de « scaling » propre à la méthode utilisée (0.9804 pour les calculs GAUSSIAN/B3LYP/BS1 et 0.9914 pour les calculs TURBOMOLE/BP86/PB).

2. *Variation d'énergie interne à T*

La variation d'énergie interne $\Delta U(0)$ de la réaction à 0 K est obtenue en ajoutant la variation d'énergie de vibration au point zéro à l'énergie de la réaction $E_{\text{él}}$:

$$\Delta U(0) = E_{\text{él}} + \Delta ZPVE$$

Pour déterminer la variation d'énergie interne à la température T à partir de l'énergie de réaction $E_{\text{él}}$ à 0 K, il faut y ajouter un terme ΔE_{th} que nous nommerons variation d'énergie thermique à T .

$$\Delta U(T) = E_{\text{él}} + \Delta E_{\text{th}}(T)$$

$$\text{avec } \Delta E_{\text{th}}(T) = E_{\text{th}}(T)(\text{prod}) - E_{\text{th}}(T)(\text{réact})$$

La valeur de $E_{th}(T)$ est obtenue pour chaque espèce chimique de la réaction par un calcul de fréquences. Notons que le terme d'énergie thermique $E_{th}(T)$ prend en compte le point zéro et ne requiert aucune correction.

3. Variation d'enthalpie à T

Nous pouvons maintenant facilement relier la variation d'énergie interne à la variation d'enthalpie :

$$\Delta H(T) = \Delta U(T) + \Delta(PV)$$

Il nous suffit d'évaluer le terme $\Delta(PV)$. Pour les réactions se produisant en phase gazeuse, nous pouvons nous placer dans l'approximation des gaz parfaits dont l'équation d'état est $PV=nRT$.

Le terme $\Delta(PV)$ est alors proportionnel à RT . Le coefficient de proportionnalité correspond à la variation du nombre de moles de produit par rapport à celui des réactifs lors de la correction : $\Delta n = n(\text{prod}) - n(\text{réact})$

Ainsi, dans le cas d'une association, nous aurons : $\Delta n = -1$ et $\Delta H(T) = \Delta U(T) - RT$ et dans le cas d'une dissociation : $\Delta n = 1$ et $\Delta H(T) = \Delta U(T) + RT$.

4. Variation d'énergie libre à T

Pour déterminer celle-ci, on utilise la relation : $\Delta G_{\text{gas}} = \Delta G(T) = \Delta H(T) - T\Delta S(T)$.

B. En phase aqueuse

Pour calculer l'énergie libre de la réaction en phase aqueuse (ΔG_{eau}), on utilise le cycle thermodynamique suivant:

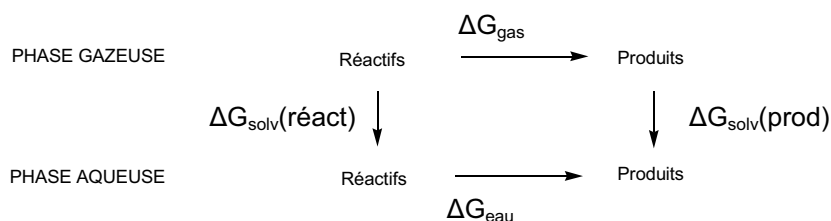


Schéma II.I

$\Delta G_{\text{solv}}(X)$ est l'énergie de solvation de chaque espèce chimique

$\Delta G_{\text{solv}}(X)$ est calculée préalablement suivant les méthodes détaillées dans la partie V de ce chapitre.

L'énergie libre de la réaction en phase aqueuse ΔG_{eau} peut ensuite être déduite à partir de la relation suivante:

$$\Delta G_{\text{eau}} = \Delta G_{\text{gas}} - \Delta G_{\text{solv}}(\text{réact}) + \Delta G_{\text{solv}}(\text{prod}).$$

Bibliographie

- Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, 162, 165.
- Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, 255, 327.
- Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, 107, 3210.
- Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, 102, 1995.
- Barone, V.; Improta, R.; Rega, N. *Theor. Chem. Acc.* **2004**, 111, 237.
- Becke, A. D. *J. Chem. Phys.* **1993**, 98, 1372.
- Born, M. *Zeitschrift für Physik.* **1920**, 1, 45.
- Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, 24, 669.
- Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, 117, 43.
- Eichkorn, K.; Treutler, O.; Ohm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, 240, 283.
- Eichkorn, K.; Treutler, O.; Ohm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, 242, 652.
- Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, 97, 119.
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; Gaussian, Inc., Wallingford CT, Gaussian 03, Revision C.02, **2004**.
- Greengard, L.; Rokhlin, V. *J. Comput. Phys.* **1987**, 73, 325.
- Hay, P. J. *J. Chem. Phys.* **1977**, 66, 4377.
- Hehre, W. J.; Radom, L.; Schleyer, P. V. R.; Pople, J. A. *Ab initio molecular orbital theory*, Wiley, **1986**.
- Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, 136 (3B), B864.
- Hurley, M. M.; Pacios, L. F.; Christiansen, P. A.; Ross, R. B.; Ermler, W. C. *J. Chem. Phys.* **1986**, 84, 6840.
- Kirkwood, J. G. *J. Chem. Phys.* **1934**, 2, 351.
- Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans 2* **1993**, 799.
- Klamt, A. *J. Phys. Chem.* **1995**, 99, 2224.
- Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem.* **1998**, 102, 5074.
- Leach, A. R. *Molecular Modelling. Principles and Applications*, 2nd Edition, Pearson Education, **2001**.
- Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev.* **1988**, B37, 785.
- Miertus, S.; Scrocco, E.; Tomasi, J. *J. Chem. Phys.* **1981**, 55, 117.

- Onsager, L. *J. Am. Chem. Soc.* **1936**, 58, 1486.
- Perdew, J. P.; Wang, W. *Phys. Rev.* **1986**, B33, 8800.
- Peterson, K. A.; Shepler, B. C.; Figgien, D.; Stoll, H. *J. Phys. Chem. A* **2006**, 110, 13877.
- Raghavachari, K.; Trucks, G. W. *J. Chem. Phys.* **1989**, 91, 1062.
- Schäfer, A.; Horn, H.; Alrichs, R. *J. Chem. Phys.* **1992**, 97, 2571.
- Sierka, M.; Hoge Kamp, A.; Ahlrichs, R. *J. Chem. Phys.* **2003**, 118, 9136.
- Smith, P. E.; Pettitt, B. M. *J. Phys. Chem.* **1994**, 98, 9700.
- Takano, Y.; Houk, K. N. *J. Chem. Theor. Comput.* **2005**, 1, 70.
- Von Armin, R. A. *J. Comput. Chem.* **1998**, 19, 1746.
- Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, 58, 200.
- Wachters, A. J. H. *J. Chem. Phys.* **1970**, 52, 1033.
- Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, 294, 143.
- Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, 7, 3297.
- White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1994**, 101, 6593.

ANNEXES**Bases initiale (BS1) et étendue (BS2) du Zinc**

<u>BS1</u>		<u>BS2</u>	
Zn 0		Zn 0	
S 6 1.00		S 6 1.00	
316336.0000	0.0006494033039	316336.0000	0.0006494033039
48561.00000	0.004911112485	48561.00000	0.004911112485
11157.40000	0.02518467188	11157.40000	0.02518467188
3205.010000	0.09870930219	3205.010000	0.09870930219
10685.80000	0.3004910850	1068.580000	0.3004910850
396.3940000	0.6600778707	396.3940000	0.6600778707
S 1 1.00		S 1 1.00	
159.8060000	1.000000000	159.8060000	1.000000000
S 1 1.00		S 1 1.00	
68.58900000	1.000000000	68.58900000	1.000000000
S 1 1.00		S 1 1.00	
23.70810000	1.000000000	23.70810000	1.000000000
S 1 1.00		S 1 1.00	
10.03720000	1.000000000	10.03720000	1.000000000
S 1 1.00		S 1 1.00	
2.810430000	1.000000000	2.810430000	1.000000000
S 1 1.00		S 1 1.00	
1.069640000	1.000000000	1.069640000	1.000000000
S 1 1.00		S 1 1.00	
1469510000	1.000000000	0.1469510000	1.000000000
S 1 1.00		S 1 1.00	
0.05114200000	1.000000000	0.05114200000	1.000000000
P 5 1.00		P 5 1.00	
2213.180000	0.003573790107	2213.180000	0.003573790107
527.0500000	0.02852211875	527.0500000	0.02852211875
172.2930000	0.1295976328	172.2930000	0.1295976328
66.08140000	0.3663134859	66.08140000	0.3663134859
27.68630000	0.5939311787	27.68630000	0.5939311787
P 1 1.00		P 1 1.00	
12.18410000	1.000000000	12.18410000	1.000000000
P 1 1.00		P 1 1.00	
4.987960000	1.000000000	4.987960000	1.000000000
P 1 1.00		P 1 1.00	
2.057910000	1.000000000	2.057910000	1.000000000
P 1 1.00		P 1 1.00	
0.7986090000	1.000000000	0.7986090000	1.000000000
D 3 1.00		D 3 1.00	
58.40840000	0.05696697377	58.40840000	0.05696697377
16.44920000	0.3095914479	16.44920000	0.3095914479
5.575700000	0.7624985047	5.575700000	0.7624985047
D 1 1.00		D 1 1.00	
1.884410000	1.000000000	1.884410000	1.000000000
D 1 1.00		D 1 1.00	
0.5723050000	1.000000000	0.5723050000	1.000000000
F 1 1.00		S 1 1.00	
1.6200000000	1.000000000	0.01534300000	1.000000000
		P 1 1.00	
		0.1624550000	1.000000000
		P 1 1.00	
		0.04776900000	1.000000000
		D 1 1.00	
		0.1620000000	1.000000000
		F 1 1.00	
		3.2400000000	1.000000000
		F 1 1.00	
		0.8100000000	1.000000000

Base initiale (BS1) et pseudopotentiel associé de l'iode

<u>BS1</u>			
I 0		I 0	
S 1 1.00		I-ECP 3 36	
0.680300000 1.000000000		F POTENTIAL	
S 1 1.00		4	
0.520800000 1.000000000		2 2.00659990 -2.60678697	
S 1 1.00		2 5.41200018 -15.30154037	
0.114300000 1.000000000		2 20.88789940 -73.32185364	
P 1 1.00		1 63.82109833 -23.43061638	
1.12500000 1.000000000		S-F POTENTIAL	
P 1 1.00		6	
0.350800000 1.000000000		2 1.53009999 -144.29801941	
P 1 1.00		2 1.85660005 252.49822998	
0.116600000 1.000000000		2 2.58430004 -200.15640259	
D 1 1.00		2 3.75419998 85.98131561	
3.95020000 1.000000000		1 1.13870001 32.24239349	
D 1 1.00		0 12.47010040 6.45168209	
2.57420000 1.000000000		P-F POTENTIAL	
D 1 1.00		6	
1.30710000 1.000000000		2 0.91109997 32.43734360	
D 1 1.00		2 1.06579995 -101.57602692	
0.511600000 1.000000000		2 1.38989997 169.25859070	
		2 1.79579997 -105.27653503	
		1 2.54789996 12.12060356	
		0 1.17250001 6.45733500	
		D-F POTENTIAL	
		6	
		2 0.86530000 3.41730189	
		2 0.89960003 -3.97894001	
		2 4.58720016 -37.31042862	
		2 5.52820015 77.44996643	
		1 22.20800018 18.01509857	
		0 19.14010048 6.86909294	

Base étendue (BS2) et pseudopotentiel associé de l'iode

<u>BS2</u>			
I 0		I 0	
S 13 1.00		ECP28MDF,4,28	
5546.5000 0.00015600000		G-Komponente	
838.21400 0.00098600000		1	
182.18700 0.0027920000		2 1.000000 0.000000	
31.212300 -0.043251000		S-G POTENTIAL	
19.531400 0.23413400		3	
8.2409900 -0.75094300		2 40.033376 49.989649	
2.1945500 0.88296800		2 17.300576 281.006556	
1.1091100 0.46206100		2 8.851720 61.416739	
0.37464100 0.022286000		P-G POTENTIAL	
0.17708000 -0.0043530000		4	
0.081061000 0.0011020000		2 15.720141 67.416239	
0.10010000 0.00000000		2 15.208222 134.807696	
0.041200000 0.00000000		2 8.294186 14.566548	
S 13 1.00		2 7.753949 28.968422	
5546.5000 -0.000073000000		D-G POTENTIAL	
838.21400 -0.00050800000		4	
182.18700 -0.0011580000		2 13.817751 35.538756	
31.212300 0.012193000		2 13.587805 53.339759	
19.531400 -0.087854000		2 6.947630 9.716466	
8.2409900 0.33820000		2 6.960099 14.977500	
2.1945500 -0.57655000		F-G POTENTIAL	
1.1091100 -0.40929800		4	
0.37464100 0.56745900		2 18.522950 -20.176618	
0.17708000 0.61248900		2 18.251035 -26.088077	
0.081061000 0.14323100		2 7.557901 -0.220434	
0.10010000 0.00000000		2 7.597404 -0.221646	
0.041200000 0.00000000			
S 1 1.00			
0.37464100 1.000000000			

S 1 1.00	
0.17708000 1.00000000	
S 1 1.00	
0.10010000 1.00000000	
S 1 1.00	
0.041200000 1.00000000	
P 12 1.00	
188.98800 0.00058500000	
21.286800 0.036923000	
10.039600 -0.23532400	
3.4518000 0.34148300	
1.9745600 0.53478800	
1.0242000 0.26514100	
0.44943700 0.025787000	
0.18664800 0.00052200000	
0.073481000 0.00060600000	
0.59810000 0.00000000	
0.096180000 0.00000000	
0.035900000 0.00000000	
P 12 1.00	
188.98800 -0.00025600000	
21.286800 -0.011682000	
10.039600 0.083192000	
3.4518000 -0.15697000	
1.9745600 -0.22451800	
1.0242000 -0.11445100	
0.44943700 0.37535600	
0.18664800 0.57513600	
0.073481000 0.24591700	
0.59810000 0.00000000	
0.096180000 0.00000000	
0.035900000 0.00000000	
P 1 1.00	
0.59810000 1.00000000	
P 1 1.00	
0.096180000 1.00000000	
P 1 1.00	
0.035900000 1.00000000	
D 9 1.00	
132.66200 0.00057200000	
37.605400 0.0044020000	
10.389100 -0.040922000	
6.4901700 0.099661000	
3.4545100 0.32266300	
1.8441300 0.40034300	
0.96247800 0.26830600	
0.47285300 0.084847000	
0.19320000 0.0076320000	
D 1 1.00	
0.47285300 1.00000000	
D 1 1.00	
0.19320000 1.00000000	
D 1 1.00	
0.078200000 1.00000000	
F 1 1.00	
0.40640000 1.00000000	
F 1 1.00	
0.18480000 1.00000000	

Chapitre III

Réactivité et mécanismes d'alkylation des complexes biomimétiques à zinc

I. Introduction

Dans cette thèse, nous nous sommes intéressés à la nouvelle classe de sites actifs à zinc ne possédant pas de molécule d'eau et subissant pourtant une alkylation au niveau d'un groupe thiolate lié au zinc (Parkin 2004). L'aspect structural du site, notamment la présence de liaisons hydrogène sur les thiolates (Smith 2002 et 2005) et la nature des ligands anioniques liés au zinc (Maynard 2001), semble avoir une influence importante sur la réactivité du site. Le mécanisme d'alkylation du thiolate n'a cependant jamais été mis en relation avec ces derniers résultats. Nous nous proposons d'établir un lien entre la structure d'un site actif à zinc et sa réactivité (nucléophilie et mécanisme).

A cette fin, nous avons utilisé une série de quatre complexes biomimétiques basés sur un squelette commun (pyrazolyl)borate (Parkin 1995, Trofimenko 1999) (Figure III.1). Ces complexes reproduisent l'arrangement tétraédrique des ligands autour du métal et modélisent la diversité chimique observée dans les sites actifs des métalloenzymes ou dans les complexes biomimétiques. Deux des complexes ont une charge totale nulle et sont équivalents aux sites $\text{Zn}(\text{Cys})_2(\text{His})_2$ ou à de nombreux complexes biomimétiques (Grappnerhaus 1998, Chiou 2000 et 2003, Warthen 2001, Ji 2005, Rombach 2006), tandis que les deux autres sont dianioniques, comme le sont les sites $\text{Zn}(\text{Cys})_4$ ou le complexe $[\text{Zn}(\text{SPh})_4]^{2-}$. De plus, deux des complexes présentent une liaison hydrogène sur l'atome de soufre réactif.

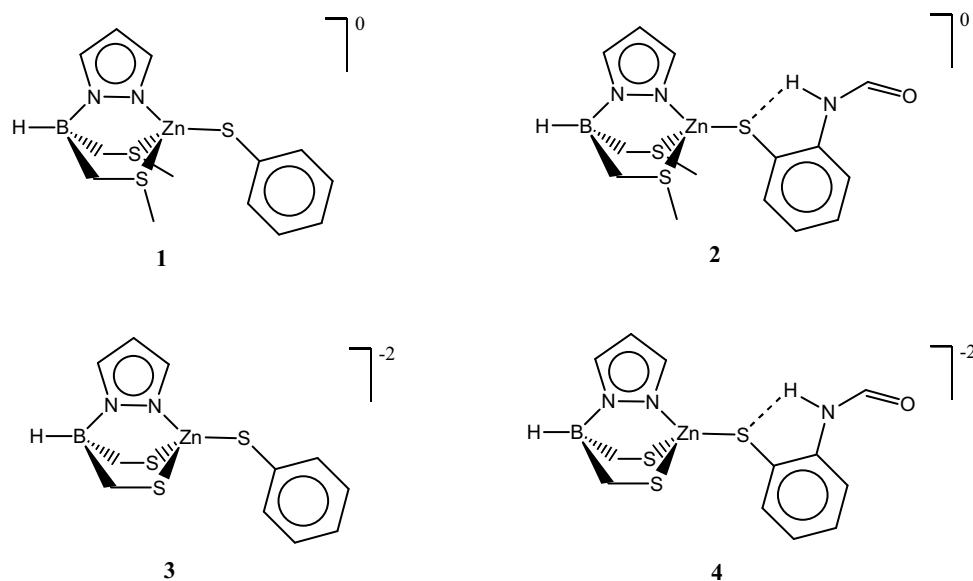


Figure III.1 : Complexes biomimétiques étudiés

Les réactions de cette série de complexes biomimétiques avec l'agent alkylant iodure de méthyle seront étudiées avec les méthodes DFT. Nous considérerons les trois chemins réactionnels d'alkylation du thiolate postulés dans la littérature : mécanismes associatifs de type $\text{S}_\text{N}2$ et à quatre centres et mécanisme dissociatif (Wilker 1997, Morlok 2005). Nous pourrions ainsi établir l'impact d'une liaison hydrogène impliquant un thiolate sur la nucléophilie du complexe, ainsi que l'influence de la nature des ligands, et donc de la charge totale du complexe sur le mécanisme de réaction. L'environnement structural du zinc sera ainsi mis en relation avec sa réactivité.

II. Résultats

Nous avons tout d'abord recherché l'influence des paramètres structuraux du complexe biomimétique sur sa nucléophilie, notamment l'influence d'une liaison hydrogène sur un thiolate coordonné au zinc. Nous avons montré que les complexes biomimétiques possédant ce type d'interaction avaient des barrières d'activation pour l'alkylation significativement plus importantes et étaient donc moins réactifs. Ceci est en accord avec les mesures cinétiques effectuées sur certains complexes biomimétiques (Chiou 2003) et confirme l'hypothèse indiquant que la nucléophilie du thiolate est réduite par sa participation à une liaison hydrogène. La présence d'une liaison hydrogène ne modifie cependant pas le mécanisme de la réaction.

La réactivité du complexe biomimétique est également accrue lorsque la charge totale négative est augmentée. Ainsi pour la réaction S_N2 associative, les complexes neutres étudiés présentent des barrières énergétiques d'alkylation plus élevées que pour les complexes dianioniques. Ainsi, plus la densité électronique d'un site à zinc est forte, c'est-à-dire plus le nombre de ligands cystéinates est important, plus le thiolate coordonné au métal sera nucléophile.

Les calculs sur les complexes biomimétiques ont montré que dans tous les cas, l'alkylation du thiolate, en phase gazeuse et aqueuse, s'effectue selon un mécanisme S_N2 . Pour les complexes neutres, le chemin réactionnel est associatif, alors que le complexe chargé négativement préférera se dissocier avant l'alkylation. La charge totale du complexe biomimétique a donc un impact considérable sur le mécanisme de réaction d'alkylation du thiolate. Ce résultat permet de rationaliser les données contradictoires issues de la littérature : ainsi, un mécanisme associatif était reporté pour les complexes neutres et positivement chargés (Grappnerhaus 1998, Bridgewater 2000, Brand 2001, Warthen 2001, Ibrahim 2005, Morlok 2005) alors que l'étude du composé $[Zn(SPh)_4]^{2-}$ était la seule à préconiser une dissociation du thiolate avant son alkylation (Wilker 1997).

Ces résultats nous apportent des informations intéressantes sur les sites actifs des enzymes. Ainsi, la nucléophilie du thiolate coordonné au métal sera influencée par la présence de liaisons hydrogène et pourra expliquer la sélectivité de l'alkylation sur un thiolate. Un exemple est la protéine Ada réparatrice de l'ADN où le soufre alkylé exclusivement dans le site $ZnCys_4$ est le seul ne possédant pas de liaison hydrogène (He 2005, Takinowaki 2006).

De même, les sites dianioniques, où toutes les cystéines seront déprotonées, sont supposés plus réactifs que les sites neutres. Ceci est cohérent avec les études rapportant l'ordre de réactivité des sites à zinc suivant : $ZnCys_4^{2-} > ZnCys_3His^- > ZnCys_2His_2$ (Wilker 1995, Maynard 2001).

De plus, nous pouvons supposer que les sites $ZnCys_4^{2-}$ réagiront de façon similaire à nos complexes dianioniques, selon un mécanisme dissociatif. Pour les complexes monoanioniques, les deux mécanismes semblent possibles car la présence d'un réseau de liaisons hydrogène pourrait réduire la charge négative du site et conduire à un mécanisme associatif. Un état de transition associatif a ainsi été proposé pour la protéine farnésyltransférase (Huang 2000). Pour les doigts de zinc de type $ZnCys_2His_2$, un mécanisme associatif sera vraisemblablement préféré.

III. Publication

The alkylation mechanism of zinc-bound thiolates depends upon the zinc ligands

Delphine Picot, Gilles Ohanessian, Gilles Frison*

Laboratoire des Mécanismes Réactionnels,
Département de Chimie, Ecole Polytechnique and CNRS,
91128 Palaiseau Cedex, France.

Fax: + 00 33 (0)1 69 33 48 03

gilles.frison@polytechnique.org

Abstract: Alkylation of zinc-bound thiolates occurs in both catalytic and structural zinc sites of enzymes. Recent biomimetic studies have led to a controversy as to which mechanism is operative in thiolate alkylation. Building on one of these biomimetic complexes, we have devised a series of models that allow for an appraisal of the roles of charge, ligand nature and hydrogen bonding to sulfur on reactivity. The reactions of these complexes with methyl iodide, leading to thioethers and zinc iodide complexes, have been examined by density functional theory (DFT) calculations, in the gas phase as well as in aqueous solution. In all cases, a S_N2 reaction is favoured over σ -bond metathesis. Both the net electronic charge and the H bond play a significant role on the nucleophilicity of the thiolate. We find that the mechanistic diversity observed experimentally can be explained by the difference in the net charge of the complexes. A dianionic complex follows a dissociative pathway whereas an associative one is preferred for a neutral system.

Introduction

Zinc is an essential constituent of more than 300 enzymes and plays both structural and functional roles in these biological systems.¹⁻⁵ The active site of the metallo-enzymes is typically characterized by a zinc-centered tetrahedral coordination, in which three or four ligands are protein residues for catalytic or structural sites respectively. Lateral chains of cysteine, histidine and to a smaller extent, of aspartic and glutamic acids, are the most common metal ligands.⁶⁻¹⁰ For a functional active site, the fourth ligand is usually a water molecule.

A class of active sites has started to emerge over the past fifteen years in which an alkylation reaction occurs on a zinc-bound thiolate.^{11, 12} The zinc-linked water molecule of a functional active site can indeed be displaced by a thiol which will then be alkylated after deprotonation. Metal coordination lowers the pK_a of the cysteine thiol to increase the concentration of the reactive thiolate at physiological pH, which functions then as the nucleophile in the alkylation step. This is the case for proteins such as farnesyl transferase^{13, 14} and geranyl geranyl transferase^{15, 16} responsible for alkyl-group transfers.

More interestingly, enzymes belonging to the structural class, thus lacking a zinc-bound water molecule in the enzyme resting state, are able to undergo the same kind of reactivity. In these cases, alkylation takes place either at a zinc-bound cysteine residue of the protein or at a thiol substrate bonded to zinc after displacement of a ligand of the metal. The first example to be characterized was the DNA repair protein Ada which possesses a Cys₄ zinc center in its N-terminal side. Among the four cysteines of its zinc core, Cys38 is specifically alkylated.^{17, 18} The methyl group transfer from the phosphotriester DNA backbone is irreversible and therefore, Ada is considered as a sacrificial agent for DNA repair.¹⁹⁻²¹ Other examples include the cobalamin-independent and cobalamin-dependent methionine synthase enzyme family, which methylates homocysteine to produce the amino acid methionine,^{22, 23} and the two retroviral HIV-1 nucleocapsid protein p7 (Ncp7) zinc fingers where one of the zinc core Cys thiolates can be alkylated, leading to inhibition of HIV-1.²⁴⁻²⁷

To study the nature and the reactivity of enzymatic cores, biomimetic models have been developed. The first prominent biological models were a series of $[ZnN_2S_2]^0$, $[ZnS_3]^-$ and $[ZnS_4]^{2-}$ complexes^{28, 29} including the dianionic complex $Zn(SPh)_4^{2-}$ (Figure 1a) used to model the $[Zn(S-Cys)_4]^{2-}$ active site of the Ada DNA repair protein.

Due to the high tendency of zinc complexes to polymerise,^{30, 31} most of the complexes developed since then reproduced the thiolate tetrahedral arrangement around the metal with macrocyclic or tripodal ligands with facial binding to the metal and one monodentate ligand. The basic building blocks of tripodal complexes are tris(pyrazolyl)borate^{32, 33} and tris(mercaptoimidazolyl)borate^{34, 35} in which the S arms exhibit electronic structures that are intermediate between those of thioketone and thiolate moiety.³⁶⁻³⁸ All the tripodal arms of these derivatives can be easily mixed to provide the desired number of nitrogen or sulfur ligand donors (Figure 1d,e).³⁹⁻⁴⁵ They may also be replaced by other nitrogen, oxygen or sulfur donor groups to provide tripodal ligands zinc complexes of the desired structures (Figure 1f,g).⁴⁶⁻⁴⁸ To the best of our knowledge, $\text{Zn}(\text{SPh})_4^{2-}$ and $\text{Zn}(\text{SPh})_3(\text{ImMe})^-$ (ImMe=Methyl imidazole) are the only negatively charged inorganic zinc species that have been used so far for biomimetic studies of thiolate alkylation, whereas other zinc thiolate complexes are all neutral (Figure 1c-g) or positively charged (Figure 1b).^{49, 50}

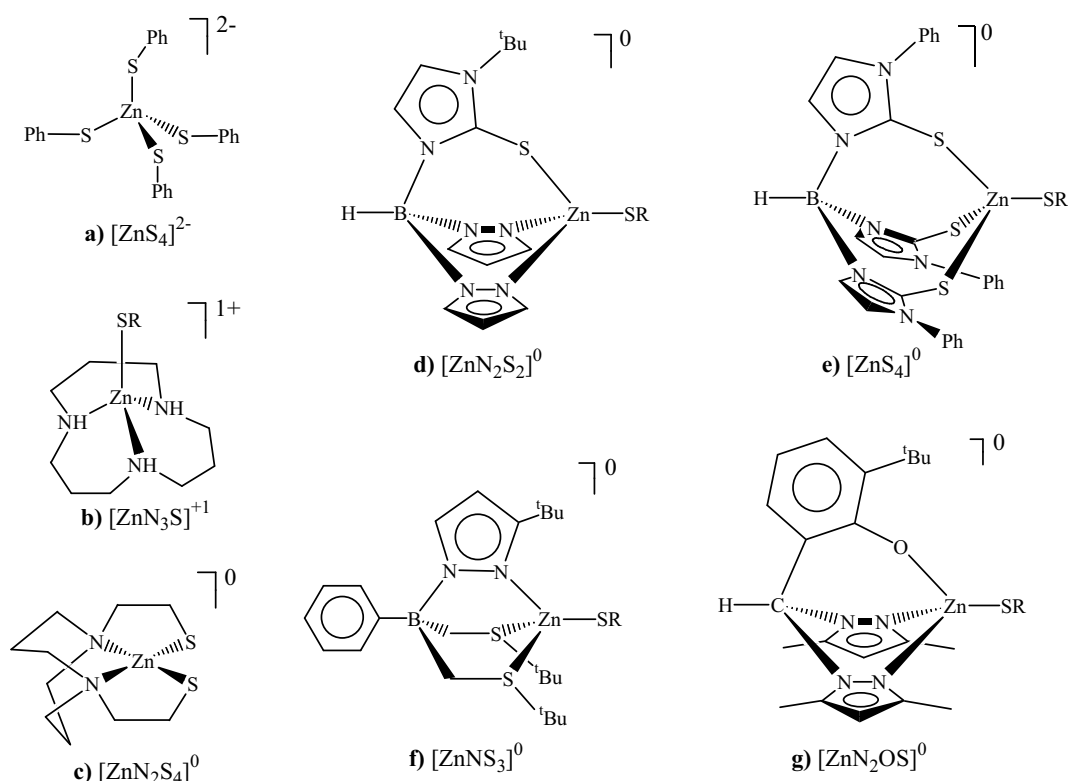


Figure 1. Examples of synthetic complexes miming the alkyl transfer activity of protein.

a) ref. ^{28, 29}; b) ref. ^{49, 50}; c) ref. ⁵¹; d) ref. ^{41, 45}; e) ref. ³⁹; f) ref. ^{46, 47}; g) ref. ⁴⁸.

The development of such biomimetic complexes has been a considerable progress in the study of biological thiolate alkylation. Indeed, this enzymatic reaction supplied by zinc

has raised numerous questions about its possible mechanisms and the relative reactivity of the thiolates located in the sphere of influence of zinc, which are difficult to answer based exclusively on enzyme studies.

A first question about enzymatic alkylation concerns the reactivity modulation of the zinc active sites due to the nature of the site itself and of its close environment. The impact of the ligand nature on the reactivity of the zinc-bound thiolate has been studied on scorpionate-zinc thiolate complexes containing diverse atom donors (N, O, S) (Figure 1g).⁴⁸ Analysis of kinetic data shows a significant increase of reaction rate, as the donor atom X in the $[N_2X]^-$ heteroscorpionate ligand is changed from O to N to S. Furthermore, experimental rates of reaction, obtained for a series of $[ZnN_3S]^0$, $[ZnN_2S_2]^0$, $[ZnNS_3]^0$, and $[ZnS_4]^0$ complexes derived from the tris(pyrazolyl)- and tris(mercaptoimidazolyl)borate (Figure 1d,e),⁴⁵ suggest that the nucleophilicity of the biomimetic complex is proportional to the number of sulfur donors of their ligands. This clearly indicates that thiolate reactivity is dependent upon the nature of the tripodal ligand. Both the nature of the zinc ligands and the electric charge are modified in the $[ZnS_4]^{2-}$, $[ZnS_3N]^-$ and $[ZnS_2N_2]^0$ series synthesized by Lippard and Wilker (Figure 1a),²⁹ who reported the following kinetic trend of alkylation ability: $Zn(SPh)_4^{2-} > Zn(SPh)_3(MeIm)^- \gg Zn(SPh)_2(MeIm)_2$.

The reactivity of “bare” zinc cores is thus dependent upon their ligands and their charges, but such sites are not found in proteins as they are embedded in a hydrogen bond network provided by the surrounding biological macromolecule which influence their reactivity.⁵²

This is for example the case in the HIV-1 nucleocapsid protein p7 (Ncp7) zinc finger. This domain contains two copies of a conserved “retroviral-type” sequence Cys-X₂-Cys-X₄-His-X₄-Cys. The zinc center of its C-terminal side is structurally similar to the one situated in its N-terminal side.⁵³ However, experimental studies reveal that the reaction is substantially faster for the former.^{25, 26} Furthermore, calculations⁵² indicate that a well-protected Cys₄ core could be less reactive than some Cys₃His or Cys₂His₂ cores. This is consistent with the finding that the Cys₃His C-terminal Ncp7 core is much more vulnerable to electrophiles than the Cys₄ core of GATA-1.⁵⁴

While several active sites can present different reactivities depending on their global environment, it has also been established that two thiolates from the same active site can

afford diverse nucleophilicities, according to their individual molecular interactions. N-H...S hydrogen bondings between sulfur atoms and backbone amide groups have been indeed proposed to influence directly the reactivity of the thiolate. This is the case in the Ada DNA repair protein, which is alkylated specifically on Cys38.¹⁸ As an explanation of this selectivity, the X-Ray structure shows that the three other zinc thiolate ligands of the enzyme (Cys42, Cys69 and Cys72) are engaged in hydrogen bondings, contrary to Cys38 which lacks potential hydrogen-bond donors.¹⁷ Kinetic studies on inorganic models have been conducted to measure the impact of a hydrogen bond on the thiolate. They showed that zinc-thiolate alkylation is markedly slower when the thiolate is involved in a hydrogen bond.^{42, 47, 55-59}

Another question has been raised about the mechanism of the sulfur alkylation which can in principle occur via a manifold of pathways. Three of them have emerged from previous studies, illustrated in Figure 2 with CH₃I as alkylating agent.^{12, 60}

In the first one, the dissociative S_N2 pathway (path a of Figure 2), the reactants first dissociate to give the reactive species **DISS**. The dissociated thiolate then adds to the alkylating agent CH₃I via an S_N2-like transition state (**TS1**) to form the methylated dissociated product. In the other nucleophilic pathway (path b of Figure 2), the alkylation reaction occurs directly between the zinc-bound thiolate (**REACT**) and methyl iodide. The S_N2-type transition state **TS2** leads to a zinc-bound thioether (**INT5**) which may dissociate from the metal to give, as in the other pathways, a free thioether and a zinc-iodine complex (**PROD**). The third pathway is based on a σ -bond metathesis reaction (path c of Figure 2). The reactant adds to the electrophile in a 4-center transition state (**TS3**). Thiolate alkylation and breaking of the zinc-thiolate bond occur simultaneously.

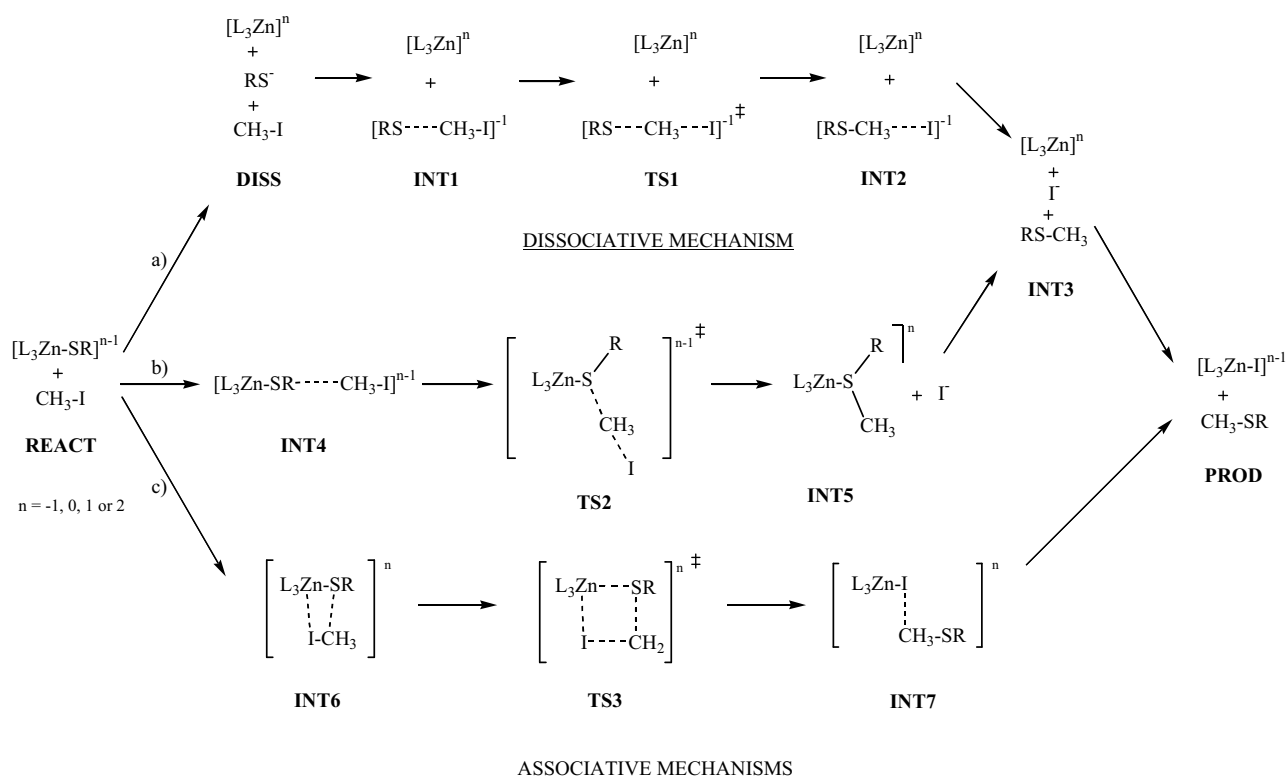


Figure 2. Possible mechanistic pathways for the alkylation of a zinc-bound thiolate by methyl iodide.

The results of numerous experiments on $[ZnN_nS_{4-n}]^0$ or $[ZnN_nS]^{1+}$ complexes (Figure 1) point to an alkylation with an associative mechanism on the basis of second order kinetics.^{45, 48, 50, 57, 61, 62} Furthermore, the temperature dependence of the alkylation is characterized by a small positive activation enthalpy and a large negative entropy whose values are in the range observed for reactions known to proceed by an associative pathway.^{48, 57}

Theoretical and experimental mechanistic studies on the farnesyl transferase enzyme or on the Ncp7 zinc finger are in agreement with this mechanism.^{13, 63-65} The most likely catalytic mechanism of the nucleophilic cysteine reaction with electrophilic reagents involves indeed a zinc-bound thiolate.

To the contrary, Lippard et al. report a dissociative alkylation of $Zn(SPh)_4^{2-}$ (Figure 1a).²⁹ The zinc complex is suggested to undergo appreciable degrees of dissociation before addition to the alkylating agent $(CH_3O)_3PO$ and thus, by the same way, that the Ada DNA repair enzyme alkylation proceeds via initial heterolytic dissociation generating a thiolate, being the active species. This is in agreement with recent kinetic isotope effects studies on a $[ZnS_4]^0$ complex⁵⁷ interpreted as indicating a dissociative mechanism but, based on DFT calculations, the authors cannot exclude thiolate alkylation before dissociation of the ligand.

Until now, there has been no study relating issues about the structure of the zinc complexes, their reactivity and their alkylation mechanisms. There is nevertheless a real need for a link between all of these aspects. This is the objective of this paper, in which we study the impact of structural components of zinc complexes, including hydrogen bonding on a thiolate, on the dissociative *vs* associative pathway for thiolate alkylation.

Chemical and computational models

1. Chemical models

We need biomimetic models that are appropriate for both structural and mechanistic studies. Based on the experimental studies summarized above, we search for a consistent series of zinc complexes in which the ligands, the electric charge, and the availability of H bonding to thiolate can be varied. In order to delineate the influence of each of these components, these complexes must be relatively rigid,⁶⁶ thus four-ligand models such as $[\text{Zn}(\text{SR})_n(\text{Im})_{4-n}]^{2-n}$ are avoided, but should keep the same steric obstruction of the zinc ion.⁵⁰ Based on these criteria, we chose the ((pyrazolyl)bis((methylthio)methyl)hydroborato)zinc complexes **1** and **2**, abbreviated $[\text{H}(\text{pz}^{\text{H}})\text{Bt}^{\text{Me}}]\text{Zn}(\text{SR})$ and depicted in Figure 3, which are $[\text{ZnNS}_3]^0$ simplified analogues of zinc complexes of phenyl(3-*tert*-butyl-pyrazolyl)bis-((*tert*-butylthio)methyl)borate (Figure 1f) that have been synthesized and characterized.^{46, 47}

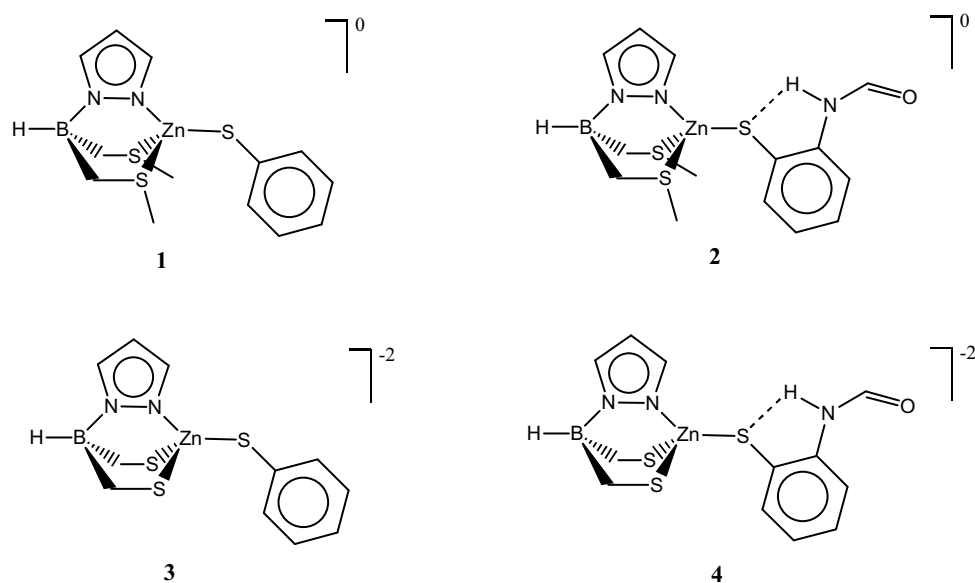


Figure 3. Biomimetic zinc complexes studied herein.

1 and **2** have a $[\text{ZnNS}_3]^0$ core which may be compared to neutral zinc enzyme active sites which include either $[\text{ZnN}_2\text{S}_2]^0$, with two histidines and two cysteinates bound to zinc, and $[\text{ZnNS}_3]^0$ and $[\text{ZnS}_4]^0$ with respectively one and two neutral cysteines bound to zinc, as considered in mass spectrometric^{67, 68} and theoretical⁶⁹ studies.

Other experimental⁷⁰⁻⁷² and theoretical^{52, 73} studies claim that all cysteines are deprotonated when bound to zinc, leading to $[\text{ZnNS}_3]^-$ and $[\text{ZnS}_4]^{2-}$ cores. In order to study the effect of ligand charge, we generate complexes **3** and **4**, abbreviated $[\text{H}(\text{pz}^{\text{H}})\text{Bt}^{2-}]\text{Zn}(\text{SR})$, $[\text{ZnNS}_3]^{2-}$ analogues of **1** and **2** by removing the sulfur methyls. Furthermore, the electronic richness and its consequence on the reactivity can then be measured effectively, by comparing neutral $[\text{ZnNS}_3]^0$ complexes **1** and **2** and negatively charged $[\text{ZnNS}_3]^{2-}$ complexes **3** and **4**.

The alkylating reagent chosen is methyl iodide, used in many previous experimental studies.^{41, 45, 48, 55, 57, 59} Its addition will be examined exclusively on the aryl thiolate of our models.

2. Level of computation

Calculations were performed with Gaussian 03.⁷⁴ Geometry optimisations were conducted using the B3LYP method at the 6-31G(d,p) level for the B, N, C, O, S, H atoms. The contracted Wachters basis [14s9p5d1f/9s5p3d1f] was used to describe the zinc atom.⁷⁵ The CRENBL relativistic effective core potential and associated valence basis set were employed to model the iodine atom.⁷⁶ This basis set is referred to as BS1.

Each stationary point has been characterized with frequency analysis and shows the correct number of negative eigenvalues (0 for a local minimum and one for a transition state). All transition states **TS1-3** (Figure 2) for all complexes were verified by stepping along the reaction coordinate (intrinsic reaction coordinates (IRC) calculations) and confirming that they transformed into the reactants/products **INT1-2**, **INT4-5** and **INT6-7**, respectively (Figure 2). Energies were calculated for the stationary points at the B3LYP level using an extended basis set labelled BS2. It consists in the 6-311+G(2d,2p) for B, N, C, O, S, H, the extended Wachters basis [15s11p6d2f/10s7p4d2f] for Zn and the Aug-cc-pVTZ-PP basis set and pseudo-potential for I.⁷⁷

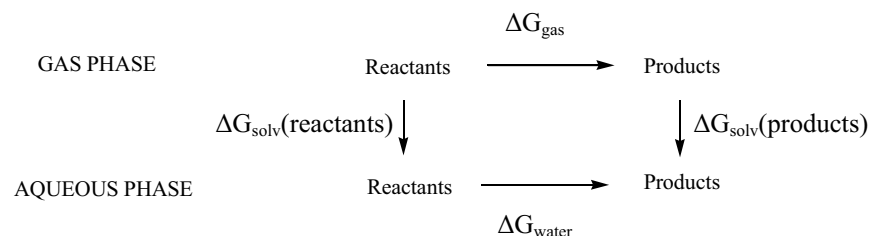
The gas phase Gibbs free energy corresponding to a given reaction was deduced from the equation:

$$\Delta G_{\text{gas}} = \Delta E_{\text{elec}} + \Delta \text{ZPE} + \Delta E_{\text{T}} - T\Delta S$$

with ΔE_{elec} , ΔZPE , ΔE_T and ΔS the differences in the electronic energy, zero-point vibrational energy, thermal energy and entropy between the products and the reactants, respectively.

The solvation free energy of each compound ΔG_{solv} was determined with the conductor polarized continuum method (CPCM).^{78, 79} The CPCM-SCRF calculations were carried out at the gas phase B3LYP/BS1 geometries at the B3LYP level with BS1+ derived from BS1 by addition of one set of diffuse functions on all atoms except H. The CPCM calculations were performed with tesserae of $0.2 \text{ \AA}^2$ size, water as solvent (dielectric constant $\epsilon=78.39$) and the UAKS cavity.⁸⁰

To calculate the reaction free energy in aqueous phase (ΔG_{water}), we used the following thermodynamic cycle:



ΔG_{gas} stands for the gas phase free energy computed previously and $\Delta G_{\text{solv}}(x)$, the free energy necessary for transferring a molecule X from the gas phase to the solvent. The free energy of the reaction in aqueous phase ΔG_{water} can then be deduced from equation (1).

$$\Delta G_{\text{water}} = \Delta G_{\text{gas}} - \Delta G_{\text{solv}}(\text{reactants}) + \Delta G_{\text{solv}}(\text{products}) \quad (1)$$

3. Evaluation of chemical model and level of computation

Before computational results can be reliably interpreted, the adequacy of the chemical modelling and of the level of computation must be assessed. Several calibration studies have been published recently for the modelling of zinc complexes.⁸¹⁻⁸⁷ To this end, we have compared the level of calculation described above, as well as the previously recommended SVWN/6-311++G(d,p)⁸³ to the available experimental data.^{46, 47} As can be seen in Table 1 (entries 1, 2 and 5), B3LYP outperforms SVWN for the structure of **1**: Zn-S bond lengths are closer to the crystallographic ones, with mean deviations of 0.032 and 0.076 Å for B3LYP/BS1 and SVWN/6-311++G(d,p), respectively. This vindicates the choice of B3LYP/BS1 for all geometry optimisations reported herein.

Modelling the t-butyl on pyrazole and the phenyl on boron by hydrogen atoms, as well as the tripod thiol t-butyls by methyls, does not have significant impact on the geometry around Zn, except on the relative positions of the tripod and phenyl thiolate ligands, since their steric repulsions are strongly diminished. This can be seen in Table 1, as measured by the N-Zn-S¹-C dihedral angle, which diminishes from 171.9 to 133.5° (Table 1, entries 1, 4 and 5). On the other hand, modelling the tripod thiol t-butyls by hydrogens leads to a strong lengthening of one of the Zn-S bonds (Zn-S³, Table 1, entry 3), accentuated by the rotation around the Zn-S¹ bond, and is therefore inadequate.

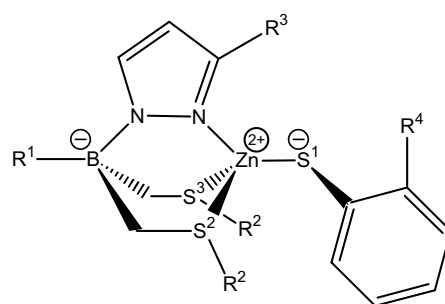


Figure 4. Labelling of the [ZnNS₃]⁰ zinc complexes.

Table 1. Main geometrical data for the [ZnNS₃]⁰ zinc model.

Entry	R ¹	R ²	R ³	R ⁴	Method	Zn-S ¹	Zn-S ²	Zn-S ³	Zn-N	N-Zn-S ¹ -C
1	H	Me	H	H	B3LYP/BS1	2.250	2.434	2.494	1.987	133.5
2	H	Me	H	H	SVWN/ 6-311++G(d,p)	2.205	2.332	2.386	1.967	143.2
3	H	H	H	H	B3LYP/BS1	2.238	2.471	2.561	1.972	117.6
4	H	^t Bu	^t Bu	H	B3LYP/BS1	2.273	2.470	2.489	1.999	163.9
5	Ph	^t Bu	^t Bu	H	Exp. ⁴⁶	2.265	2.418	2.441	2.033	171.9

See Figure 4 for the labelling of substituents R¹-R⁴ and sulfur atoms S¹-S³.

Results

1. Structure of complexes 1-4

Table 2. Main geometrical parameters of **1-4** (bond length in Å, bond angle in °)

Complex	Zn-S ¹	Zn-S ²	Zn-S ³	Zn-N	S ¹ ...HN ^a	Σ(L-Zn-L) _{tripod} ^b
1 ^c	2.250 (2.265)	2.434 (2.418)	2.494 (2.441)	1.987 (2.033)	-	286.8 (286.1)
2 ^d	2.262 (2.260)	2.423 (2.384)	2.490 (2.410)	1.984 (2.015)	2.991 (2.936)	288.2 (288.6)
3	2.458	2.324	2.320	2.141	-	309.8
4	2.496	2.309	2.309	2.129	2.946	313.1

^a distance between S¹ and N of the intramolecular hydrogen bond. ^b Sum of the N-Zn-S², N-Zn-S³ and S²-Zn-S³ angles. ^c exp. X-Ray data⁴⁶ in parentheses. ^d exp. X-Ray data⁴⁷ in parentheses.

Table 2 reports the structural characteristics of models **1-4**. In all cases, the arrangement of the four ligands around zinc is similar, corresponding to a distorted tetrahedron. The steric constraints of the tripodal ligand impose not only a rigid conformation of the ligand around the metal but also acute L-Zn-L (L : tripodal ligand arm) angles as reflected by the measure of the zinc pyramidity (last column of Table 2), lower than the 328.5° value of the tetrahedral geometry.

The complex **1** possesses two thiols linked to zinc with Zn-S bond lengths of 2.434 Å and 2.494 Å. The zinc-thiolate shows a much smaller bond length of 2.250 Å, reflecting the higher electron density of the thiolate. As already mentioned, these values, as well as the Zn-N bond length of 1.987 Å, are in agreement with the experimental ones.⁴⁶ For comparison, recent analyses of the Cambridge Structural Database indicate that the mean Zn-S bond length for dialkanethiols and phenylthiolate complexes are 2.55 Å⁶⁹ and 2.29 Å⁶⁰ respectively and that the Zn-N bond lengths are in the range 2.06-2.19 Å.⁶⁹ On this basis, the tripodal ligand seems to tighten the ligand-metal interactions. The Zn-S¹ bond length of **1** is slightly shorter than the average Zn-S bond length value of 2.34 Å determined from 22 ZnCys₄ binding sites found in the Protein Data Bank, assuming that most or all of the Cys sulfurs are anionic.^{81, 83} This could be explained by the effect of the enzyme surrounding (second coordination sphere, enzyme backbone, electrostatic environment,...) which distorts the metal sphere from its ideal geometry, thus weakening the corresponding bonds.

In our designed complex **3**, we transformed the two thiols of **1** into thiolates, leading to reduced Zn-S distances (from 2.434 to 2.324 Å and from 2.494 to 2.320 Å). This large increase in electron density around Zn leads to a Zn-S¹ lengthening of about 0.2 Å in **3** relative to **1**. The Zn-S¹ bond is now longer than those of the other two zinc-thiolates, because

the aryl thiolate can delocalize its charge over the phenyl group. This is not possible in the analogous methyl thiolate complex for which the Zn-S¹ bond length is computed to be 2.394 Å instead of 2.458 Å in **3** (data not shown). The increased electron density at Zn also leads to a Zn-N lengthening of about 0.16 Å in **3** relative to **1**.

When the phenyl thiolate is involved in a hydrogen bond as in **2** and **4**, the Zn-S¹ bond length slightly increases, and the Zn-S², Zn-S³ and Zn-N lengths slightly decrease concomitantly. This effect is almost insignificant in **2** vs. **1**, however it is larger in **4** vs. **3**, to the extent that the Zn-S¹ length in **4** is more typical of a thiol than a thiolate. This may have a significant influence on the thiolate nucleophilicity.

The larger effect of hydrogen bonding in **4** vs. **2** may be linked to the longer Zn-S¹ bond in **3** vs. **1** : the thiolate electrons are less involved in bonding to Zn in **3**, and therefore more available. It is also reflected in the hydrogen bond length which is much smaller in **4** (2.289 Å) than in **2** (2.378 Å).

2. Reaction mechanisms of complex 1

Three alkylation mechanisms have been studied as represented in Figure 2: the S_N2-type dissociative mechanism (path a), the S_N2-type associative mechanism (path b) and the 4-center associative one (path c). The optimised structures for all of these pathways for complex **1** are displayed in Figure 5.

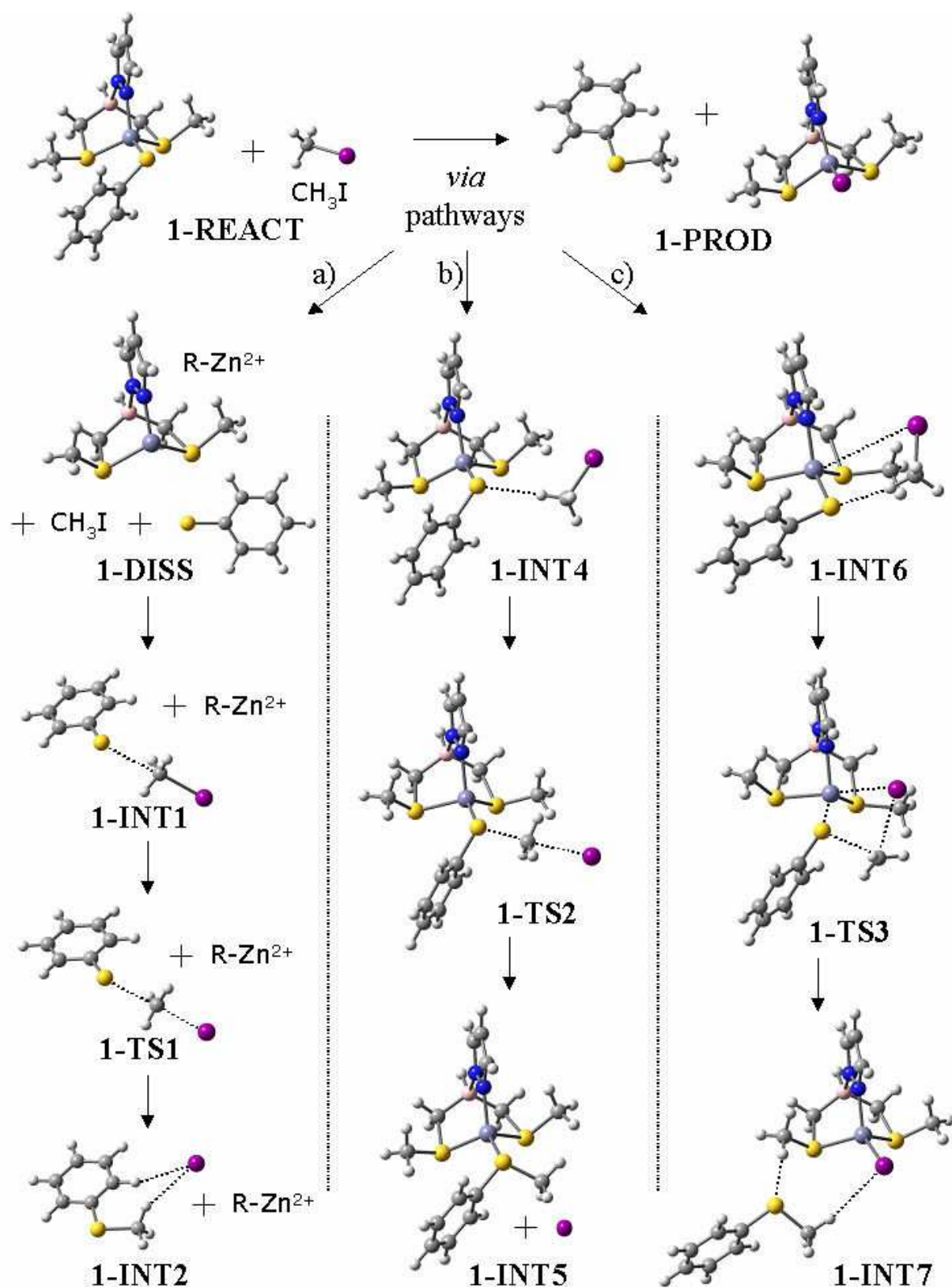


Figure 5. Geometrical structures along the reaction paths depicted in Figure 2 for complex 1. yellow : sulfur, blue : nitrogen, gray : carbon, white : hydrogen, purple : iodine, pink : boron, light blue : zinc.

The alkylation reaction is exothermic overall with a free energy release of -17.7 kcal/mol (Table 3). This corresponds to the methylation of the thiolate ligand and its detachment from the borato-zinc complex. In the final products, which is the global minimum on the calculated potential free energy surface, the iodide ion is complexed by the tripodal-ligated metal cation.

Table 3. Relative ΔG_{gas} and ΔG_{water} in kcal/mol calculated at the B3LYP/BS2//B3LYP/BS1 level using CPCM solvation model at the B3LYP/BS1+//B3LYP/BS1 level.

	1		2		3		4	
	ΔG_{gas}	ΔG_{water}	ΔG_{gas}	ΔG_{water}	ΔG_{gas}	ΔG_{water}	ΔG_{gas}	ΔG_{water}
REACT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Pathway a)								
DISS	128.9	23.0	116.5	19.1	-38.5	2.4	-40.0	-0.1
INT1	128.0	35.3	117.0	30.3	-39.5	14.6	-39.5	11.2
TS1	130.7	40.9	123.3	39.7	-36.7	20.3	-33.2	20.6
INT2	105.5	7.4	101.8	11.1	-61.9	-13.2	-54.7	-8.0
INT3	107.4	-7.5	108.9	-6.2	-60.0	-28.2	-47.6	-25.3
Pathway b)								
INT4	4.2	10.9	6.3	10.9				
TS2	32.3	26.5	34.8	26.9	4.8	23.2	8.6	25.0
INT5	87.5	-6.1	94.4	-2.4	-53.6	-14.1	-43.6	-11.0
Pathway c)								
INT6	6.9	12.4	6.6	14.0				
TS3	54.4	56.4	55.5	57.3				
INT7	-11.4	-8.4	-10.2	-7.5				
PROD	-17.7	-20.7	-16.3	-19.3	-17.4	-22.3	-5.0	-19.4

It is immediately evident from the first column of Table 3 that the dissociative pathway a) is the least favourable. This is expected since it involves the dissociation of **1** into a PhS^- and tripod zinc cation ion pair, a highly unfavourable process in the gas phase. How much this is modified by solvation will be described in the next section, yet the order remains unchanged.

We first compare the two dissociative (path a) and associative (path b) nucleophilic displacement mechanisms. In the dissociative pathway, the nucleophile is the uncoordinated phenyl thiolate **1-DISS** (Figure 5). It is expected to be more reactive than the zinc-bound thiolate **1-REACT** (Figure 5) which is the nucleophile in the associative S_N2 reaction. Indeed, a free energy barrier of 32.3 kcal/mol is associated with the transition state **1-TS2** from the zinc-bound thiolate **1-REACT**, whereas only 1.8 kcal/mol are required from the dissociated nucleophile **1-DISS** to **1-TS1**.

This difference of reactivity is also visible on the geometrical parameters along both reaction paths (Figure 6a-b). Due to the high reactivity of bare thiophenolate, **1-TS1** corresponds to an early transition state with long $S\cdots MeI$ (2.695 vs 1.821 Å in Ph-S-Me) and short $Me\cdots I$ (2.556 vs 2.188 Å in CH_3I) bonds. On the contrary, the coordination of the thiophenolate to a zinc cationic fragment in **1-REACT** induces a lower reactivity as illustrated by the late transition state **1-TS2** with short $S\cdots MeI$ (2.216 Å) and long $Me\cdots I$ (2.895 Å) bonds.

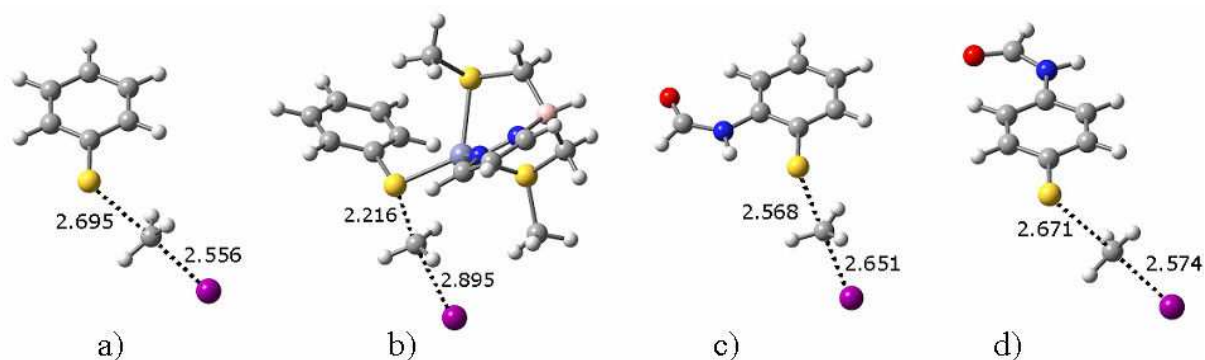


Figure 6. Transition state structures for the S_N2 reaction of CH_3I with various phenyl thiolates in the gas phase. a) thiophenolate (**1-TS1**); b) thiophenolate bound to the $[H(pz^H)Bt^{Me}]Zn$ complex (**1-TS2**); c) *o*-NHC(O)H-thiophenolate (**2-TS1**); d) *p*-NHC(O)H-thiophenolate. yellow : sulfur, red : oxygen, blue : nitrogen, gray : carbon, white : hydrogen, purple : iodine, pink : boron, light blue : zinc.

The third pathway (path c) is very different as two bonds (Zn-S and Me-I) are broken and two bonds (S-Me and Zn-I) are created at the same time. **1-TS3** (Figure 5), the transition state along this σ bond metathesis reaction, requires 54.4 kcal/mol to be reached, which is about 20 kcal/mol more than the energy need for **1-TS2** (Table 3). This high energy barrier has already been observed for 4-center transition states including an alkyl group at the β

position with respect to the metal center.⁸⁸ Consequently, the σ bond metathesis reaction is less favored than the S_N2 -type associative pathway in the gas phase.

3. Solvation effects on complex **1**

The free energies in aqueous solution, at the gas phase geometries, are given for complex **1** in the second column of Table 3. As expected, cationic ($R-Zn^{2+}$ along path a) and anionic (PhS^- in **1-DISS**, I^- in **1-INT5**) species are more strongly solvated than neutral compounds. Compared to **1-REACT**, this leads to their greater stabilisation in solution than in the gas phase.

Solvation induces indeed the reduction of the dissociation free energy of thiophenolate from the zinc complex by more than a hundred kcal/mol. Consequently, the free energy barrier to reach **1-TS1** from the reactants **1-REACT** is considerably reduced from 130.7 to 40.9 kcal/mol. We note however that the solvation of PhS^- reduces its reactivity as shown by the increased activation barrier to reach **1-TS1** from **1-DISS** (+17.9 vs +1.8 kcal/mol in the gas phase). Such dramatic solvation effects on the potential energy profile of pathway a) are likely to have some impact on the optimised geometries. It is however unlikely that such an effect would lead to a change of the most favourable mechanism. Indeed, the difference between the rate-limiting transition states for pathways a) and b) is 14.4 kcal/mol (see Table 3 and discussion below).

On the contrary, solvation induces a greater S_N2 reactivity of the zinc-bound phenylthiolate. The activation barrier to reach **1-TS2** from **1-REACT** is indeed diminished from 32.3 kcal/mol in the gas phase to 26.5 kcal/mol in solution. This may be explained by the transient formation in the S_N2 -associative pathway of two charged species (**1-INT5**) before disruption of the $Zn-S(aryl)$ bond. The associated transition state **1-TS2** has thus small apparent electronic charges and would therefore be more stabilized by the solvent than the neutral reactants.

The effect of water on the relative free energy of **1-TS3** is only 2 kcal/mol as σ bond metathesis does not lead to charge separation. Consequently the energy gap between both associative mechanisms is accentuated by solvent effects and the 4-center transition state is the least favored mechanism. The S_N2 -associative reaction is thus, by far, the most favored alkylation mechanism in water, as **1-TS2** lies 29.9 and 14.4 kcal/mol below **1-TS3** and **1-TS1**, respectively.

4. Influence of zinc ligands

The free energies in gas phase and in water solution for the three alkylation mechanisms of complexes **2-4** are indicated in Table 3. We first describe the influence of the zinc complex moiety on the potential energy surface by comparing **3** to **1** (and **4** to **2**). In the following section, we will look at the influence of the hydrogen bond on the sulfur atom by comparing **2** to **1** (and **4** to **3**).

The S_N2-dissociative mechanism (path a) is the same for **1** and **3** with the exception of the first step where the Zn-S(aryl) bond is broken. This dissociative step is exothermic from **3-REACT** in the gas phase due to favourable charge-charge separation.⁸⁹ As shown previously, this repulsive coulomb effect is visible on the longer Zn-S(aryl) bond length in **3** (2.458 Å) than in **1** (2.250 Å). Inclusion of solvent effects greatly reduces this repulsive interaction, leading to a small positive free dissociation energy of 2.4 kcal/mol. Consequently, an overall free energy barrier of 20.3 kcal/mol for path a) relative to **3-REACT** is computed, much smaller than the 40.9 kcal/mol value obtained from **1-REACT**.

The negative charge of **3-REACT** induces a more pronounced nucleophilic reactivity compared to **1-REACT**. Indeed, **3-TS2** is located only 4.8 kcal/mol higher than the reactants in the gas phase, compared to the 32.3 kcal/mol value obtained for complex **1**. As for the dissociated phenyl thiolate (*vide supra*), inclusion of solvent effect reduces the reactivity of the anionic complex **3**, increasing the free energy barrier of path b) up to 23.2 kcal/mol, much closer to the value for **1**.

We did not succeed in finding **3-TS3** and **4-TS3** on the potential energy surface. Indeed, any attempt at a 4-center transition state leads to re-formation of CH₃I. This may be explained by the very low Lewis acidity of Zn in **3** and **4**, as reflected by the long Zn-ligand bonds (*vide supra*), due to the two anionic charges of the complexes. We thus advance the hypothesis that path c) cannot be followed for electronically rich complexes such as **3** and **4**. In order to confirm the influence of the electronic charge around zinc on the metathesis reactivity, we decide to study the σ-bond mechanism of a cationic zinc complex. For this purpose, we change the (pyrazolyl)bis((methylthio)methyl)hydroborato ligand of **1** to the (pyrazolyl)bis((methylthio)methyl)methane ligand, leading to the new complex **1^C**. As expected, **1^C-TS3** is more easily reachable than **1-TS3**, with gas phase and water phase free energies barriers of, respectively, 44.9 and 51.1 kcal/mol compared to, respectively, the 54.4 and 56.4 kcal/mol values for **1-TS3**. However the S_N2 associative mechanism remains preferred for **1^C** with free energy barrier of 41.6 and 29.6 kcal/mol in the gas phase and in

solution, respectively (data not shown). In non polar solvents or buried enzyme sites, it is conceivable that the σ -bond metathesis be the most favourable in other cases, as the difference in free energy between 1^C -TS3 and 1^C -TS2 is small.

Our results indicate that, contrary to **1** for which the S_N2 -associative pathway b) is much more favorable, the S_N2 -dissociative pathway a) is preferred for **3** as illustrated in Figure 7. Furthermore, they show that complex **3** will be more easily alkylated than complex **1** due to the lower free energy of **3**-TS1 (20.3 kcal/mol) compared to **1**-TS2 (26.5 kcal/mol). This is in agreement with experimental results indicating that dianionic zinc biomimetic complexes are more reactive than neutral ones.^{28, 29} Comparison between **2** and **4** leads to the same conclusion.

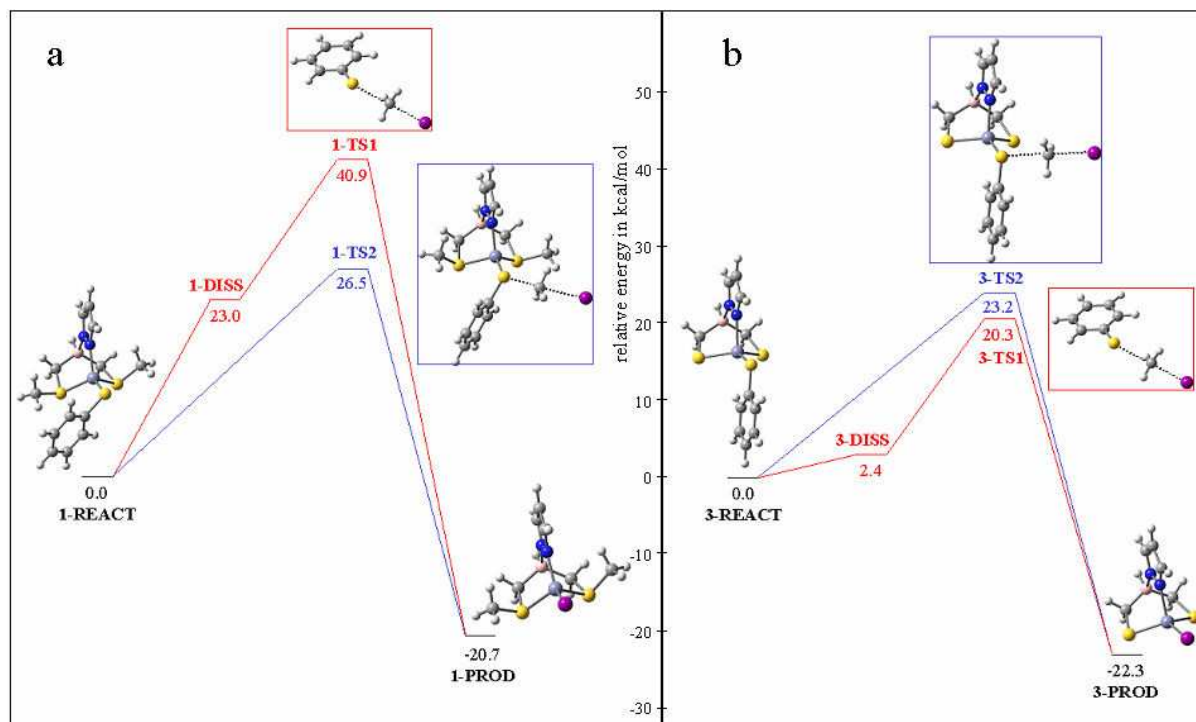


Figure 7. Free energy profiles in water solution along the dissociative (red) and associative (blue) S_N2 pathways for the reaction of MeI with complex **1** (a) and complex **3** (b).

5. Influence of the hydrogen bond

Complex **2** differs from its analog **1** only by the presence of a hydrogen bonding on the phenyl thiolate ligand. This has little effect on the dissociated intermediates, which remains high in free energy (116.5 and 19.1 kcal/mol in the gas phase and in water, respectively). The small decrease observed in comparison with **1** is due to a weaker zinc-

thiolate interaction in **2**. The Zn-S(aryl) bond length is indeed slightly longer when there is a hydrogen bond on the thiolate (2.262 Å in **2** instead of 2.250 Å in **1**, Table 2).

The effect of the hydrogen bond is marked on the transition state **2-TS1**. Due to the lower dissociation free energy, it is located below **1-TS1**, relative to the reactant complexes, both in the gas phase and in solution. However, the free energy barrier of the S_N2 step is higher from **2-DISS** (6.8 and 20.6 kcal/mol in the gas phase and in water, respectively) than from **1-DISS** (1.8 and 17.9 kcal/mol, respectively). Participation of the thiolate in a hydrogen bond reduces its nucleophilicity as already shown by the kinetics of biomimetic complexes.^{47, 55, 56} The geometries of both transition states **1-TS1** and **2-TS1** illustrate this difference of reactivity between the thiolates (Figure 6a and 6c). Indeed, **2-TS1** is a later transition state than **1-TS1**, as measured by the shorter S^{...}CH₃ (2.568 vs 2.695 Å) and the longer CH₃^{...}I (2.651 vs 2.556 Å) distances.

It should be noted that the difference between **1-TS1** and **2-TS1** is not exclusively due to the van der Waals H-bond interaction.⁴⁷ Indeed, a small part of the difference may be due to inductive or conjugative effect through the phenyl ring. This can be checked by comparing *o*- and *p*-NHC(O)H-thiophenolates since there can be no hydrogen bonding in the latter. The structure of the transition state for the S_N2 reaction of CH₃I with *p*-NHC(O)H-thiophenolate is shown in Figure 6d. Its geometrical parameters are intermediate between those of **1-TS1** and **2-TS1**, but clearly closer to the former. The associated activation barrier is 3.4 and 18.4 kcal/mol in the gas phase and in water, respectively, which confirm its greater proximity to **1-TS1** than to **2-TS1** (data not shown).

The impact of this hydrogen-bonding is less marked, but still present for the associative pathways b) and c). As in path a), the H-bond decreases the reactivity of the thiolate, leading to higher **2-TS2** and **2-TS3** compared to **1-TS2** and **1-TS3** respectively, relative to the corresponding reactants. The comparison of the potential free energy surface from **3-REACT** and from **4-REACT** shows the same results.

These changes in the activation barrier values remain small, however, compared to the differences between the three pathways. Consequently, the S_N2-associative path b) is the lower energy path from **2-REACT** as from **1-REACT**. The reactivities of **2-REACT** and **4-REACT** are nevertheless reduced compared to **1-REACT** and **3-REACT** respectively, thus showing that, in good adequacy with experimental results,^{47, 55, 56} a thiolate interacting through a hydrogen bond with an amide is less susceptible to be alkylated.

Discussion

On the reactivity of zinc site :

Model complexes have been previously used to determine second-order rate constants of thiolate alkylation. The first study demonstrates a decline of the reactivity from $\text{Zn}(\text{SPh})_4^{2-}$ to $\text{Zn}(\text{SPh})_3(\text{ImMe})^-$ and $\text{Zn}(\text{SPh})_2(\text{ImMe})_2$.²⁸ This points up the influence of the nature of the zinc-ligand and/or of the charge of the complex. Later studies on the thiolate alkylation reactivity of $\text{ZnL}_n(\text{SR})$ complexes focus (i) on the nature of the L_n zinc-ligand, by varying the zinc-bound atoms (S vs N vs O),^{45, 48} the number of zinc-bound atoms⁵⁰ or the steric requirement^{42, 45, 50, 58}, or (ii) on the nature of the R substituent,^{41, 42, 45, 48} or the presence of intramolecular H-bond in the SR moiety.^{47, 55, 56} Particularly, the rate of alkylation has been shown to increase along the N_3 , N_2S , NS_2 and S_3 series of ligands L_n . It should however be noted that in all of these cases, the comparison of the reactivity is made between complexes bearing the same electronic charge. This situation does not fit the zinc-enzyme active sites where N and S ligands are the neutral histidine and the anionic cysteine side chain respectively, even if the possible protonation of the cysteine side chain is still a matter of debate.^{67, 68} In this study, we examine both the influence of the charge of the complex and of an intramolecular H-bond in the SR moiety.

Our results show that it is possible to greatly modulate the nucleophilic reactivity of a ZnNS_3 complex by changing its global charge. The more negative a zinc site is, the more reactive it is. This is consistent with a previous study on zinc active sites indicating that $[\text{Zn}(\text{Cys})_4]^{2-}$ anionic cores are more susceptible to alkylation than the singly charged $[\text{Zn}(\text{Cys})_3(\text{His})]^-$ and the neutral $[\text{Zn}(\text{Cys})_2(\text{His})_2]$.⁵² This also means that the greater reactivity of $\text{Zn}(\text{SPh})_4^{2-}$ compared to that of $\text{Zn}(\text{SPh})_3(\text{ImMe})^-$ and $\text{Zn}(\text{SPh})_2(\text{ImMe})_2$ ²⁸ is not exclusively due to the change in the nature of the zinc-bound atoms, from ZnS_4 core to ZnS_3N and ZnS_2N_2 ,⁴⁵ but is also a consequence of its greater electronic charge.

Our statement covers two different aspects. First, a more negative environment induces a greater nucleophilicity of the zinc-bound thiolate, as illustrated by the lower free energy barrier to reach **3-TS2** compared to the corresponding value for **1-TS2** (Table 3). This is observed in water solution (23.2 and 26.5 kcal/mol, respectively) and, to a larger extent, in the gas phase (4.8 and 32.3 kcal/mol, respectively), illustrating the trend in solvents that are less polar than water, like toluene⁴⁷ or chloroform⁴⁵ used in some biomimetic studies, or in buried enzymatic active sites.

Secondly, the Zn-S bond is more easily broken in a $[\text{ZnNS}_3]^{2-}$ site than in a $[\text{ZnNS}_3]^0$ one, leading readily to the more reactive free thiolate (*vide infra*). Furthermore, thiolate dissociation from **1** is clearly easier in water, with a dissociation energy of 23.0 kcal/mol, than in the gas phase (128.9 kcal/mol). This is in agreement with the experimental observation of thiolate dissociation in $\text{Zn}(\text{SPh})_4^{2-}$ ²⁹ whereas there is no thiolate exchange in nonpolar media for $[\text{ZnN}_2\text{S}_2]^0$ and $[\text{ZnN}_3\text{S}]^0$ complexes.^{61, 90} Moreover, changing from nonpolar to polar media leads to possible fast thiolate ligand exchange between neutral zinc cores.⁹⁰ It should however be noted that thiolate exchange does not necessarily indicate dissociated thiolate because other mechanisms through bridging thiolate complexes may be responsible for the exchange.⁵⁸

Comparison of the nucleophilic reactivity of free and zinc-bound thiolates is in favour of the former, in agreement with experimental results.²⁸ Indeed, even compared to the dianionic complex **3**, for which $\text{S}_\text{N}2$ free energies barriers of 4.8 and 23.2 kcal/mol are obtained in the gas phase and in water respectively, PhS^- is more nucleophilic with corresponding free energies barriers of 1.8 and 17.9 kcal/mol, respectively. This means that the zinc cation in the anionic $[\text{H}(\text{pz}^\text{H})\text{Bt}^{2-}]\text{Zn}$ core keeps some Lewis acidity and thus kinetically stabilises the zinc-bound thiolate. We postulate that the same applies to the $[\text{Zn}(\text{Cys})_4]^{2-}$ electronically rich active sites of enzymes like the DNA repair protein Ada.

Our computational results are in agreement with several experimental studies suggesting that the reactivity of a thiolate is modulated by the presence of hydrogen bond.^{47, 55-59} The influence of the H-bond is similar for all thiolates studied here, with an increase of about 4 kcal/mol of the free energy necessary to reach the corresponding $\text{S}_\text{N}2$ transition state. This may be related to the magnitude of the $\text{N}\cdots\text{H}\cdots\text{S}$ bond.⁵⁷ Indeed, the $\text{N}\cdots\text{S}$ distance is 2.991 Å, 2.946 Å and 2.951 Å in **2**, **4** and *o*-NHC(O)H-thiophenolate respectively, indicative of a similar hydrogen bond strength.

Complexes **1** and **4** permit to compare the influence of the charge and of the hydrogen bond. In the gas phase, the total electronic charge is the dominant factor, as illustrated by the lower free energy barrier to reach **4-TS2** (8.6 kcal/mol) compared to the value for **1-TS2** (32.3 kcal/mol). By contrast, in water, electronic charge and hydrogen bond influence the nucleophilicity to approximately the same amount. Indeed, a similar free energy barrier is needed to reach the transition state TS2 from the neutral complex without hydrogen bond **1** and from the dianionic complex with hydrogen bond **4**. We may postulate that in nonpolar media or in buried enzyme active site, the situation will be intermediate with a weak

predominance of the effect of the total electronic charge. Thus most of the $[\text{Zn}(\text{Cys})_4]^{2-}$ enzyme active sites show greater reactivity than $[\text{Zn}(\text{Cys})_3(\text{His})]^-$ and $[\text{Zn}(\text{Cys})_2(\text{His})_2]^0$ cores, but some exceptions should probably occur due to hydrogen-bond networks around the active site. This nicely reflects the predicted rank of zinc finger reactivities evaluated according to protein packing and electrostatic screening of their zinc core.⁵² Indeed, in most cases, these properties are similar for $[\text{Zn}(\text{Cys})_4]^{2-}$, $[\text{Zn}(\text{Cys})_3(\text{His})]^-$ and $[\text{Zn}(\text{Cys})_2(\text{His})_2]^0$ core, reflecting the greater reactivity of the most negatively charge site. However, in some cases such as the active site of GATA-1, an important steric and electrostatic screening of the $[\text{Zn}(\text{Cys})_4]^{2-}$ core is observed compared to that of the $[\text{Zn}(\text{Cys})_3(\text{His})]^-$ core of the C-terminal HIV-1 NCp7 zinc finger, in agreement with its lower reactivity.⁵⁴

on the mechanism of zinc site:

The second issue that we approach in this study is the mechanism of thiolate alkylation. For complexes **1-4**, we have explored the potential free energy surface for the three reaction mechanisms envisaged in the literature. Our data reflect a huge difference of behaviour between models **1-2** and **3-4**, each reactivity being predominantly influenced by the electronic composition of the zinc ligands. For the neutral models **1** and **2** displaying a $[\text{ZnNS}_3]^0$ core, the favored alkylation goes through an associative $\text{S}_{\text{N}}2$ mechanism, whereas complexes **3** and **4** with a $[\text{ZnNS}_3]^{2-}$ dianionic core follow a dissociative $\text{S}_{\text{N}}2$ pathway. The difference between both models lies indeed in the dissociation energy of **1** and **3** (or **2** and **4**).

Due to the high dissociation energy for neutral core, especially in the gas phase or in non-polar media, the associative mechanism is the most competitive even if the unreachable free thiolate would be more reactive than the zinc-complexed thiolate. Among the two associative pathways, the $\text{S}_{\text{N}}2$ -type transition state is lower in energy than the 4-center one, indicating a nucleophilic reactivity. This is consistent with a previous study showing that the experimental reactivity of neutral zinc complexes toward MeI closely follows that predicted from the nucleophilicity of the zinc-bound thiolate measured as the energy of the highest occupied molecular orbital.⁴⁸

On the contrary, the electronically rich complexes **3** and **4** will undergo preferentially the dissociation, as the zinc-thiolate interaction is weak, and will thus react as a free dissociated thiolates. Even if the zinc-bound thiolate is more reactive in **3** and **4** than in **1** and **2**, the associative $\text{S}_{\text{N}}2$ transition state, preferred from **1** and **2**, seems very unlikely to form, as shown by the lower energy of the dissociative $\text{S}_{\text{N}}2$ transition state.

These results, reflecting a mechanistic difference that depends on the electronic environment around the zinc center, allow us to clarify the diverse conclusions reported until now in the literature about the associative or dissociative aspect of the mechanism. Based on NMR study, the alkylation of $\text{Zn}(\text{SPh})_4^{2-}$ by phosphotriesters was consistent with a dissociative pathway leading to de-coordination of thiolate from zinc before reaction.²⁹ On the contrary, numerous complexes show retention of thiolate coordination to zinc before alkylation.^{39, 43, 48, 51, 61} These differences in mechanism may be explained by the difference in the net charge of the complexes, the former being dianionic like **3** and **4** whereas all the latter are neutral like **1** and **2**.

All zinc enzyme active sites showing a reactivity towards alkylating agent possess a global electronic charge of -1 or -2 .¹² From our results, we can postulate that dianionic active sites such as those of the DNA repair protein Ada or of the cobalamin-dependent methionine synthase react through a dissociative pathway. For monoanionic active sites like protein farnesyl transferase or cobalamin-independent methionine synthase, we anticipate an intermediate situation where both mechanisms are possible, depending on more subtle factors like the protein hydrogen bond network. This is consistent with the proposed associative transition-state in the case of protein farnesyl transferase.⁶³ It should be noticed that the protonation state of $\text{Zn}(\text{Cys})_4$ and $\text{Zn}(\text{Cys})_3(\text{His})$ cores are still a matter of debate. If these active sites include respectively two and one additional proton to achieve neutrality,^{67, 68} an associative mechanism would be preferred in all cases.

The intramolecular $\text{NH}\cdots\text{S}$ hydrogen bond does not influence the detailed mechanism followed by **2** and **4** compared to **1** and **3**, respectively. Even if for **2**, compared to **1**, the dissociative mechanism is slightly easier (39.7 vs. 40.9 kcal/mol respectively) and the associative $\text{S}_{\text{N}}2$ mechanism is slightly more difficult (26.9 vs 26.5 kcal/mol respectively), the lower free energy pathway is still the associative one. We cannot exclude however that in some cases intramolecular H-bond may induce an inversion of the most competitive pathway, or at least lead to a situation where both pathways can occur. The recent study of the reactivity of a $[\text{ZnS}_4]^0$ complex including an intramolecular H-bond suggests a dissociative mechanism based on isotope effect even if DFT calculations cannot definitely exclude the associative mechanism.⁵⁷ It is possible that the borderline situation described above applies in this case.

Model reactions with complexes including a neutral core indicate that the thioether, produced in the zinc coordination sphere, dissociates from the metal. The zinc keeps however its tetrahedral environment by coordinating an iodide.⁵¹ Comparison of the relative energies

of **1-INT5** and **1-PROD** ($\Delta G_{\text{water}} = -14.6$ kcal/mol) is consistent with this result. This is due to the weak interaction energy between the cationic $[\text{H}(\text{pz}^{\text{H}})\text{Bt}^{\text{Me}}]\text{Zn}$ core and the thioether as reflected by a long Zn-S bond length (2.407 Å) and a weak dissociation energy (19.9 kcal/mol in the gas-phase). We even notice that inclusion of solvent effect leads to an exothermic dissociation by 1.4 kcal/mol. This trend is conserved and quantitatively slightly accentuated in the presence of an intramolecular $\text{NH}\cdots\text{S}$ hydrogen bond. In case of a dianionic core, the Zn-thioether bond is still longer (2.734 and 2.806 Å for **3-INT5** and **4-INT5**, respectively) thus leading to its easy dissociation. Furthermore, in water, we found that iodide coordination to the $[\text{H}(\text{pz}^{\text{H}})\text{Bt}^{2-}]\text{Zn}$ fragment is endothermic for **3** and **4**. This shows that the anionic $[\text{ZnS}_2\text{N}]^-$ does not bind strongly any of the available ligands. This is consistent with an experimental EXAFS study on protein farnesyl transferase indicating that the thioether is not tightly coordinated to the zinc.⁹¹ A recent theoretical study on the same enzyme confirms this aspect by showing that the thioether dissociation from zinc applies with small activation energy and is endothermic.⁹² However, although weak, this interaction may be observed if surrounding hydrogen bonds hold the substrate as in cobalamin-independent methionine synthase,⁹³ or even be significant if the thioether product belongs to the active site as in the DNA repair protein Ada.^{17, 18}

Conclusion

Using several biomimetic complexes, models of thiolate alkylating zinc enzymes, we have examined their reactivity and their reaction mechanism against methyl iodide. In accordance with experimental results, we find that intramolecular $\text{N-H}\cdots\text{S}$ hydrogen bonding reduces the nucleophilicity of the thiolate. Furthermore, we show that the net electronic charge of the complex plays a significant role not only on its reactivity, but especially on the mechanism of thiolate alkylation. The mechanism of zinc mediated alkyl group transfer to thiols has been subject to controversy both in model systems and in the enzymes. Three types of pathways have been proposed: a dissociative $\text{S}_{\text{N}}2$ mechanism with a zinc-dissociated thiolate serving as nucleophile or an associative mechanism, either a σ -bond metathesis or an $\text{S}_{\text{N}}2$ with direct alkylation on the zinc-bound thiolate. We have demonstrated that an $\text{S}_{\text{N}}2$ mechanism is followed in all cases, either with an associative or a dissociative character for $[\text{ZnS}_3\text{N}]^0$ or $[\text{ZnS}_3\text{N}]^{2-}$ cores, respectively. This agrees with experimental results and explains the opposite data obtained with neutral complexes compared to those with $[\text{Zn}(\text{SPh}_4)]^{2-}$. All of the Zn enzyme active sites showing alkyl transfer activity have at least a -1 net charge, based on the

hypothesis that zinc-bound cysteines are deprotonated. For dianionic active sites, we conclude that the dissociative S_N2 should be the preferred pathway. For monoanionic cases, we anticipate an intermediate situation where both S_N2 mechanisms are possible, depending on more subtle factors like the protein hydrogen bond network. If some zinc-bound thiolates are protonated, leading to a neutral active site, a associative mechanism will be favored.

Acknowledgment. We are grateful to the Centre National de la Recherche Scientifique (CNRS) and Ecole Polytechnique for financial support. We also thank the Institut du Développement et des Ressources en Informatique Scientifique (IDRIS, grant number 0543) and the Centre Informatique National de l'Enseignement Supérieur (CINES, grant number dcm2335) for computational resources.

Supporting Information Available: Cartesian coordinates, absolute electronic, gas phase and solvation free energies for all compounds. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Vallee, B. L.; Auld, D. S. *Biochemistry* **1990**, *29*, 5647-5659.
- (2) Vallee, B. L.; Auld, D. S. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 220-224.
- (3) Vallee, B. L.; Auld, D. S. *Acc. Chem. Res.* **1993**, *26*, 543-551.
- (4) Lipscomb, W. N.; Strater, N. *Chem. Rev.* **1996**, *96*, 2375-2433.
- (5) Coleman, J. E. *Curr. Opin. Chem. Biol.* **1998**, *2*, 222-234.
- (6) Vallee, B. L.; Auld, D. S. *FEBS Lett.* **1989**, *257*, 138-140.
- (7) Karlin, S.; Zhu, Z.-Y. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 14231-14236.
- (8) Auld, D. S. *Biometals* **2001**, *14*, 271-313.
- (9) Patel, K.; Kumar, A.; Durani, S. *Biochim. Biophys. Acta* **2007**, *1774*, 1247-1253.
- (10) Tamames, B.; Sousa, S. F.; Tamames, J.; Fernandes, P. A.; Ramos, M. J. *Proteins* **2007**, *69*, 466-475.
- (11) Matthews, R. G.; Goulding, C. W. *Curr. Opin. Chem. Biol.* **1997**, *1*, 332-339.
- (12) Penner-Hahn, J. *Curr. Opin. Chem. Biol.* **2007**, *11*, 166-171.
- (13) Hightower, K. E.; Fierke, C. A. *Curr. Opin. Chem. Biol.* **1999**, *3*, 176-181.
- (14) Strickland, C. L.; Weber, P. C. *Curr. Opin. Drug Discovery Dev.* **1999**, *2*, 475-483.
- (15) Zhang, F. L.; Casey, P. J. *Biochem. J.* **1996**, *320*, 925-932.
- (16) Lane, K. T.; Beese, L. S. *J. Lipid Res.* **2006**, *47*, 681-699.
- (17) He, C.; Hus, J. C.; Sun, L. J.; Zhou, P.; Norman, D. P. G.; Dotsch, V.; Wei, H.; Gross, J. D.; Lane, W. S.; Wagner, G.; Verdine, G. L. *Mol. Cell* **2005**, *20*, 117-129.
- (18) Takinowaki, H.; Matsuda, Y.; Yoshida, T.; Kobayashi, Y.; Ohkubo, T. *Protein Sci.* **2006**, *15*, 487-497.
- (19) Myers, L. C.; Terranova, M. P.; Ferentz, A. E.; Wagner, G.; Verdine, G. L. *Science* **1993**, *261*, 1164-1167.
- (20) Ohkubo, T.; Sakashita, H.; Sakuma, T.; Kainosho, M.; Sekiguchi, M.; Morikawa, K. *J. Am. Chem. Soc.* **1994**, *116*, 6035-6036.

- (21) Myers, L. C.; Jackow, F.; Verdine, G. L. *J. Biol. Chem.* **1995**, *270*, 6664-6670.
- (22) Peariso, K.; Goulding, C. W.; Huang, S.; Matthews, R. G.; Penner-Hahn, J. E. *J. Am. Chem. Soc.* **1998**, *120*, 8410-8416.
- (23) Matthews, R. G. *Acc. Chem. Res.* **2001**, *34*, 681-689.
- (24) Tummino, P. J.; Scholten, J. D.; Harvey, P. J.; Holler, T. P.; Maloney, L.; Gogliotti, R.; Domagala, J.; Hupe, D. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 969-973.
- (25) Loo, J. A.; Holler, T. P.; Sanchez, J.; Gogliotti, R.; Maloney, L.; Reily, M. D. *J. Med. Chem.* **1996**, *39*, 4313-4320.
- (26) Hathout, Y.; Fabris, D.; Han, M. S.; Sowder, R. C.; Henderson, L. E.; Fenselau, C. *Drug. Metab. Dispos.* **1996**, *24*, 1395-1400.
- (27) Maynard, A. T.; Huang, M.; Rice, W. G.; Covell, D. G. **1998**, *95*, 11578-11583.
- (28) Wilker, J. J.; Lippard, S. J. *J. Am. Chem. Soc.* **1995**, *117*, 8682-8683.
- (29) Wilker, J. J.; Lippard, S. J. *Inorg. Chem.* **1997**, *36*, 969-978.
- (30) Watson, A. D.; Rao, C. P.; Dorfman, J. R.; Holm, R. H. *Inorg. Chem.* **1985**, *24*, 2820-2826.
- (31) Henkel, G.; Krebs, B. *Chem. Rev.* **2004**, *104*, 801-824.
- (32) Parkin, G., Alkyl, Hydride, and Hydroxide Derivatives of the s- and p-Block Elements Supported by Poly(pyrazolyl)borato Ligation: Models for Carbonic Anhydrase, Receptors for Anions, and the Study of Controlled Crystallographic Disorder. In *Advances in Inorganic Chemistry*, Sykes, A. G., Eds. Academic Press: 1995; Vol. 42, pp 291-393.
- (33) *Scorpionates: The Coordination Chemistry of Polypyrazolylborate Ligands*. Trofimenko, S. ed.; Imperial College Press: London, 1999.
- (34) Garner, M.; Reglinski, J.; Cassidy, I.; Spicer, M. D.; Kennedy, A. R. *Chem. Commun.* **1996**, 1975-1976.
- (35) Reglinski, J.; Garner, M.; Cassidy, I. D.; Slavin, P. A.; Spicer, M. D.; Armstrong, D. R. *J. Chem. Soc., Dalton Trans.* **1999**, 2119-2126.
- (36) Frison, G.; Sevin, A. *J. Phys. Chem. A* **1999**, *103*, 10998-11003.
- (37) Frison, G.; Sevin, A. *J. Chem. Soc., Perkin Trans. 2* **2002**, 1692-1697.
- (38) Morlok, M. M.; Docrat, A.; Janak, K. E.; Tanski, J. M.; Parkin, G. *Dalton Trans.* **2004**, 3448-3452.
- (39) Bridgewater, B. M.; Fillebeen, T.; Friesner, R. A.; Parkin, G. *J. Chem. Soc., Dalton Trans.* **2000**, 4494-4496.
- (40) Melnick, J. G.; Docrat, A.; Parkin, G. *Chem. Commun.* **2004**, 2870-2871.
- (41) Ji, M.; Benkmil, B.; Vahrenkamp, H. *Inorg. Chem.* **2005**, *44*, 3518-3523.
- (42) Ibrahim, M. M.; Seebacher, J.; Steinfeld, G.; Vahrenkamp, H. *Inorg. Chem.* **2005**, *44*, 8531-8538.
- (43) Ibrahim, M. M.; He, G.; Seebacher, J.; Benkmil, B.; Vahrenkamp, H. *Eur. J. Inorg. Chem.* **2005**, 4070-4077.
- (44) Ibrahim, M. M.; Shu, M.; Vahrenkamp, H. *Eur. J. Inorg. Chem.* **2005**, 1388-1397.
- (45) Rombach, M.; Seebacher, J.; Ji, M.; Zhang, G. F.; He, G. S.; Ibrahim, M. M.; Benkmil, B.; Vahrenkamp, H. *Inorg. Chem.* **2006**, *45*, 4571-4575.
- (46) Chiou, S. J.; Innocent, J.; Riordan, C. G.; Lam, K. C.; Liable-Sands, L.; Rheingold, A. L. *Inorg. Chem.* **2000**, *39*, 4347-4353.
- (47) Chiou, S. J.; Riordan, C. G.; Rheingold, A. L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3695-3700.
- (48) Warthen, C. R.; Hammes, B. S.; Carrano, C. J.; Crans, D. C. *J. Biol. Inorg. Chem.* **2001**, *6*, 82-90.
- (49) Notni, J.; Görls, H.; Anders, E. *Eur. J. Inorg. Chem.* **2006**, 1444-1455.
- (50) Notni, J.; Gunther, W.; Anders, E. *Eur. J. Inorg. Chem.* **2007**, 985-993.

- (51) Grapperhaus, C. A.; Tuntulani, T.; Reibenspies, J. H.; Darensbourg, M. Y. *Inorg. Chem.* **1998**, *37*, 4052-4058.
- (52) Maynard, A. T.; Covell, D. G. *J. Am. Chem. Soc.* **2001**, *123*, 1047-1058.
- (53) South, T. L.; Blake, P. R.; Hare, D. R.; Summers, M. F. *Biochemistry* **1991**, *30*, 6342-6349.
- (54) Huang, M. J.; Maynard, A.; Turpin, J. A.; Graham, L.; Janini, G. M.; Covell, D. G.; Rice, W. G. *J. Med. Chem.* **1998**, *41*, 1371-1381.
- (55) Smith, J. N.; Shirin, Z.; Carrano, C. J. *J. Am. Chem. Soc.* **2003**, *125*, 868-869.
- (56) Smith, J. N.; Hoffman, J. T.; Shirin, Z.; Carrano, C. J. *Inorg. Chem.* **2005**, *44*, 2012-2017.
- (57) Morlok, M. M.; Janak, K. E.; Zhu, G.; Quarless, D. A.; Parkin, G. *J. Am. Chem. Soc.* **2005**, *127*, 14039-14050.
- (58) Melnick, J. G.; Zhu, G.; Buccella, D.; Parkin, G. *J. Inorg. Chem.* **2006**, *100*, 1147-1154.
- (59) Ibrahim, M. M. *Inorg. Chim. Acta* **2006**, *359*, 4235-4242.
- (60) Parkin, G. *Chem. Rev.* **2004**, *104*, 699-767.
- (61) Brand, U.; Rombach, M.; Seebacher, J.; Vahrenkamp, H. *Inorg. Chem.* **2001**, *40*, 6151-6157.
- (62) Fox, D. C.; Fiedler, A. T.; Halfen, H. L.; Brunold, T. C.; Halfen, J. A. *J. Am. Chem. Soc.* **2004**, *126*, 7627-7638.
- (63) Huang, C. C.; Hightower, K. E.; Fierke, C. A. *Biochemistry* **2000**, *39*, 2593-2602.
- (64) Topol, I. A.; Nemukhin, A. V.; Chao, M.; Iyer, L. K.; Tawa, G. J.; Burt, S. K. *J. Am. Chem. Soc.* **2000**, *122*, 7087-7094.
- (65) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins* **2007**, *66*, 205-218.
- (66) Frison, G.; Ohanessian, G. *submitted for publications*.
- (67) Fabris, D.; Zaia, J.; Hathout, Y.; Fenselau, C. *J. Am. Chem. Soc.* **1996**, *118*, 12242-12243.
- (68) Fabris, D.; Hathout, Y.; Fenselau, C. *Inorg. Chem.* **1999**, *38*, 1322-1325.
- (69) Simonson, T.; Calimet, N. *Proteins* **2002**, *49*, 37-48.
- (70) Chance, M. R.; Sagi, I.; Wirt, M. D.; Frisbie, S. M.; Scheuring, E.; Chen, E.; Bess, J. W.; Henderson, L. E.; Arthur, L. O.; South, T. L.; Perezalvarado, G.; Summers, M. F. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 10041-10045.
- (71) Dauter, Z.; Wilson, K. S.; Sieker, L. C.; Moulis, J. M.; Meyer, J. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8836-8840.
- (72) Clark-Baldwin, K.; Tierney, D. L.; Govindaswamy, N.; Gruff, E. S.; Kim, C.; Berg, J.; Koch, S. A.; Penner-Hahn, J. E. *J. Am. Chem. Soc.* **1998**, *120*, 8401-8409.
- (73) Dudev, T.; Lim, C. *J. Am. Chem. Soc.* **2002**, *124*, 6759-6766.
- (74) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.;

- Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, revision B.05*, Gaussian, Inc.: Pittsburgh, PA, 2003.
- (75) Wachters, A. J. *J. Chem. Phys.* **1970**, *52*, 1033-1036.
- (76) Lajohn, L. A.; Christiansen, P. A.; Ross, R. B.; Atashroo, T.; Ermler, W. C. *J. Chem. Phys.* **1987**, *87*, 2812-2824.
- (77) Peterson, K. A.; Shepler, B. C.; Figgen, D.; Stoll, H. *J. Phys. Chem. A* **2006**, *110*, 13877-13883.
- (78) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995-2001.
- (79) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669-681.
- (80) Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1*, 70-77.
- (81) Topol, I. A.; Casas-Finet, J. R.; Gussio, R.; Burt, S. K.; Erickson, J. W. *Theochem-J. Mol. Struct.* **1998**, *423*, 13-28.
- (82) Diaz, N.; Suarez, D.; Merz, K. M., Jr. *Chem. Phys. Lett.* **2000**, *326*, 288-292.
- (83) Dudev, T.; Lim, C. *J. Phys. Chem. B* **2001**, *105*, 10709-10714.
- (84) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Chem. Phys. B* **2007**, *111*, 9146-9152.
- (85) Frison, G.; Ohanessian, G. *J. Comput. Chem.* **2008**, *29*, 416-433.
- (86) Rayon, V. M.; Valdes, H.; Diaz, N.; Suarez, D. *J. Chem. Theory Comput.* **2008**, *4*, 243-256.
- (87) Amin, E. A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 75-85.
- (88) Perrin, L.; Eisenstein, O.; Maron, L. *New J. Chem.* **2007**, *31*, 549-555.
- (89) Picot, D.; Ohanessian, G.; Frison, G. *in preparation*.
- (90) Seebacher, J.; Ji, M.; Vahrenkamp, H. *Eur. J. Inorg. Chem.* **2004**, 409-417.
- (91) Tobin, D. A.; Pickett, J. S.; Hartman, H. L.; Fierke, C. A.; Penner-Hahn, J. E. *J. Am. Chem. Soc.* **2003**, *125*, 9962-9969.
- (92) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Comput. Chem.* **2007**, *28*, 1160-1168.
- (93) Ferrer, J.-L.; Ravanel, S.; Robert, M.; Dumas, R. *J. Biol. Chem.* **2004**, *276*, 44235-44238.

Bibliographie

- Brand, U.; Rombach, M.; Seebacher, J.; Vahrenkamp, H. *Inorg. Chem.* **2001**, *40*, 6151.
- Bridgewater, B. M.; Fillebeen, T.; Friesner, R. A.; Parkin, G. *J. Chem. Soc., Dalton Trans.* **2000**, 4494.
- Chiou, S. J.; Innocent, J.; Riordan, C. G.; Lam, K. C.; Liable-Sands, L.; Rheingold, A. L. *Inorg. Chem.* **2000**, *39*, 4347.
- Chiou, S. J.; Riordan, C. G.; Rheingold, A. L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3695.
- Grapperhaus, C. A.; Tuntulani, T.; Reibenspies, J. H.; Darensbourg, M. Y. *Inorg. Chem.* **1998**, *37*, 4052.
- He, C.; Hus, J. C.; Sun, L. J.; Zhou, P.; Norman, D. P. G.; Dotsch, V.; Wei, H.; Gross, J. D.; Lane, W. S.; Wagner, G.; Verdine, G. L. *Mol. Cell* **2005**, *20*, 117.
- Huang, C. C.; Hightower, K. E.; Fierke, C. A. *Biochemistry* **2000**, *39*, 2593.
- Ibrahim, M. M.; He, G.; Seebacher, J.; Benkmil, B.; Vahrenkamp, H. *Eur. J. Inorg. Chem.* **2005**, 4070.
- Ji, M.; Benkmil, B.; Vahrenkamp, H. *Inorg. Chem.* **2005**, *44*, 3518.
- Maynard, A. T.; Covell, D. G. *J. Am. Chem. Soc.* **2001**, *123*, 1047.
- Morlok, M. M.; Janak, K. E.; Zhu, G.; Quarless, D. A.; Parkin, G. *J. Am. Chem. Soc.* **2005**, *127*, 14039.
- Parkin, G., Alkyl, Hydride, and Hydroxide Derivatives of the s- and p-Block Elements Supported by Poly(pyrazolyl)borato Ligation: Models for Carbonic Anhydrase, Receptors for Anions, and the Study of Controlled Crystallographic Disorder. In *Advances in Inorganic Chemistry*, Sykes, A. G., Eds. Academic Press: **1995**; Vol. 42, pp 291.
- Parkin, G. *Chem. Rev.* **2004**, *104*, 699.
- Rombach, M.; Seebacher, J.; Ji, M.; Zhang, G. F.; He, G. S.; Ibrahim, M. M.; Benkmil, B.; Vahrenkamp, H. *Inorg. Chem.* **2006**, *45*, 4571.
- Smith, J. N.; Shirin, Z.; Carrano, C. J. *J. Am. Chem. Soc.* **2002**, *125*, 868.
- Smith, J. N.; Hoffman, J. T.; Shirin, Z.; Carrano, C. J. *Inorg. Chem.* **2005**, *44*, 2012.
- Takinowaki, H.; Matsuda, Y.; Yoshida, T.; Kobayashi, Y.; Ohkubo, T. *Protein Sci.* **2006**, *15*, 487.
- Trofimenko, S. *Scorpionates: The Coordination Chemistry of Polypyrazolylborate Ligands*. ed. Imperial College Press: London, **1999**.
- Warthen, C. R.; Hammes, B. S.; Carrano, C. J.; Crans, D. C. *J. Biol. Inorg. Chem.* **2001**, *6*, 82.
- Wilker, J. J.; Lippard, S. J. *J. Am. Chem. Soc.* **1995**, *117*, 8682.
- Wilker, J. J.; Lippard, S. J. *Inorg. Chem.* **1997**, *36*, 969.

Chapitre IV

Etude de la dissociation de liaisons zinc-thiolate

La première partie de notre étude a fait ressortir l'importance de deux caractéristiques physico-chimiques des complexes zinc-soufre vis-à-vis de leur propension à subir une alkylation. Bien sûr, la nucléophilie de l'atome du soufre, en fonction de son environnement, joue un rôle important et celle-ci sera l'objet du chapitre V de cette thèse. L'autre caractéristique essentielle concerne la possibilité de la dissociation de la liaison zinc-soufre, conduisant à un groupe thiolate décoordonné et comme nous l'avons vu, beaucoup plus réactif. L'étude des facteurs influençant cette dissociation est l'objet de ce chapitre.

De plus, l'étude précédente a fait ressortir l'importance du nombre de charges négatives autour du zinc. Ainsi les complexes **3** et **4** étudiés précédemment présentent une dissociation aisée de l'arylthiolate. La liaison zinc-soufre révèle ainsi une particularité encore non évoquée dans la littérature : sa dissociation peut, dans certains cas, être exothermique. L'abondance du zinc et de la liaison zinc-soufre dans les systèmes biologiques nous incite à regarder de plus près ce phénomène et de vérifier son occurrence dans d'autres systèmes organométalliques.

Tout d'abord, nous nous appuierons sur des structures expérimentales de petites molécules à zinc pour calibrer nos calculs quantiques et déterminer précisément l'énergie de dissociation de ces composés. Nous confronterons ensuite les résultats théoriques et les observations expérimentales, afin de mieux comprendre le phénomène de dissociation de la liaison zinc-soufre et d'en cerner les causes. L'influence des charges des ligands sera ainsi étudiée en faisant varier le nombre de ligands neutres (amines) ou anioniques (thiolates) dans les complexes choisis.

I. Introduction

A. Observation expérimentale des structures zinc-soufre

1. *Nombres de coordination du zinc*

Le zinc est caractérisé par une grande flexibilité et la possibilité de présenter plusieurs nombres de coordination. Bien qu'il soit coordonné à six molécules d'eau en solution, le zinc est généralement sous forme tétraédrique dans les doigts de zinc et les métalloenzymes.

Une étude de la Protein Data Bank (PDB) montre que pour les sites structuraux, les proportions des différentes coordinations tétra-, penta- et hexaédrique sont respectivement de 79, 6 et 12%. Pour les sites catalytiques, elles sont respectivement de 48, 44 et 6% (Alberts 1998). Des statistiques plus récentes obtenues sur des structures cristallographiques d'enzymes à zinc donnent une proportion grandissante du nombre de cystéines dans les sites publiés de la PDB : parmi 126 protéines structurales, une majorité (84%) a une coordination tétravalente pour 14% de penta- et 4% d'hexamères seulement. Les sites catalytiques sont constitués, quant à eux, de 58% de tétraèdres (Patel 2007). L'effet de stabilisation de la protéine est donc assuré par la conformation tétraédrique des ligands cystéines et histidines autour du zinc.

Les distances moyennes métal-ligand pour un site actif sont $\text{Zn-N(His)}=2.07\text{-}2.09\text{ \AA}$ et $\text{Zn-S(Cys)}=2.21\text{-}2.35\text{ \AA}$ (Dudev 2003). Une longueur Zn-S moyenne de $2.32\text{ \AA}$ a été calculée à partir de la structure de 994 enzymes à zinc (Tamames 2007) et la valeur moyenne de $2.34\text{ \AA}$ a été déterminée pour la structure de 22 sites de type ZnCys_4 de la PDB (Dudev 2001). Il a été par ailleurs montré que plus le nombre de cystéines est important dans la sphère de coordination du zinc, par rapport à celui des histidines, et plus les longueurs des liaisons correspondantes seront grandes. Ceci s'explique par le fait que les ligands thiolates sont des groupes donneurs σ et π puissants, alors que l'histidine est un faible donneur d'électrons. Les cystéines sont donc plus aptes à stabiliser la charge positive du zinc, ce qui affaiblit la liaison au métal et donc augmente la longueur de liaison correspondante (Tamames 2007).

Cette préférence pour un mode de coordination 4 se retrouve aussi dans les complexes organométalliques du zinc. En effet, des statistiques effectuées sur la base de données de Cambridge (CSD) reflètent que le zinc tétra-coordonné se retrouve dans 50% des cas, contre 26% de zinc penta-coordonné et 24% d'hexa-coordonné (Alberts 1998, Dudev 2003, Tamames 2007). En 1995, une autre étude de la CSD fait le compte des complexes dans lesquels le zinc est uniquement lié à des ligands soufrés : 20 entrées correspondent à des tétraèdres (soit ZnS_4), 3 à des pentaèdres (ZnS_5) et seulement 2 sont hexaédriques (ZnS_6) (Bock 1995). Il existe également des exemples de complexes du zinc tri-coordonnés ZnS_3 (Gruff 1989, Matsunaga 2005) et la littérature reporte un exemple de zinc di-coordonné ZnS_2 (Power 1994 et 2005). Pour ces derniers cas, la faible coordinance du zinc est obtenue par la présence de groupes thiolates portant un substituant particulièrement encombré stériquement.

2. Structures du zinc à 4 ligands soufrés

Le zinc tétra-coordonné étant le plus fréquent, nous nous sommes plus particulièrement intéressés à l'étude de ce type de sites. Parmi ceux-ci, la partie précédente a révélé un comportement particulier de la dissociation de la liaison Zn-S lorsque les quatre ligands du zinc sont anioniques. Les sites enzymatiques $[\text{Zn(Cys)}_4]^{2-}$ se trouvent donc au centre de ce chapitre. Leurs modèles chimiques ZnS_4 présentent cependant rarement une charge globale de -2 car nombre de ligands soufrés ne sont pas des thiolates. Ainsi, un récent exemple de complexe ZnS_4 , incluant un ligand tripodal dérivé du pyrazolylborate, chaque bras étant un thioimidazole, présente une charge nulle (Ibrahim 2005).

Seuls quatre composés de type ZnS_4 et de charge totale -2 sont répertoriés dans la base structurale de Cambridge. Deux de ses entrées correspondent au composé $[\text{Zn(SPh)}_4]^{2-}$ (Swenson 1978, Ueyzma 1988) (voir ci-dessous), une autre au composé $[\text{Zn(S-2-Ph-C}_6\text{H}_4)_4]^{2-}$ obtenu en ajoutant un substituant phényle en ortho des groupes phényliques de $[\text{Zn(SPh)}_4]^{2-}$ (Figure IV.1, gauche), synthétisé en présence de contre-ions tétraéthylammonium dans l'acétonitrile (Entrée LAGJIM de la CSD) (Silver 1993). Une dernière structure, $[\text{Zn}_4(\text{SC}_6\text{H}_5)_{10}]$, est en fait constituée de quatre motifs de type ZnS_4^{2-} et utilisée comme modèle pour l'étude des métallothionéines (Figure IV.1, droite) (Entrée PIVZEZ de la CSD) (Huang 1994).

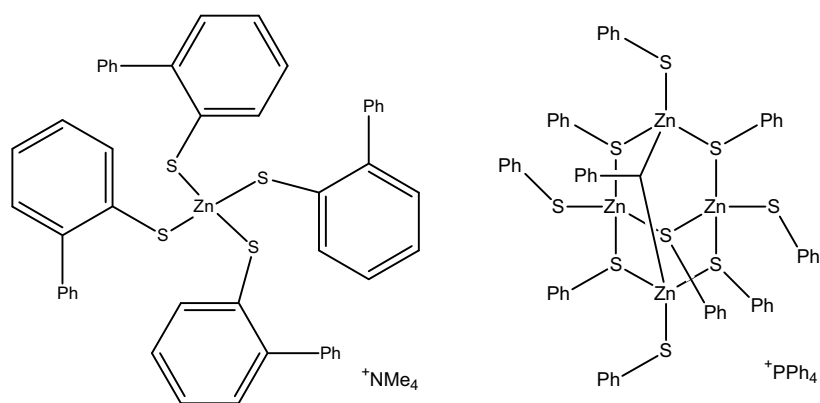


Figure IV.1 : Structures de la CSD possédant un motif ZnS_4^{2-}

Le complexe $[\text{Zn}(\text{SPh})_4]^{2-}$ est à notre connaissance le seul complexe dianionique pour lequel une réactivité nucléophile a été décrite (Wilker 1995). Sa structure cristallographique, présente dans la base de données de Cambridge, va nous servir comme structure de référence pour la suite de notre étude. Du fait de contre-ions différents, deux structures de $[\text{Zn}(\text{SPh})_4]^{2-}$ existent dans la CSD.

La structure notée A_{E1} (entrée ZNTPHP de la CSD) correspond à la première structure cristallographique de $[\text{Zn}(\text{SPh})_4]^{2-}$ réalisée en 1978 par Swenson, Baenziger et Coucovanis en présence de contre-ions tétraphénylphosphonium (Swenson 1978). Cette synthèse a été motivée par le besoin de caractérisation de complexes thiophénolates monomériques de cations divalents du manganèse, fer, cobalt, nickel, zinc et cadmium, qui permettront eux-mêmes de mieux appréhender la réactivité des sites actifs des métalloenzymes. Tous ces complexes ont été synthétisés d'après la méthode publiée en 1975 par le même groupe (Holah 1975). Elle utilise comme réactif des sels de tétraphénylphosphonium des complexes dithiosquarato (Dts) du métal M correspondant et le fait réagir avec du thiophénolate de potassium KSPH dans l'acétonitrile chaud et sous azote (Schéma IV.1).

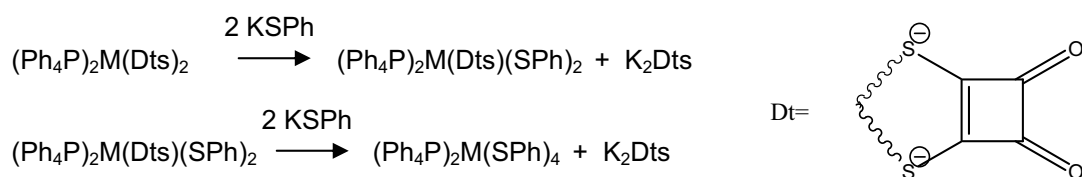


Schéma IV.1 : Synthèse de $(\text{Ph}_4\text{P})_2\text{M}(\text{SPh})_4$ ($\text{M} = \text{Zn}^{2+}, \text{Fe}^{2+}, \text{Co}^{2+}, \text{Mn}^{2+}, \text{Ni}^{2+}$)

La réaction est rapide et mise en évidence par la précipitation des sels de dithiosquarato potassium. Les complexes thiophénolates obtenus sont stables pour une période de temps courte, puis leurs solutions se décomposent rapidement à l'air. La structure observée lorsque le métal est le zinc (Figure IV.2) ressemble à un tétraèdre distordu de symétrie D_{2d} dont la valeur des angles S-Zn-S varie entre 96.7° et 121.1° . Cette distorsion consiste en une déviation par rapport à la symétrie T_d et touche deux atomes de soufre plus particulièrement. Ceci suggère que la déformation du tétraèdre

vient des forces de packing du cristal. Les longueurs des liaisons Zn-S varient entre 2.329 Å et 2.363 Å (Tableau IV.1).

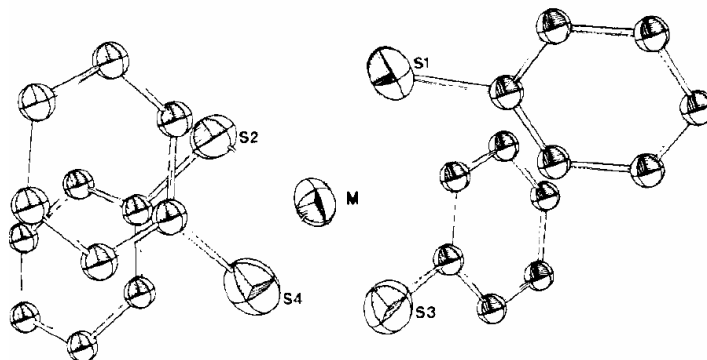


Figure IV.2 : Structure cristallographique de A_{E1} ($M=Zn$) (Swenson 1978)

Tableau IV.1 : Longueur des liaisons Zn-S pour différents complexes de type ZnS_4^{2-} et ZnS_3^- .

Structure	Composé	Zn-S1	Zn-S2	Zn-S3	Zn-S4	Zn-S moyen
A_{E1}^a	$[P(Ph)_4]_2[Zn(SPh)_4]$	2.362	2.363	2.329	2.357	2.353
A_{E2}^b	$[N(Me)_4]_2[Zn(SPh)_4]$	2.361	2.345	2.344	2.378	2.357
B_E^c	$[N(n-Pr)_4][Zn(SPh^{Me^4})_3]$	2.230	2.243	2.217	/	2.230
C_E^d	$[N(Et)_4][Zn(SAd)_3]$	2.259	2.263	2.264	/	2.262

a) Swenson 1978, b) Ueyama 1988, c) Gruff 1989, d) Matsunaga 2005 (SAd=1-adamantanethiolate)

La seconde structure cristallographique notée A_{E2} (entrée GAKKUY de la CSD) (Ueyama 1988), quant à elle, représente $[Zn(SPh)_4]^{2-}$ en présence des contre-ions tétraméthylammonium. La synthèse est effectuée à partir d'une extension des méthodes des équipes de Swenson (Swenson 1978) et de Dance (Dance 1984). Cette dernière préparation consiste à faire réagir un sel de zinc avec une solution de benzénethiol et d'une quantité équimolaire de triéthylamine, sous atmosphère d'azote, dans le méthanol en présence de chlorure de tétraméthylammonium et à température constante. Cette manière de procéder a été reprise par Lippard (Wilker 1995 et 1997).

Contrairement à la synthèse de Swenson, les cristaux du complexe de zinc obtenus sont stables à l'air pendant au moins 24h (Dance 1984). La plus grande différence concerne l'arrangement structural des ligands autour du zinc. La symétrie obtenue et apparente dans la structure cristallographique n'est plus D_{2d} mais S_4 (Figure IV.3). Les valeurs des liaisons Zn-S sont légèrement différentes par rapport à celles de la structure de conformation D_{2d} (Tableau IV.1). Nous pouvons cependant noter que la distance Zn-S moyenne est presque exactement la même pour les deux structures A et B, à 2.355 ± 0.002 Å.

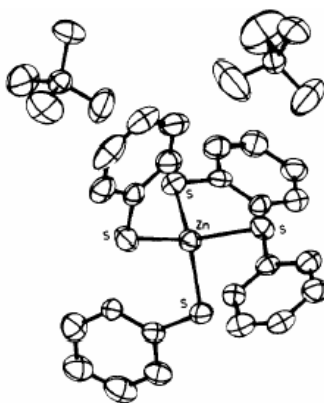


Figure IV.3 : Structure cristallographique de A_{E2}
(Ueyama 1988)

Alors que l'on aurait pu s'attendre à ce que $[Zn(SPh)_4]^{2-}$ ait une coordination tétraédrique idéale, ce composé existe donc deux conformations qui diffèrent majoritairement suivant les positions des groupes phényles.

Une raison avancée pour l'existence de deux conformations est l'apparition de forces de packing entre les ligands thiophénolates et les contre-ions $^+P(Ph)_4$ lorsque ceux-ci sont employés (Ueyama 1988). Les interactions stériques causent une distorsion du centre MS_4 par rapport à la géométrie tétraédrique, de telle façon que l'isomère S_4 est comprimé tétragonalement suivant l'axe S_4 (Figure IV.4, gauche), alors que l'isomère D_{2d} est élongé suivant l'un de ses deux axes de symétries (Figure IV.4, droite). Dans cette dernière conformation, les déviations par rapport à un tétraèdre idéal (109.5°) sont reflétées par les petites valeurs des angles S_1-Zn-S_2 (96.7°) et S_3-Zn-S_4 (99.6°) et l'ouverture des autres angles (115.6 et 121.0°). Dans l'isomère S_4 , l'angle de torsion entre le groupe phényle et le plan défini par l'angle $Zn-S-C$ devrait être nul puisqu'il existe une interaction favorable entre les électrons π du phényle et de l'orbitale p du soufre. Cependant, une légère torsion des groupes phényles apparaît et suggère une interaction de π -stacking entre les thiophénolates et les contre-ions (Ueyama 1988).

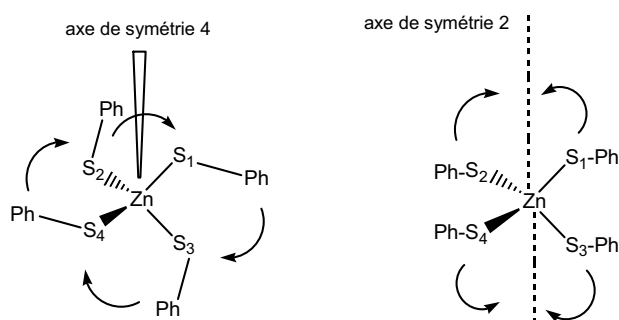


Figure IV.4 : Compression de l'isomère S_4 (à gauche)
et élongation de l'isomère D_{2d} (à droite)

Cependant, puisque l'utilisation des contre-ions $^+N(Me)_4$ laisse également apparaître une distorsion, les déviations par rapport à une coordination tétraédrique idéale semblent alors inhérentes à la nature des anions. Les groupes phényles attachés aux soufres S_1 et S_2 (Figure IV.2) se trouvent

dans un plan, lui-même orthogonal à un autre plan défini par les deux groupes phényles restants et les atomes de soufre S_3 et S_4 . Cette configuration fait naître des contacts entre les protons orthos des phényles et les atomes de soufre. Il existe alors trois conformations raisonnables et équivalentes des groupes phényles pour réduire cette gêne stérique, dont l'une est celle adoptée par la structure de symétrie D_{2d} (Figure IV.5). La conformation trouvée avec les contre-ions tétraméthylammonium est obtenue par rotation des deux groupes phényles d'environ $\pm 120^\circ$ (Coucouvani 1981). Alors que les angles S-Zn-S dans $[\text{Ph}_4\text{P}]_2[\text{Zn}(\text{SPh})_4]$ (géométrie S_4) varient dans la gamme substantielle de 97 - 121° , ceux de $[\text{Me}_4\text{N}]_2[\text{Zn}(\text{SPh})_4]$ (géométrie D_{2d}) sont dans l'intervalle 100 - 117° car la gêne stérique est réduite par la rotation des phényles (Parkin 2004).

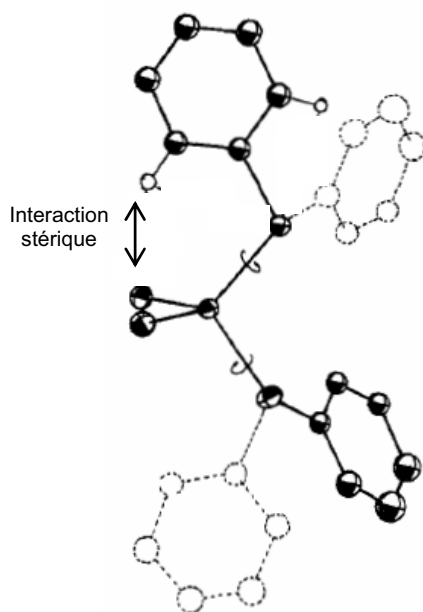


Figure IV.5: Position des groupes phényles dans les deux conformations D_{2d} (normal) et S_4 (pointillés) (Coucouvani 1981)

3. Structures du zinc à 3 ligands soufrés

Dans le cadre de notre étude portant sur la dissociation d'un groupe thiolate d'un site ZnS_4^{2-} , nous nous sommes également intéressés au produit obtenu, c'est-à-dire les complexes de type ZnS_3^- . Deux structures de ce type existent dans la littérature (Gruff 1989, Matsunaga 2005). Le premier complexe de ce type, et le plus intéressant pour notre étude de la dissociation car les ligands du zinc sont des arylthiolates, a été synthétisé et caractérisé en 1989. Il s'agit du tris(2,3,5,6-tétraméthylbenzénethiolate)zinc $[\text{Zn}(\text{SPh}^{\text{Me}_4})_3]^-$ (entrée SEJZEM de la CSD) (Gruff 1989), qui présente une structure trigonale plane (Figure IV.6).

Ce composé, noté B_E , a été synthétisé dans l'acétonitrile à partir de sulfates de zinc hydratés auxquels sont ajoutés cinq équivalents de tétraméthylbenzénethiolate de lithium et un équivalent de la base $(n\text{-Pr}_4\text{N})\text{Br}$. La somme des angles S-Zn-S de la structure recristallisée est égale à 360° et les

déviation par rapport à une symétrie C_3 résultent du π -stacking existant entre deux des anneaux aromatiques (Figure IV.6). La grande différence entre les distances Zn-S de $[\text{Zn}(\text{SPh}^{\text{Me}4})_3]^-$ et $[\text{Zn}(\text{SPh})_4]^{2-}$ reflète le changement de coordination du zinc (Tableau IV.1). On peut en effet s'attendre à ce qu'un dication du zinc coordonne plus fortement trois ligands anioniques que quatre, d'où des liaisons Zn-S plus courtes dans $[\text{Zn}(\text{SPh}^{\text{Me}4})_3]^-$ que dans $[\text{Zn}(\text{SPh})_4]^{2-}$.

Le remplacement des substituants phéniliques par des substituants plus encombrés, comme $\text{SPh}^{\text{Me}4}$, a néanmoins peu d'influence sur la longueur de la liaison zinc-soufre. En effet, les structures tétra-coordonnées $[\text{Zn}(\text{ImMe})(\text{SPh})_3]$ (Wilker 1995) et $[\text{Zn}(\text{ImMe})(\text{SPh}^{\text{Me}4})_3]$ (Gruff 1989) ont respectivement des distances moyennes Zn-S = 2.317 et 2.326 Å. On pourrait s'attendre à un allongement de la liaison métal-ligand lors de l'emploi d'un thiolate encombré. Cependant, le faible écart observé entre les deux structures suggère que l'effet stérique est négligeable ou qu'il est contrebalancé par l'effet électronique donneur des méthyles, rendant la charge anionique légèrement plus localisée sur le soufre.

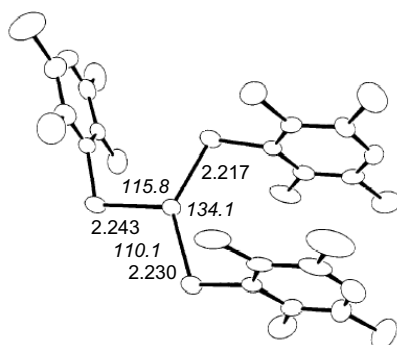


Figure IV.6 : Structure cristalllographique de B_E (Gruff 1989)
(distances en Å et angles en °)

L'obtention d'une structure tri-coordonnée du zinc n'était pas attendue étant donné l'ajout de cinq équivalents de thiolates et les connaissances en chimie de coordination, indiquant que le zinc coordonne facilement au moins quatre ligands. On peut noter que la coordination par trois ligands thiolates a également été observée pour d'autres métaux d¹⁰ : Cd(II) (Gruff 1990, Santos 1991), Hg(II) (Gruff 1990, Santos 1991, Alsina 1992, Watton 1990, Christou 1984, Bowmaker 1984), Cu(I) (Fujisawa 2000, Okamura 1993, Garner 1984, Coucouvanis 1980) et Ag(I) (Fujisawa 2000, Tang 1987).

Cependant, malgré ces différents travaux expérimentaux, les causes de la formation d'une espèce tricoordonnée du zinc ou du passage à une structure tétra-coordonnée restent floues. La nature donatrice des ligands et leurs effets stériques semblent avoir une certaine influence sur cet équilibre. En effet, Gruff et Koch notent qu'il a été nécessaire d'utiliser des ligands stériquement encombrés pour obtenir la fixation de seulement trois ligands au métal (Gruff 1990). Il faut mentionner cependant que cet encombrement stérique reste limité puisque, s'il n'est pas possible d'ajouter dans la sphère de coordination du métal un quatrième ligand $\text{SPh}^{\text{Me}4}$, ni une molécule d'acétonitrile (le solvant) comme observé pour son analogue du cadmium (Gruff 1990), le complexe $[\text{Zn}(\text{SPh}^{\text{Me}4})_3(\text{ImMe})]^-$ peut être synthétisé. En effet, l'ajout de 1-méthylimidazole dans le milieu réactionnel conduit à la molécule tétraédrique distordue $[\text{n-Pr}_4\text{N}][\text{Zn}(\text{S-2,3,5,6-Me}_4\text{C}_6\text{H})_3(1\text{-Me-imid})]$.

Depuis cette synthèse, d'autres complexes tri-coordonnés du zinc, possédant divers ligands soufrés, azotés, oxygénés, neutres et/ou chargés négativement, sont apparus (Grützmacher 1992, Bochmann 1994, Vittal, J. J. 1996, Darensbourg 1999, Prust 2002). Ce n'est cependant qu'en 2005 qu'a été publié le second exemple de complexe de type ZnS_3^- (Matsunaga 2005). De plus, ce complexe $[\text{Zn}(\text{SAd})_3]^-$ (SAd =1-adamantanethiolate), noté C_E , possède des thiolates aliphatiques. Cette structure est très intéressante dans la perspective de modéliser et de comprendre la coordination du zinc en milieu biologique, donc par des ligands cystéinates qui sont des thiolates aliphatiques. Ceci permet également de s'affranchir de l'influence des groupes aromatiques qui étaient présents dans le complexe $\text{Zn}(\text{SPh}^{\text{Me}_4})_3^-$ et en particulier des interactions de π -stacking sur l'obtention du complexe tri-coordonné. L'impact de l'effet stérique des ligands dans la formation de complexe tri-coordonné peut alors être directement apprécié. Ce dernier complexe a été synthétisé dans l'acétonitrile à partir de chlorure de zinc, de trois équivalents de thiolate et un équivalent de base (Matsunaga 2005) (Figure IV.7).

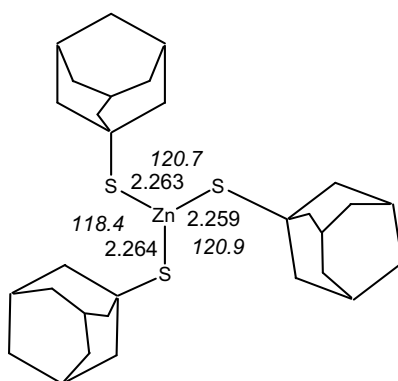


Figure IV.7 : Complexe $[\text{Zn}(\text{SAd})_3]^-$, noté C_E (Matsunaga 2005)

La structure plane trigonale est stabilisée par la demande stérique des ligands, ce qui évite une association supramoléculaire des complexes. Cette structure présente une symétrie C_3 différente de celle trouvée pour le composé $[\text{Zn}(\text{SPh}^{\text{Me}_4})_3]^-$ (Figure IV.6), pour lequel deux des substituants thiophénolates se situaient dans des plans parallèles à cause de leur interaction par π -stacking. Les substituants aliphatiques du complexe $[\text{Zn}(\text{SAd})_3]^-$ ne présentant pas cet effet électronique, préfèrent un arrangement trigonal avec des angles S-Zn-S valant chacun environ 120° . Par contre, l'encombrement stérique des ligands adamantanes est important dans cette conformation et explique l'allongement des longueurs de liaisons zinc-métal ($\text{Zn-S} = 2.262 \text{ \AA}$ en moyenne) par rapport à la structure $[\text{Zn}(\text{SPh}^{\text{Me}_4})_3]^-$ ($\text{Zn-S} = 2.230 \text{ \AA}$ en moyenne), pour laquelle la déviation par rapport à une symétrie plane trigonale laisse plus de place pour les ligands.

4. Dimères du zinc

L'utilisation de conditions expérimentales différentes et de ligands thiolates moins encombrés mène à l'obtention d'un complexe dimérique, contenant deux unités ZnS_2 connectées par deux ligands thiolates pontés (Figure IV.8). En effet, alors que l'utilisation d'un excès de thiol conduisait au composé $[\text{Zn}(\text{SPh})_4]^{2-}$ (Dance 1984), le dimère de zinc est obtenu à partir d'un rapport initial de 3 :1,

respectivement pour le thiol et le zinc. L'effet stérique des ligands soufrés influence ensuite la formation d'une espèce tri-coordonnée ou dimérique (Matsunaga, 2005).

Un dimère à zinc possède un anneau plan Zn_2S_2 et ses substituants sont en configuration *anti*. Chaque ion métallique est le centre d'un tétraèdre distordu où les angles S-Zn-S ont pour valeur moyenne 109.1° . Un tel motif a été observé pour le zinc avec trois ligands thiolates peu encombrés stériquement, tels que les groupements thioéthyl (Watson 1985), cyclohexanethiol (Matsunaga 2005) (Figure IV.8) et thiophényl (Abrahams 1987, Chung 1991). Il existe également des dimères pour les ions métalliques Cd(II) (Abrahams 1987, Watson 1985, Sola 1993, Bochmann 1990), Hg(II) (Bowmaker 1984 et 1996) et Cu(I) (Raper 1988).

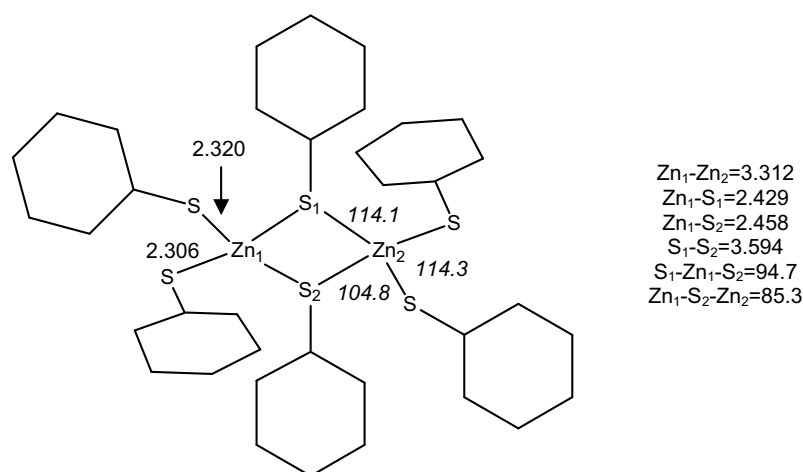


Figure IV.8 : Structure cristallographique $[Zn_2(ScHex)_6]^{2-}$ (HScHex=1-cyclohexanethiol) (Matsunaga, 2005)

Comme nous l'avons vu précédemment, le zinc possède préférentiellement un environnement tétraédrique dans les enzymes, comme par exemple dans les très nombreux sites $Zn(His)_2(Cys)_2^0$, $Zn(His)(Cys)_3^-$ et $Zn(Cys)_4^{2-}$ décrits dans la littérature (Vallee 1993, Alberts, 1998, Dudev 2003, Parkin 2004). Ces sites étant le plus souvent essentiels à la structuration tridimensionnelle de la protéine, ils possèdent une forte stabilité. A l'opposé, comparativement aux complexes neutres, il existe très peu d'exemples de complexes biomimétiques du zinc possédant une charge anionique, en particulier les sites de type ZnS_4^{2-} . En effet, dans de nombreux cas, une structure dimérique (ou polymérique) de type $Zn_2S_6^{2-}$ semble préférée (Watson 1985, Abrahams 1987), chaque atome de zinc possédant un environnement tétraédrique et monoanionique. Cet effet de polymérisation est cohérent avec l'observation de protéines polynucléaires à zinc, comme les métallothionéines. Ces protéines sont fréquemment composées des motifs Zn_3S_9 ou Zn_4S_{11} correspondant respectivement à une association de trois ou quatre motifs de type ZnS_4^{2-} (Henkel 2004). Si les groupes thiolates sont stériquement encombrants, alors une coordination de type ZnS_4^{2-} ou $Zn_2S_6^{2-}$ semble ne pas pouvoir être atteinte, même si il est possible d'obtenir une coordination 4 de type $ZnNS_3^-$ (Gruff 1989).

Ces observations nous laissent suggérer différents équilibres entre les espèces du zinc (Schéma IV.2), avec une différence notable et inexpliquée entre les systèmes chimiques et

biologiques. Au départ de ZnS_3^- , l'équilibre est déplacé vers la droite dans le premier cas et vers la gauche dans le second cas.

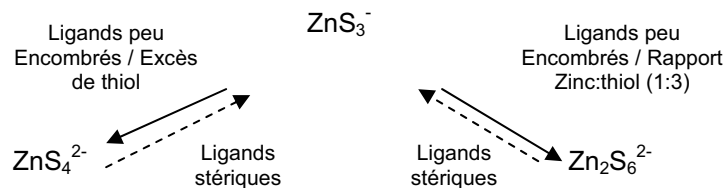


Schéma IV.2

B. Méthodes théoriques utilisées dans la littérature

Afin d'étudier la coordination et la réactivité des complexes à zinc, il est essentiel d'utiliser une méthode quantique fiable pour décrire la liaison zinc-soufre. De très nombreuses études théoriques ont montré que la liaison zinc-soufre est correctement décrite avec une fonctionnelle hybride comme B3LYP (Sousa 2007, Tamames 2007, Dimakis 2008, Simonson 2002, Rayon 2008). De même, nous avons montré dans le chapitre précédent que le niveau de calcul B3LYP/BS1 permet de reproduire correctement les structures cristallographiques des complexes **1** et **2**.

Une exception a cependant été décrite dans le cas d'un complexe ZnS_4^{2-} (Dudev 2001). La structure de la base de Cambridge servant de cible pour cette calibration de méthode est l'entrée DUFYUY qui correspond au complexe de zinc (II) bis-(ethane-1,2-dithiolato-S,S') (Rao 1986) (Figure IV.9). La longueur de liaison Zn-S moyenne, dans ce composé, est de 2.34 Å, valeur identique à la moyenne évaluée d'après 22 sites actifs ZnCys_4 de la Protein Data Bank (Dudev 2001) et très proche de celle observée dans les structures cristallographiques de $[\text{Zn}(\text{SPh})_4]^{2-}$ (Tableau IV.1). Plusieurs niveaux de théories (HF, MP2, SVWN, B3LYP, B3P86 et BVWN) et bases atomiques (de 6-31+G(d) à 6-311++G(2d,2p)) ont été testées sur cette molécule. L'objectif des auteurs était de déterminer le niveau de calcul nécessaire à la reproduction de cette structure cristallographique, pour ensuite étudier la protonation des sites ZnS_4^{2-} .

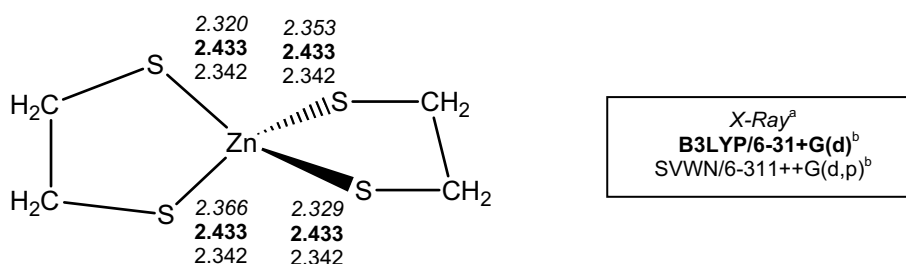


Figure IV.9 : Comparaison des distances Zn-S expérimentales et théoriques

a) Rao 1986, b) Dudev 2001

Il a été trouvé que la méthode HF ainsi que les fonctionnelles incluant un terme d'échange HF telles que BVWN, B3LYP et B3P86 surestiment les longueurs des liaisons dans le complexe, notamment pour les distances Zn-S qui peuvent différer de plus de 0.1 Å par rapport aux valeurs expérimentales (Figure IV.9). Les angles sont mieux reproduits puisqu'ils présentent des écarts de 1 à 3° par rapport à leur valeur expérimentale. De plus, augmenter la taille de la base n'améliore pas la concordance des résultats. De façon surprenante, aucune fonctionnelle GGA n'a été testée lors de cette calibration.

Par contre, les méthodes MP2 et SVWN s'avèrent mieux reproduire les paramètres structuraux du complexe. Au niveau MP2, les longueurs des liaisons Zn-S ne diffèrent au plus que de 0.02 Å des valeurs cristallographiques. Quant à la méthode SVWN, elle reproduit les valeurs expérimentales avec un écart maximum de 0.03 Å (Figure IV.9). La qualité de la base a, là encore, assez peu d'influence sur les performances de MP2 et SVWN. De cette calibration, Dudev et Lim concluent que la corrélation électronique est nécessaire pour une bonne reproduction des paramètres géométriques. Cependant, cette condition n'est pas suffisante puisque B3LYP ne s'avère pas être très performante, vraisemblablement à cause de la présence du terme d'échange HF.

SVWN/6-311++G(d,p) (plus rapide que MP2) a été déterminée comme méthode de choix pour l'étude des complexes zinc-soufre. Dans cette étude, l'utilisation d'une base triple- ξ avec fonctions diffuses et de polarisation, peu influente sur la reproduction de la géométrie des complexes à zinc, est cependant nécessaire pour les calculs d'énergies libres de déprotonation effectués sur les thiols. Ce résultat (fiabilité de SVWN et déficience de B3LYP), en contradiction avec de nombreuses autres études (Sousa 2007, Tamames 2007, Dimakis 2008, Simonson 2002, Rayon 2008), n'a pas été confirmée sur d'autres systèmes chimiques. Nous envisageons donc de tester la fonctionnelle B3LYP au cours des calculs suivants.

II. Mise en évidence de la dissociation de $[\text{Zn}(\text{SPh})_4]^{2-}$

A. $[\text{Zn}(\text{SPh})_4]^{2-}$ dans la littérature

Comme indiqué précédemment, il existe dans la littérature très peu d'exemples de complexes biomimétiques du zinc reproduisant les sites ZnS_4^{2-} . L'unique étude de réactivité de ce type de complexe a été effectuée avec le composé $[\text{Zn}(\text{SPh})_4]^{2-}$. L'objectif des auteurs était d'étudier la nucléophilie des sites $\text{Zn}(\text{His})_2(\text{Cys})_2^0$, $\text{Zn}(\text{His})(\text{Cys})_3^-$ et $\text{Zn}(\text{Cys})_4^{2-}$ par le biais de composés biomimétiques présentant une première sphère de coordination équivalente, c'est-à-dire respectivement $[\text{Zn}(\text{SPh})_2(\text{ImMe})_2]^0$ (ImMe=1-méthylimidazole), $[\text{Zn}(\text{SPh})_3(\text{ImMe})]^-$ et $[\text{Zn}(\text{SPh})_4]^{2-}$ (Wilker 1995). Il en résulte que le complexe dianionique présente des constantes de vitesses bien plus importantes pour la méthylation par $(\text{MeO})_3\text{P}(\text{O})$ que les autres complexes. Le taux de dissociation du ligand benzénethiolate a été estimé à partir de l'intégration des pics de la région aromatique des spectres ^1H RMN et indique que la vitesse d'alkylation des complexes à zinc augmente avec le nombre

de thiolates (Wilker 1997). Ceci reflète donc la plus grande capacité des sites protéiques dianioniques à subir des alkylations.

Cette facilité à être alkylé a été étudiée plus précisément d'un point de vue mécanistique: des mesures cinétiques et de conductivité ont été effectuées pendant la réaction équimolaire de $[\text{Zn}(\text{SPh})_4]^{2-}$ et $(\text{MeO})_3\text{P}(\text{O})$ dans le solvant DMSO (Wilker 1997). Alors que le benzénethiol ne conduit à aucune réaction en présence de l'agent alkylant, le benzénethiolate réagit très rapidement. A partir des constantes de vitesse d'alkylation des composés $[\text{Zn}(\text{SPh})_4]^{2-}$ et ^-SPh , il en a été déduit qu'un thiolate coordonné au métal dans un complexe à zinc de charge totale -2 montre environ 70% de la réactivité d'un thiolate libre. La réactivité nucléophile importante mesurée pour le complexe $[\text{Zn}(\text{SPh})_4]^{2-}$ a donc été attribuée, par les auteurs, au thiolate dissocié. De plus, la charge nette -2 semble promouvoir la réactivité des ligands benzénethiolates envers l'alkylation, en facilitant leur dissociation du complexe à zinc.

Selon ces études (Wilker 1995 et 1997), l'équilibre existant entre les espèces tétra-coordonnée et tri-coordonnée du zinc semble être déplacé en faveur de l'espèce monoanionique $[\text{Zn}(\text{SPh})_3]^-$. La stabilité cinétique de $[\text{Zn}(\text{SPh})_4]^{2-}$ serait alors assez faible. Cependant, deux structures cristallographiques de $[\text{Zn}(\text{SPh})_4]^{2-}$ ont été obtenues (cf IV.A.1), ce qui semble indiquer une plus grande stabilité thermodynamique de ce complexe par rapport à la dissociation d'un groupe thiolate.

Dans cette partie, nous nous sommes efforcés de décrire la dissociation aisée du complexe $[\text{Zn}(\text{SPh})_4]^{2-}$ et d'évaluer la stabilité relative des espèces $[\text{Zn}(\text{SPh})_3]^-$ et $[\text{Zn}(\text{SPh})_4]^{2-}$.

B. Calculs préliminaires sur $[\text{Zn}(\text{SPh})_4]^{2-}$

Dans un premier temps, afin d'étudier la dissociation d'un groupe thiolate à partir de $[\text{Zn}(\text{SPh})_4]^{2-}$, nous avons utilisé la même méthodologie que dans le chapitre précédent, à savoir une optimisation des géométries au niveau de calcul B3LYP/BS1 suivi d'un simple point au niveau B3LYP/BS2. Pour limiter le temps de calcul, aucun contre-ion n'a été ajouté. Deux minima A_{C1} et A_{C2} , correspondant respectivement aux structures A_{E1} et A_{E2} de la CSD, ont été déterminés (Figure IV.10). La symétrie des complexes est similaire à l'entrée de la CSD, à savoir D_{2d} pour A_{E1} et A_{C1} et S_4 pour A_{E2} et A_{C2} . Ces deux conformations ont un écart d'énergie très faible de 0.6 kcal/mol au niveau B3LYP/BS2//B3LYP/BS1 et la plus stable correspond à la structure A_{C2} .

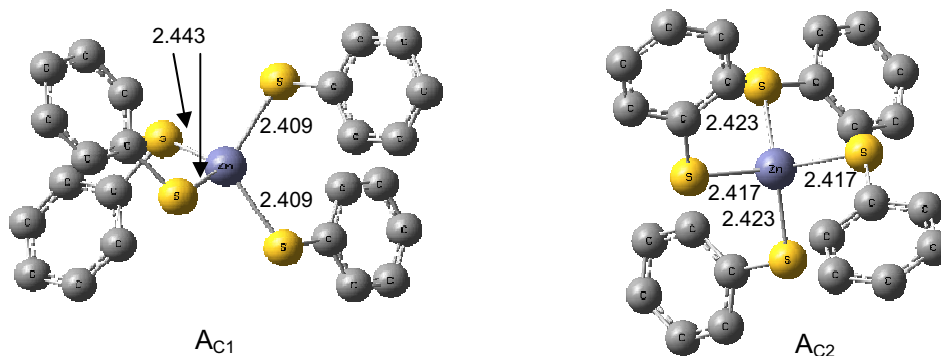


Figure IV.10 : Structures A_{C1} et A_{C2} calculées au niveau B3LYP/BS1 (distances en Å)

Afin d'étudier la dissociation de la liaison zinc-soufre, nous avons calculé un profil énergétique à partir de la structure A_{C2} par élongation progressive d'une liaison métal-thiolate (Figure IV.11). Des simples points d'énergie ont été effectués au niveau B3LYP/BS1 sur des structures non-optimisées dont la liaison zinc-soufre, initialement à 2.423 Å, a été contrainte à une valeur fixe comprise entre 2.423 Å et 20 Å avec un pas variant entre 0.1 Å (autour d'une longueur de 5 Å) et 2.5 Å (au-delà de 10 Å). Le résultat obtenu (Figure IV.11) est inattendu. En effet, la courbe obtenue ne présente pas l'allure d'une dissociation « classique » entre un fragment organométallique [M] et un ligand L à deux électrons.

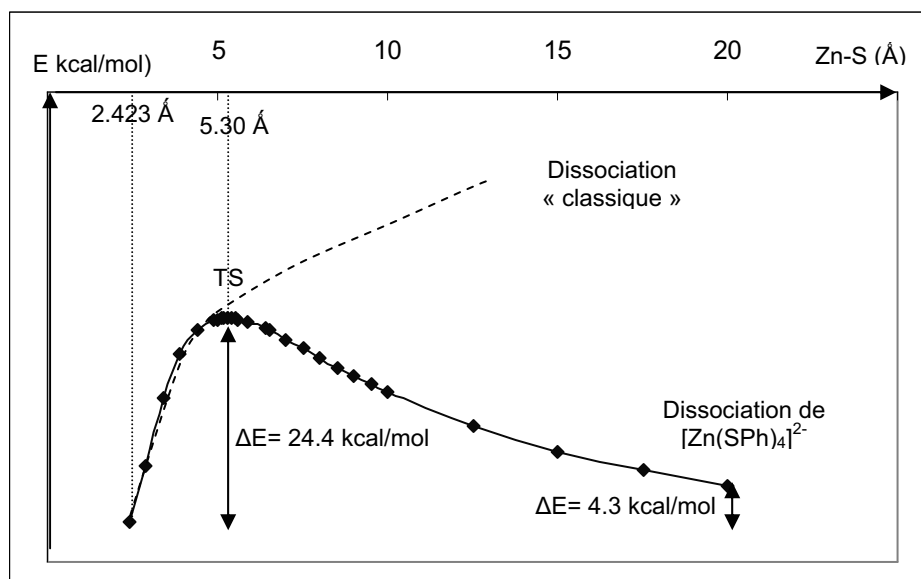


Figure IV.11: Profil énergétique de la dissociation de A_{C2} (simples points au niveau B3LYP/BS1)

Dans une dissociation "classique", l'énergie augmente continûment jusqu'à atteindre à dissociation complète, une valeur maximale qui, par différence avec l'énergie du complexe, donne l'énergie de dissociation avec réarrangement des fragments (ou énergie d'interaction). Ainsi, la liaison métal-thiolate a été étudiée par Solomon *et al.* dans un complexe modèle $[ML(SC_6F_5)]$ (L=hydrotris(3,5-diisopropyl-1-pyrazolyl)borate) (Gorelsky 2005). Pour le zinc, l'énergie de l'interaction entre les fragments ML^+ et $SC_6F_5^-$ a été estimée à 132.7 kcal/mol au niveau de calcul B3LYP/6-311+G(3df). Le zinc possède, de tous les métaux considérés dans cette étude (Mn, Fe, Co, Ni et Cu), la plus forte interaction entre l'ion métallique et le ligand thiolate (Gorelsky 2005, Solomon 2006).

Dans le cas étudié ici, la dissociation passe par un maximum d'énergie ($\Delta E = 24.4$ kcal/mol), indiquant vraisemblablement la présence d'un état de transition sur le chemin de la dissociation, lorsque la distance métal-ligand est d'environ 5 Å. A partir de ce point, l'énergie se met à décroître lorsque l'on allonge la liaison zinc-soufre et à grande distance, elle atteint une valeur proche de l'énergie du complexe initial ($\Delta E = 4.3$ kcal/mol). Nous pouvons constater que même à une distance supérieure à 15 Å, l'influence entre les deux partenaires semble non négligeable, vraisemblablement du fait des charges négatives qu'ils portent.

Nous avons alors étudié les deux fragments SPh^- et $[\text{Zn}(\text{SPh})_3]^-$ en optimisant séparément leurs géométries. La structure optimisée du fragment $[\text{Zn}(\text{SPh})_3]^-$ notée D_C , est représentée dans la Figure IV.12. Elle a une conformation similaire au complexe $[\text{Zn}(\text{SPh}^{\text{Me4}})_3]^-$ (Structure B_E , Figure IV.6), même si les ligands sont légèrement différents du fait de l'absence des substituants méthyles. Ainsi, on retrouve une interaction de type π -stacking entre deux des groupes phényles. Celle-ci n'est donc pas influencée par les substituants des cycles aromatiques.

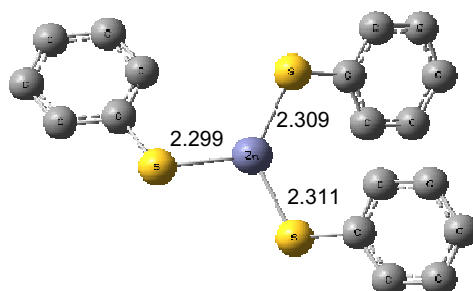


Figure IV.12 : Structure D_C calculée au niveau B3LYP/BS1 (distances en Å)

Ces calculs nous permettent d'estimer l'énergie de la réaction : $[\text{Zn}(\text{SPh})_4]^{2-} \rightarrow [\text{Zn}(\text{SPh})_3]^- + \text{SPh}^-$. Au niveau de calcul B3LYP/BS2//B3LYP/BS1, cette réaction est fortement exothermique. L'ensemble des produits dissociés, à savoir $[\text{Zn}(\text{SPh})_3]^-$ et PhS^- , est plus stable de 30.7 kcal/mol en phase gaz que le réactif non dissocié $[\text{Zn}(\text{SPh})_4]^{2-}$. Ce résultat est cohérent avec le profil énergétique de la figure IV.11. Avant l'optimisation des fragments, l'énergie d'interaction à grande distance était très faible. Après optimisation, elle devient négative puisque les fragments sont stabilisés.

Ces résultats, même cohérents entre eux, restent néanmoins surprenants vis-à-vis de la stabilité thermodynamique des sites ZnS_4^{2-} et soulèvent plusieurs questions.

C. Problématiques

La première question porte sur la précision et la fiabilité du niveau de calcul utilisé pour cette étude préliminaire, que ce soit pour la géométrie des complexes étudiés ou pour le calcul de l'énergie de dissociation. En effet, la comparaison dans le complexe $[\text{Zn}(\text{SPh})_4]^{2-}$ entre les longueurs de liaisons zinc-soufre expérimentales (2.355 ± 0.002 Å, Tableau IV.1) et théoriques (valeurs entre 2.409 et 2.443 Å, Figure IV.10) montre une différence importante qui peut avoir une influence forte sur l'énergie de cette liaison.

Si, comme nous allons le montrer, ce résultat n'est pas lié au niveau de calcul employé, nous pouvons alors nous interroger sur la généralisation de cette métastabilité à d'autres complexes du zinc, mais aussi sur l'influence de l'environnement. En effet, l'observation expérimentale en solution de $[\text{Zn}(\text{SPh})_4]^{2-}$, et non de $[\text{Zn}(\text{SPh})_3]^-$, permet d'émettre l'hypothèse que $[\text{Zn}(\text{SPh})_4]^{2-}$ n'est métastable qu'en phase gazeuse. Plus généralement, cette hypothèse semble également appropriée pour

expliquer la stabilité des sites enzymatiques à zinc de type ZnCys_4^{2-} qui semble en contradiction avec la métastabilité de $[\text{Zn}(\text{SPh})_4]^{2-}$.

Enfin, nous étendrons notre étude à des modèles de sites actifs de type $\text{Zn}(\text{His})_n(\text{Cys})_{4-n}$ afin de comprendre l'influence de la nature des ligands du zinc sur l'énergie de dissociation de la liaison zinc-soufre.

III. Structures de $[\text{Zn}(\text{SMe})_4]^{2-}$ et $[\text{Zn}(\text{SPh})_4]^{2-}$

Afin d'étudier la dissociation des complexes du zinc et de confirmer, ou non, les résultats surprenants obtenus précédemment, il est important de vérifier que la méthode théorique utilisée est adéquate. Nous avons précédemment montré que le niveau de calcul B3LYP/BS1 conduit pour le complexe $[\text{Zn}(\text{SPh})_4]^{2-}$ à une surestimation d'environ 0.06 Å des longueurs de liaisons zinc-soufre par rapport à la structure cristallographique (2.420 contre 2.357 Å). Si cette surestimation est liée à la fonctionnelle ou la base utilisée, ceci peut conduire à des énergies de dissociation erronées puisque la longueur d'une liaison reflète également sa capacité à se rompre. Il est également envisageable que cette surestimation soit due à la non prise en compte de l'environnement (contre-ions, packing dans le cristal...)

Dans le tableau IV.2 sont reportées les valeurs moyennes des distances zinc-soufre des structures expérimentales ou optimisées à différents niveaux de calcul. L'environnement du complexe organométallique a également été pris en compte avec l'inclusion de contre-ions ou d'un effet de solvant par un continuum diélectrique. Les diverses méthodes théoriques ont plus particulièrement été testées sur le complexe $[\text{Zn}(\text{SMe})_4]^{2-}$ afin de réduire les temps de calcul.

Tableau IV.2 a) Ueyama 1988

	Structure	Méthode d'optimisation	Zn-S moyen dans $\text{Zn}(\text{SR})_4^{2-}$ (Å)	E_{rel} B3LYP/BS2 (MP2/BS2)	Zn-S moyen dans $\text{Zn}(\text{SR})_4^{2-}$ (Å)	E_{rel} B3LYP/BS2 (MP2/BS2)
1	$[\text{Zn}(\text{SMe})_4]^{2-}$	B3LYP/BS1	2.432	0.0 (0.0)	2.295	0.0 (0.0)
2	$[\text{Zn}(\text{SMe})_4]^{2-}$	B3LYP/BS1+	2.434	-0.9	2.295	-0.5
3	$[\text{Zn}(\text{SMe})_4]^{2-}$	B3LYP/BS1*	2.434	-1.3	2.298	-0.3
4	$[\text{Zn}(\text{SMe})_4]^{2-}$	B3LYP/augcc-pVTZ(C,S,H)/ 6-311+G(3d,f)(Zn)	2.433	/	2.293	/
5	$[\text{Zn}(\text{SMe})_4]^{2-}$	B3LYP/PB	2.434	/	/	/
6	$[\text{Zn}(\text{SMe})_4]^{2-}$	MP2/BS1	2.389	+3.3 (+3.8)	2.295	+2.3 (+2.5)
7	$[\text{Zn}(\text{SMe})_4]^{2-}$	SVWN/6-311++G(d,p)	2.334	+8.0 (+13.6)	2.222	+4.2 (+7.3)
8	$[\text{Zn}(\text{SMe})_4]^{2-}$	BLYP/BS1	2.457	/	2.312	/
9	$[\text{Zn}(\text{SMe})_4]^{2-}$	BP86/BS1	2.419	/	2.285	/
10	$[\text{Zn}(\text{SMe})_4]^{2-}$	BP86/PB	2.419	/	/	/
11	$[\text{Zn}(\text{SMe})_4]^{2-}$	RI-BP86/PB	2.419	/	/	/
12	$[\text{Zn}(\text{SMe})_4]^{2-}$	MARI-J-BP86/PB	2.419	/	2.286	/
13	$[\text{Zn}(\text{SMe})_4]^{2-}$	COSMO-MARI-J-BP86/PB	2.385	/	2.279	/
14	$^{A_{C2}}[\text{Zn}(\text{SPh})_4]^{2-}$	B3LYP/BS1	2.420	/	2.306	/

15	$[\text{Zn}(\text{SPh})_4]^{2-}$	MARI-J-BP86/PB	2.406	/	2.304	/
16	$[\text{Zn}(\text{SPh})_4]^{2-}$	COSMO- MARI-J-BP86/PB	2.383	/	2.297	/
17	$\text{A}_{\text{E}2}$ $[\text{N}(\text{Me})_4]_2[\text{Zn}(\text{SPh})_4]$	X-ray ^a	2.357	/	/	/
	$[\text{N}(\text{Me})_4]_2[\text{Zn}(\text{SMe})_4]$	B3LYP/BS1+	2.430	/	/	/

A. Influence de la base

En optimisant le complexe $[\text{Zn}(\text{SMe})_4]^{2-}$ avec la méthode B3LYP/BS1 (Tableau IV.2, Entrée 1) utilisée précédemment, la distance zinc-soufre est de 2.432 Å, très proche de la valeur moyenne de 2.420 Å obtenue pour $[\text{Zn}(\text{SPh})_4]^{2-}$.

Quatre autres bases ont été employées avec la même fonctionnelle B3LYP pour décrire le complexe $[\text{Zn}(\text{SMe})_4]^{2-}$. L'ajout de fonctions diffuses ou de polarisation supplémentaires sur les atomes lourds, avec respectivement les bases notées BS1+ et BS1*, conduisent à des distances Zn-S égales à 2.434 Å dans les deux cas (Entrées 3 et 2 respectivement). Ceci montre que la modification des descriptions atomiques a une influence insignifiante sur la valeur des liaisons zinc-ligand.

L'optimisation de géométrie a également été effectuée avec une base plus étendue de qualité triple- ξ et incluant des fonctions de diffuses et de polarisation sur tous les atomes. Elle utilise les fonctions de base augcc-pVTZ pour les atomes des ligands et 6-311+G(3d,f) pour le métal zinc. Malgré son extension beaucoup plus grande, cette base d'orbitales atomiques ne fait varier la distance zinc-soufre que d'un millièmètre d'angström par rapport à BS1 ou PB.

Le même résultat est obtenu avec la base def2-SVP, notée PB, disponible dans le logiciel TURBOMOLE (Entrée 5). Ces résultats montrent l'influence négligeable de la base d'orbitales atomiques utilisée pour optimiser la géométrie, ce qui est en accord avec l'étude de Dudev et Lim effectuée sur le complexe de zinc(II) bis-(éthane-1,2-dithiolato-S,S') (Dudev 2001).

Par la suite, nous nous limiterons donc le plus souvent aux bases BS1 et PB pour l'optimisation des géométries.

B. Influence de la méthode

Ayant montré que le choix de la fonction de base n'influe pas sur la structure de $[\text{Zn}(\text{SMe})_4]^{2-}$, d'autres méthodes que B3LYP ont été testées pour apprécier également leur influence.

Les deux fonctionnelles du gradient de la densité électronique, BLYP et BP86, donnent des valeurs de liaisons Zn-S égales respectivement à 2.457 et 2.419 Å (Entrées 8 et 9 respectivement). Les écarts par rapport à B3LYP sont plus significatifs que pour le changement de base d'orbitales atomiques, avec un allongement (un raccourcissement respectivement) de 0.025 Å (de 0.013 Å respectivement) pour BLYP (respectivement BP86).

Les approximations de la résolution de l'identité (RI et MARI-J) implémentées pour la méthode BP86 avec la base PB n'apportent pas d'évolution de la géométrie: la longueur de la liaison reste invariablement égale à 2.419 Å pour $[\text{Zn}(\text{SMe})_4]^{2-}$ (Entrées 10-12). Le même résultat est obtenu pour $[\text{Zn}(\text{SPh})_4]^{2-}$. L'optimisation de la géométrie au niveau MARI-J-BP86/PB conduit à un raccourcissement de 0.014 Å de la liaison zinc-soufre par rapport au calcul B3LYP/BS1 (Entrées 14 et 15). Le calcul au niveau MARI-J-BP86/PB en phase gaz permet donc de se rapprocher légèrement de la valeur expérimentale (Entrée 17), tout en restant surestimée de 0.05 Å.

Du fait de cet écart persistant entre les résultats obtenus avec les méthodes du gradient de la densité et la structure cristallographique, la méthode d'approximation locale de la densité SVWN a été testée, associée à la base 6-311++G(d,p), comme préconisée par Dudev et Lim (Dudev 2001). La valeur obtenue à ce niveau de calcul pour la longueur de la liaison zinc-soufre est de 2.334 Å (Entrée 7), ce qui correspond à un raccourcissement de 0.1 Å par rapport à B3LYP/BS1. Cette valeur de 2.334 Å ne peut être directement comparée à une valeur expérimentale puisque aucune structure cristallographique de $[\text{Zn}(\text{SMe})_4]^{2-}$ n'est disponible. Pour le complexe zinc(II) bis-(éthane-1,2-dithiolato-S,S') (Dudev 2001), la distance moyenne zinc-soufre expérimentale (2.342 Å) était exactement reproduite au niveau de calcul SVWN/6-311++G(d,p). Dans notre cas, si l'on considère que le passage de $[\text{Zn}(\text{SMe})_4]^{2-}$ à $[\text{Zn}(\text{SPh})_4]^{2-}$ conduit à un raccourcissement de la liaison zinc-soufre de 0.010-0.015 Å (résultat obtenu avec B3LYP/BS1 et MARI-J-BP86/PB), alors que SVWN/6-311++G(d,p) conduit à une valeur extrapolée de 2.320 Å pour Zn-S, sous-estimant de près de 0.04 Å la valeur expérimentale.

Sur la base de ces résultats, de ceux obtenus sur le complexe **1** (cf Chapitre III) et de ceux obtenus sur les complexes $[\text{Zn}(\text{SPh})_3(\text{ImMe})]^-$ et $[\text{Zn}(\text{SPh})_2(\text{ImMe})_2]^{2-}$ (voir ci-dessous et Figure IV.13), la confiance accordée à ce niveau de calcul par Dudev et Lim (Dudev 2001) nous semble donc exagérée.

Enfin, la méthode MP2 conduit à un résultat intermédiaire entre les fonctionnelles SVWN et B3LYP ou celles du gradient de la densité avec une valeur zinc-soufre de 2.389 Å (Entrée 6). De la même façon que pour SVWN, une longueur de liaison zinc-soufre de 2.375 Å peut être extrapolée au niveau MP2/BS1 pour le complexe $[\text{Zn}(\text{SPh})_4]^{2-}$, soit une légère surestimation de 0.02 Å par rapport à la valeur cristallographique.

Ces calculs montrent une grande disparité des résultats pour la longueur de la liaison zinc-soufre sans qu'il soit définitivement possible de trancher entre les différentes méthodes puisque aucune donnée structurale en phase gazeuse n'est disponible. Afin de mieux comprendre la disparité obtenue pour les longueurs de liaison zinc-soufre de ces complexes, nous avons calculé l'énergie de $[\text{Zn}(\text{SMe})_4]^{2-}$ aux niveaux B3LYP/BS2 et MP2/BS2 aux différentes géométries optimisées (Tableau IV.2). Ceci permet de mesurer l'énergie nécessaire pour faire varier de façon concertée les quatre longueurs de liaison zinc-soufre, et donc d'estimer la forme de la surface de potentiel autour du minimum. La valeur de l'énergie relative obtenue avec le niveau de calcul SVWN/6-311++G(d,p)

(Entrée 7) est un peu surestimée par rapport à celles obtenues avec MP2 (Entrée 6) et B3LYP (Entrées 1 à 3). En effet, Dudev et Lim ont souligné la défaillance de SVWN à reproduire des énergies de déprotonation pour les complexes à zinc et préconisaient alors l'emploi de la fonctionnelle B3LYP (Dudev 2001). De plus, si l'on retire la valeur obtenue avec la méthode SVWN, les résultats montrent une faible différence d'énergie ($\Delta E = 4.2$ kcal/mol) entre les différentes géométries optimisées, ce qui explique la grande disparité des longueurs de liaison zinc-soufre calculées.

Le tableau IV.2 nous indique également que la même disparité ($\Delta E = 2.8$ kcal/mol) est observée pour les complexes ZnS_3^- avec les différents calculs de B3LYP et MP2 (Entrées 1-3 et 6). Nous pouvons cependant noter que pour tous les niveaux de calculs, un fort raccourcissement de la liaison zinc-soufre dans ZnS_3^- est calculé (entre 0.094 et 0.145 Å) par rapport au complexe ZnS_4^{2-} , ce qui est en adéquation avec la tendance expérimentale (Tableau IV.1). Dans ce cas, aucune comparaison expérimentale directe ne peut être faite, ni pour $[\text{Zn}(\text{SMe})_3]^-$, ni pour $[\text{Zn}(\text{SPh})_3]^-$.

Pour effectuer celle-ci, le complexe $[\text{Zn}(\text{SPh})_4]^{2-}$ a également été calculé (Figure IV.13) au niveau MARI-J-BP86/PB. Si le positionnement des bras thiolates est correctement reproduit par rapport à la structure cristallographique (Figure IV.5), avec en particulier une interaction de type π -stacking entre deux des cycles aromatiques, les longueurs de liaisons zinc-soufre sont là encore surestimées de 0.07 Å en moyenne. Ainsi, le comportement des méthodes utilisées est similaire pour les complexes ZnS_4^{2-} et ZnS_3^- .

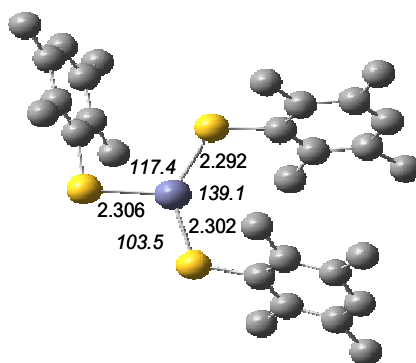


Figure IV.13 : Structure B_c calculée au niveau MARI-J-BP86/PB

Ceci nous montre que l'écart observé entre expérience et calculs n'est pas dû à la présence d'une double charge négative. La situation est donc différente de celle décrite récemment dans la littérature où la DFT échoue à reproduire la structure cristallographique d'un complexe dicationique du ruthénium mais y parvient très convenablement pour un complexe monocationique du ruthénium voisin du précédent (Calhorda 2007).

Ceci est confirmé par des calculs effectués sur les complexes $[\text{Zn}(\text{SPh})_3(\text{ImMe})]^-$ et $[\text{Zn}(\text{SPh})_2(\text{ImMe})_2]^0$ pour lesquels des structures cristallographiques existent. Sur la figure IV.14, se trouvent les structures cristallographiques des composés étudiés, ainsi que les structures optimisées au niveau B3LYP/BS1 et SVWN/6-311++G(d,p).

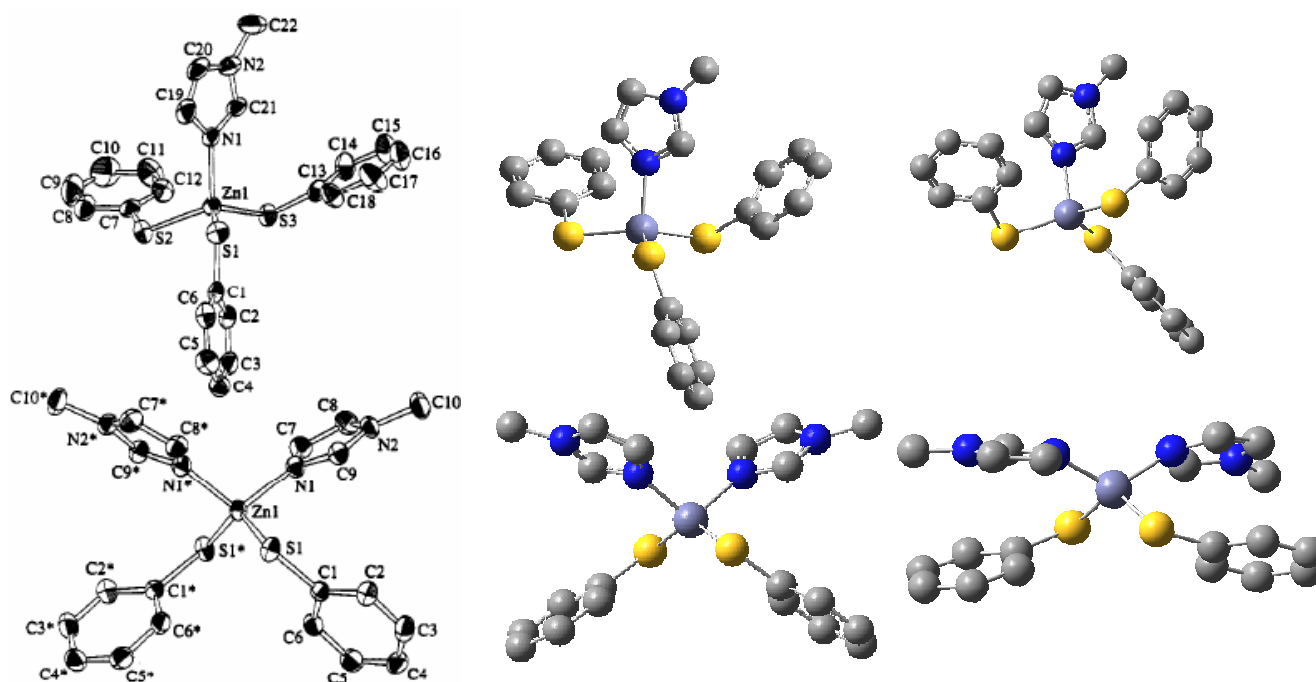


Figure IV.14 : Structures expérimentales (Wilker 1995) et théoriques de $[Zn(SPh)_3(ImMe)]^-$ et $[Zn(SPh)_2(ImMe)_2]$

Les conformations que nous avons obtenues avec la fonctionnelle B3LYP sont similaires à celles des structures cristallographiques. Les mêmes composés optimisés au niveau SVWN/6-311++G(d,p) ne correspondent toutefois pas à ces géométries, du fait de l'exagération de l'effet du π -stacking existant entre les ligands phényles. Ainsi, dans le complexe neutre $[Zn(SPh)_2(ImMe)_2]$, les cycles se rapprochent deux à deux, ce qui conduit à une distorsion de la forme tétraédrique du zinc. Ceci confirme l'imprécision de cette dernière méthode.

Tableau IV.3 : Distances moyennes Zn-S en Å

	<u>X-Ray</u>	<u>B3LYP/BS1</u>	<u>Ecart entre</u> <u>expérience et</u> <u>théorie (Å)</u>	<u>SVWN/6-</u> <u>311++G(d,p)</u>	<u>Ecart entre</u> <u>expérience et</u> <u>théorie (Å)</u>
$[Zn(SPh)_4]^{2-}$	2.357	2.420	0.063	/	/
$[Zn(SPh)_3(ImMe)]^-$	2.317 ^a	2.366	0.049	2.295	0.022
$[Zn(SPh)_2(ImMe)_2]$	2.292 ^a	2.315	0.023	2.276	0.016

a) Wilker 1995

Si l'on regarde de plus près les structures obtenues avec B3LYP/BS1, on remarque que la surestimation des distances ligand-métal diminue légèrement avec la charge du complexe. Ainsi dans le complexe dianionique $[Zn(SPh)_4]^{2-}$, la valeur calculée est supérieure de 0.063 Å à la valeur expérimentale. Cet écart reste important pour le complexe monoanionique (+ 0.049 Å). Il devient faible dans le cas du composé $[Zn(SPh)_2(ImMe)_2]$, comme cela avait déjà été observé pour le complexe **1** (Chapitre III). Ceci indique donc que B3LYP reproduit correctement les longueurs des liaisons métal-

soufre expérimentales dans les composés neutres, mais qu'elle les surestime systématiquement lorsque le complexe est chargé.

Deux hypothèses peuvent être avancées pour expliquer ce résultat : une déficience de B3LYP (et des fonctionnelles GGA, BP86 et BLYP) pour les complexes anioniques du zinc, ou un effet plus important de l'environnement et non négligeable pour les complexes anioniques par rapport aux complexes neutres. Cette seconde hypothèse est évaluée dans le paragraphe suivant.

C. Influence de l'environnement

Nous avons tout d'abord testé l'importance de la présence de contre-ions dans les calculs de géométries. Les structures cristallographiques de $[\text{Zn}(\text{SPh})_4]^{2-}$, A_{E1} et A_{E2} , issues de la base de données de Cambridge, ont été obtenues en présence de contre-ions tétraphénylphosphonium et tétraméthylammonium. Il semble alors nécessaire d'effectuer une optimisation d'un complexe métallique en présence de contre-ions, à savoir le composé $[\text{Zn}(\text{SMe})_4]^{2-}$ avec deux molécules de tétraméthylammonium. En effet, les contre-ions contribuent à des forces de packing différentes suivant leur nature et sont à l'origine, par exemple, des deux symétries distinctes A_{C1} et A_{C2} observées pour $[\text{Zn}(\text{SPh})_4]^{2-}$. L'optimisation a été effectuée avec B3LYP/BS1+, afin de tenir compte de façon plus exacte des charges des contre-ions et conduit à des valeurs de liaison métal-ligand égales à 2.430 Å dans le complexe $[\text{Zn}(\text{SMe})_4]^{2-}$. La différence est négligeable par rapport à la distance calculée avec B3LYP/BS1 sans contre-ions (Entrée 1). Cette vérification effectuée, on considérera par la suite que l'influence des contre-ions est très faible sur les géométries optimisées et donc que leur présence présente un intérêt très limité dans les calculs.

Nous avons également considéré l'impact du solvant sur les structures calculées. En effet, les structures cristallographiques nous servant de référence sont obtenues en solvant alors que par défaut, nous avons testé les méthodes théoriques en phase gazeuse. L'optimisation des complexes $[\text{Zn}(\text{SMe})_4]^{2-}$ et $[\text{Zn}(\text{SPh})_4]^{2-}$ a donc été effectuée en phase aqueuse, à l'aide de la méthode COSMO au niveau MARI-J-BP86/PB. Les valeurs des liaisons zinc-soufre obtenues sont égales à 2.385 Å pour $[\text{Zn}(\text{SMe})_4]^{2-}$ (Entrée 13) et 2.383 Å pour $[\text{Zn}(\text{SPh})_4]^{2-}$ (Entrée 16). Cette dernière valeur est plus proche de la valeur expérimentale (2.357 Å, Entrée 17) que de celle calculée en phase gazeuse (2.406 Å, Entrée 15), ce qui montre la nécessité de représenter le solvant dans les calculs de géométries. On observe un raccourcissement de 0.034 Å de la liaison zinc-soufre similaire à celui obtenu pour $[\text{Zn}(\text{SPh})_4]^{2-}$ (0.023 Å), lorsque $[\text{Zn}(\text{SMe})_4]^{2-}$ est optimisé en phase gazeuse (Entrée 12), puis aqueuse (Entrée 13).

D. Conclusions

Nous avons montré que la qualité des bases atomiques et la présence de contre-ions a peu d'importance pour la reproduction de la géométrie des complexes à zinc. Par contre, le choix de la

méthode et la prise en compte du solvant influent beaucoup sur les structures, notamment sur les longueurs des liaisons zinc-soufre.

Alors que la fonctionnelle MARI-J-BP86 donne des longueurs de liaisons légèrement surestimées par rapport aux données expérimentales, nous avons montré que cet écart peut être réduit par la prise en compte de l'environnement, notamment avec la méthode de solvation COSMO dans les calculs de géométries. Le traitement de l'environnement joue donc un rôle primordial dans l'adéquation de nos géométries avec les structures cristallographiques.

IV. Etude de la dissociation

A. Influence de la géométrie

Le phénomène qui nous intéresse est la dissociation des complexes dianioniques du zinc, puisqu'elle semble dévier du schéma traditionnel de la dissociation (Figure IV.11). Afin d'étudier la rupture de liaison d'un ligand du zinc, nous allons commencer par quantifier l'énergie de dissociation correspondante et vérifier que la géométrie n'influe pas sur le phénomène. En effet, les liaisons calculées avec MARI-J-BP86 et surtout B3LYP, sont plus longues que celles des structures cristallographiques et peuvent donc traduire une plus grande facilité à se rompre. Afin de vérifier que la dissociation aisée des complexes du zinc à quatre ligands soufrés existe réellement et n'est pas un artefact de la fonctionnelle choisie (B3LYP), nous avons effectué des calculs similaires à différents niveaux de théorie.

Tableau IV.4 : Energies de dissociation à différents niveaux de calculs

Structure	Méthode d'optimisation	Méthode de simple point d'énergie	E_{disso} (kcal/mol)
$\text{Ac}_2 : [\text{Zn}(\text{SPh})_4]^{2-}$	B3LYP/BS1	/	-29.1
	B3LYP/BS1	B3LYP/BS2	-30.7
	MARI-J-BP86/PB	/	-21.5
	MARI-J-BP86/PB	B3LYP/GB	-30.3
$[\text{Zn}(\text{SMe})_4]^{2-}$	B3LYP/BS1	/	-38.6
	B3LYP/BS1	B3LYP/BS2	-46.2
	B3LYP/BS1+	B3LYP/BS2	-45.3
	B3LYP/BS1*	B3LYP/BS2	-46.0
	B3LYP/BS1	MP2/BS2	-46.0
	MP2/BS1	/	-38.6
	MP2/BS1	B3LYP/BS2	-46.7
	SVWN/6-311++G(d,p)	/	-32.0
	SVWN/6-311++G(d,p)	B3LYP/BS2	-50.0
	SVWN/6-311++G(d,p)	MP2/BS2	-52.4
	MARI-J-BP86/PB	/	-31.9
	MARI-J-BP86/PB	B3LYP/GB	-45.9
	B3LYP/GB	/	-42.9

Tout d'abord, nous avons étudié l'impact des géométries obtenues par diverses méthodes d'optimisation, en comparant l'énergie de leur dissociation après un calcul final d'énergie au niveau B3LYP/BS2.

Nous avons précédemment montré que l'écart de géométrie était insignifiant lorsque nous avons modifié les fonctions de base, en utilisant des fonctions diffuses (base BS1+) ou des fonctions de polarisation supplémentaires (base BS1*). Il est donc attendu qu'il n'y ait pas d'impact sur l'énergie de dissociation puisque les liaisons ont une longueur semblable. En effet, les énergies après simple point au niveau B3LYP/BS2 ne sont que très légèrement différentes ($E_{\text{diss}} = -45.3$ et -46.0 kcal/mol respectivement) de celle trouvée avec la base BS1 de référence ($E_{\text{diss}} = -46.2$ kcal/mol).

De même, les structures optimisées avec différentes méthodes telles que B3LYP, SVWN, MP2 et MARI-J-BP86 ont des longueurs de liaisons zinc-soufre différentes (cf Tableau IV.2). Si on considère comme référence la valeur donnée par la méthode B3LYP/BS2//B3LYP/BS1, soit $E_{\text{diss}} = -46.2$ kcal/mol pour $[\text{Zn}(\text{SMe})_4]^{2-}$, on remarque que les structures obtenues avec MP2 et MARI-J-BP86 mènent à des énergies très proches de -46.7 et -45.9 kcal/mol, respectivement. Avec la méthode SVWN, l'écart est plus important ($E_{\text{diss}} = -50.0$ kcal/mol). Pour $[\text{Zn}(\text{SPh})_4]^{2-}$, nous retrouvons la même tendance : les structures optimisées avec B3LYP/BS1 et MARI-J-BP86 ont des énergies de dissociation semblables de -30.7 et -30.3 kcal/mol respectivement. Les géométries des structures n'ont pas d'impact significatif sur la valeur de l'énergie de dissociation.

Après avoir vérifié la cohérence des résultats après modification de la méthode d'optimisation, nous avons changé également le calcul final de l'énergie. Un changement de méthode de B3LYP/BS2 à MP2/BS2 conduit pour $[\text{Zn}(\text{SMe})_4]^{2-}$ à un écart de seulement 0.2 kcal/mol sur son énergie de dissociation. Ces méthodes sont donc de qualités semblables pour un simple point, comme le montrent les valeurs $E_{\text{diss}} = -50.0$ et -52.4 kcal/mol obtenues après optimisation des structures avec la méthode SVWN. Ces valeurs sont cependant supérieures à la moyenne des valeurs obtenues avec optimisation en MP2, B3LYP et BP86. Ceci est dû à la mauvaise adéquation de la valeur obtenue après optimisation avec SVWN ($E_{\text{diss}} = -32.0$ kcal/mol) par rapport à celles données par B3LYP et MP2 ($E_{\text{diss}} = -38.6$ kcal/mol). Il semble que la fonctionnelle SVWN n'est donc pas appropriée pour le calcul d'énergie de dissociation et les méthodes B3LYP et MP2 lui seront préférées. Dans la suite de notre étude, les simples points d'énergie seront effectués avec la fonctionnelle B3LYP, plus rapide pour les calculs que la méthode MP2.

B. Influence de la solvation

Nous avons vu dans la partie IV.C.3 que l'environnement, notamment la prise en compte du solvant, est primordial pour une bonne reproduction des géométries de $[\text{Zn}(\text{SMe})_4]^{2-}$ et $[\text{Zn}(\text{SPh})_4]^{2-}$. Nous allons maintenant étudier l'influence de la solvation dans le calcul de l'énergie de dissociation, ainsi que l'impact de la méthode choisie pour reproduire la phase aqueuse. Les énergies de dissociations calculées à différents niveaux de théorie, en phases gazeuse et aqueuse, sont reproduites ci-dessous pour $[\text{Zn}(\text{SMe})_4]^{2-}$ (Figure IV.15) et $[\text{Zn}(\text{SPh})_4]^{2-}$ (Figure IV.16).

Comme établi précédemment, l'énergie de dissociation en phase gazeuse est $E_{\text{diss}} = -46.0 \pm 0.2$ kcal/mol pour $[\text{Zn}(\text{SMe})_4]^{2-}$, les méthodes les plus appropriées pour ce calcul étant B3LYP/BS2//B3LYP/BS1 et B3LYP/GB//MARI-J-BP86/PB. Avec les mêmes méthodes, l'énergie de dissociation est de -30.5 ± 0.2 kcal/mol pour $[\text{Zn}(\text{SPh})_4]^{2-}$.

La première méthode de solvation que nous avons testée est la théorie CPCM implémentée dans le logiciel GAUSSIAN03 (cf Chapitre II) au niveau B3LYP-BS1+. Les énergies de dissociation sont alors de -18.5 kcal/mol pour $[\text{Zn}(\text{SMe})_4]^{2-}$ (Figure IV.15) et -9.6 kcal/mol pour $[\text{Zn}(\text{SPh})_4]^{2-}$ (Figure IV.16), ce qui laisse présager encore une grande aisance à la dissociation pour ces complexes. Or, d'après les observations expérimentales, il est possible d'isoler des complexes à zinc à trois ligands soufrés (Gruff 1989). Nos calculs entrent donc en contradiction avec l'expérience et ne peuvent expliquer comment se forme le complexe $[\text{Zn}(\text{SPh})_4]^{2-}$ à partir de l'espèce tri-coordonnée $[\text{Zn}(\text{SPh})_3]^-$.

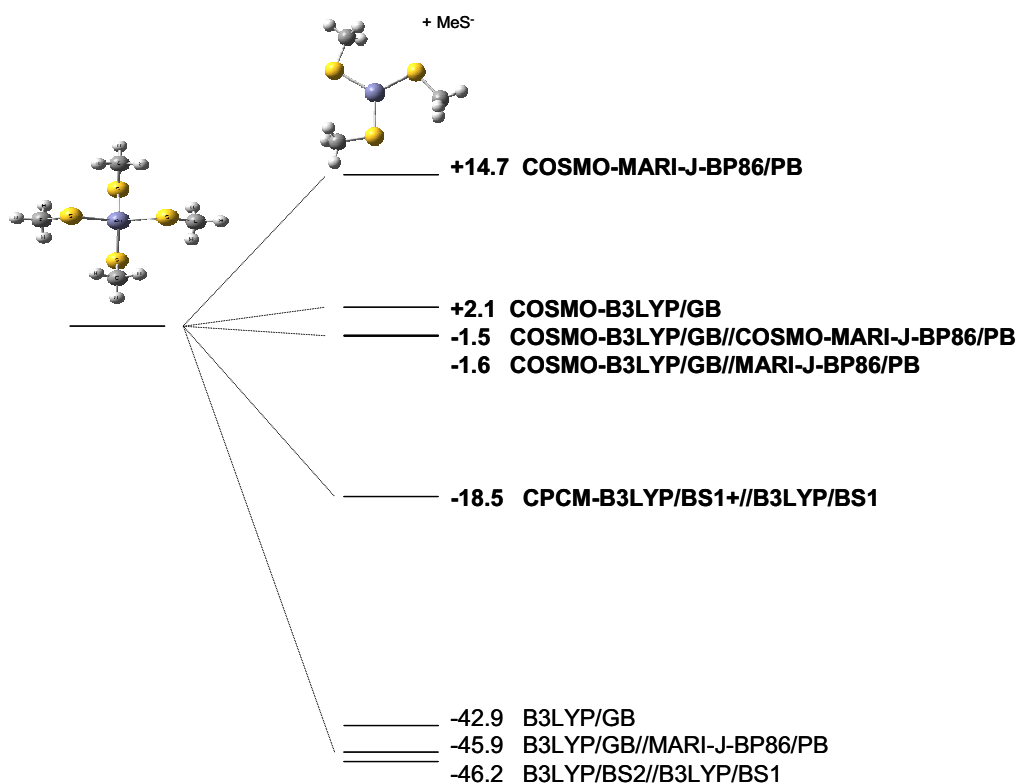


Figure IV.15 : Energies de dissociation de $[\text{Zn}(\text{SMe})_4]^{2-}$ en kcal/mol en phase gazeuse (normal) et phase aqueuse (gras)

Il semble donc nécessaire de vérifier la facilité de dissociation des complexes à zinc en phase aqueuse : la méthode COSMO a été utilisée pour calculer à nouveau l'énergie de dissociation. Tout d'abord, un simple point d'énergie au niveau COSMO-B3LYP/GB a été effectué sur les structures optimisées en phase gazeuse. Pour $[\text{Zn}(\text{SMe})_4]^{2-}$, l'énergie de dissociation devient alors proche de zéro ($E_{\text{diss}} = -1.6$ kcal/mol) (Figure IV.15) et pour $[\text{Zn}(\text{SPh})_4]^{2-}$, elle augmente également jusqu'à mener à une réaction endothermique ($E_{\text{diss}} = +8.8$ kcal/mol) (Figure IV.16). Ces résultats suggèrent un quasi-équilibre entre les complexes à zinc à trois ou quatre ligands thiométhyles et pour les espèces thiophényliques, une préférence pour le complexe dianionique à quatre ligands $[\text{Zn}(\text{SPh})_4]^{2-}$ plutôt que

pour le composé tri-coordonné $[\text{Zn}(\text{SPh})_3]^-$. Contrairement aux résultats obtenus avec la méthode CPCM, ces nouvelles données sont en accord avec l'expérience et montre que la théorie COSMO est plus adaptée pour calculer les énergies de dissociation en solvant.

D'autres calculs ont été effectués afin de vérifier la validité de la méthode COSMO et la cohérence des résultats obtenus. Ainsi, les structures ont été à nouveau optimisées en phase aqueuse. A un faible niveau de théorie (COSMO-MARI-J-BP86/PB), ces optimisations mènent à des énergies de dissociation largement supérieures aux données obtenues précédemment, à savoir +14.7 kcal/mol pour $[\text{Zn}(\text{SMe})_4]^{2-}$ et +19.0 kcal/mol pour $[\text{Zn}(\text{SPh})_4]^{2-}$. Avec une base de meilleure qualité (GB) et la méthode B3LYP, la structure optimisée de $[\text{Zn}(\text{SMe})_4]^{2-}$ en solvant conduit à une énergie de dissociation de +2.1 kcal/mol, ce qui est très proche de la valeur trouvée avec un simple point COSMO effectué sur une structure optimisée en phase gazeuse ($E_{\text{disso}} = -1.6$ kcal/mol) (Figure IV.15).

De même, des simples points COSMO effectués sur les nouvelles géométries obtenues à partir de l'optimisation en phase aqueuse mènent à des valeurs d'énergie permettant de reproduire la stabilité expérimentale des complexes à zinc à quatre ligands. Ainsi, $[\text{Zn}(\text{SPh})_4]^{2-}$ présente une dissociation endothermique ($E_{\text{disso}} = +7.9$ kcal/mol) au niveau COSMO-B3LYP/GB//COSMO-MARI-J-BP86/PB. Cette valeur est proche du résultat trouvé avec le simple point effectué sur la structure en phase gazeuse ($E_{\text{disso}} = +8.8$ kcal/mol) (Figure IV.16). Les mêmes conclusions peuvent être tirées des calculs sur $[\text{Zn}(\text{SMe})_4]^{2-}$, puisque l'énergie de dissociation est égale à -1.5 ou -1.6 kcal/mol, selon que le simple point en solvation est effectué sur une structure ré-optimisée ou non en solvant.

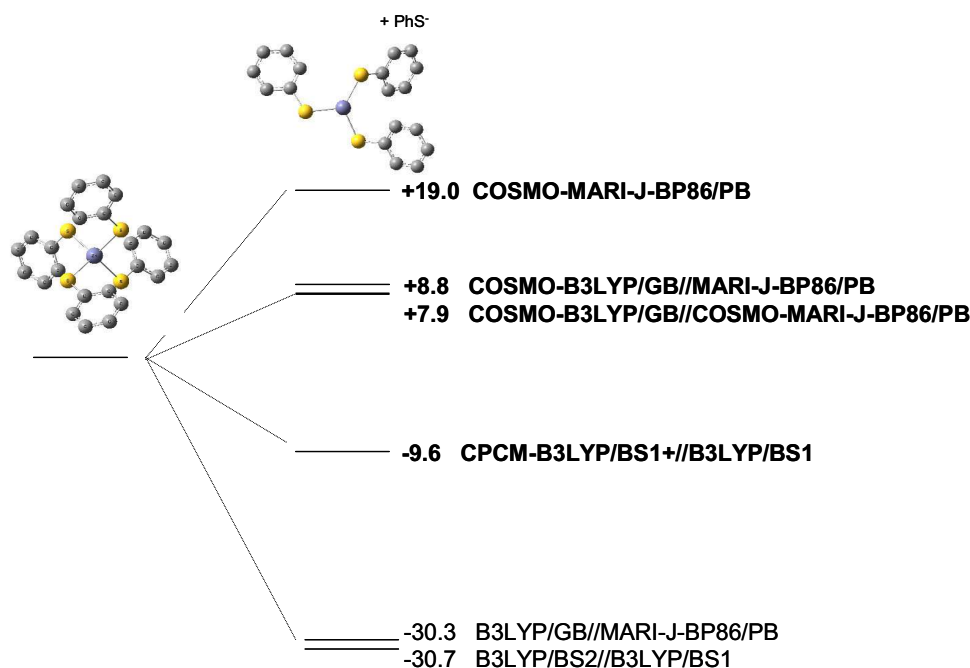


Figure IV.16 : Energies de dissociation de $[\text{Zn}(\text{SPh})_4]^{2-}$ en kcal/mol en phase gazeuse (normal) et phase aqueuse (gras)

Il semble donc nécessaire pour reproduire les résultats expérimentaux d'utiliser la méthode COSMO, soit en simple point d'énergie, soit en optimisation au niveau B3LYP/GB. Si on tient compte de la rapidité des calculs, la première solution, à savoir un simple point COSMO sur une structure optimisée en phase gazeuse, est la plus avantageuse.

C. Barrière d'activation de la dissociation

Afin de mieux comprendre la stabilité des complexes étudiés et l'équilibre existant entre les complexes tri-coordonnés et tétra-coordonnés, nous avons cherché à déterminer la barrière d'activation en phase aqueuse pour la dissociation de $[\text{Zn}(\text{SMe})_4]^{2-}$ et $[\text{Zn}(\text{SPh})_4]^{2-}$.

Nous avons montré que la méthode de solvation la plus appropriée est la théorie COSMO. Cependant, celle-ci n'est pas encore implémentée dans le logiciel TURBOMOLE pour permettre le calcul des dérivées secondes de l'énergie. Aussi, il est impossible d'effectuer des calculs de fréquence en solvant, ni d'optimiser des états de transition avec cette méthode.

Afin de détourner cette contrainte, nous déterminerons tout d'abord les énergies d'activation pour la dissociation en phase gazeuse. Un simple point d'énergie COSMO sera effectué sur l'état de transition, ainsi localisé en phase gazeuse. Une alternative consistera à effectuer un scan énergétique en phase aqueuse pour s'approcher de l'état de transition.

1. Complexe $[\text{Zn}(\text{SMe})_4]^{2-}$

L'état de transition déterminé avec la méthode B3LYP/BS2//B3LYP/BS1 révèle une barrière énergétique de +9.2 kcal/mol, ce qui montre que la dissociation en phase gazeuse est possible et aisée (Figure IV.17). La distance du ligand méthylthiolate au zinc dans l'état de transition est de 3.927 Å. Le ligand qui se détache se trouve au-dessus d'un plan où sont contenus les trois soufres toujours liés au métal et qui se rapprochent de la géométrie de l'espèce à trois ligands. L'état de transition optimisé au niveau B3LYP/GB//MARI-J-BP86/PB mène à une barrière énergétique très semblable ($E=+8.8$ kcal/mol), ce qui mène à une valeur moyenne de $E=+9.0 \pm 0.2$ kcal/mol pour $[\text{Zn}(\text{SMe})_4]^{2-}$. La dissociation de ce complexe est donc relativement aisée. La structure de l'état de transition est semblable à celle trouvée au niveau B3LYP/BS1 avec une valeur Zn-S=4.111 Å, cette valeur étant certainement la plus proche de la réalité, car utilisant la fonctionnelle du gradient de la densité BP86. Une optimisation supplémentaire a été effectuée au niveau B3LYP/GB : l'énergie d'activation y est légèrement surestimée par rapport aux autres calculs ($E=+10.4$ kcal/mol).

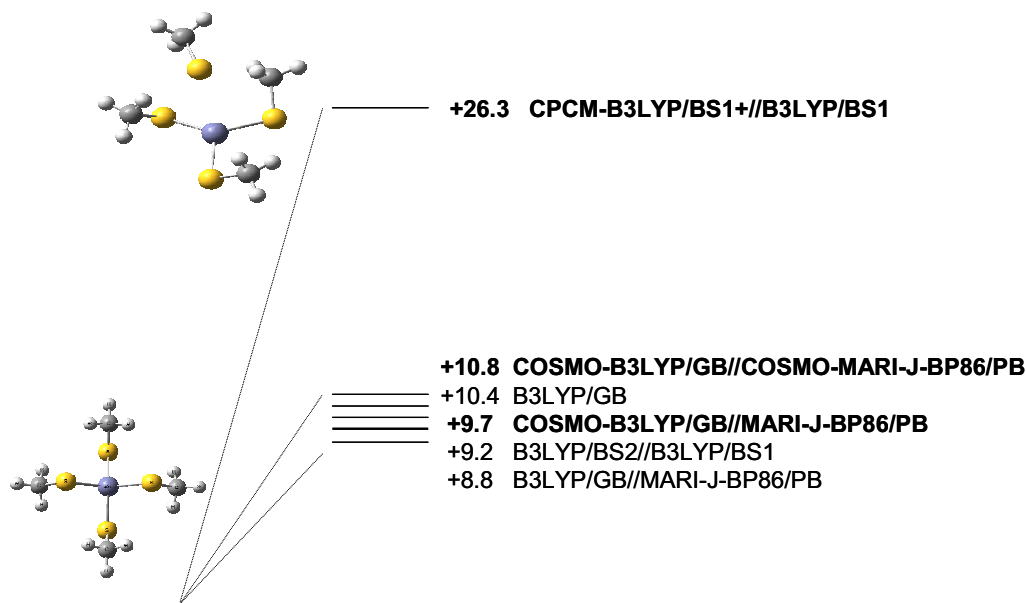


Figure IV.17 : Barrières d'activation pour la dissociation de $[\text{Zn}(\text{SMe})_4]^{2-}$ en kcal/mol

A partir de l'état de transition en phase gazeuse, nous avons déterminé la barrière d'activation en phase aqueuse par un simple point d'énergie COSMO. Ainsi au niveau COSMO-B3LYP/GB//MARI-J-BP86-PB, +9.7 kcal/mol sont nécessaires au complexe dianionique pour perdre un ligand et donner une structure tri-coordonnée (Figure IV.17).

Une deuxième solution pour déterminer l'état de transition en phase aqueuse est d'effectuer un scan énergétique au niveau COSMO-B3LYP/GB//COSMO-MARI-J-BP86/PB afin de trouver une structure s'en approchant. Les structures optimisées en solvant sont des structures de $[\text{Zn}(\text{SMe})_4]^{2-}$ pour lesquelles la liaison se détachant est contrainte à une valeur fixée (Figure IV.18). Le maximum énergétique de la dissociation est atteint pour un complexe ayant un ligand se détachant avec $\text{Zn-S} = 3.80 \text{ \AA}$ et se situant 10.8 kcal/mol au dessus du complexe $[\text{Zn}(\text{SMe})_4]^{2-}$. Les résultats obtenus avec cette méthode et le simple point COSMO effectué sur l'état de transition optimisé en phase gazeuse sont très proches et l'énergie d'activation estimée à $+10.3 \pm 0.6$ kcal/mol.

La dissociation de ce complexe métallique est donc relativement aisée et par conséquent suggère un quasi-équilibre entre les deux entités dans le milieu réactionnel. La première méthode sera privilégiée puisqu'elle évite l'étape supplémentaire de l'optimisation en solvant des structures contraintes pour le tracé du profil énergétique.

On peut remarquer que la barrière d'activation trouvée avec un simple point d'énergie CPCM sur les structures optimisées en phase gazeuse est deux fois plus importante (+26.3 kcal/mol au lieu de +10.3 kcal/mol). Ceci confirme le mauvais traitement de la solvation par la méthode CPCM et son inadéquation dans le calcul des énergies de dissociation des complexes à zinc (Figure IV.17).

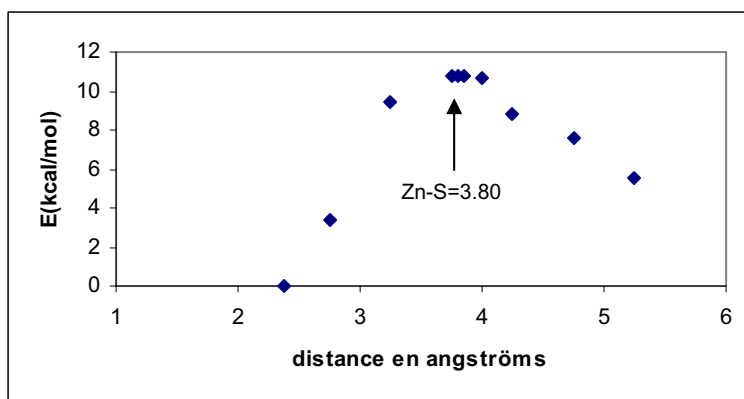


Figure IV.18 : Profil énergétique de la dissociation de $[\text{Zn}(\text{SMe})_4]^{2-}$ au niveau COSMO-B3LYP/GB//COSMO-MARI-J-BP86/PB

2. Complexe $[\text{Zn}(\text{SPh})_4]^{2-}$

Les états de transition en phase gazeuse ont été optimisés avec les méthodes B3LYP/BS1 et MARI-J-BP86/PB. Les longueurs de la liaison zinc-soufre se détachant sont respectivement de 4.012 Å et 4.381 Å et les conformations des états de transition sont semblables, malgré l'écart de 0.3 Å sur la distance ligand-métal. Les barrières énergétiques pour la dissociation sont respectivement de +11.1 kcal/mol et +12.3 kcal/mol aux niveaux B3LYP/BS2//B3LYP/BS1 et B3LYP/GB//MARI-J-BP86/PB. Pour $[\text{Zn}(\text{SPh})_4]^{2-}$, la valeur moyenne d'énergie d'activation $E = +11.7 \pm 0.6$ kcal/mol est proche de celle trouvée pour $[\text{Zn}(\text{SMe})_4]^{2-}$ ($E = +9.0 \pm 0.2$ kcal/mol). La dissociation de $[\text{Zn}(\text{SPh})_4]^{2-}$ est aussi possible en phase gazeuse, cependant un peu moins aisée que pour son homologue $[\text{Zn}(\text{SMe})_4]^{2-}$ (Figure IV.19).

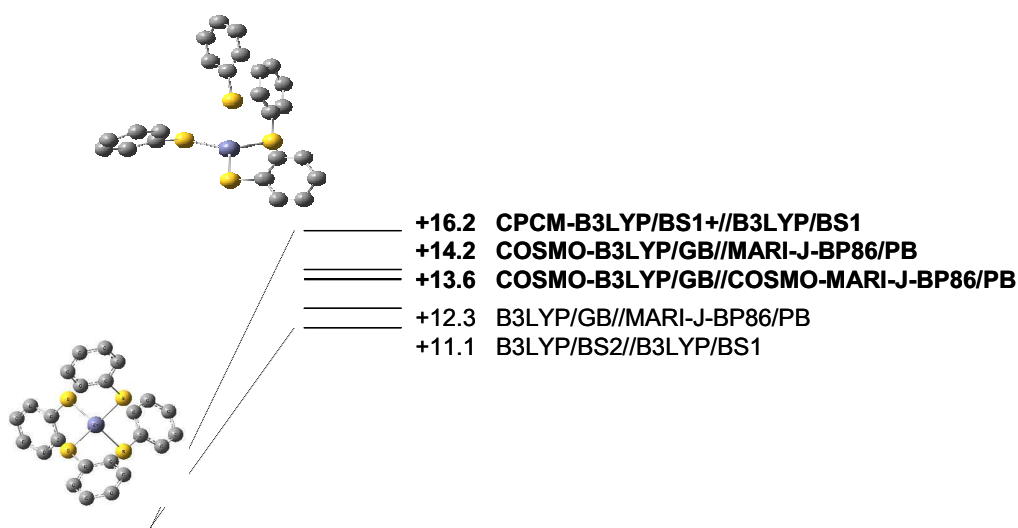


Figure IV.19 : Barrières d'activation pour la dissociation de $[\text{Zn}(\text{SPh})_4]^{2-}$ en kcal/mol en phase gazeuse (normal) et phase aqueuse (gras)

Du fait de la difficulté de trouver un état de transition correspondant à la dissociation de $[\text{Zn}(\text{SPh})_4]^{2-}$ en phase aqueuse, nous avons employé les mêmes méthodes que pour $[\text{Zn}(\text{SMe})_4]^{2-}$. Un simple point d'énergie COSMO effectué sur la structure précédemment trouvée en phase gazeuse au niveau MARI-J-BP86/PB mène à une barrière d'activation de +14.2 kcal/mol pour la dissociation de $[\text{Zn}(\text{SPh})_4]^{2-}$ (Figure IV.19).

Pour obtenir une structure se rapprochant de l'état de transition en solvant, un profil énergétique a également été tracé : des simples points au niveau COSMO-B3LYP/GB ont été effectués sur des structures optimisées en solvant ayant une élancement progressive de la liaison zinc-soufre (Figure IV.20). Le maximum d'énergie indique que la liaison métal-ligand est de 4.35 Å environ dans l'état de transition de la dissociation. La barrière énergétique avec cette méthode est de +13.6 kcal/mol. La dissociation de $[\text{Zn}(\text{SPh})_4]^{2-}$ requiert donc une énergie d'activation de $+13.9 \pm 0.3$ kcal/mol, ce qui indique que le ligand thiolate se détache assez aisément du complexe à zinc. La dissociation est cependant légèrement moins favorisée qu'avec les ligands méthylés ($E=+10.3 \pm 0.6$ kcal/mol) et le complexe dianionique $[\text{Zn}(\text{SPh})_4]^{2-}$ relativement un peu plus stable que $[\text{Zn}(\text{SMe})_4]^{2-}$.

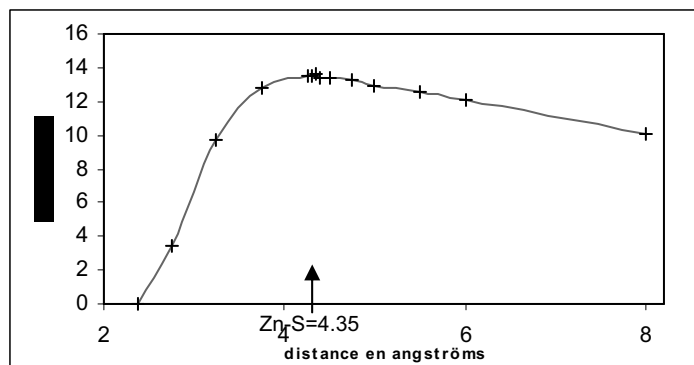


Figure IV.20 : Profil énergétique de la dissociation de $[\text{Zn}(\text{SMe})_4]^{2-}$ au niveau COSMO-B3LYP/GB//COSMO-MARI-J-BP86/PB

Le choix du traitement de la solvation pour la dissociation des complexes dianioniques à zinc est primordial puisque la méthode CPCM de GAUSSIAN03 ne permet pas de reproduire les observations expérimentales. En effet, la barrière d'activation est alors systématiquement surestimée ($E=+26.3$ kcal/mol pour $[\text{Zn}(\text{SMe})_4]^{2-}$ et $E=+16.2$ kcal/mol pour $[\text{Zn}(\text{SPh})_4]^{2-}$).

La théorie COSMO de TURBOMOLE montre cependant qu'il y a équilibre entre les entités à trois et quatre ligands soufrés en solution, avec une légère préférence pour la formation de $[\text{Zn}(\text{SPh})_4]^{2-}$, ce qui est en accord avec les observations expérimentales et la co-existence de ces deux espèces en solvant.

De plus, la réoptimisation des molécules avec COSMO n'est pas nécessaire, puisqu'un simple point d'énergie sur les structures trouvées en phase gazeuse permet d'arriver aux mêmes conclusions.

V. Discussion sur la stabilité des complexes à zinc

L'étude précédente des énergies d'activation pour la dissociation d'un ligand thiolate a indiqué que les espèces dianioniques à zinc se dissocient relativement aisément pour mener à des entités monoanioniques plus stables. Cependant, ce phénomène de dissociation semble très dépendant des conditions expérimentales comme le montre également l'impact important du choix de la méthode de solvation sur les énergies correspondantes. Ainsi, il nous semble nécessaire de considérer de façon distincte la stabilité des complexes à zinc selon leur environnement, phase gazeuse ou solvant, et d'en étudier les causes séparément.

A. Stabilité en phase gazeuse

Les énergies de dissociation de $[\text{Zn}(\text{SMe})_4]^{2-}$ et $[\text{Zn}(\text{SPh})_4]^{2-}$ sont respectivement $E_{\text{disso}} = -46.2$ et $E_{\text{disso}} = -30.7$ kcal/mol en phase gazeuse. Nous avons cherché à comprendre les raisons de la plus grande stabilité de $[\text{Zn}(\text{SPh})_4]^{2-}$ relativement à $[\text{Zn}(\text{SMe})_4]^{2-}$. Les calculs effectués en solvation ont montré que l'énergie de dissociation augmente fortement en solvant (Figure IV.15 et IV.16) et donc qu'un ligand thiolate a plus tendance à se détacher du zinc dans le gaz que dans les conditions expérimentales. Cet écart de réactivité nous a conduit à étudier l'effet électronique des ligands puisque les charges ne sont pas solvatées en phase gazeuse et peuvent donc jouer un rôle significatif dans la stabilité des espèces organométalliques.

Pour mesurer l'impact des charges des thiolates sur la dissociation, nous avons étudié, en plus des complexes $[\text{Zn}(\text{SMe})_4]^{2-}$ et $[\text{Zn}(\text{SPh})_4]^{2-}$, le composé $[\text{Zn}(\text{SPhCN})_4]^{2-}$ (Figure IV.21). Ces trois complexes à zinc ont l'avantage de posséder la même charge totale -2, les mêmes charges individuelles pour les ligands, ce qui nous permet d'étudier l'influence du groupe fonctionnel rattaché directement au thiolate. Cette série présente ainsi des ligands dont la charge anionique est plus ou moins délocalisée suivant la nature donneuse ou accepteuse du substituant thiol.

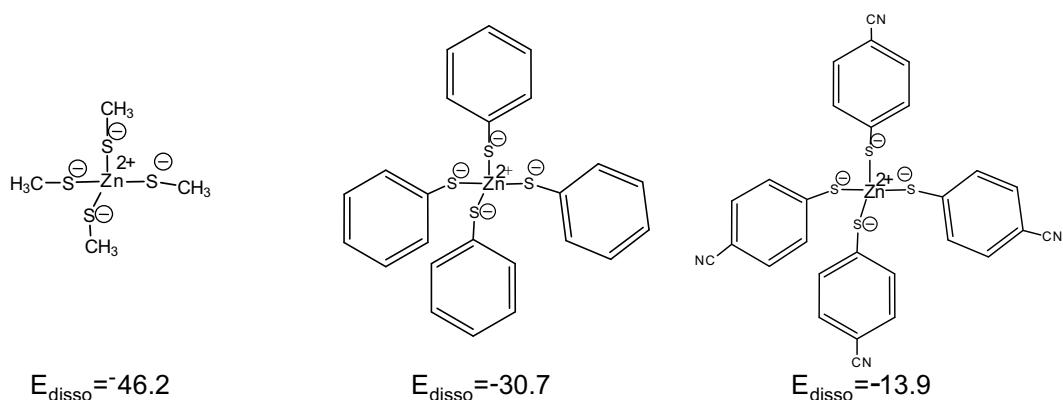


Figure IV.21 : Exemples de complexes dianioniques à zinc avec leurs énergies de dissociation calculées au niveau B3LYP/BS2//B3LYP/BS1 (en kcal/mol)

Avec un substituant méthyle donneur, $[\text{Zn}(\text{SMe})_4]^{2-}$ possède des charges négatives concentrées sur les soufres et donc proches du cation métallique zinc. $[\text{Zn}(\text{SPh})_4]^{2-}$ possède des ligands soufrés dont la charge anionique est répartie par délocalisation sur les groupements phényles. Quant à $[\text{Zn}(\text{SPhCN})_4]^{2-}$, l'effet électroattracteur de ses groupes cyano rend la charge portée par le soufre encore moins importante et plus lointaine du zinc (II). A la vue des énergies de dissociation calculées en phase gazeuse, il existe une corrélation entre la charge anionique portée par le soufre et l'énergie de dissociation du ligand : $[\text{Zn}(\text{SMe})_4]^{2-} > [\text{Zn}(\text{SPh})_4]^{2-} > [\text{Zn}(\text{SPhCN})_4]^{2-}$ (Figure IV.21).

Ainsi, pour $[\text{Zn}(\text{SMe})_4]^{2-}$, une saturation du zinc est probable au voisinage de cette forte concentration de charges négatives et par conséquent, un éloignement facilité pour une charge par la dissociation d'un thiolate ($E_{\text{disso}} = -46.2$ kcal/mol). Le zinc de $[\text{Zn}(\text{SPh})_4]^{2-}$ peut accommoder plus facilement les quatre charges délocalisées de ses ligands que celui de $[\text{Zn}(\text{SMe})_4]^{2-}$ et par conséquent, l'énergie de dissociation est moins importante ($E_{\text{disso}} = -30.7$ kcal/mol). Elle diminue encore pour le troisième complexe avec $E_{\text{disso}} = -13.9$ kcal/mol avec la diffusion importante de la densité électronique négative sur les substituants cyanophényles.

Ce résultat peut aussi être mis en relation avec les conclusions obtenues sur les complexes biomimétiques (cf Chapitre III). Le complexe **3** possède une charge totale de -2 et une énergie de dissociation de son phénylthiolate de $E_{\text{disso}} = -28.4$ kcal/mol (Figure IV.22). En changeant l'environnement électronique du phénylthiolate dans **3**, l'énergie de dissociation est directement impactée. Ainsi, le remplacement de l'atome de bore par un atome neutre de silicium ou de carbone rend l'énergie positive (Figure IV.22). La charge totale du complexe est alors égale à -1, mais surtout la concentration de densité électronique autour du zinc est diminuée et laisse la possibilité d'accommoder une charge anionique supplémentaire. Ainsi, dans **3**, les ligands anioniques sont à une faible distance de 2.141 et 2.322 Å du cation métallique, ce qui a pour effet de saturer le zinc qui « repousse » le dernier soufre à 2.458 Å. Avec un atome neutre au lieu du bore, la charge du ligand pyrazole est moindre, d'où une elongation de la liaison zinc-azote. Le métal peut alors plus facilement supporter la charge du phénylthiolate, ce qui se reflète par son rapprochement au métal (Zn-S=2.316-2.323 Å).

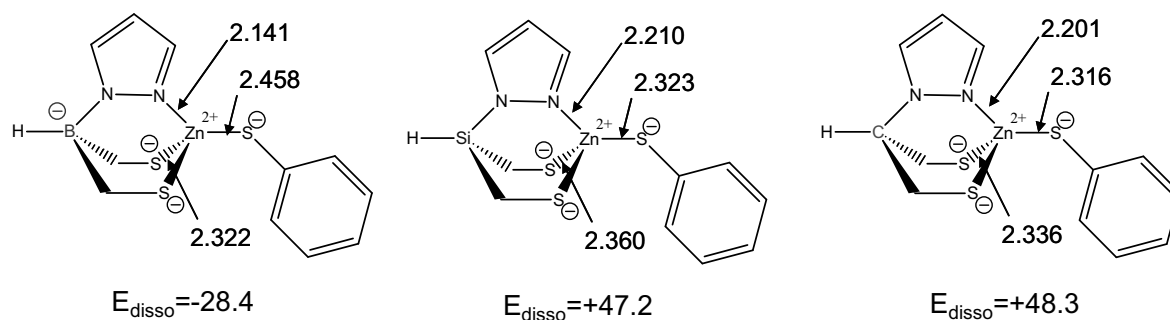


Figure IV.22 : Complexe **3** et dérivés avec leurs énergies de dissociation calculées au niveau B3LYP/BS2//B3LYP/BS1 (en kcal/mol)

Nous avons donc montré que l'énergie de dissociation d'un ligand dépend de l'environnement du zinc, à savoir la concentration de charges anioniques à sa proximité. Cette densité électronique peut se trouver localisée sur les atomes directement liés au zinc ou diffusée sur des substituants plus lointains, ce qui a une forte influence sur la capacité du métal à accepter un ligand supplémentaire.

Dans les protéines, nous pouvons supposer que le comportement chimique du zinc et donc sa coordination seront dépendants de la charge portée par les résidus au voisinage du cation métallique. Etant donné que les groupes le plus souvent coordonnés au zinc sont les acides aminés cystéine et histidine, il nous a paru intéressant d'étudier l'impact du nombre de ligands azotés et soufrés directement liés au zinc sur l'énergie de dissociation.

La famille de composés $[\text{Zn}(\text{SMe})_x(\text{NH}_3)_y]^{2-x}$, x étant le nombre de ligands soufrés et y celui des ligands azotés, a été choisie pour modéliser les sites actifs à zinc possédant dans leur sphère de coordination des cystéines et histidines. Ceci permet de limiter le temps de calcul en réduisant le nombre d'atomes et de pouvoir obtenir une série de complexes à zinc dont les charges totales varient de -2 à +2. Pour chaque composé $[\text{Zn}(\text{SMe})_x(\text{NH}_3)_y]^{2-x}$ ont été déterminées les longueurs moyennes des liaisons zinc-ligand, ainsi que les énergies de dissociation des ligands méthylthiolate et ammoniacque au niveau B3LYP/BS2//B3LYP/BS1 (Figure IV.23).

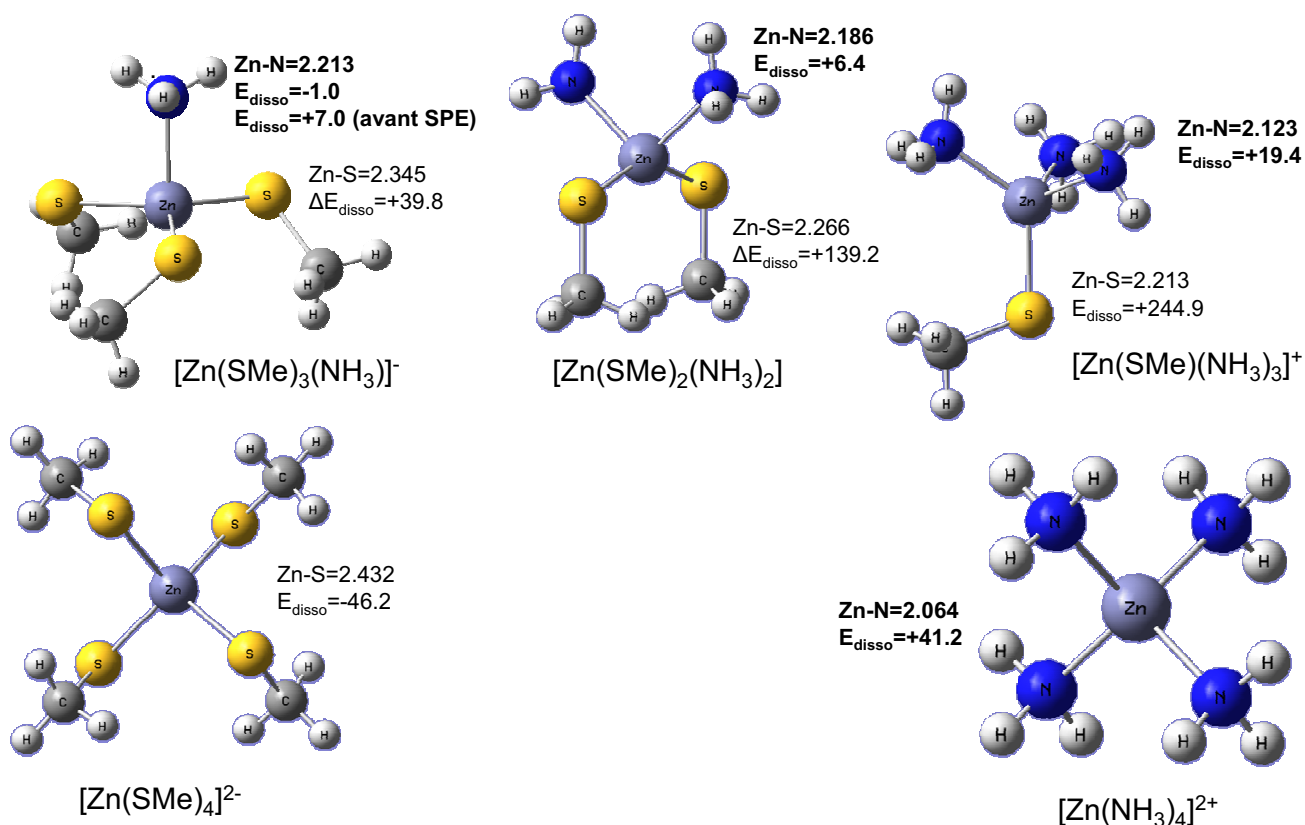


Figure IV.23 : Complexes $[\text{Zn}(\text{SMe})_x(\text{NH}_3)_y]^{2-x}$ avec les énergies de dissociation d'un ligand soufré (normal) et azoté (en gras) calculées au niveau B3LYP/BS2//B3LYP/BS1 (en kcal/mol)

Si l'on considère le cas d'un ligand soufré, on remarque que la distance au métal diminue avec la charge totale du complexe. Ceci est en accord avec l'hypothèse prévoyant une saturation du zinc dans les complexes anioniques se traduisant par une élévation de la liaison zinc-soufre ($\text{Zn-S}=2.432 \text{ \AA}$ dans $[\text{Zn}(\text{SMe})_4]^{2-}$ et $2.345 \text{ \AA}$ dans $[\text{Zn}(\text{SMe})_3(\text{NH}_3)]^-$). Dans le complexe neutre $[\text{Zn}(\text{SMe})_2(\text{NH}_3)_2]$, elle est de $2.266 \text{ \AA}$ pour diminuer encore dans les complexes cationiques $[\text{Zn}(\text{SMe})(\text{NH}_3)_3]^+$ et $[\text{Zn}(\text{NH}_3)_4]^{2+}$ où le zinc peut supporter le rapprochement des soufres chargés négativement ($\text{Zn-S}=2.213$ et $2.064 \text{ \AA}$). Ainsi, la charge totale des complexes, modulée par la nature et le nombre des ligands, peut faire varier la liaison zinc-soufre de $0.219 \text{ \AA}$. De même, en corrélation avec la diminution de cette distance, l'énergie de dissociation du méthylthiolate augmente de $E_{\text{diss}}=-46.2 \text{ kcal/mol}$ pour le complexe dianionique jusqu'à $E_{\text{diss}}=+244.9 \text{ kcal/mol}$ pour le composé dicationique. L'effet des charges proches du zinc a donc un impact remarquable sur l'énergie de dissociation du thiolate, puisqu'à chaque ligand anionique en moins, celle-ci augmente de près de $+100 \text{ kcal/mol}$.

Le phénomène est semblable pour les ligands azotés : la distance de l'amine au métal diminue avec la concentration de charge négative autour du zinc, passant de $\text{Zn-N}=2.213 \text{ \AA}$ dans $[\text{Zn}(\text{SMe})_3(\text{NH}_3)]^-$ à $\text{Zn-N}=2.064 \text{ \AA}$ dans $[\text{Zn}(\text{NH}_3)_4]^{2+}$. Cependant, la dissociation n'est pas aisée pour un ligand azoté du fait de sa compatibilité électronique avec le zinc. Seulement dans le cas où le zinc est déjà un peu saturé par les trois ligands anioniques dans $[\text{Zn}(\text{SMe})_3(\text{NH}_3)]^-$, l'énergie de dissociation est égale à $E_{\text{diss}}=-0.1 \text{ kcal/mol}$. Elle devient ensuite positive dans les complexes neutres ($E_{\text{diss}}=+6.4 \text{ kcal/mol}$) et les complexes cationiques ($E_{\text{diss}}=+19.4 \text{ kcal/mol}$ pour $[\text{Zn}(\text{SMe})(\text{NH}_3)_3]^+$ et $+41.2 \text{ kcal/mol}$ pour $[\text{Zn}(\text{NH}_3)_4]^{2+}$). Ce ligand azoté aura tendance à stabiliser les complexes à zinc du fait de sa neutralité et de sa paire d'électrons libres qui compensent l'électrodéficiences de l'ion zinc (II). Alors que l'ajout d'un ligand soufré faisait perdre environ 100 kcal/mol à la stabilité du complexe tétra-coordonné, chaque ligand amine supplémentaire apporte environ 10 kcal/mol d'énergie de stabilisation.

B. Stabilité en phase aqueuse

Alors que les complexes à quatre ligands thiolates, comme $[\text{Zn}(\text{SPh})_4]^{2-}$, se dissocient facilement en phase gazeuse pour donner des espèces tri-coordonnées plus stables, les calculs effectués avec la méthode COSMO ont montré que leurs dissociations étaient beaucoup moins aisées en solvant.

Par exemple, l'énergie de dissociation du complexe $[\text{Zn}(\text{SPh})_4]^{2-}$ devient positive en phase aqueuse ($E_{\text{diss}}=+8.8 \text{ kcal/mol}$, partie IV.B.3) et indique alors une légère préférence pour la stabilisation de l'espèce à quatre ligands plutôt que pour l'espèce monoanionique $[\text{Zn}(\text{SPh})_3]^-$. Ceci explique l'observation expérimentale du composé $[\text{Zn}(\text{SPh})_4]^{2-}$ et l'absence de structure cristallographique du composé tri-coordonné correspondant.

Au vu de l'étude de la stabilité des complexes à zinc en phase gazeuse, il est aisé de supposer que la solvation des charges anioniques des ligands permet de limiter le phénomène de

saturation autour du zinc. Cependant, la stabilité des espèces à quatre ligands reste faible, notamment pour $[\text{Zn}(\text{SPh})_4]^{2-}$. Par ailleurs, un complexe à zinc possédant uniquement trois thiolates et que nous avons noté $[\text{Zn}(\text{SPh}^{\text{Me4}})_4]^{2-}$, a été isolé expérimentalement (Gruff 1989) (cf figure IV.6). La nature encombrée de ses ligands suppose que leur effet électronique, n'ayant que peu d'influence sur la dissociation en phase aqueuse, pourrait être contrebalancé par leur effet stérique. Ainsi, la nature des substituants des thiolates pourrait avoir un impact important sur la stabilité des complexes à zinc en phase aqueuse. Pour vérifier cette hypothèse, nous avons déterminé l'énergie de dissociation du complexe $[\text{Zn}(\text{SPh}^{\text{Me4}})_4]^{2-}$ (Figure IV.24).

L'énergie de dissociation de $[\text{Zn}(\text{SPh}^{\text{Me4}})_4]^{2-}$ est égale à $E_{\text{diss}} = -37.4$ kcal/mol en phase gazeuse (Figure IV.24), ce qui équivaut à environ 7 kcal/mol de moins que l'énergie déterminée précédemment pour $[\text{Zn}(\text{SPh})_4]^{2-}$ ($E_{\text{diss}} = -30.3$ kcal/mol). Ceci indique que $[\text{Zn}(\text{SPh}^{\text{Me4}})_4]^{2-}$ est relativement plus stable que $[\text{Zn}(\text{SPh})_4]^{2-}$. En solvant, cet écart de 7 kcal/mol se retrouve pour les énergies de dissociation entre les deux composés dianioniques : elle est déterminée à -4.8 kcal/mol et -0.01 kcal/mol, respectivement aux niveaux COSMO-B3LYP/GB//MARI-J-BP86/PB et COSMO-B3LYP/GB//COSMO-MARI-J-BP86/PB, alors que les valeurs pour $[\text{Zn}(\text{SPh})_4]^{2-}$ étaient $E_{\text{diss}} = +8.8$ et $+7.9$ kcal/mol (Figure IV.16).

Nous pouvons donc en conclure que l'effet stérique des substituants abaisse considérablement l'énergie de dissociation des complexes à zinc. Ainsi, si $[\text{Zn}(\text{SPh})_4]^{2-}$ possédait une stabilité assez faible, $[\text{Zn}(\text{SPh}^{\text{Me4}})_4]^{2-}$ est encore moins aisément isolable, car en équilibre avec $[\text{Zn}(\text{SPh}^{\text{Me4}})_3]^-$. Ceci est confirmée par l'obtention de la structure cristallographique de $[\text{Zn}(\text{SPh}^{\text{Me4}})_3]^-$, dont l'encombrement des ligands favorise une tri-coordination au zinc.

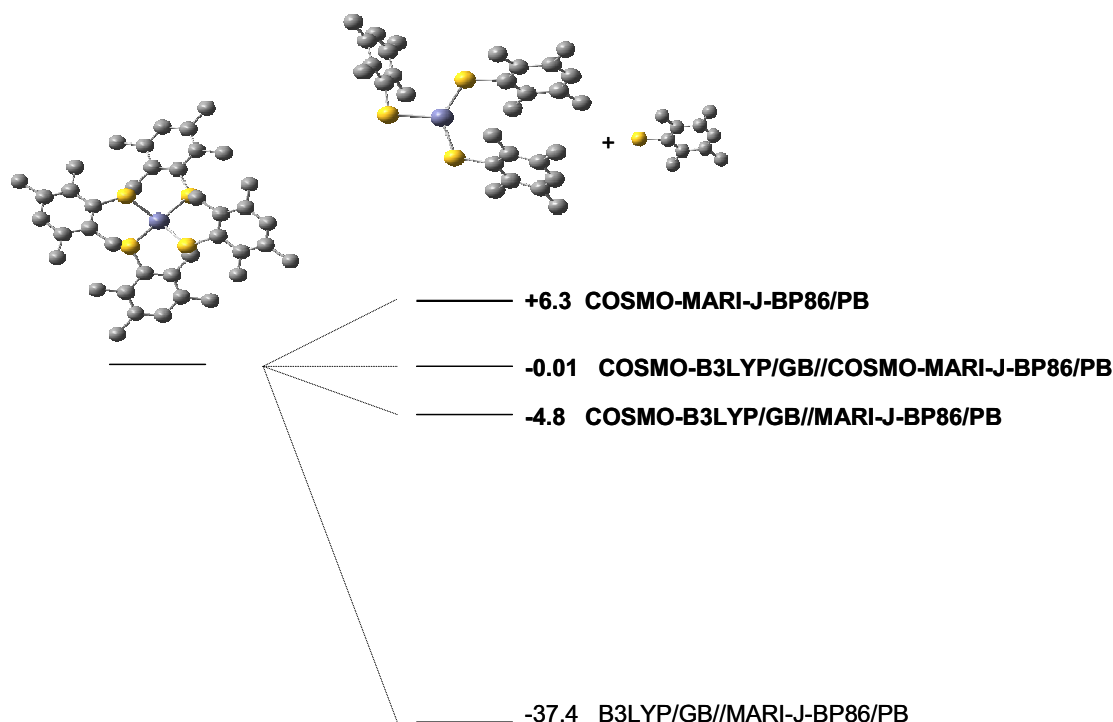


Figure IV.24 : Energies de dissociation du complexe $[\text{Zn}(\text{SPh}^{\text{Me4}})_4]^{2-}$ en phase gazeuse (normal) et phase aqueuse (gras) en kcal/mol

C. Conséquences pour la stabilité des sites actifs ZnCys₄ des protéines

Nous avons établi que la stabilité des complexes de type $[\text{Zn}(\text{SR})_4]^{2-}$ en phase aqueuse est dépendante du substituant R des ligands thiolates. L'influence des charges négatives est importante en phase gazeuse et nuancée en solvant, puisque le milieu permet de mieux diffuser les charges. Cet effet électronique est à contrebalancer avec l'encombrement stérique qui peut prohiber l'approche d'un quatrième thiolate.

Si l'on considère $[\text{Zn}(\text{SMe})_4]^{2-}$ comme un modèle simplifié des sites protéiques ZnCys₄, nous pouvons avoir une idée plus précise de la stabilité des sites actifs à zinc. En effet, le substituant méthyle permet un arrangement spatial tétraédrique autour du zinc. Cependant, l'absence de groupe encombrant est compensée par une concentration de charge anionique importante autour du zinc, favorisant la dissociation d'un quatrième ligand. La stabilité du site dianionique sera aussi dépendante de l'environnement : si le site est situé à la surface de la protéine, comme pour les petits doigts de zinc, les charges seront mieux solvatées car accessibles par le solvant. Ainsi, la stabilité de ZnCys₄ sera faible ($E_{\text{diss}} = -1.5$ kcal/mol en phase aqueuse), mais pourra probablement être compensée par l'effet stabilisateur de liaisons hydrogènes ou par la rigidité du squelette protéique. Il a en effet été montré que les sites dianioniques ZnCys₄ possèdent un réseau amplifié de liaisons hydrogènes reliant les thiolates et les résidus du squelette (Maynard 2001).

Par contre, si le site à zinc est enfoui profondément au cœur de la protéine, on se rapprochera de conditions en phase gazeuse où les charges, concentrées sur les thiolates, empêcheront le métal de coordonner quatre cystéinates. Il se peut alors que la protéine ne puisse exister dans un état stable avec un site à zinc tétravalent et qu'elle préfère alors une conformation à deux cations métalliques qui permettront d'absorber l'excès de charge négative. Cette hypothèse permettrait de rationaliser l'existence de nombreuses enzymes contenant plusieurs atomes de zinc, comme la vaste famille des métallothionéines (Figure IV.25).

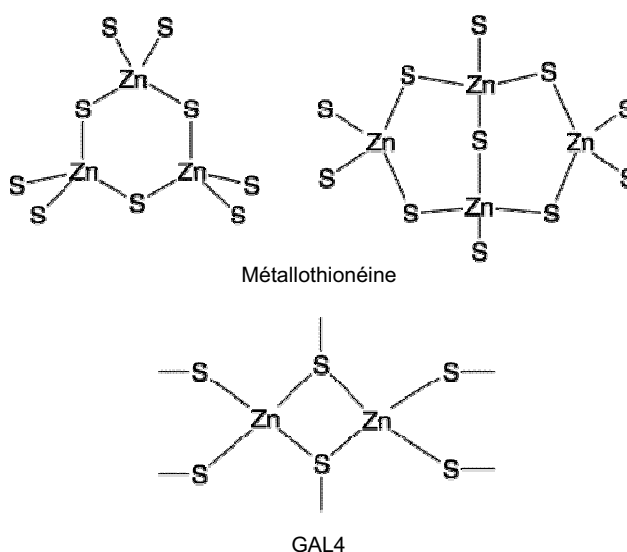


Figure IV.25 : Sites à zinc polynucléaires (Parkin 2004)

Ces protéines possèdent des clusters à zinc et cadmium, chacun contenant deux à quatre ions métalliques zinc. La particularité de ces clusters est leur forte proportion de métal et soufre (10% en poids chacun) et leur constant arrangement tétraédrique des ligands thiolates autour du zinc (Kägi 1960, Romero-Isart 2002, Henkel 2004). Ces résidus cystéines peuvent être aussi bien reliés à la chaîne peptidique qu'à un ou plusieurs atomes de zinc. Le rôle des métallothionéines concerne le stockage du zinc et l'élimination des radicaux libres. Le facteur de transcription GAL4 (Figure IV.25) est un analogue de ces métallothionéines. Cette structure polynucléaire pour les sites protéiques à zinc est probablement une alternative pour pallier le manque de stabilité causé par la concentration importante de thiolates autour du zinc.

Bibliographie

- Abrahams, I. L.; Garner, C. D.; Clegg, W. *J. Chem. Soc. Dalton Trans.* **1987**, 1577.
- Alberts, I. L.; Nadassy, K.; Wodak, S. *J. Prot. Sci.* **1998**, 7, 1700.
- Alsina, T.; Clegg, W.; Fraser, K. A.; Sola, J. *J. Chem. Soc. Dalton Trans.* **1992**, 1393.
- Bochmann, M.; Webb, K.; Harman, M.; Hursthouse, M. B. *Angew. Chem. Int. Ed. Engl.* **1990**, 29, 638.
- Bochmann, M.; Bwembya, G. C.; Grinter, R.; Powell, A. K.; Webb, K. J.; Hursthouse, M. B.; Bock, C. W.; Katz, A. K.; Glusker, J. P. *J. Am. Chem. Soc.* **1995**, 117, 3754.
- Bowmaker, G. A.; Dance, I. G.; Dobson, B. C.; Rogers, D. A. *Aust. J. Chem.* **1984**, 37, 1607.
- Bowmaker, G. A.; Dance, I. G.; Harris, R. K.; Henderson, W.; Laban, I.; Scudder, M. L.; Oh, S. W. *J. Chem. Soc. Dalton Trans.* **1996**, 2381.
- Calhorda, M. J.; Pregosin, P. S.; Veiros, L. F. *J. Chem. Theory Comput.* **2007**, 3, 665.
- Christou, G.; Folting, K.; Huffman, J. C. *Polyhedron* **1984**, 3, 1247. Dance, I. G. *J. Am. Chem. Soc.* **1980**, 102, 3445.
- Chung, W. P.; Dewan, J. C.; Walters, M. A. *Inorg. Chem.* **1991**, 30, 4280.
- Coucovanis, D.; Murphy, C. N.; Kanodia, S. K. *Inorg. Chem.* **1980**, 19, 2993.
- Coucovanis, D.; Swenson, D.; Baenziger, N. C.; Murphy, C.; Holah, D. G.; Sfarnas, N.; Simopoulos, A.; Kostikas, A. *J. Am. Chem. Soc.* **1981**, 103, 3350.
- Dance, I. G.; Choy, A.; Scudder, M. L. *J. Am. Chem. Soc.* **1984**, 106, 6285.
- Darensbourg, D. J.; Niezgod, S. A.; Draper, J. D.; Reibenspies, J. H. *Inorg. Chem.* **1999**, 38, 1356.
- Dimakis, N.; Farooqi, M. J.; Garza, E. S.; Bunker, G. *J. Chem. Phys.* **2008**, 128, 11, 115104.
- Dudev, T.; Lim, C. *J. Phys. Chem. B* **2001**, 105, 43, 10709.
- Dudev, T.; Lim, C. *Chem. Rev.* **2003**, 103, 773.
- Fujisawa, K.; Imai, S.; Suzuki, S.; Moro-Oka, Y.; Miyashita, Y.; Yamada, Y.; Okamoto, K. *J. Inorg. Biochem.* **2000**, 82, 229.
- Garner, C. D.; Nicholson, J. R.; Clegg, W. *Inorg. Chem.* **1984**, 23, 2148.
- Gorelsky, S. I.; Basumallick, L.; Vura-Weis, J.; Sarangi, R.; Hodgson, K. O.; Hedman, B.; Fujisawa, K.; Solomon, E. I. *Inorg. Chem.* **2005**, 44, 4947.
- Gruff, E. S.; Koch, S. A. *J. Am. Chem. Soc.* **1989**, 111, 8762.
- Gruff, E. S.; Koch, S. A. *J. Am. Chem. Soc.* **1990**, 112, 1245.
- Grützmacher, H.; Steiner, M.; Pritzkow, H.; Zsolnai, L.; Huttnner, G.; Sebald, A. *Chem. Ber.* **1992**, 125, 2199.
- Henkel, G.; Krebs, B. *Chem. Rev.* **2004**, 104, 801.
- Holah, D. G.; Coucovanis, D. *J. Am. Chem. Soc.* **1975**, 97, 23, 6917.
- Huang, Z.; Hu, H.; Gu, W.; Wu, G. *J. Inorg. Biochem.* **1994**, 54, 147.
- Ibrahim, M. M.; Seebacher, J.; Steinfeld, G.; Vahrenkamp, H. *Inorg. Chem.* **2005**, 44, 8531.
- Kägi, J. H. R.; Vallee, B. L. *J. Biol. Chem.* **1960**, 235, 3460.
- Matsunaga, Y.; Fujisawa, K.; Ibi, N.; Amir, N.; Miyashita, Y.; Okamoto, K. *Bull. Chem. Soc. Jpn.* **2005**, 78, 1285.
- Maynard, A. T.; Covell, D. G. *J. Am. Chem. Soc.* **2001**, 123, 1047.

- Okamura, T.; Ueyama, N.; Nakamura, A.; Ainscough, E. W.; Brodie, A. M.; Waters, J. M. *J. Chem. Soc. Chem. Commun.* **1993**, 1658.
- Parkin, G. *Chem. Rev.* **2004**, 104, 699.
- Patel, K.; Kumar, A.; Durani, S. *Biochim. Biophys. Acta* **2007**, 1774, 1247.
- Power, P.P.; Ellison, J. J. *Inorg. Chem.* **1994**, 33, 4231.
- Power, P.P.; Nguyen, T.; Panda, A.; Olmstead, M. M.; Richards, A.F.; Stender, M.; Brynda, M. J. *Am. Chem. Soc.* **2005**, 127, 8545.
- Prust, J.; Hohmeister, H.; Stasch, A.; Roesky, H. W.; Magull, J.; Alexopoulos, E.; Uson, I.; Schmidt, H. G.; Noltemeyer, M. *Eur. J. Inorg. Chem.* **2002**, 2156.
- Rao, C. P.; Dorfman, J. R.; Holm, R. H. *Inorg. Chem.* **1986**, 25, 428.
- Raper, E. S.; Creighton, J. R.; Robson, D.; Wilson, J. D.; Clegg, W.; Milne, A. *Inorg. Chim. Acta* **1988**, 143, 95.
- Rayon, V. M.; Valdes, H.; Diaz, N.; Suarez, D. *J. Chem. Theory Comput.* **2008**, 4, 243.
- Romero-Isart, N.; Vasak, M. *J. Inorg. Biochem.* **2002**, 88, 388.
- Santos, R. A.; Gruff, E. S.; Koch, S. A.; Harbison, G. S. *J. Am. Chem. Soc.* **1991**, 113, 469.
- Silver, A.; Koch, S. A.; Millar, M. *Inorg. Chim. Acta.* **1993**, 205, 9.
- Simonson, T.; Calimet, N. *Proteins, Struct. Funct. Bioinf.* **2002**, 49, 37.
- Sola, J.; Gonzalez-Duarte, P.; Sanz, J.; Casals, I.; Alsina, T. Sobrados, I.; Alvarez-Larena, A.; Piniella, J. F.; Solans, X. *J. Am. Chem. Soc.* **1993**, 115, 10018.
- Solomon, E. I.; Gorelsky, S. I.; Dey, A. *J. Comput. Chem.* **2006**, 27, 1415.
- Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J Phys. Chem. B*, **2007**, 30, 9146.
- Swenson, D.; Baenziger, N. C.; Coucouvanis, D. *J. Am. Chem. Soc.* **1978**, 100, 1932.
- Tamames, B.; Sousa, S. F.; Tamames, J.; Fernandes, P. A.; Ramos, M. J. *Proteins, Struct. Funct. Bioinf.* **2007**, 69, 466.
- Tang, K.; Aslam, M.; Block, E.; Nicholson, T. *J. Zubieta Inorg. Chem.* **1987**, 26, 1488.
- Ueyama, N.; Sugawara, T.; Sasaki, K.; Nakamura, A.; Yamashita, S.; Wakatsuki, Y.; Yamazaki, H.; Vahrenkamp, H. *Dalton Trans.* **2007**, 4751.
- Vallee, B. L.; Auld, D. S. *Acc. Chem. Res.* **1993**, 26, 543.
- Vittal, J. J.; Dean, P. A. W. *Inorg. Chem.* **1996**, 35, 3089.
- Watson, A. D.; Rao, C. P.; Dorfman, J. R.; Holm, R. H. *Inorg. Chem.* **1985**, 24, 2820.
- Watton, S. P.; Wright, J. G.; MacDonnell, F. M.; Bryson, J. W.; Sabat, M.; O'Halloran, T. V. *J. Am. Chem. Soc.* **1990**, 112, 2824.
- Wilker, J. J.; Lippard, S. J. *J. Am. Chem. Soc.* **1995**, 117, 33, 8682.
- Wilker, J. J.; Lippard, S. J. *Inorg. Chem.* **1997**, 36, 969.

Chapitre V

Indices de nucléophilie du motif zinc-thiolate

La compréhension des mécanismes enzymatiques et de la réactivité des sites protéiques à zinc passe par le développement d'un moyen de mesure de la nucléophilie des thiolates liés au métal. En effet, il est montré que l'alkylation d'un thiolate est parfois sélective envers un unique ligand, supposé plus nucléophile du fait de son environnement. Par exemple pour la protéine Ada, la cystéine 38 est l'unique soufre ne participant pas à des liaisons hydrogènes, et par conséquent, ce thiolate, plus réactif, est exclusivement alkylé (He 2005, Takinowaki 2006).

La connaissance du site d'alkylation et du thiolate impliqué dans une réaction enzymatique nécessite donc un critère de réactivité fiable pour les thiolates. Dans le chapitre III, nous avons utilisé les énergies d'alkylation calculées avec le logiciel GAUSSIAN pour évaluer la nucléophilie des arylthiolates de complexes biomimétiques à zinc. La taille des systèmes protéiques rend impossible la recherche d'un état de transition et donc du mécanisme réactionnel. Pour les sites actifs à zinc, un critère simple et rapide permettant de déterminer le thiolate le plus nucléophile est nécessaire.

Cette partie est consacrée à la nucléophilie du thiolate et à la recherche d'un indice permettant de mesurer cette réactivité. Il s'agit de considérer la réactivité intrinsèque du composé soufré et de tirer uniquement cette information de la composition du réactif. L'idée principale est d'obtenir un facteur applicable à tous les thiolates liés au zinc, indépendant de la réactivité des partenaires chimiques et qui permettrait alors d'éviter l'étude parfois laborieuse des chemins réactionnels.

Nous utiliserons les résultats obtenus sur les complexes biomimétiques (Chapitre III) pour lesquels la nucléophilie a déjà été mesurée grâce à l'énergie d'activation requise en phase gazeuse pour effectuer une alkylation par l'iodure de méthyle via un mécanisme S_N2 . Cette barrière énergétique nous servira de référence pour effectuer des calculs préliminaires de corrélation avec d'autres indices issus de la littérature. Le critère de réactivité choisi sera ensuite confirmé sur une large famille de thiolates liés au zinc dans des complexes biomimétiques variés.

I. Introduction

Afin d'établir une corrélation entre la barrière énergétique pour l'alkylation d'un thiolate et un indice de nucléophilie intrinsèque au réactif, il est nécessaire de repérer tout d'abord les indices de réactivité déjà existants et d'en cerner les applications et les limites.

De nombreuses tentatives de prédictions de la réactivité d'une structure existent dans la littérature. Cependant, les familles de groupes chimiques ont fréquemment été traitées ensemble, ce qui rend l'objectif très ambitieux. En effet, si l'on considère la variété des composés chimiques, présentant des caractères structuraux et des liaisons différentes, il est difficile de concevoir un critère de réactivité universel. C'est pourquoi il n'existe encore actuellement aucun indice de nucléophilie absolu et reconnu par la totalité de la communauté chimiste.

L'étude bibliographique suivante retrace les constants efforts des chimistes pour définir la réactivité, nucléophilie ou électrophilie, et la mesurer quantitativement. Après avoir présenté les

échelles de réactivité qui se basent principalement sur des observations expérimentales pour prédire la réactivité d'une molécule (Cf I.B), nous considérerons les indices développés à partir de calculs théoriques, tels que les charges ou encore les coefficients de Fukui (Cf I. C et D).

A. Nucléophilie et basicité

Les termes empiriques de « nucléophile » et « électrophile » sont apparus dans les années trente et désignent les espèces chimiques respectivement riches, et pauvres en électrons (Ingold 1929). Découlant directement des concepts développés par Lewis (Lewis 1923) et Lowry et Bronsted (Bronsted 1923, Lowry 1923), la définition suivante est communément acceptée : un nucléophile est une base de Lewis qui utilise une paire d'électrons disponible pour former une liaison avec son partenaire réactionnel (un électrophile). Les facteurs influençant le plus sa réactivité sont son énergie de solvation, la force de la liaison formée, son électronégativité et sa polarisabilité (Carey 1990).

Il apparaît alors clairement que les concepts empiriques de basicité et nucléophilie sont reliés. Cependant, alors que le terme nucléophilie est généralement employé pour décrire des tendances dans les cinétiques de réaction et à quelle vitesse un nucléophile peut former une nouvelle liaison, la basicité, quand à elle, est reliée à la position de l'équilibre de la réaction avec un proton ou un autre acide (Carey 1990). La nucléophilie est une grandeur cinétique, alors que la basicité est un concept thermodynamique. La différence est clairement établie du moment que l'on utilise des données cinétiques (vitesses de réactions) pour mesurer la nucléophilie et des données thermodynamiques (affinité protonique ou énergie libre de Gibbs) pour quantifier la basicité.

Cependant, la distinction entre les deux est toujours sujette à débat (Jaramillo 2007) et les échelles expérimentales développées par les premiers chercheurs utilisent aussi bien des propriétés thermodynamiques que des vitesses de réactions.

B. Les échelles de réactivité

Les échelles de réactivité sont les seules tentatives essayant de relier la nucléophilie à une observation expérimentale, celle-ci étant généralement une constante de vitesse.

Le premier essai d'évaluation de la réactivité a donné l'équation de Hammett, qui inclut la constante du même nom. Sa technique appelée LFER (*Linear Free Energy Relationship*) a alors pour but d'évaluer l'effet des substituants et des groupes partants dans une réaction d'hydrolyse d'ester (Hammett 1937). D'autres relations similaires incluant des paramètres cinétiques ont été proposées dans cette lignée et ont constitué un travail de pionnier dans le domaine des relations structure-activité (Johnson 1973, Hine 1981, Taft 1987, Waisser 2000).

Le concept de « nucléophilie » est ensuite généralisé à des réactifs et non plus uniquement à des substituants (Swain 1953). Les vitesses relatives des réactions de substitution de divers agents nucléophiles avec des halogénures de méthyle ont été mesurées en milieu aqueux. L'équation

suivante permet de corréler les constantes de réaction avec la nucléophilie n , constante intrinsèque du réactif nucléophile et le paramètre s spécifique à chaque électrophile:

$$\log(k / k_{H_2O}) = s \times n$$

k : Constante de réaction S_N2 d'un électrophile avec un nucléophile donné

k_{H_2O} : Constante de réaction avec un nucléophile de référence (l'eau)

Un modèle basé sur la réaction de carbocations et ions diazoniums avec divers nucléophiles est ensuite proposé avec un indice de réactivité voulant s'affranchir de la dépendance de la nature de l'électrophile (Ritchie 1972). Un système nucléophile particulier peut être alors caractérisé par un paramètre N_+ suivant l'équation suivante:

$$\log(k / k_{H_2O}) = N_+$$

Cependant, les corrélations semblent bien meilleures quand différentes familles d'électrophiles sont traitées séparément (Ritchie 1986).

Un autre développement d'indices de réactivité qui seraient indépendants des partenaires de réaction est réalisé (Mayr 1994). Il est montré que les vitesses mesurées pour les réactions de carbocations avec des nucléophiles neutres obéissent à la relation suivante :

$$\log k = s_N (E + N)$$

où E et N sont les paramètres d'électrophilie et de nucléophilie respectivement

s_N : paramètre supplémentaire spécifique au réactif nucléophile

Les effets de solvant sont faibles et également négligés (Mayr 1990). La force des électrophiles est déterminée par une analyse des réactions entre ces composés et des nucléophiles π de référence (Mayr 2001). Sur cette base, les auteurs ont construit des échelles relatives expérimentales d'électrophilie et de nucléophilie pour une large diversité de composés organiques et organométalliques (Mayr 2003, Kempf 2003, Bug 2003, Hofmann 2004, Brotzel 2007 et Tumanov 2007).

Tous ces modèles expérimentaux montrent la difficulté d'obtenir une échelle absolue ou un indicateur universel de la nucléophilie. Cependant, une équation globale regroupant toutes les études précédentes a été établie récemment (Schéma V.1) (Phan 2006).

Les paramètres nucléophiles et électrophiles étant obtenues à partir de réactions impliquant des carbones nucléophiles ou électrophiles, cette équation globale est applicable uniquement à des réactions où au moins un des sites réactionnels est un atome de carbone. Même sous cette condition, des déviations peuvent être attendues, si des interactions stabilisantes ou déstabilisantes avec des substituants géminaux sont créées ou détruites dans l'état de transition. Par exemple, les ions carboxoniums réagissent plus rapidement avec les alcools que les équations globales et de Patz le prédisent, du fait de la stabilisation du produit par effet anomère (Juaristi 1992).

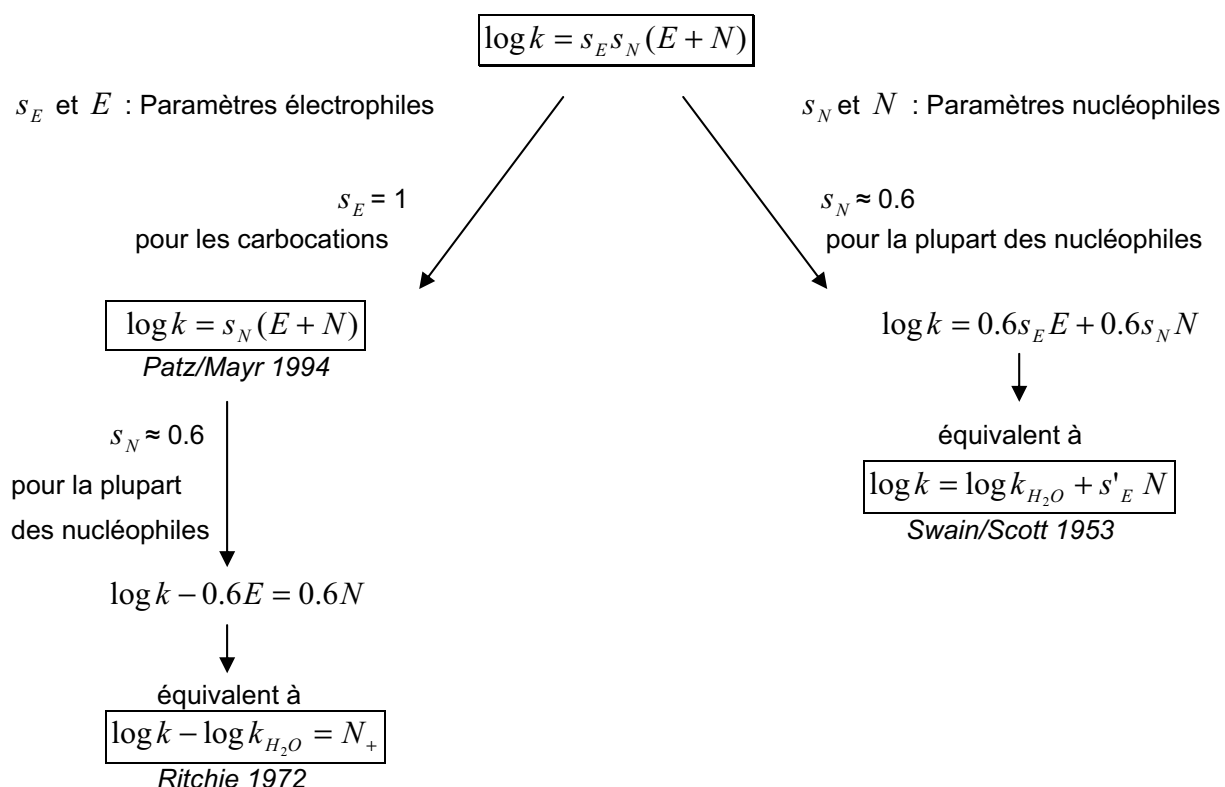


Schéma V.1 : Equation globale pour la relation nucléophilie/électrophilie (Phan 2006)

Alors que les essais basés sur des mesures cinétiques restent infructueux du fait de la dépendance aux conditions réactionnelles (partenaire chimique), d'autres indices sont développés qui reposent sur des facteurs expérimentaux moins conventionnels.

Une équation à quatre paramètres, plus générale, relie cette fois les constantes de réaction aux différentes propriétés du nucléophile comme la basicité, la polarisabilité et le potentiel d'oxydation (Edwards 1954 et 1956). Bien qu'il permette de corréler de nombreuses données expérimentales, des problèmes apparaissent quand le système possède plusieurs atomes donneurs, une paire d'électrons libres ou un facteur stérique important (Pearson 1968). Le principe des bases dures et molles selon la théorie HSAB (Klopman 1968) est alors préféré pour expliquer les interactions électrophiles-nucléophiles (Pearson 1968).

Une autre échelle relative de la nucléophilie est le modèle, moins conventionnel, basé sur le caractère électrostatique de la liaison hydrogène formé entre une base de Lewis B et un électrophile HX dans plusieurs dimères B-HX (Legon 1987).

Il faut attendre le développement des méthodes de chimie quantique pour s'affranchir complètement des observations expérimentales (partenaires de réaction, conditions expérimentales...) qui compliquent la quantification de la nucléophilie. La constante de vitesse expérimentale n'est plus utilisée comme référence et on voit alors apparaître les premiers indices de réactivité purement théoriques.

C. Les charges

Les progrès effectués dans les algorithmes de calculs et l'essor associé de la chimie quantique ont vu apparaître de nombreuses méthodes pour calculer les charges d'une molécule. Obtenue de façon directe grâce aux logiciels tels que GAUSSIAN et TURBOMOLE, la charge d'un atome reste un critère simple pour évaluer rapidement sa force nucléophile. Diverses analyses de la population électronique de la molécule peuvent être choisies, permettant ainsi une répartition des charges différentes suivant la méthode sélectionnée. Parmi les plus populaires se trouvent les charges issues d'une analyse de population de Mulliken, de Hirshfeld ou d'une population NPA.

1. Charges de Mulliken

L'analyse de population de Mulliken (MPA) est le schéma de population le plus utilisé et le plus trivial (Mulliken 1955). Il permet de partitionner la densité électronique de façon arbitraire sur les atomes, en considérant que toute la densité électronique $P_{\mu\mu}$ d'une orbitale ϕ_μ est assignée à l'atome sur laquelle elle est localisée. Le reste de la densité, associée à la population de recouvrement $P_{\mu\nu}$, est également partagée entre les deux atomes sur lesquelles les orbitales ϕ_μ et ϕ_ν sont situées. La charge nette de l'atome A est alors calculée en soustrayant le nombre d'électrons de la charge nucléaire Z_A :

$$q_A = Z_A - \sum_{\mu=1; \mu \text{ sur } A}^K P_{\mu\mu} - \sum_{\mu=1; \mu \text{ sur } A}^K \sum_{\nu=1; \nu \neq \mu}^K P_{\mu\nu} S_{\mu\nu}$$

Cependant, ce partitionnement arbitraire peut causer des erreurs. Par exemple, le partage égal des électrons entre une paire d'atomes d'électronégativités très différentes peut conduire à des valeurs improbables de charge sur les atomes. Cette analyse de Mulliken est également fortement dépendante de l'effet de base (Leach 2001).

2. Charges NBO

Une alternative à la traditionnelle population de Mulliken est le calcul des charges suivant la théorie NPA (*Natural Population Analysis*). Cette analyse de population commence avec la génération des « Natural Bond Orbitals » ou NBO, correspondant à la structure de Lewis la plus proche du composé étudié. Ces orbitales sont des orbitales localisées de paires d'électrons correspondant à des paires liantes ou libres d'électrons. L'hybridation des atomes et le poids de chaque atome dans chaque liaison localisée sont calculés selon cette structure de Lewis idéale. Pour chaque orbitale est donnée son type (liante, non-liante ou anti-liante) et son occupation (entre 0 and 2.0000 électrons).

Cette méthode permet une meilleure distribution électronique que celle de Mulliken, notamment pour les composés à caractère ionique élevé ou contenant des atomes métalliques (Foster 1980, Rives 1980, Reed 1983 et 1985).

3. Charges de Hirshfeld

L'analyse des charges suivant la méthode de Hirshfeld (Hirshfeld 1977) a aussi été beaucoup employée, notamment pour le calcul des coefficients de Fukui.

Pour obtenir les charges de Hirshfeld d'une molécule, celle-ci est tout d'abord décomposée en fragments atomiques. La méthode de choix, naturelle et générale, est de partager la densité électronique localement entre les différents atomes et ce, proportionnellement à la densité de chaque atome individuel et à une certaine distance du noyau. Cette méthode permet d'obtenir des distributions électroniques localisées et liant les atomes, qui de plus, ressemblent approximativement à la densité électronique moléculaire. L'intégration des densités entourant chaque atome définit sa charge atomique nette et les moments multipolaires de la molécule.

D. Les indices de réactivité issus des méthodes DFT

Un nombre important de critères est ensuite apparu dans la littérature avec l'émergence de la DFT conceptuelle (Geerlings 2003). A la différence des paramètres développés précédemment avec les échelles de réactivité et à l'instar des charges, ces indices sont issus et dépendent uniquement de la structure électronique des molécules. Ils ont donc l'avantage d'être totalement dissociés des partenaires de réactions.

1. Coefficient de Fukui

a. Définition

Le coefficient de Fukui a été introduit par Parr and Yang dès 1984 (Parr 1984) et permet de généraliser la théorie des orbitales frontières (Fukui 1952 et 1954) grâce à la fonctionnelle de la densité. Ainsi, si l'on considère une molécule à N électrons, la fonctionnelle de Fukui $f(\vec{r})$ peut être interprétée soit comme la variation de densité électronique lorsque le nombre total d'électrons N est modifié, soit comme la réponse du potentiel chimique μ du système à une perturbation du potentiel externe V et ceci, à un point particulier r .

$$f(\vec{r}) = [\delta\mu / \delta V(\vec{r})]_N = [\delta\rho(\vec{r}) / \delta N]_{V(\vec{r})}$$

La fonction de Fukui est donc une quantité locale et une forte valeur reflète une importante réactivité du site chimique considéré. Pour chaque nombre total d'électrons N , l'énergie E de la molécule et sa densité électronique ρ montrent des discontinuités (Perdew 1982). De même, il a été déduit trois formes différentes de la fonction de Fukui, selon qu'elle gouverne :

- une attaque électrophile : $f^-(\vec{r}) = [\delta\rho(\vec{r}) / \delta N]_v^-$

- une attaque nucléophile : $f^+(\vec{r}) = [\delta\rho(\vec{r}) / \delta N]_v^+$

- une attaque radicalaire : $f^0(\vec{r}) = [\delta\rho(\vec{r}) / \delta N]_v^0$

ce qui donne, après application de la théorie des orbitales frontières, respectivement :

$$f^-(\vec{r}) \approx \rho_{HOMO}(\vec{r})$$

$$f^+(\vec{r}) \approx \rho_{LUMO}(\vec{r})$$

$$f^0(\vec{r}) \approx 1/2[\rho_{HOMO}(\vec{r}) + \rho_{LUMO}(\vec{r})]$$

b. Calcul du coefficient de Fukui

Depuis plusieurs années, différentes méthodes ont été développées pour approximer cette dérivée. Une forme condensée de la fonction de Fukui est introduite (Yang 1986) et consiste à assigner une valeur f_k à chaque atome k d'une molécule. Dans l'approximation de « différence finie » (LDA), les coefficients de Fukui « condensés » sont :

$$f_k^- = q_k(N) - q_k(N-1)$$

$$f_k^+ = q_k(N+1) - q_k(N)$$

$$f_k^0 = 1/2[q_k(N+1) - q_k(N-1)]$$

Où $q_k(N)$ est la charge de Mulliken du $k^{\text{ième}}$ atome de la molécule à N électrons, et $q_k(N-1)$ et $q_k(N+1)$ sont les charges du $k^{\text{ième}}$ atome de la molécule à $N-1$ et $N+1$ électrons dans une approximation d'orbitale gelée, respectivement (Mulliken 1955, Balawender 1998). Il est également possible d'appliquer cette équation avec les charges issues d'une population de Hirshfeld (Roy 1999) ou d'une population NBA ou encore CHELPG (Breneman 1990).

Les avantages et inconvénients de ces différentes méthodes de partitionnement ont été largement étudiés (Thanikaivelan 2002, Roy 2000, 2001 et 2003), notamment l'apparition de coefficients de Fukui négatifs et donc incohérents (Roy 1999). Ce problème semble moins présent avec la population de Hirshfeld, qui pourtant ne prédit pas de façon fiable les ordres de réactivité (Thanikaivelan 2002, Roy 1999).

La méthode de choix pour calculer les coefficients de Fukui est donc celle, néanmoins moins aisée, décrite par les équation ci-dessous (Contreras 1999). Ce modèle est basé sur l'exacte relation entre le descripteur de réactivité local et les orbitales moléculaires frontières (HOMO et LUMO). Ainsi, pour un atome k , le coefficient de Fukui pour une attaque électrophile (ou nucléophile) peut s'exprimer à l'aide des coefficients C_α des orbitales atomiques de sa HOMO (respectivement, de sa LUMO):

$$f_k^- \approx |\phi_{HOMO}|^2 = \sum_{\alpha \in k} |C_{\alpha HOMO}|^2 + \sum_{\beta \neq \alpha} |C_{\beta HOMO} C_{\alpha HOMO}| S_{\beta\alpha}$$

$$f_k^+ \approx |\phi_{LUMO}|^2 = \sum_{\alpha \in k} |C_{\alpha LUMO}|^2 + \sum_{\beta \neq \alpha} |C_{\beta LUMO} C_{\alpha LUMO}| S_{\beta\alpha}$$

avec $S_{\beta\alpha}$ l'intégrale de recouvrement entre les orbitales atomiques $\chi_\alpha(\vec{r})$ et $\chi_\beta(\vec{r})$

Cette définition a l'avantage d'éviter l'obtention de valeurs négatives des coefficients, ainsi que des calculs additionnels sur les espèces ionisées associée à la molécule, maintenant ainsi la multiplicité du système. Un unique simple point permet d'accéder directement aux quantités f_k^- et f_k^+ .

c. Limites du coefficient de Fukui

Le bon usage des coefficients de Fukui est remis en question, ainsi que son manque d'universalité. Son lien étroit avec la théorie FMO, mais également avec la théorie HSAB, est souligné : cependant, si la théorie FMO peut être interprétée comme un exemple du principe HSAB, ces deux principes entrent parfois en conflit : alors qu'un coefficient de Fukui maximum pourra prédire correctement le composé le plus nucléophile dans une réaction « soft-soft », il se peut qu'une réaction entre deux composés qualifiés de « durs », et donc placée sous contrôle de charge et ne suivant pas la théorie des orbitales frontières, ne puisse être prédite par les coefficients de Fukui (Li 1995).

Ceux-ci sont admis comme indicateurs fiables de la réactivité dans des réactions entre composés mous, mais d'autres indices, comme les charges et la dureté locale $\eta_{(\vec{r})}$ (cf I.D.2), sont préférables dans le cas des interactions dites « hard-hard » (Melin 2004). C'est le cas notamment dans les réactions de protonation des alkylamines pour lesquelles le site de la protonation n'est pas celui présentant la fonction de Fukui maximale (Yang 1986) et des hydroxylamines, dont le site de protonation est mieux décrit par les charges que les coefficients de Fukui (Melin 2004).

2. Dureté, mollesse et indices dérivés

D'autres descripteurs de réactivité sont basés plus directement sur les principes de dureté et mollesse afin de pallier au manque d'universalité du coefficient de Fukui, notamment en ce qui concerne les réactions dites dures.

La dureté η est définie comme la dérivée seconde de l'énergie E d'une molécule par rapport au nombre d'électrons, à un potentiel externe fixe (Parr 1983) ce qui est interprété comme la résistance du potentiel chimique μ aux changements de nombre d'électrons N .

$$\eta = \left(\delta^2 E / \delta N^2 \right)_{V(\vec{r})} = \left(\delta \mu / \delta N \right)_N$$

La dureté locale est définie comme :

$$\eta_{(\vec{r})} = -(V_{el}(\vec{r}) / 2N)$$

avec $V_{el}(\vec{r})$ la contribution électronique du potentiel électrostatique moléculaire (Bonaccorsi 1970, Berkowitz 1988). Ce paramètre est jugé fiable pour les réactions dirigées par les transferts de charge et est aussi préconisé pour la prédiction des réactivités de séquences intermoléculaires (Roy 1998), notamment

lorsqu'une molécule présente plusieurs sites possibles d'attaque électrophile. Dans ce cas, les coefficients de Fukui sont en effet incapables de discerner le site le plus réactif (Melin 2004, Roy 2004). La nucléophilie d'une série de composés cétoniques et aldéhydiques, dont certains possèdent une double liaison en alpha du carbonyle, est ainsi mieux décrite avec la dureté locale (Roy 1998 et 2004). Il en est de même pour une famille de chlorures insaturés (Roy 2005 et 2006, Meneses 2007).

Etant donné la capacité de la dureté locale de prédire la réactivité de molécules avec plusieurs sites nucléophiles, un autre indice basé sur la mollesse a été introduit.

La mollesse globale S d'une molécule est exprimée comme l'inverse de la dureté (Yang 1985) et traduit une mesure de la polarisabilité de la molécule :

$$S = 1 / 2\eta = (\delta N / \delta \mu)_{v(\bar{r})}$$

La mollesse locale s , soit celle condensée en un atome k , est alors égale à :

$$s_k = S \times f_k$$

Elle correspond donc au produit de la mollesse totale S et du coefficient de Fukui. Il est alors évident que la mollesse locale contient les mêmes informations sur la réactivité du site k considéré que le coefficient de Fukui, avec en supplément une information sur la mollesse globale de la molécule. C'est pourquoi ce critère s_k est considéré plus fiable pour la comparaison de réactivité de différentes molécules avec un partenaire de réaction commun (Roy 1998 et 1999). Cependant, cette quantité s_k conduira au même problème de non-positivité que celui du coefficient de Fukui si celui-ci est calculé à partir des charges des atomes et elle est également sensible aux changements de fonction de bases et aux effets de corrélation (Roy 1998).

Des indices de relative électrophilie (ou relative nucléophilie) ont donc été définis, respectivement par les rapports s_k^+ / s_k^- et s_k^- / s_k^+ (Roy 1998) pour annihiler ces effets de bases. Le site ayant le plus grand quotient s_k^- / s_k^+ serait donc le site préférentiel pour une attaque électrophile.

Cependant, même si ce nouveau descripteur prédit l'ordre de réactivité correcte de façon intramoléculaire, il n'en est pas forcément de même pour les comparaisons intermoléculaires. L'indice de relative nucléophilicité est dit uniquement applicable sur des sites k dont les valeurs s sont déjà importantes (Roy 2005). Du fait de ce manque d'universalité, ce facteur de réactivité n'a pas été largement exploité dans la littérature et de la même façon, les descripteurs locaux faisant intervenir le coefficient de Fukui présenteront cette même limitation.

3. Indice d'électrophilie

Il est parfois préféré aux index de mollesse S et s_k les facteurs d'électrophilie global et local définis respectivement par :

$$\omega = \mu^2 / 2\eta \text{ avec } \mu \text{ le potentiel chimique de la molécule (Parr 1999)}$$

$$\text{et } \omega_k = \omega \times f_k \text{ (Chattaraj 2003).}$$

La valeur globale de ω est naturellement proportionnelle au coefficient de Fukui comme l'indique la

$$\text{relation suivante : } \omega_k = \omega \times f_k = \frac{\mu^2}{2\eta} f_k = \mu^2 S f_k = \mu^2 s_k$$

D'où des ordres de réactivité intramoléculaire de tendance similaire donnés par l'un et l'autre coefficient et une fiabilité médiocre dans ce domaine (Roy 2004).

La philicité globale ω apporte néanmoins une information supplémentaire sur l'électronégativité de la molécule. Pour une comparaison intermoléculaire, à défaut d'avoir deux molécules d'électronégativité similaire, il semble donc nécessaire de comparer la réactivité de leur meilleur site, pour une réaction donnée, à l'aide de la philicité globale, plutôt que du coefficient de Fukui (Chattaraj 2003). Cependant malgré cette apparente supériorité du coefficient d'électrophilie, des contre-exemples sont apparus dans la prédiction de réactions intermoléculaires (Roy 2004 et 2005).

E. Application aux composés soufrés

Peu d'études ont été publiées sur la réactivité du soufre et des composés correspondants. Des informations sont néanmoins disponibles en filigrane dans la littérature.

Dans les études les plus anciennes, les molécules soufrées apparaissent dans les échelles parmi les plus composés les plus réactifs. L'ion donneur S^- est présenté comme le plus nucléophile parmi une famille de réactifs simples (F^- , H_2O , Cl^- , Br^- , OH^- et I^-). Notamment, la constante de nucléophilie, calculée à partir de facteurs de basicité et polarisabilité, est supérieure à toutes les autres (Edwards 1956). De même, la classification de Ritchie fait apparaître les plus grandes constantes de nucléophilie N_+ mesurées par cinétique pour le composé phénylbenzoate (Ritchie 1972). Ces valeurs expérimentales ont d'ailleurs été comparées récemment avec les énergies d'ionisation I_0 calculées au niveau MP2/6-311G(2d,p) pour diverses amines, alcools et thiolates (Contreras 2003). Il en résulte que les thiolates présentent une nucléophilie élevée avec $N_+ > 8.5$ alors que les amines et alcools ont une constante $N_+ < 6.0$.

Une étude plus récente s'efforce de relier le pouvoir accepteur de liaison hydrogène du soufre avec sa réactivité pour une famille de pyridinethiols possédant des substituants variés. Il apparaît que les constantes d'Hammett de ces divers substituants corrélaient avec l'énergie de dissociation de la liaison hydrogène et également avec l'affinité protonique du soufre (Nam 2006). Ainsi, l'affinité protonique constitue un bon indicateur de la basicité des soufres de cette série. Etant donné l'étroite relation existant entre la nucléophilie et la basicité, l'affinité protonique sera retenue comme un indice éventuel de la nucléophilie à tester sur nos complexes.

II. Calculs préliminaires sur les complexes biomimétiques

La première étape dans notre recherche d'un indice de réactivité des thiolates liés au zinc consiste à tester les critères utilisés dans la littérature sur un petit jeu de molécules. Nous utiliserons les complexes biomimétiques **1** à **4** étudiés précédemment (Chapitre III) puisque les énergies d'alkylation de leur ligand soufré ont déjà été calculées. Ces barrières énergétiques permettent de façon fiable d'apprécier la nucléophilie du thiolate lié au zinc et pourront ainsi nous aider à déterminer l'indice de nucléophilie le plus adéquat pour les composés soufrés. Nous ajouterons également deux complexes **1_c** et **3_c** dérivés respectivement des complexes **1** et **3** et pour lesquels l'atome de bore a été remplacé par un carbone.

L'ordre de réactivité nucléophile des arylthiolates est présenté sur la figure V.1. Les énergies d'alkylation entre le complexe considéré et l'iodure de méthyle ont été déterminées en phase gaz au niveau B3LYP/BS2//B3LYP/BS1.

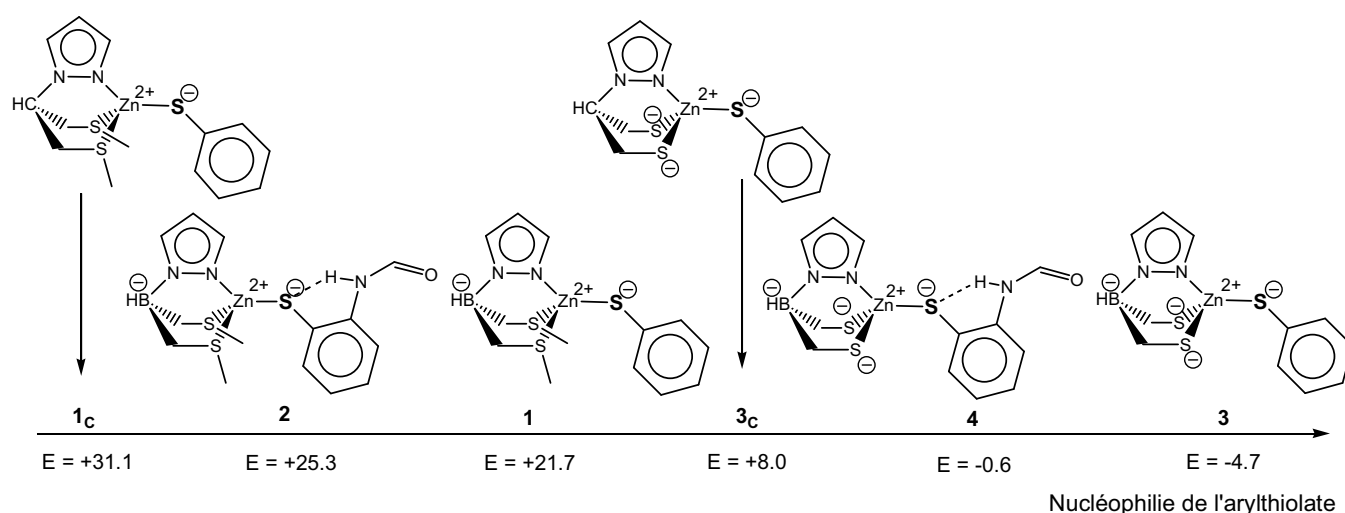


Figure V.1 : Complexes biomimétiques avec l'énergie d'alkylation de leur arylthiolate (en kcal/mol)

Ces complexes biomimétiques ont également pour avantage de présenter plusieurs sites possibles d'alkylation. La capacité de l'indice à distinguer différents soufres nucléophiles pourra éventuellement être évaluée. Cependant, dans un premier temps, seul l'arylthiolate sera considéré et sa nucléophilie mesurée.

Les barrières énergétiques de référence (Figure V.1) seront comparées avec des indices calculés à l'aide du logiciel GAUSSIAN et basés sur la configuration électronique seule du site réactif, comme les charges et les coefficients de Fukui, ainsi que l'affinité protonique du phénylthiolate correspondant.

A. Charges

Nous avons testé la capacité des charges de Mulliken et charges NBO à évaluer la nucléophilie des thiolates de nos complexes. Sur la figure V.II sont représentées les charges de Mulliken et NBO des arylthiolates, ainsi que leurs énergies d'alkylation.

Il est évident qu'aucune des populations de Mulliken ou NPA ne corrèle avec l'énergie d'alkylation de l'arylthiolate et par conséquent, aucune n'est proportionnelle à la nucléophilie du soufre. Ainsi, dans le complexe **4**, le thiolate lié au zinc participe à une liaison hydrogène et présente une énergie d'alkylation plus grande que le thiolate « nu » du complexe **3**. Ce dernier est donc plus nucléophile et devrait posséder une charge plus importante que le thiolate du complexe **3**. Alors que les charges de Mulliken suivent correctement cette tendance, les charges NBO prévoient que le soufre du complexe **4** présente une charge plus négative que celui du composé **3**. De même, les charges du phénylthiolate dans **1** devraient être plus importantes que celle du soufre, participant à une liaison hydrogène dans **2**. Les deux théories ne représentent donc pas correctement l'ordre de nucléophilie des thiolates. De plus, les valeurs déterminées pour les six complexes couvrent un petit intervalle ($\Delta_q = 0.114$), ce qui ne permet pas réellement de les distinguer suivant leur capacité nucléophile.

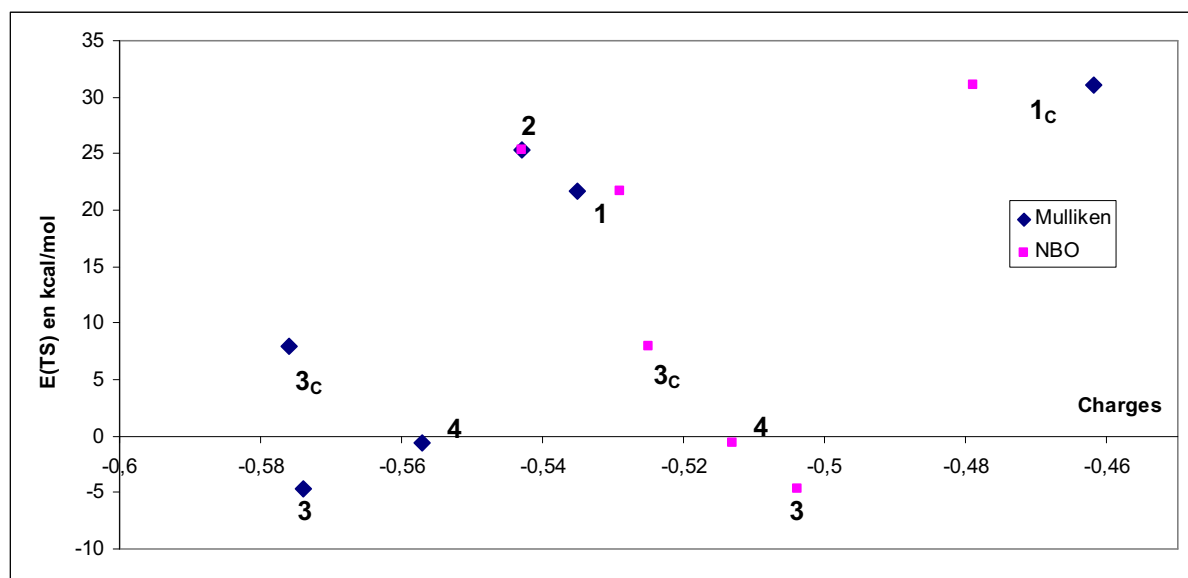


Figure V.2 : Charges de Mulliken et NBO des complexes biomimétiques

B. Coefficient de Fukui et mollesse locale

Les coefficients de Fukui ont été déterminés pour les six arylthiolates à l'aide de leurs coefficients orbitaux, de la matrice de densité et de l'analyse de population de Mulliken disponibles à partir d'un calcul de population du point stationnaire correspondant. L'équation utilisée pour calculer le coefficient de Fukui condensé sur le soufre est la suivante :

$$f_k^- = \sum_{\alpha \in k}^{AO} |C_{\alpha HOMO}|^2 + \sum_{\beta \neq \alpha}^{AO} |C_{\beta HOMO} C_{\alpha HOMO}| S_{\beta \alpha}$$

Afin de rendre compte des réactions se déroulant sous contrôle non-orbitaire et donc avec une orbitale d'énergie moins élevée que la HOMO, il nous a paru intéressant de donner les coefficients de Fukui associés à l'orbitale HOMO-1 dont l'énergie est située juste en-dessous de la HOMO :

$$f_k^{HOMO-1} = \sum_{\alpha \in k} |C_{\alpha HOMO-1}|^2 + \sum_{\beta \neq \alpha} |C_{\beta HOMO-1} C_{\alpha HOMO-1}| S_{\beta\alpha}$$

En effet, les alkylthiolates des complexes **3**, **3c** et **4** peuvent également effectuer des alkylations. Leur charge est plus localisée que celle des phénylthiolates et la HOMO pourrait donc présenter un plus gros coefficient sur ces soufres. Les phénylthiolates pourraient alors posséder un important coefficient de la HOMO-1, mais non maximal, compatible avec leur capacité nucléophile plus modérée.

Les calculs des coefficients atomiques de Fukui sont effectués à l'aide d'un script Perl (cf Annexe) et les résultats sont donnés sur le schéma suivant.

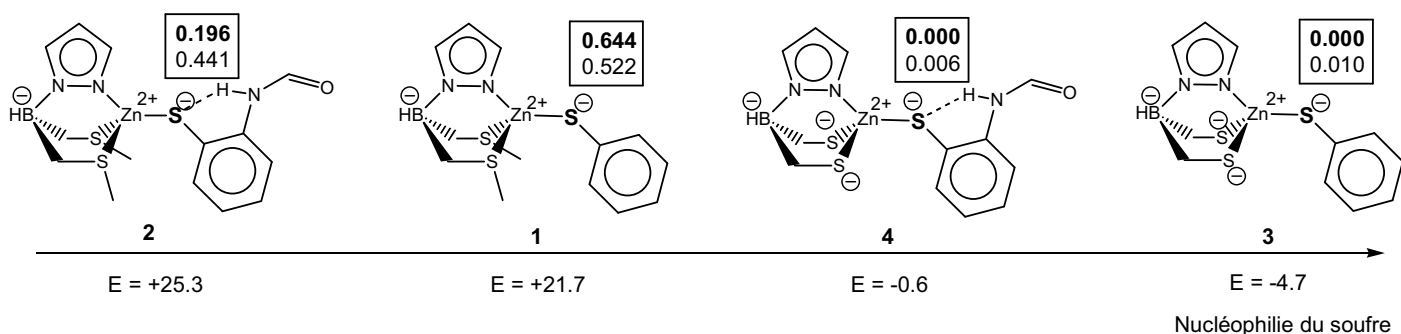


Figure V.3 : Coefficients de Fukui de l'arylthiolate pour la HOMO (gras) et HOMO-1 (normal)

Des premiers calculs effectués sur les complexes **1** à **4**, il est clair que le coefficient de Fukui ne permet pas d'évaluer, même globalement, la nucléophilie de l'arylthiolate considéré.

En effet, les thiolates les plus nucléophiles des complexes **3** et **4** possèdent un coefficient de Fukui nul, alors que le soufre le moins réactif présente un coefficient de 0.196. Le plus gros coefficient ($f = 0.644$) est atteint pour le phénylthiolate du complexe **1** qui présente toutefois une énergie d'alkylation élevée ($E = 21.7$ kcal/mol). La plus petite valeur trouvée pour le thiolate « nu » dans **1** par rapport à celui dans **2** peut éventuellement s'expliquer par un nombre plus important d'atomes dans ce dernier complexe. En effet, le coefficient de Fukui est normalisé à la valeur 1 pour chaque orbitale. Ainsi pour une orbitale donnée, les coefficients de Fukui seront d'autant plus petits que l'orbitale sera répartie entre plus d'atomes.

Par ailleurs, les coefficients de la HOMO sur les alkylthiolates des complexes **3** et **4** sont respectivement $f = 0.480$ et 0.455 (données non représentées sur la figure V.3). Alors que ces valeurs suggèrent la supériorité nucléophile attendue des alkylthiolates par rapport aux arylthiolates, elles ne représentent pourtant pas la plus grande réactivité attendue pour le thiolate dans **4** par rapport à celui de **3**.

Les coefficients de Fukui pour la deuxième orbitale considérée pour une attaque nucléophile ne permet pas non plus de classer les arylthiolates suivant leur ordre de réactivité. Ceux-ci sont importants pour les complexes **1** et **2** ($f^{HOMO-1} = 0.441$ et 0.522), indiquant une densité électronique et une nucléophilie très centrée sur ses thiolates par rapport aux autres soufres de la molécule. Les

faibles valeurs trouvées pour les composés **3** et **4** ($\epsilon^{\text{HOMO}-1} = 0.010$ et 0.006) suggèrent une répartition plus homogène de la réactivité entre les différents thiolates des complexes.

Les coefficients de Fukui, plus longs à obtenir que les charges, ne sont donc pas des indicateurs fiables de la nucléophilie des thiolates au zinc. Nous ne tenterons pas de calculer cet indice sur d'autres complexes.

La mollesse locale a également été déterminée pour ces phénylthiolates à partir de la formule suivante : $s_k = S \times f_k$ avec $S = \frac{1}{2} \times (E_{\text{LUMO}} - E_{\text{HOMO}})$. Les coefficients obtenus pour les composés **1** et **2** sont respectivement $s = 6.468$ et 4.489 , alors que pour les complexes riches en électrons **3** et **4**, ces indices sont nuls. Nous pouvons en conclure que tout coefficient dérivé du coefficient de Fukui ne sera pas adapté à l'estimation de la nucléophilie.

C. Affinité protonique (AP)

Le dernier critère testé est l'affinité protonique du thiolate considéré. Celles-ci ont été calculées en faisant la différence des énergies de l'espèce protonée sur l'arythiolate et de l'espèce non protonée. Les valeurs des affinités protoniques sont représentées sur la figure suivante :

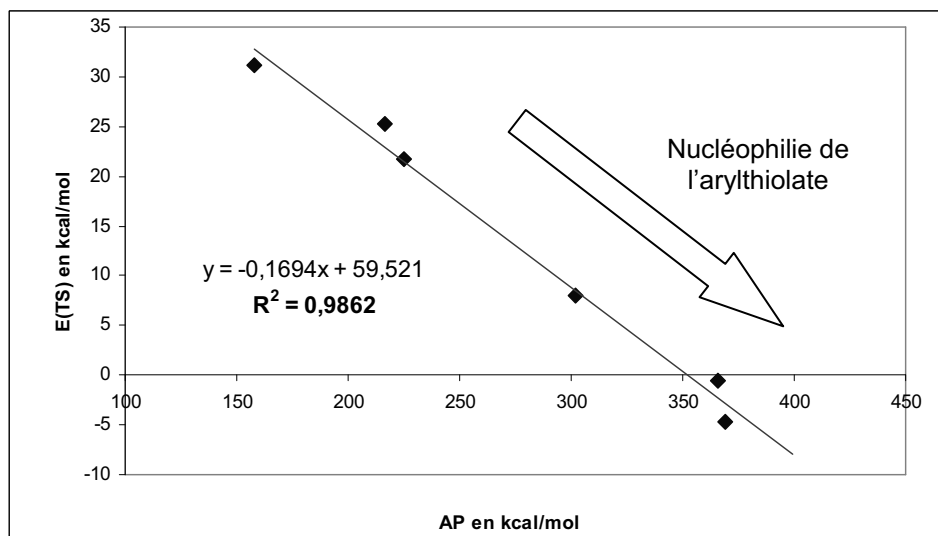


Figure V.4 : Corrélation entre l'affinité protonique des arylthiolates et leur énergie d'alkylation

Les affinités protoniques obtenues permettent d'ordonner les phénylthiolates suivant quatre classes. Les valeurs d'affinités protoniques les plus grandes (> 350 kcal/mol) sont obtenues pour les sulfures les plus nucléophiles, à savoir ceux des complexes riches en électrons **3** et **4**. Une valeur intermédiaire proche de 300 kcal/mol est obtenue pour le complexe monoanionique **3c**. Dans la gamme inférieure (200 à 250 kcal/mol) se trouvent les affinités protoniques des complexes neutres **1** et **2** et dans un quatrième intervalle (< 200 kcal/mol), celle du composé **1c**.

De plus, ces affinités protoniques corrélaient parfaitement avec les énergies d'alkylation des arylthiolates, comme l'indique le facteur de régression supérieur à 0.98. La droite obtenue peut aisément servir pour la prédiction d'autres thiolates liés au zinc dont l'affinité protonique est connue. Ainsi, le calcul du chemin réactionnel et de l'énergie d'alkylation sera évité et remplacé par la détermination, plus rapide, de l'affinité protonique.

Ceci sera particulièrement utile pour l'étude de la réactivité des thiolates d'un site actif protéique, puisque la grande taille de ces systèmes rend la location d'un état de transition impossible. L'ordre de réactivité des thiolates liés au zinc, par exemple dans un centre structural ZnCys₄, pourra être déterminé par leurs affinités protoniques respectives et expliquer la sélectivité de l'alkylation envers un unique thiolate.

Afin de valider l'affinité protonique comme indicateur fiable de la nucléophilie des thiolates liés au zinc, il est préférable de valider ce critère sur un jeu plus important de thiolates. La troisième partie de ce chapitre est consacrée à l'étude plus précise d'une large famille de complexes biomimétiques.

III. Nucléophilie des thiolates liés au zinc

A. Introduction

La sélectivité de l'alkylation enzymatique envers un thiolate dans un site actif possédant plusieurs ligands de même nature nécessite un critère de nucléophilie fiable et rapide des soufres liés au zinc. Les complexes biomimétiques étudiés dans le chapitre III, possédant un squelette tripodal, ont l'avantage de présenter également plusieurs sites d'alkylation possibles. La nucléophilie de leurs phénylthiolates a été mesurée par l'énergie libre requise pour leur alkylation par mécanisme S_N2. Cette barrière énergétique s'est révélée être une référence utile pour évaluer la nucléophilie du thiolate lié au zinc et a montré qu'une liaison hydrogène impliquant le thiolate réduisait sa capacité à être alkylé (Chapitre III).

Les échelles de réactivité existantes et diverses études cinétiques n'ont pas permis à ce jour de développer un critère de nucléophilie pour les thiolates des complexes du zinc, du fait de conditions expérimentales très variées (Wilker 1997, Chiou 2003, Smith 2005, Morlok 2005). Cependant, le développement des méthodes de chimie quantique et notamment de la DFT conceptuelle a mené au développement de critères de réactivité permettant de s'affranchir aisément des partenaires réactionnels et des conditions de l'alkylation (cf V.I). Nous avons testé la cohérence de tels indices sur nos complexes biomimétiques et vérifié leur corrélation avec l'énergie d'alkylation de leur arylthiolate, déterminée dans le chapitre III. Alors que les charges atomiques et les coefficients de Fukui ne permettent pas d'évaluer la nucléophilie d'un thiolate lié au zinc, l'affinité protonique est proportionnelle à la barrière d'activation et semble être ainsi un indicateur correct et pratique de sa réactivité (cf V.II).

Afin de confirmer cette hypothèse, une famille plus large de composés biomimétiques du zinc sera développée, en faisant varier la nature des ligands du zinc (N ou S) et la charge totale du complexe (-2 à +2) par modification des substituants de la cage tripodale. Les énergies libres de Gibbs pour l'alkylation des phénylthiolates et des alkylthiolates seront calculées et la corrélation avec l'affinité protonique du soufre considéré sera vérifiée. La structure des complexes sera également mise en relation avec les variations de nucléophilie afin de mettre en avant les facteurs influençant la réactivité des complexes biomimétiques du zinc.

B. Résultats

Les énergies libres d'alkylation entre l'iodure de méthyle et seize composés biomimétiques, de charge totale -2 à +2 et possédant un ligand tripodal de type S_3 , NS_2 ou N_2S , ont été calculées en phase gazeuse au niveau B3LYP/BS2//B3LYP/BS1.

Les barrières énergétiques trouvées pour l'alkylation de l'arylthiolate sont comprises dans l'intervalle 15-243 kJ/mol, ce qui montre l'importance des ligands du zinc sur la nucléophilie intrinsèque d'un thiolate lié au métal. Cette réactivité est, par ailleurs, fonction des substituants des ligands et surtout de la charge totale du complexe biomimétique, alors qu'elle n'est pas influencée par la nature (N/S) des ligands du zinc comme il avait été suggéré dans de précédentes études (Wilker 1995, Rombach 2006).

La présence d'une liaison hydrogène dirigée vers l'arylthiolate réduit sa nucléophilie, comme il a été montré dans le chapitre III. Pour la première fois dans la littérature, nous avons montré que cette interaction a également un effet sur la réactivité globale du complexe, puisque la nucléophilie des autres thiolates présents dans la sphère de coordination du zinc est aussi diminuée.

Par analogie, la nucléophilie des sites protéiques ne sera pas fonction uniquement de la nature des résidus coordonnés au zinc (histidine ou cystéinate), mais sera impactée par l'ensemble constitué des substituants, des charges et des liaisons hydrogènes présentes dans la sphère de coordination du métal.

Enfin, la longueur zinc-soufre dans les complexes biomimétiques ne reflète pas la réactivité nucléophile du thiolate considéré. Par contre, l'énergie de la HOMO du composé organométallique est proportionnelle à sa capacité d'alkylation. L'affinité protonique corrèle également avec la réactivité des thiolates liés au zinc et permet, en outre, de distinguer les différents ligands selon leur nucléophilie. Nous avons donc confirmé l'hypothèse précédente et montré que l'affinité protonique du thiolate lié au zinc est un indicateur fiable de sa nucléophilie.

Son calcul permet d'envisager une prédiction rapide et directe de la réactivité des différents thiolates présents dans un même site actif et la rationalisation de la sélectivité de l'alkylation. La dernière partie de cette thèse montrera la mise en application de ce résultat, à savoir la prédiction de la nucléophilie des thiolates du C-terminus du doigt de zinc ZnCys₃His de la protéine nucléocapside p7 du virus VIH.

C. Publication

Nucleophilicity of zinc-bound thiolates

Delphine Picot, Gilles Ohanessian, Gilles Frison*

Laboratoire des Mécanismes Réactionnels,

Département de Chimie, Ecole Polytechnique and CNRS,

91128 Palaiseau Cedex, France.

Fax: + 00 33 (0)1 69 33 48 03

gilles.frison@polytechnique.org

Abstract: The nucleophilicity of biomimetic zinc-thiolate complexes against methyl iodide has been studied. The activation barrier has been computed with the B3LYP functional. The reactivity depends on the global charge, the nature of the ligand set and the presence of hydrogen bonds. This shows that the understanding of zinc site nucleophilicity can not be achieved by the knowledge of the atom donor set for zinc only, as currently done in the literature. Moreover, we show that sulfur proton affinity and activation barrier are directly proportional, thus providing a good nucleophilicity index for zinc-bound thiolate.

Keywords: Density functional calculations, Zinc, Nucleophilicity, Hydrogen bonds, S ligands, Tridentate ligands, Bioinorganic chemistry

Résumé: Nucléophilie du motif zinc-thiolate

Nous avons étudié la réactivité nucléophile de complexes biomimétiques du zinc, incluant le motif Zn-thiolate, vis-à-vis de l'iodure de méthyle. La barrière d'activation a été calculée en utilisant la fonctionnelle de la densité B3LYP. Les facteurs influençant la réactivité (charge totale, nature des ligands du métal, liaisons hydrogène) montrent que la seule connaissance des atomes liés au zinc, comme souvent fait dans la littérature, ne permet pas de comprendre la nucléophilie du site. De plus, l'affinité protonique du soufre corrèle avec la barrière d'activation, fournissant ainsi un indice de nucléophilie performant pour le motif zinc-thiolate.

Mots-clés: Calculs DFT, Zinc, Nucléophilie, Liaisons hydrogène, Ligands S, Ligands tridentate, Chimie bioinorganique

Introduction

Metal-thiolate bonds have been found in many sites of proteins where they play important biochemical functions.[1-3] For example, nature utilizes iron- and nickel-ligated thiolates to promote superoxide reductase[4] and superoxide dismutase[5, 6] functions, respectively. Besides its fundamental role in concentration modulation,[7] structural[8] and redox[7, 9] functions, metal-bound thiolate also reveals their nucleophilic character.

Zinc-bound thiolate has been shown to promote alkyl group transfer in several zinc enzymes.[10] This reactivity is reproduced by many zinc biomimetic complexes[11] as well as by iron-, [12] cadmium- and mercury-thiolate complexes.[13] The influence of the nature of the metal cation on the nucleophilicity of *p*-toluenethiolate has been studied both experimentally and by DFT computation, revealing that nickel- and zinc-thiolate complexes are more reactive than iron- and cobalt-thiolate complexes. In the case of zinc complexes, both the composition of the ligand set,[14-17] in terms of donating capability and steric hindrance, and the presence of hydrogen bonding towards the reactive thiolate[18-21] have been shown to modulate the reactivity against electrophiles.

Prediction of the nucleophilicity remains a challenging task.[22] Attempts to theoretically define a general nucleophilicity index,[23, 24] based for example on atomic charges, energy of the highest occupied molecular orbital, force constants or, in the context of conceptual density functional theory,[25] chemical hardness and Fukui function, have been shown to be conclusive only for a limited variety of compounds. In the case of zinc-bound thiolates, several studies provide the rate constants of their reaction against electrophile but it is difficult from these data to have a general view of their relative nucleophilicities as the experimental conditions are not the same.[16, 18, 20, 21, 26] Such a determination would however be of great significance in order to evaluate the relative reactivity of enzymes and to determine which thiolate is reactive in zinc active sites including more than one cysteinate.

In a previous study,[27] we have shown that the computed Gibbs free energy of the S_N2 activation barrier can explain the relative reactivity of zinc-thiolate complexes against methyl iodide. In this study, we have determined the S_N2 activation barrier for a variety of zinc complexes including a relatively rigid tripodal core in order to conserve as much as possible along the reaction path the tetrahedral arrangement around zinc observed in biomimetic complexes and enzymatic active sites.[28] This will allow us to compare the donating capability of tripodal ligands and to determine a good correlation between the zinc-bound thiolate nucleophilicity and the thiolate gas phase basicity.

Computational methods

Calculations were performed with Gaussian 03.[29] Geometry optimisations were conducted using the B3LYP method at the 6-31G(d,p) level for the B, N, C, O, S, H atoms. The contracted Wachters basis [14s9p5d1f/9s5p3d1f] was used to describe the zinc atom.[30] The CRENBL relativistic effective core potential and associated valence basis set were employed to model the iodine atom.[31] This basis set is referred to as BS1. Each stationary point has been characterized with frequency analysis and shows the correct number of negative eigenvalues (0 for a local minimum and one for a transition state).

Energies were calculated for the stationary points at the B3LYP level using an extended basis set labelled BS2. It consists in the 6-311+G(2d,2p) for B, N, C, O, S, H, the extended Wachters basis [15s11p6d2f/10s7p4d2f] for Zn and the Aug-cc-pVTZ-PP basis set and pseudo-potential for I.[32] We have demonstrated previously that this level of calculation gives reliable geometries and relative energies on zinc complexes.[27, 33]

The relative gas phase Gibbs free energy was deduced from the equation:

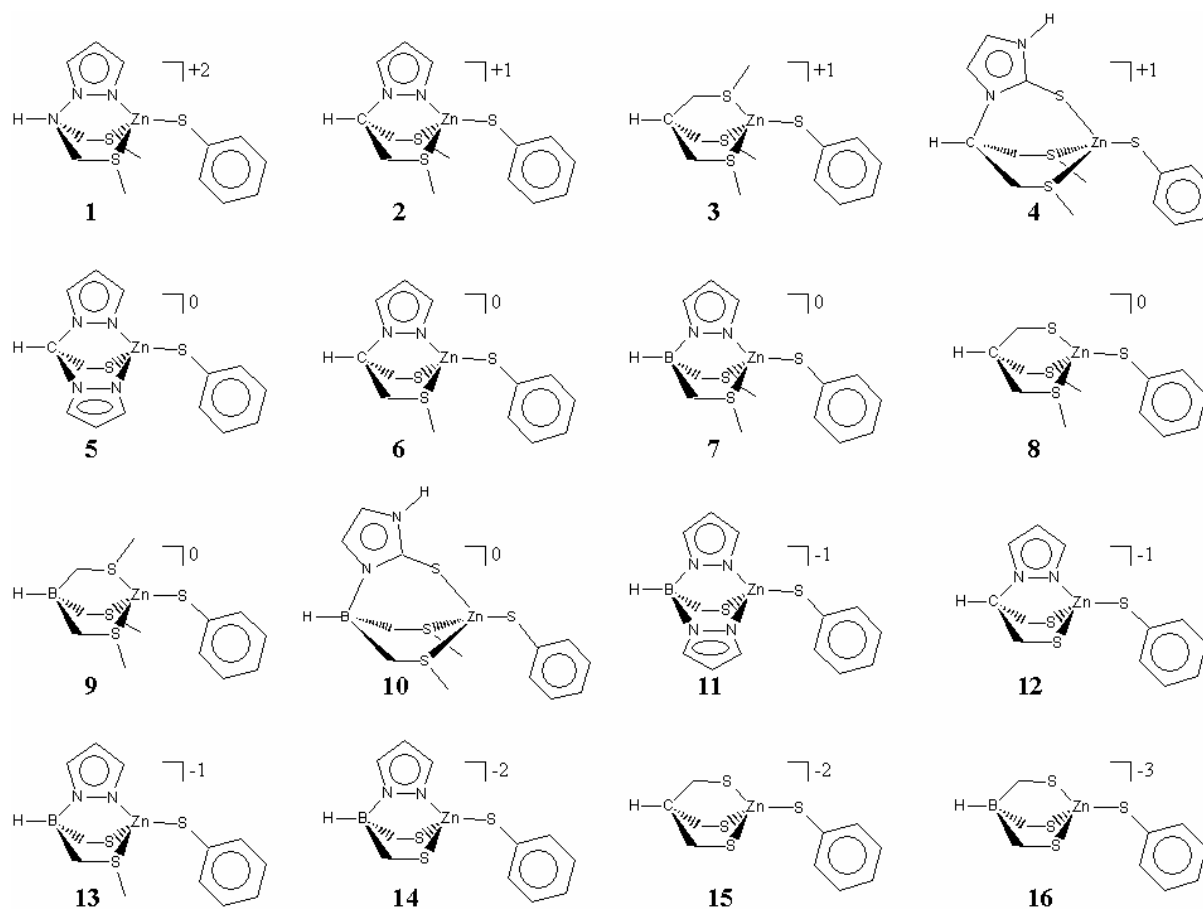
$$\Delta G_{\text{gas}} = \Delta E_{\text{elec}} + \Delta \text{ZPE} + \Delta E_{\text{T}} - T\Delta S$$

with ΔE_{elec} , ΔZPE , ΔE_{T} and ΔS the differences in the electronic energy, zero-point vibrational energy, thermal energy and entropy between products and reactants, respectively.

All calculations have been done in the gas phase, thus permitting to know the intrinsic properties and reactivities of the complexes. Furthermore, we have shown previously[27] that inclusion of solvation effect does not modify the relative reactivity of zinc thiolate complexes. The gas phase nucleophilicity trends can thus be extrapolated to solvated biomimetic complexes or enzymatic active sites.

Results and discussion

A series of 16 zinc biomimetic complexes depicted in scheme 1 has been examined in this work.

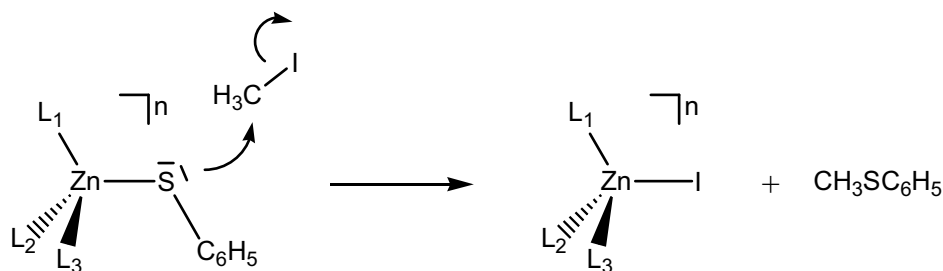


Scheme 1. thiophenolate zinc complexes studied herein

They are derived from experimental complexes bearing a tripodal ligand and one monodentate phenylthiolate ligand. The tripod ligands possess a mixture of pyrazole-type nitrogen, and of thioimidazole-, thioether- or alkylthiolate-type sulfur as the donors for zinc. This allows to obtain N_2S_2 , NS_3 or S_4 donor sets which model respectively the His_2Cys_2 , $HisCys_3$ and Cys_4 coordination observed in many active sites of zinc enzymes.[34-36] Furthermore, the tripodal core is either a borate, a methylene or an ammonium moiety, thus permitting to modulate the global charge of the complexes. Consequently, our series of zinc thiolate complexes goes from the dicationic $[ZnNS_3]^{2+}$ **1** to the trianionic $[ZnS_4]^{3-}$ **16** species. Optimisation of the geometry of these complexes led to a tetrahedral coordination around zinc in all cases except **16** for which the phenylthiolate ligand dissociates from the metal, leading to a tricoordinated zinc complex. The high electronic density around zinc in **16** explains this incapability to keep four Zn-S bonds. Compared to **15** for which a ZnS_4 core could be obtained, this also means that the negative charge of the borate group in **16** influences the zinc atom even in the absence of conjugated arms. We can thus hypothesize that the presence of a negatively charged group in the neighbourhood of a $[Zn(Cys)_4]^{-2}$ enzyme active site facilitates the sulfur-zinc bond dissociation. Reciprocally, the shorter Zn-thiolate bond in **1** (2.182 Å) compared to **2** (2.204 Å) shows that a cationic site close to the coordination sphere tightens the Zn-S bonds.

a. Reactivity of zinc-bound phenylthiolate

For complexes **1-15**, we have determined the Gibbs free energy barrier of their S_N2 reaction with methyl iodide (Scheme 2 and Figure 1).



Scheme 2. S_N2 alkylation of zinc-bound phenylthiolate with MeI

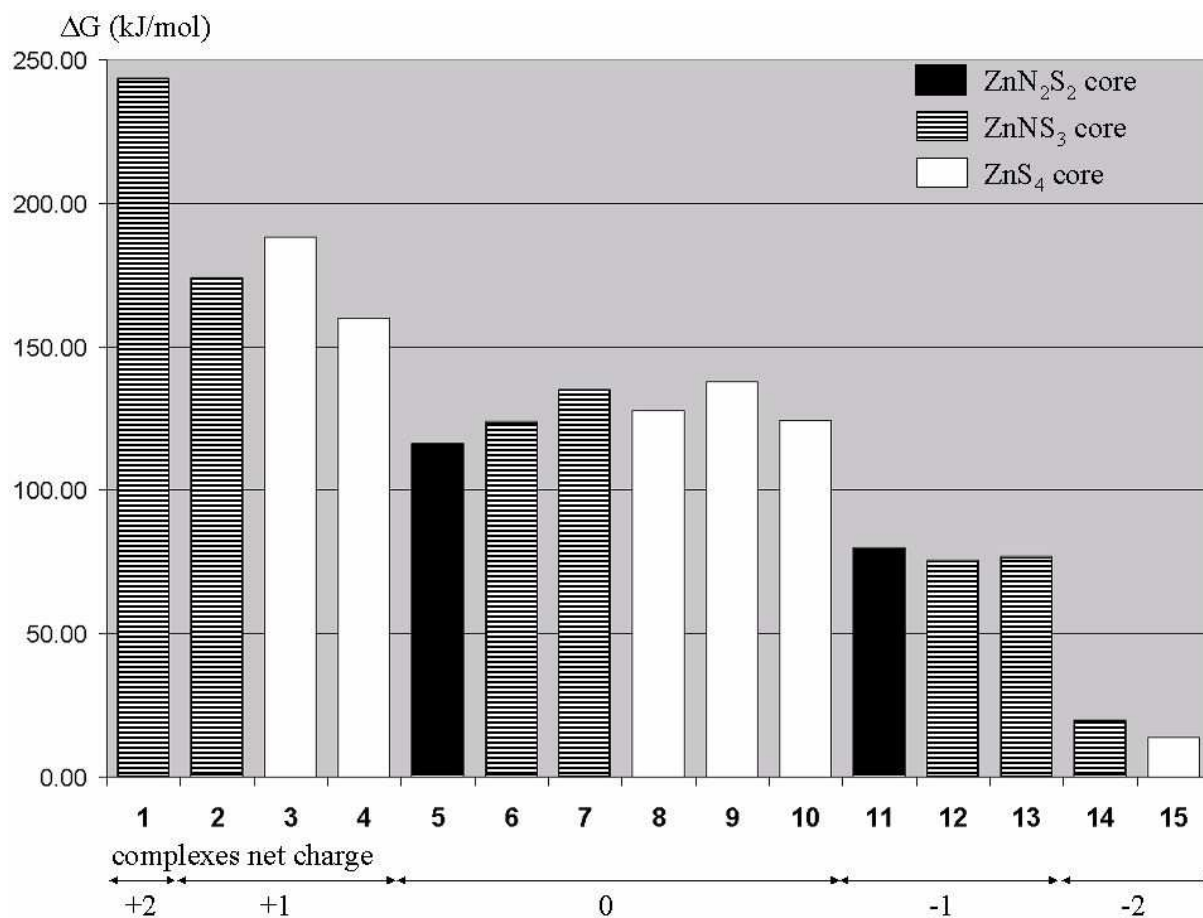


Figure 1. Gibbs free energy barrier (ΔG) for the reaction between MeI and complexes **1-15**, calculated at the B3LYP/BS2//B3LYP/BS1 level (in kJ/mol).

The less and the more reactive complexes are **1** and **15** with a barrier of 243 and 15 kJ/mol, respectively. The big difference between these extreme values illustrates the importance of the coordination sphere around zinc on the intrinsic nucleophilicity of the thiolate. Figure 1 shows however that the nature of only the atom donor set for zinc does not determine the

nucleophilicity of the complex. Indeed, a ZnN_2S_2 core may be either more or less reactive than ZnN_3S or ZnS_4 cores, and the same holds for the same comparison between ZnN_3S and ZnS_4 cores. This shows that the conclusion obtained in previous studies[14, 16] that ZnS_4 complexes are more reactive than ZnNS_3 and ZnN_2S_2 complexes cannot be generalized to a broader series. It should be noticed that a counter-example of the previous rule has already been mentioned.[16]

Figure 1 shows that the nucleophilicity depends on the complex net charge. Indeed dianionic, monoanionic, neutral, monocationic and dicationic complexes have their Gibbs free energy barrier in the range 0-50, 50-100, 100-150, 150-200 and 200-250 kJ/mol, respectively. This is in agreement with the decreased reactivity order measured from $[\text{ZnS}_4]^{-2}$ to $[\text{ZnNS}_3]^{-1}$ and $[\text{ZnN}_2\text{S}_2]^0$ complexes.[14, 26]

As for the complete set of complexes, the relative reactivity between complexes bearing the same charge does not depend on the nature of the atom donors for zinc. This is illustrated for example by the fact that **2**, a $[\text{ZnNS}_3]^{+1}$ complex, is respectively more and less reactive than **3** and **4**, which are both $[\text{ZnS}_4]^{+1}$ complexes. On the contrary, the nucleophilicity of the zinc-bound phenylthiolate is influenced by the nature of the tripod arms. Substituting a pyrazolyne arm by a thioether arm, as from **2** to **3** or from **5** to **6**, reduces indeed the reactivity of the phenylthiolate. By comparing the activation barrier for all complexes, it is thus possible to determine a scale of the ligands which induce the larger nucleophilicity (Figure 2).

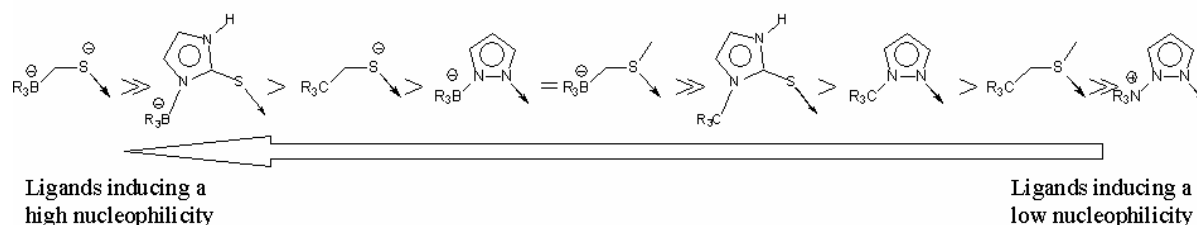


Figure 2. Scale of ligands influencing the nucleophilicity of zinc-bound thiolate

This scale is in agreement with experimental results indicating that substituting a pyrazolyl nitrogen ligand by a thioimidazole sulfur ligand increases the reactivity.[16] It also shows that thioimidazole is intermediate between a thioether and a thiolate group in terms of their capability of thiolate enrichment.[37, 38] On the contrary we may notice that a thioimidazole borate is a more donating ligand than an alkylthiolate.

b. Reactivity of zinc-bound alkylthiolate

Some of the tripodal ligands possess an alkylthiolate arm which can be alkylated, as already observed experimentally.[20] For complexes **5**, **6**, **11**, **12** and **13**, we have calculated the transition state corresponding to the S_N2 reaction between the zinc-bound alkylthiolate and methyl iodide. The values of the Gibbs free energy barrier for this reaction are given in Table 1 in parallel to the values obtained for the same reaction on the phenylthiolate.

Table 1. Gibbs free energy barrier (ΔG) for the reaction between MeI and zinc-bound alkyl- or aryl-thiolate, calculated at the B3LYP/BS2//B3LYP/BS1 level (in kJ/mol).

Complex	ΔG for the arylthiolate	ΔG for the alkylthiolate
5	116	124
6	124	135
6'	137	141
11	80	52
12	76	64
13	77	46
13'	92	53

These results indicate that the alkylthiolate nucleophilicity in a given complex stands in the same range as the one observed for the phenylthiolate in the same complex. Furthermore, the alkylthiolate reactivity may be either higher (in **11**, **12** and **13**) or smaller (in **5** and **6**) than the reactivity of the phenylthiolate. The order of reactivity of these zinc-bound thiolates is thus not determined by their substituents but by the nature of the ligand set around zinc. In our

series, it seems that the presence of a second negative charge in the tripod ligand, as in **11**, **12** and **13**, induces a higher reactivity of the alkylthiolate arm.

c. Influence of hydrogen bonding

We and others have shown in previous studies[18-21, 27] that substituting the phenylthiolate by the *o*-NHC(O)H-phenylthiolate, inducing a H bond with the sulfur atom of the arylthiolate, reduces the reactivity of the corresponding zinc-bound arylthiolate. The same aryl-substitution on complexes **6** and **13** leads to two new complexes, noted **6'** and **13'**, respectively, for which it is possible to evaluate the influence of the H bond toward the arylthiolate on the nucleophilicity of both the aryl- and the alkyl-thiolate.

As expected (Table 1), moving from **6** to **6'** or from **13** to **13'** induces an increase by 13-15 kJ/mol of the Gibbs free energy barrier for the reaction on the arylthiolate. More surprisingly, this also leads to a decrease of reactivity of the alkylthiolate as the Gibbs free energy barrier increases by 6-7 kJ/mol. The optimized geometry of complexes **13** and **13'** (or **6** and **6'**) (Figure 3) gives an explanation of this trend.

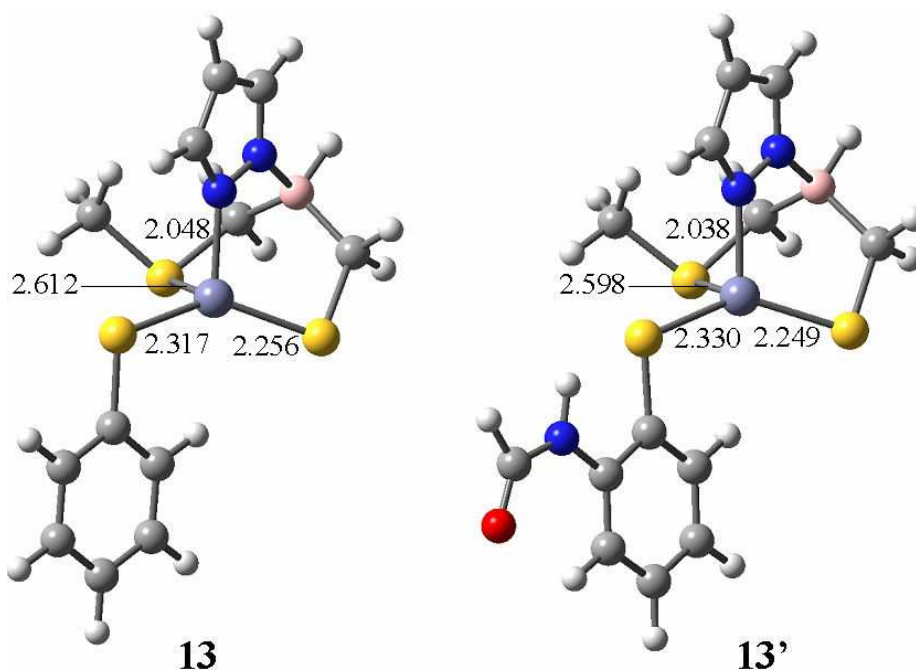


Figure 3. Optimized structures of **13** and **13'** at the B3LYP/BS1 level. bond length in Å. yellow : sulfur, blue : nitrogen, gray : carbon, white : hydrogen, red : oxygen, pink : boron, light blue : zinc.

The H bond reduces the Lewis basicity of the *o*-NHC(O)H-phenylthiolate compared to the bare phenylthiolate. Consequently, the Zn-S(aryl) bond is slightly longer in **13'** than in **13**. Due to the d^{10} electronic structure of zinc(II), we observe a “push-pull” effect between the ligand set around the metal.[28] Indeed, the lengthening of the Zn-S(aryl) bond induces a shortening of all other Zn-ligand bonds. Thus, the alkylthiolate of **13'** is slightly more tightly coordinated to zinc than the same alkylthiolate of **13**. In other words, the charge transfer between the alkylthiolate and Zn^{2+} is larger in **13'** than in **13**, thus explaining the lower nucleophilicity of the alkylthiolate of **13'** compared to the same thiolate in **13**.

These results indicate that hydrogen bond has not only a local effect on the reactivity of the thiolate group toward which it is directed, but also a global effect on the complex reactivity.

d. Nucleophilicity index

Having in hand the calculated reactivity of a large panel of zinc-bound thiolates, is it possible to find a zinc-bound thiolate nucleophilicity scale based on the structure of the complex ? It has been postulated that the Zn-S bond lengths[16] or the HOMO energy of the complex[15] could correlate with the complex reactivity.

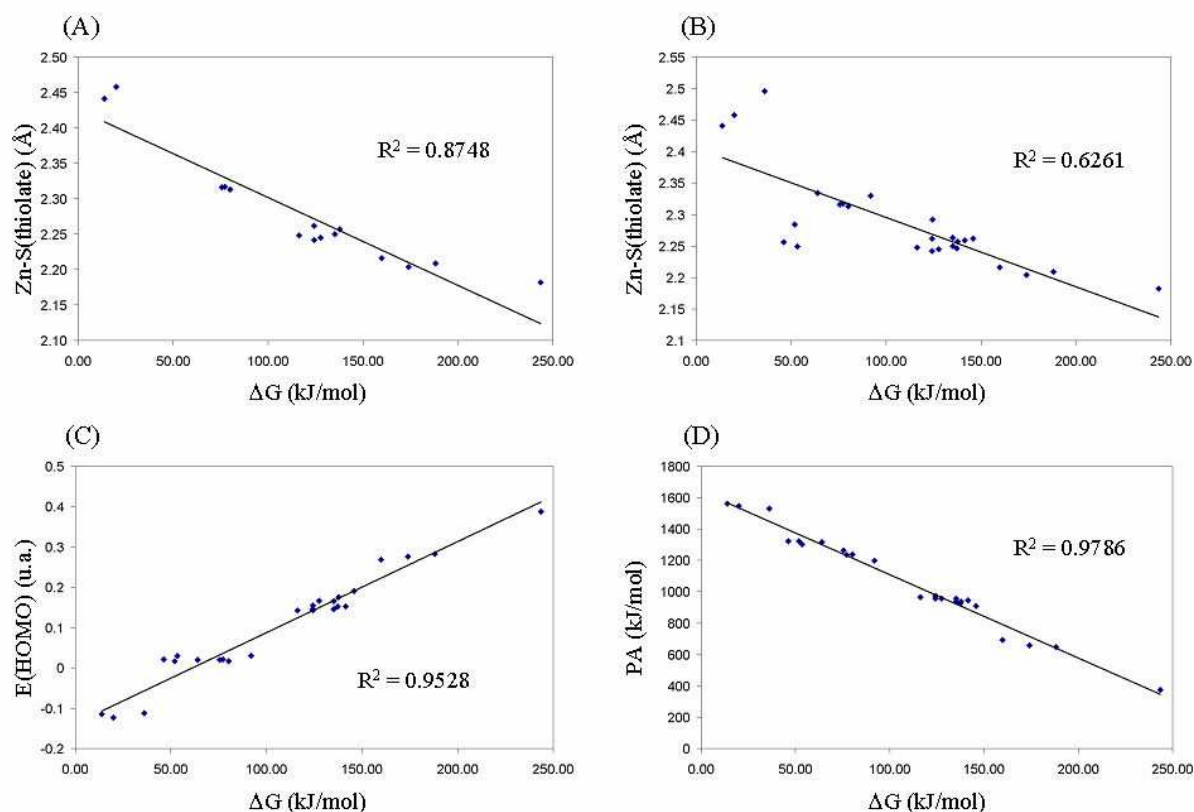


Figure 4. Plot of calculated Gibbs free energy barrier (ΔG) versus (A) Zn-S bond length for phenylthiolate (B) Zn-S bond length for all thiolates (C) energy of the HOMO for all complexes (D) proton affinity for all thiolates.

Figures 4A shows the relationship obtained between the energy barrier and the Zn-S(phenylthiolate) bond length for complexes **1-15**. As expected, an increase of the metal-sulfur bond length corresponds approximately to an increase of the reactivity of the phenylthiolate. The linear correlation remains however modest. Extension of this relationship to all zinc-thiolate sites, thus including alkylthiolate and *o*-NHC(O)H-phenylthiolate (Figure 4B) shows a clear disruption of the previous trend. This means that the reactivity of various zinc-bound thiolates cannot be deduced from the zinc-thiolate bond length.

On the contrary, the energy of the HOMO of all the complexes studied correlates reasonably with the calculated energy barrier (Figure 4C) and may thus be used to compare the reactivity of different zinc-thiolate complexes. This however does not allow to determine the most reactive site of a zinc complex if several thiolate ligands are bound to the metal.

In order to be able to predict the most reactive thiolate of a complex, we also examined the proton affinity of the zinc-bound sulfur atoms. It is well known that basicity and nucleophilicity are related.[15] However, due to the differences between these concepts, they are often not directly proportional.[39] Figure 4D shows a good correlation between the proton affinity and the activation barrier of all the zinc-thiolate moieties studied here. This shows that the zinc-bound thiolate basicity can be used as a pertinent nucleophilicity index in order to predict the relative reactivity of these coordinated thiolates.

Conclusion

Based on the calculated reactivity of 15 zinc-thiolate complexes, we have shown that the dominant factor inducing the nucleophilicity is the charge of the ligand set and the nature of the chemical groups bound to zinc. Thus the N/S ratio of atom donor set is not a sufficient parameter to determine the reactivity if the N and S ligands are not exclusively histidine and cysteinate side chains, respectively. The substituant of the thiolate does not give a clear indication of its reactivity, whereas the presence of H bond reduces the nucleophilicity of all reactive sites of a complex. The good correlation between basicity and nucleophilicity of a large series of zinc-bound thiolates gives confidence to the pertinence of the proton affinity as a nucleophilicity index. Work is in progress to use this index in order to predict the relative reactivity of enzymatic zinc active sites and to determine which cysteinate would be the most reactive in these sites.

Acknowledgment

We are grateful to the Centre National de la Recherche Scientifique (CNRS) and Ecole Polytechnique for financial support. We also thank the Institut du Développement et des Ressources en Informatique Scientifique (IDRIS, grant number 0543) and the Centre Informatique National de l'Enseignement Supérieur (CINES, grant number dcm2335) for computational resources.

References

- [1] G. Henkel, B. Krebs, Chem. Rev. 104 (2004) 801.
- [2] J. A. Kovacs, Chem. Rev. 104 (2004) 825.
- [3] P. V. Rao, R. H. Holm, Chem. Rev. 104 (2004) 527.
- [4] J. A. Kovacs, L. M. Brines, Acc. Chem. Res. 40 (2007) 501.
- [5] D. P. Barondeau, C. J. Kassmann, C. K. Bruns, J. A. Tainer, E. D. Getzoff, Biochemistry 43 (2004) 8038.
- [6] J. Wuerges, J. W. Lee, Y. I. Yim, H. S. Yim, S. O. Kang, K. D. Carugo, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 8569.
- [7] A. Krezel, Q. Hao, W. Maret, Arch. Biochem. Biophys. 463 (2007) 188.
- [8] D. Jantz, B. T. Amann, G. J. Gatto, Jr., J. M. Berg, Chem. Rev. 104 (2004) 789.
- [9] E. I. Solomon, S. I. Gorelsky, A. Dey, J. Comput. Chem. 27 (2006) 1415.
- [10] J. Penner-Hahn, Curr. Opin. Chem. Biol. 11 (2007) 166.
- [11] G. Parkin, Chem. Rev. 104 (2004) 699.
- [12] J. J. Wilker, S. J. Lippard, Inorg. Chem. 38 (1999) 3569.
- [13] D. Rabinovich, Struct. Bonding 120 (2006) 143.
- [14] J. J. Wilker, S. J. Lippard, J. Am. Chem. Soc. 117 (1995) 8682.

- [15] C. R. Warthen, B. S. Hammes, C. J. Carrano, D. C. Crans, *J. Biol. Inorg. Chem.* 6 (2001) 82.
- [16] M. Rombach, J. Seebacher, M. Ji, G. F. Zhang, G. S. He, M. M. Ibrahim, B. Benkmil, H. Vahrenkamp, *Inorg. Chem.* 45 (2006) 4571.
- [17] J. Notni, W. Gunther, E. Anders, *Eur. J. Inorg. Chem.* (2007) 985.
- [18] S. J. Chiou, C. G. Riordan, A. L. Rheingold, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 3695.
- [19] J. N. Smith, Z. Shirin, C. J. Carrano, *J. Am. Chem. Soc.* 125 (2003) 868.
- [20] J. N. Smith, J. T. Hoffman, Z. Shirin, C. J. Carrano, *Inorg. Chem.* 44 (2005) 2012.
- [21] M. M. Morlok, K. E. Janak, G. Zhu, D. A. Quarless, G. Parkin, *J. Am. Chem. Soc.* 127 (2005) 14039.
- [22] T. B. Phan, M. Breugst, H. Mayr, *Angew. Chem. Int. Ed.* 45 (2006) 3869, and references cited therein.
- [23] P. Jaramillo, P. Perez, R. Contreras, W. Tiznado, P. Fuentealba, *J. Phys. Chem. A* 110 (2006) 8181.
- [24] F. De Vleeschouwer, V. Van Speybroeck, M. Waroquier, P. Geerlings, F. De Proft, *Org. Lett.* 9 (2007) 2721.
- [25] P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.* 103 (2003) 1793.
- [26] J. J. Wilker, S. J. Lippard, *Inorg. Chem.* 36 (1997) 969.
- [27] D. Picot, G. Ohanessian, G. Frison, submitted for publication.
- [28] G. Frison, G. Ohanessian, submitted for publication.
- [29] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida,

T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian 03, revision C.02, Gaussian, Inc., Wallingford CT, 2003.

[30] A. J. Wachters, *J. Chem. Phys.* 52 (1970) 1033.

[31] L. A. Lajohn, P. A. Christiansen, R. B. Ross, T. Atashroo, W. C. Ermler, *J. Chem. Phys.* 87 (1987) 2812.

[32] K. A. Peterson, B. C. Shepler, D. Figgen, H. Stoll, *J. Phys. Chem. A* 110 (2006) 13877.

[33] G. Frison, G. Ohanessian, *J. Comput. Chem.* 29 (2008) 416.

[34] B. L. Vallee, D. S. Auld, *Acc. Chem. Res.* 26 (1993) 543.

[35] D. S. Auld, *Biometals* 14 (2001) 271.

[36] W. Maret, *Biochemistry* 43 (2004) 3301.

[37] G. Frison, A. Sevin, *J. Chem. Soc., Perkin Trans. 2* (2002) 1692.

[38] M. M. Morlok, A. Docrat, K. E. Janak, J. M. Tanski, G. Parkin, *Dalton Trans.* (2004) 3448.

[39] P. Jaramillo, P. Perez, P. Fuentealba, *J. Phys. Org. Chem.* 20 (2007) 1050.

IV. Comparaison avec la réactivité des thiols et des thiolates

La partie précédente nous a permis de sélectionner un indice de réactivité du motif zinc-thiolate et de pouvoir les comparer les uns relativement aux autres suivant leurs caractères structuraux (charge des ligands, présence d'une liaison hydrogène dirigée vers le thiolate). Cependant, nous n'avons encore actuellement aucune indication absolue sur leur nucléophilie comparativement aux autres formes de soufres tels que les thiols et les thiolates. Il a été montré que l'ensemble des ligands du zinc peut considérablement faire varier la réactivité d'un complexe biomimétique à zinc. Nous nous proposons d'évaluer maintenant l'amplitude des variations de réactivité des thiolates de ces complexes à l'aide de leur affinité protonique et de caractériser leur comportement chimique par rapport à ceux des thiols et des thiolates.

Une échelle théorique de réactivité a donc été établie pour les thiolates liés au zinc, pour lesquels nous avons calculé précédemment l'affinité protonique et dont la charge totale varie de -2 à +2 (Figure V.5). Les familles des thiols et des thiolates ont été représentées pour référence, avec respectivement, trois composés simples (NH_2SH , CH_3NHS , CH_3SNH_2) et les méthyl- et phényl-thiolates. D'autres thiolates présentant une liaison hydrogène avec une fonction amide et dont la nucléophilie est donc amoindrie, ont également été inclus.

Comme il avait été remarqué auparavant pour la barrière énergétique d'alkylation, les complexes du zinc sont rangés en classes de réactivité croissante suivant leur charge totale, avec le complexe dicationique étant le moins réactif ($\text{AP} < 100 \text{ kcal/mol}$) et les complexes dianioniques les plus nucléophiles ($\text{AP} > 350 \text{ kcal/mol}$).

Nous pouvons maintenant constater que les soufres des complexes chargés positivement présentent une affinité protonique inférieure à ceux des thiols ($\text{AP} < 170 \text{ kcal/mol}$), ce qui montre la très faible réactivité de ces motifs zinc-thiolate. Ainsi, le métal ne joue plus ici un rôle d'activateur envers le phénylthiolate. Ceci explique vraisemblablement l'absence de site actif de ce type dans les protéines assurant un rôle fonctionnel, puisque le métal ne peut catalyser la réaction enzymatique. Un résidu cystéine sous forme neutre est alors plus susceptible d'effectuer une alkylation.

Les complexes biomimétiques neutres présentent une nucléophilie légèrement supérieure à celles des thiols puisque leurs affinités protoniques sont comprises dans l'intervalle 200-250 kcal/mol. Les soufres de ces composés ont une réactivité comparable à ceux des doigts de zinc dont les sites $\text{ZnCys}_2\text{His}_2$ ont une charge totale nulle. La présence du métal permet alors de faciliter l'alkylation des thiolates liés au zinc, comparativement à celle de thiols non activés et la réactivité, encore faible, assure une stabilité conformationnelle à l'ensemble.

Les complexes monoanioniques, dont les affinités protoniques sont comprises entre 270 et 320 kcal/mol), présentent une réactivité bien supérieure à celle des composés neutres. Leur nucléophilie se rapprochent alors de celles des thiolates et est même dans certains cas, comparable à celle des thiolates impliqués dans une liaison hydrogène. Le zinc active le thiolate pour l'attaque nucléophile et le soufre présente alors un comportement proche des thiolates dissociés.

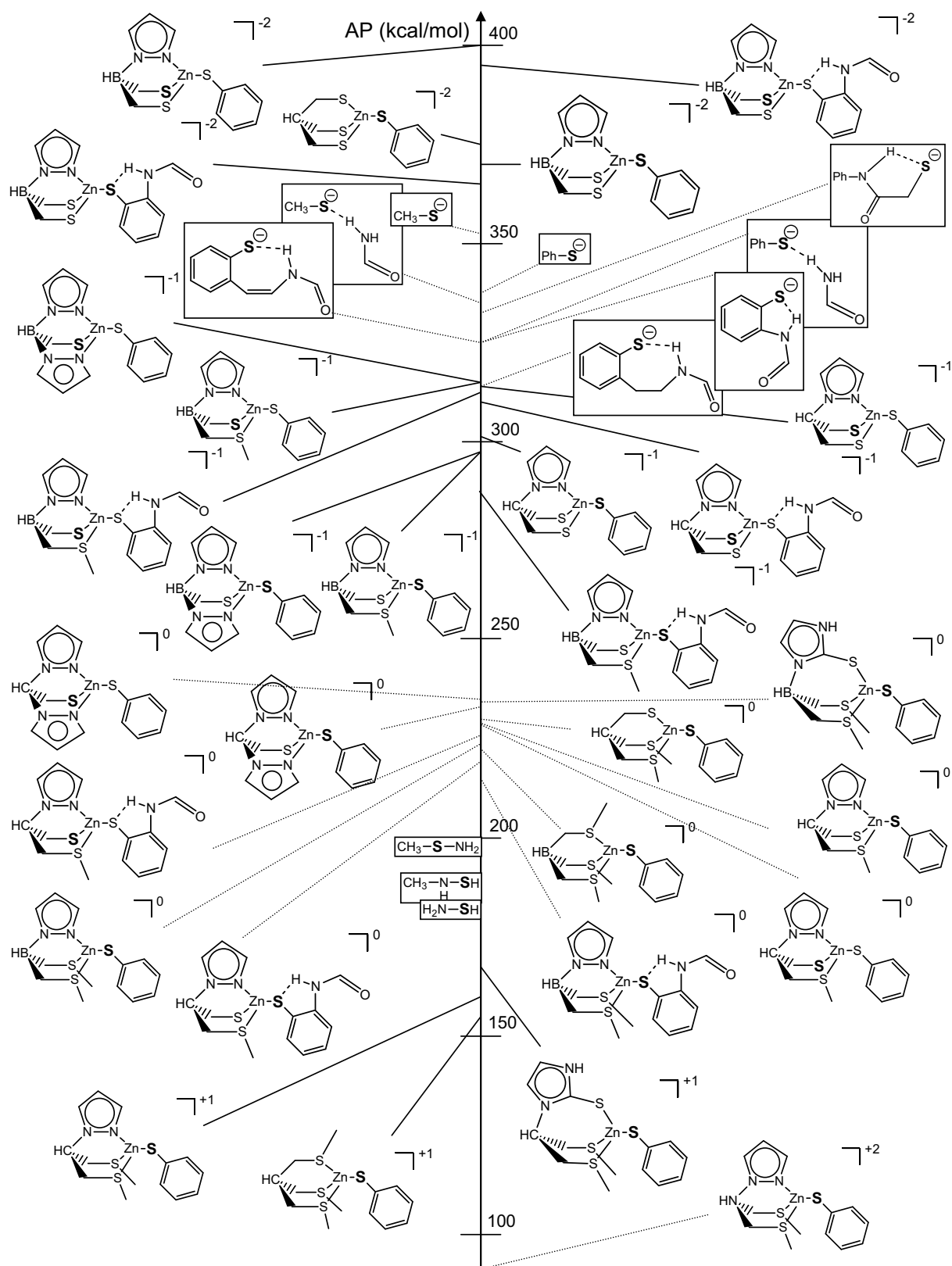


Figure V.5 : Affinités protoniques des différentes formes de soufres

Enfin, les complexes dicationiques ($AP > 350$ kcal/mol) sont supérieurs en nucléophilie aux thiolates dépourvus de liaison hydrogène tels que le méthylthiolate et le phénylthiolate. La présence du zinc et de ses ligands dans ces complexes biomimétiques permet d'exacerber la nucléophilie d'un

thiolate. Si l'on considère que les résidus cystéines sont sous forme déprotonée dans les sites actifs de type ZnCys_4 , il est à supposer que les thiolates liés au métal possèdent une réactivité effectivement intensifiée par rapport aux autres soufres. C'est pourquoi de tels motifs zinc-thiolate se retrouvent dans les protéines ayant une fonction catalytique essentielle, comme par exemple la protéine Ada qui est alkylée par l'ADN endommagé et en général dans toutes les enzymes où un thiolate « nu » ne pourrait assurer une attaque nucléophile. Ces motifs de coordination impliquant l'ion zinc(II) permettent de pallier un manque de réactivité des cystéinates.

Bibliographie

- Balawender, R.; Komorowski, L. *J. Chem. Phys.* **1998**, *109*, 5203.
- Berkowitz, M.; Parr, R. G. *J. Chem. Phys.* **1988**, *88*, 2554.
- Bonaccorsi, R.; Scrocco, E.; Tomasi, J. *J. Chem. Phys.* **1970**, *52*, 5270.
- Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
- Bronsted, J. N. *Recl. Trav. Chim.* **1923**, *42*, 718.
- Brotzel, F.; Kempf, B.; Singer, T.; Zipse, H.; Mayr, H. *Chem. Eur. J.* **2007**, *13*, 336.
- Bug, T.; Hartnagel, M.; Schlierf, C.; Mayr, H. *Chem. Eur. J.* **2003**, *9*, 4068.
- Carey, F. A.; Sundberg, R. J. *Advanced organic chemistry*, 3rd Ed., Part A, Structure and Mechanisms. Plenum Press: New York, **1990**.
- Chattaraj, P. K.; Maiti, B.; Sarkar, U. *J. Phys. Chem. A* **2003**, *107*, 4973.
- Chiou, S. J.; Riordan, C. G.; Rheingold, A. L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3695.
- Contreras, R.; Fuentealba, P.; Galvan, M.; Perez, P. *Chem. Phys. Lett.* **1999**, *304*, 405.
- Contreras, R.; Andres, J.; Safont, V. S.; Campodonico, P.; Santos, J. G. *J. Phys. Chem. A* **2003**, *107*, 5588.
- Edwards, J. O. *J. Am. Chem. Soc.* **1954**, *76*, 1540.
- Edwards, J. O. *J. Am. Chem. Soc.* **1956**, *78*, 1819.
- Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211.
- Fukui, K.; Yonezawa, T.; Shingu, H. *J. Chem. Phys.* **1952**, *20*, 722.
- Fukui, K.; Yonezawa, T.; Nagata, C.; Shingu, H. *J. Chem. Phys.* **1954**, *22*, 1433.
- Geerlings, P.; De Proft, F.; Langenaeker, W. *Chem. Rev.* **2003**, *103*, 1793.
- Hammett, L. P. *J. Am. Chem. Soc.* **1937**, *59*, 96.
- He, C.; Hus, J. C.; Sun, L. J.; Zhou, P.; Norman, D. P. G.; Dotsch, V.; Wei, H.; Gross, J. D.; Lane, W. S.; Wagner, G.; Verdine, G. L. *Mol. Cell* **2005**, *20*, 117.
- Hine, J. *Structural effects on Equilibria in Organic Chemistry*. Robert E. Krieger Publishing Company: Huntington, New-York, **1981**.
- Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129.
- Hofmann, M.; Hampel, N.; Kanzian, T.; Mayr, H. *Angew. Chem. Int. Ed.* **2004**, *43*, 5402.
- Ingold, C. K. *Recl. Trav. Chim. Pays-Bas* **1929**, *48*, 797.
- Jaramillo, P.; Perez, P.; Fuentealba, P. *J. Phys. Org. Chem.* **2007**, *20*, 1050.
- Johnson, C. D. *The Hammett equation*. Cambridge University Press: New York, **1973**.
- Juaristi, E.; Cuevas, G. *Tetrahedron* **1992**, *48*, 5019.
- Kempf, B.; Hampel, N.; Ofial, A. R.; Mayr, H. *Chem. Eur. J.* **2003**, *9*, 2209.
- Klopman, G. *J. Am. Chem. Soc.* **1968**, *90*, 2, 223.
- Leach, A. R. *Molecular Modelling. Principles and Applications*, 2nd Edition, Pearson Education, **2001**, 79.
- Legon, A. C.; Millen, D. J. *J. Am. Chem. Soc.* **1987**, *109*, 356.
- Lewis, G. N. *Valence and the structure of atoms and molecules*, The chemical catalog Co., 1923.
- Li, Y.; Evans, J. N. S. *J. Am. Chem. Soc.* **1995**, *117*, 7756.

- Lowry, T. M. *Chem. Ind.* **1923**, 42, 43.
- Mayr, H. *Angew. Chem. Int. Ed. Engl.* **1990**, 29, 1371.
- Mayr, H.; Patz, M. *Angew. Chem.* **1994**, 106, 990.
- Mayr, H.; Patz, M. *Angew. Chem. Int. Ed. Engl.* **1994**, 33, 938.
- Mayr, H.; Bug, T.; Gotta, M. F.; Hering, N.; Irrgang, B.; Janker, B.; Kempf, B.; Loos, R.; Ofial, A. R.; Remennikov, G.; Schimmel, H. *J. Am. Chem. Soc.* **2001**, 12, 9500.
- Mayr, H.; Kempf, B.; Ofial, A. R. *Acc. Chem. Res.* **2003**, 36, 66.
- Melin, J.; Aparicio, F.; Subramanian, V.; Galvan, M.; Chattaraj, P. K. *J. Phys. Chem. A* **2004**, 108, 2487.
- Meneses, L.; Araya, A.; Pilaquinga, F.; Contreras, R.; Fuentealba, P. *Chem. Phys. Lett.* **2007**, 446, 170.
- Morlok, M. M.; Janak, K. E.; Zhu, G.; Quarless, D. A.; Parkin, G. *J. Am. Chem. Soc.* **2005**, 127, 14039.
- Mulliken, R. S. *J. Chem. Phys.* **1955**, 10, 1833.
- Mulliken, R. S. *J. Chem. Phys.* **1955**, 12, 2338.
- Nam, P. C.; Nguyen, M. T. *J. Phys. Chem. A* **2006**, 110, 10904.
- Parr, R. G.; Pearson, R. G. *J. Am. Chem. Soc.* **1983**, 105, 7512.
- Parr, R. G.; Yang, W. T. *J. Am. Chem. Soc.* **1984**, 106, 4049.
- Parr, R. G.; Szentpaly, L. V.; Liu, S. *J. Am. Chem. Soc.* **1999**, 121, 1922.
- Pearson, R. G.; Sobel, H.; Songstad, J. *J. Am. Chem. Soc.* **1968**, 90, 319.
- Perdew, J. P.; Parr, R. G.; Leby, M.; Balduz, J. L. *Phys. Rev. Lett.* **1982**, 49, 1691.
- Phan, T. B.; Breugst, M.; Mayr, H. *Angew. Chem. Int. Ed.* **2006**, 45, 3869.
- Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, 78, 4066.
- Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, 83, 735.
- Ritchie, C. D. *Acc. Chem. Res.* **1972**, 5, 348.
- Ritchie, C. D. *Can. J. Chem.* **1986**, 64, 2239.
- Rives, A. B.; Weinhold, F. *Int. J. Quantum Chem. Symp.* **1980**, 14, 201.
- Rombach, M.; Seebacher, J.; Ji, M.; Zhang, G. F.; He, G. S.; Ibrahim, M. M.; Benkmil, B.; Vahrenkamp, H. *Inorg. Chem.* **2006**, 45, 4571.
- Roy, R. K.; Krishnamurti, S.; Geerlings, P.; Pal, S. *J. Phys. Chem. A* **1998**, 102, 3746.
- Roy, R. K.; Pal, S.; Hirao, K. *J. Chem. Phys.* **1999**, 110, 8236.
- Roy, R. K.; Hirao, K.; Pal, S. *J. Chem. Phys.* **2000**, 113, 1372.
- Roy, R. K.; Tajima, N.; Hirao, K. *J. Phys. Chem.* **2001**, 105, 211.
- Roy, R. K. *J. Phys. Chem. A* **2003**, 107, 397 et 10428.
- Roy, R. K. *J. Phys. Chem. A* **2004**, 108, 4934.
- Roy, R. K.; Usha, V.; Paulovic, J.; Hirao, K. *J. Phys. Chem. A*, **2005**, 109, 4601.
- Roy, R. K.; Parthasarathi, R.; Padmanabhan, J.; Sarkar, U.; Subramanian, V.; Chattaraj, P. K. *J. Phys. Chem. A* **2006**, 110, 1084.
- Smith, J. N.; Hoffman, J. T.; Shirin, Z.; Carrano, C. J. *Inorg. Chem.* **2005**, 44, 2012.
- Swain, C. G.; Scott, C. B. *J. Am. Chem. Soc.* **1953**, 75, 141.
- Taft, R. W.; Topsom, R. D. *Prog. Phys. Org. Chem.* **1987**, 16, 1.

- Takinowaki, H.; Matsuda, Y.; Yoshida, T.; Kobayashi, Y.; Ohkubo, T. *Protein Sci.* **2006**, *15*, 487.
- Thanikaivelan, P.; Padmanabhan, J.; Subramanian, V.; Ramasami, T. *Theor. Chem. Acc.* **2002**, *107*, 326.
- Tumanov, V. V.; Tishkov, A. A.; Mayr, H. *Angew. Chem. Int. Ed.* **2007**, *46*, 3563.
- Waisser, K.; Polasek, M.; Nemec, I.; Exner, O. *J. Phys. Org. Chem.* **2000**, *13*, 127.
- Wilker, J. J.; Lippard, S. J. *J. Am. Chem. Soc.* **1995**, *117*, 33, 8682.
- Wilker, J. J.; Lippard, S. J. *Inorg. Chem.* **1997**, *36*, 969.
- Yang, W. T.; Paar, R. G. *Proc. Natl. Acad. Sci. USA*, **1985**, *82*, 6723.
- Yang, W. T. ; Mortier, W. J. *J. Am. Chem. Soc.* **1986**, *108*, 5708.

ANNEXE**Script Perl utilisé pour le calcul des coefficients de Fukui**

```
#!/usr/bin/perl -s

use strict;
use Spreadsheet::ParseExcel;

my $oExcel = new Spreadsheet::ParseExcel;
die "You must provide a filename to $0 to be parsed as an Excel file" unless @ARGV;
my $oBook = $oExcel->Parse($ARGV[0]);

my ($iR, $new_iR, $newiR, $iC, $new_iC, $newiC, $oWkS, $oWkC, $entre, $i, $l);
my ($NAtoms, $k_atom, $OM, $xOA, $nbk_OA, $nbtot_OA, $prod, $totprod, $om, $dest);
my (@contenu_cell, @first_OA, @k, @valeur_coeff, @j, @valeur_density, @valeur_mulpop, @OA, @s,
    @f, @fukui);
my ($atom, $a, $b, $c, $d, $e);

print ("Opening the file Fukui_Coeff.txt\n");
unless (open(OUTPUT, ">Fukui_Coeff.txt")) {die ("Cannot open file Fukui_Coeff.txt\n")}

#####
##  INITIALISATION DES ATOMES ET OA
#####

for( my $oWkS = 3 )
    { $oWkS = $oBook->{Worksheet}[3];
      $nbtot_OA = $oWkS->{MaxRow}+1;
      for(my $iR = $oWkS->{MinRow} ;
          defined $oWkS->{MaxRow} && $iR <= $oWkS->{MaxRow} ; $iR++) {

          for(my $iC = $oWkS->{MinCol} ;
              defined $oWkS->{MaxCol} && $iC <= $oWkS->{MaxCol} ; $iC++) {
              $oWkC = $oWkS->{Cells}[$iR][$iC];
              $contenu_cell[$iR][$iC] = $oWkC->{Val};
          }
          if ( $contenu_cell[$iR][$iC] == $contenu_cell[$iR][1] && $contenu_cell[$iR][$iC] =~ /[0-9]/)
              {
                  $k_atom = $contenu_cell[$iR][$iC];
                  $first_OA[$k_atom] = $contenu_cell[$iR][0];
              }
          }
      $NAtoms = $k_atom;
    }

print ("Writing initial data in the file Fukui_Coeff.txt\n");

### ECRITURE DES DONNEES INITIALES DANS LE FICHIER OUTPUT

select (OUTPUT);
print ("Le nombre total d'atomes de cette molécule est: $NAtoms\n");
print ("Le nombre total d'orbitales atomiques de cette molécule est: $nbtot_OA\n");
for ($k_atom = 1; $k_atom <= $NAtoms; $k_atom++){
#       print ("La premiere OA de l'atome $k_atom est: $first_OA[$k_atom]\n");
    }
print ("\n");

print (" ");
for (my $k_atom=1; $k_atom <= $NAtoms; $k_atom++){
    print ("      $k_atom      ");
}
print ("\n");

close (OUTPUT);
print ("Assigning MO coefficients from the excel file\n");

#####
#  INITIALISATION DES VALEURS_COEFF
#####

$k[0]=$nbtot_OA;
for ($l =1; $l < ($nbtot_OA/5); $l++){
    $k[$l] = $k[0] + $l*($k[0]+3);
}
```

```
#           print (" $k[$l]\n");
}

for( my $oWkS = 0 )
{ $oWkS = $oBook->{Worksheet}[0];

  for(my $iR = $oWkS->{MinRow} ;
    defined $oWkS->{MaxRow} && $iR < $k[0] ; $iR++) {
    for(my $iC = $oWkS->{MinCol} ;
      defined $oWkS->{MaxCol} && $iC <= $oWkS->{MaxCol} ; $iC++) {
      $oWkC = $oWkS->{Cells}[$iR][$iC];
      $valeur_coefff[$iR][$iC] = $oWkC->{Val};
#       print (" $iR  $iC  $valeur_coefff[$iR][$iC]\n");
    }
  }

  for ($l = 1; $l < ($nbtot_OA/5); $l++){
    for(my $iR = $k[$l-1] +3;
      defined $oWkS->{MaxCol} && $iR < $k[$l]; $iR++){

      for(my $iC = $oWkS->{MinCol} ;
        defined $oWkS->{MaxCol} && $iC <= $oWkS->{MaxCol} ; $iC++) {

        my $newiR = $iR - $l*($k[0]+3);
        my $newiC = $iC + $l*5;

        if ($newiR == $nbtot_OA-1 && $newiC==$nbtot_OA){last;}

        $oWkC = $oWkS->{Cells}[$iR][$iC];
        $valeur_coefff[$newiR][$newiC] = $oWkC->{Val};

#         print (" $iR $iC  $newiR $newiC  $valeur_coefff[$newiR][$newiC]\n");
      }
    }
  }
# print (" $valeur_coefff[0][4]\n");

print ("Assigning density coefficients from the excel file\n");

#####
##  INITIALISATION DES VALEURS_DENSITY
#####

$j[0]= $nbtot_OA;
for ($i = 1; $i < ($nbtot_OA/5); $i++){
  $j[$i] = $j[0] + 1 + $j[$i-1] - $i*5;}

for( my $oWkS = 1 ) { $oWkS = $oBook->{Worksheet}[1];

  for(my $iR = $oWkS->{MinRow} ;
    defined $oWkS->{MaxRow} && $iR <  $j[0]; $iR++) {

    for(my $iC = $oWkS->{MinCol} ;
      defined $oWkS->{MaxCol} && $iC <= $oWkS->{MaxCol} ; $iC++) {

      if ($iR >= $iC){
        $oWkC = $oWkS->{Cells}[$iR][$iC];
        $valeur_density[$iR][$iC] = $oWkC->{Val};}
        $valeur_density[$iC][$iR] = $valeur_density[$iR][$iC];
# print (" $iR  $iC  $valeur_density[$iR][$iC]\n");
      }
    }

    for ($i = 1; $i < ($nbtot_OA/5); $i++){
      for(my $iR = $j[$i-1] +1;
        defined $oWkS->{MaxCol} && $iR < $j[$i]; $iR++){

        for(my $iC = $oWkS->{MinCol} ;
          defined $oWkS->{MaxCol} && $iC <= $oWkS->{MaxCol} ; $iC++) {

          my $new_iR = $iR -1 - $j[$i-1] +5*$i;
```

```

my $new_iC = $iC + 5*$i ;

if ($new_iR == $nbtot_OA-1 && $new_iC==$nbtot_OA){last;}

if ($new_iR >= $new_iC){
    $oWkC = $oWKS->{Cells}[$iR][$iC];
    $valeur_density[$new_iR][$new_iC] = $oWkC->{Val};
    $valeur_density[$new_iC][$new_iR] =
}
}

}

#print (" $valeur_density[0][7]\n");

print ("Assigning Mulliken population coefficients from the excel file\n");

#####
### INITIALISATION DES VALEURS_MULPOP
#####

$j[0]= $nbtot_OA;
for ($i = 1; $i < ($nbtot_OA/5); $i++){
    $j[$i] = $j[0] + 1 + $j[$i-1] - $i*5;}

for( my $oWkS = 2 ) {$oWkS = $oBook->{Worksheet}[2];

    for(my $iR = $oWkS->{MinRow} ;
        defined $oWkS->{MaxRow} && $iR < $j[0]; $iR++) {

        for(my $iC = $oWkS->{MinCol} ;
            defined $oWkS->{MaxCol} && $iC <= $oWkS->{MaxCol} ; $iC++) {

            if ($iR >= $iC){
                $oWkC = $oWKS->{Cells}[$iR][$iC];
                $valeur_mulpop[$iR][$iC] = $oWkC->{Val};
                $valeur_mulpop[$iC][$iR] = $valeur_mulpop[$iR][$iC];
                #print (" $iR $iC $valeur_mulpop[$iR][$iC]\n");
            }
        }

        for ($i = 1; $i < ($nbtot_OA/5); $i++){
            for(my $iR = $j[$i-1] +1;
                defined $oWkS->{MaxCol} && $iR < $j[$i]; $iR++){

                for(my $iC = $oWkS->{MinCol} ;
                    defined $oWkS->{MaxCol} && $iC <= $oWkS->{MaxCol} ;
                    $iC++) {

                        my $new_iR = $iR -1 - $j[$i-1] +5*$i;
                        my $new_iC = $iC + 5*$i ;

                        if ($new_iR == $nbtot_OA-1 &&
                            $new_iC==$nbtot_OA){last;}

                        if ($new_iR >= $new_iC){
                            $oWkC = $oWKS->{Cells}[$iR][$iC];
                            $valeur_mulpop[$new_iR][$new_iC] = $oWkC->{Val};
                            $valeur_mulpop[$new_iC][$new_iR] =
                        }

                        #print (" $new_iR $new_iC
                        $valeur_mulpop[$new_iR][$new_iC]\n");
                    }
                }
            }
        }

    }

    #print (" $valeur_mulpop[0][7]\n");

    print ("Calculating MO overlap coefficients from the excel file\n");

    #####
    ##### CALCUL DES OVERLAP
    #####

```

```

for (my $k_atom = 1; $k_atom < $NAtoms; $k_atom++){
#print ("L'atome considéré est $k_atom\n");

    for (my $OA = $first_OA[$k_atom] -1 ; $OA < $first_OA[$k_atom+1]-1; $OA++){
#print ("L'OA considéré de l'atome $k_atom est $OA\n");

        for (my $xOA = 0; $xOA < $nbtot_OA; $xOA++){
#print ("L'xOA considéré est :$xOA\n");

            if ($valeur_density[$OA][$xOA] == 0 || $valeur_mulpop[$OA][$xOA] == 0){
                $s[$OA][$xOA]= 0;
            } else {
                $s[$OA][$xOA]= $valeur_mulpop[$OA][$xOA] / $valeur_density[$OA][$xOA];
                #print (" Le coefficeint d'overlap est:  $s[$OA][$xOA]\n");
            }
        }
    }
}

for (my $k_atom = $NAtoms){
    for (my $OA = $first_OA[$k_atom] -1 ; $OA < $nbtot_OA; $OA++) {
#print ("L'OA considéré de l'atome $k_atom est $OA\n");

        for (my $xOA = 0; $xOA < $nbtot_OA; $xOA++){
#print ("L'xOA considéré est :$xOA\n");

            if ($valeur_density[$OA][$xOA] == 0 || $valeur_mulpop[$OA][$xOA] == 0){
                $s[$OA][$xOA]= 0;
            } else {
                $s[$OA][$xOA]= $valeur_mulpop[$OA][$xOA] / $valeur_density[$OA][$xOA];
                #print (" Le coefficeint d'overlap est:  $s[$OA][$xOA]\n");
            }
        }
    }
}

#print (" Le coefficeint d'overlap est:  $s[0][7]\n");

print ("Calculating the fukui coefficients\n");

#####
#####  CALCUL DES COEFFICIENTS DE FUKUI
#####

$fukui[$k_atom-1][$OM]=0;

for (my $k_atom = 1; $k_atom < $NAtoms; $k_atom++){
#print ("L'atome considéré est $k_atom\n");

    for (my $OM = 0; $OM < $nbtot_OA ; $OM++){
#print ("L'OM considérée est $OM\n");

        for (my $OA = $first_OA[$k_atom] -1; $OA < $first_OA[$k_atom+1]-1; $OA++){
#print ("L'OA considéré de l'atome $k_atom est $OA\n");

            for (my $xOA = 0; $xOA < $nbtot_OA; $xOA++){
#print ("L'xOA considéré est :$xOA\n");

                $f[$OA][$xOA]=
                $valeur_coeff[$OA][$OM]*$valeur_coeff[$xOA][$OM]*$s[$OA][$xOA];
                $fukui[$k_atom-1][$OM]+= $f[$OA][$xOA];
            }
        }
    }
}

for (my $k_atom = $NAtoms){
    for (my $OM = 0; $OM < $nbtot_OA ; $OM++){
#print ("L'OM considérée est $OM\n");

        for (my $OA = $first_OA[$k_atom] -1 ; $OA < $nbtot_OA; $OA++) {
#print ("L'OA considéré de l'atome $k_atom est $OA\n");

            for (my $xOA = 0; $xOA < $nbtot_OA; $xOA++){
#print ("L'xOA considéré est :$xOA\n");

```

```
                $f[$OA][$xOA]=
$valeur_coef[$OA][$OM]*$valeur_coef[$xOA][$OM]*$s[$OA][$xOA];
                $fukui[$k_atom-1][$OM]+=$f[$OA][$xOA];
            }
        }
        $om= $OM + 1;
        #print ("Le coefficient de Fukui pour l'atome $k_atom sur l'orbitale $om est:
$fukui[$k_atom-1][$OM]\n");
    }
}

print ("Writing fukui coefficients in text file\n");

#####
##### ECRITURE DES COEFF DE FUKUI DANS UN FICHIER TEXTE
#####

unless (open(OUTPUT, ">>Fukui_Coeff.txt")) {die ("Cannot open file Fukui_Coeff.txt\n")}
select (OUTPUT);

for (my $om =1; $om <= $nbtot_OA; $om++){
    print (" $om ");

    for (my $k_atom = 1; $k_atom <= $NAtoms; $k_atom++){
        $atom= $k_atom;
        $a=$fukui[$k_atom-1][$om-1];
        printf (" % .4f ", $a );}
    print (" \n");
}

close (OUTPUT);
```

Chapitre VI

Modélisation d'un site actif à zinc

I. Introduction

A. Importance de l'environnement du zinc dans les sites actifs

Dans les premiers chapitres de cette étude, nous avons mis en évidence l'influence de la nature des ligands du zinc sur le mécanisme d'alkylation des thiolates des complexes biomimétiques. Ainsi, si le zinc est entouré de groupes anioniques, il aura tendance à se dissocier du thiolate avant réaction, alors que si le complexe a une charge neutre totale, l'alkylation du thiolate procédera via un mécanisme associatif.

De plus, la littérature suppose que dans les protéines, toutes les cystéines d'un site actif à zinc sont déprotonées (Bulaj 1998). Dans ce cas, les complexes **3** et **4** étudiées dans le chapitre III donnent directement une idée de la réactivité du site métallique. La figure VI.1 nous montre la prédominance des mécanismes dissociatifs et associatifs type S_N2 dans des sites « nus » à zinc chargés respectivement -1 et -2. Alors que pour les complexes neutres et cationiques, le mécanisme est clairement dissociatif, les sites protéiques pourront présenter des mécanismes associatifs ou dissociatifs suivant la charge globale de leur environnement.

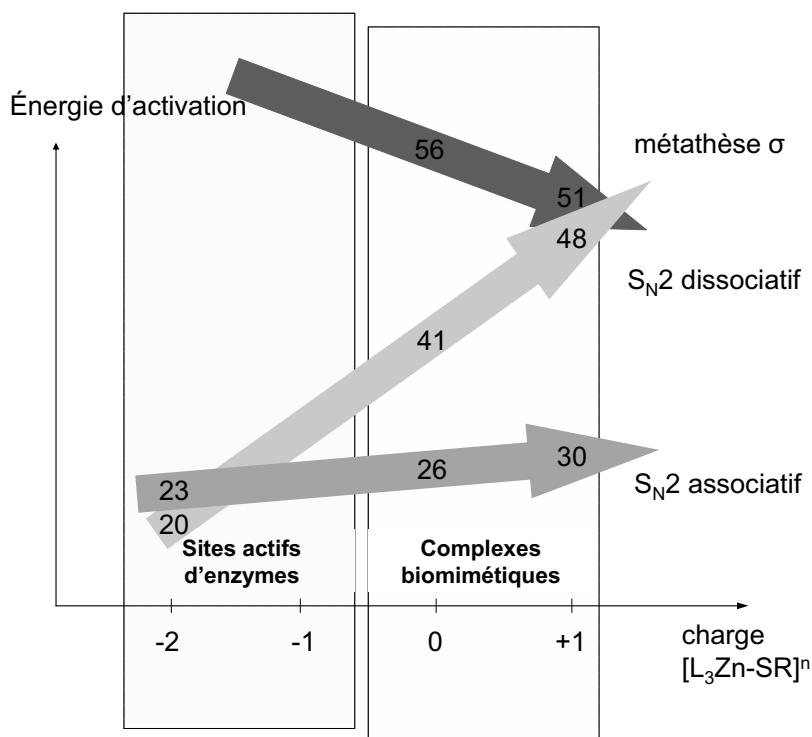


Figure VI.1 : Energie libre d'activation en solution (kcal/mol) des complexes biomimétiques **1**, **3** et **1^c** (Chapitre III)

Si les mécanismes d'alkylation du thiolate varient avec l'environnement direct du zinc, il en est de même pour la nucléophilie du site. Ce phénomène a été démontré sur les doigts de zinc, en expliquant comment les sites à quatre cystéines étaient supposés être intrinsèquement plus nucléophiles que les sites à trois cystéines et enfin à deux cystéines dans un rapport 1000 :100 :1

(Wilker 1995, Maynard 2001). L'environnement proche du zinc a donc aussi un impact important sur la nucléophilie des complexes organométalliques.

De plus, l'écrantage électrostatique et l'effet stérique de l'environnement autour du métal ont été calculés et corrélient avec cet ordre de réactivité, puisque les sites anioniques les plus nucléophiles se trouvent être les plus compacts et ceux possédant le plus de liaisons hydrogènes (Maynard 2001). Pour les structures protéiques, la réactivité des thiolates est donc aussi à mettre en relation avec l'influence de l'environnement plus distant. Ainsi, le site à quatre cystéines, le plus nucléophile intrinsèquement, peut voir sa réactivité être modulée fortement par la présence de résidus basiques ou donneurs de liaisons hydrogènes (cf Chapitre I Partie III.C.2)

Nous avons montré dans le chapitre V que la réactivité du motif zinc-thiolate peut être mesurée via l'affinité protonique correspondante. Ainsi, les valeurs des barrières énergétiques des alkylations qui reflètent la nucléophilie du soufre n'ont plus besoin d'être calculées. Ceci est beaucoup plus rapide notamment pour les systèmes de grande taille comme les sites actifs protéiques. Le but de cette partie est de modéliser un site actif à zinc et d'utiliser les calculs d'affinités protoniques pour déterminer la nucléophilie de ses cystéinates. Nous pourrons ainsi mieux appréhender l'influence de l'environnement sur la réactivité des différents thiolates, notamment la rigidité du squelette protéique, la présence de liaisons hydrogènes et de résidus basiques.

Le choix du site à modéliser s'est porté sur l'un des deux sites structuraux à zinc de la nucléocapside p7 du virus VIH. Cette protéine possède en effet deux sites métalliques appartenant à la famille des doigts de zinc, dont nous présenterons tout d'abord les caractéristiques structurales et fonctionnelles.

La compréhension du mécanisme d'alkylation de cette nucléocapside a des perspectives thérapeutiques intéressantes puisque la protéine est impliquée dans la réplication du virus du SIDA. De plus, le doigt de zinc que nous modéliserons, possède l'avantage de présenter plusieurs thiolates dans le même site, donc plusieurs sites d'alkylation possibles. La sélectivité de son alkylation est d'ailleurs le sujet de nombreuses études, théoriques et expérimentales, ainsi que son mécanisme réactionnel.

B. Les doigts de zinc de la nucléocapside p7 du virus VIH

1. La famille des doigts de zinc

Parmi les protéines possédant des sites structuraux à zinc se trouve la large famille des doigts de zinc. Il s'agit de petites unités peptidiques liées à un atome de zinc qui jouent un rôle primordial dans la régulation des gènes en promouvant la reconnaissance et l'attachement spécifique à des séquences d'acide nucléique (Berg 1997, Choo 1997). Tous les doigts de zinc contiennent un site constitué de résidus histidine et cystéine coordonnées de façon tétraédrique à un ion zinc, qui permet de s'accrocher de façon sélective à l'ADN. Ils sont nécessaires à la stabilité et à l'arrangement structural

des protéines (structure secondaire) et de nombreuses expériences ont montré que la coordination au zinc était primordiale pour la fonction de la protéine (Berg 1996, Choo 1997, Wolfe 1999).

Le motif général accepté pour les doigts de zinc est $X_i-\Omega-X_j-\Omega-X_k-\Omega-X_l-\Omega-X_m$ où Ω sont des ligands du zinc (His ou Cys) et X_i sont des segments peptidiques de longueur i . Les résidus conservés d'une espèce à l'autre et situés dans les boucles X_i contribuent également au repliement du doigt de zinc et à la spécificité de la ligature à l'ADN. Au moins dix classes fonctionnelles de doigts de zinc, distinctes de par leurs ligands et boucles, ont été caractérisées (Klug 1994).

2. La protéine nucléocapside p7

a. Généralités

Toutes les protéines nucléocapsides rétrovirales, formées à partir du clivage de la protéine précurseur *gag*, possèdent une ou deux copies de séquences conservées C-X₂-C-X₄-H-X₄-C (C=Cystéine, H=Histidine et X=acide aminé variable) (Henderson 1981, Copeland 1984). Ces domaines forment des doigts de zinc qui se lient au métal avec une stoechiométrie 1 :1 et une très forte affinité ($K_a \sim 10^{14} M^{-1}$) (Mély 1991). Les nucléocapsides sont essentiels pour la transcription du génome de l'ARN, ainsi que pour le développement de l'infection (Roques 1997).

La nucléocapside p7, un des constituants de la protéine *gag*, contient elle-même deux doigts de zinc de type rétroviral (South 1990, Summers 1992) (Figure VI.2). Elle est impliquée dans l'assemblage des protéines virales et leur maturation, ainsi que dans l'encapsidation spécifique de l'ARN viral, un processus qui nécessite la reconnaissance d'une séquence spécifique de l'ARN (Lever 1989, Aldovini 1990).



Figure VI.2 : Structure d'un doigt de zinc (1mfs.pdb) (Lee 1998)

Les thérapies médicamenteuses actuelles contre le SIDA sont actuellement constituées d'inhibiteurs puissants de la transcriptase reverse et des protéases de la protéine *gag*. Alors qu'elles retardent le développement de l'infection en réduisant significativement la réplication du virus VIH, d'autres traitements sont requis pour éradiquer complètement la maladie. De plus, les effets indésirables sont nombreux et des lignées de virus émergent, résistantes aux thérapies actuelles (Cohen 1997, Tozser 1997). C'est pourquoi le développement d'agents antiviraux affectant des cibles uniques et ne démontrant pas de résistance avec les médicaments existants est nécessaire.

La protéine NCp7 peut réagir avec des agents alkylants et mener à l'inhibition du HIV (Tummino 1996, Loo 1996, Hathout 1996, Maynard 1998, Huang 1998). Les agents antiviraux agissent au niveau du doigt de zinc, provoquant la perte de la structure native de la protéine et de sa capacité à se lier aux brins d'acide nucléique et par conséquent, la disruption de la réplication du virus. De plus, la forte conservation de séquence de Ncp7, sa forte résistance aux mutations, et le rôle critique qu'elle joue dans de multiples phases du cycle de réplication du VIH en font une cible particulièrement intéressante pour le développement de nouveaux traitements thérapeutiques (Jenkins 2007).

b. Structure

La protéine nucléocapside ncp7 est constituée de 55 acides aminés et présente un poids moléculaire de 6 430 Da (Henderson 1992). Sa petite taille a permis sa synthèse et l'analyse structurale par de nombreuses méthodes telles que la RMN et les spectroscopies CD et EXAFS. Elle est ainsi la troisième protéine du VIH à être caractérisée structuralement et l'unique protéine pour laquelle une structure a été déterminée à partir d'un échantillon directement isolé de particules du VIH (Summers 1992).

La structure de NCp7 a été déterminée indépendamment par trois groupes de recherche (Omichinski 1991, Summers 1992, Morellet 1992). Les structures individuelles des doigts de zinc y sont similaires et comportent chacun 14 résidus. Les ions divalents zinc ont une coordination tétravalente avec une histidine et trois cystéinates. Dans le doigt de zinc du C-terminus qui sera modélisé lors de cette étude, les résidus métalliques concernés sont les cystéines Cys 36, Cys 39 et Cys 49 et l'histidine 44.

Des contradictions sont apparues quant à la dynamique de ces domaines à zinc, initialement décrits comme indépendants structuralement, pour finalement s'accorder sur l'existence d'interactions transitoires entre les deux doigts de zinc (Lee 1998). Ceux-ci sont stabilisés par un réseau interne extensif de liaisons hydrogène NH-O et NH-S (Figure VI.3) et par la conservation de petits résidus glycine à des positions suivant et précédant immédiatement la deuxième cystéine et l'histidine (South 1990).

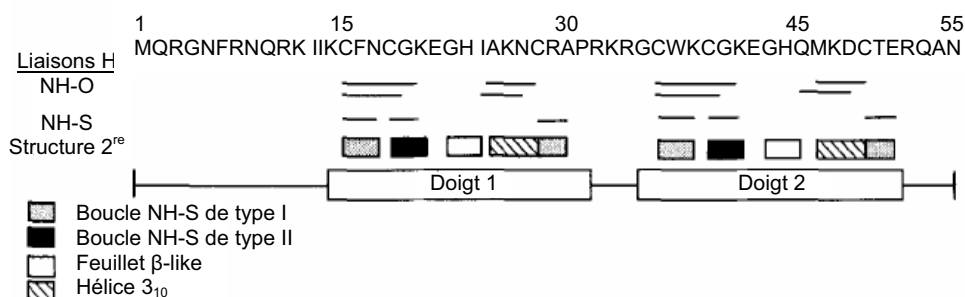


Figure VI.3 : Séquence avec liaisons hydrogène (1aaf.pdb) (Summers 1992)

Les doigts de zinc sont, par ailleurs, composés de résidus hydrophobes conservés et localisés sur leurs surfaces et séparés l'un de l'autre par un « linker » composé d'acides aminés basiques, ceci

permettant l'attachement du squelette phosphodiester des acides nucléiques. Alors que les domaines à zinc sont stabilisés par le réseau de liaisons hydrogènes et donc structurés (Figure VI.3), les parties constituées du N-, du C-terminus ainsi que du « linker » sont plus flexibles et existent dans plusieurs conformations aléatoires (Summers 1992).

c. Réactivité

La protéine NCp7 peut réagir avec des agents alkylants. Il a été montré expérimentalement que les composés antiviraux « attaquent » les doigts de zinc et qu'ils causent leur destruction et l'éjection du métal, provoquant l'inhibition du virus HIV (Rice 1993 et 1995). Par ailleurs, les agents antiviraux agissent sélectivement sur les doigts de zinc rétroviraux de la nucléocapside, sans affecter les autres protéines cellulaires à zinc, telles que la polymérase poly(ADP-ribose) et les facteurs de transcription Sp1 et GATA-1 (Huang 1998).

Mécanisme de l'alkylation

Les agents antiviraux possèdent généralement une fonction disulfure ou un soufre déficient en électrons. Une exception concerne l'agent antiviral ADA (azocarbonamide) dont l'atome d'azote central protoné peut aussi jouer le rôle de site électrophile (Topol 2001). Cette caractéristique commune a donc suggéré qu'une première étape dans l'expulsion du zinc est l'attaque nucléophile de ces soufres par une cystéinate, dont la nucléophilie est exacerbée par sa coordination au zinc.

Les intermédiaires et produits de la réaction entre la protéine p7 et des agents alkylants ont été piégés par spectrométrie de masse et analysés par HPLC (Hathout 1996) ou RMN (Loo 1996). La caractérisation des produits de la réaction a montré que ceux-ci possédaient des liaisons disulfures intramoléculaires. Il peut en être conclu que l'éjection du zinc est accompagnée de la formation de complexes covalents du soufre.

Un premier mécanisme a été proposé dans lequel les cystéinates des deux doigts de zinc sont impliquées, puisqu'il a été montré que ceux-ci étaient alkylés successivement (Hathout 1996). Cependant, il n'y a pas de preuve formelle concernant la participation du doigt de zinc du côté N-terminal dans l'alkylation se produisant du côté C-terminal.

Un mécanisme, plus vraisemblable, consiste en l'attaque nucléophile de la cystéinate sur le soufre électrophile de l'agent antiviral, suivie par l'éjection du zinc avec formation d'un pont disulfure intramoléculaire et/ou réarrangement (Loo 1996).

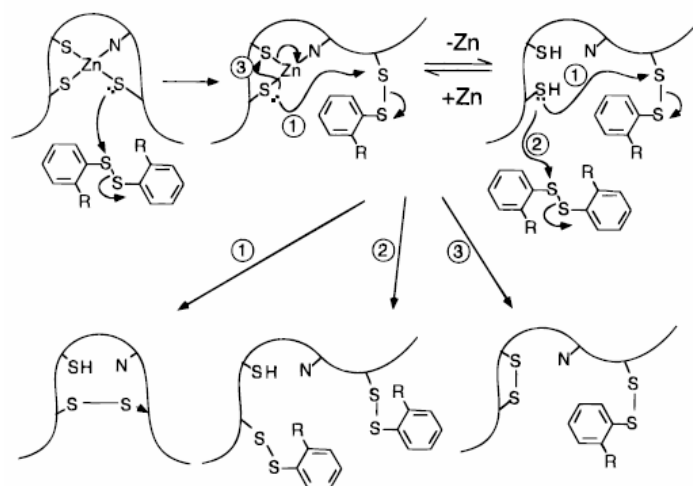


Figure VI.4 : Mécanismes de réactions possibles pour l'alkylation de Ncp7 par un agent électrophile dithio(benzamide) (Loo 1996)

Différents mécanismes sont alors possibles selon que la formation du groupe disulfure soit concertée (1) ou successive (2) à l'expulsion du métal. L'observation des graphes obtenus par spectrométrie de masse supporterait le mécanisme (2) du fait de la faible abondance du pic correspondant aux produits [Ncp7 + 2Zn + antiviral].

Cependant, une étude du mécanisme par les méthodes DFT (Topol 2000) indique que les agents alkylants provoqueraient la formation du pont disulfure avant l'éjection du zinc. En effet, la formation de la liaison covalente S-S au sein du site tétraédrique s'accompagne d'un changement de géométrie de celui-ci : le fragment protéique restant ZnS_2N a alors une configuration plane qui facilite l'expulsion du métal (3). Celle-ci peut être provoquée par l'arrivée d'une deuxième molécule « extractante ». Ce mécanisme est également applicable aux agents antiviraux ne possédant pas de groupe disulfure, telle que la molécule ADA, qui détruit similairement la sphère de coordination du zinc, par l'élimination du ligand thiolate nucléophile lié au zinc.

Sélectivités de l'alkylation

- Sélectivité entre les deux doigts de zinc

Cette découverte de la réactivité de la nucléocapside NCp7 s'est accompagnée d'une problématique sur la sélectivité entre les deux doigts de zinc. Les cinétiques de l'éjection du métal, provoquées par la réaction de la nucléocapside avec des benzamides disulfides, ont en effet été mesurées par une sonde fluorescente spécifique au zinc et par la fluorescence du tryptophane. Il a été ainsi montré que le C-terminus expulsait le zinc sept fois plus vite que le N-terminus (Tummino 1996). Cette sélectivité envers l'un des deux sites à zinc est confirmée par des études utilisant les spectrométries de masse et RMN (Hathout 1996, Loo 1996). L'alkylation des doigts de zinc par des agents antiviraux a ainsi été suivie cinétiquement et a montré que les deux doigts de zinc étaient alkylés, mais que le côté C-terminal réagissait en premier.

Malgré leur similarité structurale (South 1991), les deux doigts de zinc du nucléocapside Ncp7 de la protéine HIV présentent ainsi une réactivité différente envers l'alkylation. Il est tout d'abord suggéré,

après avoir mesuré les constantes d'affinité du zinc avec les domaines spécifiques, que cette différence de réactivité vient d'une plus faible chélation du métal du côté C-terminal (Tummino 1996). De plus, une modélisation des deux sites métalliques de la nucléocapside indique que celui du C-terminus est plus accessible pour l'approche d'inhibiteurs (Maynard 1998). L'effet électronique des sites a également été considéré dans cette étude, en plus de l'effet stérique. Ainsi, les constantes de l'alkylation de la nucléocapside avec des agents électrophiles ont été déterminées par fluorescence du tryptophane Trp 37. Elles corrélaient avec un indice de réactivité défini par le rapport entre l'électronégativité et la dureté (χ^2/η), une quantité qui relie la capacité d'un électrophile à promouvoir une interaction molle, c'est-à-dire une réaction covalente. Le doigt de zinc de la partie C-terminale est donc décrit comme possédant une densité électronique plus riche que celui du côté N-terminal.

Ceci est cohérent avec une autre rationalisation de la sélectivité, considérant le nombre de liaisons hydrogènes autour des sites à zinc. Le nombre de ces interactions impliquant les cystéinates liées au zinc est en effet moins important dans le site plus réactif (Hathout 1996). Ainsi, les écrantages électrostatique et stérique, tels qu'ils ont été définis par Maynard et Covell (Maynard 2001) sont moindres dans le C-terminus, ce qui explique sa plus grande capacité à l'alkylation, alors que les interactions intramoléculaires et l'environnement compact autour du site à zinc diminuent sa réactivité dans le N-terminus.

Cette hypothèse est à mettre en relation avec notre étude sur l'influence de la liaison hydrogène sur la nucléophilie du motif zinc-thiolate (Chapitre V). Ainsi, nous avons conclu que de telles interactions atténuent de façon importante la réactivité des thiolates liés au zinc, ce qui rejoint l'idée d'un doigt métallique plus écranté et donc moins nucléophile.

- *Sélectivité entre les thiolates d'un même doigt de zinc*

La sélectivité de l'alkylation se retrouve également entre les trois cystéinates du doigt de zinc du côté C-terminal. Les intermédiaires et produits de la réaction entre la protéine p7 et des agents alkylants ont été piégés par spectrométrie de masse et analysés par HPLC (Hathout 1996). Il en résulte que les trois cystéinates du doigt de zinc du C-terminus sont présentes sous forme alkylée dans les produits et qu'elles sont donc toutes des cibles initiales pour l'attaque électrophile. Cependant, le peptide obtenu par l'alkylation de la cystéine 49 est le plus abondant, suivi de celui obtenu par alkylation de la cystéine 36, ce qui suggère que la cystéine 49 réagit préférentiellement avec les molécules antivirales.

Les causes avancées pour cette sélectivité sont l'accès plus facile au solvant de la cystéine et son plus grand caractère nucléophile mesuré par des coefficients de Fukui (Maynard 1998, Huang 1998). De plus, un examen des liaisons hydrogènes indique que la cystéine 49 se situe dans une région plus flexible du doigt de zinc que les deux autres cystéinates. Sa plus faible participation dans les interactions intramoléculaires le rend plus accessible et surtout plus riche en électrons (Summers 1992, Hathout 1996). Dans le doigt de zinc du côté N-terminal, la cystéine correspondante (Cys 28) est

d'ailleurs impliquée dans une liaison hydrogène supplémentaire avec le groupe amide du squelette peptidique deux résidus plus loin.

Ces études supportent la nucléophilie supérieure de la cystéinate 49, coordonnée au zinc dans le C-terminus, par rapport aux autres thiolates du site. Nous avons montré précédemment que l'affinité protonique était un indice fiable de réactivité du motif zinc-thiolate, alors que le coefficient de Fukui ne permettait pas une estimation correcte de la tendance à l'alkylation (Chapitre V). Dans ce chapitre, nous déterminerons les affinités protoniques de ces thiolates et rationaliseront la sélectivité envers un unique thiolate grâce à cet indice.

II. Evaluation du niveau de calcul

Au cours des travaux précédemment décrits dans ce manuscrit, nous avons étudié différents systèmes chimiques de taille relativement modeste (moins de 100 atomes) à l'aide de fonctionnelles de la DFT présentes dans les logiciels GAUSSIAN et TURBOMOLE. L'utilisation de ces outils pour ce type de molécule est très fréquente depuis de nombreuses années et est réalisable sans difficulté majeure.

L'utilisation des mêmes outils pour des molécules de plus grande taille, comme envisagée dans ce chapitre, pose différents problèmes techniques qu'il est nécessaire d'aborder avant toute étude d'intérêt chimique et/ou biologique. Le premier problème concerne la faisabilité de ces calculs, que ce soit d'un point de vue du temps de calcul nécessaire ou de la mémoire disponible. Nous verrons que la taille des systèmes étudiés nous a conduit à privilégier l'approximation *marij* avec une fonctionnelle GGA pour les optimisations de géométrie.

Le second aspect porte sur la précision des résultats obtenus, que ce soit au niveau des géométries obtenues ou des énergies relatives. Il a été nécessaire d'évaluer cette précision non seulement du fait de l'approximation utilisée pour les optimisations de géométrie, mais surtout du fait de la plus grande flexibilité des systèmes étudiés. En effet, il est apparu que suivant les paramètres utilisés au cours de l'optimisation de géométrie, différentes structures pouvaient être obtenues avec une proximité plus ou moins grande avec la structure expérimentale.

L'ensemble de ces travaux d'évaluation a été effectué sur le fragment de la protéine Ncp7 incluant le doigt de zinc en position C-terminale et constitué de 15 acides aminés (36-50). Il a en effet été montré expérimentalement par des études RMN que la structure de ce type de peptide est la même que la structure du fragment équivalent inclus dans la protéine entière (voir Partie I.B.2.b).

A. Faisabilité des calculs

Comme mentionné précédemment, la protéine Ncp7 contient 2 doigts de zinc. Nous nous limiterons ici à l'étude du doigt de zinc en position C-terminale pour lequel l'enchaînement des acides aminés est rappelé dans la figure VI.5.

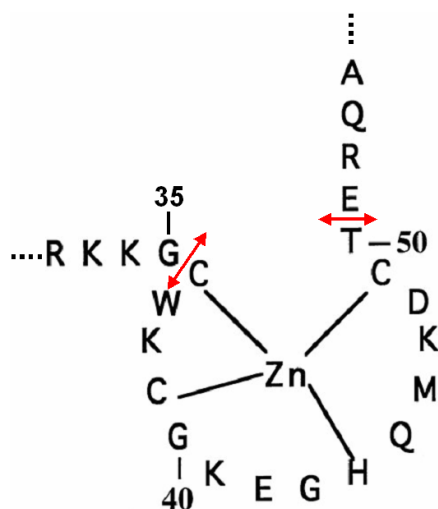


Figure VI.5 : Séquence de la protéine Ncp7 incluant le doigt de zinc en position C-terminale
Les flèches indiquent le fragment modélisé.

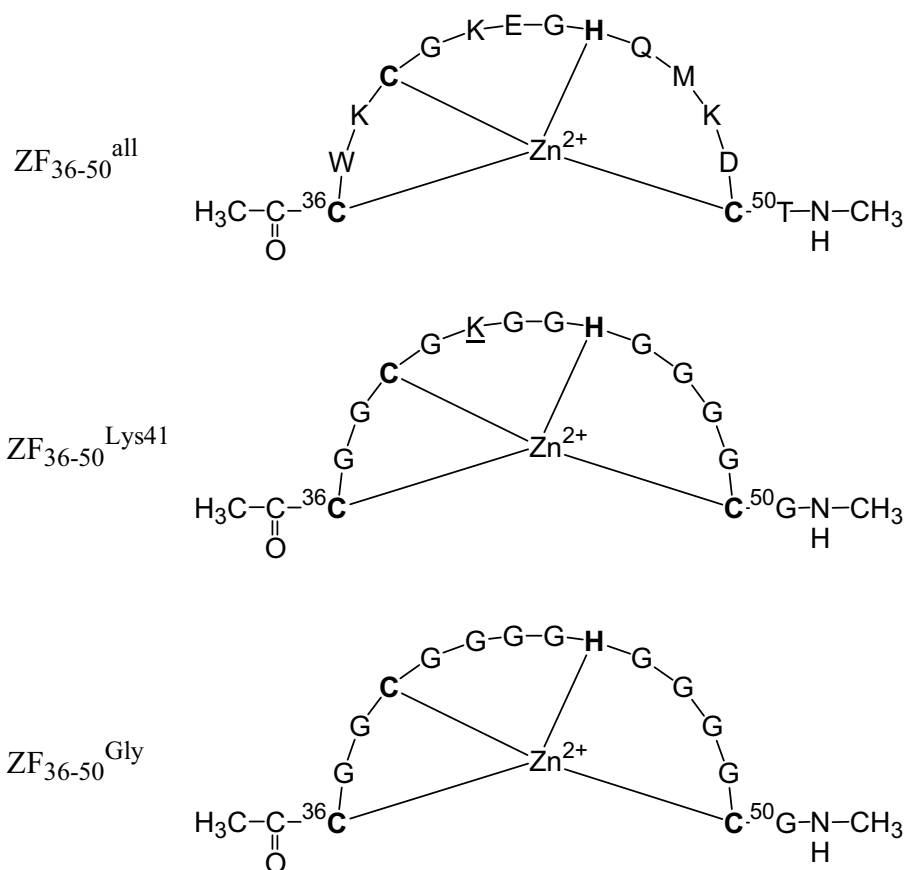


Figure VI.6 : Fragments modèles du doigt de zinc en position C-terminale de Ncp7

En incluant des terminaisons amides aux deux extrémités du fragment 36-50 (groupes CH₃-NH- et CH₃-C(O)- du côté C-terminal et N-terminal respectivement), de façon à éviter l'isomérisation possible entre des terminaisons neutres et zwitterioniques, nous obtenons le fragment noté ZF₃₆₋₅₀^{all}

qui contient 239 atomes (Figure VI.6). Ce système nous a paru encore de trop grande taille pour évaluer la faisabilité de la modélisation à l'aide de plusieurs méthodes. Nous avons en conséquence encore simplifié ce modèle en substituant tous les acides aminés de ZF_{36-50}^{all} , à l'exception de Cys36, Cys39, His44 et Cys49 qui sont liés au zinc par leur chaîne latérale, par des glycines. Cette simplification conduit au système noté ZF_{36-50}^{Gly} qui contient 137 atomes (Figure VI.6).

B. Structure tridimensionnelle

La structure tridimensionnelle de ZF_{36-50}^{Gly} , obtenue après simplification de la structure RMN de Ncp7 (structure pdb 1AAF, molécule 1) (Summers 1992) est représentée dans la figure VI.7. Le repliement de la chaîne peptidique autour du zinc pour former le doigt de zinc implique une structuration interne de ce fragment avec, en particulier, plusieurs liaisons hydrogène impliquant un groupe NH peptidique et un atome de soufre d'une cystéine ou un groupe CO peptidique. Certaines de ces interactions sont indiquées sur la figure VI.7.

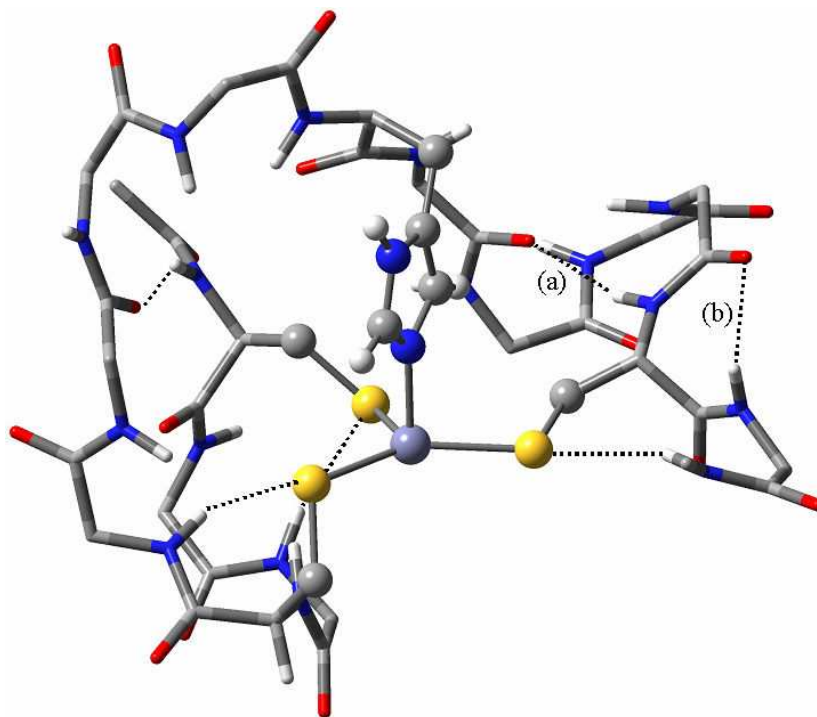


Figure VI.7 : Structure RMN de ZF_{36-50}^{Gly} (structure RMN 1AAF, molécule 1). Les atomes d'hydrogène aliphatiques ont été supprimés par souci de simplification. Le squelette peptidique est en représentation « Tube » tandis que le Zinc et ses ligands (chaînes latérales de Cys36, Cys39, His44 et Cys49) sont en représentation « Ball & Stick ».

Il est néanmoins difficile de faire un décompte exhaustif de ces interactions. En effet, il existe un continuum énergétique entre une liaison H existante et une liaison H inexistante, la force de l'interaction étant liée non seulement à la distance entre le groupe donneur et le groupe accepteur, mais aussi à leur orientation relative.

Nous aurons cependant par la suite besoin de définir plus précisément l'ensemble des liaisons H qui se trouvent présentes dans cette structure. En ce qui concerne les liaisons H entre un groupe NH peptidique et un atome de soufre lié au zinc d'une cystéine, nous avons évalué la distance entre

tous les H peptidiques et les atomes de soufre de Cys36, Cys39 et Cys49 (figure VI.8). Cette figure nous montre que les H peptidiques des résidus 38, 39, 40, 41 et 51 se trouvent tous à une distance inférieure à 3,5 Å d'un atome de soufre. D'autres résidus (35, 36, 46, 49 et 50) présentent une interaction vraisemblablement non négligeable avec un atome de soufre car la distance S...H est comprise entre 3,5 et 4,5 Å. Ce graphique montre enfin que Cys36, Cys39 et Cys49 présentent respectivement deux, trois et une liaisons H forte avec un NH peptidique.

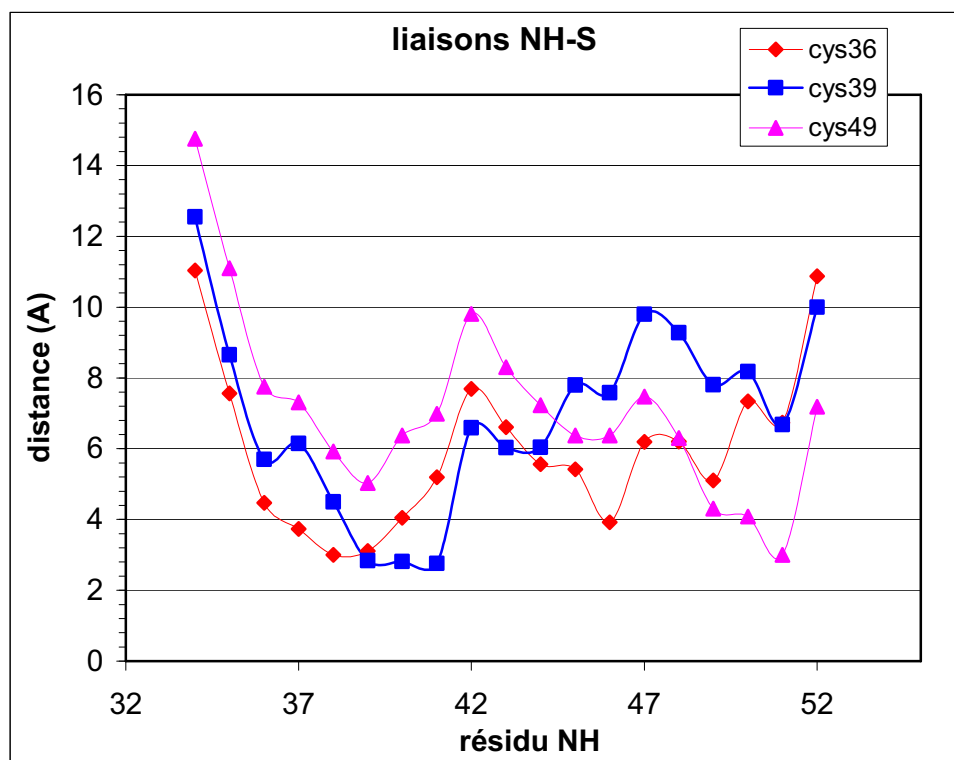


Figure VI.8 : Distance entre les H peptidiques et les atomes de soufre des cystéines 36, 39 et 49, pour la portion 34-52 de Ncp7 (structure 1AAF, molécule 1).

Ces interactions sont également importantes pour la structuration tridimensionnelle du doigt de zinc. Il faut noter que plusieurs structures expérimentales de la protéine Ncp7 ont été déterminées par RMN et sont disponibles dans la PDB : code 1NCP (Omichinski 1991), code 1AAF (Summers 1992), code 1MFS (Lee 1998) et code 1ESK (Morellet 2000). Trois de ces fichiers PDB contiennent en fait plusieurs structures RMN (entre 9 et 30 structures), ce qui montre une forte flexibilité de cette petite protéine. Cependant, ces études montrent aussi que cette flexibilité est limitée essentiellement aux chaînes latérales et, concernant le squelette peptidique, aux portions qui ne sont pas incluses dans un doigt de zinc. Ainsi, le squelette peptidique sur la partie 36-50 que nous souhaitons modéliser conserve, d'après ces travaux expérimentaux, une structure relativement rigide (figure VI.9).

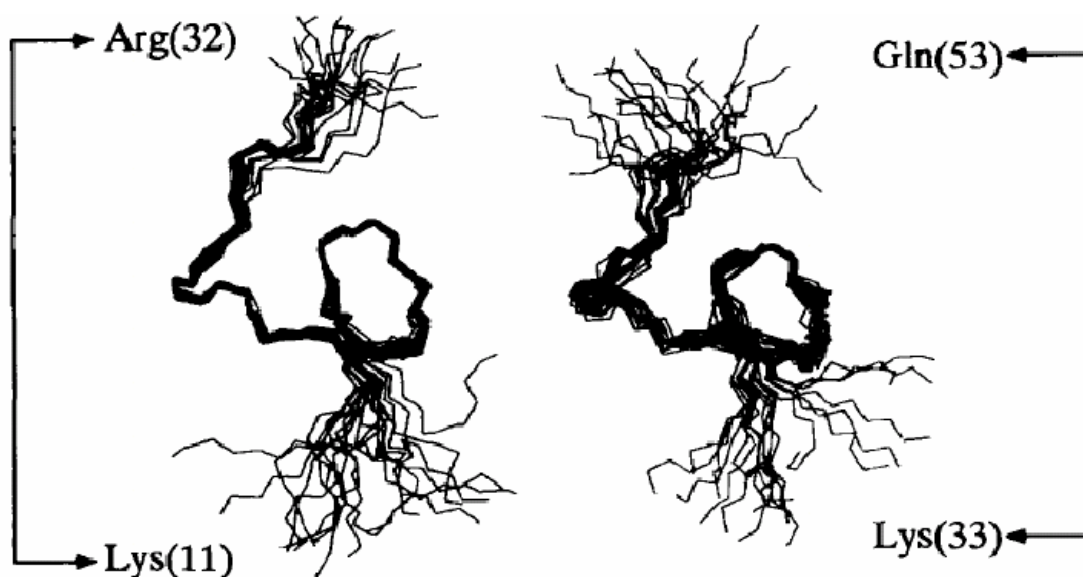


Figure VI.9 : Superposition des fragments 11-32 (gauche) et 33-53 (droite) pour 30 structures RMN de la protéine Ncp7 (Summers 1992).

Afin d'établir un critère quantitatif de ressemblance, nous avons calculé le RMSD (*Root Mean Square Deviation*) de la chaîne principale du peptide 36-50 pour les vingt structures RMN 1AAF (Summers 1992) ainsi que pour la première structure de 1ESK, structure décrite comme représentative de l'ensemble des structures par les auteurs (Morellet 2000).

Le tableau VI.1, qui rassemble l'ensemble des valeurs de RMSD, montre que les différentes structures présentent en moyenne le même écart entre elles, compris entre 0,85 et 1,15 Å, soit une valeur moyenne d'environ 1 Å. Cette valeur de 1 Å peut donc être considérée comme une valeur de référence pour estimer la pertinence de nos modélisations si l'on souhaite reproduire convenablement la structure tridimensionnelle de ce peptide. Nous pouvons cependant remarquer que des écarts plus importants existent entre certaines structures. Ainsi, on trouve comme écart le plus important une distance de 1,67 Å entre les molécules 4 et 11 du fichier PDB 1AAF.

Malgré leur grande ressemblance, au moins deux différences existent entre les deux doigts de zinc présents sur la protéine Ncp7. La première consiste en un enchaînement légèrement différent des acides aminés (voir figure VI.3). La seconde correspond à de faibles différences structurales au niveau de la chaîne principale comprise entre les acides aminés 46 et 49 pour le côté C-terminal et la partie 25-28 pour le côté N-terminal (South 1991). Ces deux différences peuvent expliquer la différence de réactivité entre les deux doigts de zinc. Il est donc crucial de correctement reproduire la structure spatiale d'un doigt de zinc si l'on veut correctement reproduire sa réactivité.

Dans cette partie, nous avons donc recherché le meilleur compromis entre la précision et le temps de calcul pour reproduire au mieux la structure expérimentale de ZF_{36-50}^{Gly} . La cible choisie correspond à la molécule 1 du fichier PDB 1AAF, structure « moyenne » par rapport à l'ensemble des structures disponibles.

Tableau VI.1 : RMSD entre différentes structures RMN de la chaîne principale du peptide 36-50 de Ncp7.

Code PDB	Numéro	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1esk	1	0,00																			
1aaf	1	1,07	0,00																		
1aaf	2	1,10	0,48	0,00																	
1aaf	3	0,94	1,01	1,12	0,00																
1aaf	4	1,26	0,77	0,60	1,37	0,00															
1aaf	5	1,01	0,84	0,75	0,87	0,84	0,00														
1aaf	6	0,98	0,93	0,89	0,65	1,13	0,72	0,00													
1aaf	7	0,85	0,88	0,95	0,51	1,23	0,93	0,54	0,00												
1aaf	8	1,04	1,10	1,18	0,48	1,44	0,91	0,56	0,56	0,00											
1aaf	9	1,21	0,85	0,86	1,27	0,90	0,84	1,06	1,15	1,19	0,00										
1aaf	10	1,09	1,12	1,16	0,90	1,26	0,95	0,66	0,78	0,75	1,05	0,00									
1aaf	11	1,42	1,52	1,59	1,29	1,67	1,52	1,13	1,06	1,15	1,52	0,82	0,00								
1aaf	12	1,14	1,19	1,19	1,28	1,15	1,26	1,01	1,01	1,24	1,13	0,90	0,95	0,00							
1aaf	13	1,14	1,19	1,19	1,28	1,15	1,26	1,01	1,01	1,24	1,13	0,90	0,95	0,00	0,00						
1aaf	14	1,24	1,19	1,17	1,36	1,12	1,30	1,08	1,05	1,31	1,63	0,95	1,03	0,54	0,00	0,00					
1aaf	15	1,27	0,73	0,68	1,18	0,73	0,87	0,88	0,95	1,16	0,80	0,88	1,27	1,01	1,00	1,00	0,00				
1aaf	16	1,20	1,00	1,13	0,96	1,17	1,03	0,86	0,82	0,96	1,08	0,65	0,86	1,01	0,97	0,97	0,77	0,00			
1aaf	17	1,20	1,34	1,46	0,69	1,63	1,18	0,76	0,76	0,61	1,48	0,74	0,93	1,27	1,35	1,35	1,30	0,83	0,00		
1aaf	18	1,26	1,22	1,19	1,27	1,24	1,29	0,99	0,92	1,18	1,35	0,86	0,89	0,92	0,68	0,68	0,99	0,87	1,14	0,00	
1aaf	19	1,33	1,16	1,25	1,16	1,24	1,16	0,94	0,98	1,12	1,21	0,59	0,78	1,00	0,95	0,94	0,82	0,41	0,92	0,79	0,00
1aaf	20	1,12	1,19	1,17	1,03	1,20	1,02	0,80	0,85	0,90	1,04	0,61	0,69	0,68	0,82	0,82	0,93	0,80	0,94	0,89	0,80
	moyenne	1,14	0,99	1,00	0,98	1,09	0,98	0,83	0,85	0,95	1,08	0,83	1,08	0,94	0,90	0,97	0,90	0,86	1,03	0,97	0,91

C. Précision des résultats

La structure ZF_{36-50}^{Gly} a été calculée à différents niveaux de calcul :

- (1) en utilisant le logiciel GAUSSIAN03 avec la fonctionnelle B3LYP. Dans ce cas, l'utilisation de la base BS1 pour toute la molécule s'est révélée impossible car alors le système à étudier était de taille trop importante. Nous avons donc diminué la taille de la base pour une partie du système. Les chaînes latérales des résidus Cys36, Cys39, His44 et Cys49, l'atome de zinc et les atomes NH-CO des liaisons peptidiques reliant les résidus 37-38, 38-39, 39-40, 40-41 et 50-51 ont été calculés avec la base BS1. Ces liaisons peptidiques correspondent à celles pour lesquelles une forte liaison H est observée avec un atome de soufre. Tous les autres atomes ont été calculés avec la base 3-21G. Cette base mixte sera notée par la suite BS1'.
- (2) En utilisant le logiciel TURBOMOLE avec les fonctionnelles BP86 et B3LYP, la base PB et, éventuellement, les approximations RI ou MARI-J.

Outre le niveau de calcul utilisé, un second problème est apparu au cours de cette calibration. En effet, nous avons remarqué que suivant le système de coordonnées utilisé pour représenter la structure de la molécule (coordonnées internes redondantes (CIR) ou coordonnées cartésiennes (CC)), les résultats pouvaient être différents (Koslover 2007).

Le tableau VI.2 présente les longueurs de liaison au niveau du site ZnS_3N en fonction de la méthode d'optimisation.

Tableau VI.2 : Structure du cœur ZnS_3N suivant la méthode d'optimisation (distances en Å)

Entrée	méthode d'optimisation ^a	Zn-S(Cys36)	Zn-S(Cys39)	Zn-S(Cys49)	Zn-N(His44)
1	Structure expérimentale	2.302	2.300	2.302	2.010
2	B3LYP/BS1' (CIR)	2.404	2.375	2.290	2.154
3	B3LYP/PB (CC)	2.418	2.368	2.324	2.166
4	RI-B3LYP/PB (CC)	2.407	2.353	2.341	2.136
5	MARI-J-B3LYP/PB (CC)	2.405	2.350	2.346	2.128
6	BP86/PB (CC)	2.402	2.356	2.307	2.145
7	RI-BP86/PB (CC)	2.402	2.357	2.309	2.147
8	MARI-J-BP86/PB (CC)	2.402	2.357	2.307	2.149
9	MARI-J-BP86/PB (CIR)	2.372	2.346	2.341	2.123
10	MARI-J-BP86/PB (CIR+12)	2.389	2.336	2.337	2.106
11	MARI-J-BP86/PB (CIR+18)	2.397	2.348	2.316	2.140
12	MARI-J-BP86/PB (CIR+34)	2.389	2.334	2.338	2.106

a. CC : coordonnées cartésiennes ; CIR : coordonnées internes redondantes par défaut ; CIR+X : X correspond au nombre de coordonnées internes redondantes ajoutées

Plusieurs constatations peuvent être faites à partir de ces résultats. Premièrement, les longueurs de liaisons Zn-S(Cys36), Zn-S(Cys39) et Zn-N(His44) expérimentales sont

systématiquement assez sensiblement inférieures aux distances calculées. L'absence de l'effet de solvant et la simplification de la molécule calculée par rapport à la protéine peuvent être une raison de cet écart. Il faut également noter que la résolution de ces structures RMN est faite en utilisant des calculs de modélisation moléculaire incluant des paramètres prédéfinis pour les distances métal-ligands. Nous pensons que cet effet représente un biais important dans la structure expérimentale qui explique les écarts des structures calculées par rapport à celle-ci. En effet, les trois longueurs de liaison Zn-S expérimentales sont identiques, ce qui est assez surprenant vu la différence d'environnement des atomes de soufre. Ceci correspondrait donc bien au fait que la distance Zn-S est paramétrée à 2,3 Å.

Les structures calculées sont toutes relativement proches. Les écarts les plus importants sont de 0,046 Å pour Zn-S(Cys36) (entrées 3 et 9), 0,041 Å pour Zn-S(Cys39) (entrées 2 et 12), 0,056 Å pour Zn-S(Cys49) (entrées 2 et 5) et 0,060 Å pour Zn-N(His49) (entrées 3 et 10/12). En examinant plus finement les résultats, certaines tendances peuvent être notées : (i) les approximations RI et MARI-J avec la méthode B3LYP induisent de faibles modifications par rapport à l'absence d'approximation (entrées 3-5) alors qu'avec la fonctionnelle BP86, les résultats sont identiques avec ou sans approximation (entrées 5-7) ; (ii) les longueurs de liaison métal-ligand sont assez peu sensibles à la fonctionnelle BP86 ou B3LYP (entrées 3-5 vs. 6-8) ; (iii) l'utilisation des coordonnées internes redondantes ou des coordonnées cartésiennes conduit à des structures différentes du site métallique. Le tableau VI.3 nous permet de constater que cela est lié à des conformations différentes du squelette peptidique.

Tableau VI.3 : Structure globale de ZF₃₆₋₅₀^{Gly} suivant la méthode d'optimisation (distances en Å, énergies en kcal/mol)

Entrée	méthode d'optimisation ^a	(a) ^b	(b) ^c	RMSD ^d	RMSD ^e	RMSD ^f	ΔE ^g
1	Structure expérimentale	1,973	2.658	0.00	0.98	0.99	
2	B3LYP/BS1' (CIR)	3,473	1.826	0.92	0.56	0.56	0.00
3	B3LYP/PB (CC)	3,748	2.268	0.99	0.29	0.00	-19.28
4	RI-B3LYP/PB (CC)	4,537	4.308	1.31	0.70	0.66	-22.18
5	MARI-J-B3LYP/PB (CC)	4,886	4.752	1.38	0.82	0.70	-22.98
6	BP86/PB (CC)	3,754	2.123	1.00	0.28	0.39	-6.60
7	RI-BP86/PB (CC)	3,743	2.167	1.01	0.28	0.08	-6.65
8	MARI-J-BP86/PB (CC)	3,742	2.103	0.98	0.00	0.29	-6.54
9	MARI-J-BP86/PB (CIR)	6,789	5.098	1.69	1.12	1.10	-15.32
10	MARI-J-BP86/PB (CIR+12)	4,786	4.934	1.43	0.94	0.85	-10.65
11	MARI-J-BP86/PB (CIR+18)	3,913	2.854	1.27	0.74	0.63	-8.31
12	MARI-J-BP86/PB (CIR+34)	4,646	4.937	1.42	0.94	0.85	-10.42

a. CC : coordonnées cartésiennes ; CIR : coordonnées internes redondantes par défaut ; CIR+X : X correspond au nombre de coordonnées internes redondantes ajoutées. b. distance O^{··}H (a) figure VI.7. c. distance O^{··}H (b) figure VI.7. d. RMSD par rapport à la structure expérimentale. e. RMSD par rapport à la structure calculée au niveau MARI-J-BP86/PB (CC). f. RMSD par rapport à la structure calculée au niveau B3LYP/PB (CC). g. Energie relative au niveau B3LYP/GB.

Nous avons vu précédemment l'importance de la structuration tridimensionnelle du squelette peptidique, et en particulier de la portion comprise entre les résidus 46 et 49. Nous avons considéré deux distances CO $\cdots$ HN correspondant à des liaisons hydrogène (interactions notées (a) et (b) dans la figure VI.7). Le tableau VI.3 permet de comparer l'évolution de ces deux distances en fonction de la méthode d'optimisation. La distance (a) est assez fortement surestimée, quelque soit la méthode d'optimisation. On note cependant que l'utilisation des coordonnées internes redondantes au niveau MARI-J-BP86/PB (entrée 9) allonge très fortement cette liaison par rapport au même calcul effectué avec les coordonnées cartésiennes (entrée 8). La distance (b) n'est pas non plus très précisément reproduite, avec une surestimation ou une sous-estimation importante suivant la méthode. Là encore, le passage des coordonnées cartésiennes aux coordonnées internes redondantes modifie très fortement le résultat de l'optimisation, conduisant à un fort allongement de cette distance.

Le Tableau VI.3 montre également l'écart (RMSD) entre les structures calculées et la structure expérimentale. Par rapport à cette dernière, seuls les calculs effectués au niveau B3LYP sans approximation (entrées 2 et 3) ou au niveau BP86, en coordonnées cartésiennes et en incluant ou non les approximations RI et MARI-J (entrées 6-8), conduisent à des RMSD inférieurs ou égaux à 1 Å. Les écarts sont plus faibles si l'on prend pour référence des structures calculées, mais dans tous les cas, les calculs effectués en coordonnées internes redondantes avec le logiciel TURBOMOLE (entrées 9-12) présentent un écart sensible avec les autres structures.

Cette influence des paramètres d'optimisation nous a conduit à examiner de plus près la nature des coordonnées internes redondantes définies par défaut par les programmes GAUSSIAN03 et TURBOMOLE. Dans le premier cas, de nombreuses distances à longue portée (c'est à dire entre des atomes interagissant par liaison hydrogène) sont définies, alors que cela n'est pas le cas avec TURBOMOLE. Ainsi, Cette absence de paramètres à longue portée conduit à des structures optimisées où certaines de ces interactions ont été rompues, pour conduire à un autre minimum local sur la surface de potentiel. Afin de pallier à ce problème, nous avons modifié les coordonnées internes redondantes par défaut en ajoutant d'autres distances à longue portée. Suivant le critère pour définir ces interactions à longue portée (distances de 4 à 5 Å, angle plus ou moins aigu), nous avons ainsi ajouté 12, 18 ou 34 nouveaux paramètres (entrées 10-12). Le résultat est peu satisfaisant car, même si l'ajout de ces nouveaux paramètres permet de se rapprocher de la structure expérimentale, suivant les paramètres ajoutés, la structure est à chaque fois différente. Il s'est donc révélé impossible de définir un critère universel pour inclure de nouvelles coordonnées internes redondantes et dans la suite de ce travail, tous les calculs effectués avec le logiciel TURBOMOLE seront effectués avec les coordonnées cartésiennes, même si le nombre de pas d'optimisation pour aboutir à une structure convergée est sensiblement augmenté avec ce choix de paramètres.

Enfin, nous avons effectué une comparaison au niveau B3LYP/GB de l'énergie relative de l'ensemble de ces structures. Comme attendu, les optimisations B3LYP conduisent à des structures de plus basse énergie que les optimisations BP86. Il faut noter cependant que le temps d'optimisation est sensiblement plus long avec la fonctionnelle hybride par rapport à la fonctionnelle GGA, en particulier du fait de l'impossibilité de paralléliser le calcul lorsque l'on inclue les approximations RI ou MARI-J. Comme indiqué dans la partie précédente, l'optimisation en coordonnées internes

redondantes avec BP86 (entrées 9-12) conduit à différents minima locaux légèrement plus bas en énergie que le minimum proche de la structure expérimentale obtenu avec les coordonnées cartésiennes (entrée 8).

L'ensemble de ces considérations nous conduit à privilégier pour la suite de cette étude la méthode d'optimisation MARI-J-BP86/PB en utilisant comme paramètres les coordonnées cartésiennes.

III. Calcul de l'affinité protonique d'un fragment de protéine

Ayant défini le protocole de calcul, nous avons pu débiter le cœur de ce chapitre, à savoir l'estimation de la nucléophilie relative de groupes thiolates au sein d'un même site réactif. Notre cible reste le site de type $\text{Zn}(\text{Cys})_3(\text{His})^-$ côté C-terminal de la protéine Ncp7. Il a été démontré expérimentalement que, pour cette protéine, la cystéine qui subit l'alkylation est Cys49 (Hathout 1996). Notre objectif est de mettre en évidence l'influence des différents éléments de la protéine sur cette sélectivité. L'ensemble des calculs effectués ici, sauf indication contraire, a été réalisé au niveau MARI-J-BP86/PB.

A. Référence

Le premier modèle de notre site actif est le modèle le plus simple envisageable, c'est à dire la molécule $\text{Zn}(\text{SMe})_3(\text{Im})^-$ (Figure VI.10).

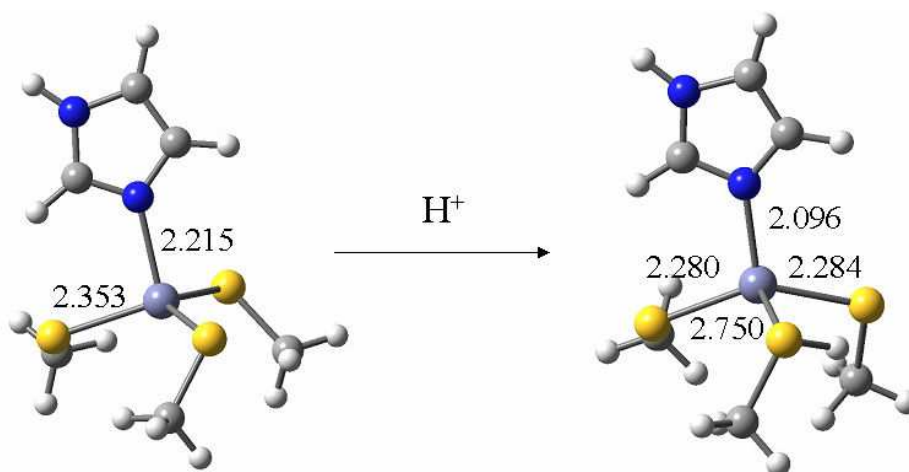


Figure VI.10 : Structures de $\text{Zn}(\text{SMe})_3(\text{Im})^-$ et $\text{Zn}(\text{HSMe})(\text{SMe})_2(\text{Im})$.

Cette molécule reproduit exactement la première sphère de coordination du doigt de zinc de Ncp7. Ceci permet d'estimer l'influence des différents ligands du zinc sur sa structure et sa basicité. De plus, elle nous servira de référence pour estimer l'influence des différents éléments de l'environnement autour de cette première sphère de coordination.

La structure de $\text{Zn}(\text{SMe})_3(\text{Im})^-$ (Figure VI.10, gauche) montre que les trois ligands thiolates sont (quasi)-équivalents. Il n'est donc pas possible sur ce modèle d'estimer une quelconque sélectivité pour la réactivité d'alkylation. Celle-ci est donc bien liée à l'environnement du site réactif. La protonation de $\text{Zn}(\text{SMe})_3(\text{Im})^-$ conduit à $\text{Zn}(\text{HSMe})(\text{SMe})_2(\text{Im})$ (Figure VI.10, droite). A l'exception du groupe protoné pour lequel la distance avec le métal augmente, les liaisons des trois autres ligands avec le zinc se trouvent renforcées. On retrouve ici les observations effectuées sur d'autres modèles biomimétiques dans les chapitres précédents et qui montrent l'effet « push-pull » entre les différents ligands du métal : lorsque l'interaction entre un des ligands et le métal est amoindrie, ici par l'effet de la protonation, alors l'interaction des autres ligands avec le métal devient plus forte.

Enfin, la mesure de l'affinité protonique de $\text{Zn}(\text{SMe})_3(\text{Im})^-$ conduit à une valeur, qui nous servira de référence pour la suite, de 337 kcal/mol.

B. Influence du squelette peptidique

La première étape de cette évaluation de l'environnement a consisté à étudier l'affinité protonique de $\text{ZF}_{36-50}^{\text{Gly}}$. Outre l'effet des ligands de la première sphère de coordination, ce modèle inclut les contraintes de rigidité imposées par le squelette peptidique sur l'arrangement des ligands du métal. De plus, comme nous l'avons indiqué précédemment, ce modèle intègre également les liaisons hydrogènes entre les NH peptidiques et les atomes de soufre (figure VI.8). Ainsi, 2, 3 et 1 groupes NH sont présents dans une sphère de 3,5 Å autour des atomes de soufre de Cys36, Cys39 et Cys49 respectivement.

La figure VI.11 rassemble les structures optimisées au niveau MARI-J-BP86/PB de $\text{ZF}_{36-50}^{\text{Gly}}$ et de ses acides conjugués après protonation d'un des trois atomes de soufre de Cys36, Cys39 et Cys49 (structures notées respectivement $\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{36}$, $\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{39}$, et $\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{49}$). Pour chaque atome de soufre, la protonation peut se faire des deux côtés du plan défini par le plan C-S-Zn (noté a et b), ce qui conduit au final à un total de 6 structures protonées.

La figure VI.11 montre que la protonation d'un atome de soufre ne modifie que très faiblement la structuration tridimensionnelle du doigt de zinc. Un examen plus précis des structures montre cependant deux évolutions non négligeables. Premièrement, les liaisons hydrogène qui pointent vers l'atome de soufre ayant subi la protonation sont affaiblies. Ceci est particulièrement visible pour les structures $\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{49}$ pour lesquelles la liaison hydrogène provient du fragment $-\text{C}(\text{O})-\text{NH}-\text{CH}_3$ terminal. En effet, l'absence de contrainte de structure sur cette extrémité lui permet d'adopter une conformation assez différente dans laquelle la distance $\text{S}(\text{Cys49})\cdots\text{HN}$ est fortement allongée (4,27 et 5,03 Å respectivement pour $\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{49a}$ et $\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{49b}$, contre 3,28 Å pour $\text{ZF}_{36-50}^{\text{Gly}}$). Cet effet est généralement atténué, mais toujours présent, dans les autres structures car les NH peptidiques en liaison H avec un soufre sont dans des régions plus contraintes du squelette protéique. Deuxièmement, comme pour le complexe biomimétique $\text{Zn}(\text{SMe})_3(\text{Im})^-$, une forte évolution des longueurs de liaison métal-ligand est observée (tableau VI.4).

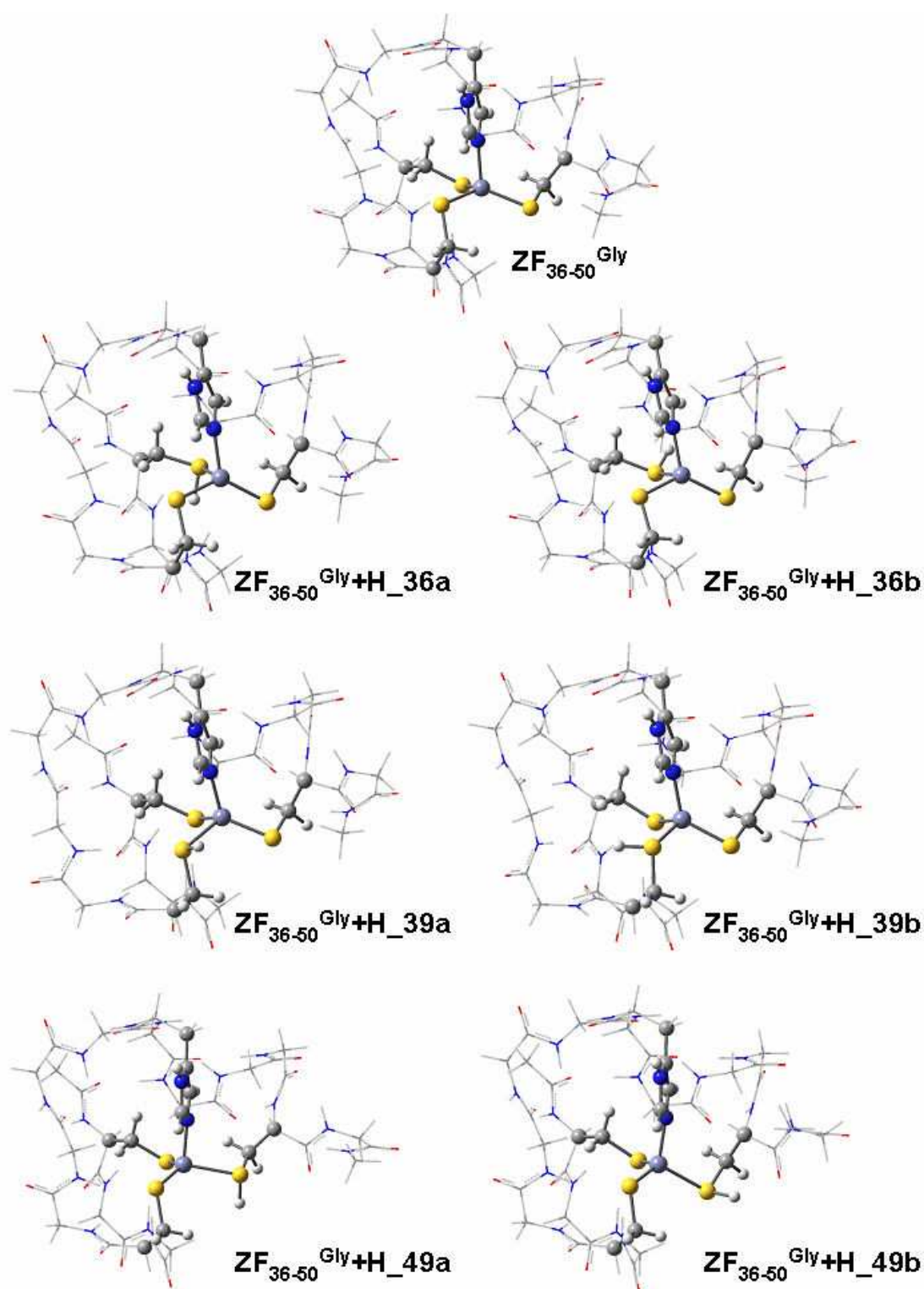


Figure VI.11 : Structures optimisées de ZF_{36-50}^{Gly} et de ses acides conjuguées. Seuls les ligands directs du métal sont représentés en « Ball & Stick », les autres atomes étant en « wireframe » pour des raisons de lisibilité.

Si l'on compare tout d'abord ZF_{36-50}^{Gly} à $Zn(SMe)_3(Im)^-$, on constate que la présence du squelette peptidique conduit à différencier les trois groupes thiolates, ceux-ci ayant des longueurs de liaisons avec le métal comprises entre 2,307 et 2,402 Å. Contrairement à ce qui pouvait être attendu d'après l'étude sur l'influence d'une liaison H sur des complexes biomimétiques (Chapitre V), on

constate que cette différenciation n'est pas corrélée avec le nombre de liaisons hydrogène. En effet, même si Cys49 présente la distance la plus courte car elle ne possède qu'une seule liaison H, ce n'est plus le cas pour Cys36 qui possède moins de liaison H que Cys39 mais présente une liaison Zn-S plus longue. Nous pouvons donc avancer l'hypothèse que deux phénomènes influencent la longueur de la liaison Zn-S, le nombre de liaison H, mais aussi les contraintes stériques du squelette peptidique.

Tableau VI.4 : Structure du cœur ZnS_3N de $\text{ZF}_{36-50}^{\text{Gly}}$ et de ses acides conjugués (distances en Å).

structure	Zn-S(Cys36)	Zn-S(Cys39)	Zn-S(Cys49)	Zn-N(His44)
$\text{ZF}_{36-50}^{\text{Gly}}$	2.402	2.357	2.307	2.149
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{36\text{a}}$	2.656	2.299	2.273	2.088
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{36\text{b}}$	2.693	2.295	2.265	2.089
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{39\text{a}}$	2.338	2.581	2.256	2.086
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{39\text{b}}$	2.336	2.547	2.266	2.082
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{49\text{a}}$	2.326	2.278	2.539	2.086
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{49\text{b}}$	2.337	2.278	2.532	2.085

Dans tous les cas, la protonation d'un atome de soufre conduit à un allongement notable (environ 0,25 Å) de la liaison Zn-S correspondante et à un raccourcissement des trois autres liaisons métal-ligand. Il faut noter que cet allongement de la liaison Zn-S est sensiblement moins important ici que pour le complexe biomimétique $\text{Zn}(\text{SMe})_3(\text{Im})^-$ (+0,40 Å), ce qui nous indique là encore l'influence du squelette peptidique : la rigidité qu'il impose au site ne permet pas aux ligands du métal de s'éloigner du métal aussi facilement qu'en présence de la seule première sphère de coordination.

La différence d'énergie entre $\text{ZF}_{36-50}^{\text{Gly}}$ et ses différents acides conjugués permet d'obtenir l'affinité protonique des différents sites de protonation possible du doigt de zinc (Tableau VI.5).

Tableau VI.5 : Affinités protoniques des différents sites de protonation possible de $\text{ZF}_{36-50}^{\text{Gly}}$ (kcal/mol).

Structure de l'acide	AP ^a	AP ^b	AP ^c	AP ^d
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{36\text{a}}$	286.0	286.7	300.3	289.2
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{36\text{b}}$	286.7	288.9	300.0	287.0
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{39\text{a}}$	291.1	292.7	302.4	292.2
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{39\text{b}}$	295.1	293.3	304.1	293.3
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{49\text{a}}$	301.0	299.7	311.1	298.2
$\text{ZF}_{36-50}^{\text{Gly}}+\text{H}_{49\text{b}}$	301.0	297.6	310.5	296.7

Affinités protoniques calculées au niveau de calcul : a : MARI-J-BP86/PB ; b : B3LYP/GB//MARI-J-BP86/PB ; c : B3LYP/BS1' ; d : B3LYP/GB//B3LYP/BS1'.

Outre l'affinité calculée au niveau MARI-J-BP86/PB, différents autres niveaux de calcul ont également été réalisés pour affermir et vérifier ces résultats. Nous pouvons ainsi constater que le niveau de calcul MARI-J-BP86/PB donne des affinités protoniques similaires à un calcul plus précis au niveau B3LYP/GB//MARI-J-BP86/PB. Nous pourrions donc par la suite nous limiter aux valeurs données par le niveau d'optimisation MARI-J-BP86/PB. Une optimisation de la géométrie au niveau B3LYP/BS1' conduit à des affinités protoniques de 10 kcal/mol supérieures environ aux précédentes. Cependant, un calcul d'énergie finale au niveau B3LYP/GB redonne des valeurs équivalentes. Ceci nous montre que le niveau de calcul utilisé pour effectuer les optimisations a peu d'importance sur les valeurs finales de l'affinité protonique si un simple-point d'énergie au niveau B3LYP/GB est effectué.

L'affinité d'un atome de soufre est peu dépendante de l'angle d'approche du proton. En effet, les écarts entre les conformères a et b de $ZF_{36-50}^{Gly}+H_{36}$, $ZF_{36-50}^{Gly}+H_{39}$ et $ZF_{36-50}^{Gly}+H_{49}$ sont très faibles. L'écart pour $ZF_{36-50}^{Gly}+H_{39}$ est le plus important au niveau de calcul MARI-J-BP86/PB (4 kcal/mol) mais celui-ci est fortement abaissé aux autres niveaux de calcul. Ceci peut sembler assez surprenant car les deux faces des plans C-S-Zn ne sont pas équivalentes puisqu'il existe différentes liaisons hydrogène sur les atomes de soufre. Nous avons cependant vu précédemment que ces liaisons hydrogène sont assez fortement allongées par la protonation de l'atome de soufre. Ceci nous confirme donc que l'énergie de ces liaisons hydrogène impliquant un groupe thiol lié au zinc sont faibles et ne contribuent que très faiblement à la stabilité de l'espèce protonée. Dans la suite de ces travaux, nous nous limiterons donc, dans la plupart des cas, au calcul d'une seule affinité protonique pour chaque atome de soufre.

Les écarts entre les différents isomères $ZF_{36-50}^{Gly}+H_{36}$, $ZF_{36-50}^{Gly}+H_{39}$ et $ZF_{36-50}^{Gly}+H_{49}$ sont plus significatifs (environ 5 kcal/mol) avec un ordre identique quel que soit le niveau de calcul utilisé. Ainsi, l'atome de soufre le plus basique de ZF_{36-50}^{Gly} est celui de Cys49, suivi par celui de Cys39, puis en dernier celui de Cys36. Cet ordre de basicité corrèle avec la longueur de la liaison Zn-S dans ZF_{36-50}^{Gly} . En effet, plus la liaison Zn-S est courte, plus la basicité de l'atome de soufre est forte. Inversement, le nombre de liaison hydrogène par atome de soufre ne permet pas de prédire l'ordre complet de la réactivité de ceux-ci. En effet, si S(Cys49) est non seulement l'atome le plus basique et aussi celui qui possède le plus faible nombre de liaison hydrogène, cette comparaison n'est plus pertinente pour classer S(Cys36) et S(Cys39).

Par comparaison avec $Zn(SMe)_3(Im)^-$, l'affinité protonique de ZF_{36-50}^{Gly} est fortement abaissée d'environ 35 kcal/mol, soit 10 % de sa valeur. Cette valeur peut être comparée à ce que nous avons obtenu dans le cas de complexes biomimétiques (Chapitre V). En passant d'un phénylthiolate à un arylthiolate incluant une liaison H intramoléculaire sur le soufre, nous avons calculé une diminution de l'affinité protonique de cet atome d'au plus 10 kcal/mol. Cette comparaison nous permet de mettre en évidence que l'influence du squelette peptidique sur la basicité d'un groupe thiolate lié au zinc est bien supérieure à celle d'une seule liaison H, et cela même si, comme dans le cas de Cys49, une unique liaison H est présente sur l'atome de soufre. Ce phénomène peut s'expliquer non seulement par les contraintes stériques imposées par le squelette peptidique, mais surtout, à notre avis, par le fait que l'ensemble des liaisons hydrogène autour du site à zinc influence la réactivité de chaque groupe thiolate. Ainsi, comme nous l'avons montré sur les complexes biomimétiques, chaque liaison H a un

rôle global, ce qui implique par exemple que celles formées avec les soufre de Cys36 et Cys39 diminuent également la basicité de l'atome S(Cys49). L'inclusion des effets dus aux chaînes latérales (voir plus loin) confirmera ce rôle global.

Il est enfin intéressant de noter que nos calculs permettent de trouver que la cystéine 49 est la cystéine la plus réactive du site. Ceci permet de reproduire avec succès les résultats expérimentaux qui ont montré que seule Cys49 est alkylée. Ceci nous montre toute l'influence du squelette peptidique sur la réactivité de ces sites. En effet, non seulement il permet de différencier les différents groupes thiolates, mais surtout, il ne semble pas nécessaire d'inclure l'effet des chaînes latérales pour expliquer la réactivité relative des ligands en première sphère de coordination. Dans le plus de confirmer cette hypothèse, nous avons dans la suite de ce chapitre chercher à inclure une ou plusieurs chaînes latérales dans le modèle ZF_{36-50}^{Gly} .

C. Influence de la chaîne latérale de Lys41

Après avoir étudié le modèle ZF_{36-50}^{Gly} n'incluant aucune chaîne latérale autres que celles liées au zinc, nous avons décidé de construire un nouveau modèle incluant une chaîne latérale formant une liaison H avec l'un des atomes de soufre. Parmi toutes les chaînes latérales d'acides aminés, les chaînes chargées positivement à pH physiologique (Lys, Arg et éventuellement His) semblent les plus intéressantes puisque ce sont celles qui, potentiellement, formeraient les liaisons H les plus fortes avec les groupes thiolates liés au zinc. L'examen de la séquence de Ncp7 montre que plusieurs résidus satisfont à ce critère au voisinage du doigt de zinc : Arg32, Lys33, Lys34, Lys38, Lys41, Lys47 et Arg52. L'examen des différentes structures RMN disponibles dans la PDB montre que seules Lys41 et Arg52 présentent, pour certaines de ces structures, une liaisons H avec un des atomes de soufre. Afin de ne pas allonger la chaîne peptidique de notre modèle, nous avons donc sélectionné la chaîne latérale de Lys41. Le modèle ainsi construit sera noté ZF_{36-50}^{Lys41} (Figure VI.6).

L'examen détaillé des différentes structures RMN de Ncp7 montre que Lys41 adopte différentes conformations. Trois catégories de conformations peuvent être définies, détaillées dans la figure VI.12.

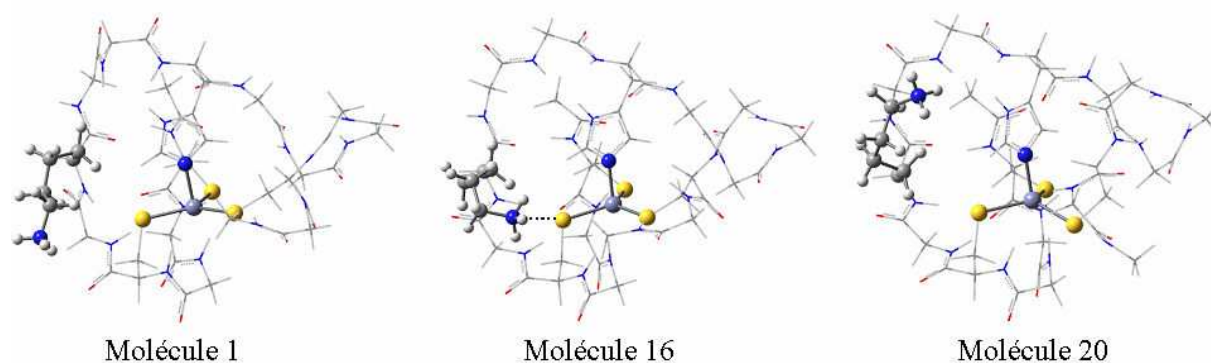


Figure VI.12 : Conformations différentes de Lys41 présente dans la structure PDB 1AAF. Seuls les atomes directement liés au zinc et la chaîne latérale de Lys41 sont représentés en « Ball & Stick », les autres atomes étant en « wireframe » pour des raisons de lisibilité.

La première conformation, illustrée par la molécule 1 de la structure PDB 1AAF, correspond à une chaîne latérale sans interaction de type liaison H avec le doigt de zinc. Elle se trouve cependant dans le voisinage de Cys39 (distances $N^{\cdots}S$ et $N^{\cdots}OC_{\text{peptidique}}$ de 4,95 et 5,11 Å respectivement). La seconde conformation, illustrée par la molécule 16, correspond à une chaîne latérale de Lys41 formant une liaison hydrogène avec l'atome de soufre de Cys39 (distance $N^{\cdots}S$ de 3,64 Å). Enfin, la molécule 20 illustre la troisième conformation où la chaîne latérale de Lys41 est dans le voisinage du cycle imidazole et de la chaîne peptidique au niveau des résidus 42 et 43 (distances $N^{\cdots}OC_{\text{peptidique}}$ de 4,72 et 4,71 Å respectivement).

Afin d'explorer l'influence de la chaîne latérale de Lys41 dans ces différentes conformations, nous avons optimisé $ZF_{36-50}^{\text{Lys41}}$ au départ de ces trois géométries. Les résultats obtenus sont représentés dans la figure VI.13.

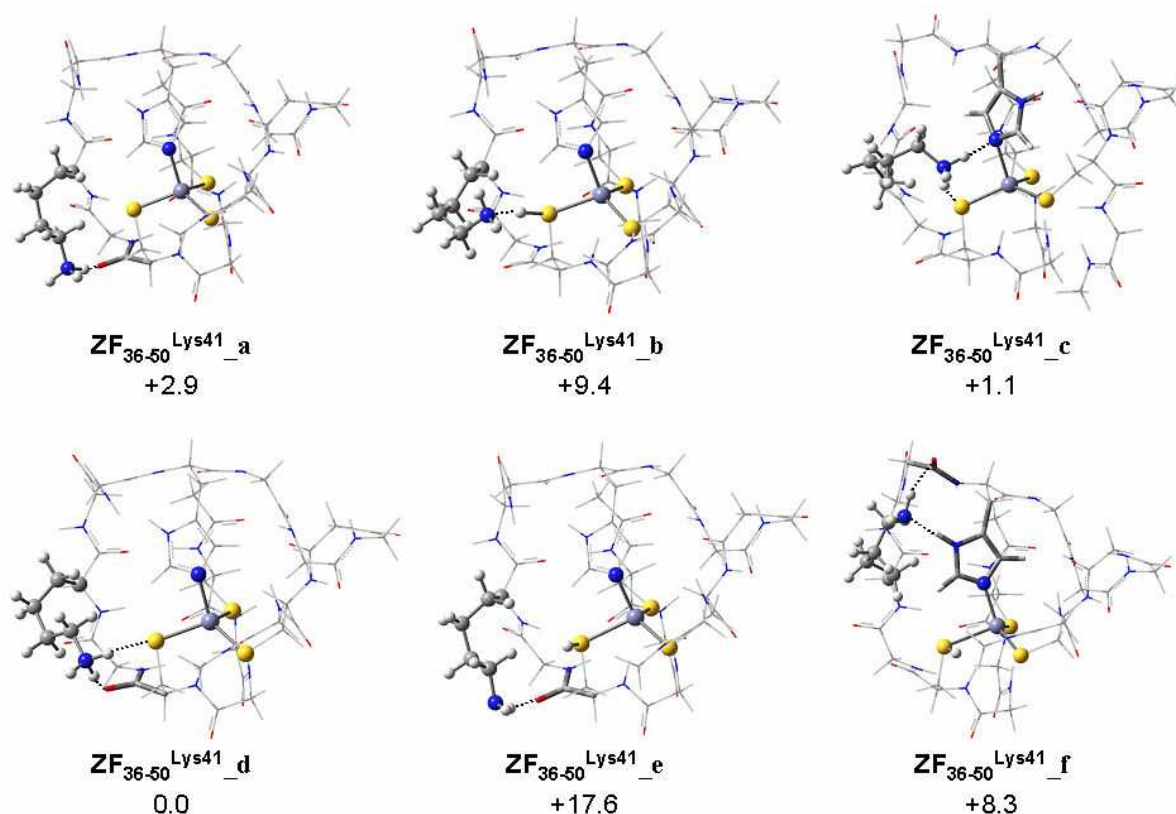


Figure VI.13 : Structures optimisées de $ZF_{36-50}^{\text{Lys41}}$ au niveau MARI-J-BP86/PB.

Les atomes directement liés au zinc et la chaîne latérale de Lys41 sont représentés en « Ball & Stick », les atomes interagissant avec la chaîne latérale de Lys41 sont représentés en « tube » et les autres atomes sont en « wireframe ». Les énergies relatives sont en kcal/mol.

Contrairement à ce que nous avons observé lors des optimisations de géométrie du modèle ZF_{36-50}^{Gly} , les trois structures de $ZF_{36-50}^{\text{Lys41}}$ (notées a, b et c), optimisées respectivement à partir des structures expérimentales des molécules 1, 16 et 20 de 1AAF.pdb, présentent des différences notables avec les structures expérimentales. Ces différences proviennent de la chaîne latérale de Lys41. Vraisemblablement du fait de l'absence de solvant pour modérer la charge du groupe ammonium et former des liaisons hydrogène avec lui, celui-ci interagit beaucoup plus fortement avec

le doigt de zinc que dans les structures expérimentales. Cela se traduit, suivant la conformation considérée, par différents effets :

- $ZF_{36-50}^{Lys41}_a$ possède une forte liaison hydrogène entre le groupe ammonium et le CO peptidique de Cys39 (distance $N^{\cdots}OC_{peptidique}$ de 2,515 Å). Comme dans la structure de la molécule 1, l'interaction entre le groupe ammonium et le soufre de Cys39 reste présente (distance $N^{\cdots}S$ de 4,050 Å).
- $ZF_{36-50}^{Lys41}_b$ possède, comme la structure expérimentale, une liaison hydrogène entre le groupe ammonium et l'atome de soufre de Cys39 (distance $N^{\cdots}S$ de 3,121 Å). Cependant, cette liaison hydrogène a une polarité inversée puisque l'atome d'hydrogène est porté par l'atome de soufre et non par l'atome d'azote. Cette inversion de la liaison hydrogène conduit à un site ZnS_3N dans lequel l'un des atomes de soufre est protoné. Ce résultat nous semble à mettre en relation avec la controverse de la littérature portant sur l'éventuelle protonation de certains soufres dans les sites ZnS_3N et ZnS_4 . Nous y reviendrons dans la partie VI.3.d.
- Dans $ZF_{36-50}^{Lys41}_c$, le groupe ammonium forme deux liaisons hydrogène, l'une avec l'atome de soufre de Cys39 (distance $N^{\cdots}S$ de 3,057 Å) et l'autre avec l'un des azotes du cycle imidazole (distance $N^{\cdots}N$ de 2,931 Å). Cette seconde interaction semble peut vraisemblable car, à notre connaissance, une telle interaction n'a jamais été décrite dans une enzyme. Cette structure est cependant la plus stable des trois structures a, b et c.

Cette forte flexibilité de la chaîne latérale de la lysine 41 nous a incité à rechercher d'autres minima sur la surface de potentiel à partir des trois structures précédentes. Ceci d'autant plus que la structure la plus stable obtenue parmi les trois décrites ci-dessus ne correspondait pas à une structure proche d'une structure expérimentale. Nous avons ainsi obtenu trois autres structures (notées d, e et f. Figure VI.13) :

- Dans $ZF_{36-50}^{Lys41}_d$, le groupe ammonium forme deux liaisons hydrogène, l'une avec l'atome de soufre de Cys39 et l'autre avec le CO peptidique de Cys39. Cette structure est donc assez proche de la structure expérimentale de la molécule 1 qui présente également ces deux interactions, même si elles sont plus fortes dans la structure calculée (distances $N^{\cdots}S$ et $N^{\cdots}O$ calculées de 3,02 et 2,62 Å contre 4,95 et 5,11 Å dans la molécule 1).
- Les structures $ZF_{36-50}^{Lys41}_e$ et $ZF_{36-50}^{Lys41}_f$ présentent une chaîne latérale Lys41 déprotonée en interaction respectivement avec le CO peptidique de Cys39 (distance $N^{\cdots}OC_{peptidique}$ de 3,043 Å) ou avec à la fois le cycle imidazole et le CO peptidique du résidu 43 (distances $N^{\cdots}N$ et $N^{\cdots}OC_{peptidique}$ de 2,778 et 3,101 Å respectivement). Dans ces deux structures, le proton est porté par l'atome de soufre de Cys39, comme dans la structure $ZF_{36-50}^{Lys41}_b$.

L'ensemble de ces 6 structures montre une grande variété d'interaction possible entre la chaîne latérale de Lys41 et le doigt de zinc. Dans tous les cas, cette (ou ces) interaction est plus forte que l'interaction équivalente dans la structure expérimentale. Nous pouvons émettre l'hypothèse que cela est dû à l'absence du solvant dans nos modélisations. Nous reviendrons sur ce point dans la partie VI.3.d. Enfin, nous pouvons remarquer que la structure la plus stable ($ZF_{36-50}^{Lys41}_d$) correspond

à une orientation de la chaîne latérale de Lys41 intermédiaires entre les structures expérimentales des molécules 1 et 16, ce qui est satisfaisant pour la pertinence de nos modélisations.

Si cette grande flexibilité de la chaîne latérale de Lys41 complique l'étude de ce système, elle permet d'explorer son influence pour différentes positions, plus ou moins proche de la première sphère de coordination du site. En effet, sur la base des géométries obtenues, nous pouvons classer ces structures en fonction de la proximité entre le groupe ammonium (ou le groupe amine suivant les cas) et l'atome de soufre de Cys39. Ceci donne, par ordre croissant d'influence de la chaîne latérale de Lys41 sur le site :

- ZF_{36-50}^{Lys41} _a ne présente pas de liaison H entre la chaîne latérale de Lys41 et la première sphère de coordination. L'influence du groupe ammonium est donc purement électrostatique, il est situé à une distance de 4 Å de S(Cys39).
- ZF_{36-50}^{Lys41} _d présente une liaison H orientée $NH \cdots S$ entre l'ammonium et S(Cys39).
- ZF_{36-50}^{Lys41} _b présente une liaison H orientée $N \cdots HS$ entre l'ammonium et S(Cys39).
- ZF_{36-50}^{Lys41} _e et ZF_{36-50}^{Lys41} _f présentent une protonation sans liaison H sur l'atome de soufre de Cys39.

Pour ces différents cas (la structure ZF_{36-50}^{Lys41} _c n'a pas été considérée), nous avons alors calculé l'affinité protonique des différents atomes de soufre lié au zinc. Pour tous ces calculs, nous avons été particulièrement vigilants avec les structures obtenues, celles-ci devant conserver une interaction entre la chaîne latérale de la lysine 41 et le doigt de zinc, équivalente à celle présente dans la structure non protonée. Un exemple des structures ainsi obtenues est représenté dans la figure VI.14 pour le cas ZF_{36-50}^{Lys41} _d.

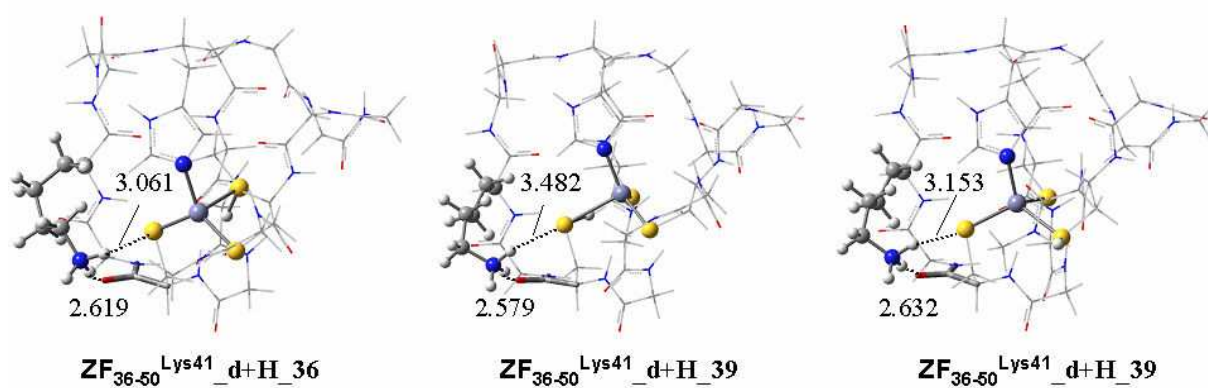


Figure VI.14 : Structures optimisées de ZF_{36-50}^{Lys41} _d+H au niveau MARI-J-BP86/PB. Les atomes directement liés au zinc et la chaîne latérale de Lys41 sont représentés en « Ball & Stick », les atomes interagissant avec la chaîne latérale de Lys41 sont représentés en « tube » et les autres atomes sont en « wireframe ». Les distances $N \cdots OC$ et $N \cdots S$ sont en Å.

On constate sur ces trois structures, par rapport à la structure de ZF_{36-50}^{Lys41} _d (Figure VI.13), une très bonne conservation de l'interaction entre le groupe ammonium et à la fois le CO peptidique de Cys39 et l'atome de soufre de Cys39. Comme attendu, l'écart le plus important est observé lorsque

le site de protonation est la cystéine 39. On a alors un allongement de la liaison N...S (3,482 contre 3,025 Å dans ZF₃₆₋₅₀^{Lys41}_d) et un léger raccourcissement de la liaison N...OC (2,579 contre 2,615 Å dans ZF₃₆₋₅₀^{Lys41}_d).

L'ensemble des résultats obtenus pour les différentes structures est présenté dans le tableau VI.6. Par rapport aux affinités protoniques calculées pour ZF₃₆₋₅₀^{Gly}, nous constatons dans tous les cas une forte diminution liée à la chaîne latérale de Lys41. Celle-ci étant chargée positivement (ou ayant apportée un proton localisé sur S(Cys39)), elle introduit une interaction coulombienne répulsive avec le proton ajouté, ce qui induit une diminution de l'affinité protonique des systèmes ZF₃₆₋₅₀^{Lys41}.

Cette diminution n'est cependant pas équivalente dans tous les cas. En effet, elle dépend à la fois du site de protonation et de la conformation considérée.

Tableau VI.6 : Affinités protoniques des différents sites de protonation possible de ZF₃₆₋₅₀^{Lys41} calculées au niveau MARI-J-BP86/PB (kcal/mol).

Site de protonation Structure de la base	S(Cys36)	S(Cys39)	S(Cys49)
ZF ₃₆₋₅₀ ^{Gly}	286.7	295.1	301.0
ZF ₃₆₋₅₀ ^{Lys41} _a	246.5	243.7	260.4
ZF₃₆₋₅₀^{Lys41}_d	245.1	241.0	265.2
ZF ₃₆₋₅₀ ^{Lys41} _b	247.8	-	255.7
ZF ₃₆₋₅₀ ^{Lys41} _e	239.5	-	251.3
ZF ₃₆₋₅₀ ^{Lys41} _f	240.3	-	247.4

Les valeurs correspondant aux structures de plus basses énergies sont en gras.

Par rapport aux valeurs obtenues pour ZF₃₆₋₅₀^{Gly}, ZF₃₆₋₅₀^{Lys41}_a présente une diminution d'environ 40 kcal/mol des affinités protoniques de S(Cys36) et S(Cys49) et une diminution plus marquée, d'environ 50 kcal/mol, de l'affinité protonique de S(Cys46). Ainsi, la plus grande proximité de la charge pour la cystéine 36 induit une plus grande réduction de son affinité protonique. Le renforcement de l'interaction entre la chaîne latérale de Lys41 et le doigt de zinc, et en particulier Cys39, conduit alors à des structures où S(Cys39) ne peut plus être protoné, l'addition d'un proton sur cet atome conduisant par exemple à des structures où le proton ajouté a migré sur un CO voisin. La même tendance à la réduction de l'affinité protonique est observée pour les deux autres atomes de soufre, de ZF₃₆₋₅₀^{Lys41}_a à ZF₃₆₋₅₀^{Lys41}_f. Celle-ci reste cependant relativement modeste, avec une diminution d'au plus 13 kcal/mol.

Ces résultats nous permettent de confirmer l'influence locale et globale d'une liaison H sur la réactivité d'un site. En effet, le groupement considéré dans notre cas est situé à proximité de S(Cys39). Celui-ci a un effet important sur la cystéine 39, mais aussi, de façon non négligeable, sur les deux autres atomes de soufre S(Cys36) et S(Cys49). Nous constatons néanmoins que l'effet local prédomine et conduit à une inversion de la réactivité des cystéines 36 et 39, l'atome S(Cys36) ayant dans le modèle ZF₃₆₋₅₀^{Lys41} une plus grande affinité protonique que S(Cys39).

Ces résultats sont très intéressants car ils montrent l'importance de l'environnement sur la réactivité d'un site enzymatique. Ils sont cependant problématiques en ce qui concerne nos modélisations car dans celles-ci, nous n'avons pas introduit l'ensemble de l'environnement. L'introduction de la totalité de celui-ci (ensemble des chaînes latérales, solvation) semble donc nécessaire si l'on veut obtenir une description fiable et quantitative de la réactivité d'un site enzymatique et de la réactivité relative des différentes cystéines présentes. Outre cette question de la réactivité, nous avons également vu que la structure elle-même du site semble dépendre de l'environnement. En effet, les structures calculées de ZF_{36-50}^{Lys41} présentent une interaction beaucoup plus forte, par rapport aux structures expérimentales, entre la chaîne latérale de Lys41 et le doigt de zinc. Ceci conduit dans certains cas à un transfert de proton entre le groupe ammonium en deuxième sphère de coordination et un atome de soufre lié au zinc. L'hypothèse du rôle du solvant sur ce comportement reste à vérifier. Les deux parties suivantes vont donc chercher à aborder ces différentes questions. Nous rechercherons tout d'abord à comprendre l'effet du solvant sur la structure du site enzymatique, puis nous introduirons l'ensemble de l'environnement pour déterminer la réactivité relative des différents atomes de soufre du site.

D. Influence de la solvation sur la structure du doigt de zinc

Nous avons vu précédemment que l'optimisation de la géométrie en phase gazeuse du modèle ZF_{36-50}^{Lys41} conduit non seulement à des interactions de type liaisons H trop fortes, mais surtout à des structures non observées expérimentalement ($ZF_{36-50}^{Lys41_c}$). Dans cette partie, nous avons voulu vérifier que ces interactions non observées dans les structures expérimentales étaient effectivement dues à l'absence du solvant. Pour cela, nous avons ré-optimisées les structures de ZF_{36-50}^{Lys41} en incluant un continuum polarisable avec la méthode COSMO. Les structures les plus stables ainsi obtenues sont représentées dans la figure VI.15.

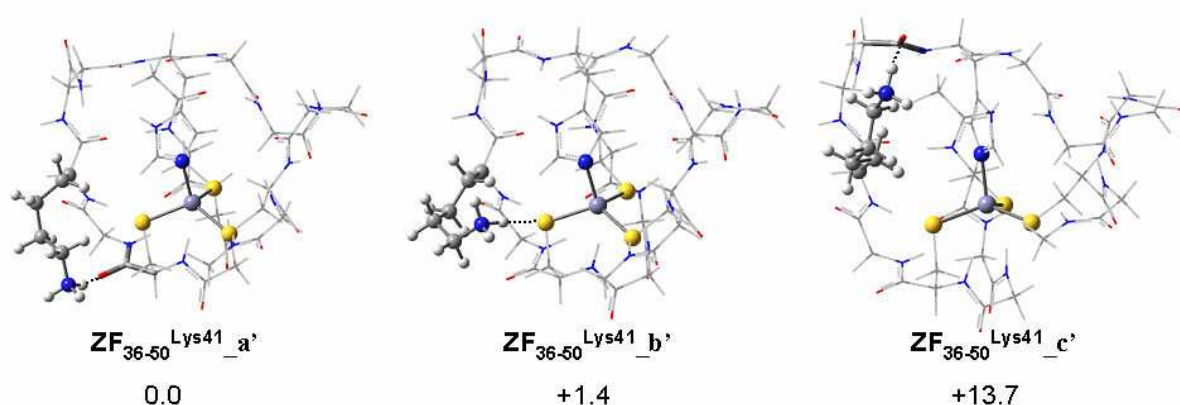


Figure VI.15 : Structures optimisées de ZF_{36-50}^{Lys41} au niveau COSMO-MARI-J-BP86/PB.

Les atomes directement liés au zinc et la chaîne latérale de Lys41 sont représentés en « Ball & Stick », les atomes interagissant avec la chaîne latérale de Lys41 sont représentés en « tube » et les autres atomes sont en « wireframe ». Les énergies relatives sont en kcal/mol.

Comparées aux structures obtenues en phase gazeuse (figure VI.13), ces structures se rapprochent sensiblement des structures expérimentales :

- $ZF_{36-50}^{Lys41_a'}$ présente une liaison hydrogène $NH\cdots OC$ allongée par rapport à $ZF_{36-50}^{Lys41_a}$ (2,692 contre 2,515 Å) et se rapproche ainsi de la structure de la molécule 1 (Figure VI.12).
- La polarité de la liaison H entre la chaîne latérale de Lys41 et S(Cys39) est inversée dans $ZF_{36-50}^{Lys41_b'}$ par rapport à $ZF_{36-50}^{Lys41_b}$ et retrouve ainsi la même orientation que dans la molécule 16 (Figure VI.12). La force de la liaison H reste surestimée par le calcul, avec une distance $N\cdots S$ de 3,17 Å contre 3,64 Å dans la structure expérimentale.
- Le bras ammonium de $ZF_{36-50}^{Lys41_c'}$ ne fait plus, comme dans $ZF_{36-50}^{Lys41_c}$, une interaction avec le groupe imidazole et se retrouve dans une position voisine de celle observée dans la molécule 20 (Figure VI.12).

D'autres structures, équivalentes aux structures $ZF_{36-50}^{Lys41_e}$ et $ZF_{36-50}^{Lys41_f}$ ont été obtenues en solution. Elles sont localisées plus de 15 kcal/mol au dessus de $ZF_{36-50}^{Lys41_a'}$ et ne sont donc pas représentées ici. Enfin, les optimisations de géométries pour obtenir une structure proche de $ZF_{36-50}^{Lys41_d}$ conduisent invariablement aux structures $ZF_{36-50}^{Lys41_a'}$ ou $ZF_{36-50}^{Lys41_b'}$.

Ces résultats nous montre que l'introduction des effets de solvant permet de s'approcher de façon plus raisonnable des structures expérimentales que les optimisations en phase gazeuse. Nous constatons néanmoins des différences significatives entre les structures calculées ici et les structures RMN. Nous attribuons ces différences au fait que l'effet de solvant a été modélisé par un continuum. En effet, aucune molécule d'eau discrète n'a été introduite, alors que celles-ci forment en solution un réseau de liaisons hydrogène avec la protéine. En particulier, le groupe ammonium de Lys41 est vraisemblablement fortement solvato par les molécules d'eau, effet que l'on sait être mal reproduit par un continuum diélectrique. Ceci expliquerait les plus faibles distances entre ce groupe chargé et le doigt de zinc dans nos calculs par rapport aux structures expérimentales. Au niveau de calcul utilisé ici, il n'était cependant pas possible d'inclure un nombre significatif de molécules d'eau pour solvater le doigt de zinc et, même si cela aurait été réalisable par d'autres méthodes (modélisation moléculaire par exemple), nous n'avons donc pas recherché à vérifier leur rôle précis sur la structure du site actif.

Un point a cependant attiré plus particulièrement notre attention. Nous avons précédemment calculé en phase gazeuse une inversion de la liaison hydrogène entre Lys41 et Cys39 dans la structure $ZF_{36-50}^{Lys41_b}$. Celle-ci ne se retrouve pas en solution puisque, malgré plusieurs essais, seule la structure $ZF_{36-50}^{Lys41_b'}$, présentant une liaison H de type $NH\cdots S$, a pu être obtenue. Nous pensons que ces résultats sont à même d'expliquer la controverse de la littérature concernant l'état de protonation des doigts de zinc (voir Chapitre I Partie III.B). En effet, l'étude expérimentale qui a montré que les sites $Zn(Cys)_3(His)$ et $Zn(Cys)_4$ sont respectivement simplement et doublement protonés a été effectuée par spectrométrie de masse (Fabris 1996, 1999). Ces études montraient par exemple que la méthylation d'un doigt de zinc de type $Zn(Cys)_3(His)$ n'augmente la masse de celui-ci que de 14 Da, indiquant ainsi que, avant méthylation, le doigt de zinc était protoné.

Nous émettons ici l'hypothèse que ces résultats ne tiennent pas compte de l'influence de la solvation sur la structure des doigts de zinc. En effet, en l'absence de solvant, et seulement dans ce cas, un transfert de proton est possible depuis la deuxième sphère de coordination vers un atome de soufre lié au zinc. Lors du passage de la phase aqueuse à la phase gazeuse, ce transfert peut donc s'effectuer. De plus, la chaîne latérale basique se retrouvant ainsi déprotonée, elle peut alors accepter un nouveau proton lors de l'électrospray. Ainsi, un fragment $\text{Lys}^{\cdots}\text{Cys-Zn}$, qui n'accueille qu'un proton en solution, pourrait en accepter un second lors de la mise en phase gazeuse. Par contre, si la cystéine est protonée, alors le transfert de proton entre la deuxième et la première sphère de coordination du métal est impossible et le fragment $\text{Lys}^{\cdots}\text{Cys}(\text{méthylée})\text{-Zn}$ ne possède donc, en solution comme en phase gaz, qu'un seul proton. La différence de masse serait donc de 15 Da en solution, mais seulement de 14 Da en phase gaz, comme observée par spectrométrie de masse.

Nous n'avons pas exploré davantage cette hypothèse au cours de cette thèse car elle se situait en dehors de notre périmètre d'étude.

E. Influence de l'ensemble de l'environnement sur la réactivité du site

Nous avons vu ci-dessus qu'une chaîne latérale peut fortement moduler la réactivité d'un site enzymatique et que la solvation joue un rôle non négligeable dans la structure de ces sites. Pour décrire de façon fiable le site réactif de la protéine Ncp7, il faut donc inclure l'ensemble de ces éléments sous peine d'ignorer un facteur important de la réactivité. La stratégie adoptée jusqu'ici, c'est à dire l'optimisation des géométries et le calcul des énergies au niveau DFT, n'était cependant pas adaptée pour cela. En effet, l'introduction de toutes les chaînes latérales conduit à un système de taille trop importante (pour le fragment 35-53, cela correspond à 300 atomes) pour être étudié au niveau de calcul utilisé jusque là. De plus, nous avons vu que l'introduction d'une unique chaîne latérale induit une forte flexibilité et donc une multitude de minima. Ceci sera d'autant plus vrai avec plusieurs chaînes latérales et l'étude d'un tel système ne peut plus être envisagée de façon statique.

Face à ces difficultés, nous avons décidé d'utiliser le matériel expérimental à notre disposition. En effet, comme nous l'avons noté précédemment, les fichiers PDB de Ncp7 contiennent chacun plusieurs conformations différentes de la protéine, les différences au niveau des doigts de zinc s'effectuant essentiellement sur les chaînes latérales. Ainsi, les structures RMN de Ncp7 fournissent un échantillon des différentes conformations possibles des chaînes latérales. De plus, la pertinence des structures RMN est incontestable, même si leur précision atomique est liée à l'utilisation de paramètres structuraux prédéfinis.

Nous avons donc décidé d'utiliser trois de ces structures (correspondant aux molécules 1, 16 et 20 de 1AAF.pdb) sans procéder à aucune ré-optimisation. Comme nous l'avons vu précédemment, ces structures présentent une variété de conformation pour la chaîne latérale de Lys41. Il en est de même pour les autres chaînes latérales qui adoptent des conformations très différentes entre ces trois structures. Cependant, dans tous les cas sauf pour Arg52, ces chaînes latérales sont peu ou ne sont pas en interaction avec le site enzymatique ou avec le squelette peptidique mais pointent vers l'extérieur, c'est à dire vers le solvant, ou forment des liaisons H entre elles. Nous avons néanmoins conservé l'ensemble de ces chaînes latérales, en particulier car certaines sont chargées (positivement

ou négativement) et peuvent donc jouer un rôle non négligeable, même à grande distance du site à zinc. De plus, sur la molécule 20, la chaîne latérale de Arg52 forme une liaison H avec S(Cys49). Nous avons donc décidé d'inclure légèrement la taille du fragment, pour inclure ce résidu et pour diminuer les erreurs liées à un fragment trop restreint sur le doigt de zinc. Nous avons donc considéré le fragment incluant la totalité des acides aminés 35 à 53 et présentant des terminaisons amides. Ce système est noté ZF_{35-53}^{all} .

La protonation de ce fragment a été effectuée en incluant manuellement un H^+ sur l'une des trois atomes de soufre dans la position (distance S-H, angle Zn-S-H et angle dièdre N-Zn-S-H) qu'il occupait dans nos modèles $ZF_{36-50}^{Gly}+H_{36}$, $ZF_{36-50}^{Gly}+H_{39}$ et $ZF_{36-50}^{Gly}+H_{49}$. Afin de pouvoir mesurer l'influence des différents facteurs de l'environnement, ce protocole, basé sur les structures RMN et sans optimisation de géométrie, a aussi été effectué pour les systèmes ZF_{36-50}^{Gly} et ZF_{36-50}^{Lys41} . Les calculs d'énergies ont été tous effectués au niveau MARI-J-BP86/BP à la fois en phase gazeuse et en solution à l'aide de l'approximation COSMO. Pour le modèle ZF_{35-53}^{all} , il n'a pas été possible, pour aucune des conformations considérées, d'obtenir la convergence SCF inférieure à 10^{-4} du calcul DFT pour les calculs en phase gazeuse. Nous attribuons ce problème à la présence d'un trop grand nombre de charges (positives et négatives). En conséquence, ces résultats ne sont pas considérés ici. L'ensemble des résultats obtenus pour ces différents modèles est présenté dans le tableau VI.7.

Tableau VI.7 : Affinités protoniques des modèles ZF_{36-50}^{Gly} , ZF_{36-50}^{Lys41} et ZF_{35-53}^{all} calculées, pour leur structure RMN, au niveau MARI-J-BP86/PB et COSMO-MARI-J-BP86/PB (kcal/mol).

		Molécule 1			Molécule 16			Molécule 20		
		S(36)	S(39)	S(49)	S(36)	S(39)	S(49)	S(36)	S(39)	S(49)
AP										
1	ZF_{36-50}^{Gly}	284.7	296.5	279.9	287.3	281.3	293.7	290.0	287.9	275.2
2	ZF_{36-50}^{Lys41}	243.0	222.3	257.4	239.7	208.9	246.8	247.3	238.9	235.8
3	$ZF_{36-50}^{Gly} + \text{COSMO}$	271.6	278.2	268.3	275.0	265.0	277.1	273.8	269.2	256.1
4	$ZF_{36-50}^{Lys41} + \text{COSMO}$	269.2	261.5	276.7	271.1	246.0	274.6	271.4	265.8	254.7
5	$ZF_{35-53}^{all} + \text{COSMO}$	263.9	260.4	273.8	271.1	243.9	271.7	269.6	263.8	246.2
6	$ZF_{35-53}^{all} + \text{COSMO corr}^a$	265.9	259.0	294.9	270.6	257.7	278.9	266.3	270.9	272.0
ΔAP										
7	$ZF_{36-50}^{Gly} \rightarrow ZF_{36-50}^{Lys41}$	-41.7	-74.3	-22.5	-47.5	-72.4	-46.9	-42.7	-49.0	-39.4
8	$ZF_{36-50}^{Gly} + \text{COSMO} \rightarrow$ $ZF_{36-50}^{Lys41} + \text{COSMO}$	-2.4	-16.7	+8.4	-3.9	-19.0	-2.5	-2.3	-3.4	-1.4
9	$ZF_{36-50}^{Lys41} + \text{COSMO} \rightarrow$ $ZF_{35-53}^{all} + \text{COSMO}$	-5.3	-1.1	-2.9	0.0	-2.1	-2.9	-1.8	-2.0	-8.5
10	$ZF_{36-50}^{Gly} \rightarrow$ $ZF_{36-50}^{Gly} + \text{COSMO}$	-13.1	-18.3	-11.6	-12.2	-16.3	-16.6	-16.3	-18.7	-19.1
11	$ZF_{36-50}^{Lys41} \rightarrow$ $ZF_{36-50}^{Lys41} + \text{COSMO}$	+26.1	+39.2	+19.3	+31.4	+37.1	+27.7	+24.1	+26.8	+18.9

^a valeurs corrigées sur la base des données obtenues après optimisation des géométries pour le modèle ZF_{36-50}^{Gly} (Tableau VI.6), soit $[ZF_{35-53}^{all} \text{ optimisé} + \text{COSMO}] \approx [ZF_{35-53}^{all} + \text{COSMO corr}]$
 $= [ZF_{35-53}^{all} \text{ fixe} + \text{COSMO}] + [ZF_{36-50}^{Gly} \text{ optimisé}] - [ZF_{36-50}^{Gly} \text{ fixe}]$

Les entrées 1 et 2 du tableau VI.7 donnent les valeurs des affinités protoniques au niveau MARI-J-BP86/BP respectivement pour les modèles ZF_{36-50}^{Gly} et ZF_{36-50}^{Lys41} . Ces données ont été calculées dans les parties précédentes après optimisation des géométries (tableaux VI.5 et VI.6). Les résultats obtenus sans optimisation des géométries et avec optimisation des géométries montrent des différences significatives, en particulier pour l'atome de soufre S(Cys49) du modèle ZF_{36-50}^{Gly} qui présente systématiquement une affinité protonique sous évaluée (de 8 à 26 kcal/mol) en l'absence d'optimisation de la géométrie. Ces différences sont à mettre en relation avec les variations de la géométrie observées lors de l'optimisation des modèles ZF_{36-50}^{Gly} et $ZF_{36-50}^{Gly}+H$ (figure VI.11). En effet, nous avons noté que la protonation sur l'atome S(Cys49) conduisait à une modification notable des interactions de cet atome avec son environnement, contrairement au cas où la protonation s'effectue sur S(Cys36) ou S(Cys39). Cette différence de comportement lors de l'optimisation des géométries n'est pas, par principe, incluse dans les résultats du tableau VI.7, ce qui explique les différences obtenues. Les valeurs absolues d'affinité protonique du tableau VI.7 (entrées 1-5) sont donc à prendre avec précaution. Différentes conclusions peuvent néanmoins être obtenus de ces données. Tout d'abord, l'examen des affinités protoniques pour chaque atome de soufre montre que la basicité de S(Cys36) est très peu sensible à la géométrie du système. En effet, les valeurs pour les trois conformations différentes sont au plus différentes de 8 kcal/mol, que ce soit en l'absence ou en présence des chaînes latérales. Ceci est en accord avec la position de S(Cys36) dans l'enzyme, complètement entouré par le site à zinc et le squelette peptidique et donc relativement éloigné de toutes les chaînes latérales. A l'inverse, S(Cys39) et S(Cys49) présentent des affinités protoniques très différentes (jusqu'à des écarts de 30 kcal/mol) entre les différentes conformations. Ceci montre bien que la basicité de ces sites est très dépendante de leur environnement, ce qui semble en accord avec leur plus grande accessibilité au solvant, leur interaction possible avec des chaînes latérales chargées et, pour Cys49, la proximité avec l'extrémité du doigt de zinc et donc une zone plus flexible du squelette peptidique.

Pour comparer entre elles les valeurs des affinités protoniques des différents sites, une solution, pour minimiser l'erreur due à la non optimisation des géométries, est de modifier ces valeurs en y ajoutant un terme correctif pour reproduire pour chaque molécule et pour chaque atome de soufre l'affinité protonique du modèle ZF_{36-50}^{Gly} (entrée 6). Ces valeurs corrigées confirment la plus forte affinité protonique de S(Cys49) par rapport aux autres atomes de soufre, et cela quelque soit la conformation considérée. Ceci est en très bon accord avec les résultats expérimentaux montrant que cette position est la plus réactive pour l'alkylation. Nous constatons cependant que pour la conformation de la molécule 20, l'affinité protonique de S(Cys49) est très proche de celle des autres atomes de soufre, du fait de la présence de la chaîne latérale de Arg52 dans son voisinage. Sur la base de nos résultats, il est difficile de conclure quant à la réactivité relative des sites S(Cys36) et S(Cys39) puisque cela est fonction de la conformation considérée. Il est clair qu'il serait nécessaire d'évaluer l'entropie du système, et en particulier celle liée à la chaîne latérale de Lys41, pour pouvoir définitivement conclure sur ce point.

Si la non-optimisation des géométries conduit à des valeurs absolues de l'affinité protonique « contestables », leurs variations (entrées 7-11) ne sont pas sujettes à ce défaut du fait d'une

compensation des erreurs. Ainsi, l'addition dans nos modèles de la chaîne latérale de Lys41 conduit à une diminution de l'affinité protonique des différents sites (entrées 7 et 8). Cette variation est plus forte en phase gazeuse (entrée 7) qu'en solution (entrée 8) ce qui s'explique par une atténuation de l'influence de la charge du groupe ammonium sur le doigt de zinc du fait de sa solvation. Comme nous l'avons observé dans la partie III.C, l'influence de Lys41 est liée à sa proximité avec les différents atomes de soufre. Ainsi, son influence est maximale dans les conformations des molécules 1 et 16 pour l'atome S(Cys39) car elle se trouve à moins de 5 Å de celui-ci. L'inclusion de la totalité des chaînes latérales (entrée 9) ne conduit pas à de grandes variations de l'affinité protonique, celles-ci étant pour la plupart de quelques kcal/mol, c'est à dire du même ordre de grandeur que l'influence de Lys41 sur les atomes S(Cys36) et S(Cys49) situés à grande distance du groupe ammonium. Une différence notable concerne l'atome S(Cys49) dans la conformation de la molécule 20. En effet, dans ce cas, la présence de la chaîne latérale de Arg52 conduit à une diminution plus importante de son affinité protonique. Ces différents résultats montrent que, dans leur ensemble, les chaînes latérales ne jouent un rôle non négligeable sur l'affinité protonique des différents sites que lorsqu'elles s'en rapprochent à courte distance, par exemple en formant une liaison hydrogène avec l'un d'entre eux. Dans le cas contraire, leur influence, si elle n'est pas nulle, est fortement atténuée par l'effet du solvant et est de plus peu sélective entre les différents sites. Nous pouvons enfin noter que dans tous les cas (entrée 9), l'ajout des différentes chaînes latérales conduit à une diminution de l'affinité protonique de tous les sites, et cela même si, globalement, l'ensemble des chaînes latérales a une charge nulle.

Le solvant joue un rôle important sur les valeurs absolues des affinités protoniques. Dans le cas du modèle ZF_{36-50}^{Gly} (entrée 10), il diminue la basicité des groupes thiolates, comme nous l'avons observé par exemple dans le cas des complexes biomimétiques (chapitre IV). La situation est différente pour le modèle ZF_{36-50}^{Lys41} où l'effet du solvant conduit à une augmentation de l'affinité protonique des différents sites (entrée 11). Ceci est dû à la présence du groupe ammonium. En l'absence d'effet de solvant, son influence est maximale et il conduit à une forte diminution de l'affinité protonique des différents sites (entrée 7). En présence du solvant, sa charge est solvatée et perd ainsi une grande partie de son influence (entrée 8). Deux effets contraires sont alors présents : le solvant diminue, comme pour ZF_{36-50}^{Gly} , la nucléophilie des groupes thiolates, tendant ainsi à faire baisser les affinités protoniques, mais diminue également l'effet coulombien lié à la charge de l'ammonium, tendant ainsi à faire augmenter les affinités protoniques. Notre résultat montre que l'effet sur la charge de l'ammonium est ici prépondérant.

IV. Conclusion

L'objectif de ce chapitre était d'étendre nos études précédentes afin d'aborder un site de protéine dans toutes sa complexité. Les chapitres précédents nous avaient permis de mettre au point une méthodologie afin d'évaluer la nucléophilie relative des différents motifs zinc-thiolates dans un site réactif. Pour aborder l'étude de notre cible, le doigt de zinc en position C-terminale de la nucléocapside du virus HIV, il nous a tout d'abord fallu vérifier la faisabilité de l'étude d'un système de

grande taille (plus de 100 atomes) à l'aide de méthode de la DFT. Nous avons ainsi pu montrer que les outils à notre disposition nous permette d'étudier avec précision de tels systèmes en utilisant le niveau de calcul MARI-J-BP86/PB.

Passé ces problèmes méthodologiques, nous avons abordé par une stratégie « bottom-up » l'étude du doigt de zinc. En partant d'un modèle n'incluant que la première sphère de coordination du métal, nous avons introduit progressivement (i) le squelette peptidique, (ii) une chaîne latérale chargée présentant dans certaines conformations une forte interaction avec le doigt de zinc, (iii) l'ensemble des chaînes latérales du peptide et enfin (iv) le solvant à l'aide du modèle de continuum COSMO. Cette stratégie nous a permis de mettre en évidence le rôle de ces différents éléments. Nous avons pu montrer que chacun de ces éléments joue un rôle non négligeable dans la basicité des atomes de soufre du site actif. Ainsi, le squelette peptidique, du fait de contraintes stériques et de liaisons hydrogène avec les atomes de soufre via ses NH peptidiques, diminue globalement la réactivité du site et différencie les différents groupes thiolates, Cys49 se trouvant alors le site le plus réactif. L'inclusion d'une chaîne latérale proche de Cys39 réduit là encore la réactivité globale du site ainsi que, plus fortement, la réactivité de S(Cys39). Nous avons ainsi montré sur cet exemple que Lys41, par l'intermédiaire de son bras ammonium, joue un rôle local (sur S(Cys39)) et un rôle global (sur l'ensemble du site). L'inclusion de toutes les chaînes latérales conduit à de multiples conformations qui présentent des réactivités différentes du fait de la présence plus ou moins proche de groupes chargés au voisinage des différents motifs zinc-thiolates. En nous limitant à une étude statique, nous avons néanmoins pu montrer que Cys49 restait le site le plus réactif. La réactivité relative entre les deux autres sites n'a pu être tranchée définitivement. Enfin, le solvant modère l'effet des différentes chaînes latérales sans modifier les tendances observées en phase gazeuse.

Bibliographie

- Aldovini, A.; Young, R. A. *J. Virol.* **1990**, 64, 1920.
- Berg, J. M.; Shi, Y. *Science* **1996**, 271, 1081.
- Berg, J. M.; Godwin, H. A. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, 26, 357.
- Bulaj, G.; Kortemme, T.; Goldenberg, D. *Biochemistry* **1998**, 37, 8965.
- Choo, Y.; Klug, A. *Curr. Opin. Struct. Biol.* **1997**, 7, 117.
- Cohen, J. *Science*, **1997**, 277, 32.
- Copeland, T. D.; Morgan, M. A.; Oroszlan, S. *Virology* **1984**, 133, 137.
- Fabris, D.; Zaia, J.; Hathout, Y.; Fenselau, C. *J. Am. Chem. Soc.*, **1996**, 118, 12242.
- Fabris, D.; Hathout, Y.; Fenselau, C. *Inorg. Chem.*, **1999**, 38, 1322.
- Hathout, Y.; Fabris, D.; Han, M. S.; Sowder, R. C.; Henderson, L. E.; Fenselau, C. *Drug Metabol. Dispos.* **1996**, 24, 1395.
- Henderson, L. E.; Copeland, T. D.; Sowder, R. C.; Smythers, G.W.; Oroszlan, S. *J. Biol. Chem.* **1981**, 256, 8400.
- Henderson, L. E.; Bowers, M. A.; Sowder, R. C.; Sarabyn, S. A.; Johnson, D. G.; Bess, J. W.; Arthur, L. O.; Bryant, D. K.; Fenselau, C. *J. Virol.* **1992**, 66, 1856.
- Huang, M.; Maynard, A.; Turpin, J. A.; Graham, L.; Janini, G. M.; Covell, D. G.; Rice, W. G. *J. Med. Chem.* **1998**, 41, 1371.
- Jenkins, L. M. M.; Hara, T.; Durell, S. R.; Hayashi, R.; Inman, J. K.; Piquemal, J. P.; Gresh, N.; Appella, E. *J. Am. Chem. Soc.* **2007**, 129, 11067.
- Klug, A.; Schwabe, J. W. R. *Nat. Struct. Biol.* **1994**, 1, 345.
- Koslover, E. F.; Wales, D. J. *J. Chem. Phys.*, **2007**, 127, 234105.
- Lee, B. M.; De Guzman, R. N.; Truner, B. G.; Tjandra, N.; Summers, M. F. *J. Mol. Biol.* **1998**, 279, 633.
- Lever, A.; Gottlinger, H.; Haseltine, W.; Sodroski, J. *J. Virol.* **1989**, 63, 4085.
- Loo, J. A.; Holler, T. P.; Sanchez, J.; Gogliotti, R.; Maloney, L.; Reily, M. D. *J. Med. Chem.* **1996**, 39, 4313.
- Maynard, A. T.; Huang, M.; Rice, W. G.; Covell, D. G. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, 95, 11578.
- Maynard, A. T.; Covell, D. G. *J. Am. Chem. Soc.* **2001**, 123, 1047.
- Mély, Y.; Cornille, F.; Fournié-Zaluski, M. C.; Darlix, J. L.; Roques, B. P.; Gérard, D. *Biopolymers* **1991**, 31, 899.
- Morellet, N.; Jullian, N.; De Roquigny, H.; Maignet, B.; Darlix, J. L.; Roques, B. P. *EMBO J.* **1992**, 11, 3059.
- Morellet, N.; Demene, H.; Teilleux, V.; Huynh-Dinh, T.; de Rocquigny, H.; Fournie-Zaluski, M.-C.; Roques, B.P. **2000**, résultats non publiés.
- Omichinski, J. G.; Clore, G. M.; Sakaguchi, K.; Appella, E.; Gronenborn, A. M. *FEBS Letters* **1991**, 92, 25.

- Rice, W. G.; Schaeffer, C. A.; Harten, B.; Villinger, F.; South, T. L.; Summers, M. F.; Henderson, L. E.; Bess, J. W.; Arthur, L. O.; McDougal, J. S.; Orloff, S. L.; Mendeleyev, J.; Kun, E. *Nature* **1993**, 361, 473.
- Rice, W. G.; Supko, J. G.; Malspeis, L.; Buckheit, R. W.; Clanton, D.; Bu, M.; Graham, L.; Schaeffer, C. A.; Turpin, J. A.; Domagala, J.; Gogliotti, R.; Bader, J. P.; Hallidays, S. M.; Coren, L.; Sowder, R. C.; Arthur, L. O.; Henderson, L. E. *Science* **1995**, 270, 1194.
- Roques, B. P.; Morellet, N.; De Rocquigny, H.; Déméné, H.; Schueler, W.; Jullian, N. *Biochimie* **1997**, 79, 673.
- South, T. L.; Kim, B.; Hare, D. R.; Summers, M. F. *Biochem. Pharmacol.* **1990**, 40, 123.
- South, T. L.; Blake, P. R.; Hare, D. R.; Summers, M. F. *Biochemistry* **1991**, 30, 6342.
- Summers, M. F.; Henderson, L. E.; Chance, M. R.; Bess, J. W.; South, T. L.; Blake, P. R.; Sagi, I.; Perez-Alvardo, I.; Sowder, R. C.; Hare, D. R.; Arthur, L. O. *Protein Sci.* **1992**, 1, 563.
- Topol, I. A.; Nemukhin, A. V.; Chao, M.; Iyer, L. K.; Tawa, G. J.; Burt, S. K. *J. Am. Chem. Soc.* **2000**, 122, 7087.
- Topol, I. A.; Nemukhin, A. V.; Dobrogorskaya, Y. I.; Burt, S. K. *J. Phys. Chem. B* **2001**, 105, 11341.
- Tozser, J. *Ann. N. Y. Acad. Sci.* **2001**, 946, 145.
- Tummino, P. J.; Scholten, J. D.; Harvey, P. J.; Holler, T. D.; Maloney, L.; Gogliotti, R.; Domagala, J.; Hupe, D. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, 93, 969.
- Wilker, J. J.; Lippard, S. J. *J. Am. Chem. Soc.* **1995**, 117, 8682.
- Wolfe, S.; Greisman, H.; Ramm, E.; Pabo, C. J. *Mol. Biol.* **1999**, 285, 1917.

Conclusions et perspectives

Le but de ce projet était la meilleure compréhension de la chimie des protéines à zinc et plus spécifiquement, de la réaction d'alkylation du motif zinc-thiolate dans les sites "structuraux". A cette fin, nous avons étudié cette réactivité (mécanisme et nucléophilie) sur des complexes biomimétiques à l'aide des méthodes DFT, puis étendu et confirmé les résultats obtenus par la modélisation d'un fragment protéique.

Notre premier travail (Chapitre III) a porté sur la détermination du mécanisme d'alkylation du motif zinc-thiolate au sein des complexes biomimétiques. En effet, les observations expérimentales étaient contradictoires à ce sujet, supportant soit un mécanisme S_N2 dissociatif, soit un mécanisme associatif. Cette divergence dans la littérature a pu être expliquée en prenant en compte l'influence de l'ensemble de la sphère de coordination du zinc sur le mécanisme de la réaction. Pour la première fois en effet, les paramètres structuraux des complexes, à savoir la nature et la charge des différents ligands du métal, ont été mis en relation avec la réactivité nucléophile du complexe. Nous avons montré que les complexes possédant une charge totale neutre ou cationique réagissent suivant un mécanisme associatif de type S_N2 , alors que les complexes dianioniques procèdent à la dissociation du motif zinc-thiolate avant l'alkylation. Ces résultats sont cohérents avec la dissociation observée précédant l'alkylation de $[Zn(SPh)_4]^{2-}$ et le mécanisme associatif déterminé pour de nombreux composés organométalliques tripodaux neutres et cationiques de la littérature. Ces résultats permettent également de postuler un mécanisme pour les sites enzymatiques, celui-ci dépendant de la nature Glu/Asp/Cys/His des ligands.

Concernant la réactivité de ces complexes, nous avons montré que celle-ci s'accroît avec la densité électronique du système, c'est-à-dire avec le nombre de ligands chargés négativement. Ceci est en parfait accord avec l'ordre de réactivité trouvé expérimentalement. La réactivité est également dépendante de l'environnement du site actif. Ainsi, nous avons montré qu'une liaison hydrogène dirigée sur un thiolate lié au zinc réduisait sa nucléophilie. La présence d'une telle interaction ne modifie cependant pas le mécanisme de la réaction.

Cette étude a fait ressortir l'importance de deux caractéristiques physico-chimiques des complexes zinc-soufre vis-à-vis de leur propension à subir une alkylation, à savoir la dissociation possible de la liaison zinc-thiolate et la nucléophilie du soufre. Ces deux aspects ont été étudiés plus particulièrement dans les parties suivantes.

Au vu des résultats du chapitre III, une dissociation aisée du thiolate lié au zinc dans les complexes dianioniques est attendue. Nous avons appréhendé en détail ce phénomène pour les composés $[Zn(SPh)_4]^{2-}$ et $[Zn(SMe)_4]^{2-}$ dans le chapitre IV.

Les calculs quantiques en phase gazeuse ont indiqué que la dissociation d'un de leurs ligands thiolates était exothermique et déviait donc du schéma type d'une dissociation classique de ligand métallique. Ce résultat inattendu nous a conduit à évaluer la précision de diverses méthodes. Ainsi, nous avons vérifié que les fonctionnelles B3LYP et BP86 étaient bien adaptées aux calculs des géométries, malgré une légère surestimation des longueurs de liaisons zinc-soufre dans les

complexes chargés, et qu'elles n'influençaient pas les résultats concernant le phénomène de dissociation.

Cette métastabilité des complexes dianioniques du zinc apparaît néanmoins uniquement en phase gazeuse. En effet, les calculs incluant un effet de solvant ont indiqué un quasi-équilibre entre les complexes du zinc à trois ou quatre ligands thiométhyles et, pour les espèces thiophényles, une préférence pour le complexe dianionique à quatre ligands $[\text{Zn}(\text{SPh})_4]^{2-}$ plutôt que pour le composé tri-coordonné $[\text{Zn}(\text{SPh})_3]^-$. Ceci est en accord avec la plus grande stabilité de $[\text{Zn}(\text{SPh})_4]^{2-}$ par rapport à $[\text{Zn}(\text{SPh})_3]^-$. Les barrières énergétiques de la dissociation déterminées pour les complexes $[\text{Zn}(\text{SMe})_4]^{2-}$ et $[\text{Zn}(\text{SPh})_4]^{2-}$ sont environ de 10 et 15 kcal/mol respectivement et confirment l'existence d'un équilibre avec les espèces tri-coordonnées en solution. L'environnement, à savoir l'inclusion d'un solvant, a donc un impact important sur la thermodynamique de la dissociation. Enfin, il a été montré que si la méthode CPCM échouait dans le calcul de cette énergie relative, un simple point effectué avec la méthode COSMO représentait un moyen rapide et fiable de l'obtenir.

Les causes de cette dissociation aisée ont ensuite été considérées en modifiant l'environnement proche du zinc, à savoir la nature de ses ligands. Il apparaît qu'en phase gazeuse l'effet des charges à proximité du zinc a un impact remarquable sur l'énergie de dissociation du thiolate : une densité électronique trop importante venant des ligands thiolates peut provoquer un effet de saturation pour le cation métallique et l'empêcher d'accepter un ligand supplémentaire. En solution, l'effet stérique des ligands primera sur leur effet électronique. Les ligands possédant des substituants stériquement encombrés favoriseront ainsi la formation d'espèces tri-coordonnées plutôt que tétravalentes, ce qui est tout à fait en accord avec les observations expérimentales, comme par exemple, l'isolation du complexe $[\text{Zn}(\text{SPh}^{\text{Me4}})_3]^-$.

Ces résultats apportent des informations importantes sur la stabilité des sites structuraux ZnCys_4^{2-} des enzymes à zinc : leur tendance à se dissocier est modérée soit par la présence de liaisons hydrogène entre les atomes de soufre et des résidus basiques situés à proximité du métal, permettant ainsi à ces derniers d'accommoder l'excès de charge anionique, soit par la formation de sites polynucléaires à zinc, comme pour la famille des métallothionéines.

Le chapitre V a été consacré à la détermination d'un indice de nucléophilie du motif zinc-thiolate. La barrière énergétique de l'alkylation servant de référence, il a été montré que l'affinité protonique d'un thiolate lié au zinc corrélait avec sa nucléophilie. Cet indice de réactivité a été validé sur une série de seize complexes biomimétiques à zinc, dont la charge totale varie de -2 à +2 et possédant, pour certains, plusieurs thiolates liés au zinc. L'affinité protonique permet non seulement de classer les complexes suivant leur réactivité, mais également de distinguer les différents sites d'alkylation d'un même complexe à zinc suivant leur nucléophilie.

L'étude de la nucléophilie du motif zinc-thiolate à l'aide de cet indice a montré que la réactivité du complexe est fonction des substituants des ligands et surtout de la charge totale du complexe biomimétique, alors qu'elle n'est pas influencée par la nature (N/S) des ligands du zinc comme il avait été suggéré dans la littérature. De plus, la présence d'une liaison hydrogène sur un ligand thiolate

réduit non seulement sa nucléophilie, en accord avec les résultats du chapitre III, mais également la nucléophilie des autres ligands thiolates présents dans cette même sphère de coordination du zinc.

Avec l'affinité protonique comme critère de référence, nous avons pu établir une échelle de réactivité incluant les thiols et les thiolates. Les complexes biomimétiques du zinc présentent une grande amplitude de nucléophilie, ce qui montre l'impact important de l'environnement du zinc sur la modulation de la réactivité du motif zinc-thiolate. Les complexes dianioniques apparaissent plus nucléophiles que les thiolates, alors que les complexes cationiques sont moins réactifs que des thiols. Ceci est cohérent avec l'abondance naturelle des sites anioniques dans les protéines devant assurer une fonction catalytique et la rareté des motifs chargés positivement.

Enfin, l'indice de nucléophilie sélectionné dans la partie précédente nous a permis d'étudier la réactivité d'un site protéique dans le chapitre VI, ainsi que sa sélectivité pour l'alkylation envers un unique thiolate. Nous avons ainsi effectué la modélisation du doigt de zinc du côté C-terminal de la nucléocapside p7 du virus VIH, pour lequel nous pouvions nous appuyer sur des données expérimentales précises. Nous avons montré que la méthode de calcul la plus adaptée pour ce type de système de grande taille, en termes de précision et rapidité, est la méthode MARI-J-BP86/PB employant les coordonnées cartésiennes, qui peut être utilisée aussi bien pour les calculs de géométries que pour l'obtention de l'énergie finale.

Un modèle du doigt de zinc a été construit à partir de la structure $[Zn(SMe)_3(ImMe)]^-$ représentant le site métallique ZnCys₃His, pour introduire progressivement le squelette peptidique, une chaîne latérale chargée à proximité d'une cystéine, puis l'ensemble des chaînes latérales et le solvant. Nous avons déterminé l'influence de chacun de ces paramètres sur la réactivité du site, en déterminant l'affinité protonique des trois cystéinates. En accord avec les résultats précédents, l'influence d'une liaison hydrogène, entre le squelette peptidique ou une chaîne latérale chargée et un thiolate lié au zinc, présente une influence locale (diminution de la nucléophilie du thiolate) et globale (diminution de la réactivité du site). Il a été montré que le squelette peptidique, notamment de par ses contraintes stériques et un éventuel réseau de liaisons hydrogène, a un impact important sur l'ordre de réactivité des thiolates, mais que celui-ci peut difficilement être prédit par le nombre de leurs liaisons H. De plus, l'inclusion de l'ensemble des chaînes latérales mène à de multiples conformations protéiques de réactivités différentes. L'influence des diverses chaînes latérales est fortement atténuée par l'effet du solvant et, à l'exception de celles formant une liaison hydrogène avec un groupe thiolate, leur effet sur la réactivité locale et globale du site est négligeable. Tous les paramètres environnementaux du doigt de zinc contribuent globalement à la diminution de la réactivité du site.

La nucléophilie des cystéinates considérées individuellement est dépendante des conformations des chaînes latérales à leur proximité. La cystéine 49 apparaît néanmoins, selon notre indice de réactivité et quels que soient les paramètres environnementaux pris en compte, comme le site d'alkylation privilégié, ce qui est en accord avec la littérature.

Alors, que les problématiques concernant l'alkylation du motif zinc-thiolate ont été élucidées, telles que le mécanisme de la réaction, la possibilité de dissociation de la liaison zinc-soufre et l'influence de l'environnement sur la nucléophilie du thiolate, les perspectives concernent principalement la réactivité de ce motif dans les métalloprotéines. La modélisation du fragment protéique de la nucléocapside p7 a permis de mieux appréhender la sélectivité de l'alkylation envers un thiolate et l'influence de l'environnement. Cependant, une étude plus exhaustive concernant le mécanisme et/ou l'environnement, éventuellement à l'aide de la dynamique moléculaire, serait à envisager. Il pourrait s'agir, par exemple, d'inclure des molécules de solvant explicites dans le site actif et/ou de tenir compte de la dynamique entre divers domaines protéiques pouvant interagir entre eux. Notamment, l'inclusion du doigt de zinc du côté N-terminal de la nucléocapside p7 pourrait mettre en lumière une éventuelle participation de ses cystéinates, comme proposé dans la littérature. Une étude comparative des doigts de zinc, comprenant l'inventaire des différences structurales et des interactions hydrogène et électrostatiques, serait également nécessaire pour comprendre la préférence pour une alkylation du doigt de zinc côté C-terminal. Les affinités protoniques des six cystéinates de la nucléocapside permettraient l'évaluation de leur nucléophilie relative et également la comparaison avec des motifs zinc-thiolate d'autres sites dont la charge totale est différente. Par exemple, le site ZnCys_4 de la protéine GATA-1, malgré sa nucléophilie intrinsèque plus importante, présente une réactivité moindre, du fait d'un important écrantage stérique et électronique. Il serait particulièrement intéressant de mesurer et mieux cerner l'impact de l'environnement sur la réactivité d'un site à zinc dans ce cas précis.

Les perspectives à plus long terme concernent la modélisation d'inhibiteurs des motifs zinc-thiolate des protéines. Une meilleure compréhension de la catalyse enzymatique des sites à zinc permettront de mieux cibler les paramètres structuraux (groupes fonctionnels, substituants) nécessaires pour inhiber la réaction d'alkylation. Les applications thérapeutiques sont nombreuses et incluent le développement d'agents antiviraux du virus VIH et de molécules anticancéreuses.