



HAL
open science

Préparation et étude structurale d'une thiorédoxine h de triticum aestivum et d'un mutant de la thiorédoxine d'Escherichia coli : développement d'outils pour l'étude structurale des protéines par RMN

Christian Pruvost

► To cite this version:

Christian Pruvost. Préparation et étude structurale d'une thiorédoxine h de triticum aestivum et d'un mutant de la thiorédoxine d'Escherichia coli : développement d'outils pour l'étude structurale des protéines par RMN. Biologie structurale [q-bio.BM]. Université Montpellier 1, 1996. Français. NNT: . tel-02137355

HAL Id: tel-02137355

<https://hal.science/tel-02137355>

Submitted on 22 May 2019

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.

Public Domain

Université de Montpellier I
CNRS - INSERM
UNITES DE FORMATION ET DE RECHERCHE PHARMACEUTIQUES

THESE

présentée
pour obtenir

Le grade de DOCTEUR EN SCIENCES
de l'UNIVERSITE MONTPELLIER I

par
Christian PRUVOST

Sujet: Préparation et étude structurale d'une thiorédoxine h de *Triticum Aestivum* et d'un mutant de la thiorédoxine de *Escherichia Coli*. Développement d'outils pour l'étude structurale des protéines par RMN.

Imprimée le 2 décembre 1996

soutenue le 12 décembre 1996 devant la commission d'examen:

- | | | | |
|-----------------------------|------------------|--------------------|---------------------|
| • Paulette Decottignies | R IBTP Orsay | • Patrick Gilard | L'Oréal |
| • Frédéric de Lamotte-Guéry | INRA Montpellier | • Gérard Grassy | Univ. Montpellier I |
| • Marc-André Delsuc | CBS Montpellier | • Dominique Marion | R IBS Grenoble |

SOMMAIRE :

CHAPITRE I. LA RÉSONANCE MAGNÉTIQUE NUCLÉAIRE :	11
A. INTRODUCTION :	11
B. LE PRINCIPE DE LA RMN :	11
1. L'enregistrement d'un spectre :	12
2. Notions relatives aux spectres enregistrés :	15
3. Les constantes de couplage :	16
C. L'IDENTIFICATION DES SIGNAUX OBTENUS :	19
1. Les systèmes de spins :	20
2. Le rôle de l'attribution :	20
D. L'INFORMATION GÉOMÉTRIQUE OBTENUE :	21
1. Les distances :	21
2. Les angles dièdres :	21
CHAPITRE II. AMÉLIORATION DE LA MÉTHODE DE GÉOMÉTRIE DE DISTANCE :	23
A. INTRODUCTION :	23
B. LA GÉOMÉTRIE DE DISTANCE :	23
1. Théorie de la géométrie de distance :	24
C. UN EXEMPLE :	26
1. Hypothèses et données du problème :	26
2. Résultats :	27
D. LE « PLONGEMENT » :	28
E. LES CONTRAINTES :	30
F. LA PROPAGATIONS DES INÉGALITÉS DE DISTANCE :	31
1. L'inégalité du triangle :	32
2. L'inégalité du tétrangle :	32
G. LA MÉTRISATION :	34
1. La métrisation complète ordonnée :	34
2. La métrisation complète aléatoire :	35
3. La métrisation partielle aléatoire :	35
H. APPLICATION DE LA GÉOMÉTRIE DE DISTANCE AUX MOLÉCULES :	35
1. Le raffinement de structure :	36
2. Le problème des images topologiques :	37
I. ETUDE DE L'EFFET DE L'APPLICATION DES INÉGALITÉS DU TÉTRANGLE :	38
1. Matériel et Méthode :	38
2. Les résultats :	40
3. Discussion :	45
4. Conclusion :	46
J. IDÉE ÉTUDIÉE POUR RECONSTRUIRE DES STRUCTURES PAR RMN :	47
1. Matériel et Méthodes :	47
2. Résultats :	49
3. Le problème du positionnement de la chaîne principale :	52
4. Conclusion :	54
CHAPITRE III. LES THIORÉDOXINES :	57
A. INTRODUCTION :	57
B. LES ACTIVITÉS DIVERSES DES THIORÉDOXINES :	57
C. DIFFÉRENTS TYPES DE THIORÉDOXINES :	59
1. Les thiorédoxines m :	59
2. Les thiorédoxines f :	59
3. Les thiorédoxines h :	59
D. LA STRUCTURE TRIDIMENSIONNELLE DES THIORÉDOXINES :	60
E. LE SITE ACTIF DES THIORÉDOXINES :	61
1. Séquence caractéristique :	61
2. Position dans les structures connues :	62
F. LA THIORÉDOXINE CHEZ <i>ESCHERICHIA COLI</i> :	63

1. Position dans la cellule :	63
2. Aspects Structuraux :	63
3. Relation Structure-Activité :	64
CHAPITRE IV. EFFET DE LA MUTATION EN ALANINE DE LA PROLINE 40 DANS LA THIORÉDOXINE D'ESCHERICHIA COLI : (P40A)	65
A. INTRODUCTION :	65
B. MATÉRIEL ET MÉTHODES :	66
1. Modélisation des structures des thiorédoxines P40A (mutant) et WT (« Wild Type ») :	66
C. RÉSULTATS :	67
1. Analyse de la structure modélisée des thiorédoxines P40A et WT :	67
2. Discussion :	71
D. CONCLUSION :	72
CHAPITRE V. ETUDE DE LA DÉNATURATION THERMIQUE DE P40A PAR RMN :	74
A. INTRODUCTION :	74
B. LES ÉCHANTILLONS EXPÉRIMENTÉS :	74
1. Préparation et purification de P40A :	74
2. Analyse par spectrométrie de masse :	79
3. Analyse par RMN :	80
4. Discussion :	82
C. EXPÉRIENCE MISE EN ŒUVRE :	82
1. Matériel et Méthode :	82
2. Résultats sur la thiorédoxine d'E.Coli (WT) et P40A :	83
3. Analyse :	83
4. Discussion :	89
CHAPITRE VI. ETUDE STRUCTURALE DE P40A PAR RMN :	91
A. INTRODUCTION :	93
B. POUR L'ÉTUDE STRUCTURALE :	93
C. EXPÉRIENCE DE COSY :	93
D. EXPÉRIENCE DE TOCSY :	94
E. EXPÉRIENCE DE NOESY :	94
F. ATTRIBUTION :	95
1. La recherche des systèmes de spins :	95
2. L'attribution séquentielle :	95
G. LA STRUCTURE DE P40A :	95
CHAPITRE VII. LA THIORÉDOXINE H DE TRITICUM AESTIVUM (TRX H) :	96
A. PRÉSENTATION DES PROTÉINES DE BLÉ :	97
B. D'OU VIENT LA PROTÉINE ÉTUDIÉE ? :	97
C. L'INTÉRÊT DE CETTE ÉTUDE :	98
1. Les différences avec les trx connues :	98
D. LA FABRICATION D'UN MODÈLE :	98
1. Présentation :	98
2. La prédiction de structure sur les 1 ^{ers} résidus :	98
3. L'homologie de structure :	100
4. Reconstruction 3D de la trx h par homologie :	102
5. Conclusion :	107
E. PRÉPARATION DE LA TRX H :	107
1. Préparation et purification de Triticum Aestivum :	107
2. Tests d'activité biologique :	119
3. Analyse par spectrométrie de masse :	120
4. conclusion :	121
F. ÉTUDE PAR RÉSONANCE MAGNÉTIQUE NUCLÉAIRE (RMN) :	121
1. Introduction :	121
2. Partie expérimentale :	121
3. Simulation des spectres de RMN de Triticum Aestivum :	127

CHAPITRE VIII. ANNEXES :	133
A. LES DIFFÉRENTS TYPES DE SYSTÈMES DE SPINS :	133
B. LES ALGORITHMES DE GÉOMÉTRIE DE DISTANCE EN MATLAB :	136
1. Les inégalités du triangle :	136
2. Les inégalités du tétrangle :	136
3. La métrisation :	137
4. Le plongement :	138
5. La minimisation :	138
6. Autres :	139
C. LES ALGORITHMES DE GÉOMÉTRIE DE DISTANCE EN FORTRAN :	142
1. Le fichier makefile :	142
2. Le fichier principal dgf :	142
3. Les inégalités du triangle :	145
4. Les inégalités du tétrangle :	145
5. La métrisation :	146
6. Le plongement :	148
7. La minimisation :	149
8. L'inversion de matrice :	152
9. Les tirages aléatoires :	153
10. Autres parties :	155
D. ASPECTS GÉNÉRAUX EN BIOLOGIE :	159
1. La préparation des macromolécules biologiques :	159
2. La biochimie structurale ou l'étude de structure des molécules biologiques :	160
E. LE GEL D'ÉLECTROPHORÈSE SDS PAGE :	161
F. LE « WESTERN BLOTTING GRADE » :	162
G. L'ATTRIBUTION DE P40A :	165
H. PUBLICATION SUR LES EFFETS DE LA MUTATION P40A DANS <i>ESCHERICHIA COLI</i> :	167
I. ATTRIBUTION DE <i>TRITICUM AESTIVUM</i> :	179
CHAPITRE IX. RÉFÉRENCES :	183

Avant-propos et remerciements :

Tout le travail présenté dans ce document a été réalisé au Centre de Biochimie Structurale à Montpellier, au sein de l'équipe de Résonance Magnétique Nucléaire et de Modélisation, en collaboration avec l'Unité de Biochimie et de Biologie Moléculaire des Céréales de l'INRA de Montpellier. Je tiens bien sûr à exprimer ma reconnaissance à Monsieur Marc-André Delsuc, Directeur de Recherche au CNRS, pour tous ses conseils et ses critiques, pour son incroyable gentillesse, y compris dans les moments difficiles et pour sa disponibilité de tous les instants malgré le nombre considérable de sollicitations dont il est l'objet. Ces trois années passées ensemble ont vraiment été d'une richesse culturelle et scientifique sans pareille.

Je tiens à exprimer ma gratitude envers Monsieur Jean-Marc Lhoste, Directeur du Centre de Biochimie Structurale de Montpellier, pour m'avoir accueilli dans son nouveau laboratoire, pour la confiance qu'il m'a témoigné au cours de ces trois années et pour les diverses recommandations dont il m'a fait part.

Je tiens également à remercier tous les autres membres du jury :

Monsieur Dominique Marion, qui a accepté sans hésiter la lourde tâche d'être rapporteur.

Madame Paulette Decottignies, qui a également accepté la tâche laborieuse de rapporteur.

Monsieur Gérard Grassy, qui a accepté de juger ce travail et sans qui je n'aurais pu effectuer de thèse au Centre de Biochimie Structurale.

Monsieur Frédéric de Lamotte-Guéry pour son dynamisme et sa collaboration active à la plupart des travaux présentés dans ce document.

Monsieur Patrick Gilard qui également accepte de juger ce travail.

Un grand merci à ma femme Kamna, à tous mes amis et à mes parents, ma grand-mère qui ont eu également beaucoup de travail et qui n'ont pas ménagé leur soutien au cours de ces dernières années.

Mes remerciements vont également à tous les membres du laboratoire pour leurs conseils, leurs encouragements ou tout simplement leur sympathie de tous les jours : Thérèse Malliavin, Jean-Luc Pons, Laurent Chiche, David Crégut, Marie-Paule Strub, Mireille Boyer, Adrien Cavé, Annie Heitz, André Aumelas, Jean-René Alattia, Christian Dumas, François Hoh, Magali Jullien, André Padilla, Xavier Manival, Christian Roumestand, Jonas Uppenberg, Herman Van Tilbeurgh, Stéphane Vigo, Aziz Yasri.

Je remercie l'Unité de Biochimie et de Biologie Moléculaire des Céréales à l'INRA de Montpellier, qui m'a fourni son soutien tout au long de ces trois années.

Le Centre de Biochimie Structurale:

Le Centre de Biochimie Structurale (600 m²) est une création relativement récente - Mars 1993 - proposée par l'université de Montpellier, fortement soutenue par les collectivités régionales et réalisée en collaboration étroite avec le CNRS et son programme interdisciplinaire de recherche IMABIO (Ingénierie des MACromolécules BIOlogiques). Le CBS est associé au CNRS (Département des Sciences Chimiques), à l'université de Montpellier I (Unité Mixte de Recherche 9955) et à l'INSERM (Unité de recherche U 414).

Le CBS a disposé progressivement des moyens nécessaires à l'étude structurale, dynamique et fonctionnelle des macromolécules biologiques, les protéines en particulier, et de leurs cibles. Il a rassemblé des membres de différents laboratoires du CNRS, de l'INSERM et des Universités de Montpellier I (médecine et pharmacie) et II (sciences) disposant ainsi de personnel de qualité en biochimie des protéines et en biologie moléculaire, ce qui a représenté un point capital pour le développement de la RMN à très haute résolution (en 1993) et de la cristallographie par diffraction des rayons X (en 1994). L'informatique biologique, opérationnelle dès 1993 et bien organisée, a permis au CBS de se spécialiser également vers les méthodes de modélisation moléculaires, statistiques et dynamiques, couvrant l'ensemble des approches structurales et des études des relations structure-activité (RSA).

Le fonctionnement du CBS est associé prioritairement aux laboratoires Montpelliérains pour le choix de ses thèmes de recherche, mais sans exclusivité. Des collaborations nationales et internationales y sont initiées ainsi que des participations industrielles.

Il est donc aujourd'hui fonctionnel et autonome. Il ne lui reste plus qu'à trouver son identité scientifique à travers ses collaborations, mais aussi par l'initiation et le développement de thèmes propres rendus plus aisés par son autonomie sur le plan de la préparation des protéines grâce à la mise en place (en 1995) des moyens nécessaires à la culture cellulaire et la préparation des protéines et des acides nucléiques, en vue de leur cristallisation ou de leur marquage isotopique pour une étude par RMN.



Chapitre I. La résonance magnétique nucléaire :

A. Introduction :

La structure et la fonction des protéines ainsi que celle des acides nucléiques est un thème central en biologie dont l'intérêt s'est accru par le fait que les biopolymères peuvent maintenant être modifiés de façon efficace pour améliorer ou rechercher de nouvelles activités biologiques. La RMN s'est révélée être une méthode de choix pour l'étude de protéines et d'acides nucléiques. Au tout début [Bloch, et al., 1946] [Abragam, 1961], cela était plutôt difficile et lent à cause du matériel. L'intensité du champ magnétique des spectromètres ne permettait pas de telles études. Néanmoins, la RMN fut de plus en plus représentée dans la littérature scientifique avec l'augmentation croissante des champs magnétiques et la découverte de nouvelles séquences d'impulsions. Cependant il aura fallu attendre 1977 [Nagayama, et al., 1977] pour voir la mise au point de la première NOESY et le début des années 1980 pour les études de RMN bidimensionnelle sur des macromolécules biologiques. La RMN est maintenant devenue, avec désormais plusieurs milliers de publications scientifiques par an, une méthode complémentaire à la cristallographie par diffraction des rayons X [Lattman, 1994], les deux méthodes se recouvrant parfois.

La RMN permet entre autre l'étude d'échantillons en solution aqueuse ce qui est très intéressant en particulier lorsqu'on ne sait pas cristalliser l'échantillon pour une étude cristallographique. De plus, dans le cas où l'on sait déterminer la structure spatiale par RMN et que celle-ci a aussi été déterminée en cristallographie, une comparaison intéressante est alors envisageable. En RMN, il est également possible de faire varier facilement les conditions de la solution (pH, température, force ionique, solvant) et d'observer leurs effets. Pour la caractérisation de paramètres de dynamique intramoléculaire, la RMN fournit aussi des valeurs quantitatives sur la fréquence de certains mouvements et il est même possible d'obtenir des informations sur l'échange de protons avec le solvant. Toutes ces informations peuvent alors être utilisées pour essayer de comprendre l'activité de la protéine.

B. Le principe de la RMN :

En RMN haute résolution, un tube de verre contient une solution de la molécule à étudier. Celui-ci est placé dans un champ magnétique B_0 intense et uniforme d'une bonne dizaine de Tesla (~ 14 T pour un champ à 600MHz) et est soumis à des irradiations d'une ou plusieurs radiofréquences. On étudie le signal de retour à l'équilibre pour obtenir des informations. Les informations intéressantes obtenues correspondent à des couplages entre les différents spins de certains atomes. Cet échange d'aimantation entre les spins peut être détecté et analysé.

I. L'enregistrement d'un spectre :

a) Préparation de l'échantillon :

En RMN, il faut avoir un échantillon relativement concentré de l'ordre de 1 à 2 mM. Dans le cas contraire, cela est un inconvénient, surtout pour accéder aux expériences de RMN bidimensionnelles. Une quantité insuffisante de produit signifie qu'il faut accumuler un grand nombre de balayages (« scans » en anglais) pour espérer observer quelques pics de très mauvaise qualité et vite noyés dans le bruit de fond. Néanmoins, il est également judicieux de faire tout aussi attention de ne pas trop concentrer l'échantillon. En effet, dans ce cas, celui-ci devient relativement visqueux et les raies observées s'élargissent ce qui diminue fortement la résolution obtenue.

Le choix de la température pour les expériences de RMN est également un facteur déterminant sur la résolution des raies, une température élevée permet d'obtenir des raies plus fines mais peut aussi détériorer l'échantillon. Cela dépend de ses propriétés physico-chimiques.

Les impuretés de taille très inférieure ne sont pas vraiment gênantes tant qu'elles ne surpassent pas en concentration la molécule que l'on désire étudier. Elles provoquent alors l'apparition de raies très fines.

En RMN, il faut faire attention au solvant utilisé pour solubiliser la molécule car si celui-ci donne un signal, il sera bien difficile de détecter quelque chose là où se trouvera le signal du solvant. Ainsi les solvants contenant en grande quantité les noyaux que l'on désire étudier sont à prescrire dans la mesure du possible. Malheureusement en biochimie, le solvant est pratiquement imposé ; on doit travailler dans l'eau pour respecter au mieux les conditions in vivo de la macromolécule étudiée.

Quant au pH de la solution, il est préférable d'avoir un pH acide de façon à ralentir l'échange des protons amides avec l'eau. En effet, les pics des protons qui s'échangent voient leur intensité diminuer, à cause des diverses méthodes employées pour éliminer le pic de l'eau.

Par ailleurs, il est possible de placer une substance de référence dans l'échantillon pour calibrer l'échelle des fréquences et mesurer par la suite les déplacements chimiques des autres noyaux. Une raie issue des petites molécules organiques utilisées dans le tampon (le TRIS par ex.) peut être aussi utilisée.

b) Le signal de précession libre (Free Induction Decay) :

Après une impulsion d'excitation, un système de spins tend à retourner à l'équilibre, c'est le phénomène de relaxation (voir Notions relatives aux spectres enregistrés). Durant ce retour à l'équilibre un signal de précession libre (FID) est détecté par le spectromètre (Figure 1). C'est le phénomène de résonance magnétique observé.



Figure 1 : Allure du signal de précession libre (FID) enregistré après plusieurs balayages.

Le retour à l'équilibre est gouverné par un terme exponentiel $\exp(-t/T2^*)$, où :

- $T2^* = 1/T2s + 1/T2i$
- $T2s$ est le temps de relaxation spin-spin du système de spins couplés.
- $T2i$ caractérise un déphasage instrumental du spin dû aux inhomogénéités du champ magnétique.

Ce signal de précession libre ne peut pas être interprété directement en terme de déplacements chimiques ou de constantes de couplage. C'est la transformée de Fourier de ce signal qui conduit aux spectres de RMN.

c) La transformée de Fourier :

De façon générale, c'est cette procédure de transformation du signal qui permet de déterminer les fréquences des signaux constituant le signal de précession libre. Elle transforme un signal $h(t)$ enregistré sur une certaine échelle de temps en un signal $H(\omega)$ dans l'échelle des fréquences.

$$H(\nu) = \int_{-\infty}^{+\infty} h(t) e^{-i2\pi\nu t} dt \quad \omega = 2\pi\nu \text{ d'où} \quad H(\omega) = \int_{-\infty}^{+\infty} h(t) e^{-i\omega t} dt$$

$$H(\omega) = \text{Re}(H(\omega)) + i \text{Im}(H(\omega)).$$

$h(t)$ est la fonction dans l'échelle de temps.

$H(\nu)$ est la fonction dans l'échelle des fréquences.

Le signal de précession est composé de sinusoïdes amorties par un terme exponentiel, chacune d'entre elle correspondant au signal de précession d'un noyau.

$$h(t) = \sum_{k=0}^{k=N} A_k e^{i\omega_k t} e^{-t/T_2}$$

Prenons un noyau i, alors

$$\text{Re}(H(\omega)) = \frac{a_i T_2}{1 + ((\omega - \omega_i) T_2)^2} \quad \text{et} \quad \text{Im}(H(\omega)) = \frac{-a_i (\omega - \omega_i) T_2^2}{1 + ((\omega - \omega_i) T_2)^2}$$

Cette fonction $\text{Re}(H(\omega)) + i \cdot \text{Im}(H(\omega))$ est appelée Lorentzienne et est caractérisée par un pic à la fréquence $\omega_i/2\pi$.

L'analyse dans le domaine des fréquences du FID permet donc à priori de retrouver tous les noyaux.

d) Le spectre à 1 dimension :

La transformée de Fourier obtenue à partir de l'enregistrement d'un FID se présente de la façon suivante (Figure 2) :

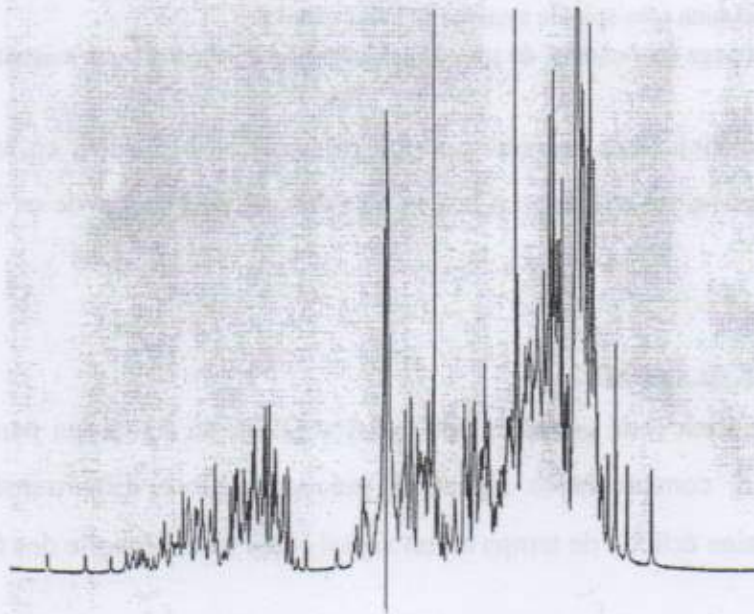


Figure 2 : Spectre à une dimension de protéine obtenu après transformée de Fourier.

e) Le FID et les manipulations de spectres :

Tous les spectromètres RMN commerciaux offrent un programme ou une série de modules permettant les manipulations des spectres ou des signaux de précession libre (FID). Ces opérations ont diverses conséquences sur le spectre final. Elles peuvent être employées par exemple pour éviter la distorsion de la ligne de base, pour augmenter la résolution des raies, ou pour améliorer le rapport

signal sur bruit.

La distorsion de la ligne de base provient du système d'impulsions lui-même ou de la sonde de réception, en particulier avec les phénomènes transitoires (hautes fréquences) avec lesquels il y a parfois saturation du convertisseur analogique (ADC) pour les premiers points. Une ligne de base plate peut alors être obtenue en déplaçant sur la gauche le signal de précession libre et en "oubliant" les premiers points, ce qui revient à ajouter un délai supplémentaire et un déphasage avant l'acquisition des données.

On modifie en général le FID en le multipliant par une exponentielle ou une courbe de sinus de façon à améliorer le rapport signal sur bruit. La fin du FID prend alors une valeur quasiment nulle indiquant un retour à l'équilibre.

Quant à l'augmentation de la résolution elle peut se faire en augmentant le nombre de points du FID par un ajout de zéros à la suite du signal enregistré (« zerofilling »). On effectue alors la transformée de Fourier sur plus de points.

2. Notions relatives aux spectres enregistrés :

a) Les temps de relaxation :

i) La relaxation T1 (relaxation longitudinale) :

Il s'agit du temps de relaxation longitudinal qui décrit la vitesse à laquelle la composante M_z de la magnétisation totale revient à son état d'équilibre thermodynamique le long de $B_0 \hat{z}$ (Figure 3).

$$dM_z(t)/dt = - (M_z(t) - M_0) / T1.$$

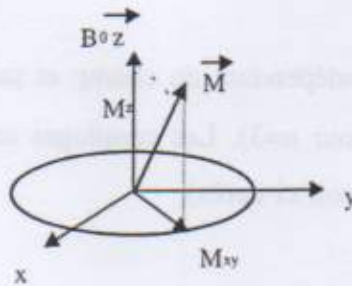


Figure 3 : Représentation du vecteur aimantation dans le référentiel du laboratoire.

ii) La relaxation T2 (relaxation transverse) :

Il s'agit d'une constante de temps qui caractérise la perte de cohérence de phase pour un système de spin donné, après excitation dans un plan perpendiculaire à B_0 , qui amène à une décroissance selon xy de la composante M_{xy} de la magnétisation totale.

$$dM_{xy}(t)/dt = - M_{xy}(t)/T2$$

T2 est directement relié à la largeur du pic de résonance. Pour un pic Lorentzien, $\Delta\nu_{1/2} = 1/(\pi T2)$

b) Le déplacement chimique :

Le déplacement chimique est une grandeur relative à la fréquence de résonance d'un noyau, par définition, le déplacement chimique est un nombre sans dimension correspondant à :

$$\delta_{\text{ppm}} = ((\nu - \nu_{\text{ref}}) / \nu_{\text{ref}}) * 10^6 = \Delta \nu \text{ (Hz)} / \nu_{\text{ref}} \text{ (MHz)}$$

où ν est la fréquence de résonance du pic observé et ν_{ref} est la fréquence de résonance du pic de référence [Emsley, 1966-73].

Convention :

La convention IUPAC définit clairement l'échelle des déplacements chimiques de la façon suivante :

- $\delta > 0$ pour tout pic détecté à une fréquence supérieure à celle de la référence (déblindage, champ faible).
- $\delta < 0$ pour tout pic détecté à une fréquence inférieure à celle de la référence (blindage, champ fort).

Cette grandeur est utilisée pour repérer les fréquences de résonance dans les spectres de RMN.

3. Les constantes de couplage :

En RMN, il est possible d'obtenir des informations sur la structure des molécules grâce à deux types de couplage : le couplage scalaire et le couplage dipolaire qui sont utilisés dans les expériences de RMN à 2 dimensions.

a) Le couplage scalaire :

C'est un couplage qui passe par les orbitales moléculaires qui se recouvrent (les liaisons chimiques). Par exemple, dans le cas de 3 liaisons, on peut définir un angle dièdre θ autour de la liaison N-C α (courbe de Karplus [Pardi, 1984]) et décrire le couplage entre HN et H α . (voir Figure 4).

${}^nJ_{AB}$ est un *couplage scalaire* indépendant du champ et sa mesure permet la détermination de certains angles dans les structures (pour $n=3$). Les couplages scalaires sont mis en oeuvre dans les expériences de COSY et de TOCSY (voir ci-après).

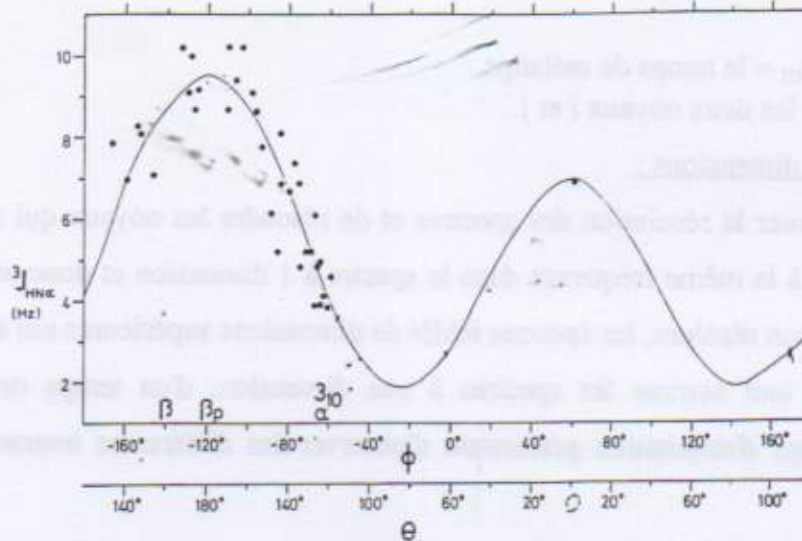


Figure 4 : $^3J_{\text{HN-H}\alpha} = 6.4 \cos^2\theta - 1.4 \cos\theta + 1.9$
 Constante de couplage scalaire entre un hydrogène amide HN et un hydrogène H α

b) Le couplage dipolaire :

Il s'agit là d'un couplage différent ne considérant pas les liaisons chimiques mais simplement la distance entre deux noyaux. Le couplage dipolaire correspond à un couplage entre deux dipôles magnétiques. Son intensité varie en fonction de la distance et du temps laissé à l'aimantation pour se propager entre les deux noyaux (temps de mélange, noté τ_m).

Ce type de couplage est mis en place dans les expériences de NOESY (voir ci-après) pour déterminer des distances interatomiques.

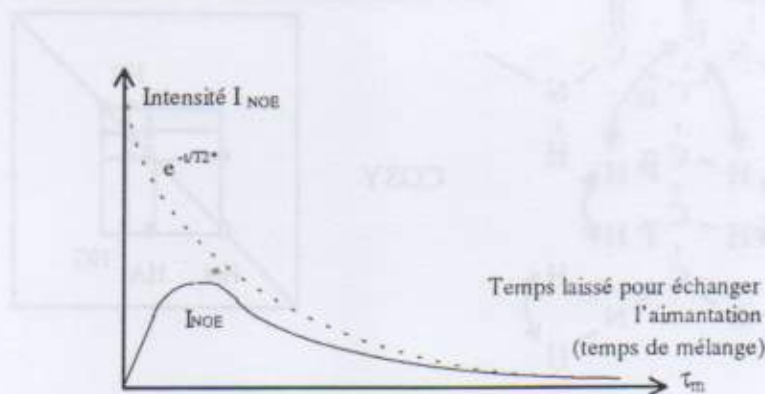


Figure 5 : évolution de l'intensité NOE en fonction du temps de mélange τ_m .

La pente à l'origine de I_{NOE} (Figure 5) en fonction du temps de mélange pour deux noyaux i et j représente la vitesse de relaxation R_{ij} du phénomène d'échange d'aimantation entre i et j .

$$I_{\text{NOE}}(\text{origine}) = R_{ij} \tau_m \quad \text{or} \quad R_{ij} = k/r_{ij}^6 \quad \text{pour une rotation isotopre.}$$

d'où :

$$I_{\text{NOE}}(\text{origine}) = (k \cdot \tau_m) / r_{ij}^6$$

avec :

k = une constante ; τ_m = le temps de mélange.

r_{ij} = la distance entre les deux noyaux i et j .

c) Les spectres à 2 dimensions :

Dans le but d'améliorer la résolution des spectres et de résoudre les noyaux qui sont dégénérés, c'est à dire apparaissant à la même fréquence dans le spectre à 1 dimension et donc avec des pics de fréquences similaires et non résolus, les spectres RMN de dimensions supérieures ont été développés.

Ils sont constitués, tout comme les spectres à une dimension, d'un temps de préparation à l'acquisition et d'un temps d'acquisition permettant d'observer les différentes interactions entre les noyaux.

En RMN homonucléaire, il existe trois types d'expériences de base qui permettent l'étude de structure des macromolécules biologiques. Ces expériences apportent des informations diverses au sein des résidus et aussi entre les résidus. Elles sont décrites ci-dessous.

Dans les expériences de RMN à 2 dimensions, après transformée de Fourier de chacune des dimensions d'enregistrement, le spectre correspond à un plan avec des pics sur une diagonale (équivalents au spectre 1D) et avec des pics hors diagonale correspondants aux couplages scalaires ou dipolaires suivant les expériences de RMN effectuées.

i) L'expérience de COSY :

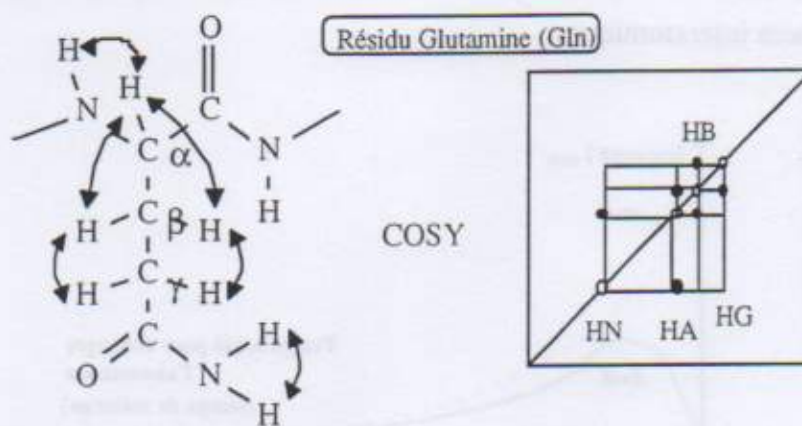


Figure 6 : L'expérience de COSY.

L'expérience de COSY (COrelated SpectroscopY) fait intervenir les couplages scalaires entre les noyaux et dans le cas d'une chaîne d'acides aminés ne permet d'observer que les protons qui se trouvent à une distance de 3 liaisons maximum l'un de l'autre (voir Figure 6).

ii) L'expérience de TOCSY :

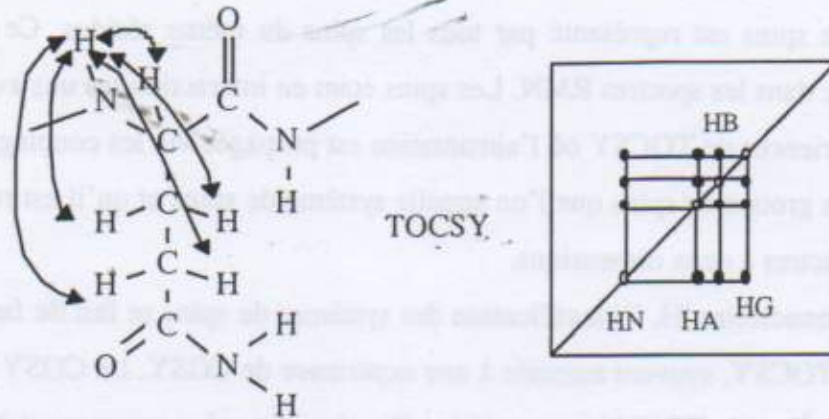


Figure 7 : L'expérience de TOCSY.

L'expérience de TOCSY (Total Correlated SpectroscopY) fait également intervenir les couplages scalaires mais dans cette expérience, l'aimantation peut se propager sur d'autres noyaux eux-même situés à 3 liaisons et ainsi de suite tant que des couplages à une distance de 3 liaisons sont possibles.. Cela permet donc d'observer un couplage entre l'hydrogène amide (HN) de la glutamine ci-dessus et tous les hydrogènes de la chaîne latérale (voir Figure 7).

iii) L'expérience de NOESY :

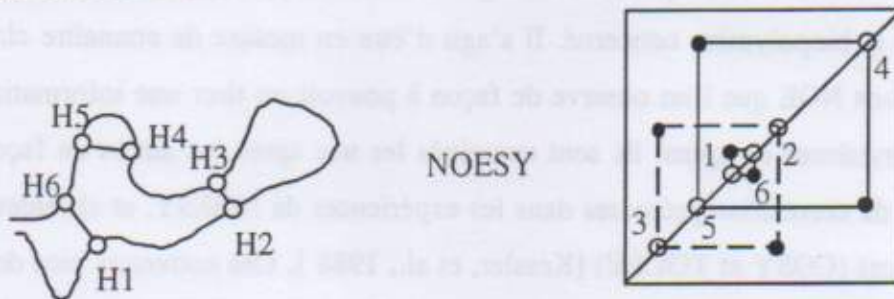


Figure 8 : L'expérience de NOESY.

La NOESY (Nuclear Overhauser Effect SpectroscopY) fait intervenir les couplages dipolaires et permet donc d'observer des échanges entre hydrogènes (dipôles) proches dans l'espace (typiquement à une distance inférieure à 5 Å). c'est l'expérience qui permet de déterminer des distances inter-hydrogènes dès que l'on a une référence dans le spectre via l'Equation 1:

$$r_1/r_2 = \sqrt[6]{I_2(\text{origine})/I_1(\text{origine})}$$

Equation 1

C. l'identification des signaux obtenus :

En RMN, l'étape la plus longue de l'analyse est généralement l'attribution des spectres obtenus. Elle consiste à identifier, grâce à la séquence connue de la molécule, les divers pics obtenus après transformée de Fourier des signaux enregistrés dans les expériences de RMN.

1. Les systèmes de spins :

Un système de spins est représenté par tous les spins du même résidu. Ce sont eux que l'on identifie en premier dans les spectres RMN. Les spins étant en interaction les uns avec les autres, entre autre dans les expériences de TOCSY où l'aimantation est propagée via les couplages scalaires 2J et 3J , ils forment alors un groupe de spins que l'on appelle système de spins et qu'il est relativement aisé de repérer dans les spectres à deux dimensions.

En RMN homonucléaire 1H , l'identification des systèmes de spins se fait de façon simple grâce à une expérience de TOCSY, souvent associée à une expérience de COSY. La COSY permet de détecter les interactions 2J et 3J , et la TOCSY permet l'identification des spins appartenant à un même système de spins (ou résidu). Ainsi, tant que l'on n'a pas de problèmes de dégénérescence, on sait identifier tous les spins et les systèmes correspondants. (voir Les différents types de systèmes de spins : page 133). Les problèmes de dégénérescence apparaissent dès que la taille de la molécule augmente, donc dès que le nombre de résidus augmente. En effet, dans ce cas, il est fortement probable de trouver des spins de même nature dans des environnements électromagnétiques semblables et qui résonnent alors à des fréquences similaires.

2. Le rôle de l'attribution :

Après l'identification des systèmes de spins, nous devons considérer les effets nucléaires Overhauser dans le biopolymère concerné. Il s'agit d'être en mesure de connaître clairement quelles sont les interactions NOE que l'on observe de façon à pouvoir en tirer une information géométrique. Connaissant les systèmes de spins, ils sont examinés les uns après les autres de façon à détecter de nouvelles taches de corrélation présentes dans les expériences de NOESY, et absentes dans les autres types d'expériences (COSY et TOCSY) [Kessler, et al., 1988]. Ces nouveaux pics de corrélation qui sont alors détectés, puis identifiés, mènent à une information géométrique de distance (voir le couplage dipolaire). Un NOE entre deux atomes d'hydrogène (ou groupe d'atomes d'hydrogène) n'est observé que si ces hydrogènes sont proches dans l'espace, à moins de 5 Å l'un de l'autre dans la structure moléculaire. Typiquement un grand nombre de NOE est observé dans les protéines globulaires ou dans les acides nucléiques en double hélice, indiquant bien qu'un certain nombre de paires d'atomes sont proches dans l'espace.

Combiné à l'attribution des résonances, ces contraintes de distances obtenues pour des parties spécifiques le long de la chaîne du biopolymère définissent alors des boucles et des repliements de la chaîne de longueur variable (Figure 9). Avec la construction de modèles et l'utilisation de procédures mathématiques convenables, notamment la géométrie de distance (voir page 23), il est alors possible de rechercher la conformation et l'arrangement spatial du biopolymère.

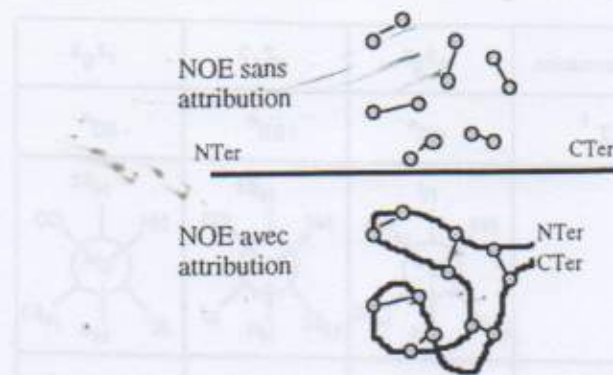


Figure 9 : Information contenue dans les NOE inter-hydrogènes avec et sans attribution. Les cercles représentent les atomes d'hydrogènes de la chaîne du biopolymère.

D. L'information géométrique obtenue :

1. Les distances :

L'information sur les distances, essentielle pour déterminer une famille de structures, s'obtient à partir du spectre NOESY de la façon suivante : Après avoir déterminé les pics intéressants, représentant des interactions inter-résidus et intramoléculaires pouvant donner une information structurale, la distance entre les deux spins couplés est déterminée en fonction de l'intensité du pic observé dans l'expérience [voir Equation 1 page 19]. Pour cela on calibre généralement l'intensité du couplage sur des pics représentant des interactions bien connues entre lesquels on connaît la distance de façon précise, par exemple sur le pic NOE observé pour l'interaction entre eux de deux hydrogènes en alpha ou en bêta d'un groupement carboxyle (GLY, MET, etc.) ou entre deux hydrogènes d'un noyau aromatique.

2. Les angles dièdres :

Dans l'étude de structure de protéines en solution par RMN, le manque d'attributions stéréospécifiques de protons diastéréoisomères tels que les hydrogènes des groupes CH_2 , les méthyles de leucine ou de valine crée en fait une incertitudes sur les contraintes de distance. De telles attributions stéréospécifiques peuvent donc être intéressantes dans certains cas, et en fait sont réalisables en observant la structure fine des pics dans les expériences de COSY associé à l'intensité relative du même pic observé dans une expérience de NOESY. Pour ces protons, l'attribution stéréospécifique [Beckman et al., 1993] [Polshalov et al., 1995] [Cai et al., 1995] [Constantine et al., 1994] est reliée à la définition de l'angle de rotation χ^1 autour de la liaison $\text{C}\alpha\text{-C}\beta$. Les conformations énergétiquement les plus stables pour ces atomes sont les conformations en position éclipsée et il en existe trois qui sont préférentielles et qui permettent donc la détermination des angles dièdres autour de $\text{C}\alpha\text{-C}\beta$ (Figure 10).

Conformation	g^2g^3	g^2t^3	t^2g^3
χ^1	60°	180°	-60°
$^3J_{\alpha\beta^2}$ (Hz)	2.6-5.1	2.6-5.1	11.8-14.0
$^3J_{\alpha\beta^3}$ (Hz)	2.6-5.1	11.8-14.0	2.6-5.1
NOE(α, β^2)	Strong	Strong	Weak
NOE(α, β^3)	Strong	Weak	Strong
NOE(NH, β^2)	Weak	Strong-Medium	Strong
$d(\text{NH}, \beta^2)$ (Å)	3.5-4.0	2.5-3.4	2.2-3.1
NOE(NH, β^3)	Strong-Medium	Strong	Weak
$d(\text{NH}, \beta^3)$ (Å)	2.5-3.4	2.2-3.1	3.5-4.0

Figure 10 : Caractérisation des 3 conformations possibles autour de la liaison $C\alpha-C\beta$. Les orientations gauche et trans sont indiquées par g et t, respectivement. Les positions β^2 et β^3 , sont définies en fonction de la convention IUPAC. Une projection de Newmann est représentée pour chacune des conformations. L'angle χ , est indiqué, ainsi que l'échelle des distances observées.

Chapitre II. Amélioration de la méthode de Géométrie de Distance :

A. Introduction :

La procédure actuelle de détermination de structures par RMN est tributaire de l'attribution des spectres RMN de la macromolécule considérée. Cette étape est en général une étape plutôt ardue et fastidieuse, et pourtant incontournable pour interpréter les pics obtenus dans les expériences de NOESY permettant d'obtenir des distances interatomiques. Bien souvent, cette étape requiert non seulement l'acquisition d'un grand nombre de spectres RMN mais aussi parfois le recours au marquage isotopique (^{15}N ou/et ^{13}C) pour parvenir à lever certaines indéterminations, soit parce que l'on veut observer des détails très précis de la structure, soit parce que la molécule est très grosse et que de nombreux noyaux ont le même déplacement chimique (dégénérescence). Puis, de façon systématique, les distances interatomiques obtenues sont soumises à un algorithme de géométrie de distance pour améliorer les distances obtenues expérimentalement et en même temps vérifier la cohérence des données.

L'algorithme de géométrie de distance [Havel, 1991] [Oshiro and Kuntz, 1993] utilisé est rarement accessible de façon simple dans les programmes l'utilisant. De plus, il est généralement écrit pour un type bien particulier d'application. C'est pourquoi, suivant un souci de compréhension de cet algorithme et de son amélioration, il a été analysé et réécrit de façon à pouvoir être implanté avec ses améliorations dans les programmes développés au laboratoire. Mais tout d'abord il est bon de connaître quelques notions de bases pour bien appréhender la géométrie de distance, c'est ce qui est expliqué dans un premier temps avant la présentation des algorithmes.

Nous avons aussi envisagé une nouvelle méthodologie pour la reconstruction de structures par RMN sans passer par l'étape d'attribution du spectre. Elle est basée sur l'utilisation d'un sous-ensemble d'hydrogènes de la protéine (les hydrogènes amides, les hydrogènes en α , et les hydrogènes en β) et la reconstruction du nuage d'atomes d'hydrogènes par géométrie de distance, ceci précédant la mise en place de tout le squelette de la molécule.

B. La Géométrie de Distance :

Avant de passer à la théorie, il est bon de remarquer que, l'on peut assimiler les différents atomes d'une molécule à des points dans l'espace. Entre ces points, dont les coordonnées sont connues, il est possible possible de calculer des distances et d'écrire un certain nombre d'équations les reliant et

décrivant la structure.

1. Théorie de la géométrie de distance :

Les distances interatomiques d'une molécule, que l'on assimilera désormais à des distances entre des points dans un espace défini, ne peuvent pas avoir simultanément des valeurs arbitraires, mais sont restreintes à des combinaisons très particulières décrivant ce qui est essentiel dans une approche des problèmes de structure par géométrie de distance. Un certain nombre d'équations peuvent alors être mises en place. La forme générale de ces équations a été découverte par A. Cayley en 1841 [Cayley, 1841], mais ne fut pas étudiée avant 1928, date à laquelle K. Menger [Menger, 1928] montra comment la convexité et plusieurs autres propriétés géométriques de base pouvaient être définies et étudiées en termes de distances entre points. Menger proposa alors un nouvel adage pour la géométrie Euclidienne basée sur ces relations [Menger, 1931]. En 1935, Schoenberg découvrait un équivalent de cette caractérisation des distances Euclidiennes [Schoenberg, 1935], et devenait le premier à réaliser la connexion avec les formes bilinéaires. Tous ces travaux furent rassemblés et clarifiés par L. Blumenthal dans "Distance Geometry" [Blumenthal, 1953] qui reste une bonne introduction à ce sujet. Dans les années 1980, la théorie générale des formes bilinéaires définies dans des espaces affines fut découverte pour amener aujourd'hui à une présentation élégante dans laquelle tous les résultats peuvent être présentés de la même façon.

a) Généralités

i) Notion de distance

Soit A l'ensemble des atomes d'une molécule.

Soit $p: A \rightarrow \mathbb{R}^n$, une fonction qui à tout élément de A associe un élément de \mathbb{R}^n , en fait ses coordonnées.

Soit « $\| \cdot \|$ » une norme et « \cdot » un produit scalaire.

$$\forall a, b \in \mathbb{R}^n, \|a\|^2 = a \cdot a.$$

$$\|a+b\| \leq \|a\| + \|b\|.$$

$$0 \leq \|a\|$$

On peut alors définir une autre fonction $d_p: A \times A \rightarrow \mathbb{R}$ qui,

$$\forall a, b \in A, \text{ vérifie } d_p(a, b) = \|p(a) - p(b)\|.$$

Cette fonction est alors appelée *fonction de distance à n dimensions sur A* .

d_p vérifie les propriétés suivantes :

$$\forall a, b \in A, d_p(a, b) = d_p(b, a);$$

$$\forall a, b \in A, d_p(a, b) = 0 \Leftrightarrow a = b;$$

$$\forall a, b \in A, d_p(a, b) \geq 0;$$

Si $\|x\| = \sqrt{x^2}$ est la norme Euclidienne d'un vecteur de l'espace \mathbb{R}^n , alors $d_p: A^2 \rightarrow \mathbb{R}$ est appelée *fonction de distance Euclidienne à n dimensions* et (\mathbb{R}^n, d_p) est appelé *espace Euclidien*.

ii) Inégalités de Cauchy-Schwarz

Les équations utilisées en géométrie de distance dérivent tout naturellement des inégalités de Cauchy-Schwarz:

$$\text{Soit } \|\cdot\| \text{ une norme, } \forall u, v \in \mathbb{R}^n, \|u \cdot v\| \leq \|u\| \cdot \|v\|$$

Soit $d_p: A \times A \rightarrow \mathbb{R}$ une fonction de distance à n dimensions sur A.

Alors d_p vérifie non seulement les propriétés suivantes :

$$\forall a, b \in A, d_p(a, b) = d_p(b, a);$$

$$\forall a, b \in A, d_p(a, b) = 0 \Leftrightarrow a = b;$$

$$\forall a, b \in A, d_p(a, b) \geq 0;$$

mais aussi :

$$\forall a, b, c \in A, d_p(a, b) \leq d_p(a, c) + d_p(c, b). \text{ (propriété utilisée en géométrie de distance)}$$

démonstration:

$$\begin{aligned} d_p(a, b)^2 &= \|p(a) - p(b)\|^2 = \|p(a) - p(c) + p(c) - p(b)\|^2 \\ &= \|p(a) - p(c)\|^2 + \|p(c) - p(b)\|^2 + 2 \cdot \|p(a) - p(c)\| \cdot \|p(c) - p(b)\| \\ &\leq \|p(a) - p(c)\|^2 + \|p(c) - p(b)\|^2 + 2 \|p(a) - p(c)\| \cdot \|p(c) - p(b)\| \rightarrow \text{(Cauchy-Schwarz)} \\ &\leq (\|p(a) - p(c)\| + \|p(c) - p(b)\|)^2 = (d_p(a, c) + d_p(c, b))^2 \end{aligned}$$

d'où $d_p(a, b) \leq d_p(a, c) + d_p(c, b)$ puisque $\forall a, b \in A, d_p(a, b) \geq 0$

qui est reprise dans **les inégalités du triangle**.

$$\forall a, b, c \in A, d_p(a, b) \leq d_p(a, c) + d_p(c, b).$$

Equation 2 : première inégalité du triangle

De la même manière et toujours à partir des inégalités de Cauchy-Schwarz, il est possible de décrire l'inégalité suivante :

$$\forall a, b, c \in A, \|d_p(a, c) - d_p(c, b)\| \leq d_p(a, b).$$

Equation 3 : deuxième inégalité du triangle

On parle donc des inégalités du triangle.

iii) Notion de métrique

Définition d'un espace pré-métrique

Un espace *prémétrique* E_p est un couple (A, d_p) constitué d'un ensemble d'éléments A associés à une fonction $d_p : A \times A \rightarrow \mathbb{R}$ qui satisfait les conditions :

$$\forall a, b \in A, \quad d_p(a, b) = d_p(b, a);$$

$$\forall a, b \in A, \quad d_p(a, b) = 0 \Leftrightarrow a = b;$$

Définition d'un espace semi-métrique et d'un espace métrique :

Si d_p satisfait la condition : $\forall a, b \in A, d_p(a, b) \geq 0$, alors E_p est un espace *semi-métrique*. Et si en plus la fonction d_p satisfait les inégalités du triangle (voir précédemment) alors E_p est appelé espace *métrique*. Dans ce dernier cas, $d_p(a, b)$ est la distance de a à b ; et si N est le nombre d'éléments de A , il est alors possible d'indexer ces éléments de la façon suivante : $A = [a_1, a_2, \dots, a_N]$, et de construire la matrice (symétrique) des distances représentant le couple (A, d_p) .

$$M_{d_p}(A) = \begin{bmatrix} 0 & d_p(a_1, a_2) & \dots & d_p(a_1, a_N) \\ d_p(a_2, a_1) & 0 & \dots & d_p(a_2, a_N) \\ \dots & \dots & \dots & \dots \\ d_p(a_N, a_1) & d_p(a_N, a_2) & \dots & 0 \end{bmatrix}$$

C. Un exemple :

1. Hypothèses et données du problème :

- Soit un ensemble $A = [a_1, a_2, a_3, a_4]$ de 4 atomes dont on connaît les distances entre eux sauf la distance X entre a_1 et a_4 .
- Et soit $d : A \times A \rightarrow \mathbb{R}$ une fonction de distance Euclidienne à n dimensions sur A qui à tous points a_i et a_j de A associe la distance $d(a_i, a_j) = \sqrt{\sum_{i,j} (x_i - x_j)^2}$.

Soit alors la matrice de distances associée :

$$M_d(A) = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0 & 1 & \sqrt{2} & X \\ 1 & 0 & 1 & \sqrt{2} \\ \sqrt{2} & 1 & 0 & 1 \\ X & \sqrt{2} & 1 & 0 \end{bmatrix} \end{matrix}$$

- L'espace E , dans lequel l'objet doit être décrit par $M_d(A)$, est alors défini comme étant l'ensemble des éléments de A associés à la distance Euclidienne d ; $E = (A, d)$.

Que se passe-t-il suivant les valeurs prises par X ?

2. Résultats :

- Si $X < 0$, l'espace E est pré-métrique mais pas semi-métrique. Les inégalités du triangle ne sont pas vérifiées.
- Si $X = 0$, l'espace E est pré-métrique si et seulement si $a_1 = a_4$, mais c'est impossible car dans ce cas $d(a_1, a_2) = d(a_4, a_2)$, or $\sqrt{2} \neq 1$ donc l'espace n'est pas pré-métrique.
- Si $0 < X < (\sqrt{2} - 1)$, l'espace E est semi-métrique sans être métrique puisque l'inégalité du triangle n'est pas respectée ($d(a_1, a_2) + d(a_1, a_4) < d(a_2, a_4)$). [$\max(1 + X) < \sqrt{2}$].
- Si $(\sqrt{2} - 1) \leq X < 1$, l'espace E est alors un espace métrique mais n'est pas un espace Euclidien (bien que respectant toutes les inégalités du triangle), puisque $d(a_1, a_4)$ est toujours plus petite que la valeur atteinte lorsque l'angle dièdre θ autour de 2-3 est nul (voir Figure 11). Cet espace n'est pas « projetable » (voir Le Plongement) dans R^n .

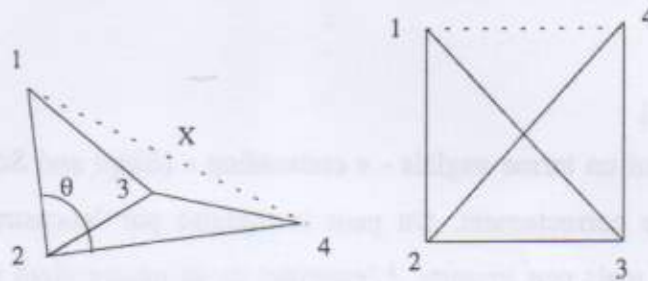


Figure 11 : Figure explicative sur la position des 4 points dans l'espace E

- Si $X = 1$, alors E est Euclidien et « projetable » dans R^2 sous la forme d'un carré
En fait 4 points peuvent toujours être représentés dans un espace à 3 dimensions, tout comme 3 points peuvent être représentés dans un espace à 2 dimensions et 2 points dans un espace à 1 dimension, sans perte de généralité.
- Si $\sqrt{5} \geq X > 1$, E est un espace Euclidien à 3 dimensions.

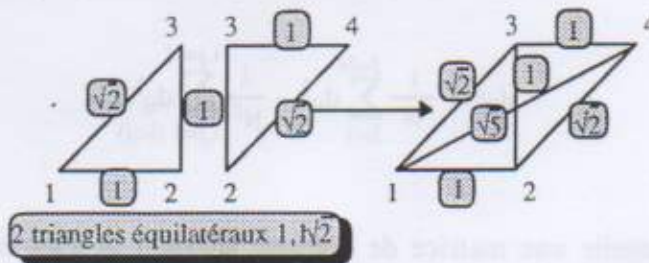


Figure 12 : Représentation des 4 points dans un espace Euclidien.

La plus grande distance entre a_1 et a_4 se recalcule via la distance Euclidienne, c'est au maximum deux fois la distance entre un des sommets du triangle et le milieu de l'arête opposée.

$$[2*(1^2+0.5^2)^{1/2} = \sqrt{5}]$$

- Si $\sqrt{5} < X \leq (\sqrt{2} + 1)$, l'espace E considéré n'est pas Euclidien. Mais les inégalités du triangle sont toujours respectées. Nous sommes toujours dans un espace métrique. Mais dans ce cas, l'ensemble représenté par A n'est plus un objet Euclidien, c'est un objet métrique.
- Si $X > (\sqrt{2} + 1)$, E n'est plus un espace métrique.

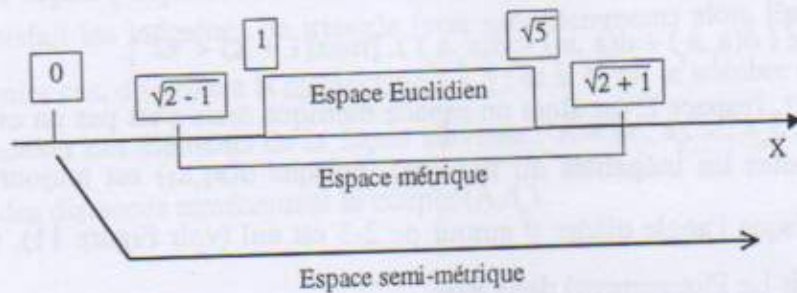


Figure 13 : Correspondance entre E et les valeurs possibles de X

D. Le « plongement » :

C'est au premier abord un terme anglais - « embedding » [Sippl and Scheraga, 1985] - qu'il est assez difficile de traduire correctement. On peut le traduire par 'encastrement par projection' ou encore par 'plongement', mais peu importe. L'essentiel de sa nature vient du fait qu'il représente le passage d'une matrice de distance $D=(d_{ij})$ (avec N points) à une matrice de coordonnées dans un espace de dimension p ($p \leq N-1$), matrice à partir de laquelle nous pouvons éventuellement recalculer les distances. Cette étape permet de représenter ainsi le même objet dans un espace de dimension inférieure p, généralement l'espace Euclidien à 3 dimensions pour une structure moléculaire.

Le principe du 'plongement' est le suivant, la matrice de distance D est utilisée pour calculer chacune des distances au carré d_{oi}^2 des points i par rapport au centre de masse O de tous les points (Figure 14).

$$d_{oi}^2 = \frac{1}{N} \sum_{j=1}^{j=N} d_{ij}^2 - \frac{1}{N^2} \sum_{i,j=1}^{i,j=N} d_{ij}^2$$

Puis nous calculons ensuite une matrice de distance $M=(m_{ij})$ appelée *matrice métrique* définie par :

$$m_{ij} = \frac{1}{2} (d_{O_i}^2 + d_{O_j}^2 - d_{ij}^2)$$

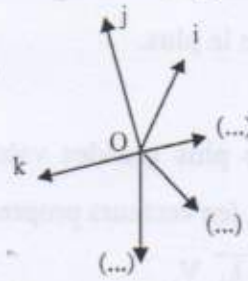


Figure 14 : Les points par rapport au centre de masse

Considérons les 3 vecteurs $O_i \vec{}$, $O_j \vec{}$ et $ij \vec{}$:

$$(ij \vec{})^2 = (-O_i \vec{} + O_j \vec{})^2 = O_i^2 + O_j^2 - 2 O_i \vec{}.O_j \vec{}$$
 donc $O_i \vec{}.O_j \vec{} = \frac{1}{2} (d_{O_i}^2 + d_{O_j}^2 - d_{ij}^2)$

De plus $O_i \vec{}.O_j \vec{} = O_i.O_j \cos(\theta_{ij})$ ainsi (m_{ij}) est une *matrice de produits scalaires*.

Si a_{ik} sont les coordonnées de $O_i \vec{}$ dans la base des $uk \vec{}$ on peut donc décrire m_{ij} :

$$m_{ij} = O_i \vec{}.O_j \vec{} = \sum_{k=1,p} a_{ik}.a_{kj} \text{ avec } O_i \vec{} = \sum_{k=1,p} a_{ik} uk \vec{}$$

Il est alors possible de définir M comme étant le produit $X.X^t$ où X est une matrice $N \times p$ représentant les p coordonnées des N points dans la base des $uk \vec{}$.

M peut être diagonalisée. Il est alors possible de passer de la matrice $M=(m_{ij})$ à une matrice $\Lambda=(\Lambda_{ij})$ via une matrice de passage P , tout en faisant apparaître le produit scalaire :

$$M = (m_{ij}) = P^t . (\Lambda_{ij}) . P = (P^t . (\sqrt{\Lambda_{ij}}) . P) . (P^t . (\sqrt{\Lambda_{ij}}) . P)$$

$$M = X.X^t \text{ avec } X = P^t . (\sqrt{\Lambda_{ij}}) . P$$

Enfin, les N valeurs propres et les N vecteurs propres associés de cette matrice (Λ_{ij}) $N \times N$ et diagonale sont déterminés.

Pour qu'il y ait des racines, il faut que Λ_{ii} soit positif ou nul. Dans ce cas si $\forall i, \Lambda_{ii} \geq 0$ et si les valeurs propres sont positives ou nulles, l'objet est Euclidien et nous pouvons obtenir ses coordonnées dans R^p (voir plus loin). Si une des valeurs propres est négative, cela signifie que l'objet est un objet métrique mais non Euclidien ; nous sommes dans le cas de l'exemple où $(\sqrt{2} - 1) \leq X < 1$ ou bien $\sqrt{5} < X \leq (\sqrt{2} + 1)$.

NB : Λ_{ij} est appelé 'rayon de giration' de M le long de l'axe $O_i \vec{}$.

L'objet, représenté par la matrice initiale de distances D , est alors obtenu dans un espace à p dimensions en calculant les coordonnées de chaque points dans le repère constitué par les p vecteurs propres obtenus correspondants aux p plus grandes valeurs propres. Cela revient un peu à projeter l'objet dans un espace de dimension inférieure le long des axes qui déforment le moins cet objet, le

long desquels l'objet s'étale le plus.

- Soit $\lambda_1, \lambda_2, \dots, \lambda_p$ les p plus grandes valeurs propres trouvées pour une matrice de distance D données et V_1, V_2, \dots, V_p les vecteurs propres associés.

Alors $X_1(i) = \sqrt{\lambda_1} V_1$

$$X_2(i) = \sqrt{\lambda_2} V_2$$

$$X_p(i) = \sqrt{\lambda_p} V_p \quad \text{sont les coordonnées obtenues dans un espace Euclidien à } p \text{ dimensions.}$$

Néanmoins, les choses ne sont pas aussi simples expérimentalement, étant donnés les différents cas vus dans l'exemple. Même si l'objet initial (une molécule par exemple ; dans ce cas $p=3$) est un objet à p dimensions, les erreurs expérimentales et le choix de distances lors de la métrisation peuvent modifier la « nature » de l'objet. Si l'on avait un véritable objet à p dimensions, toutes les valeurs propres calculées devraient être nulles exceptées p d'entre elles. En pratique, on considère les p plus grandes valeurs propres mais il y en a généralement d'autres, l'objet obtenu est alors déformé car il est, en quelque sorte, encastré dans le nouvel espace, exactement comme si on effectuait une projection de l'objet dans un espace de dimension inférieure. Cela se comprend bien lorsque l'on imagine la projection d'un livre (objet 3D) sur un écran (2D), si le livre est projeté selon sa tranche, la projection ne ressemble plus à un livre. C'est pour éviter cela que les objets sont projetés selon leur plus grand rayon de giration et qu'une étape de raffinement de la structure doit être envisagée ensuite.

D'autres cas peuvent se présenter, ce sont souvent des incohérences dans les contraintes initiales qui amène à ce genre de situations, ou tout simplement dans le cas où nous avons un ensemble de données dans un espace métrique sans pour autant être Euclidien (voir l'exemple, nous sommes dans le cas où de nouveau $(\sqrt{2} - 1) \leq X < 1$ ou bien $\sqrt{5} < X \leq (\sqrt{2} + 1)$). En particulier si une des plus grandes valeurs propres est négative, cela indique qu'il y a un sérieux problème sur la réalité Euclidienne de l'objet décrit par la matrice D obtenue.

Ainsi, en sorti de ce type d'*encastrement par projection*, certaines structures doivent être rejetées car elles ne sont pas représentables dans un espace Euclidien et les autres doivent être soumises à un raffinement de structures.

E. Les contraintes :

Ce sont les informations géométriques que l'on peut obtenir expérimentalement et traduire en terme de distances plus ou moins une certaine valeur (erreur expérimentale entre autre) pour une partie de l'ensemble des points représentés. Ces contraintes, que nous appellerons *contraintes de distance* désormais, sont alors représentées par un intervalle de distance, c'est là qu'est tout le problème :

La distance de a à b est dans $[d_0(a,b) - d_{\text{inf}}, d_0(a,b) + d_{\text{sup}}]$

où :

$d_0(a,b)$ est la vraie valeur de la distance de a à b .

d_{inf} est l'écart sur le minimum par rapport à $d_0(a,b)$.

d_{sup} est l'écart sur le maximum par rapport à $d_0(a,b)$.

Ce qui nous amène non plus à une matrice de distances simple $M_{dp}(a,b)$, mais à deux matrices de distances : Celle des limites inférieures $L_{dp}(A)$ et celle des limites supérieures $U_{dp}(A)$ qui sont des matrices symétriques que l'on définit de la façon suivante :

$$L_{dp}(A) = \begin{bmatrix} 0 & dp(a_1, a_2) - d_{\text{inf}}(a_1, a_2) & \dots & dp(a_1, a_n) - d_{\text{inf}}(a_1, a_n) \\ dp(a_1, a_2) - d_{\text{inf}}(a_1, a_2) & 0 & \dots & dp(a_2, a_n) - d_{\text{inf}}(a_2, a_n) \\ \dots & \dots & \dots & \dots \\ dp(a_1, a_n) - d_{\text{inf}}(a_1, a_n) & dp(a_2, a_n) - d_{\text{inf}}(a_2, a_n) & \dots & 0 \end{bmatrix}$$

et

$$U_{dp}(A) = \begin{bmatrix} 0 & dp(a_1, a_2) + d_{\text{sup}}(a_1, a_2) & \dots & dp(a_1, a_n) + d_{\text{sup}}(a_1, a_n) \\ dp(a_1, a_2) + d_{\text{sup}}(a_1, a_2) & 0 & \dots & dp(a_2, a_n) + d_{\text{sup}}(a_2, a_n) \\ \dots & \dots & \dots & \dots \\ dp(a_1, a_n) + d_{\text{sup}}(a_1, a_n) & dp(a_2, a_n) + d_{\text{sup}}(a_2, a_n) & \dots & 0 \end{bmatrix}$$

Généralement, nous ne disposons pas de toutes les informations possibles sur les intervalles de distances entre tous les points. Il convient donc de donner certaines valeurs aux distances inconnues. La valeur inférieure prend la valeur d'une constante que l'on sait être une valeur minimale entre tous les points de l'ensemble, comme par exemple la somme des deux plus petits rayons de Van der Waals dans une molécule. La valeur supérieure est également majorée par une constante qui représente la plus grande distance possible entre deux points. Cela se calcule par exemple en considérant dans une molécule un état non structuré de cette molécule sous sa forme étendue.

F. La propagation des inégalités de distance :

Comme nous l'avons copieusement évoqué dans les paragraphes qui précèdent, l'ensemble des intervalles de distances obtenu est le plus souvent incomplet à cause de la partie expérimentale et pratique, et celui-ci est complété par des intervalles de distances qui sont très grands.

Dans le but d'obtenir une meilleure information sur ces ensembles de distances minimales et

maximales, nous employons certaines inégalités bien particulières issues de l'inégalité de Cauchy-Schwarz.

1. L'inégalité du triangle :

Cette inégalité a été décrite précédemment (voir Théorie de la Géométrie de Distance) et permet de réduire considérablement les intervalles de distances.

Les inégalités du triangle :

$$\forall a,b,c \in \{\text{ensemble des atomes}\}, |d(a,c) - d(c,b)| \leq d(a,b) \leq d(a,c) + d(c,b).$$

Cette relation entre trois points dans l'espace décrit tout simplement le fait que, si « a » est proche de « b », et que si « c » est aussi proche de « b », alors « a » ne peut pas être loin de « c ». Et bien sûr, cette relation est alors répétée entre tous les atomes dans l'espace considéré jusqu'à ce que cette relation géométrique simple soit vérifiée par chacun d'entre eux. La rapidité des calculs évolue en $N \times N \times N (=N^3)$ avec le nombre de points.

2. L'inégalité du tétrangle :

Elle est basée sur les mêmes principes géométriques que l'inégalité du triangle [Easthope and Havel, 1989], mais au lieu de considérer une inégalité entre trois points dans l'espace, elle en considère quatre (Figure 15). Ceci la rend nettement plus performante que l'inégalité du triangle au niveau de la réduction des intervalles de distances décrits par les matrices U_d et L_d , mais les calculs deviennent vite beaucoup plus lourds et varient alors en N^4 avec le nombre de points.

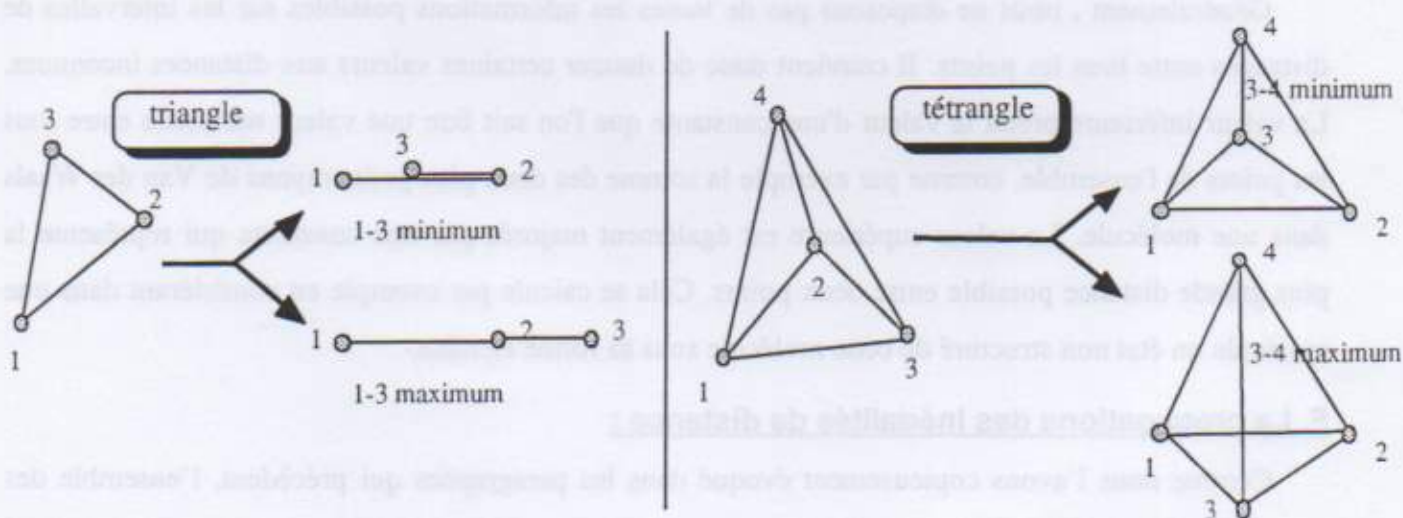


Figure 15 : Schéma comparatif représentant le principe de la méthode des inégalités du triangle et de celle du tétrangle.

L'inégalité du tétrangle [Crippen and Havel, 1988] se présente sous la forme d'un déterminant d'une matrice de distances prises au carré, où la distance (au carré) entre les points a_1 et a_4 est inconnue et représentée par « X ».

$$D_{34}(a_1, a_2, a_3, a_4, X) = 1/8 \cdot \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & d_{12}^2 & d_{13}^2 & d_{14}^2 \\ 1 & d_{21}^2 & 0 & d_{23}^2 & d_{24}^2 \\ 1 & d_{31}^2 & d_{32}^2 & 0 & X \\ 1 & d_{41}^2 & d_{42}^2 & X & 0 \end{vmatrix} \geq 0 \quad \text{Equation 4}$$

$D_{34}(a_1, a_2, a_3, a_4, X)$ est appelé déterminant de Caley-Menger [Purisima and Scheraga, 1986] relatif aux 4 points considérés [Sippl and Scheraga, 1986]. Ce déterminant doit être positif dans toutes les situations qui peuvent se présenter. C'est la définition même de l'inégalité du tétrangle[†].

$D_{34}(a_1, a_2, a_3, a_4, X) \geq 0$ peut-être développé en $A_{34} X^2 + B_{34} X + C_{34} \geq 0$,

où :

$$A_{34} = -1/8 \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & d_{12}^2 \\ 1 & d_{21}^2 & 0 \end{vmatrix}$$

$$B_{34} = -1/4 \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & d_{12}^2 & d_{14}^2 \\ 1 & d_{21}^2 & 0 & d_{24}^2 \\ 1 & d_{31}^2 & d_{32}^2 & 0 \end{vmatrix} \quad C_{34} = 1/8 \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & d_{12}^2 & d_{13}^2 & d_{14}^2 \\ 1 & d_{21}^2 & 0 & d_{23}^2 & d_{24}^2 \\ 1 & d_{31}^2 & d_{32}^2 & 0 & 0 \\ 1 & d_{41}^2 & d_{42}^2 & 0 & 0 \end{vmatrix}$$

A_{34} , B_{34} et C_{34} sont également simplifiables, avec un peu de patience, en :

$$A_{34} = -1/4 (d_{12}^2)$$

$$B_{34} = -1/4 [(d_{12}^2)^2 - d_{12}^2 \cdot (d_{13}^2 + d_{14}^2 + d_{23}^2 + d_{24}^2) - (d_{13}^2 - d_{23}^2) \cdot (d_{24}^2 - d_{14}^2)]$$

$$C_{34} = +1/4 [(d_{13}^2 \cdot d_{24}^2 - d_{14}^2 \cdot d_{23}^2) \cdot (v_1 + v_2) + d_{12}^2 \cdot (v_1 + v_2)] \text{ avec } v_1 = d_{23}^2 - d_{24}^2 \text{ \& } v_2 = d_{14}^2 - d_{13}^2$$

Cette simplification permet un gain de temps considérable dans les calculs puisque le calcul des différents déterminants est plus long que le calcul simplifié sous forme d'additions et de multiplications. Cette étape devant être répétée de nombreuses fois comme nous le verront par la suite, il est important qu'elle soit calculée le plus vite possible.

Il reste maintenant à considérer tous les atomes de la molécule, et à déterminer les racines de l'Equation 4. Deux cas peuvent se présenter : Soit il n'y a pas de racines réelles dans ce cas l'intervalle

de distance entre a_3 et a_4 , soit il y a deux racines et dans ce cas, la plus grande racine devient la borne supérieure et l'autre la borne inférieure. Cependant, cela n'est pas chose facile car toutes les distances entre les différents atomes sont comprises dans un intervalle défini par une matrice des bornes inférieures L_d et une matrice des bornes supérieures U_d . La

Figure 15 ne tient pas compte de ce problème, nous devons en fait calculer les racines de l'Equation 4 pour tous les cas, c'est à dire en considérant les d_{12} , d_{13} , d_{14} , d_{23} , d_{24} tantôt affectés de leur plus grande valeur ou de leur plus petite valeur, ce qui fait un total de 32 situations différentes (toutes les combinaisons d_{12} , d_{13} , d_{14} , d_{23} , d_{24} avec la distance choisie soit dans U_d soit dans L_d).

Cependant, lorsqu'on examine toutes les possibilités, c'est à dire en considérant les maxima et les minima de chacune des distances, il reste uniquement 7 déterminants de Caley-Menger à considérer :

Tétrangle	distance 1-2	distance 1-3	distance 1-4	distance 2-3	distance 2-4
déterminant 1	L_d	U_d	U_d	U_d	U_d
déterminant 2	U_d	L_d	L_d	U_d	U_d
déterminant 3	U_d	U_d	U_d	L_d	L_d
déterminant 4	U_d	U_d	L_d	L_d	U_d
déterminant 5	U_d	L_d	U_d	U_d	L_d
déterminant 6	L_d	L_d	U_d	L_d	U_d
déterminant 7	L_d	U_d	L_d	U_d	L_d

Nous obtenons donc 7 couples de racines (u_i , l_i) et nous en déduisons :

- $U_{34} = \max(u_1, u_2, u_3, u_4, u_5, u_6, u_7)$ la nouvelle borne supérieure pour d_{34} .
- $L_{34} = \min(l_1, l_2, l_3, l_4, l_5, l_6, l_7)$ la nouvelle borne inférieure pour d_{34} .

En effet, il faut prendre le maximum de toutes les racines u_i trouvées de façon à ce que quelque soit la valeur que l'on ait, le déterminant de Caley-Menger relatif aux quatre points ne soit jamais négatif dans aucun des 7 cas, et réciproquement avec les l_i pour le minimum [Crippen et Havel, p.264-267, 1988].

G. La métrisation :

Les matrices U_d et L_d obtenues après avoir propagé l'inégalité sur tous les points représentent alors un intervalle de distances dans lequel il va falloir choisir une distance fixe de façon à obtenir une matrice de distance unique représentant les distances entre tous les points (atomes) de la molécule, c'est l'étape de métrisation.

C'est une étape clef dans les algorithmes de géométrie de distance, car c'est l'étape où nous choisissons les distances dans les intervalles.

I. La métrisation complète ordonnée :

² Un déterminant similaire, à 3 points, peut être développé pour retrouver l'inégalité du triangle.

Les premiers algorithmes de géométrie de distance se contentaient, à cause de la simplicité de cette programmation, de choisir dans l'ordre (eg. $\sum_{i=1, N} (\sum_{j=1, N})$) une valeur aléatoire de la distance dans la fourchette définie, suivi d'une nouvelle propagation des inégalités de la géométrie de distance pour resserrer les intervalles en fonction du choix effectué, et finalement obtenir une matrice de distance D unique. Il s'est avéré que cette façon de procéder crée un biais dans la distribution des distances obtenues et donc dans l'échantillonnage de l'espace des structures. Cette méthode, appelée *métrisation complète ordonnée*, ne doit donc pas être utilisée.

2. La métrisation complète aléatoire :

Cette méthode varie en N^{n+2} avec le nombre N de points, $n=3$ avec les inégalités du triangle, $n=4$ avec les inégalités du tétrangle. Comme il n'y a à priori aucune raison de choisir de fixer une distance entre deux points plutôt qu'entre deux autres, cette décision est prise au hasard. De même, quand une distance est choisie, on doit prendre soin de propager à nouveau les inégalités de distances entre tous les points de façon à améliorer tous les intervalles en fonction des nouvelles valeurs avant de passer au choix, toujours au hasard, de la distance entre deux autres points, et ainsi de suite. Ce processus, appelé *métrisation complète aléatoire*, est réitéré jusqu'à ce que U_d et L_d soit égales. Nous obtenons alors la matrice de distances D pour les atomes de la molécule.

C'est une méthode effectuant des tirages aléatoires où il est judicieux d'effectuer plusieurs métrisations à partir du même jeu de matrices U_d et L_d . Nous obtenons alors une série de matrices de distances différentes à partir desquelles il sera possible, après le « plongement » d'extraire une série de structures.

3. La métrisation partielle aléatoire :

Elle est différente de l'étape de métrisation complète aléatoire dans le sens où l'on ne propage les inégalités de la géométrie de distance que par rapport à p atomes choisis aléatoirement, au lieu de le faire par rapport à tous les atomes aléatoirement. Il s'agit donc d'une *métrisation partielle aléatoire* et cela permet d'avoir des calculs qui sont plus rapides dans la mesure où moins de points sont utilisés, l'évolution se fait alors en $N^{n+1} \times p$ au lieu de N^{2n} .

H. Application de la géométrie de distance aux molécules :

L'idée générale est la suivante ; à partir d'un ensemble de points dans l'espace (typiquement des atomes) pour lesquels nous connaissons des contraintes géométriques telles qu'un ensemble de distances entre certains points, l'algorithme estime toutes les distances encore inconnues et, à partir des matrices de distances calculées par « plongement », construit une structure dans l'espace Euclidien à 3 dimensions. Cette structure peut alors être interprétée.

1. Le raffinement de structure :

Celui-ci s'avère indispensable en sortie d'une étape de géométrie de distance, les différents points des objets ayant des distances entre eux inférieures ou égales à ce qu'elles étaient avant la « projection ». L'étape de raffinement essentielle est celle de minimisation [Glunt, et al., 1993], elle consiste à minimiser une énergie dite de géométrie de distance E_{DG} basée sur les contraintes de distances initialement fournies de façon à ce qu'elles soient à nouveau toutes respectées après le 'plongement'.

$$E_{DG} = \sum_{ij} k_{DG} (d_{ij} - u_{ij})^2 \text{ pour } d_{ij} > u_{ij}$$

$$E_{DG} = \sum_{ij} k_{DG} (l_{ij} - d_{ij})^2 \text{ pour } d_{ij} < l_{ij}$$

$$E_{DG} = 0 \text{ dans les autres cas.}$$

a) La minimisation :

Pour l'étape de minimisation, on peut se contenter de minimiser la valeur de E_{DG} , mais en général dans le cas d'une structure moléculaire, nous faisons intervenir une énergie E basée sur un champ de force bien particulier (GROMOS, CHARMM et AMBER en sont les plus populaires) rendant compte des longueurs de liaisons atomiques, des angles de valence, des angles dièdres, des interactions non covalentes, etc., et de E_{DG} bien sûr.

$$E = E_{\text{van der Waals}} + E_{\text{coulomb}} + E_{\text{dièdres}} + E_{\text{torsion}} + E_{\text{liaison}} + E_{\text{angle}} + E_{\text{croucée}} + E_{DG}$$

Où :

$$E_{\text{liaison}} = \sum_{\text{liaisons}} \frac{1}{2} k_{\text{liaison}} (r - r_0)^2.$$

$$E_{\text{angle}} = \sum_{\text{angles}} \frac{1}{2} k_{\theta} (\theta - \theta_0)^2.$$

$$E_{\text{croucée}} = \sum_{\text{angles liaisons}} \frac{1}{2} k_{r\theta} (r - r_0) (\theta - \theta_0).$$

$$E_{\text{torsion}} = \sum_{\text{torsions}} \frac{1}{2} k_{\xi} (\xi - \xi_0)^2.$$

$$E_{\text{dièdres}} = \sum_{\text{dièdres}} \frac{1}{2} k_{\varphi} [1 + \cos(n\varphi - \delta)].$$

$$E_{\text{van der Waals}} = \sum_{ij} (C_{12}/r_{ij}^{12} - C_{6}/r_{ij}^6) \text{ et } E_{\text{coulomb}} = \sum_{ij} q_i q_j / 4\pi\epsilon_0 \epsilon_r r_{ij}.$$

Diverses méthodes de minimisation, s'appliquant soit à E_{DG} seule ou à E , peuvent être utilisées pour cette étape : Monte-Carlo, gradient, Powel, quasi Newton, Newton-Raphson, etc. [Pruvost Ch., 1992].

b) La dynamique moléculaire :

Nous nous servons ici de la dynamique moléculaire non pas pour regarder comment évoluent les atomes mais pour éviter des miminsa locaux de l'énergie E dans lesquels nous avons de fortes chances de rester en effectuant une simple minimisation.

Toute étape de dynamique moléculaire démarre à partir d'un jeu de coordonnées des atomes, souvent obtenu en sortie d'une étape de minimisation. Des vitesses initiales sont alors données aux atomes suivant une distribution de Maxwell pour la température désirée, et l'évolution de la structure moléculaire complète est étudiée en fonction du temps. Durant cette étape, la température est maintenue constante en réajustant les vitesses v_i des atomes à chaque pas Δt du calcul.

$$v_i \rightarrow \lambda v_i$$

$\lambda = [1 + ((T_0/T) - 1)\Delta t/\tau_T]^{1/2}$ où τ_T est la constante de temps pour le couplage avec le bain à la température constante T_0 .

$$T = 2E_{\text{cinétique}}/(3N - 6)k_B$$

avec k_B la constante de Boltzmann et N le nombre total d'atomes. La valeur de l'énergie cinétique étant calculée à partir de la distribution des masses et des vitesses.

$$E_{\text{cinétique}} = \frac{1}{2} \sum_i m_i v_i^2$$

Les équations du mouvement sont résolues sur de petits intervalles de temps Δt durant lesquels les forces appliquées sont supposées constantes. On en déduit alors l'évolution des vitesses et des atomes (eg. Algorithme de leap-frog [Hockney and Eastwood, 1981]).

Les étapes de dynamique moléculaire sont aussi associées aux contraintes de distances initiales. E_{DG} est donc incluse dans le terme d'énergie et cela permet d'éviter de s'éloigner trop des contraintes initiales dans les simulations. L'ensemble géométrie de distance/ dynamique moléculaire mène généralement à un échantillonnage relativement exhaustif de l'espace conformationnel [Oppenheimer and James, 1989] et permet ainsi l'obtention de 'bonnes' structures.

2. Le problème des images topologiques :

Lorsque l'on travaille en géométrie de distance, on est systématiquement confronté au problème des inversions miroir. En effet une structure et son image miroir sont décrites par le même jeu de distances. Nous savons donc, dans le cas de vraies images miroir, qu'il faudra choisir une structure plutôt qu'une autre en regardant les sens dans lequel tourne les hélices par exemple.

Les images topologiques [Pastore et al., 1991] sont des inversions à un niveau de structure et pas à l'autre. Elles sont différentes des vraies images miroir. Cette inversion vient du fait que nous utilisons des distances. Si, à une certaine échelle, nous n'avons pas toutes les distances (cas de l'expérimentation), il se peut que certaines parties de la structure obtenue soient bonnes et que d'autres soit l'image miroir de ce qu'elle devrait être. A une plus grande échelle, si le jeu de distances utilisé

est trop lacunaire ou vraiment mauvais, on peut observer une rotation de deux hélices α , décrites par les mêmes distances, qui respectent la chiralité des $C\alpha$ mais qui tournent dans des sens inverses l'une de l'autre. Il y a alors une inversion miroir au niveau de la structure secondaire mais pas au niveau de la structure primaire (chiralité). Par ailleurs, cette inversion peut se produire avec 3 hélices dans une structure tertiaire (voir Figure 16).

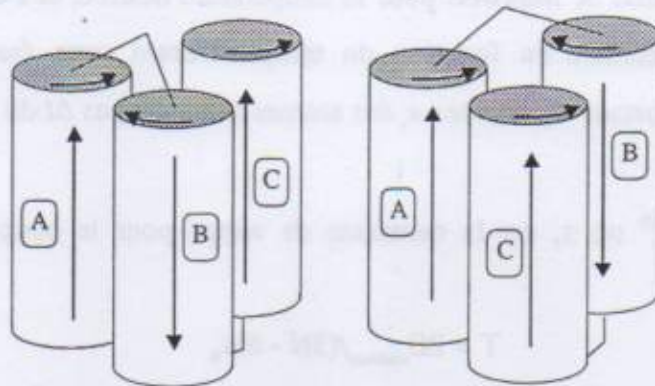


Figure 16 : Les images topologiques - ex : conservation de la structure secondaire avec une inversion au niveau de la structure tertiaire.

Les images topologiques apparaissent dans le cas où nous manquons d'informations. Cela se produit souvent lorsqu'on travaille avec des distances, surtout si les données initiales sont assez lacunaires.

Les images topologiques se repèrent facilement au niveau des hélices α car on peut constater une rotation d'une ou plusieurs hélices dans des sens différents. On peut aussi les repérer grâce à la valeur de leur énergie interne qui est généralement plus grande à cause des tensions internes qui apparaissent quand une partie de la structure est du 'bon côté du miroir' et pas l'autre.

I. Etude de l'effet de l'application des inégalités du tétrangle :

I. Matériel et Méthode :

Les algorithmes écrits concernent l'intégration de contraintes de distances dans les matrices, la propagation des inégalités du triangle et du tétrangle, la métrisation et le « plongement » suivi d'une étape de minimisation (Monte-Carlo Métropolis [Raj, et al., 1994] et Algorithme de « Majorization ». [Glunt, et al., 1993] [Liu, et al., 1992]). Nous nous sommes intéressés au resserrement des différents intervalles de distance en fonction de la méthode employée : Triangulation/Tétrangulation. Dans le but de gagner du temps dans les calculs, le développement de base a été effectué dans Matlab (voir en annexe B) puis a été repris en FORTRAN (voir en annexe C page 135).

Les tests concernant la comparaison Triangulation/Tétrangulation ont été effectués sur une structure 3D très simple de façon à pouvoir comparer les résultats et conserver des vitesses de calcul très raisonnables. La structure choisie est un cube d'arête de longueur 1 pour lequel certaines distances

ont été fixées (voir Figure 17) :

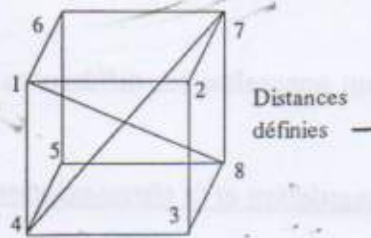


Figure 17 : représentation des distances définies sur l'objet initial (cube).

Nous avons ensuite affiné les distances non définies à partir d'une fourchette de distances très large [0.5-500] en utilisant les deux types d'inégalités.

a) Le développement sous Matlab :

Il n'est présenté qu'en annexe car il fournit les mêmes résultats que le développement en FORTRAN. Néanmoins, il est bon de savoir que Matlab est un outil de travail très performant pour le développement rapide de calculs compliqués, mais qu'il est nécessaire de réécrire les algorithmes dans un langage plus rapide (compilé par ex.). Les calculs effectués dans Matlab (langage interprété) sont extrêmement longs !

b) Le développement en FORTRAN : (voir annexes page 135)

Ce développement peut sembler délicat au premier abord, mais il est relativement aisé quand toute la structure du programme est déjà développée dans Matlab. Le problème du calcul matriciel a été résolu en employant LAPACK [Chergui, 1993], une bibliothèque de calcul matriciel en FORTRAN qui après avoir été installée s'est avérée des plus utiles pour notre travail.

Le but d'un développement algorithmique en FORTRAN est d'arriver à l'obtention de résultats rapidement. En effet nous pensons appliquer ces algorithmes sur des macromolécules (protéines entre autre), nous ne travaillerons alors plus avec huit points mais avec plusieurs centaines. Nous avons en fait gagné un facteur 30 en effectuant nos calculs en FORTRAN au lieu de les faire dans Matlab. Les mois de calculs en Matlab deviennent des jours de calculs en FORTRAN. Ce résultat est plutôt appréciable et permet d'espérer pouvoir travailler sur des macromolécules pour lesquelles les calculs sont bien plus longs.

c) Le type de métrisation employée :

Le type de métrisation employée est la *métrisation totale aléatoire* pour les études sur huit points, néanmoins nous nous sommes aussi intéressés à l'étape de *métrisation partielle aléatoire* car c'est une étape avec laquelle il est possible de gagner du temps. En effet cela s'avère nécessaire car la propagation de l'inégalité du tétrangle est bien plus longue que la propagation de l'inégalité du triangle. De plus, si nous considérons seulement 100 résidus, en fait il faut compter en moyenne 500 hydrogènes, ce qui nous fait cinquante fois plus de points. Nous avons donc effectué une analyse de la

métrisation partielle par rapport à 4 atomes.

2. Les résultats :

Les résultats obtenus font clairement apparaître les différences entre les inégalités du triangle et du tétrangle (voir Figure 18).

La différence observée entre la triangulation et la tétrangulation :

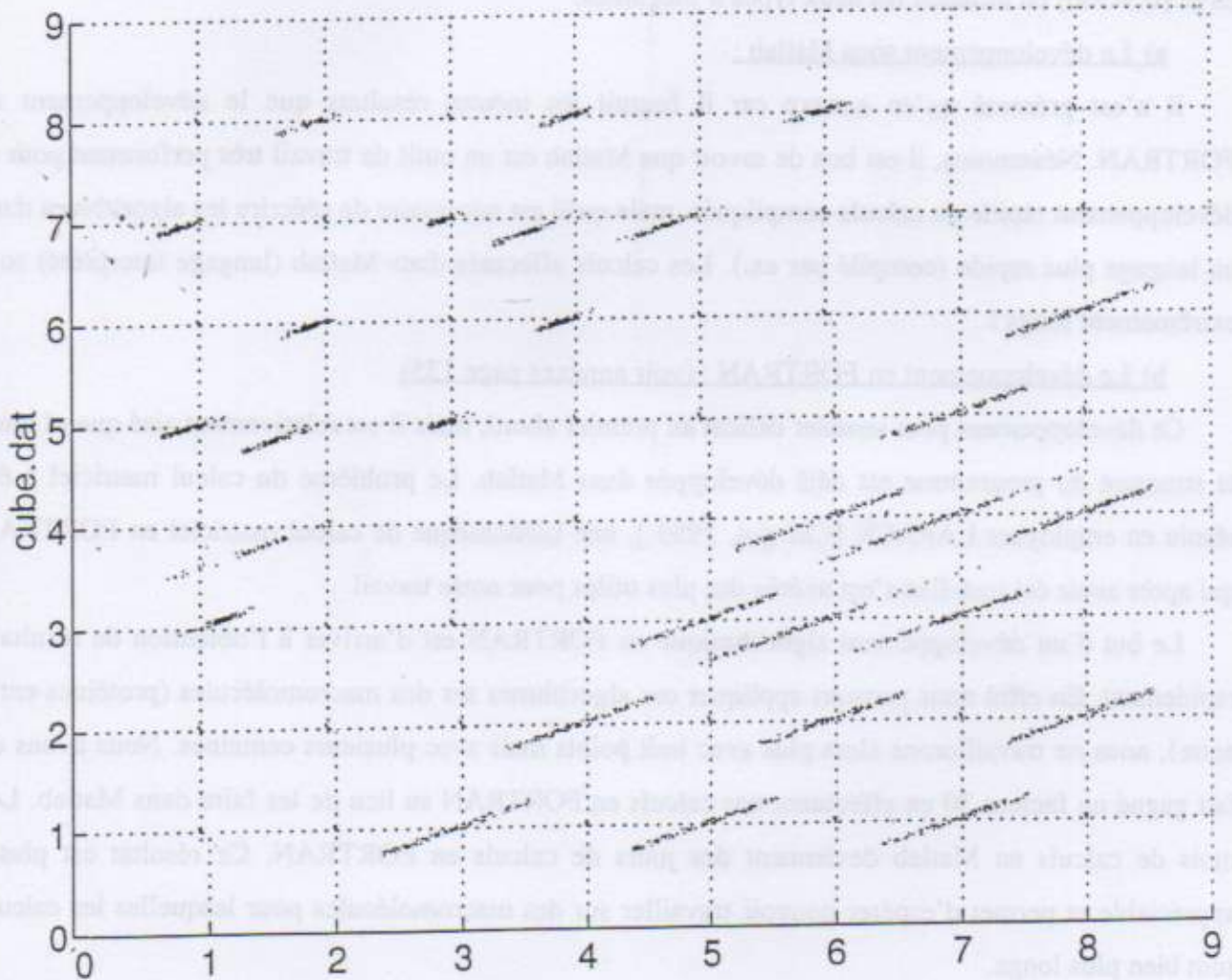


Figure 18 : résultats obtenus après *métrisation complète aléatoire* sur 100 structures calculées. Représentation des résultats de l'inégalité du **tétrangle** dans la moitié supérieure et du **triangle** dans la moitié inférieure. L'étalement de la fourchette de distance, par rapport à la bonne valeur située à l'intersection du quadrillage, est représenté par le nuage des 100 points obtenus.

Dans le cas présent, nous vérifions que les fourchettes de distances obtenues sont plus étroites après propagation de l'inégalité du tétrangle. L'échantillonnage des distances obtenues semble tout à

fait correct dans la majorité des cas mais on constate un écart certain parfois par rapport à la valeur de la distance dans le cube. (par ex : $d(2,4)$ avec le tétrangle ou $d(3,6)$ avec le triangle).

Les valeurs de distance retrouvées pour le cube :

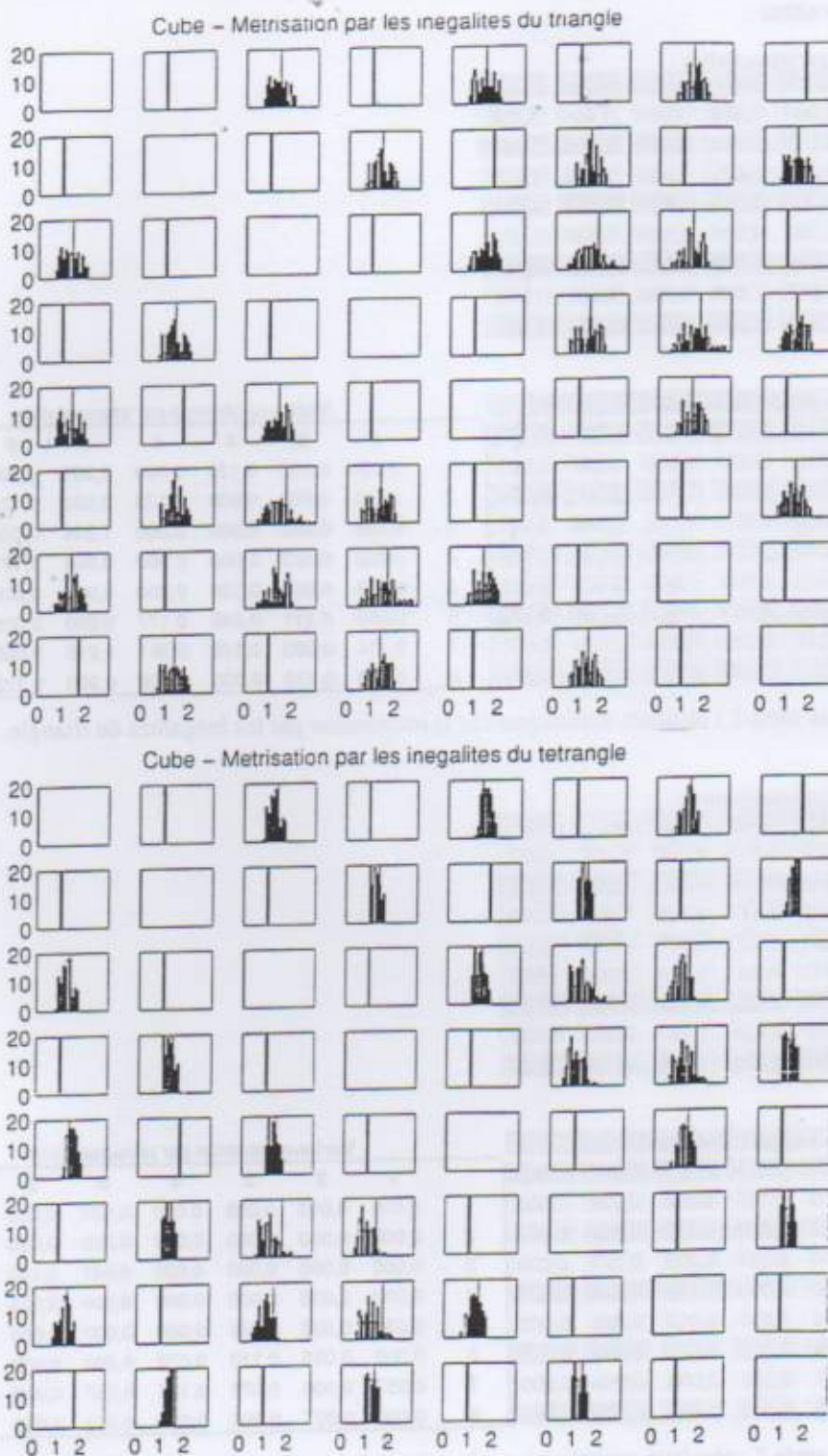


Figure 19 : Histogrammes de la distribution des distances obtenues après métrisation. Cas des inégalités du triangle et du tétrangle. Le nombre de tirage est en ordonnée dans chacun des sous-graphiques en fonction de la distance obtenue. Chacun des sous-graphiques représente de gauche à droite et de bas en haut les points 1 à 8 du cube.

La distribution de distance obtenues fait apparaître cette fois les valeurs trouvées de façon claire. Les valeurs $\sqrt{2}=1.414$ et $\sqrt{3}=1.732$, correspondant respectivement aux diagonales des faces et aux

grandes diagonales non définies initialement sont retrouvées. Nous avons une distribution des distances qui est relativement centrée sur la bonne valeur.

Les statistiques obtenues (Tableaux série 1 et Tableaux série 2) sur les cents jeux de distances recalculés sont les suivantes :

Moyenne des distances obtenues par triangulation :

	1	2	3	4	5	6	7	8
1	0,000	1,000	1,309	1,000	1,335	1,000	1,401	1,732
2	1,000	0,000	1,000	1,347	1,732	1,411	1,000	1,342
3	1,309	1,000	0,000	1,000	1,425	1,416	1,378	1,000
4	1,000	1,347	1,000	0,000	1,000	1,349	1,512	1,429
5	1,335	1,732	1,425	1,000	0,000	1,000	1,436	1,000
6	1,000	1,411	1,416	1,349	1,000	0,000	1,000	1,352
7	1,401	1,000	1,378	1,512	1,436	1,000	0,000	1,000
8	1,732	1,342	1,000	1,429	1,000	1,352	1,000	0,000

Déviaton obtenue par triangulation

	1	2	3	4	5	6	7	8
1	0,000	0,000	0,355	0,000	0,372	0,000	0,337	0,000
2	0,000	0,000	0,000	0,350	0,000	0,334	0,000	0,366
3	0,355	0,000	0,000	0,000	0,353	0,496	0,386	0,000
4	0,000	0,350	0,000	0,000	0,000	0,420	0,511	0,358
5	0,372	0,000	0,353	0,000	0,000	0,000	0,349	0,000
6	0,000	0,334	0,496	0,420	0,000	0,000	0,000	0,349
7	0,337	0,000	0,386	0,511	0,349	0,000	0,000	0,000
8	0,000	0,366	0,000	0,358	0,000	0,349	0,000	0,000

Variance obtenue par triangulation

	1	2	3	4	5	6	7	8
1	0,000	0,000	0,126	0,000	1,383	0,000	0,114	0,000
2	0,000	0,000	0,000	0,123	0,002	0,111	0,000	0,134
3	0,126	0,000	0,000	0,000	1,244	0,246	0,149	0,000
4	0,000	0,123	0,000	0,000	0,000	0,177	0,261	0,128
5	1,383	0,002	1,244	0,000	0,000	0,000	0,121	0,000
6	0,000	0,111	0,246	0,177	0,000	0,000	0,000	0,122
7	0,114	0,000	0,149	0,261	1,215	0,000	0,000	0,000
8	0,000	0,134	0,000	0,128	0,000	0,122	0,000	0,000

Tableaux série 1 : résultats statistiques sur la métrisation par les inégalités du triangle.

Moyenne des distances obtenues par tétrangulation :

	1	2	3	4	5	6	7	8
1	0,000	1,000	1,253	1,000	1,545	1,000	1,307	1,732
2	1,000	0,000	1,000	1,271	1,732	1,290	1,000	1,425
3	1,253	1,000	0,000	1,000	1,317	1,165	1,072	1,000
4	1,000	1,271	1,000	0,000	1,000	0,997	1,212	1,293
5	1,545	1,732	1,317	1,000	0,000	1,000	1,301	1,000
6	1,000	1,290	1,165	0,997	1,000	0,000	1,000	1,281
7	1,307	1,000	1,072	1,212	1,301	1,000	0,000	1,000
8	1,732	1,425	1,000	1,293	1,000	1,281	1,000	0,000

Déviaton obtenue par tétrangulation

	1	2	3	4	5	6	7	8
1	0,000	0,000	0,235	0,000	0,187	0,000	0,238	0,000
2	0,000	0,000	0,000	0,189	0,000	0,187	0,000	0,185
3	0,235	0,000	0,000	0,000	0,202	0,332	0,275	0,000
4	0,000	0,189	0,000	0,000	0,000	0,279	0,348	0,196
5	0,187	0,000	0,202	0,000	0,000	0,000	0,239	0,000
6	0,000	0,187	0,332	0,279	0,000	0,000	0,000	0,198
7	0,238	0,000	0,275	0,348	0,239	0,000	0,000	0,000
8	0,000	0,185	0,000	0,196	0,000	0,198	0,000	0,000

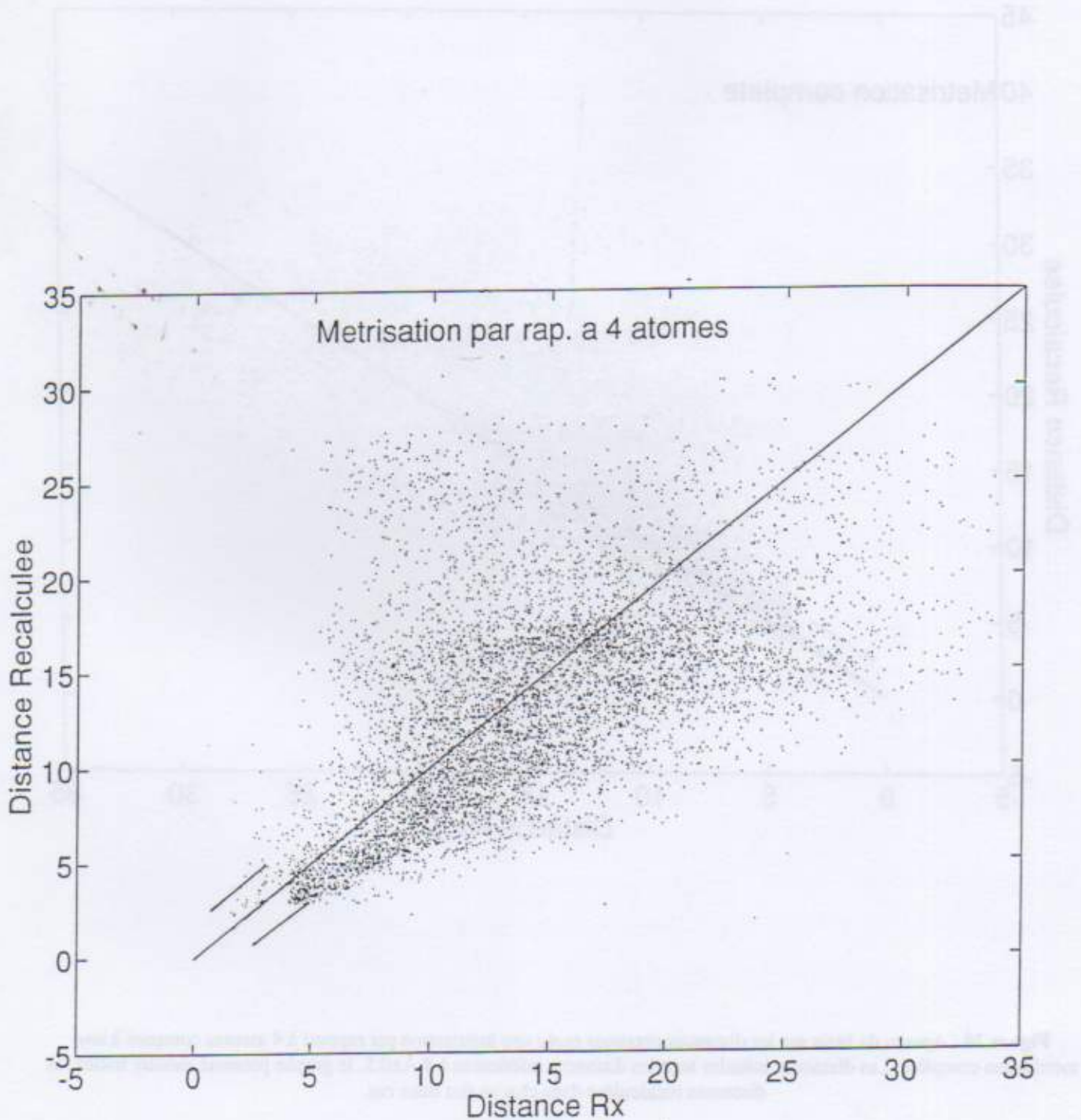
Variance obtenue par tétrangulation

	1	2	3	4	5	6	7	8
1	0,000	0,000	0,055	0,000	0,035	0,000	0,057	0,000
2	0,000	0,000	0,000	0,036	0,000	0,035	0,000	0,027
3	0,055	0,000	0,000	0,000	0,041	0,110	0,075	0,000
4	0,000	0,036	0,000	0,000	0,000	0,078	0,121	0,038
5	0,035	0,000	0,041	0,000	0,000	0,000	0,057	0,000
6	0,000	0,035	0,110	0,078	0,000	0,000	0,000	0,039
7	0,057	0,000	0,075	0,121	0,057	0,000	0,000	0,000
8	0,000	0,027	0,000	0,038	0,000	0,039	0,000	0,000

Tableaux série 2 : résultats statistiques sur la métrisation par les inégalités du tétrangle.

On retrouve une déviaton et une variance plus faible, donc meilleure dans le cas de la propagation par les inégalités du tétrangle.

La méthode de métrisation partielle par rapport à p atomes comporte un biais dans les résultats, elle fournit des distances qui sont globalement trop courtes par rapport à la réalité. Ceci a été mis en évidence sur le BPTI, en propageant l'inégalité du triangle et en métrisant soit complètement, soit par rapport à 4 atomes.



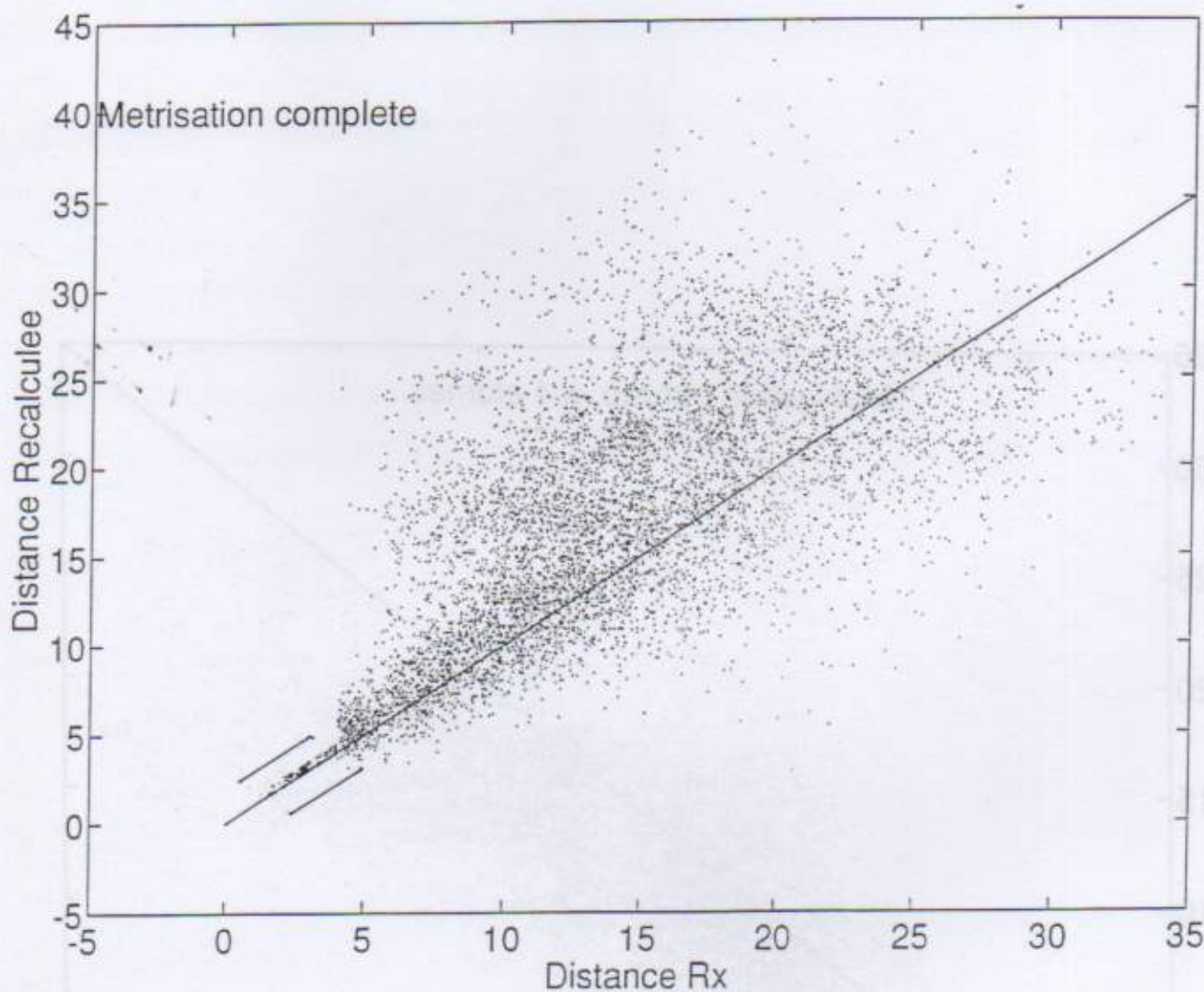


Figure 20 : Aperçu du biais sur les distances obtenues après une métrisation par rapport à 4 atomes comparé à une métrisation complète. Les distances initiales sont les distances inférieures à $4 \text{ \AA} \pm 0.5$, le graphe présenté montre toutes les distances recalculées dans chacun des deux cas.

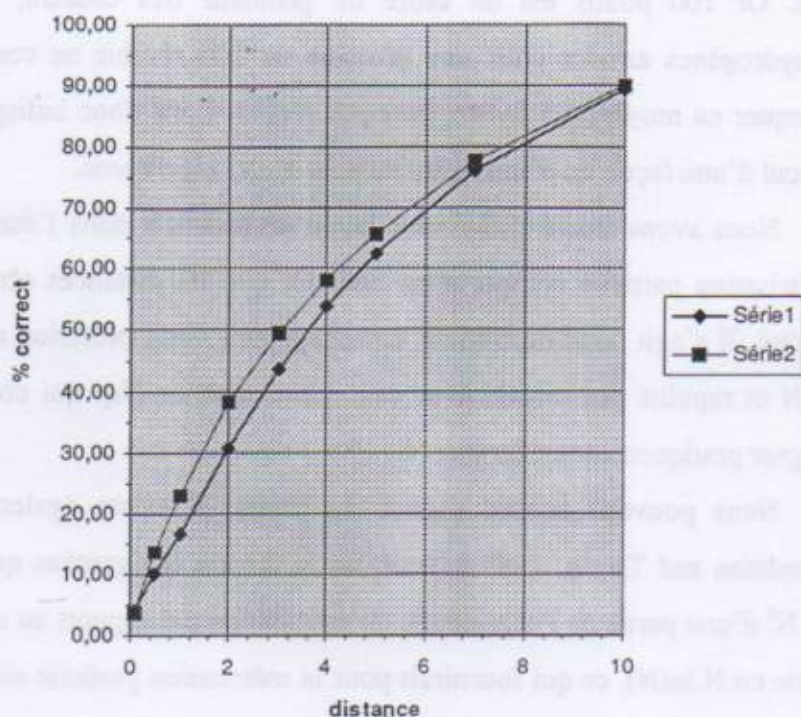
L'écart statistique de la distribution des distances obtenues a été étudié de la façon suivante. Nous avons calculés l'écart entre la valeur de la distance calculée (en ordonnée) et la valeur de la distance dans la structure cristallographique (en abscisse), soit $y-x$. Puis nous avons comptabilisé le nombre de

points vérifiant $|x-y| \leq d_{\text{choisie}}$ où d_{choisie} varie entre 0.1Å et 10Å. Nous avons obtenu les résultats suivants :

métrisation / 4 atomes		Série 1	
d choisie	nbre distances $\leq d$ choisie	d choisie	%
0,1	288	4,10	%
0,5	692	9,86	%
1	1172	16,69	%
2	2136	30,42	%
3	3054	43,50	%
4	3779	53,82	%
5	4389	62,51	%
7	5337	76,01	%
10	6263	69,20	%

métrisation totale		Série 2	
d choisie	nbre distances $\leq d$ choisie	d choisie	%
0,1	250	3,56	%
0,5	954	13,59	%
1	1598	22,76	%
2	2673	38,07	%
3	3474	49,48	%
4	4083	58,15	%
5	4607	65,62	%
7	5453	77,67	%
10	6323	90,06	%

comparaison métrisation totale/partielle



Série2-Série1	-0,54	3,73	6,07	7,65	5,98	4,33	3,10	1,65	0,85
d choisie	0,1	0,5	1	2	3	4	5	7	10

Tableaux série 3 : comparaison sur le BPTI entre métrisation par rapport à 4 atomes et métrisation complète.

Ce résultat indique que le nombre de distances proches de la valeur correcte est plus important lorsqu'une métrisation complète est utilisée. L'évolution de l'écart (Série2-Série1, voir Tableaux série 3) montre donc bien la différence entre la métrisation aléatoire par rapport à 4 atomes et la métrisation aléatoire complète. On peut remarquer qu'il y a jusqu'à 7.6% des distances calculées qui sont plus mauvaises dans le cas d'une métrisation par rapport à 4 atomes et qu'il y a un écart de environ 0.8Å sur ces distances par rapport à celles trouvées en utilisant une métrisation complète.

3. Discussion :

Les résultats obtenus nous montrent que l'inégalité du tétrangle est bien plus performante que celle du triangle. Les fourchettes de distances obtenues sont nettement moins étalées. Les calculs effectués dans Matlab sont extrêmement longs par rapport au FORTRAN. Il y a une grande différence de temps de calcul entre le tétrangle et le triangle, car l'inégalité du triangle varie en N^3 en fonction du nombre N total d'atomes et l'inégalité du tétrangle en N^2 . De plus lorsqu'on y ajoute l'étape indispensable de métrisation totale aléatoire, il faut multiplier le temps de calcul par un facteur N^2 .

puisque l'on choisit la distance d_{ij} [$N(N-1)/2$ choix à faire] et que l'on effectue une nouvelle propagation des inégalités. Ainsi, pour la méthode employant le tétrangle, s'il faut 2min de calcul sur une Hewlett Packard HP735 pour 8 points, il en faudra $2 \cdot (100/8)^6$ pour 100 points, ce qui fait un peu plus de 14,5 ans.⁴ Or 100 points est un ordre de grandeur très courant, c'est juste l'estimation du nombre d'hydrogènes amides pour une protéine de 100 résidus ne comportant pas de proline. Or, il faut compter en moyenne 5 hydrogènes par résidu. Il est donc indispensable de gagner du temps dans le calcul d'une façon ou d'une autre au sein de cet algorithme.

Nous avons étudié différentes façon de procéder dans l'étape de métrisation, il est clair que la métrisation partielle comporte un biais tel que les distances obtenues sont plus courtes que dans la réalité. Il s'agit donc de trouver un compromis entre précision des résultats avec une métrisation en $N \times N$ et rapidité des calculs avec une métrisation en $N \times p$ qui comporte un biais mais qui permet de gagner pratiquement un facteur N .

Nous pouvons encore gagner du temps. Il existe également des méthodes plus complexes [Fredman and Tarjan, 1987] basée sur la théorie des graphes qui permettent de modifier l'évolution en N^2 d'une partie de l'algorithme de métrisation par rapport au nombre N d'atomes en une valeur qui varie en $N \cdot \ln(N)$, ce qui fournirait pour la métrisation partielle aléatoire par rapport à 4 atomes avec le tétrangle une variation en $4 \cdot N^4 \ln(N)$. La durée du calcul serait alors de environ 2 mois et demi pour 100 atomes.

4. Conclusion :

Dans la mesure où de nombreuses structures devront être calculées, il sera nécessaire d'effectuer plusieurs fois les calculs, en évitant autant que possible ce biais causé par la métrisation partielle. Et dans la mesure où nous désirons poursuivre cette étude jusqu'à une métrisation utilisant l'inégalité du pentangle (par rapport à 5 points), il nous faut des moyens de calculs nettement plus performants. Ces travaux sont actuellement à l'étude à l'Institut d'Informatique Parallèle (à Lyon) pour tenter, en parallélisant les calculs notamment l'étape de métrisation, d'obtenir un gain supplémentaire de vitesse ce qui rendrait ces études possibles sur une plus courte échelle de temps. Pour pouvoir bénéficier des effets bénéfiques de l'inégalité du tétrangle, voir du pentangle, par rapport aux inégalités du triangle, il convient donc d'accélérer le déroulement des algorithmes d'une manière ou d'une autre.

J. Idée étudiée pour reconstruire des structures par RMN :

L'idée qui est à la base de cette étude est la reconstruction de structures moléculaires par RMN sans passer par l'étape d'attribution des spectres RMN. Nous avons donc travaillé sur le type d'informations récupérables très facilement à partir des expériences de RMN. Il s'agit de la présence de spins (hydrogènes) à une certaine position dans le spectre RMN et de leur interaction avec un autre spins de fréquence différente. Il est aisé, dans la plupart des cas, de distinguer si l'on a affaire à un HN, un H α ou un H β , ceci sans tenir compte de l'attribution des spectres.

Ainsi, à partir d'interactions observées par RMN entre ces spins, il est possible comme décrit précédemment de positionner entre eux par géométrie de distance l'ensemble des spins considérés. Une fois leur position, ou plutôt les séries de positions, obtenues il ne resterait qu'à trouver les HN ou les H α les plus proches l'un de l'autre deux à deux pour reconstituer un squelette carboné complet qui soit le plus probable pour la molécule.

1. Matériel et Méthodes :

Dans cette optique, nous nous sommes intéressés à l'étude suivante : Nous sommes partis d'une structure connue dont les coordonnées sont disponibles dans la « Protein Data Bank ». Nous avons extrait de ces données les interactions observables en RMN puis nous nous sommes intéressés à la qualité des structures obtenues par géométrie de distance. Nous avons considéré les interactions limitées aux HN entre eux, aux HN avec les H α et aux HN avec les H α et les H β dans une certaine limite de distance (4 ou 5 Å). Le programme employé pour cette étude est X-plor [Brünger, 1988].

a) X-plor3.1 et son module dgsa.inp :

X-plor est un programme constitué d'une série de modules ('fichiers'.inp voir Figure 21) très bien adaptés pour la reconstruction de structures moléculaires compte tenu de contraintes de distances et de contraintes d'angles dièdres, typiquement celles que l'on obtient en RMN. C'est aussi un moteur général pour tous les types de modélisation moléculaire.

Le principe de fonctionnement général du module dgsa.inp et de ses modules satellites est de partir de la séquence de ce que l'on veut reconstruire, d'en préparer une structure complète sous forme étendue puis d'y appliquer les contraintes de distances et d'angles dièdres fournies, améliorées via des étapes de géométrie de distance (inégalités du triangle) pour en déduire finalement une série de structures vérifiant un maximum de contraintes définies, ces structures étant par la suite raffinées via des simulations de dynamiques moléculaires sous contraintes.

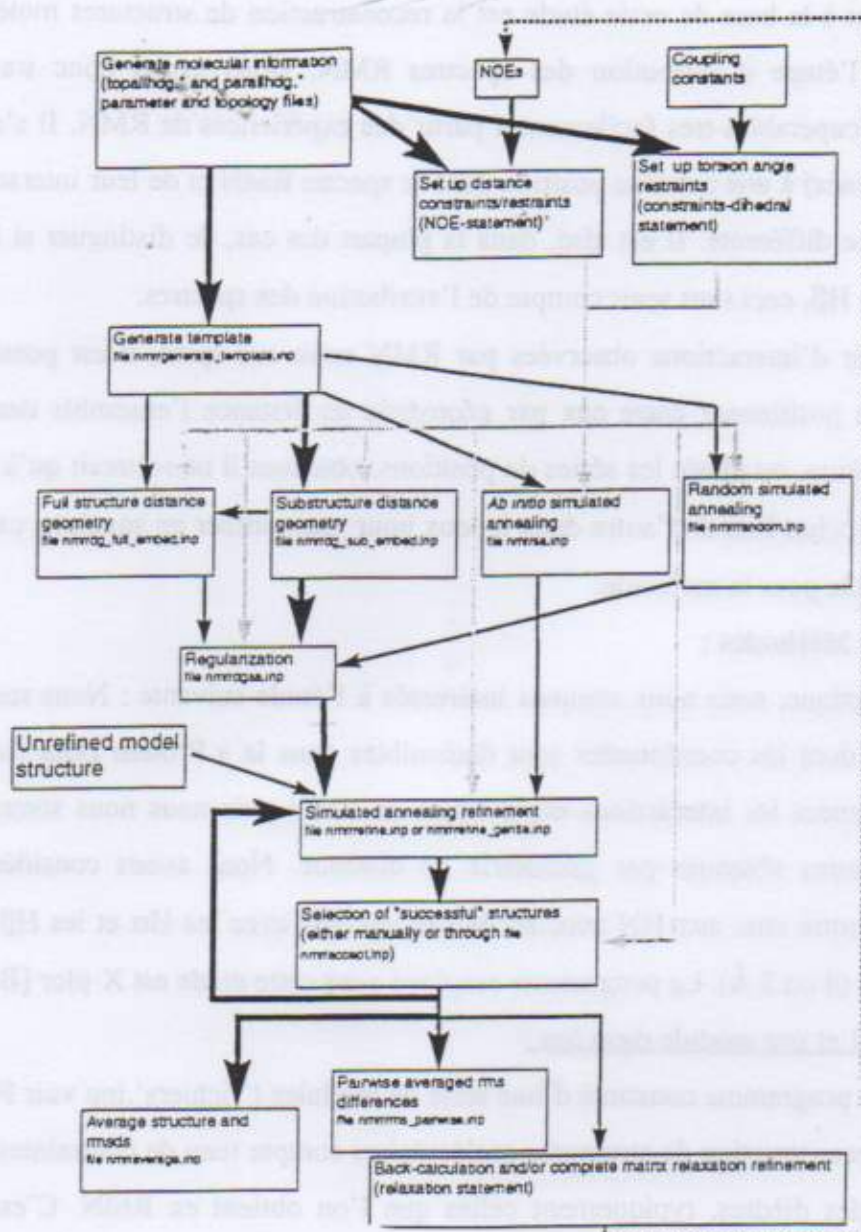


Figure 21 : Vue générale de la détermination de structure RMN dans X-plor3.1

[Brünger, 1988]

b) Le travail effectué :

Comme indiqué ci-dessus, nous avons considéré uniquement certains types d'atomes d'hydrogène, et eux seuls. Nous n'avons donc travaillé qu'avec des atomes non liés entre eux, avec un 'gaz d'atomes' en quelque sorte. Il n'existe donc aucune topologie entre les atomes, ce sont des points dans l'espace.

Dans X-plor, habitué à travailler avec des molécules complètes et une topologie bien définie, nous avons utilisé le langage de commande pour définir une 'topologie de gaz d'atomes' et nos paramètres

pour un point dans l'espace ont été les suivants : Nous avons appelé le point PT et l'avons inclus dans un résidu PTT, aucune charge électrique ne lui a été affecté. A partir d'une structure connue dans la PDB (la parvalbumine - 109 résidus), nous avons récupéré tous les hydrogènes qui nous intéressaient (HN, H α et H β), nous les avons débaptisés en PT et puis nous avons calculé toutes les distances inférieures à 4 Å ou à 5 Å entre chacun de ces points.

Nous avons considéré en tant que contraintes le jeu de distances obtenu soit avec les hydrogènes amides HN seuls, soit avec les HN et les H α , soit avec les HN, H α et H β . Ces jeux de distances ont été considérés comme étant connues à ± 0.5 Å près. Une fois les distributions de distances ayant été optimisées dans les modules de géométrie de distance de X-plor, les nuages de points obtenus ont été comparés par rapport à la position qu'ils occupent réellement dans la structure moléculaire pour examiner si les contraintes de distances fournies en entrée dans le programme sont suffisantes pour retrouver la structure initiale prise dans la base de données (PDB).

Les résultats obtenus avec X-plor ont été comparés avec la structure de départ en utilisant le module de calcul de rms (« root mean square ») présent dans le programme. Ainsi notre critère de qualité des structures est basé sur la racine carré de la somme des carrés des écarts de distance (rms).

Dans un souci de comparaison de nos résultats avec les positions des hydrogènes dans la PDB, et surtout de savoir si les positions que nous obtenions pour les hydrogènes étaient très loin de la réalité ou pas, nous avons calculé la valeur du rms entre deux cubes d'arête R et entre deux sphères de même rayon R dans lesquelles les points étaient aléatoirement placés. La parvalbumine a été choisie car c'est une protéine globulaire de rayon moyen égal à 20 Å et que la comparaison avec un nuage globulaire de points extrêmement mal reconstruit ou un 'gaz d'atomes' sphérique aléatoirement calculé est similaire.

2. Résultats :

Structure	Nombre de points	Rayon R (Å)	Rms (Å)	Comparaison
Parvalbumine	109	20	~0.5	Structure réelle
Gaz d'atomes	109	20	~1.5	Structure aléatoire

Sphères aléatoires	position des 100 points	rms
rayon R = 20 Å	$r = \sqrt{\text{rand}(1)} * R ; (\theta, \varphi)$	18.75 Å
rayon R = 40 Å	$r = \sqrt{\text{rand}(1)} * R ; (\theta, \varphi)$	37.7 Å
Cubes aléatoires	position des 100 points	rms
arête R = 20 Å	$x, y, z = \text{rand}(1) * R$	13.3 Å
arête R = 40 Å	$x, y, z = \text{rand}(1) * R$	26 Å

Tableau 1 : Comparaison des rms entre 2 cubes d'arête R (20 ou 40 Å) et entre 2 sphères de rayon R (20 ou 40 Å) constitués de points répartis aléatoirement et de façon homogène ; rand(1) est un nombre aléatoire compris entre 0 et 1. R est la distance délimitant les différents objets.

NB : une loi en $\sqrt{\text{rand}(1)}$ permet de simuler à peu près une sphère homogène. Les calculs ont été effectués dans Matlab.

Ainsi, pour deux sphères remplies de 100 points aléatoirement positionnés et de façon quasi homogène dans la sphère, le rms est proche du rayon de la sphère. Dans le cas d'un cube constitué lui aussi de 100 points, le rms est voisin de la moitié de la longueur des arêtes (voir Tableau 1)

Résultats obtenus dans X-plor :

La parvalbumine est une protéine comportant 6 hélices, les différents rms présentés dans les tableaux ci-dessous représentent les valeurs trouvées dans chacun des cas comprenant des jeux de distances initiaux différents ($d \leq 4$ Å : Tableau 2 ; $d \leq 5$ Å : Tableau 3), soit en considérant les (HN), les (HN, H α) ou les (HN, H α , H β). Lorsque deux valeurs sont présentes dans une case du tableau, elles représentent la valeur moyenne du rms trouvé pour chaque structure (colonne *global*) ou sous-structure (colonne *hélice 1 à 6*), et la valeur du rms correspondant à la structure image miroir respective (*sym*).

hydrogènes / rms	dist. à.	global	(hélice 1)	(hélice 2)	(hélice 3)	(hélice 4)	(hélice 5)	(hélice 6)	moy/hélice
HN et sym.	4 Å	35-45 Å	2-4.2 Å	2-3.5 Å	25-60 Å	2-3.5 Å	1.2-3.8 Å	10-50 Å	14 ± 17 Å
HN, H α et sym	4 Å	10-14 Å	2-4 Å	2-4 Å	2-4 Å	2-4 Å	2-4 Å	2-4 Å	3 ± 1 Å
HN, H α , H β	4 Å	12 Å	3.75 ± 0.8 Å	3.75 ± 0.8 Å	3.75 ± 0.8 Å	3.75 ± 0.8 Å	3.75 ± 0.8 Å	3.75 ± 0.8 Å	3.75 ± 0.8 Å
HN, H α , H β sym	4 Å	3 Å	3 Å	1 Å	2.5 Å	2 Å	1.5 Å	1.5 Å	1.9 ± 0.7 Å

Tableau 2 : calcul des rms obtenus sur 10 structures comparées à la structure cristallographique. sym. représente la structure image miroir observée. Le rms a été calculé pour la structure globale et chaque hélice. Distance considérée : 4 Å.

hydrogènes / rms	dist. à.	global	(hélice 1)	(hélice 2)	(hélice 3)	(hélice 4)	(hélice 5)	(hélice 6)	moy/hélice
HN et sym	5 Å	9-18 Å	1.6-4 Å	1.6-4 Å	1.6-4 Å	1.6-4 Å	1.6-4 Å	1.6-4 Å	2.8 ± 1.3 Å
HN, H α et sym	5 Å	8-10 Å	2-4 Å	1-3 Å	2-4 Å	2-4 Å	2-4 Å	2-4 Å	2.8 ± 1.0 Å
HN, H α , H β	5 Å	12 Å	4 Å	4 Å	4 Å	4 Å	4 Å	4 Å	4 Å
HN, H α , H β sym	5 Å	1.3 Å	0.5 Å	0.5 Å	0.5 Å	0.5 Å	0.5 Å	0.5 Å	0.5 Å

Tableau 3 : calcul des rms obtenus sur 10 structures comparées à la structure cristallographique. sym. représente la structure image miroir. Le rms a été calculé pour la structure globale et chaque hélice. Distance considérée : 5 Å.

Le premier rms obtenu qui permette d'affirmer que nous avons une structure similaire à celle de

départ (rms = 1.3 Å) n'apparaît que lorsque l'on considère les HN, H α et H β et leur position relative deux à deux à 5 ± 0.5 Å près. C'est à dire celui comportant le jeu initial de distance le plus contraignant, les contraintes étant quant à elles, toutes correctes.

Dans le cas où l'on considère seulement les HN et les H α (à 4 ou à 5 Å), il n'y a pas vraiment de différences, les rms obtenus sont mauvais quelque soit la structure globale ou son symétrique considéré.

Dans le cas où les HN, H α et H β sont impliqués dans les contraintes, nous observons une différence de rms entre la structure globale et son symétrique, ainsi qu'entre les différentes hélices et leurs symétriques respectifs.

Les valeurs des rms observés sont inférieures au rayon moyen de la parvalbumine (R= 20 Å) sauf dans le cas du jeu de distances où l'on considère des distances à 4 Å sur les HN seuls. Dans ce dernier cas, la structure obtenue est donc très mauvaise et quasi aléatoire étant donné les valeurs de rms trouvées précédemment entre des sphères aléatoires comportant un nombre similaire de points.

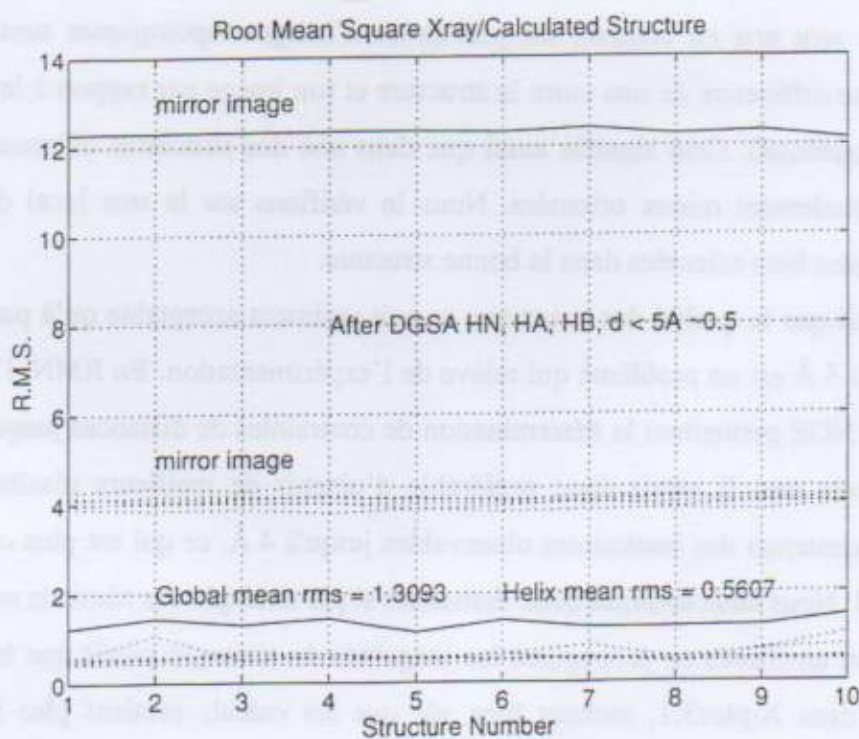


Figure 22 : Courbe des rms obtenus par rapport à la structure cristallographique sur 10 structures calculées dans le module dgsa.inp (X-plor) en considérant comme contraintes les distances inter HN, H α et H β inférieures à 5 Å et prises ± 0.5 Å.

a) Discussion :

Les résultats obtenus indiquent clairement que si l'on veut reconstruire une structure moléculaire

en ne considérant que les interactions entre les HN, H α et H β , il faut identifier des effets NOE entre tous ces atomes jusqu'à une distance de 5 Å de façon à obtenir des contraintes de distance suffisamment adéquates pour reconstruire la structure d'une protéine globulaire de façon correcte et pour ne pas être confronté au problème de l'apparition d'images topologiques dans les structures reconstruites (voir Théorie de la géométrie de distance). En effet, le fait de ne pas avoir d'attribution sur les atomes implique l'absence de topologie (liaisons chimiques et angles). Le 'gaz d'atomes' créé ainsi est donc très susceptible vis à vis des problèmes de chiralité. En effet, dans le cas où nous considérons de fixer comme contraintes les distances jusque 4 Å, nous obtenons des structures globales avec des rms supérieurs à 10 Å (cas des HN + H α) alors que les rms locaux au niveau des structures secondaires hélicoïdales sont autour 3 Å. La structure obtenue est donc partiellement incorrecte. Ce qui est intéressant de remarquer est la similitude des résultats obtenus tant que les H β ne sont pas pris en compte dans le 'gaz d'atomes'. Bien sûr la qualité des structures obtenues est meilleure au niveau des structures secondaires si on prend des contraintes initiales dans les distances jusque 5 Å (rms=2.8 Å au lieu de rms=3 Å pour des contraintes initiales prises 4 Å maximum).

Lorsque les H β sont pris en compte, les problèmes d'images topologiques sont différents. En effet, il y a une grosse différence de rms entre la structure et son image par rapport à la structure vraie (structure cristallographique). Cela signifie aussi que dans une des structures obtenue, les structures secondaires sont globalement mieux orientées. Nous le vérifions sur le rms local de chacune des hélices, elles sont toutes bien orientées dans la bonne structure.

Cependant, le fait que la qualité des structures ne soit vraiment acceptable qu'à partir d'un jeu de contraintes complet à 5 Å est un problème qui relève de l'expérimentation. En RMN, l'obtention d'un jeu complet d'effets NOE permettant la détermination de contraintes de distances jusqu'à 5 Å n'arrive que dans les meilleurs cas. Il serait donc préférable d'obtenir de meilleurs résultats d'une autre manière pouvant se contenter des contraintes observables jusqu'à 4 Å, ce qui est plus courant dans les expériences de RMN. Nous nous sommes donc demandés si par exemple ces résultats ne pouvaient pas être considérablement améliorés en propageant les inégalités du tétrangle plutôt que les inégalités du triangle implantées dans X-plor3.1, sachant bien sûr que les calculs seraient plus longs. Il serait intéressant de pouvoir effectuer cette étude dès que les contraintes temporelles dues à la longueur des calculs ne seront plus un facteur limitant.

3. Le problème du positionnement de la chaîne principale :

Dans l'étude de reconstruction de structures par RMN sans attribution, il existe également, à la suite du repositionnement correct des hydrogènes dans l'espace 3D, le problème du placement de la chaîne principale des résidus selon un repliement correct en fonction des positions trouvées pour les

atomes d'hydrogènes. C'est un problème que nous n'avons pas encore résolu, mais que nous avons envisagé. En effet, il semble possible lorsque l'on connaît la position des hydrogènes HN, H α et H β d'un même système de spins de mettre en place le squelette carboné. Connaissant HN et H α , nous pouvons déterminer l'angle dièdre N-C α (voir Figure 23) et de même obtenir une approximation sur la position du C β puisque la position des H β ou du pseudo-atôme correspondant est connue, la position du carbone du groupement carboxyle s'en déduit suite à l'hybridation sp³ du C α .

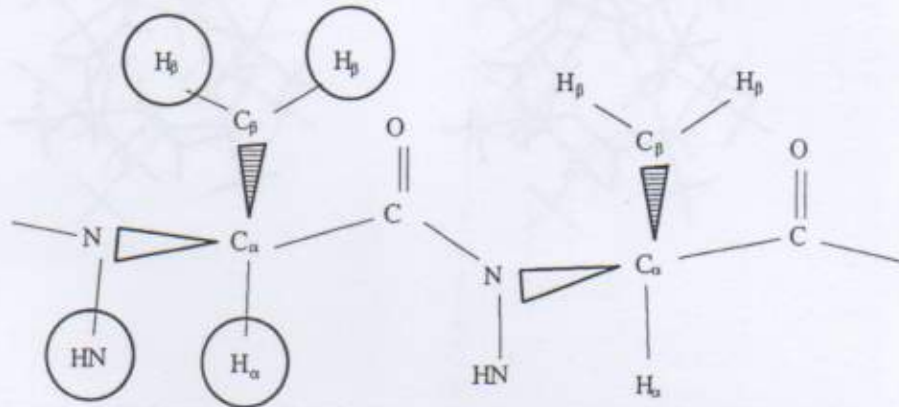


Figure 23 : Le problème du positionnement de la chaîne principale

Ainsi, sans pour autant former les liaisons peptidiques, il est relativement possible de placer le squelette carboné de chaque résidus dans un premier temps, minimiser son énergie interne en priorité tout en optimisant sa position en fonction des positions des hydrogènes placés par géométrie de distance, puis ensuite former les liaisons peptidiques et minimiser l'énergie du squelette carboné en fonction des longueurs de liaison à respecter, ceci suivant un champ de forces donné. Nous avons réalisé cette partie du travail sur un polymère (poly-Alanine de 10 Alanines), normalement en forme d'hélice en écrivant un programme dans Matlab qui permet de replacer les résidus suivant cette méthode, la structure étant ensuite relaxée et minimisée dans AMBER. (voir résultat Figure 24).

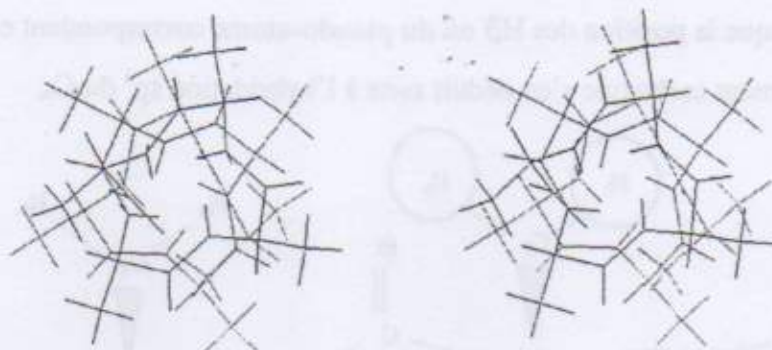


Figure 24 : Vue stéréoscopique de la reconstruction d'un polymère d'Alanine en hélice via la méthode décrite ci-dessus.

4. Conclusion :

Nous avons vu que la géométrie de distance est une méthode très efficace pour la reconstruction de structures moléculaires. Cependant c'est une méthode qui est encore trop gourmande en temps de calcul et ce sera à mes successeurs qui disposeront d'outils de calcul plus rapides de poursuivre cette étude passionnante de reconstruction.

Des résultats obtenus avec X-plor, il en ressort que l'obtention de résultats de qualité avec les inégalités du triangle nécessite la connaissance de contraintes de distances sur les HN, H α et H β à 5 Å. On ne retrouve alors plus de problèmes d'images topologiques sur le squelette carboné car les contraintes obtenues sont suffisantes pour la chiralité des structures secondaires en hélice, et aussi pour la structure globale.

Concernant le positionnement de la chaîne principale des résidus à partir de la connaissance de la position des hydrogènes, les résultats obtenus pour un polymère d'Alanine indique qu'il est possible de reconstruire le polymère complet à partir de ces informations. Bien sûr, le cas étudié est d'une simplicité relative et il serait nécessaire d'envisager une étude plus complexe avec deux résidus différents ou plus dans le polymère.

L'amélioration de la méthode de reconstruction de structures moléculaires via la géométrie de distance et la RMN sans passer par l'attribution semble réalisable dans la mesure où les spins ne sont pas dégénérés et où l'information obtenue ne comporte pas d'ambiguïtés. Dans le cas contraire, il faut envisager un marquage isotopique à l'azote ^{15}N ou au carbone ^{13}C pour obtenir la suppression des

dégénérescences et l'obtention de meilleures informations géométriques. Cette approche à partir de données en RMN hétéronucléaire a également été développée par Kraulis en 1994 et a révélée la possibilité de déterminer près de 95% des spins dans le cas de deux protéines, GAL4 et la trypsine pancréatique du bœuf [Kraulis, 1994].



Chapitre III. Les thiorédoxines :

A. Introduction :

Les thiorédoxines [Holmgren, 1985] forment une famille de protéines ayant une grande variété d'activités. Ce sont des protéines rédox, de petite taille (de 12.5 à 13.5 kDa), stables thermiquement et caractérisées par la présence d'un pont disulfure qui leur donne leurs propriétés oxydo-réductrices.

Des thiorédoxines ont été isolées depuis un éventail très large d'espèces eucaryotes et procaryotes [Nisbet and Fowler, 1996] [Gray, 1996] ; on les trouve au niveau de la membrane des cellules. Les thiorédoxines de mammifères montrent environ 25% de similarité de séquence avec la thiorédoxine bien connue de *Escherichia Coli*. En particulier les thiorédoxines des mammifères possèdent 2 cystéines de plus, éloignées l'une de l'autre et non catalytiques, conservées dans la séquence. La thiorédoxine humaine possède même 3 de ces cystéines. Chez l'homme, ces résidus structuraux, non catalytiques, peuvent cependant être oxydés pour conduire à une inactivation de la protéine et à une dimérisation.

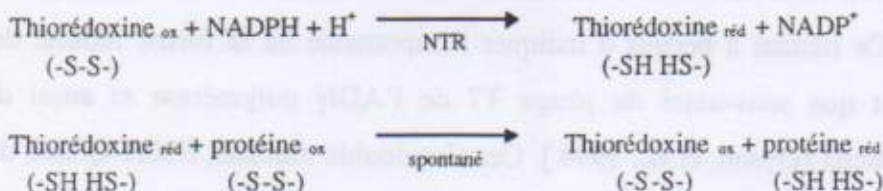
Quelques structures de thiorédoxines ont été déterminées par diffraction des rayons X ou par résonance magnétique nucléaire (la thiorédoxine de *Escherichia Coli*: 108 Résidus, la thiorédoxine humaine: 105 résidus, la thiorédoxine de *Chlamydomonas Reinhardtii*: 113 résidus).

Au niveau moléculaire, ce sont des protéines très hydrophiles et leur fonction est d'agir en tant que donneur d'hydrogène.

B. Les activités diverses des thiorédoxines :

Les thiorédoxines sont des donneurs d'atomes d'hydrogène pour diverses macromolécules biologiques, relativement à leur fort potentiel rédox $E_0 = -270\text{mV}$ (valeur ^{de} dans la thiorédoxine d'*E.Coli*).

Les thiorédoxines interviennent dans la régulation de nombreux systèmes par l'intermédiaire d'un contrôle rédox de thiols: changement d'activité d'une enzyme, d'un récepteur, d'un facteur de transcription.



Elles agissent sur les ribonucléotides réductases [Darling, 1988] [Hopper and Iurlano, 1983] [Mirandavizueté et al., 1994] et influent donc sur la régulation de la synthèse d'ADN puisque ceux-ci y jouent un rôle essentiel. Ce sont donc des protéines influant donc sur la reproduction des cellules et

pouvant jouer un rôle dans la régulation de nombreux systèmes.

Dans le système ferrédoxine/thiorédoxine [Schurmann,] [Salamon et al., 1995], les thiorédoxines f activent des enzymes de régulation de la photosynthèse [Buchanan, 1991]. [Jacquot et al., 1984], telles que la fructose-1,6-biphosphate phosphatase [Pla and Lopez-Gorge, 1981]. Les thiorédoxines m activent des enzymes de régulation tels que la NADP malate deshydrogénase impliquée dans l'exportation du pouvoir réducteur hors du chloroplaste [Issakidis et al., 1993], ou la glucose-6-phosphate deshydrogénase [Udvardy et al., 1984]. Elles interviennent aussi dans l'infection des bactéries [Masutani et al., 1992].

Enfin, il s'avère que des thiorédoxines jouent un rôle dans l'assimilation de l'azote [Vargas et al., 1994], dans la défense de l'épiderme contre les radicaux libres chez l'homme [Schallreuter, 1996], dans la fibrillation ventriculaire [Aota, 1996], la coagulation sanguine [Blombäck et al., 1986] [Biguet, 1994], l'accroissement de solubilité de protéines étrangères dans *E.Coli* [Yasukawa, 1995]. De plus, la thiorédoxine humaine est susceptible d'interagir avec NFκB [Matthews, et al., 1992] [Schenk, et al., 1994] [Schulzeosthoff, et al., 1995] [Qin, et al., 1995] [Qin, et al., 1996] facteur de transcription de l'ADN intervenant dans la propagation du virus HIV. C'est une protéine de régulation avec des propriétés intéressantes et des résultats prometteurs en technologie et en médecine [Buchanan, 1994] [Buchanan, 1994].

Pour information, le nombre de publications sur les thiorédoxines depuis les années 1960 rend bien compte de leur intérêt (voir Figure 25).

De nombreux mutants des thiorédoxines ont également été étudiés dans le but de mieux comprendre cette famille de protéines. La structure d'un mutant de la thiorédoxine d'*E.Coli*, K36E [Nikkola, et al., 1993], a été résolue montrant des changements structuraux dans la séquence primaire avant le site actif, ces derniers modifiant la dépendance de la fluorescence en fonction du pH mais pas l'activité de la protéine. D'autres mutants tels que C62A, C69A, C73A dans la thiorédoxine humaine [Forman-Kay, et al., 1992] montrent également une conservation de l'activité indiquant que ces cystéines supplémentaires n'interviennent pas sur l'activité du site actif. Un double mutant C32S-C35S de la thiorédoxine d'*E.Coli* fut aussi étudié et montra une activité voisine de celle de la protéine sauvage sous forme réduite. Ce mutant a permis d'indiquer l'importance de la forme réduite de la thiorédoxine d'*E.Coli* en tant que sous-unité du phage T7 de l'ADN polymérase et aussi dans l'assemblage du phage filamenteux [Dyson, et al., 1994]. Certains double mutants, E30H-Q62H, de la thiorédoxine d'*E.Coli* sont même employés dans l'élaboration sur des colonnes contenant des ions métalliques, de nouvelles méthodes de purification de protéines fusionnées avec ce type de thiorédoxine [Lu, et al., 1996].

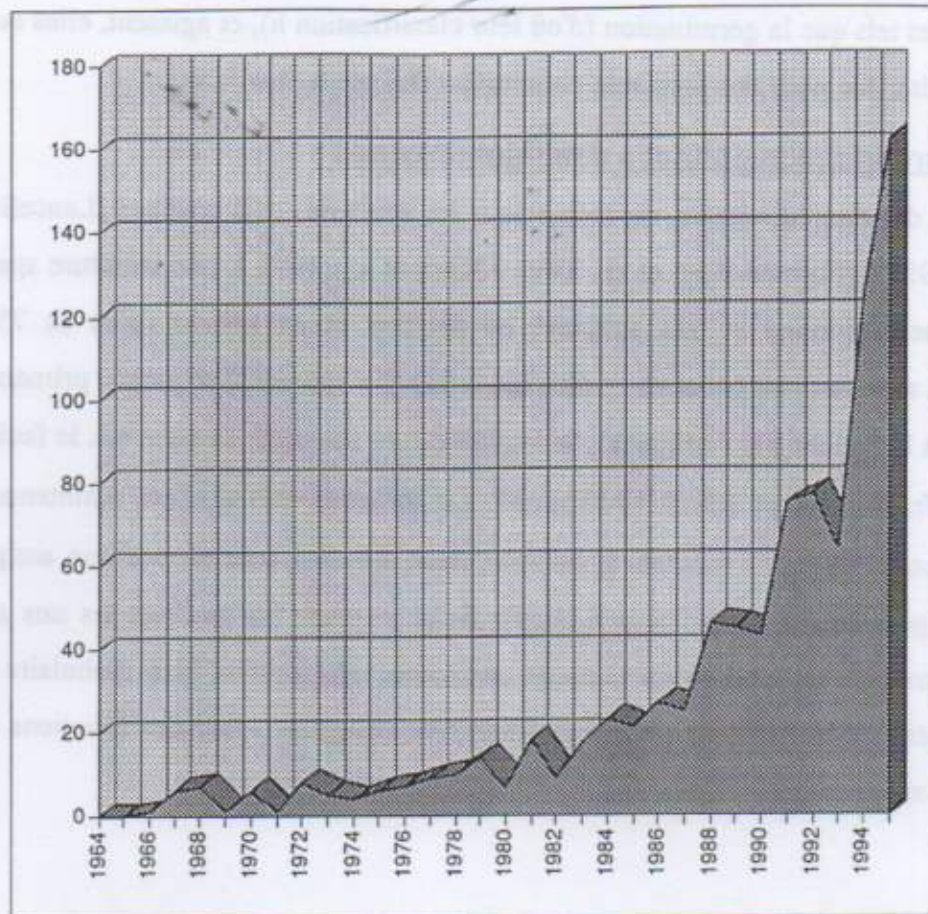


Figure 25 : Evolution du nombre de publication annuelles relatives aux thiorédoxines et disponibles dans la base de données Medline (<http://www3.ncbi.nlm.nih.gov/>) depuis les années 1960.

C. Différents types de thiorédoxines :

Il existe trois types de thiorédoxines, les types m, f et h. Les thiorédoxines m et f ne se trouvent que dans les plantes et sont localisées dans le chloroplaste des cellules. Les thiorédoxines h se trouvent dans tous les organismes, aussi bien chez les animaux que chez les plantes.

1. Les thiorédoxines m :

L'activité des thiorédoxines m est basée sur leur capacité d'activer dans le chloroplaste la malate déshydrogénase NADP⁺-dépendante (NADP⁺-MDH) quand elles sont réduites photochimiquement (en présence de ferrédoxine et de Ferrédoxine Thiorédoxine Réductase) ou en présence de DTT.

2. Les thiorédoxines f :

L'activité des thiorédoxines f est basée sur leur capacité d'activer dans le chloroplaste la fructose-1,6-bisphosphatase (FBPase) quand elles sont réduites photochimiquement en présence de ferrédoxine et de Ferrédoxine Thiorédoxine Réductase, ou aussi en présence de DTT.

3. Les thiorédoxines h :

Les thiorédoxines h, présentes dans les animaux et les plantes, sont associées à des processus hétérotrophiques tels que la germination (d'où leur classification h), et agissent, elles aussi, en tant que donneurs d'hydrogène pour des protéines comportant des ponts disulfures.

D. La structure tridimensionnelle des thiorédoxines :

La famille des thiorédoxines a, du moins pour les structures déjà connues [Lancelin, et al., 1993] [Jeng, et al., 1994] [Forman-Kay, et al., 1991] [Qin, et al., 1995], une structure spatiale constituée d'un arrangement ordonné et très structuré en feuillets et en hélices (plus de 75% des résidus participent aux structures secondaires - voir Figure 27). En suivant la séquence primaire, on rencontre successivement le feuillet $\beta 1$, l'hélice $\alpha 1$, le feuillet $\beta 2$, le site actif, l'hélice $\alpha 2$, le feuillet $\beta 3$, l'hélice $\alpha 3$, puis les feuillets $\beta 4$ et $\beta 5$ et l'hélice $\alpha 4$. La structure tertiaire est maintenue par un cœur hydrophobe constitués par 5 feuillets β dont les deux derniers sont en position antiparallèle. Les 4 hélices s'arrangent autour de ce cœur hydrophobe et joignent les feuillets les uns aux autres pour compléter la structure tertiaire en lui donnant une forme relativement bien globulaire. Des structures similaires ont été observées dans un grand nombre de protéines ayant des fonctions différentes, par exemple la flavodoxine (voir Figure 26).

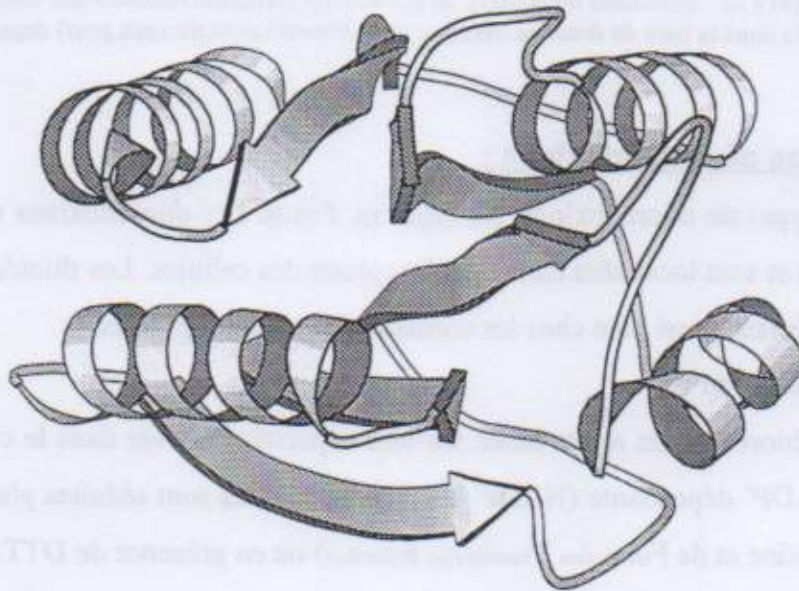


Figure 26 : Représentation dans Molscript de l'allure de la structure des flavodoxines.

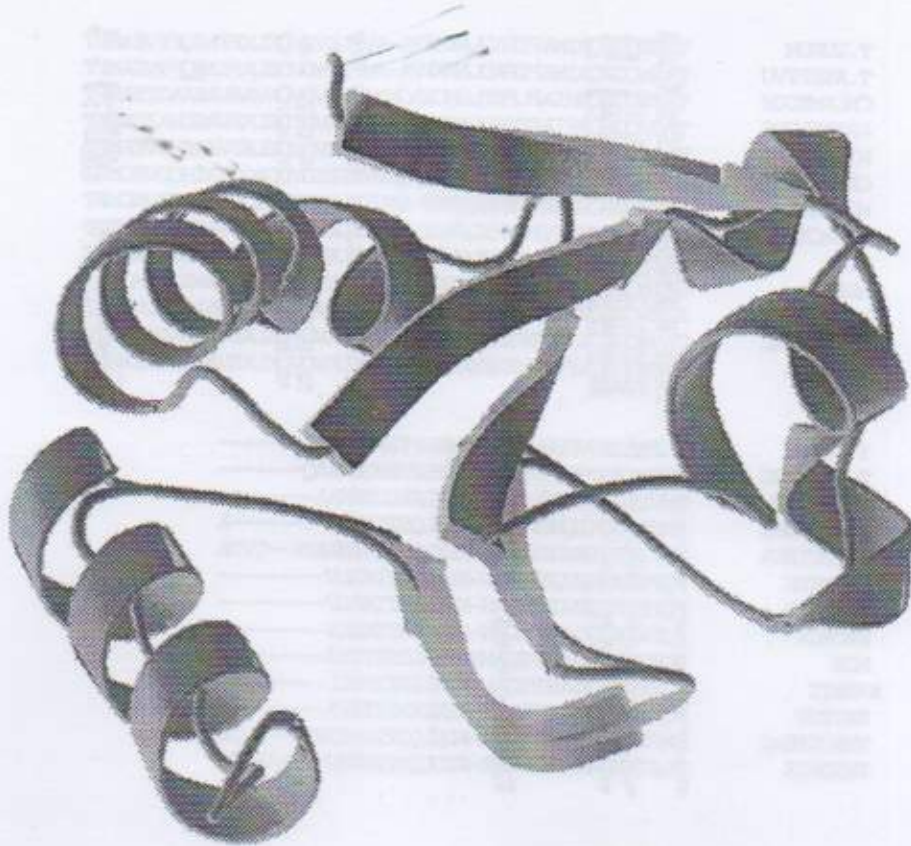


Figure 27 : Représentation de la structure tertiaire de la thiorédoxine humaine - (Rasmol - Molscrip - Raster3D).

E. Le site actif des thiorédoxines :

1. Séquence caractéristique:

Le site actif des thiorédoxines est constitué par l'enchaînement caractéristique des acides aminés Trp-Cys-Gly-Pro-Cys (ou WCGPC). Cet enchaînement varie dans deux séquences connues où la proline est remplacée par un autre résidu X. Ci-dessous (Figure 28) est présenté un alignement de séquences entre les diverses thiorédoxines, montrant bien la conservation du site actif et d'autres résidus, en autre ceux environnant le site actif dans la structure tertiaire.

CLUSTAL V multiple sequence alignment

T. DUFILM	MAAAATATTTAAATAAAVGFGEIISVHSLQWIMQIEEANAARKLWVDF
T. AESTIVU	MAASA---ATATATAAAVGADEVISVHSLQWIMQIEEANAARKLWVDF
CHLAMYDOH	-----GGSVVIDSKAAWDAQLAKGKEEHKPTWDF
ARABIDOPS	MA-----SEE-----GNTACHIVEIWNELQKQNESKILWVDF
NICOTIANA	MAANDATISSEE-----GQVEGCHKVEEANEYFKGVEIKKLWVDF
CHICKEN	-----VKSUGNLADFEAELKAA--GEKLWVDF
HIMDSAP	-----VKQIESKIAFQEALDAA--GEKLWVDF
MACACA	-----VKQIESKAAFQEALIDA--GEKLWVDF
MUS	-----VKLIESKEAFQEALAAA--GEKLWVDF
RABBIT	-----VKQIESKSAFQEVLDAA--GEKLWVDF
RATTUS	-----VKLIESKEAFQEALAAA--GEKLWVDF
THIOTABAC	MA-----EE-----GQVIGMHTVDWANEHLQKGIIDKRLTWDF
THIORIZ	MAA-----EE-----GNTACHNKDEFDAQMTKAKEAGKVVILDF

T. DURUM	TASWCGPCRIMAPIFADLAKKFFA-AVFLKVDVDELKPIAEQFSVEAMPT
T. AESTIVU	TASWCGPCRIMAPIFADLAKKFFA-AVFLKVDVDELKPIAEQFSVEAMPT
CHLAMYDOH	TATWCGPCRMIAPIFETLSNDVAGKVFELKVDVAVAAVAEAGITAMPT
ARABIDOPS	TASWCGPCRFIAPFFADLAKKLEN-VLEFLKVDVDELKSVASDWAIQAMPT
NICOTIANA	TASWCGPCRFIAPILADIAKKMPH-VIFLKVDVDELKTVSAEWSVEAMPT
CHICKEN	SATWCGPCRMKPPFFHSLDRK-GEVVFLEIDVDDAQDVATHCDVKCMPT
HMO5AP	SATWCGPCRMKPPFFHSLSEKY-SNVVFLEVDVDDQQDVASECEVKCMPT
MACACA	SATWCGPCRMKPPFFHSLSEKY-SNVVFLEVDVDDQQDVASECEVKCMPT
MJS	SATWCGPCRMKPPFFHSLDRY-SNVVFLEVDVDDQQDVADCEVKCMPT
RABBIT	SATWCGPCRMKPPFFHSLSEKF-SNVVFLEVDVDDCKDIAABCEVKCMPT
RATTUS	SATWCGPCRMKPPFFHSLDRY-SNVVFLEVDVDDQQDVADCEVKCMPT
THIOTABAC	TASWCGPCRFIAPVFAEYAKKFFG-AVFLKVDVDELKSVAITWAVEAMPT
THICRIZ	TASWCGPCRFIAPVFAEYAKKFFG-AVFLKVDVDELKEVAEKYINVEAMPT
T. DURUM	FLFMKEGVKDRVWGAI-KEELTIKVGHLHAA-----
T. AESTIVU	FLFMKEGVKDRVWGAI-KEELTIKVGHLHAAQ-----
CHLAMYDOH	FHVYKLVKADLVGAS-QDKLKVAKHAA-----
ARABIDOPS	FMFLKEGKILKVGAK-KDELQSTIAKHL-----A
NICOTIANA	FVFIKDGKEVDRVWAK-KEELQTTVKAAPA--TVIA
CHICKEN	FQFYKGGKVGESGAN-KEKLEATTIKSLV-----
HMO5AP	FQFFKGGKVGESGAN-KEKLEATTINELV-----
MACACA	FQFFKGGKVGESGAN-KEKLEATTINELV-----
MJS	FQFYKGGKVGESGAN-KEKLEATTISITEYS-----
RABBIT	FQFFKGGKVGESGAN-KEKLEATTINELL-----
RATTUS	FQFYKGGKVGESGAN-KEKLEATTITEFA-----
THIOTABAC	FMFLKEGKIVKVGAK-KDELQTTIAKHSST--S-TA
THICRIZ	FLFIKGAEADKVGAR-KDELQNTIVKHVGATAASASA

Figure 28 : résultat de l'homologie de séquence observée entre diverses thiorédoxines (Clustal V)

2. Position dans les structures connues:

Le site actif est situé dans un boucle externe située à la surface de la protéine (voir Figure 29) et qui est comprise entre l'extrémité d'un feuillet et l'hélice qui le suit (feuillet $\beta 2$ et hélice $\alpha 2$ dans *E.Coli*). Cette situation est très commune car il est courant d'avoir un site actif près de l'extrémité d'une hélice, celle-ci créant un fort moment dipolaire.

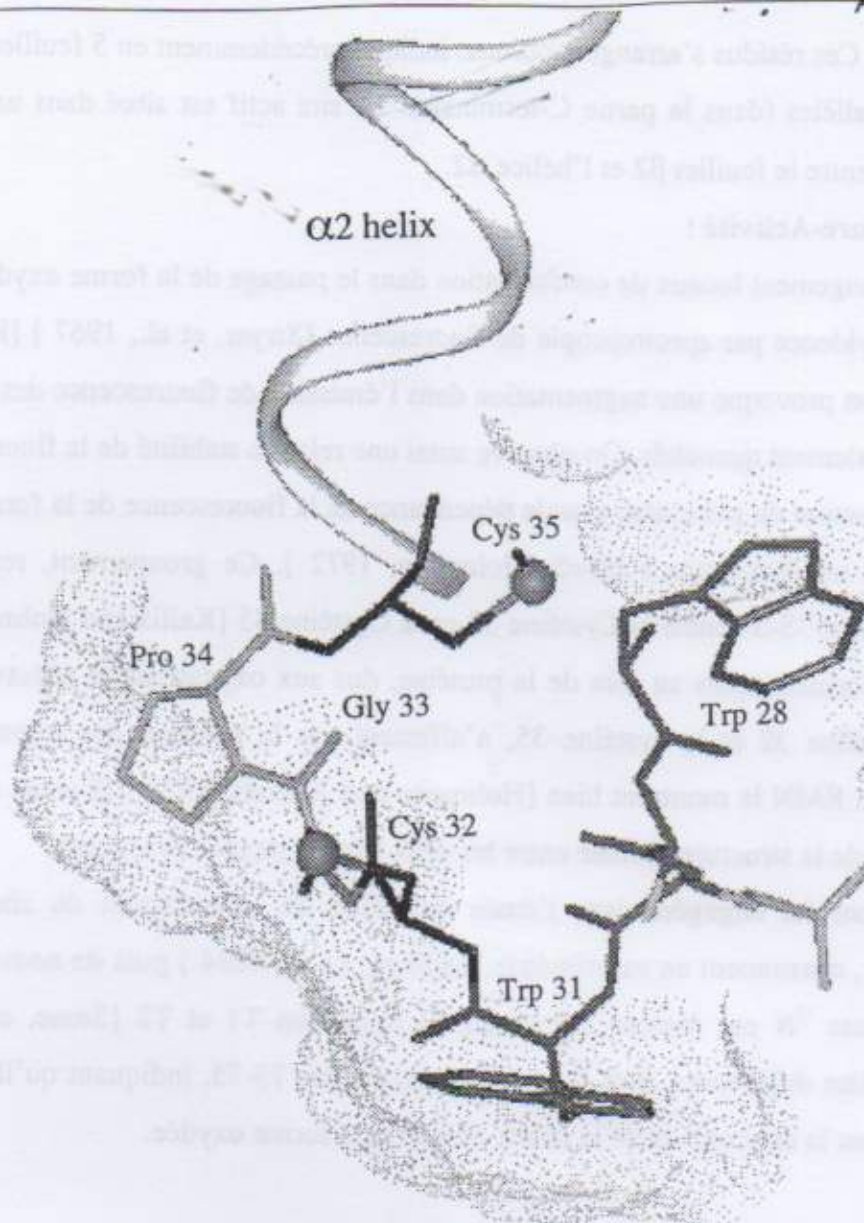


Figure 29 : Représentation dans Insight [Biosym, 1993] du site actif des thiorédoxines. Les parties représentées par des nuages de points indiquent la surface accessible au solvant autour du site actif.

F. La thiorédoxine chez *Escherichia Coli* :

1. Position dans la cellule :

Les thiorédoxines se comportent comme des protéines très solubles après extraction des cellules. Elles se situent à la périphérie locale de la membrane interne des cellules, comme l'indique leur résistance relativement importante aux chocs osmotiques et leur association partielle avec une fraction de la membrane après une coupure douce [Lunn and Pigiet, 1982]. La démonstration de cette localisation périphérique, proche de la membrane, dans la cellule a été obtenue par microscopie électronique [Nygren, et al., 1981]. Néanmoins, dans certaines cellules, une localisation dans la région des nucléotides a été observée.

2. Aspects Structuraux :

La thiorédoxine d'*E.Coli* est une chaîne polypeptidique de 108 acides aminés, de masse

moléculaire 11,7 kDa. Ces résidus s'arrangent comme indiqué précédemment en 5 feuillets dont 3 sont parallèles et 2 antiparallèles (dans la partie C-terminale). Le site actif est situé dans une saillie à la surface de la protéine entre le feuillet $\beta 2$ et l'hélice $\alpha 2$.

3. Relation Structure-Activité :

L'évidence de changement locaux de conformation dans le passage de la forme oxydée à la forme réduite a été mis en évidence par spectroscopie de fluorescence [Stryer, et al., 1967] [Reutimann, et al., 1981]. La réduction provoque une augmentation dans l'émission de fluorescence des tryptophanes (Trp28 notamment) fortement quenchés. On observe aussi une relative stabilité de la fluorescence dans la forme oxydée en fonction du pH et une grande dépendance de la fluorescence de la forme réduite en fonction du pH, avec un maximum à pH=5 [Holmgren, 1972]. Ce groupement, responsable du 'quenching', est la liaison -S-S- entre la Cystéine 32 et la Cystéine 35 [Kallis and Holmgren, 1980]. Les changements conformationnels au sein de la protéine, dus aux oxydations et réductions du pont disulfure entre la cystéine 32 et la cystéine 35, n'affectent pas la position des tryptophanes. Les structures observées en RMN le montrent bien [Holmgren and Roberts, 1976] [Dyson, et al., 1988] avec une conservation de la structure globale entre les deux formes oxydée et réduite.

D'autres études ont été engagées dans l'étude des effets de la réduction du site actif de la thiorédoxine d'*E.Coli*, notamment en calorimétrie [Ladbury, et al., 1994] puis de nouveau en RMN hétéronucléaire à l'azote ^{15}N par mesure des temps de relaxation T1 et T2 [Stone, et al., 1993] montrant quelques petites différences dans la mobilité des résidus 73-75, indiquant qu'il n'y a aucun changement majeur dans la structure entre la forme réduite et la forme oxydée.

Chapitre IV. Effet de la mutation en Alanine de la Proline 40 dans la thiorédoxine d'*Escherichia Coli*: (P40A)

A. Introduction :

Dans le but de comprendre le rôle physiologique des thiorédoxines, plusieurs laboratoires ont séquencé des thiorédoxines provenant de différentes sources. La séquence primaire de nombreuses thiorédoxines est désormais connue et toutes possèdent le même site actif caractéristique [Le Marechal et al., 1992] (voir aussi page 62). Parmi les résidus conservés, Pro 40 (numérotation par rapport à la thiorédoxine d'*E.Coli*) est présente dans toutes les thiorédoxines séquencées jusqu'à maintenant à part dans une thiorédoxine présente dans l'algue rouge [Reynolds et al., 1994] et deux types de thiorédoxines h de *Nicotiana tabacum* dans lesquelles une sérine ou une arginine est présente [Brugidou et al., 1993]. Cette proline est proche du site actif (32-35) dans l'hélice α_2 . Or, pour maintenir le réseau de liaisons hydrogène servant à structurer le squelette carboné en hélice, il faut que chacun des résidus ait son hydrogène amide (HN). C'est pourquoi ce résidu s'insère difficilement dans les structures secondaires en hélice dans les protéines.

Dans le cas des thiorédoxines, cette proline induit un coude dans la structure de l'hélice α_2 mais ne brise pas la régularité de la structure secondaire. Dans l'étude qui suit, nous nous sommes intéressés au rôle de cette proline très conservée aussi bien biochimiquement que structurellement. La mutagenèse dirigée a été employée pour remplacer la proline 40 par une alanine, dans le but de déterminer si ce changement affecte la structure et/ou la fonctionnalité de la thiorédoxine. Nous avons effectué des simulations de dynamique moléculaire pour explorer les changements structuraux que la mutation P40A induit dans la molécule. Une étude par RMN a également été effectuée dans le but de distinguer des changements conformationnels entre P40A et la thiorédoxine d'*E.Coli* . Les raisons possibles pour la grande conservation de cette proline pourrait être simplement que les propriétés stéréospécifiques de ce résidu sont nécessaires pour l'interaction de la thiorédoxine dans une conformation particulière avec ses différents partenaires. Au vu de cela, les conséquences d'une telle substitution pourrait avoir un grand nombre d'effets sur les cibles considérées [de Lamotte-Guéry et al., 1991]. D'un autre côté, la position de la proline 40 au sein de l'hélice α_2 directement reliée au site actif pourrait suggérer que ce résidu est structurellement impliqué dans le contrôle de la réactivité du pont disulfure de la thiorédoxine.

B. Matériel et méthodes :

Pour la partie de biologie moléculaire, se reporter à l'article dans les annexes.

I. Modélisation des structures des thiorédoxines P40A (mutant) et WT (« Wild Type ») :

La structure cristallographique de la thiorédoxine d'*E.Coli* [Katti et al., 1990] a été le point de départ de toutes les simulations. La substitution P40A a été effectuée en mutant directement la proline 40 en alanine dans le fichier contenant la structure au format PDB. La structure mutée a ensuite été minimisée très prudemment et la structure cristallographique originelle a été soumise à la même procédure de minimisation que le mutant. Pour les étapes de minimisation et les simulations de dynamique moléculaire dans l'eau, le solvant a été incorporé sous la forme d'une sphère d'eau de rayon 25 Å autour du centre de masse de chaque protéine, pour finalement ne retenir que 1540 molécules d'eau, chacune des molécules de solvant ayant un atome d'oxygène à une distance inférieure à 2 Å du soluté ou un atome d'hydrogène à une distance inférieure à 1.8 Å ayant été éliminée. La charge de l'hydrogène dans l'eau a été fixée à 0.417e.

Pour nous assurer d'un bon équilibre du système soluté-solvant, la sphère d'eau a été minimisée étape par étape : d'abord 200 pas de minimisation en ne laissant libre que les molécules d'eau et en figeant toutes les molécules du soluté, ensuite 5 ps de dynamique moléculaire à une température de 10 K ont été effectuées puis suivies de 500 autres pas de minimisation toujours en figeant le soluté. Enfin le processus de minimisation de l'ensemble soluté-solvant a été effectué pour affiner le système avant d'entrer dans les étapes de simulation en dynamique moléculaire. D'abord toujours en laissant les molécules d'eau libres de se déplacer, 500 pas de minimisation ont été effectués en fixant tous les atomes du soluté sauf les hydrogènes, puis 1000 pas de minimisation en fixant le « backbone » (squelette NCαCO) sauf les chaînes latérales qui sont alors mobiles, suivi de 5000 pas de minimisation similaires aux conditions précédentes mais avec des contraintes plus souples sur le « backbone », et finalement 5000 pas de minimisation sont effectués sans aucune contrainte ni sur le solvant, ni sur le soluté. Nous obtenons ainsi un système soluté-solvant qui est prêt pour être étudié en dynamique moléculaire.

Les simulations ont été réalisées sur la forme réduite de la protéine native et du mutant obtenus selon le protocole ci-dessus. La durée des simulations de dynamique moléculaire dans l'eau représente 400 ps dans un bain à une température de 300 K. Les vitesses initiales ont été attribuées aux atomes suivant une distribution de Maxwell à une température de 10 K et le système obtenu a été couplé au bain de température de façon à atteindre la température de 300 K en 20 ps. La valeur du temps de contact pour les couplages en température du soluté, du solvant et du bain est de 0.2 ps. La commande « shake » n'a pas été activée, autorisant donc les mouvements pour toutes les liaisons atomiques (hydrogènes compris), et un pas temporel de 1 fs a été programmé. Les calculs ont été effectués avec

une constante diélectrique distance-dépendante $\epsilon = r_{ij}^{-1}$ et une distance de « cut-off » de 10 Å a été incluse pour fixer une limite aux interactions de non-liaison. Les résultats ont été analysés sur les dernières 100 ps des trajectoires de dynamique moléculaire après obtention de la stabilisation relative de l'énergie totale et de la déviation des écarts structuraux par rapport à la structure de l'étape de calcul précédente (rms_d). (voir Figure 30 et Figure 31).

Les simulations ont été menées à bien grâce au programme Amber 4.0 [Pearlman et al., 1991]. Les programmes Insight 2.0 [Biosym, 1993] et Access [Eisenberg and McLachan, 1986] ont été utilisés pour la comparaison des structures, leur première évaluation et l'estimation de l'énergie libre de solvation, ainsi que de la surface accessible au solvant. L'angle de l'hélice α_2 a été calculé en considérant l'angle entre deux lignes, chacune d'entre elle étant positionnée dans la direction de l'axe de l'hélice à chacune de ses extrémités. Tous les calculs ont été effectués sur une machine Hewlett Packard HP735 sous OS HPUX 9.01.

C. Résultats :

I. Analyse de la structure modélisée des thiorédoxines P40A et WT :

a) Comparaison des formes minimisées :

Les structures obtenues des thiorédoxines P40A et WT, minimisées dans des sphères d'eau, ont été analysées. L'énergie totale durant le raffinement des structures a atteint un minimum d'énergie de -5750 kCal.mol⁻¹ pour la forme WT et de -5780 kCal.mol⁻¹ pour la forme P40A avec un rms_d de 0.01 Å pour chacune des structures. Les structures ont été estimées correctes dans PROCHECK (de chez Oxford Molecular, [Laskowski, et al., 1993]) en comparant à ceux de la structure obtenue par diffraction des rayons X les diagrammes de Ramachandran et les différents pourcentages obtenus pour chaque type de résidus trouvés dans les différentes régions.

b) Comparaison des trajectoires, évolution des structures :

Le protocole utilisé pour les simulations de dynamique moléculaire a été appliqué sur chacune des structures : P40A et WT. L'équilibre a été atteint après 300 ps de simulation lorsqu'une stabilisation raisonnable de l'énergie totale fut observée. Seules les dernières 100 ps sur les 400 ps de trajectoire en dynamique moléculaire ont été utilisées pour la comparaison des structures.

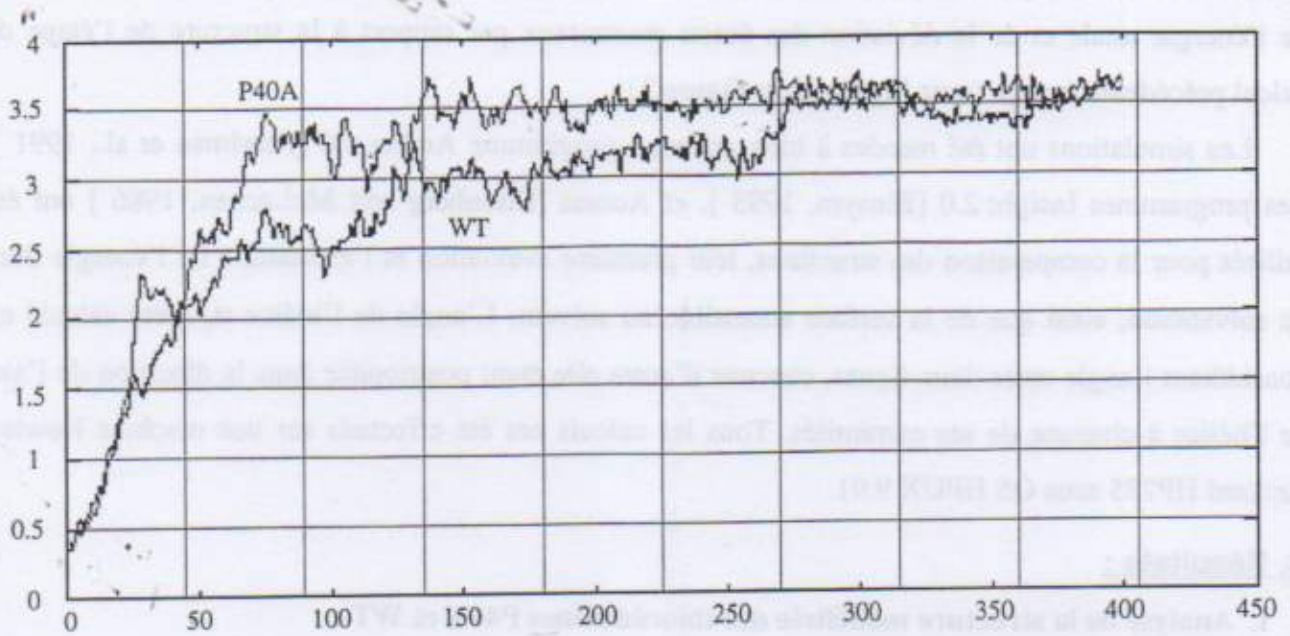


Figure 30 : évolution du rms au cours des simulations de dynamique moléculaire.

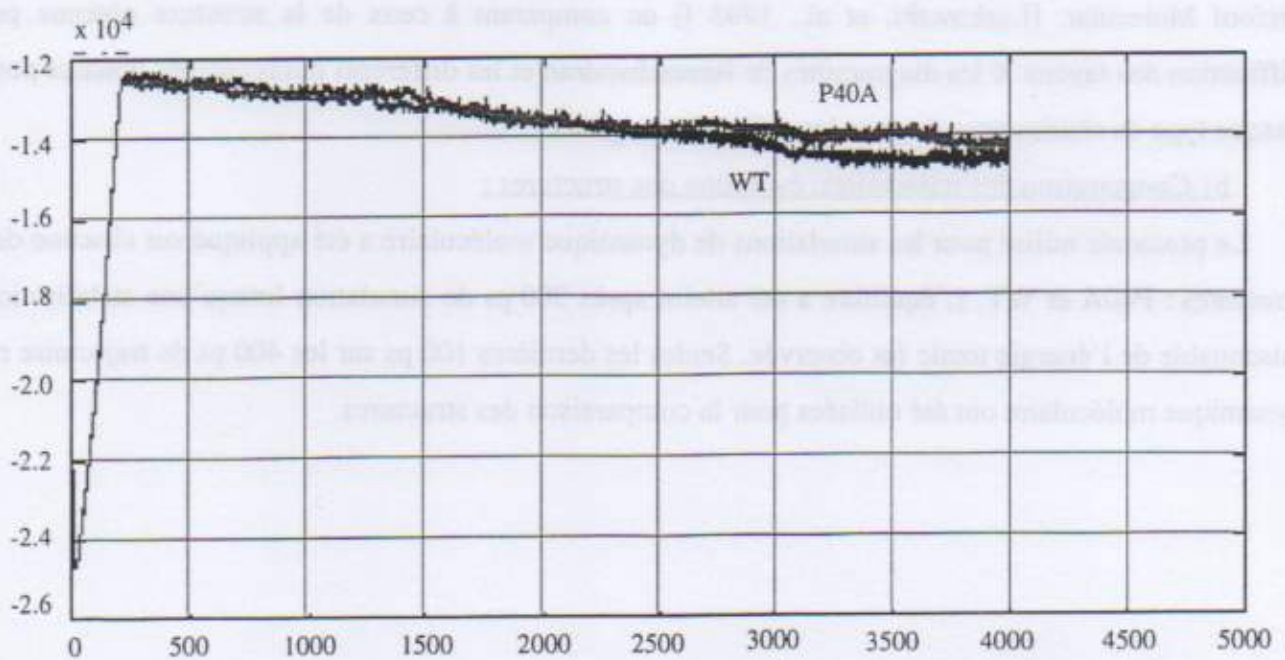


Figure 31 : évolution de l'énergie totale au cours des simulations de dynamique moléculaire.

Considérant l'ensemble des trajectoires, la structure globale est conservée aussi bien dans la protéine P40A et la protéine WT. Les deux structures gardent le même repliement, avec une conservation de la forme globale et des structures secondaires, mis à part le petit feuillet $\beta 1$ dans la partie N-terminale qui se débobine un peu dans la protéine sauvage. Toutes les autres structures secondaires sont parfaitement conservées avec un rms de 1.54 Å sur le squelette carboné des feuillets $\beta 2$ à $\beta 5$ entre les deux structures. Toutes les hélices restent parfaitement structurées et conservées durant les simulations néanmoins quelques petits changements locaux sont observés. L'hélice $\alpha 1$ est un peu moins conservée durant les étapes de dynamique moléculaire de la protéine sauvage (rms=1.38 Å) à cause de la petite différence d'orientation due au déplacement du feuillet $\beta 1$. Un effet intéressant a été observé dans l'hélice $\alpha 2$ qui devient plus droite dans P40A (rms=0.91 Å), forçant le site actif à se déplacer et tirant sur toute la structure. L'hélice $\alpha 3$ n'a pas vraiment subi de changements (rms=0.66 Å). Quant à l'hélice $\alpha 4$ de la partie C-terminale, elle se déplace légèrement le long de son axe mais conserve parfaitement sa géométrie (rms=0.88 Å).

Quelques différences dans l'hélice $\alpha 2$ peuvent être détectées aisément autour de la position de la mutation de la proline 40. Le réseau de liaisons hydrogène est plus important dans l'hélice $\alpha 2$ du mutant, permettant à cette hélice de se rapprocher d'une structure plus régulière. L'angle dans cette hélice est de 59° dans WT et de 13° dans P40A. (voir fig.7 de la publication sur P40A en annexe).

La mutation n'affecte pas la structure du site actif car sa géométrie interne est très bien conservée (rms_{rms31,35}=0.93 Å rms_{rms32,35}=0.27 Å le long du squelette carboné), mais elle le déplace de sa position originelle de environ 6Å. Ce déplacement entraîne des contacts très différents dans l'environnement du site actif entre les deux structures. La différence la plus frappante concerne le tryptophane W31. Les contacts hydrophobes qui maintiennent W31 et la boucle entre $\beta 2$ et $\alpha 3$ dans WT ont été supprimés dans P40A, effaçant ainsi les contacts avec I5, P30, I60, E62, D63, P64 et I72. Cette modification induit plusieurs changements dans le réseau de liaisons hydrogène (voir Figure 32), par exemple : dans la thiorédoxine mutée, la liaison W31HNe-I60O disparaît, la liaison W28HNe-K57O est affaiblie tandis que la liaison W28HN-K57O apparaît. Ces changements drastiques modifient la mobilité de W31.

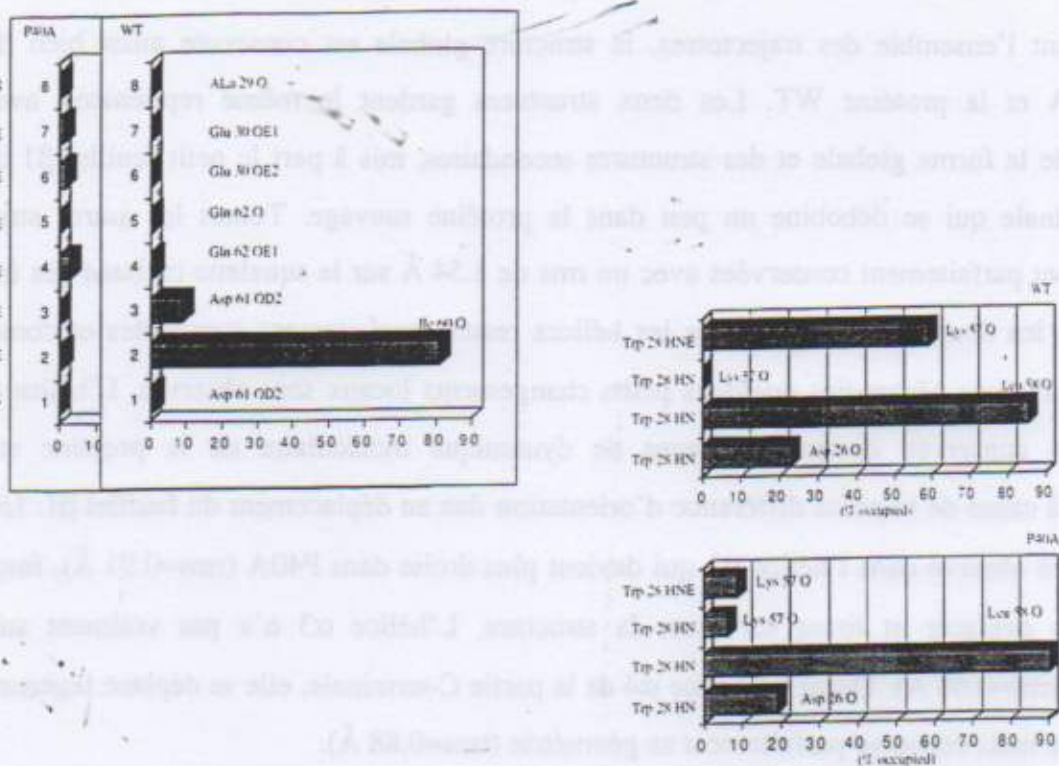


Figure 32 : Présence des liaisons H avec les tryptophanes Trp28 et Trp31 pendant les 100 dernières ps des simulations de dynamiques

Ce résidu se montre beaucoup plus statique dans la structure de WT, coincé dans un chemin hydrophobe constitué de I60, P64 et I72 alors que dans P40A il est plus détaché de la surface de la molécule et libre de ces mouvements. Nous avons calculés la surface accessible au solvant pour les résidus W pendant les dernières 100 ps. La surface d'accessibilité augmente de 45.5 \AA^2 à 74.4 \AA^2 pour W28 et de 112.7 \AA^2 à 233.0 \AA^2 pour W31. L'énergie libre de solvation (Solvation Free Energy - Figure 33) a été calculée sur les cents dernières ps montrant une surface plus hydrophile pour WT que pour P40A avec une différence de $2.9 \text{ kcal.mol}^{-1}$.

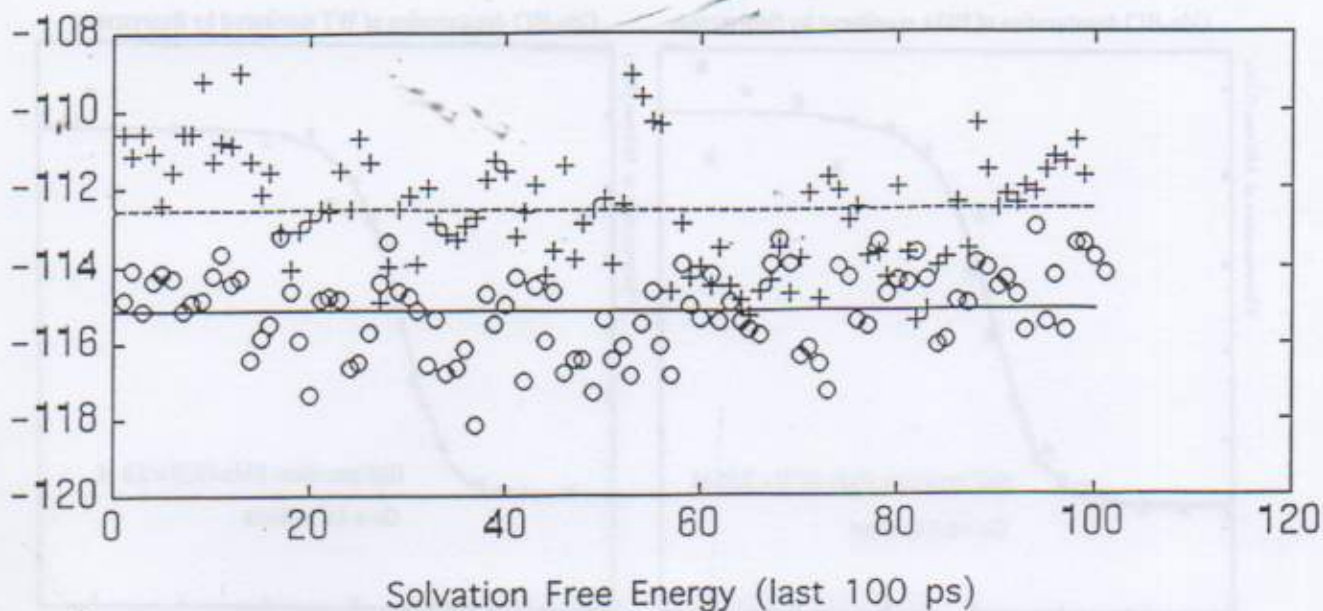


Figure 33 : Courbe de l'évolution de l'énergie libre de solvation (kCal.mol^{-1}) au cours des 100 dernières ps de la dynamique. (o - Wild Type) (+ - P40A)

2. Discussion :

Les effets de cette mutation sur les propriétés biophysiques et biologiques de la thiorédoxine ont été analysées et comparées aux simulations de dynamique moléculaire. La modélisation prévoit que le remplacement de la proline 40 par une alanine induit un déplacement du site actif qui expose le tryptophane 31 au solvant et ouvre une crevasse entre l'hélice α_2 et l'hélice α_3 . Les calculs d'énergie libre de solvation indiquent également que P40A (le mutant) est plus hydrophobe, puisque Trp 31 devient plus accessible. Ces prédictions sont parfaitement en accord avec les résultats expérimentaux. Le mutant P40A a des affinités chromatographiques et des propriétés en fluorescence très différentes de celles de la protéine sauvage WT. La détermination de l'énergie libre de dénaturation de P40A montre que le mutant est globalement moins stable. Néanmoins, les effets de la mutation sur la courbe de transition (voir Figure 34) sont inhabituels puisque la valeur du point à demi-hauteur sur la courbe de transition augmente indiquant que certaines structures locales seraient stabilisées par la mutation.

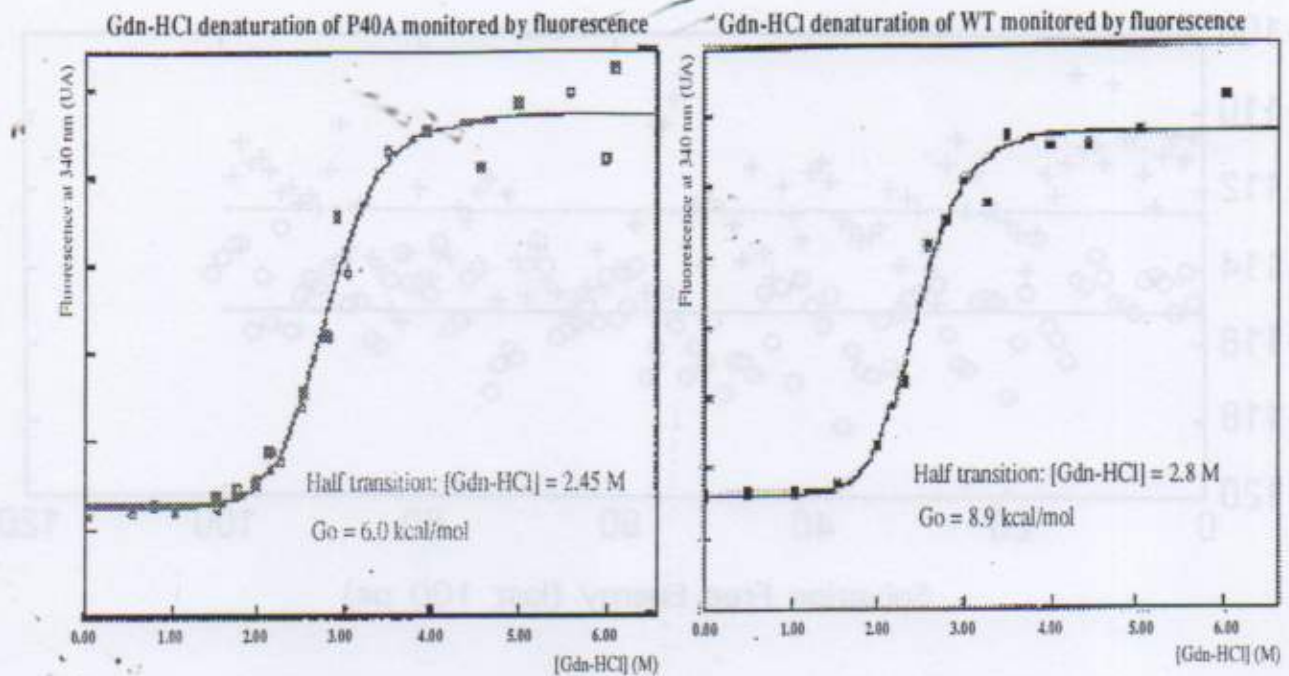


Figure 34 : Courbes de dénaturation au Chlorure de Guanidine pour WT (■ – ■ dénaturation, ● – ● renaturation) suivi par fluorescence à 340 nm. La dénaturation et la renaturation ont été effectuées à 25°C dans un tampon phosphate à pH 7.2.

Ni les capacités de la protéine pour réduire une enzyme chloroplastique, ni sa réactivité vis à vis de la bactérie réductase ne sont affectées, la seule différence fonctionnelle se situe au niveau de la plus grande stabilité de P40A dans l'activation par la lumière du NADP-malate déshydrogénase sous air, ce qui suggère que P40A est moins rapidement ré-oxydé que WT.

D. Conclusion :

Nous avons modifié un résidu extrêmement conservé : la proline 40. Ce résidu, situé près du site actif, induit un coude dans l'hélice $\alpha 2$. Il été muté en alanine, et observé par dynamique moléculaire. La prédiction indique la conservation de la structure globale dans P40A (vérifié par dichroïsme circulaire - voir annexes). La modélisation prévoit que le coude de l'hélice est supprimé dans P40A et qu'une structure plus régulière est adoptée. Le site actif subit un déplacement de 6 Å sans aucune modification de sa géométrie intrinsèque et ce déplacement expose le tryptophane 31 du site actif au solvant. Cela ouvre également une crevasse hydrophobe sur la surface de la protéine entre les hélices $\alpha 2$ et $\alpha 3$. De tels changements impliquent bien entendu la suppression d'un grand nombre de liaisons engagées avec le tryptophane 31. Néanmoins, les résidus qui sont considérés être en interaction avec les protéines cibles (les résidus 33-34, 75-76, 91-93) ([Eklund et al., 1984]) ne subissent pas de

modifications dans notre modélisation, et leurs liaisons hydrogène ne sont pas affectées par la mutation. Cela peut expliquer que biologiquement, les activités de P40A soient identiques à la thiorédoxine sauvage WT : ni sa capacité à réduire les enzymes chloroplastiques, ni sa réactivité avec la bactériale-réductase ne sont modifiées. Le calcul de l'énergie libre de solvation (SFE) indique que le mutant est plus hydrophobe que la forme sauvage, ceci étant dû en partie à l'augmentation de l'accessibilité au tryptophane W31. Dans la forme oxydée de la thiorédoxine sauvage, la fluorescence de W28 et de W31 est connue pour être fortement quenchée à cause du pont disulfure et, dans la forme réduite, on observe une augmentation en fluorescence due aux changements quantiques de W28. La modélisation prévoit une plus grande accessibilité pour W31 et donc un résidu moins quenché qui participerait à une augmentation supplémentaire de la fluorescence dans P40A par rapport à WT. Ce résultat est en accord avec les résultats expérimentaux en fluorescence (voir fig.2 dans la publication en annexe).

Toutes les observations faites sur le mutant concordent avec les habituelles sur les thiorédoxines. Il a également été montré [Hellings et al., 1992] que structurellement les thiorédoxines mutantes possèdent des propriétés thermodynamiques différentes sans que leur activité biologique n'en soit affectée. Ces mutations, localisées dans le cœur hydrophobe de la protéine, furent expérimentées *in vivo* pour l'assemblage du bactériophage filamentueux pour lequel l'activité rédox n'est pas nécessaire. Les auteurs conclurent que, dans les thiorédoxines, le lien entre stabilité et fonction n'était pas si aussi fort que dans les autres systèmes. Plus récemment, le remplacement de W28 dans *E.Coli* a été décrit comme induisant des effets uniquement sur la stabilité et les propriétés non fonctionnelles [Slaby, et al., 1996].

De plus, nos résultats pourraient être reliés à une observation récente faite par Qin et al (1995) sur l'interaction de la thiorédoxine humaine avec des peptides cibles [Qin et al., 1995] et [Qin et al., 1996]. Les auteurs de cette observation décrivent l'ouverture d'une crevasse similaire à la surface de la protéine dans laquelle l'interaction prend place. Cette interaction modifiant évidemment les interactions hydrophobes de la protéine. Ainsi, même si cela est causé par la mutation de la proline 40 en alanine, les résultats observés et les différences qui se sont manifestées pourraient bien refléter le mouvement naturel des thiorédoxines tout comme il a été observé par Qin et al. Cette hypothèse est supportée par le fait que l'activité biologique du mutant n'est pas altérée, la mutation P40A serait en faveur de la conformation que les thiorédoxines adoptent naturellement dans leur interaction avec leur cible.

Chapitre V. Etude de la dénaturation thermique de P40A par RMN :

A. Introduction :

Cette étude a été effectuée dans le but de comparer le comportement thermodynamique de la dénaturation thermique de P40A et de la thiorédoxine d'*Escherichia Coli* non mutée (WT). Cette étude a pu être réalisée par RMN grâce aux facilités d'emploi des appareils pour ce type d'expérimentation et grâce à la présence au laboratoire de logiciels tels que GIFA [Delsuc, 1996], très bien adaptés pour ce type d'études.

B. Les échantillons expérimentés :

I. Préparation et purification de P40A :

a) La culture bactérienne :

La culture (4 litres) a été effectuée dans la bactérie JM109, 20g de cellules ont pu être extraites du milieu après centrifugation.

b) Les étapes de purification :

i) La casse des cellules:

Les 20g de cellules obtenus sont solubilisés après congélation-décongélation dans le tampon de casse BB (Breaking Buffer). Ce tampon est nécessaire pour solubiliser l'ensemble du matériel génétique et protéique contenu dans les cellules. Il est constitué de RN-ase et DN-ase (benzamidine), d'un inhibiteur de métalloprotéase (EDTA), d'un agent réducteur puissant pour éviter l'agrégation des macromolécules par formation de liaisons covalentes S-S (DTT) et d'un produit maintenant le pH de la solution (TRIS pH=8).

BB est préparé de la façon suivante (pour un volume total de 50ml) :

- 5mM EDTA pH=8 (prélèvement de 50 µl d'EDTA 1M pH=8)
- 5mM de benzamidine (prélèvement de 0.0397g)
- 5mM de DTT (0.077g de DTT)
- 50mM de TRIS pH=8 (prélèvement de 5ml TRIS/HCl pH=8)
- Le tout est complété à 50ml avec de l'eau milli Q

50 ml de BB sont nécessaires. Les cellules restent dans un bêcher métallique à 0°C (bain de glace) pour éviter l'activation de réactions de dégradation de la protéine par des protéases. Le tout est mixé et

broyé à 0°C durant 30 min.

ii) La sonication :

Les cellules sont soumises à des ultrasons pendant deux séries de 2 min avec 40 % de puissance, 0.2s d'impulsions suivies de 0.2s de pause, et toujours dans le bain de glace à 0°C sous agitation. Cela permet de terminer la casse des cellules et de libérer tout leur contenu en solution.

iii) Les étapes de centrifugation :

L'ensemble de la solution contenant un grand nombre de débris cellulaires est centrifugé à 16460g (10000 tours par minute) sur un rotor de 147 mm pendant 30 min puis à $\gamma=40000g$ sur un rotor de 112 mm pendant 45 min. Ceci permet d'éliminer toutes les particules insolubles. Le surnageant est conservé pour les étapes de purification suivantes, il représente 75 ml de solution.

iv) La chromatographie d'échange d'ions :

La colonne utilisée est une colonne XK26 (100cm x Ø2,6cm) de chez Pharmacia. Le support est de la "Q Sépharose Fast Flow" et le tampon d'élution est du TRIS/HCl 50mM pH=8, EDTA 5mM (Solution A). La solution B utilisée pour éluer les protéines qui entrent en interaction avec le support est du TRIS/HCl 50mM pH=8, EDTA 5mM, NaCl 1M. Le gradient de NaCl utilisé pour décrocher sélectivement les protéines a été programmé comme suit :

Chargement de 75 ml du surnageant		Lavage avec 80 ml de solution A					
puis programme...	(ml)	(ml)	A	B	fractions	débit total	
	V_{charge}	V_{total}					
	0	100	100%	0%	9 ml	1 ml/min	
	100	101	100%	0%	5 ml	1 ml/min	
	101	500	0% → 100%		80x5 ml	1 ml/min	
	500	500++	100%	0%	9 ml	1 ml/min	

La courbe d'enregistrement obtenue grâce à un détecteur d'absorption dans l'UV ($\lambda=280$ nm) est la suivante (Figure 35) :

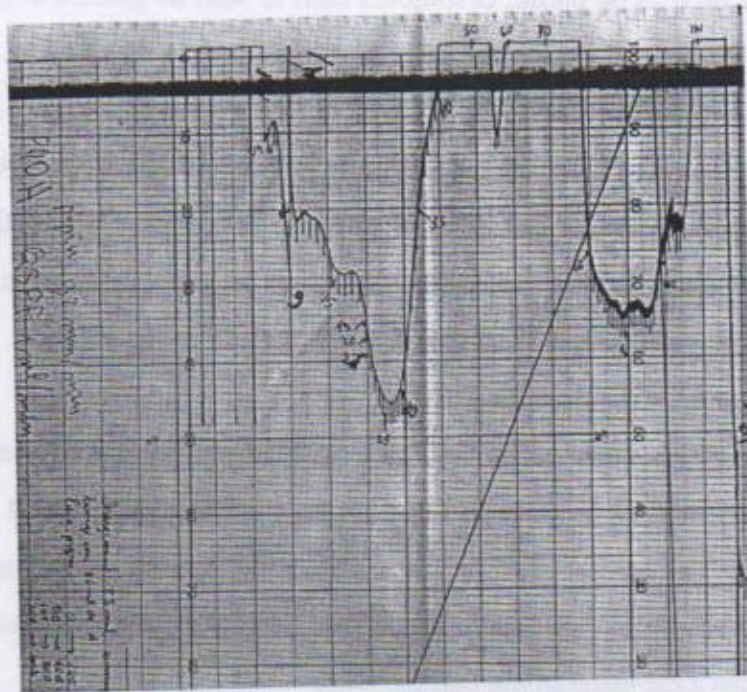


Figure 35 : Enregistrement observés en sortie de la colonne Q Sépharose Fast Flow.

Les diverses fractions obtenues ont été analysées par électrophorèse (voir en annexe page 160) sur gel de polyacrylamide (gel SDS page 15%), puis révélées au bleu de Coomassie (Figure 36).

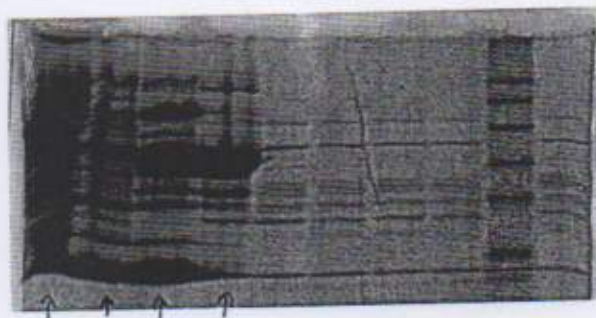


Figure 36 : Gel SDS 15% des fractions 30 à 59

Si on se réfère aux marqueurs de masse moléculaires déposés sur le gel et à la masse de P40A (12.5 kDa), la présence de bandes intenses sur le gel à la masse de 12.5kDa indique que les fractions

40 à 59 se révèlent être les fractions susceptibles de contenir la protéine mutée. Néanmoins, P40A est loin d'être pure puisque le nombre de bandes obtenues sur le gel est considérable et montrent la présence d'autres macromolécules, ces fractions sont donc rassemblées en vue d'une étape supplémentaire de séparation.

v) La 1^{ère} étape de reconcentration :

Cette étape est reliée à l'étape de purification suivante. Etant donné que la manipulation de trop grands volumes de liquide en chromatographie d'exclusion provoque un étalement des pics et une mauvaise séparation, il est nécessaire pour avoir une bonne séparation de réduire le volume de solution à filtrer à 1% du volume total de la colonne de filtration utilisée (soit un maximum de 2.4 ml dans notre cas).

Les fractions 40 à 59 sont reconcentrées sur une membrane filtrant les masses moléculaires de moins de 3kDa (MWCO YM3). Un microconcentrateur (Amicon) de 50 ml, puis de 10 ml sont utilisés pour cette opération de façon à ramener le volume des 19 fractions rassemblées à 2 ml.

vi) La chromatographie d'exclusion : (filtration sur gel)

La colonne utilisée est une colonne (120cm x Ø1.6cm) de chez BIORAD. Le support est du Sépharose Haute résolution (S100HR). Les 2 ml de solution concentrée sont déposés en haut de cette colonne et sont élués avec du TRIS/HCl 50mM pH=8 à un débit de 1.1 ml/min. (Voir Figure 37).

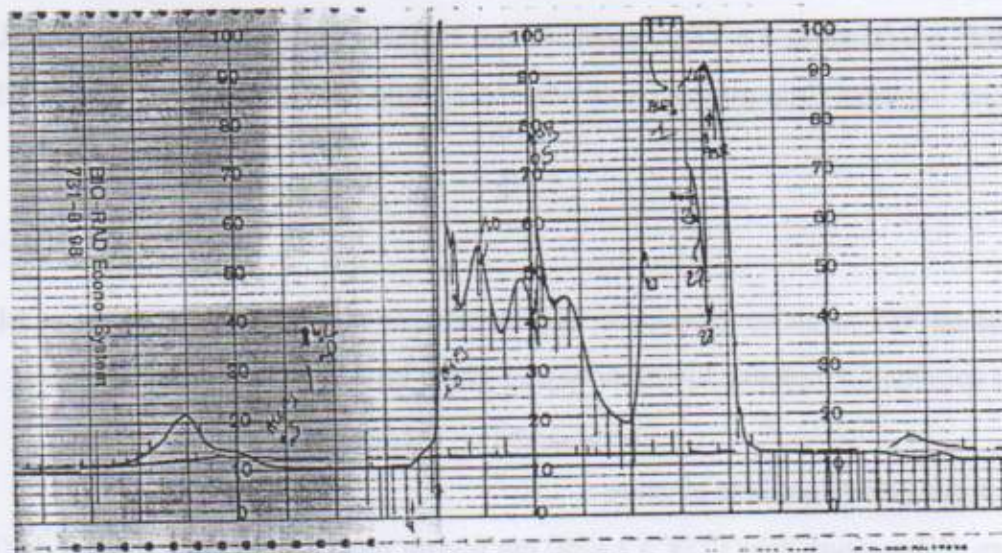


Figure 37 : Courbe d'élué des protéines en sortie de la colonne S100HR

Comme précédemment, les diverses fractions sont analysées par électrophorèse sur gel de polyacrylamide (gel SDS page 15% - Figure 38) et révèlent la présence d'une protéine correspondant à la masse moléculaire attendue dans les fractions 23 à 26.

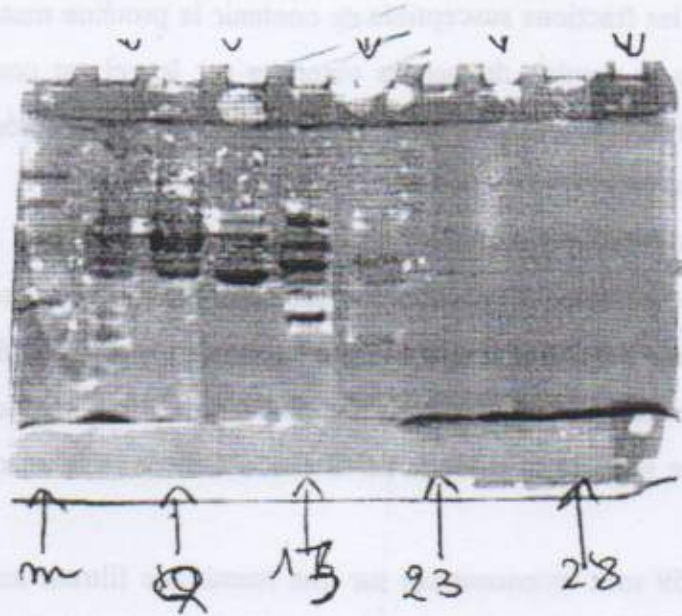


Figure 38 : Gel SDS Page 15% des fractions prélevées en sortie de la colonne de filtration.

vii) La 2^{ème} reconcentration :

Les fractions 23 à 26 sont rassemblées et reconcentrées sur membrane MWCO YM3 dans un amicon de 10 ml. Un centricon de 2 ml est ensuite utilisé pour ramener le volume total à 1 ml. Et finalement une solution de K_2HPO_4/HCl 0.1M pH=5.7 est utilisée pour changer le tampon de la protéine par lavage successifs en diluant à 2ml puis reconcentrant à 1ml dix fois de suite. De cette façon la concentration en TRIS de 50 mM au départ est ramenée à une valeur nettement inférieure à 1 mM ($50 \text{ mM}/2^{10} = 0.05 \text{ mM}$). Le TRIS, molécule qui est loin d'être « transparente » en RMN, n'est alors plus qu'une impureté (impureté qui sera néanmoins utile pour calibrer les spectres RMN par la suite).

Remarque :

Dans le cas où WT serait en proportion trop grande par rapport à P40A, il est possible de les séparer sur une colonne C4, colonne hydrophobe en « phase inverse ». (voir publication.)

Phényl chromatographie d'Interaction hydrophobe

viii) La quantité de protéine obtenue :

Une mesure d'absorption dans l'UV à $\lambda=280 \text{ nm}$ fournit une DO (densité optique) de 0.790 dans une cuve dont le trajet optique est de 0.5 mm, ceci pour 13 μl de notre millilitre de préparation qui ont été dilués à 500 μl pour la mesure. Donc compte tenu de la valeur calculée du $\epsilon=0.769$ (1 mg de protéine par ml pour une DO de 1.3) nous pouvons en déduire notre quantité finale de P40A à environ 93.5 mg.

Nous avons voulu vérifier si la solution de P40A sur laquelle nous avons travaillé était uniquement composée de P40A sans WT, de tampon phosphate et de quelques traces de TRIS. Dans *E.Coli*, il existe également la forme sauvage (WT) de notre mutant et on peut se poser la question qui est de savoir s'il la solution obtenue ne contient pas la forme naturelle WT qui serait bien sûr à séparer.

Le clone qui nous a été fourni¹ pour la production de P40A a été conçu pour obtenir la protéine mutée de façon prépondérante (sauvage/mutant = 1/200). Pour vérifier si P40A était majoritaire dans la solution, nous avons effectués un spectre de masse et un spectre RMN de la solution de P40A.

2. Analyse par spectrométrie de masse :

Une partie de la solution contenant P40A a été observée par spectrométrie de masse (Figure 39) indiquant la présence d'un pic majoritaire à 70% d'une masse de 11640 g.mol^{-1} et d'un pic minoritaire à 30% correspondant à une masse de 11850 g.mol^{-1} .

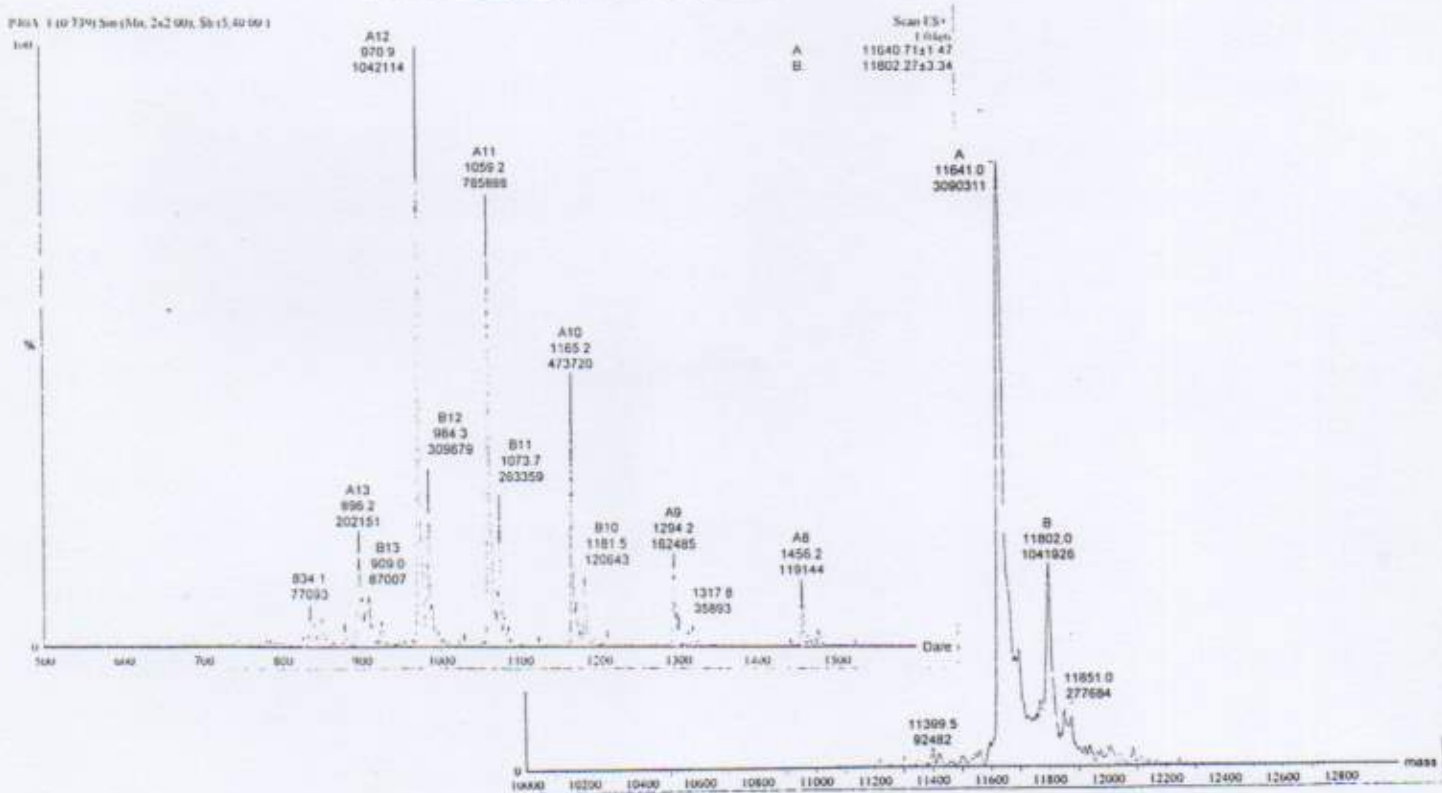


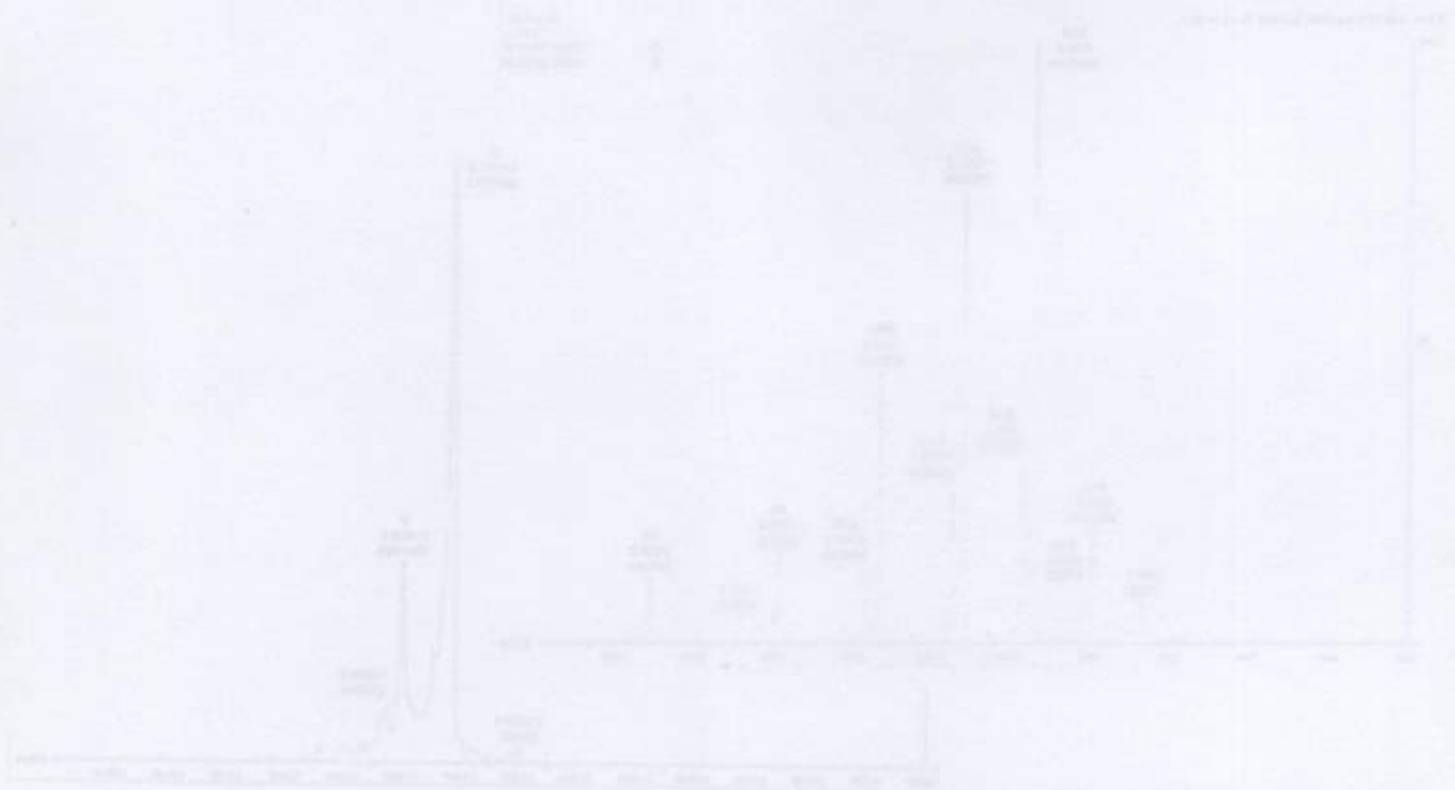
Figure 39 : Spectre de masse obtenue pour la solution contenant la protéine mutée.

¹ Par l'INRA de Montpellier et par l'Institut de Recherche sur les plantes d'ORSAY.

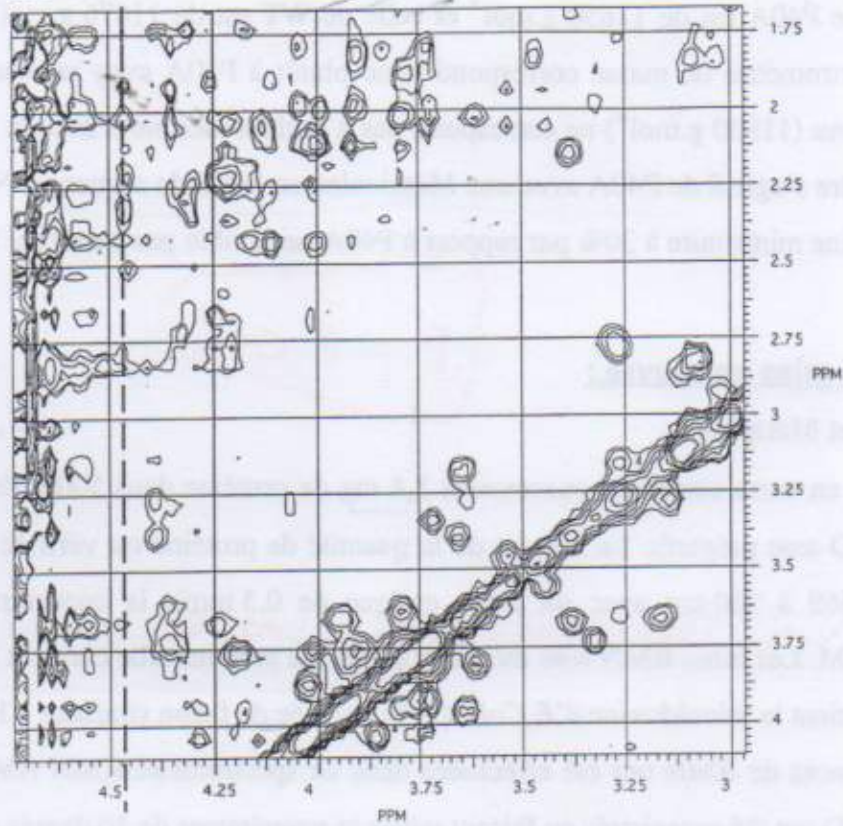
3. Analyse par RMN :

L'attribution de la thiorédoxine d'*E.Coli* étant connue, nous avons décidé de rechercher la présence des pics de corrélation de la proline 40 dans le spectre du mutant. Les déplacements chimiques des hydrogènes dans WT à pH=5.7, T=308K sont 4.5, 2.33, 1.92, 2.03, 2.08, 3.64, et 3.82 ppm respectivement pour H α , H β 1, H β 2, H γ 1, H γ 2, H δ 1, et H δ 2 [Dyson, et al., 1989]. Sur le spectre de la thiorédoxine WT, les pics des couplages scalaires entre H α et H δ 1, H δ 2 ne sont pas observés ce qui peut s'expliquer par une distance trop grande entre les hydrogènes (voir Figure 40).

Les deux spectres ont des rapports signal/bruit différents et sont assez difficiles à comparer. Une comparaison de deux expériences de TOCSY aurait également été plus judicieuse. Après avoir regardé précisément les spectres, nous n'avons pas retrouvé le système de spins de la Proline 40 dans la solution de P40A alors qu'il semble être présent dans WT, il faut noter que l'intensité du signal de l'eau dans cette zone du spectre (4.5 ppm) gêne l'analyse.



Spectre de WT |



Spectre de P40A |

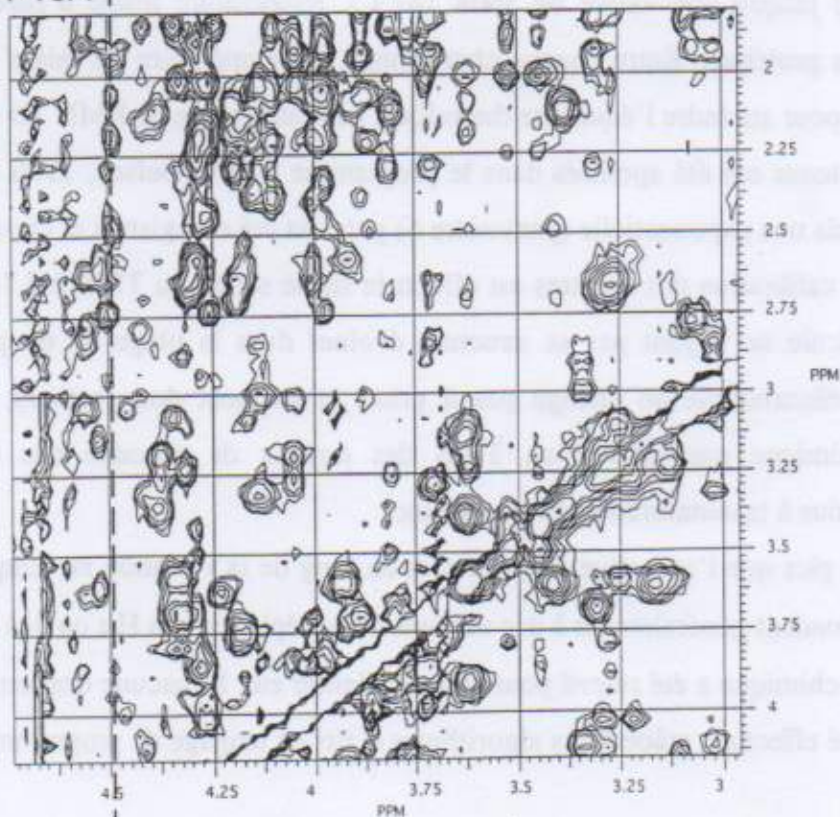


Figure 40 : Spectres NOESY de la thioredoxine WT et de P40A dans la zone du spectre RMN correspondant à la Proline 40. Le spectre de P40A est celui du bas.

4. Discussion :

La masse de P40A est de 11650 g.mol^{-1} et celle de WT est de 11676 g.mol^{-1} . Le pic majoritaire obtenue en spectrométrie de masse correspond donc plutôt à P40A avec sa masse de 11640 g.mol^{-1} . L'autre pic obtenu (11850 g.mol^{-1}) ne correspond pas à la thiorédoxine d'*E.Coli*. c'est donc une autre protéine. Peut-être s'agit-il de P40A avec une Méthionine en début de séquence. Nous ne le savons pas mais cette protéine minoritaire à 30% par rapport à P40A ne semble pas perturber les spectres de RMN obtenus.

C. Expérience mise en œuvre :

1. Matériel et Méthode :

Deux tubes en verre contenant exactement 3,4 mg de protéine dans $500 \mu\text{l}$ de $\text{K}_2\text{HPO}_4/\text{HCl}$ 0.1M $\text{pH}=5.7$, $8\% \text{ D}_2\text{O}$ sont préparés. La mesure de la quantité de protéine est vérifiée par absorption dans l'UV ($\text{DO}=0.0569$ à 280 nm avec un trajet optique de 0.5 mm), la concentration est bien celle attendue : 0.5 mM . Les tubes RMN sont mis sous argon. Le premier tube contient le mutant P40A et le second tube contient la thiorédoxine d'*E.Coli* (WT) purifiée de façon similaire à P40A.

Les expériences de RMN ont été effectuées dans un spectromètre RMN AMX 600 Bruker. Des spectres RMN 1D ont été enregistrés en faisant varier la température de 10 degrés en 10 degrés à partir de 300K (27°C) jusqu'à une valeur de 360K (87°C), température limite à partir de laquelle il y a précipitation des protéines. Entre chaque changement de température on laisse reposer l'échantillon pendant 30 min pour atteindre l'équilibre thermique, puis un spectre de RMN 1D est enregistré.

Les FID obtenus ont été apodisés dans le programme GIFA [Delsuc, 1996] par une gaussienne (paramètre 5) puis une exponentielle (paramètre 6) puis ont été enregistrés et complétés avec des zéros jusqu'à 32K . La calibration des spectres est effectuée sur le signal du TRIS à 3.7 ppm . Le TRIS étant une petite molécule ne voyant pas sa structure évoluer dans la plage de température étudiée, son environnement électronique ne change pas a priori et on peut donc estimer que la variation de déplacement chimique par rapport au TRIS des noyaux de chacune des deux protéines sera essentiellement due à la dénaturation de la protéine.

Chacun des pics que l'on arrive à suivre tout au long de la variation en température a été étudié. Ces pics correspondent généralement à des méthyles très déplacés, des $\text{H}\alpha$ ou des $\text{HN}\epsilon$ de tryptophane. Le déplacement chimique a été relevé pour chacun d'entre eux à chacune des températures. L'analyse des résultats a été effectuée grâce à des algorithmes écrits en langage de programmation Matlab.

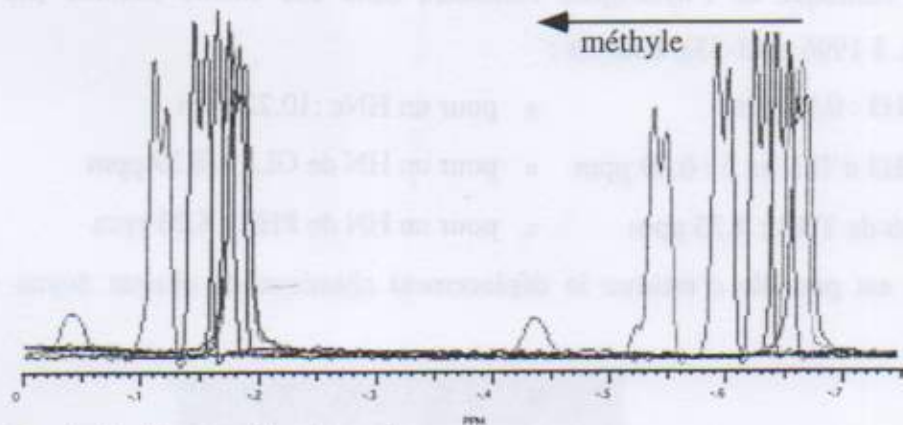
D-E-C-L

0,8 mg/ml

2. Résultats sur la thiorédoxine d'*E.Coli* (WT) et P40A :

<i>E.Coli</i> température	déplacement chimique (ppm)						
	300 K	310 K	320 K	330 K	340 K	350 K	360 K
1 ^{er} méthyle	-0.6965	-0.6945	-0.6844	-0.6800	-0.6510	-0.6003	-0.5132
2 ^{ème} méthyle	-0.2057	-0.2091	-0.2074	-0.2112	-0.1955	-0.1634	-0.1060
3 ^{ème} méthyle	0.0277	0.0469	0.0545	0.0548	0.0700	0.0912	0.1137
4 ^{ème} méthyle δ	0.1750	0.1757	0.1808	0.1754	0.1846	0.1981	0.2096

$D = \epsilon c l$
0,030, 0,564



Etude de dénaturation thermique;
évolution des raies dans les spectres 1D au cours de
l'augmentation de température.

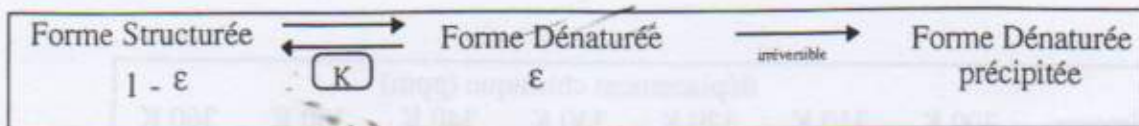
Figure 41

P40A température	déplacement chimique (ppm)						
	300 K	310 K	320 K	330 K	340 K	350 K	360 K
1 ^{er} méthyle	-0.6621	-0.6567	-0.6458	-0.6266	-0.5948	-0.5388	-0.4357
2 ^{ème} méthyle	-0.1793	-0.1791	-0.1752	-0.1655	-0.1468	-0.1114	-0.0415
3 ^{ème} méthyle	0.0510	0.0564	0.0686	0.0808	0.0975	0.1235	0.1538
4 ^{ème} méthyle δ	0.1980	0.2021	0.2082	0.2167	0.2274	0.2398	0.2547

3. Analyse :

On s'intéresse aux courbes $\delta(\text{ppm})=f(T)$ de variation du déplacement chimique en fonction de la température (Figure 41) pour en déduire des paramètres thermodynamiques caractérisant la dénaturation.

Pour cela, nous commençons par considérer la dénaturation thermique comme étant un équilibre chimique entre la forme structurée et la forme dénaturée, cet équilibre évoluant vers la forme dénaturée lorsque l'on chauffe.



ϵ est l'avancement de la dénaturation.

$K = (a_{\text{forme dénaturée}} / a_{\text{forme structurée}})$ est la *constante d'équilibre* thermodynamique entre la forme structurée et la forme dénaturée.

La valeur du déplacement chimique de la forme dénaturée ($\epsilon=1$), que l'on ne peut pas mesurer sur notre protéine puisqu'elle précipite à haute température, est fixée à la valeur théorique qu'aurait le déplacement chimique de l'hydrogène considéré dans une chaîne étendue (réf Nature Structural Biology, Vol. 3 1996 p.45-53), à savoir :

- o pour un CH3 : 0.94 ppm
- o pour un HNe : 10.22 ppm
- o pour un CH3 d'ILE en δ : 0.89 ppm
- o pour un HN de GLY : 8.39 ppm
- o pour un H α de THR : 4.35 ppm
- o pour un HN de PHE : 8.23 ppm

Ainsi, il est possible d'estimer le déplacement chimique de chaque noyau observé suivant la relation :

$$\delta_{\text{ppm}} = \delta_s + \epsilon \cdot (\delta_b - \delta_s)$$

où δ_s est le déplacement chimique observé dans la forme structurée et δ_b le déplacement chimique dans la forme étendue.

Nous pouvons alors en déduire une courbe $\epsilon = f(T)$ qui est une courbe croissante dans l'intervalle [0...1] et qui représente l'avancement de la dénaturation en fonction du temps. C'est de cette courbe que nous allons extraire les paramètres thermodynamiques de la dénaturation : ΔH , ΔS et ΔG .

Pour cela reprenons les lois de la thermodynamique chimique et plus particulièrement celle qui concerne la constante d'équilibre K. Elle vérifie la relation :

$$K = e^{-\Delta G/RT}$$

Or :

$$K = ([a_{\text{forme dénaturée}}] / [a_{\text{forme structurée}}]) = [\epsilon / (1-\epsilon)]$$

et sachant que $\Delta G = \Delta H - T \cdot \Delta S$, on en déduit l'évolution de ϵ en fonction de la température.

$$\epsilon = \frac{e^{\Delta S/R} \cdot e^{-\Delta H/RT}}{1 + e^{\Delta S/R} \cdot e^{-\Delta H/RT}}$$

Il ne reste donc plus qu'à déterminer les valeurs de ΔS et ΔH qui permettent de se rapprocher au mieux des valeurs expérimentales obtenues. L'optimisation des ces valeurs est effectuées dans MATLAB (voir Figure 42, Figure 43 et Figure 44) pour chacun des pics que nous avons pu suivre. $T_{1/2}$ (voir Tableau 4) représente la valeur estimée de la température de demi-dénaturation pour $\epsilon = 0.5$.

Noyau	Protéine	ΔH (kcal.mol ⁻¹)	$T_{1/2}$	ΔS (cal.mol ⁻¹ .K ⁻¹)	ΔG_0 (kcal.mol ⁻¹)	$\Delta\Delta G_0(298K)$
1 ^{er} méthyle VAL16	WT	75.89	392	193.6	18.19	-2.03 kcal.mol ⁻¹
	P40A	67.78	391.3	173.2	16.16	
2 ^{ème} méthyle VAL16	WT	111	384.3	288.9	24.91	-6.18 kcal.mol ⁻¹
	P40A	80.53	388.4	207.4	18.73	
3 ^{ème} méthyle LEU103	WT	35.93	443.8	80.99	11.8	1.15 kcal.mol ⁻¹
	P40A	46.24	414.0	111.7	12.95	
4 ^{ème} méthyle VAL25	WT	64.59	418.5	154.3	18.59	-5.34 kcal.mol ⁻¹
	P40A	41.54	437.6	94.94	13.25	

Tableau 4 : Valeurs obtenues pour les paramètres thermodynamiques des méthyles observés au cours de la dénaturation.

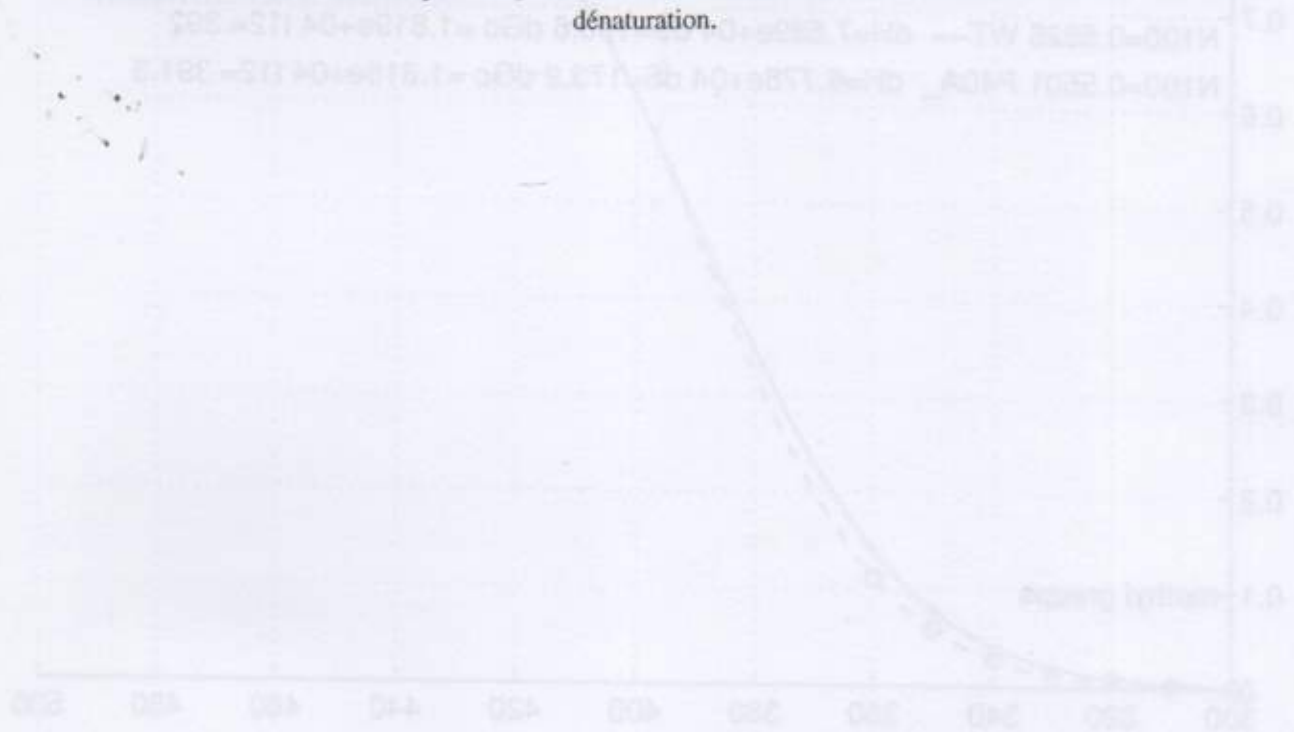


Figure 43 : Représentation de l'ajustement de la dénaturation de la thiorédoxine de la souches WT et P40A. Les courbes sont ajustées avec les paramètres thermodynamiques obtenus dans le Tableau 4. Les données expérimentales sont représentées par des points noirs et les courbes continues sont les ajustements. Les courbes en pointillés sont les ajustements pour les méthyles 1 et 2, et les courbes en traits pleins sont les ajustements pour les méthyles 3 et 4.

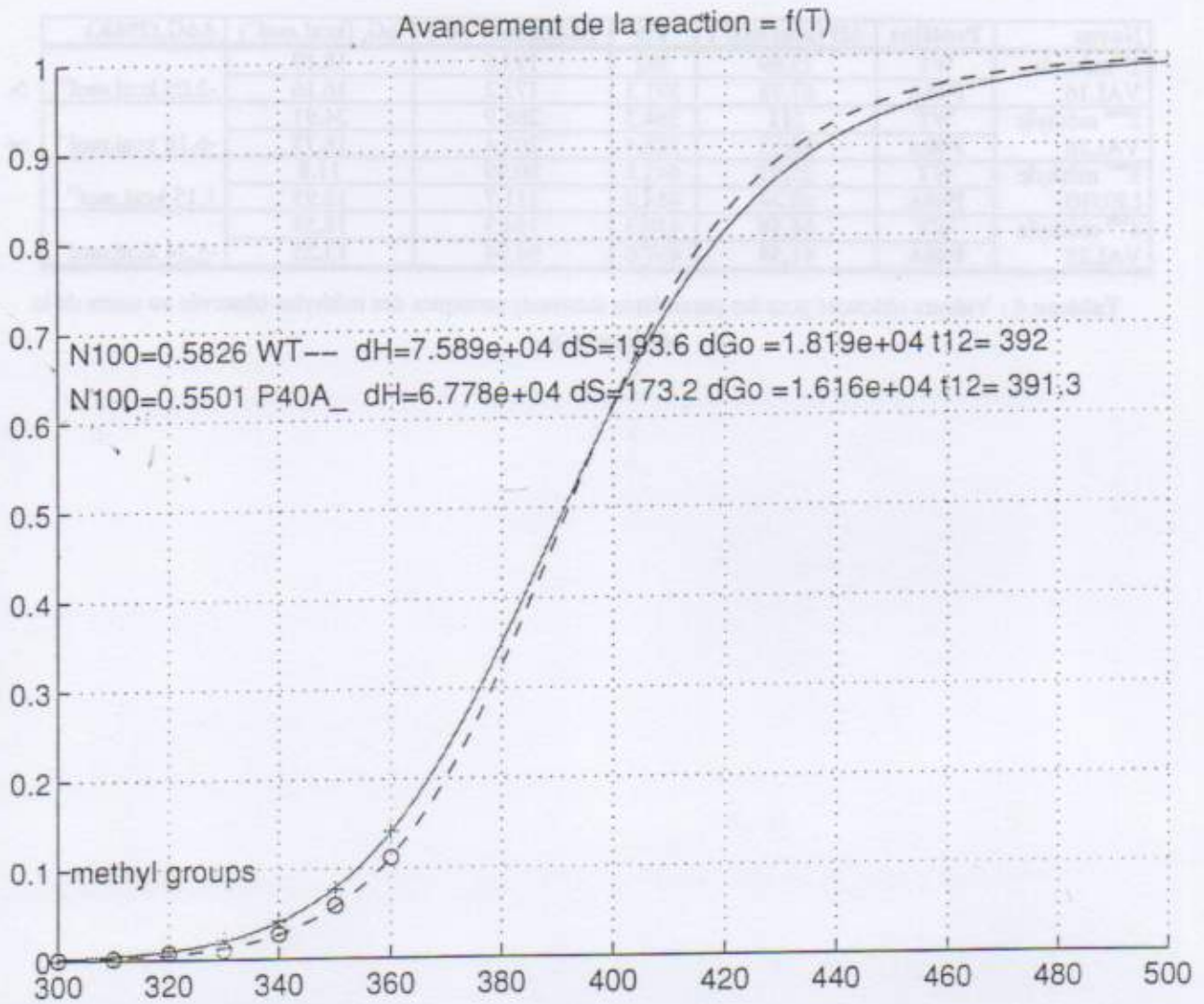


Figure 42 : Représentation de l'avancement de la dénaturation de la Valine 16 (méthyle 1) durant la montée en température. La courbe calculée de la forme sauvage (WT) est en pointillée, l'exactitude de l'approximation est à 5.86^{E-09} . La courbe en trait continu représente l'évolution du mutant P40A, l'exactitude de l'approximation est à 5.50^{E-09} . dH représente l'enthalpie libre ; dS est l'entropie ; ΔG_0 est l'énergie libre ; et t_{12} est la température de $\frac{1}{2}$ dénaturation.

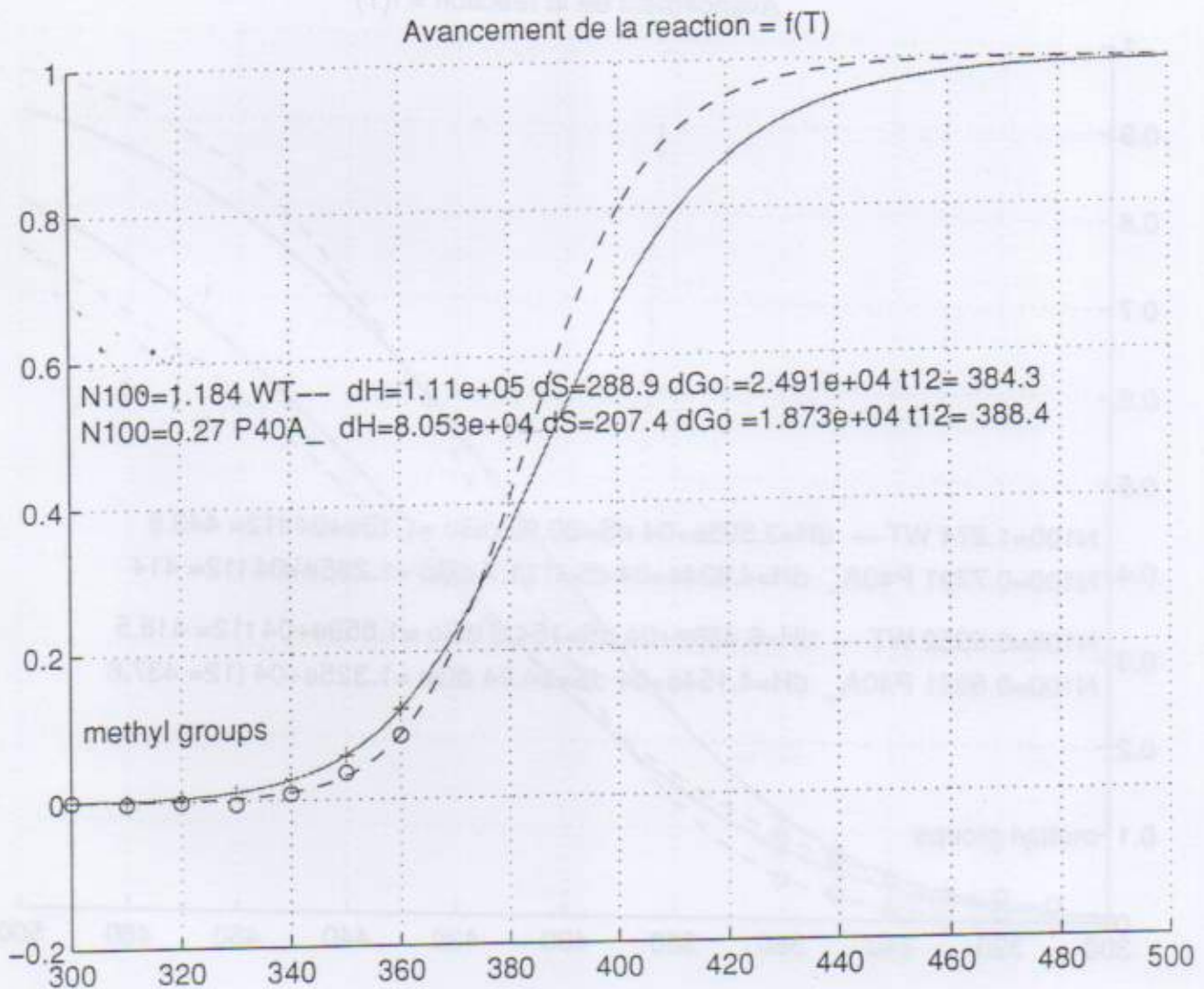


Figure 43 : Représentation de l'avancement de la dénaturation de la Valine 16 (méthyle 2) durant la montée en température. La courbe calculée de la forme sauvage (WT) est en pointillée, l'exactitude de l'approximation est à 11.84^{E-03} . La courbe en trait continu représente l'évolution du mutant P40A, l'exactitude de l'approximation est à 2.7^{E-03} . dH représente l'enthalpie libre ; dS est l'entropie ; ΔG_0 est l'énergie libre ; et t_{12} est la température de $\frac{1}{2}$ dénaturation

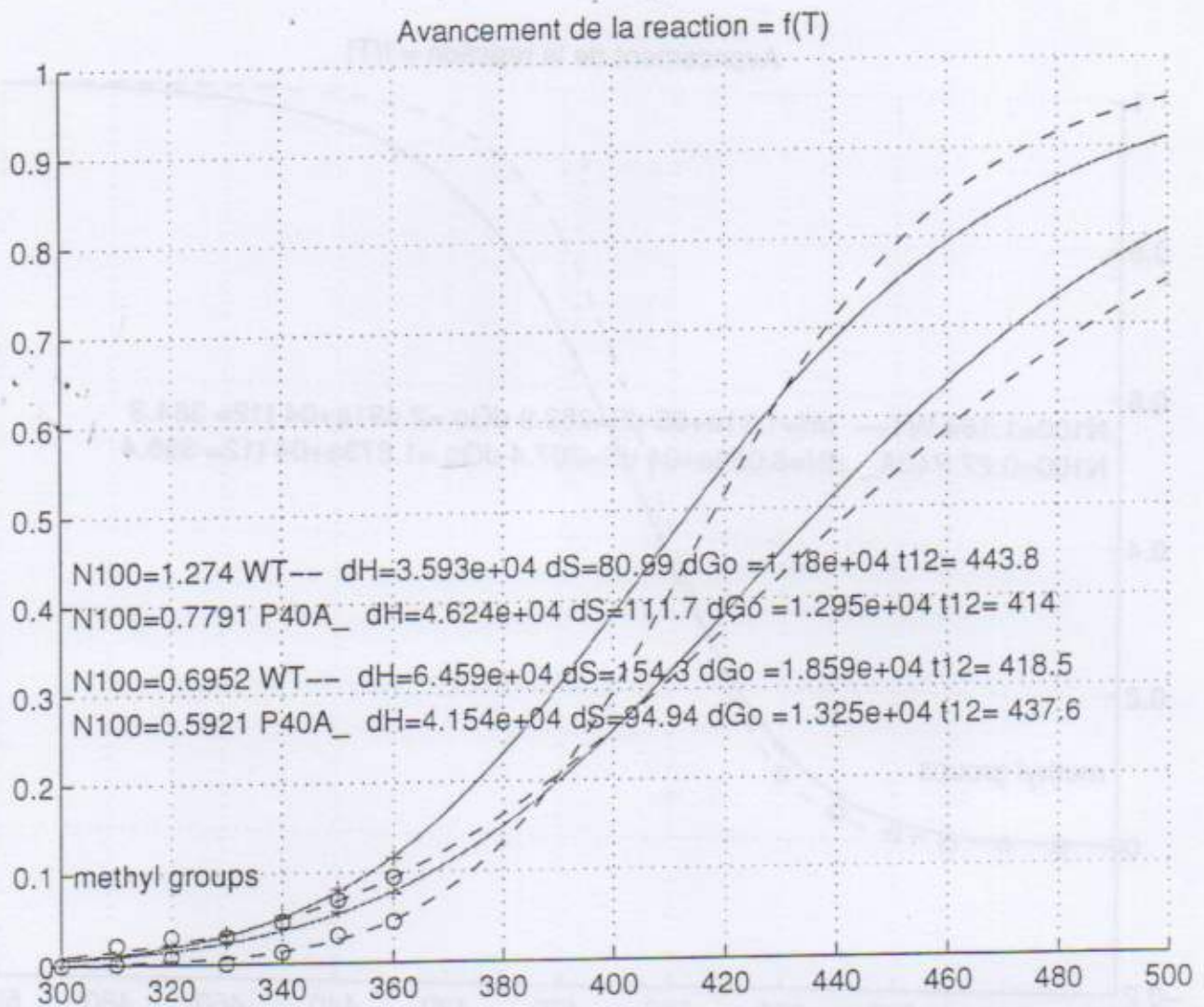


Figure 44 : Représentation de l'avancement de la dénaturation de la Valine 25 (méthyle 1) et de la Leucine 103 (Méthyle) durant la montée en température. La courbe calculée de la forme sauvage (WT) est en pointillée, l'exactitude des approximations est à 12.74^{E-03} (LEU103) et à 6.95^{E-03} (VAL25). La courbe en trait continu représente l'évolution du mutant P40A, l'exactitude de l'approximation est à 7.79^{E-03} (LEU103) et à 5.92^{E-03} (VAL25)

dH représente l'enthalpie libre ; dS est l'entropie ; ΔG_0 est l'énergie libre ; et t_{12} est la température de $\frac{1}{2}$ dénaturation

4. Discussion :

Les résultats obtenus en utilisant cette technique ne sont réellement appréciables que dans le cas où ceux-ci montrent une certaine concordance. Or si nous considérons tous les spins pouvant être suivis dans le spectre RMN 1D, certaines des courbes obtenues sont totalement différentes des autres (HN de GLY65, PHE81, THR.1 et HNe des TRP- non présenté). Nous nous sommes alors posés la question de savoir s'il fallait prendre en compte tous les pics détectables. En effet, les hydrogènes amides et les autres hydrogènes formant une liaison covalente avec un azote peuvent s'échanger avec l'eau. Le pic observé en RMN pour l'eau est un pic qui se déplace considérablement en fonction de la température, et tous les protons qui s'échangent fortement avec l'eau ont tendance à suivre ce déplacement dans le spectre RMN, les résultats obtenus sur ces protons-là ne suivent donc plus du tout les équations précédentes puisque d'autres facteurs interviennent. Ainsi les variations détectées pour les HN en général ne seront pas prises en compte.

La dénaturation observée est surtout caractérisée par le déplacement des méthyles. En effet, ceux-ci maintiennent la structure tridimensionnelle de la protéine grâce à des interactions hydrophobes et, étant relativement faibles car elles sont fonctions de la distance en $1/r^6$, ce sont les premières détruites lors de la dénaturation et donc, celles qu'il convient d'étudier dans notre cas.

Sur les quatre méthyles analysés, nous constatons que deux d'entre eux (LEU103 et VAL25) fournissent des résultats où la courbe de dénaturation obtenue a une pente plus faible. Cela augmente ainsi la valeur de la température $T_{1/2}$ de demi dénaturation qui devrait être à priori unique pour la protéine. Les courbes obtenues pour ces 2 méthyles ne permettent par ailleurs pas l'obtention d'un avancement ϵ nul à 300 K, comme cela est intrinsèquement supposé dans notre démarche puisque notre plus basse température est de 300 K. Nous avons des points qui sont pratiquement alignés et où il est alors bien difficile de déterminer précisément la courbe $\epsilon=f(T)$, une droite ferait aussi bien l'affaire. Ces courbes concernant la Leucine 103 et la Valine 25 doivent donc être prises en compte avec un certain recul, voir même ne pas être retenues du tout pour le calcul des valeurs thermodynamiques.

Nous constatons que sur chacune des courbes comparées (Figure 42, Figure 43 et Figure 44), les 'ronds' sont en dessous des 'croix'. Cela signifie donc que les déplacements chimiques de P40A sont toujours plus importants que ceux de WT. L'avancement de la dénaturation est donc plus rapide pour P40A que pour WT. Nous en concluons donc que P40A se dénature avant WT.

En ne considérant que les méthyles avec un ϵ nul à 300 K et des valeurs expérimentales un peu plus significatives, la valeur moyenne des différences d'énergies libres standards calculées (à 298 K) est de $\Delta\Delta G_0 = -4.1 \text{ kCal.mol}^{-1}$ et, la température $T_{1/2}$ de demi-dénaturation pour P40A est de 390 K et pour WT de 388 K. On peut également noter que la valeur du $\Delta\Delta G_0$ observé, d'un ordre de grandeur

raisonnable [Hellinga, et al., 1992], est proche de la valeur mesurée par dénaturation au Chlorure de guanidine ($-2.9 \text{ kCal.mol}^{-1}$ - voir publication sur les effets de la mutation P40A page 167). La petite différence de $\Delta\Delta G_0$ implique une différence de température de demi dénaturation mais ce qui est intéressant de remarquer est que l'on retrouve le même ordre de grandeur pour cette variation d'énergie libre en ayant employé deux méthodes différentes.

La valeur obtenue pour la différence d'énergie libre entre P40A et WT est donc $\Delta\Delta G_0 = \Delta G_0(\text{P40A}) - \Delta G_0(\text{WT}) < 0$, ce qui implique $K(\text{P40A}) > K(\text{WT})$. La protéine mutée serait donc moins stable vis à vis d'une augmentation de température que la protéine sauvage. Néanmoins, on constate que la température $T_{1/2}$ pour P40A est supérieure à celle de WT ($\Delta T = 2 \text{ K}$ en faveur de P40A). Cela montre ainsi que la protéine mutée résiste un peu plus longtemps à l'augmentation de température avant de se trouver complètement dénaturée. Ainsi il y aurait bien une stabilisation de certaines structures secondaires locales dans P40A puisque nous constatons une dénaturation plus lente pour P40A malgré une déstabilisation globale de la structure ($-4.1 \text{ kCal.mol}^{-1}$).

Au niveau entropique, $\Delta S_0^{\text{moy}}(\text{WT}) = 241.2 \text{ Cal.mol}^{-1}.\text{K}^{-1}$ et $\Delta S_0^{\text{moy}}(\text{P40A}) = 190.3 \text{ Cal.mol}^{-1}.\text{K}^{-1}$ ce qui indique une diminution de l'entropie dans P40A par rapport à WT. Au niveau enthalpie, nous avons également une diminution : $\Delta H_0^{\text{moy}}(\text{WT}) = 93.4 \text{ kCal.mol}^{-1}$ et $\Delta H_0^{\text{moy}}(\text{P40A}) = 74.2 \text{ kCal.mol}^{-1}$. Nous sommes donc dans un cas où ΔH et ΔS diminuent et où ΔG_0 reste constant car il y a une compensation entre ΔH et $-\Delta S$. La diminution de ΔH indique une diminution des interactions stabilisantes dans P40A. La diminution de ΔS indique que P40A a plus de degrés de liberté disponibles, mais cette diminution de ΔS est en contradiction avec le fait que la surface hydrophobe augmente dans P40A par rapport à WT (voir publication). Nous devrions voir l'entropie de notre système augmenter à cause des molécules d'eau (elles n'apprécient pas les zones hydrophobes) et cela devrait plutôt être un indicateur de la disparition de degrés de liberté. Néanmoins, il se peut aussi que l'entropie perdue à cause de l'accroissement de la surface hydrophobe soit compensée par une meilleure structuration de l'hélice $\alpha 2$.

P40A serait donc mieux organisée, et donc plus structurée que WT au niveau de ses structures secondaires.

Si nous nous intéressons maintenant à la position des groupements méthyles étudiés au sein de la thiorédoxine elle-même, la VAL16 est située dans la partie terminale de l'hélice $\alpha 1$ avec ses méthyles orientés vers le cœur hydrophobe de la protéine, aidant au maintien de cette hélice sur le reste de la structure. La LEU103 est placée dans l'hélice C-terminale $\alpha 4$ avec sa chaîne latérale orientée elle-aussi vers le cœur hydrophobe. Quant à la VAL25, elle est placée dans le milieu du feuillet $\beta 2$ au cœur même de la protéine juste à côté de l'ASP26, [Knut, et al., 1991] [Ladbury, et al., 1993] [Jeng and

Dyson, 1996] [Wilson, et al., 1995] qui favorise la déstabilisation de la protéine. Les résultats que nous avons sélectionnés concernent la VAL16 et sont peut-être, tout en étant associés à la dénaturation de la protéine, plus particulièrement associés à celle de la partie N-terminale au niveau de l'hélice α_2 , ce qui pourrait expliquer les différences observées entre les différents groupes méthyles.

Les résultats obtenus sur la VAL16 dans cette étude de dénaturation thermique permettent de retrouver les mêmes différences entre WT et P40A que celles qui sont décrites dans la publication page 167. Néanmoins, nous sommes bien conscients du fait que le nombre de points expérimentaux dont nous disposons est relativement faible. Des conditions de pH ou de concentration différentes auraient peut-être pu permettre l'obtention de points supplémentaires. Nous savons également qu'il est bien difficile de généraliser (sur les protéines P40A et WT) un résultat obtenu sur simplement deux points d'un ensemble (les groupements méthyles de la VAL16).



Chapitre VI. Etude structurale de P40A par RMN :

A. Introduction :

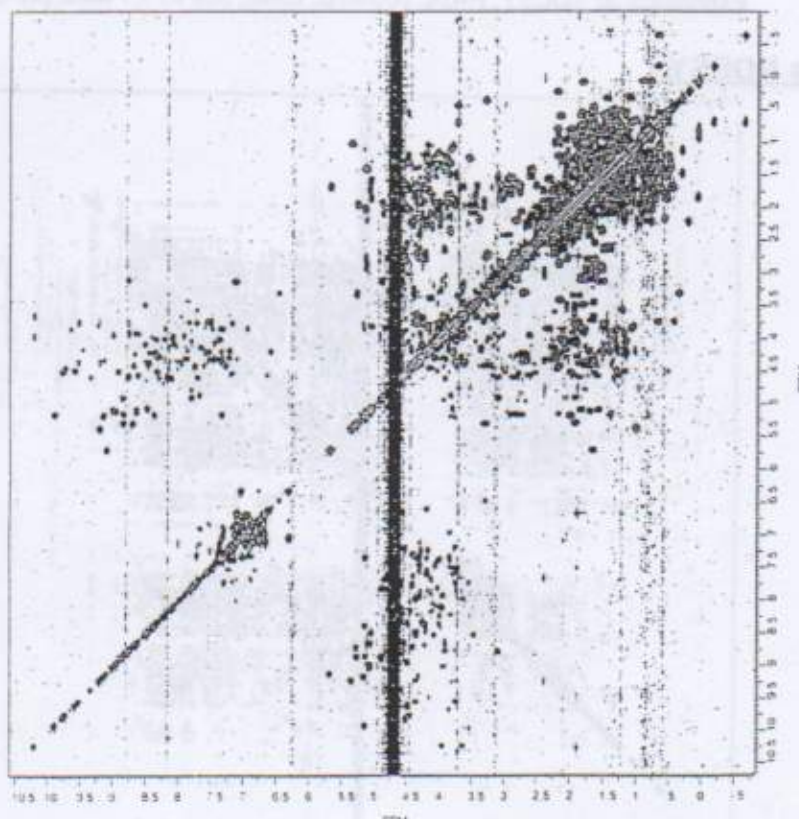
Une étude structurale de la protéine mutée a donc été envisagée dans le but d'observer directement des différences structurales entre les deux thiorédoxines (P40A et WT). Un échantillon, préparé comme indiqué précédemment, a donc été soumis à plusieurs séries d'expériences en RMN dans le but d'effectuer l'attribution des spectres et de déterminer des différences structurales. Les spectres RMN à deux dimensions obtenus sont présentés et interprétés dans les paragraphes suivants.

B. Pour l'étude structurale :

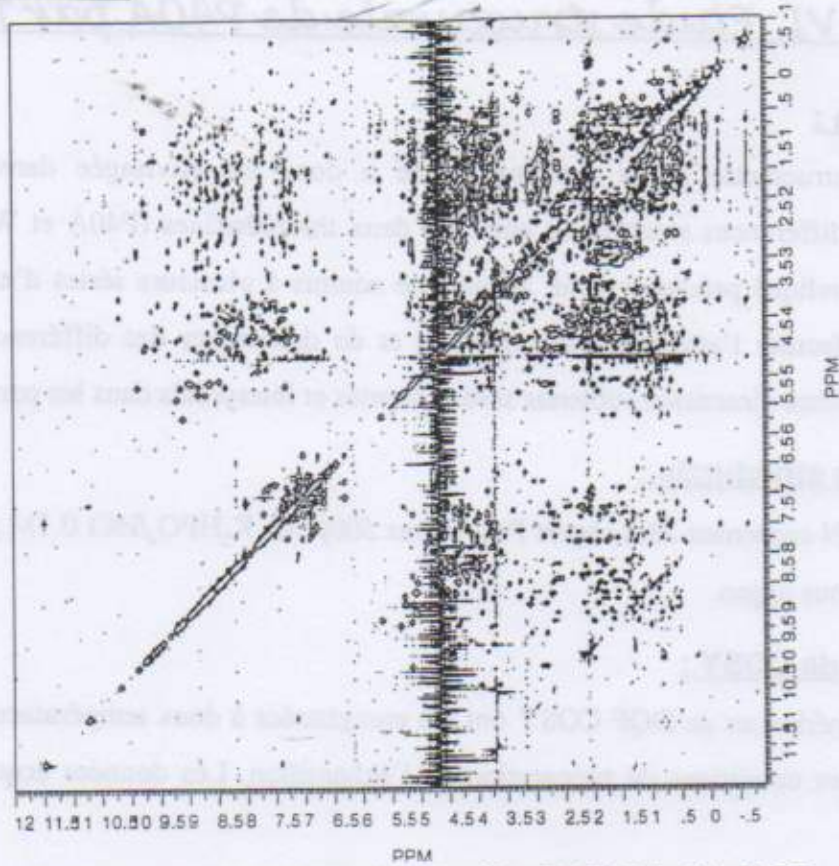
Un tube RMN contenant 23,4 mg de P40A dans 500 μ l de K_2HPO_4/HCl 0.1M pH=5.7, 8% D_2O est préparé est mis sous argon.

C. Expérience de COSY :

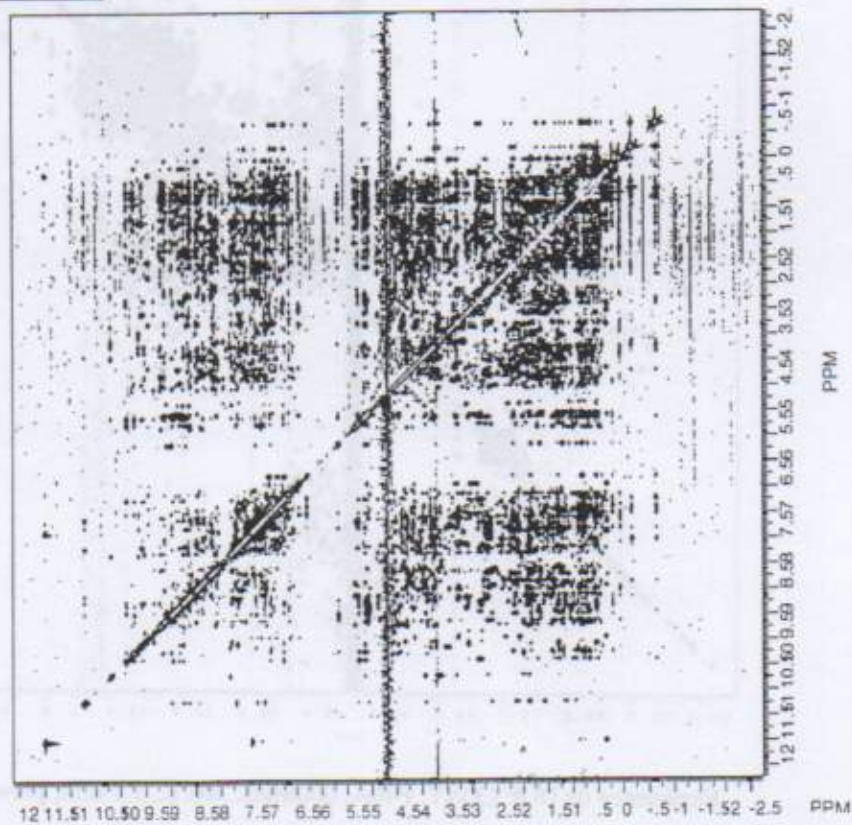
Plusieurs expériences de DQF COSY ont été enregistrées à deux températures différentes (308 K et 330 K) dans les conditions de préparation de l'échantillon. Les données acquises ont été traitées dans GIFA :



Expérience de DQF COSY, P40A, 600MHz, 308 K, 3.7 mM, pH=5.7.

D. Expérience de TOCSY :

Expérience de TOCSY, P40A, 600MHz, 65ms, 308 K. 3.7 mM, pH=5.7

E. Expérience de NOESY :

Expérience de NOESY, P40A, 600MHz, 150ms, 308 K. 3.7 mM, pH=5.7.

F. Attribution :

1. La recherche des systèmes de spins :

De nombreux systèmes de spins ont pu être trouvés facilement, nous disposons en effet de l'attribution de la thiorédoxine non mutée dans les mêmes conditions expérimentales [Dyson et al., 1989] et cela nous a permis de reconnaître de nombreux systèmes de spins qu'il n'était pas toujours aisé de retrouver. Un total de 62 systèmes de spins a pu être identifié dans les expériences de RMN 2D homonucléaires ^1H effectuées.

2. L'attribution séquentielle :

Sur les 108 résidus de la protéine mutée, 41 résidus ont été identifiés. (voir en annexe L'attribution de P40A).

G. La structure de P40A :

Pour le moment, étant donné l'avancement actuel de l'attribution, il est relativement difficile de reconstruire la structure de façon suffisamment précise pour observer des différences avec la structure sauvage de la thiorédoxine d'*E.Coli*. De plus nous sommes confrontés à un problème de superposition des spins rendant délicate l'identification des taches dans les expériences de NOESY. Un marquage isotopique à l'azote ^{15}N serait des plus utiles pour aider à attribuer tous les spins.

Il semble que le nombre de taches NOE observées sur les chaînes latérales des Tryptophanes 28 et 31 soient les mêmes que dans le cas de la thiorédoxine non mutée, elles se retrouvent aux mêmes déplacements chimiques (Figure 45). Cependant, les résultats sont préliminaires à ce niveau et il est encore un peu tôt pour dire que les différences observées dans les simulations au niveau des groupements TRP ne soient pas vérifiées. Si cela se vérifiait, cela signifierait que les différences entre P40A et WT se situent ailleurs dans la structure. Il est donc nécessaire de poursuivre l'étude structurale de P40A pour lever les ambiguïtés.

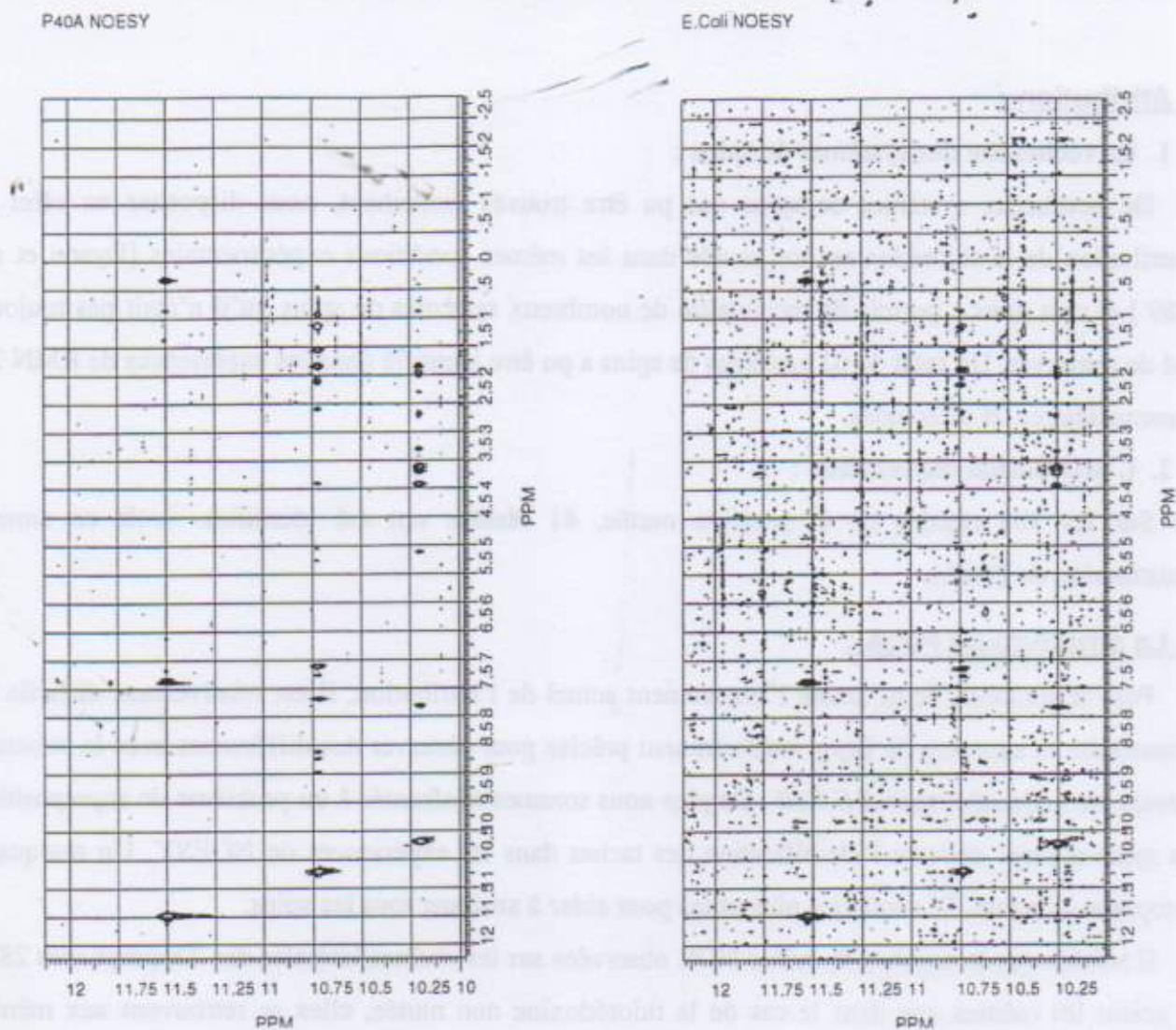


Figure 45 : Représentation des différentes NOE observés sur les HNe des TRP 28 et 31 dans WT et dans P40A.

Chapitre VII. La thiorédoxine h de *Triticum Aestivum*

(trx h) :

A. Présentation des protéines de blé :

Les protéines de réserve qui sont formées et stockées au sein des graines servent de source principale d'azote et de source primaire de carbone pour la germination et la croissance des germes. Les protéines de réserve sont synthétisées après la pollinisation et sont accumulées durant la maturation des graines. Les protéines sont mobilisées, dégradées par protéolyse et utilisées quand les conditions sont favorables à la germination. Dans le cas des céréales, la majorité des protéines de réserve sont des protéines insolubles après maturation des graines. Dans la plupart des autres types de graines, les protéines restent solubles et prisonnières de la membrane créant ainsi une structure connue appelée corps protéique. Dans certaines céréales, la membrane du corps protéique se rompt durant la maturation et la graine se dessèche exposant ainsi les protéines de réserve aux protéases endogènes de l'endosperme. Néanmoins, dans les céréales non panifiables (ne contenant pas de gluten) telles que le maïs, le riz, le sorgho ou le millet, les protéines de réserve restent dans le corps protéique quand les graines sont mures.

Les ponts disulfures sont une des principales forces stabilisant les protéines de réserve. La formation des liaisons disulfures apparaît être une fonction importante pour permettre l'augmentation de la stabilité structurale d'un côté et la diminution de la solubilité de l'autre côté. Ces deux techniques offrent une bonne protection contre la protéolyse. Les protéines de réserves sont constituées de grosses protéines insolubles (60-100 kDa) mais aussi de petites protéines (5-30 kDa) généralement moins abondantes, toutes très riches en ponts disulfures avec 4 cystéines ou plus par molécule.

Différents auteurs [Wong, et al., 1995] ont montré que dans les graines les petites protéines soufrées, aussi bien que les grosses protéines insolubles de réserve sont dégradées durant la germination. Le système NADP-thiorédoxine joue un rôle dans la réduction de groupes disulfures critiques dans les graines ($S-S \rightarrow 2SH$) et, en outre, entraîne la germination.

B. D'où vient la protéine étudiée ? :

elle provient d'une thiorédoxine de blé tendre () elle
 La thiorédoxine de blé étudiée est une protéine de blé tendre (*Triticum Aestivum*). Cette protéine intervient dans les étapes initiales du processus de germination et a été isolée à partir des grains de blé. Par la suite, un clone préparé à l'INRA de Montpellier (Institut de Biochimie et de Biologie Moléculaire des Céréales) a été employé pour la production. Nous avons donc travaillé avec cette protéine recombinante exprimée dans *E.Coli*.

C. L'intérêt de cette étude :

1. Les différences avec les trx connues :

La trx h de *Triticum Aestivum* comporte 127 résidus, c'est une des thiorédoxines comportant la plus longue séquence d'acides aminés connue pour une thiorédoxine. Cette famille de protéines, constituée généralement de 105 à 112 résidus est d'ors et déjà bien étudiée. Ceci nous a donc poussé à en savoir plus sur ces résidus supplémentaires. Nous avons donc commencé par un alignement de séquence pour déterminer les différences séquentielles et les parties communes ou similaires. (voir Figure 28 page 62). Il s'avère que les résidus supplémentaires observés dans la séquence constituent la partie N-terminale de la protéine. Dans le reste de la séquence, il n'y a pas de différences majeures observées. On trouve 37% de similarité entre la trx h de *Triticum Aestivum* et la thiorédoxine h humaine. Sachant que les thiorédoxines montrent 25% d'homologie de séquence en général entre elles (voir page 57), à priori la structure des résidus 20 à 127 de la trx h devrait être similaire aux structures connues de thiorédoxines (5 feuillets β entourés de 4 hélices α). L'intérêt d'une étude structurale sur cette thiorédoxine réside donc en la détermination de sa partie N-terminale, sachant qu'avec les 19 résidus supplémentaires par rapport à la thiorédoxine d'*E.Coli* [MAASAATATATAAAVGAGE] il est possible d'avoir par exemple un feuillet et une hélice supplémentaires, ou 2 feuillets supplémentaires.

D. La fabrication d'un modèle :

1. Présentation :

a) L'idée de ce développement :

Dans un premier temps, nous avons donc essayé de voir si des méthodes de prédiction de structure pouvaient nous fournir des informations sur une hypothétique structure hélicoïdale ou en feuillet.

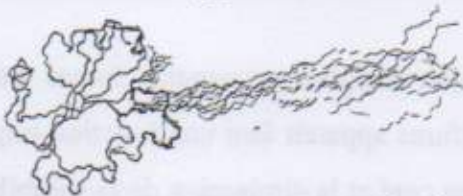
Nous avons également tenté de déterminer une structure modèle de la trx h, à partir des structures de thiorédoxines connues, ce modèle pouvant nous aider par la suite pour voir si la partie N-terminale n'adopte pas une conformation particulière de façon naturelle en faisant une dynamique moléculaire. Egalement, ce modèle pourrait être d'une grande utilité pour l'attribution des expériences de RMN qui seront faites ultérieurement sur cette protéine.

2. La prédiction de structure sur les 1^{ers} résidus :

a) Qu'est-ce que la prédiction de structure :

La prédiction de structure consiste à employer une méthode empirique pour déterminer la structure secondaire probable d'un enchaînement d'acides aminés. Cette méthode est basée sur des statistiques calculées sur les structures connues stockées dans des bases de données telles que la « Protein Data Bank ».

b) Les méthodes employées :



Diverses méthodes ont été utilisées dans le but d'obtenir un résultat pouvant être interprété de façon précise en terme de structure secondaire :

- La méthode de Gibrat [Gibrat, 1987]
- La méthode de Levin [Levin, 1986]
- La méthode DPM [Dayhoff, et al., 1978]
- La méthode SOPM (Self Optimized Prediction Method) [Geourjon and Deleage, 1993].

Toutes ces méthodes de prédictions sont regroupées dans le module SOPM [Geourjon and Deleage, 1994] implanté à l'Institut de Biologie et Chimie des Protéines à Lyon (France).

c) Le consensus :

Le résultat des différentes méthodes employées pour la prédiction de structures est rapporté dans une table qui contient pour chaque acide aminé la prédiction la plus souvent observée (Feuillet ou Hélice). Cette table représente le consensus et représente la prédiction de structure secondaire la plus probable pour la protéine.

d) Résultats :

Numero du résidu :	1	10	20	30	40	50	60	127
<i>Triticum Aestivum</i>	MAASAATATATAAAVGAGEVISVHSLEQWTMQIEEANAAKLVVIDFTASWCGPCRIMAP							
Méthode de Gibrat	HHHHHHHHHHHHHHHHHHHCCCEEEEEENHHHHHHHHHHHHHHHHHHHEEECCCCCCCCCECH							
Méthode de Levin	CCCCHHHHHHHHHHHHHHHHHCCCEEEECSSHHHHHHHHHHHHHHHHHTEEEEECCSCCCSHHHHHH							
Méthode DPM	CCHHHHHHHHHHHHHHHHHHEENHHHHHHHHHHHHHHHHHEENHHHHHHHTCHHHHHH							
Prediction SOPM	EECCSEEEEEECCECCCC EEEEEE EEEEEEHHHHHEEC EEEEEE CCCCCHHHHHH							
Consensus	CCCCHHHHHHHHHHHHHHHCCCEEEENHHHHHHHHHHHHHHHHHEEECCCCCCCCCHHHHH							
Structure vraie prob. EEEEEE CCHHHHHHHHHHHHHHHH EEEEEE CCCCCHHHHHH							
<i>Triticum Aestivum</i>	IFADLAKKFPAAVFLKVDVDELKPIAEQFSVEAMPTFLPMKEGDKORVGAIKEELTTKVLGHAHQ							
Méthode de Gibrat	HHHHHHHHHHHHHHHHHHHCHHHHHHHHHCHHHHHHHCCCCHHHHHEHHHHHHHHHHHHHHHH							
Méthode de Levin	HHHHHHHHSSNHEEECCSNNHHSSNNHHHCECCSHEEENCSCHSSTHHHHHHHHHHHHHHHHHTH							
Méthode DPM	HHHCC							
Prediction SOPM	HH CHHHH CCCC EEEEEE CCCC HHHH CCCC EEEEEE CC EEEEEE EE EEEEEE EE EEEEEE EE							
Consensus	HHHHHHHHSSNNHHHHHHHHHHSSNNHHHCHSSCHHHHHHHCCCCHHHHHHHHHHHHHHHHHH							
Structure vraie prob.	HH HHHHHHHH CC EEEEEE EE EN HHHHHH CC EE EEEE CC CC EEEEEE CC EEEEEE CC EEEEEE CHHHHHHHHHHHHH							

Tableau 5 : Représentation de la prédiction de structure obtenue par la méthode de prédiction SOPM : Dans la prédiction SOPM et le consensus, les hélices sont représentées par H, les feuillets par E et les coudes par C. La 'Structure vraie prob.' sous le consensus de *Triticum Aestivum* représente la structure qu'elle devrait avoir si on se réfère à la structure observée dans la famille des thiorédoxines.

On constate que la méthode SOPM est capable de retrouver la structure secondaire de la famille des thiorédoxines dans la séquence de *Triticum Aestivum*, du moins pour la partie comprise entre les résidus 20 et 106 (voir Tableau 5). Les parties terminales restent assez mal définies. Si on regarde le consensus, il donne de bons résultats dans la première moitié de la séquence mais n'indique la présence d'aucun feuillet dans la deuxième moitié de la séquence.

e) Discussion :

Il semble bien qu'après cette prédiction, la méthode SOPM donne de meilleurs résultats pour *Triticum Aestivum*, en se basant sur l'hypothèse que la structure est commune dans la famille des

thiorédoxines. Cependant les prédictions observées présentent des différences majeures dans les parties N-terminale et C-terminale, et ce sont justement celles qui nous intéressent.

Lorsque le consensus et SOPM donne le même résultat, la structure prédite est plus vraisemblable, mais lorsque ces deux méthodes fournissent des résultats différents, on ne peut pas conclure. Cela se vérifie en s'appuyant la structure commune observée dans la famille des thiorédoxines et représentée par 'Structure vraie prob.' (Tableau 5). Ainsi la structure secondaire de la partie N-terminale qui nous intéresse tant, ne peut pas être prédite comme étant en feuillet ou en hélice.

3. L'homologie de structure :

a) Présentation :

L'homologie de structure consiste à comparer deux structures, tout comme on effectue de l'homologie de séquence. Dans notre cas, nous employons les deux types d'homologie non pas pour comparer deux structures, mais pour reconstruire la structure d'une protéine A (*Triticum Aestivum*), en s'appuyant sur une structure connue (celle de la famille des thiorédoxines) et en imposant cette structure à la séquence de la protéine A. Cela implique de connaître l'alignement de la séquence de A avec la séquence d'une structure connue de la famille.

Les problèmes essentiels pouvant survenir sont des problèmes de délétion et d'insertion entre les deux séquences, il n'est pas évident de supprimer ou d'ajouter un résidu dans une structure tertiaire telle qu'un feuillet béta, il y a alors un décalage ou une modification plus importante dans le feuillet, il convient donc d'être prudent.

b) Les méthodes utilisées :

i) Première méthode (géométrie de distance) :

Dans un premier temps, nous avons déterminé l'alignement des séquences des protéines cibles entre elles. Le programme Clustal V [Higgins et al., 1991] nous a permis de réaliser tous les alignements dont nous avons besoin. Les statistiques sur les résultats de l'homologie (voir plus loin) ont été calculées séparément.

Pour construire la structure de *Triticum Aestivum*, nous avons employé le programme X-plor et l'idée suivante : Partant de la structure connue d'une protéine de la même famille, il ne devrait pas être difficile de replacer, dans un premier le squelette carboné [-N-C α -C-O-] par géométrie de distance. (voir Reconstruction 3D par homologie)

ii) Le programme Modeller :

Modeller est un programme qui modélise la structure 3D de protéines en satisfaisant à certaines contraintes spatiales. Les données fournies au programme sont la structure spatiale d'une séquence d'acides aminés et la séquence d'une structure à modéliser en suivant le repliement de la structure précédente. C'est un outil performant adapté pour l'étude de structures homologues suivant leur

alignement de séquence avec la séquence dont la structure est connue.

Les modèles de structures sont obtenus par optimisation de la fonction de densité de probabilité de la molécule via la méthode du gradient conjugué et une dynamique moléculaire avec un recuit simulé [Sali A., 1994].

c) Résultats de l'homologie de séquence :

i) Alignement entre les thiorédoxines connus :

Résultats :

Les résultats sont présentés sous forme de tableaux (Tableau 6) et permettront de choisir la meilleure structure connue pour son utilisation lors de l'homologie de structure, telle que cela a été expliqué précédemment.

1srx.PDB	-----SDKIIHLTD-DSFDTDLVKADGA---ILVDFWAEWCGPCKMIAPIL
4trx.PDB	MVKQIESKTAFQEALDA-----AGDKLVVDFSATWCGPCKMIKPPF
<i>Triticum Aestivum</i>	MAASAATATATAAAVGAGEVISVHSLEQWTMQIEEANAACKLVVIDFTASWCGPCRIMAPIF
Structure vraie prob.	HHHCEEEEEEECCCCCHHHHHHHHH
alignement obtenu*.....**.....*
1srx.PDB	DEIADEYQGKLTVAKLNIDQNPGTAPKYIERGIPTLLLFKNGEVAATKVGALSKGQLKEFLD-ANLA
4trx.PDB	HSLSEKYSNVIFL-EVDVDDCQDVASECEVKCTPTFQFFKKGQKVGEFSGAN-KEKLEATI--NELV
<i>Triticum Aestivum</i>	ADLAKKFPAAVFL-KVDVDELKPIAEQFSVEAMPTFLFMKEGDVKDRVVGAI-KEELTTKVGLHAAQ
Structure vraie prob.	HHHHHHHHCCEEEEEEEHHHHHHHCCCCEEEEEEEECECCCEEEECCHHHHHHHHHHHHHHH
alignement obtenu*.....**.....*

Tableau 6 : Alignement de la thiorédoxine *Triticum Aestivum* avec la thiorédoxine de *Escherichia Coli* (1srx.PDB) et la thiorédoxine humaine (4trx.PDB), réalisé avec le programme Clustal V. Les « * » représentent les résidus conservés et les « . » les résidus similaires souvent observés lors de mutations simples. (H - hélice ; E - feuillet ; C - coude)

Lorsque l'on effectue un alignement de séquence entre la thiorédoxine humaine, la thiorédoxine d'*E.Coli* et *Triticum Aestivum*, deux à deux, nous obtenons les résultats suivants :

	Δ (Nombre de résidus)	résidus homologues	résidus similaires	%homologie globale	%homologie avec mutations	%homologie non comptés (xx) Acides aminés
4trx trx h	105 127	41	24	32,3%	51,2%	39,0% (22)
4trx 1srx	105 108	36	30	33,3%	61,1%	34,3% (3)
1srx trx h	108 127	30	39	26,6%	54,3%	27,0% (16)

Tableau 7 : Résultats des alignements de séquences dans Clustal V entre la thiorédoxine *Triticum Aestivum* (trx h), la thiorédoxine de *Escherichia Coli* (1srx.PDB) et la thiorédoxine humaine (4trx) prises 2 à 2.

Choix d'une structure homologue :

Compte tenu des structures de thiorédoxines disponibles pour cette étude d'homologie (la trx

humaine et celle d'*E.Coli*), nous n'avons que deux choix possibles pour la structure à utiliser comme point de départ de la reconstruction de la trx de *Triticum Aestivum*. L'ensemble des alignements de séquence effectués nous a permis de choisir la thiorédoxine humaine parce qu'elle est plus semblable à *Triticum Aestivum* que celle d'*E.Coli*. En effet, ce choix s'explique par un nombre plus important de résidus homologues et donc un pourcentage d'homologie plus grand qu'avec la trx d'*E.Coli* (Tableau 7). Il est vrai que la thiorédoxine d'*E.Coli* dispose d'un nombre de résidus similaires un peu plus élevé mais ces résidus devront être mutés par la suite pour reconstruire la structure. Or, il est préférable d'avoir un maximum de résidus homologues pour ne pas avoir à imposer trop de contraintes initiales sur le squelette carboné et les chaînes latérales de la future structure.

ii) Alignement entre la trx h de blé et la trx humaine :

L'alignement de séquence obtenu :

L'alignement de séquence calculé dans le programme Clustal V entre la thiorédoxine humaine (4trx) et la thiorédoxine h de *Triticum Aestivum* (trx h) est le suivant :

4trx.PDB	-----MVKQLESKTAFOEALDAA--GDKLVVVDPSATWCGPCIMIKPPFPHSLSEKYSNV
<i>Triticum Aestivum</i>	CUT-GEVLSVHSLBQWTMQIEEANAAKKLVVDFASWCGPCRIMAPIFADLAKKFPAA
alignement	* . . * . . * . . * . . * . . * . . * . . * . . * . . * . . * . . * . .
4trx.PDB	IFLEVDVDDCQDVASECEVKCTPTFOFFKKGQKVGEFSGANKEKLEATI--NELV
<i>Triticum Aestivum</i>	VFLKVDVDELKPIAEQFSVEAMPTFLFMKEGDVKDRVVGAIKEEL/TKVGLHAAQ
alignement	. * * * * . . * . . * . . * . . * . . * . . * . . * . . * . . * . .

Tableau 8 : Alignement de séquence entre la thiorédoxine humaine et *Triticum Aestivum* réalisé dans Clustal V.

Le problème des résidus N-terminaux :

Pour l'alignement de séquence précédent, les résidus supplémentaires situés dans la partie N-terminale de la trx h n'ont pas été pris en compte. En effet, comme il n'existe aucune information structurale dans la trx d'*E.Coli* ou la trx humaine pouvant être utilisée pour cette partie, nous avons donc décidé de ne pas la prendre en compte pour la reconstruction de structure puisque cela n'aurait fait que ralentir les calculs sans pour autant apporter de nouvelles informations.

Les résultats observés montrent une homologie de séquence relativement plus grande entre la thiorédoxine humaine et la trx h (39,0% - voir Tableau 8), comparée à l'homologie observée généralement dans la famille (25%). La thiorédoxine humaine semble donc être un bon point de départ pour reconstruire une structure de *Triticum Aestivum* sans ses 19 premiers résidus.

4. Reconstruction 3D de la trx h par homologie :

a) Choix des contraintes pour la géométrie de distance :

Nous avons appliqué des contraintes de distances provenant de distances observées dans la structure connue. Toutes les distances observées entre les résidus parfaitement conservés dans la

séquence ont été imposées à plus ou moins 0.5 Å. Quand d'autres résidus sont impliqués, seules les distances observées entre les atomes du squelette carboné [-N-C α -C-O-] ont été imposées à la séquence de la trx h. En cas d'insertion, aucune contrainte n'a été imposée sur les résidus, et les contraintes concernant les résidus voisins de l'insertion ont été supprimées de façon à laisser suffisamment de souplesse.

Dans notre cas nous avons observé deux insertions, une dans l'hélice C-terminale et une autre juste après la première hélice α 1 dans la structure connue de la famille.

b) Insertion dans X-plor3.1 :

Le programme X-plor dispose d'une batterie de modules successifs pour la reconstruction de structure par géométrie de distance (inégalité du triangle) que nous avons utilisés. Nous sommes partis de la structure sous forme étendue de la trx h, cette structure à été construite dans le programme grâce aux modules *generate* puis *generate_template*. Ensuite, les contraintes déterminées comme indiqué dans le paragraphe précédent ont été stockées dans les fichiers de contraintes d'X-plor (*.tbl) et ont été intégrées dans les calculs. L'ensemble des distances appliquées a été soumis aux algorithmes de géométrie de distance disponible dans X-plor (*dg_sub_embed* et *dg_full_embed*) pour en améliorer les valeurs et se rapprocher d'un jeu de distances possibles pour les résidus de la trx h.

i) La mise en place du « backbone » N, C α , C, O :

Dans un premier temps seules les contraintes concernant les squelette carboné (« backbone ») ont été considérées, tout résidus confondus, exceptés ceux situés dans l'insertion et leurs voisins pour lesquels il n'y a aucune contrainte fixée. C'est le module *dg_sub_embed* de X-plor qui prend cette étape en compte.

ii) Ajout des contraintes sur les chaînes latérales :

Dans cette étape de calcul, le module de géométrie de distance *dg_full_embed* prend en charge l'application de toutes les contraintes imposées. La reconstruction repart des structures obtenues lors de la mise en place du squelette carboné et est donc progressive. C'est seulement à ce moment que les contraintes concernant les chaînes latérales des résidus homologues sont prises en compte.

Elle permet l'obtention d'une série de structures sur lesquelles nous pouvons contrôler le nombre de « bonnes » et de « mauvaises » structures, c'est à dire en fait le nombre et le type de violations constatées sur les contraintes de distances imposées.

c) Résultats des étapes de dynamique sous contraintes & minimisation :

i) Emploi de X-plor3.1 :

Le module *dg_sa.inp* (« Distance Geometry Simulated Annealing ») est un module de dynamique moléculaire sous contrainte des distances fournies dans les fichiers *.tbl (voir précédemment). En sortie de ce module, nous obtenons une série de structures qui sont ensuite améliorées par un autre

recuit simulé plus doux (module *refine.inp*) avant d'être rejetées ou acceptées suivant le nombre de violations de contraintes de distance, et les valeurs obtenues pour les longueurs de liaisons ou les angles de valences (module *accept.inp*). Aucune structure n'a été reconnue correcte. La structure la plus acceptable est de mauvaise qualité. (Figure 46 et Figure 47).

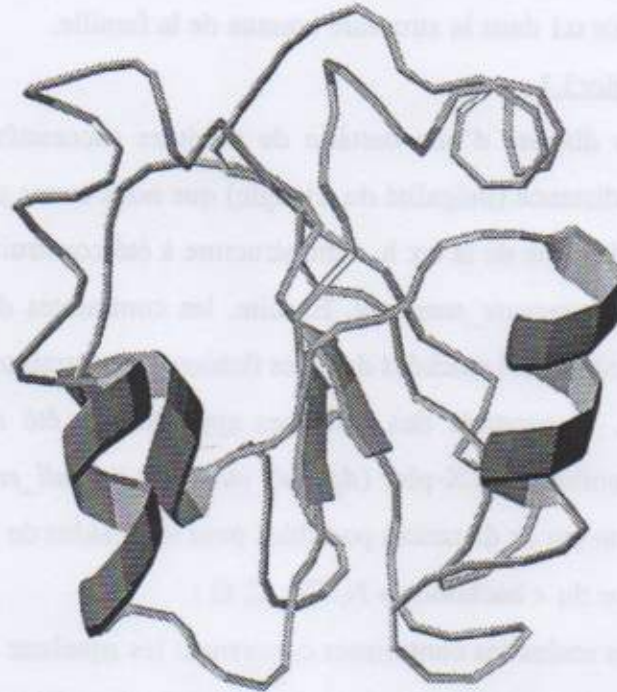


Figure 46 : Structure tertiaire obtenue dans X-plor

La figure de Ramachandran obtenue pour la structure moyenne, calculée dans molmol, indique que la structure obtenue n'est pas des plus réjouissantes. En effet, de nombreux résidus possèdent des angles (ϕ, ψ) avec des valeurs inhabituelles dans le diagramme (Figure 47).

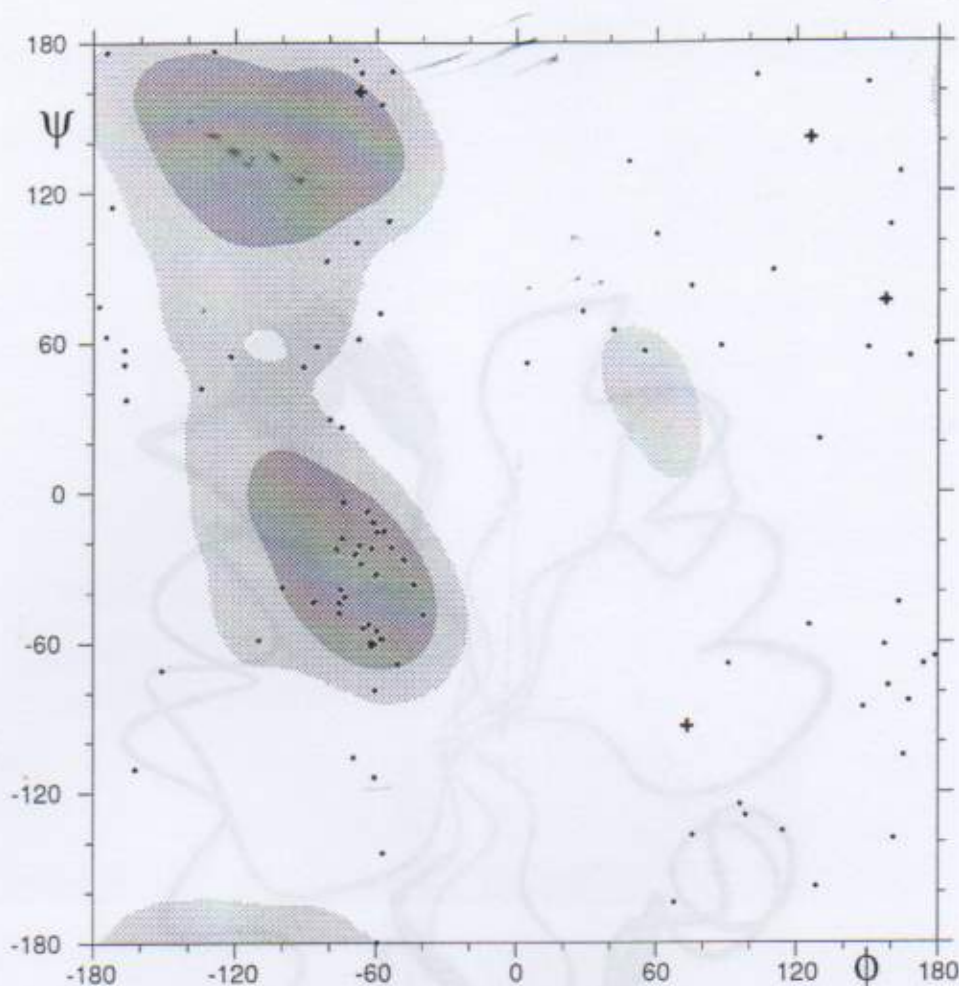


Figure 47 : Figure de ramachandran obtenue pour la 'meilleure' structure obtenue dans X-plor.

ii) Emploi du programme Modeler :

Comme précédemment, nous avons utilisé la thiorédoxine humaine comme structure de base pour la reconstruction de notre modèle. La commande ALIGN3D, employée pour l'homologie de structure, utilise un alignement de séquence courant, obtenu grâce à la commande ALIGN, comme point de départ des différentes itérations successives pour superposer les deux structures 3D. La reconstruction se fait d'abord en superposant les résidus équivalents obtenus dans l'alignement de séquence et contenant la liste des atomes à utiliser pour cette étape (FIT_ATOMS = <atomes définis>). Ensuite, une matrice de distance résidu-résidu est obtenue en calculant les distances Euclidiennes entre toutes les paires d'atomes sélectionnées dans les deux structures. L'alignement des structures est calculé par optimisation de cette matrice durant une étape de simulation de dynamique moléculaire. Une pénalité énergétique étant appliquée lorsque 2 atomes équivalents sélectionnés dans les 2 résidus alignés sont à plus d'une certaine distance (3.5 Å) l'un de l'autre. Une pénalité nulle signifie que les résidus sont équivalents. La procédure de dynamique employant cette pénalité s'arrête lorsqu'il n'y a plus de changements dans le nombre de résidus équivalents. Les résultats obtenus sont les suivants (Figure

48) :

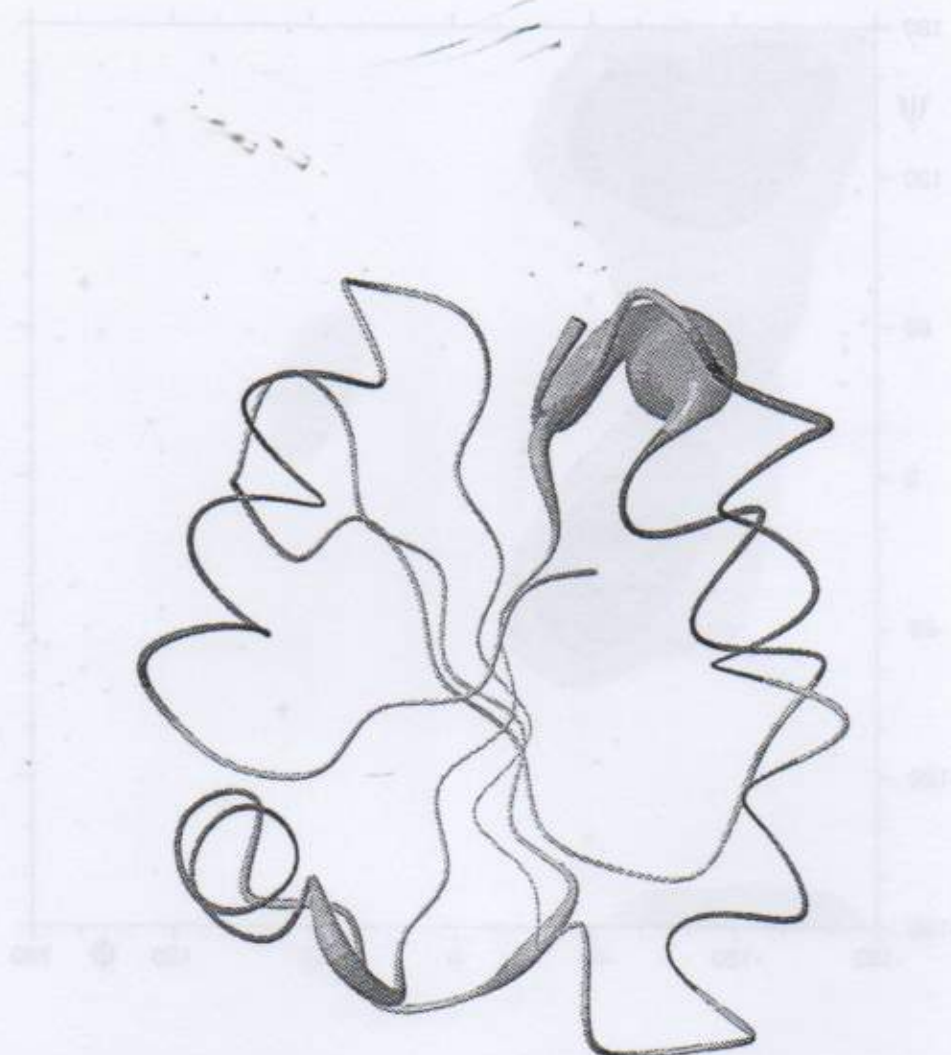
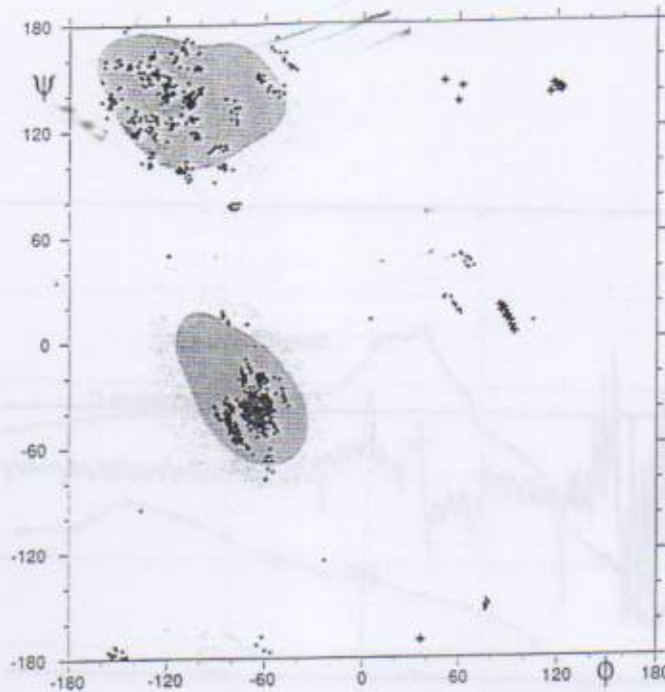


Figure 48 : Structure moyenne de la thiorédoxine h de *Triticum Aestivum* avec les domaines d'erreur.



Ramachandran Plot obtenu pour 10 structures calculées et optimisées dans Modeller.

5. Conclusion

Nous disposons désormais d'un modèle de thiorédoxine h avec certaines réserves à considérer en particulier au niveau des insertions séquentielles, car ces zones restent assez mal définies. Ce modèle n'apporte pas vraiment de nouveauté structurale mais néanmoins, il pourra être utile lors de l'étape d'attribution des expériences de RMN, notamment des taches NOESY. La structure reconstruite par géométrie de distance dans X-plor est de qualité très inférieure à celle obtenue dans Modeller. Nous conservons donc la structure simulée dans Modeller pour la suite de notre étude structurale.

E. Préparation de la trx h :

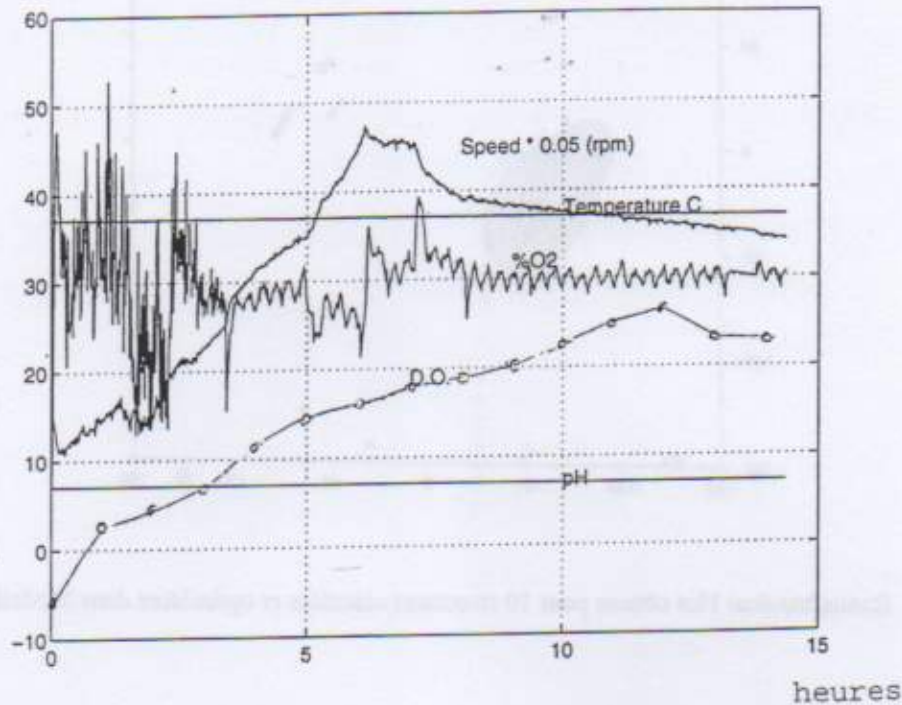
1. Préparation et purification de *Triticum Aestivum* :

a) Culture :

Elle est produite de la façon suivante. La première étape est une phase de production, un plasmide contenant l'ADN complémentaire de la trx h de *Triticum Aestivum* permettant une expression dans *E.Coli* (pET-trxTa) est introduit dans une bactérie (BL21 DE3 pLysS ou E). Cette bactérie produit, en présence d'IPTG*, une ARN polymérase particulière, la T7 ARN polymérase, qui ne reconnaît que le promoteur de pET en amont de l'ADNc. Ainsi cette polymérase ne transcrit pas tous les gènes mais seulement celui porté par pET. On a ainsi une production violente qui démarre lorsqu'on ajoute de

* isopropyl-2-D-thiogalactopyranoside, un analogue non métabolisable du galactose. Il induit les promoteurs sensibles au

l'IPTG.



Evolution de la D.O. du milieu, du pH, de la température et du %O₂ au cours de l'expression de *trx h*.

b) Les étapes de purification :

i) Le protocole initial :

Les toutes premières préparations de *trx h* se faisaient suivant un premier protocole qui avait fait ses preuves pour de petites quantités de protéine, mais qui, semblait-il, n'était pas reproductible de façon suffisamment satisfaisante à la longue pour obtenir les quelques dizaines de milligrammes nécessaires pour une étude structurale par RMN. Ce protocole est le suivant :

Après l'étape de culture cellulaire, on récolte par centrifugation les cellules puis on les fait éclater à la presse de French. Après centrifugation, le surnageant est chauffé 5 min à 60°C. Après une nouvelle centrifugation, le surnageant est additionné de sulfate d'ammonium qui insolubilise les protéines sans les dénaturer (on en ajoute pour arriver à 40% de saturation en sulfate d'ammonium en solution). Après centrifugation, le surnageant obtenu est additionné de sulfate d'ammonium pour atteindre 60% de saturation en sulfate d'ammonium en solution. C'est ce qui précipite entre 40% et 60% de saturation en sulfate d'ammonium qui nous intéresse. Après centrifugation, le culot est repris dans 10ml de TRIS 30mM, pH=8, EDTA 1mM, NaCl 50mM et EtOH 10% puis chargé sur une galactose sans pour autant intervenir dans le métabolisme.

B. Kucapko

colonne Séphadex G50 (5cm x 100cm) équilibré dans ce même tampon. Après filtration, l'activité de la thiorédoxine est mesurée par un test d'activation de la MDH à NADP⁺ (voir publication sur P40A en annexe). Les fractions actives sont concentrées et dialysées contre du TRIS 10mM, pH=8, EDTA 1mM puis chargées sur une colonne Q sépharose Fast Flow, rincées puis éluées par un gradient de NaCl 0-200 mM. La trx h sort à 120mM de NaCl. En solution. Néanmoins après plusieurs mois d'essais, cette méthode ne s'est pas avérée capable de produire la protéine pure. Nous avons essayé d'aller plus loin dans cette méthode de purification en ajoutant une étape de séparation en HPLC sur une colonne d'échange d'ion Q8HR mais il n'a pas été possible d'obtenir d'échantillon valable pour une étude par RMN. Nous avons donc envisagé de reprendre complètement le protocole de purification, notamment au niveau de la purification à basse pression avec les colonnes de chromatographie d'échange d'ions et d'exclusion.

{ DEAE Sepharose
HPLC, ITC

ii) La méthode de purification remise au point :

Cette méthode de purification nous a pris 5 mois, et a nécessité 6 cultures bactériennes, avant d'obtenir un échantillon convenable pour des expériences de RMN, avec une concentration de environ 1mM ce qui relativement peu. La méthode est décrite pour la sixième culture bactérienne qui nous a permis d'enregistrer les expériences de RMN avec le meilleur rapport signal/bruit.

La casse des cellules :

La culture provenant d'une souche BL21 DE3 pLysE (22g de cellules) a été reprise dans le tampon de casse de cellules (BB) suivant :

- NaCl 100mM (4ml NaCl 5M dans $V_{total}=200ml$)
- EDTA 5mM pH=8 (1ml EDTA 1M pH=8 pris dans $V_{total}=200ml$)
- DTT 5mM (0.1543g DTT dans $V_{total}=200ml$)
- TRIS 50mM pH=8 (10ml de TRIS 1M pH=8 dans $V_{total}=200ml$)
- 5mM benzamidine (0.1566g dans $V_{total}=200ml$)
- 100mg de lysozyme dans $V_{total}=200ml$ complété avec de l'eau mQ.

→ réduction TRX → moins stable
(T=proté)

L'EDTA sert contre les éventuelles métalloprotéases, le lysozyme sert à démanteler les parois cellulaires et la benzamidine à couper en morceaux les ARN et ADN libérés en solution, le TRIS à tamponner le pH et le DTT à réduire les liaisons covalentes S-S susceptibles de se former et de provoquer la précipitation de certaines macromolécules.

Il nous a fallu 70ml de BB pour les 22g de cellules. Celles-ci une fois reprises en solution ont été agitées et broyées avec un *mixeur* durant 30min dans un bêcher métallique plongé dans un bain de glace à 0°C. A la suite de ce mixage, une étape de *sonication* aux ultrasons a été utilisée pour compléter la destruction des parois cellulaires : 2 fois [2 min à 50%, 2s de pulse, 2s stop sous agitation

à 0°C].

La centrifugation :

Les débris cellulaires sont précipités par centrifugation à 4°C sur un rotor de 112mm à 10730 rpm durant 45min. Nous obtenons donc un culot $C_{\text{débris}}$ et un surnageant $S_{\text{débris}}$.

Le choc thermique :

Cette étape s'explique par le fait que les thiorédoxines sont des protéines thermiquement stables. Le surnageant $S_{\text{débris}}$, qui contient la trx h, est soumis à un *choc thermique* à 60°C durant 3 min. Un peu d'eau (2 cm d'eau) est chauffée à 62°C dans un cristalliseur de diamètre 20 cm. Le chauffage est arrêté et un erlen contenant $S_{\text{débris}}$ est mis en contact avec le bain d'eau durant 3 min, temps après lequel la température chute à environ 59-60°C, puis il est aussitôt plongé dans un bain de glace pendant 5 min. Nous procédons ensuite à une centrifugation à 40000g (45 min) pour obtenir un culot C_{Δ} et un surnageant S_{Δ} .

Les étapes de précipitations au sulfate d'ammonium :

Le surnageant S_{Δ} , qui contient la trx h, est soumis à deux étapes de précipitation au sulfate d'ammonium, une première durant laquelle la quantité de $(\text{NH}_4)_2\text{SO}_4$ en solution est progressivement augmentée à 30%. La quantité de sulfate à ajouter, qui est fonction du volume de la solution et de la température, est déterminée dans les tables de Dixon (voir Tableau 9).

don extraction thermique en présence de DTI ?

Table 8. Ammonium Sulfate Usage at 0°C.

		Final Percent Ammonium Sulfate Solution at 0 C																
		20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
0	106	134	164	194	226	258	291	326	361	398	436	476	516	559	603	650	697	
5	79	108	137	166	197	229	262	296	331	368	405	444	484	526	570	615	662	
10	53	81	109	139	169	200	223	266	301	337	374	412	452	493	536	581	627	
15	26	54	82	111	141	172	204	237	271	306	343	381	420	460	503	547	592	
20		27	55	83	113	143	175	207	241	276	312	349	387	427	469	512	557	
25			27	56	84	115	146	179	211	245	280	317	355	395	436	478	522	
30				28	57	86	117	148	181	214	249	285	323	362	402	445	488	
35					28	57	87	118	151	184	218	254	291	329	369	410	453	
40						29	58	89	120	153	187	222	258	296	335	376	418	
45							29	59	90	123	156	190	226	263	302	342	383	
50								30	60	92	125	159	194	230	268	308	348	
55									30	61	93	127	161	197	235	273	313	
60										31	62	95	129	164	201	239	279	
65											31	63	97	132	168	205	244	
70												32	65	99	134	171	209	
75													32	66	101	137	174	
80														33	67	103	139	
85															34	68	105	
90																34	70	
95																	35	

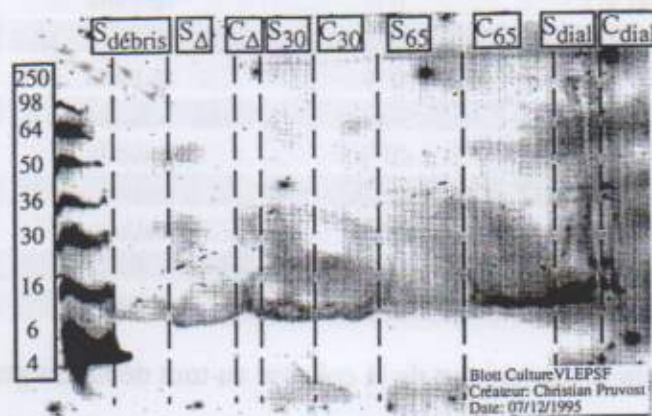
This table provides the amount (in grams) of ammonium sulfate to add to a solution to adjust it to a new percentage. To determine the correct amount of ammonium sulfate to add, find the row for initial percentage of ammonium sulfate in the solution. Then move right along that row to the column containing the desired percentage. **Note:** A saturated ammonium sulfate solution is 4.1M (at 25°C) and requires 767g per liter. Ammonium sulfate solubility decreases by approximately 2.44g/°C. (England, S. and Scifter, S. (1990) Meth. Enzymol. 182, 285.)

Tableau 9 : Nomogramme de Dixon à 0°C

Nous prenons bien soin d'ajouter lentement le sulfate d'ammonium cristallisé pour éviter de provoquer des chocs acides ou des excès locaux de sulfate dans la solution et de précipiter ainsi des protéines qui à priori devraient rester bien solubles dans cette solution à 30% de sulfate d'ammonium. Une fois le sulfate dissous, la solution obtenue est mise sous agitation à 0°C durant 15 min. Une étape de centrifugation à 40000g (45 min) est effectuée ensuite pour séparer un culot C_{30} de molécules qui précipitent, d'un surnageant S_{30} .

Une deuxième étape de précipitation est effectuée de manière similaire, avec cette fois 65% de sulfate d'ammonium en solution. Nous récupérons alors, après centrifugation à 40000g (45 min), un culot C_{65} contenant la trx h recherchée et un surnageant S_{65} .

Ces 2 étapes de précipitations permettent d'éliminer un grand nombre de molécules organiques mais doivent être réalisées avec prudence. En effet, lors des essais, la première précipitation avait lieu avec une coupe à 30%, puis une à 40% de sulfate, étape dans laquelle nous perdions un peu de thiorédoxine h ; cette coupe a été éliminée pour limiter les pertes protéiques, l'inconvénient étant bien sûr de ne pas avoir une séparation aussi performante qu'à 40% mais nous avons l'avantage d'une



Gel/W-Blott 1 : Western Blott obtenu dans les différentes étapes initiales de purification.

La trx h a une masse moléculaire voisine de 13.5 kDa, elle doit donc apparaître entre les marqueurs de masse moléculaire 6 et 16. C'est bien ce qu'on observe. Néanmoins, il a toujours été difficile de détecter la protéine sur gel, car semble-t-il elle fixe très mal le bleu de coomassie. Il nous a fallu, de façon systématique dans toutes nos analyses, transférer sur une feuille de cellulose les diverses bandes obtenues après migration sur le gel, et effectuer une révélation avec des anticorps spécifiques de la trx h.. C'est l'analyse par Western Blott obtenue ci-dessus. On constate que le culot C_{65} une fois repris dans un minimum de BBII est la fraction qui contient le plus de trx h. Cependant, nous perdons un peu de trx h dans C_{30} .

La chromatographie d'échange d'ions :

Un support d'échange d'ions est préparé dans une colonne C16/20 de chez pharmacia. Le support employé est de la Q Sépharose Fast Flow (QSFF). Ce support est lavé et préparé comme indiqué par le fournisseur puis mis en solution dans la solution A.

Solution A :

- β -mercapto éthanol 14mM
- EDTA 5mM
- Tris pH=8 50mM

Solution B :

solution A + NaCl 1M

Le culot surnageant $S_{dialyse}$ est chargé en une seule fois sur la colonne. Nous rinçons alors la colonne avec $2 V_0$ de solution A (V_0 est le volume de la colonne) puis un gradient de concentration en sel NaCl est programmé jusqu'à 1M en NaCl.

Volume des fractions	pourcentage de solution A	pourcentage de solution B	Débit de la solution	volume échantillonné
9 ml	100 %	0 %	1ml/min	0 ml
9 ml	100 %	0 %	1ml/min	119 ml
4 ml	100 %	0 %	1ml/min	120 ml
4 ml	60 %	40 %	1ml/min	319 ml
5 ml	60 %	40 %	1ml/min	320 ml
5 ml	0 %	100 %	1ml/min	420 ml
5 ml	0 %	100 %	1ml/min	430 ml
5 ml	100 %	0 %	1ml/min	440 ml
5 ml	100 %	0 %	0.1ml/min	450 ml

La thiorédoxine h se détache du support de la colonne au tout début du gradient (vers 50mM NaCl - Figure 49). La détection en sortie de colonne se fait grâce aux absorptions dans l'UV à $\lambda=280\text{nm}$ des tryptophanes de la protéine.

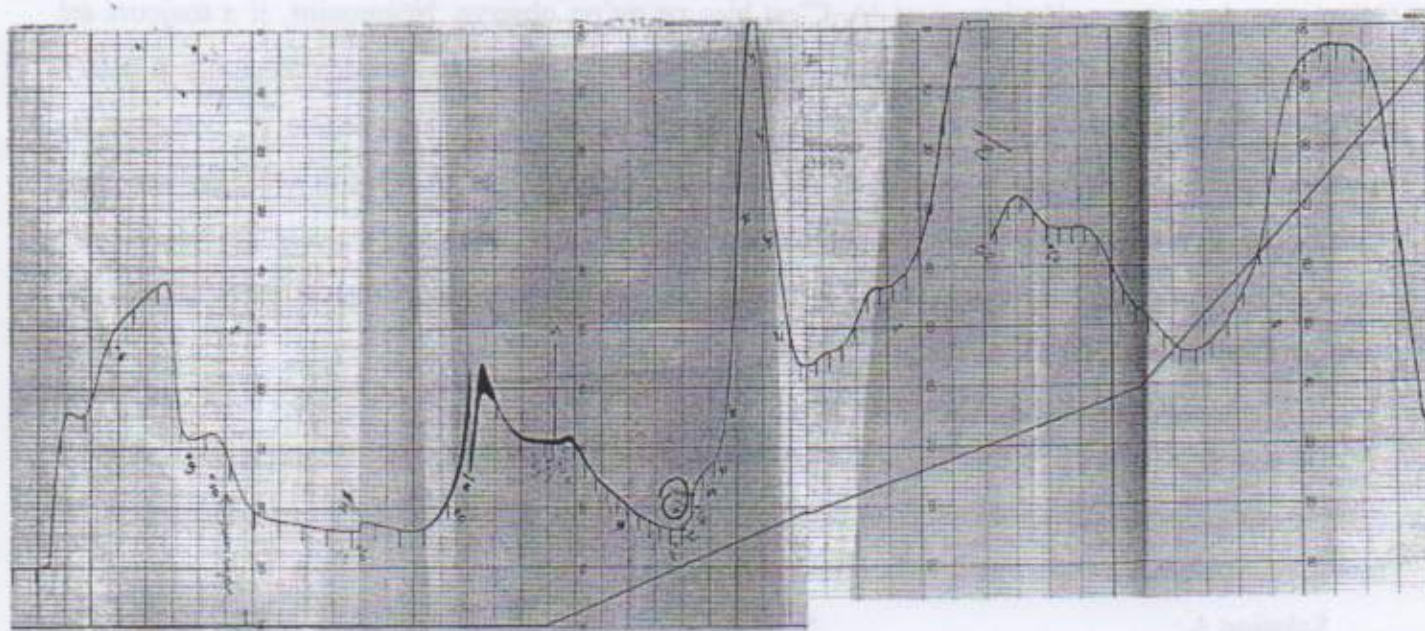
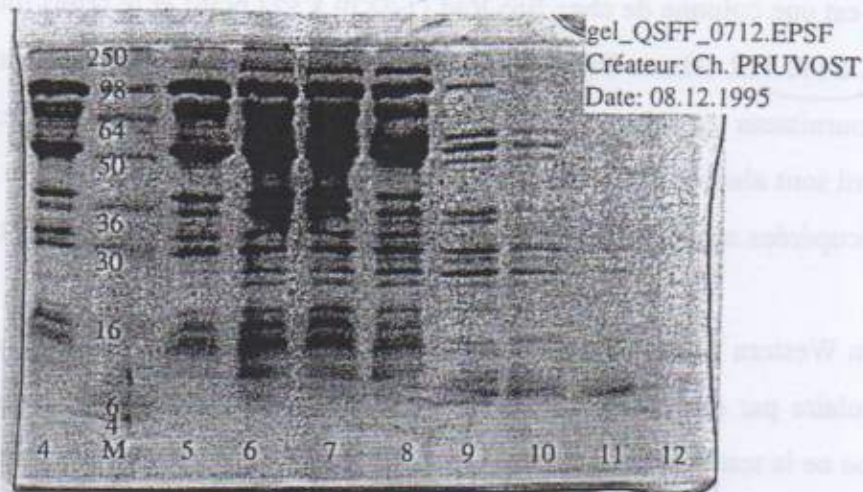
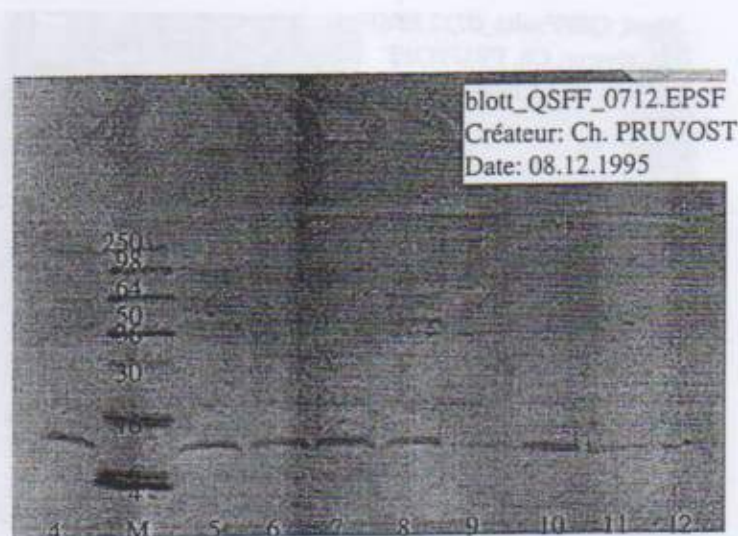


Figure 49 : Courbe détectée à 280 nm en sortie de la colonne QSFF.

Les fractions des échantillons 4 à 12 et 21 à 64 sont déposées sur gel de polyacrylamide et mises à migrer par électrophorèse. Nous effectuons ensuite un transfert des gels obtenus sur membrane de cellulose pour faire apparaître la trx h par révélation aux anticorps.



Gel/W-Blott 2 : Gels SDS Page obtenus pour la sortie de la QSFF



Gel/W-Blott 3 : Western Blotts obtenus pour la sortie de la QSFF.

La détection aux anticorps nous indique la présence de protéines qui interagissent dans les fractions 4 à 12, ainsi que dans les fractions 20 à 30 (voir **Figure 50**).

La chromatographie d'exclusion :

Les fractions 3 à 12 (81 ml) sont finalement concentrées sur centriprep par élimination à travers une membrane YM3 de toutes les molécules de tailles inférieures à 3kDa, entre autre le solvant. Le volume obtenu n'était pas assez faible et la reconcentration très lente, aussi les protéines ont toutes été précipitées au sulfate d'ammonium à 80% en solution puis reprise dans 2.5 ml de tris 50mM ph=8. Celui-ci est chargé sur une colonne de chromatographie d'exclusion pour une étape de séparation complémentaire. En effet, les fractions 3 à 12 contiennent de nombreuses macromolécules de masse moléculaire supérieure à 16 kDa (voir Gel/W-Blott 1).

La colonne est une colonne de chez Bio-Rad (120cm x Ø2.6cm) et le tamis moléculaire utilisé est de la Sépharose S100 HR (Haute Résolution) de chez pharmacia. La colonne est préparée selon les consignes du fournisseur et mise en équilibre dans du Tris 50mM pH=8. Les fractions 3 à 12 concentrées à 2ml sont alors chargées puis éluées à un débit de 1.1 ml/min avec du Tris 50mM pH=8.. Les fractions récupérées en sortie de colonne sont de 9ml. La détection se fait toujours grâce à un détecteur UV.

Un gel et un Western Blott des fractions obtenues est effectué. C'est la zone où, après calibration du tamis moléculaire par des molécules de taille connue, est censée apparaître la trx h de *Triticum Aestivum* mais on ne la retrouve dans aucune des fractions (les gels ne sont pas présentés). Elle se situe en fait dans les fractions 20 à 30.

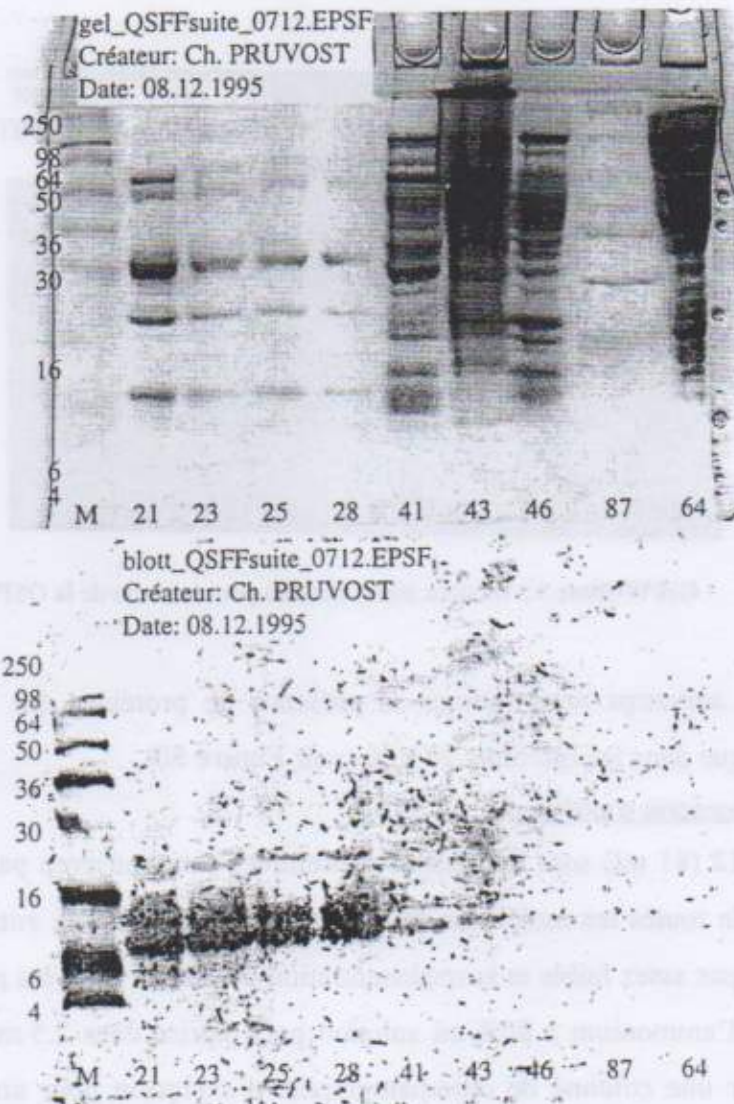


Figure 50 : Gel et Blott des fractions 20 et suivantes en sortie de QSFF

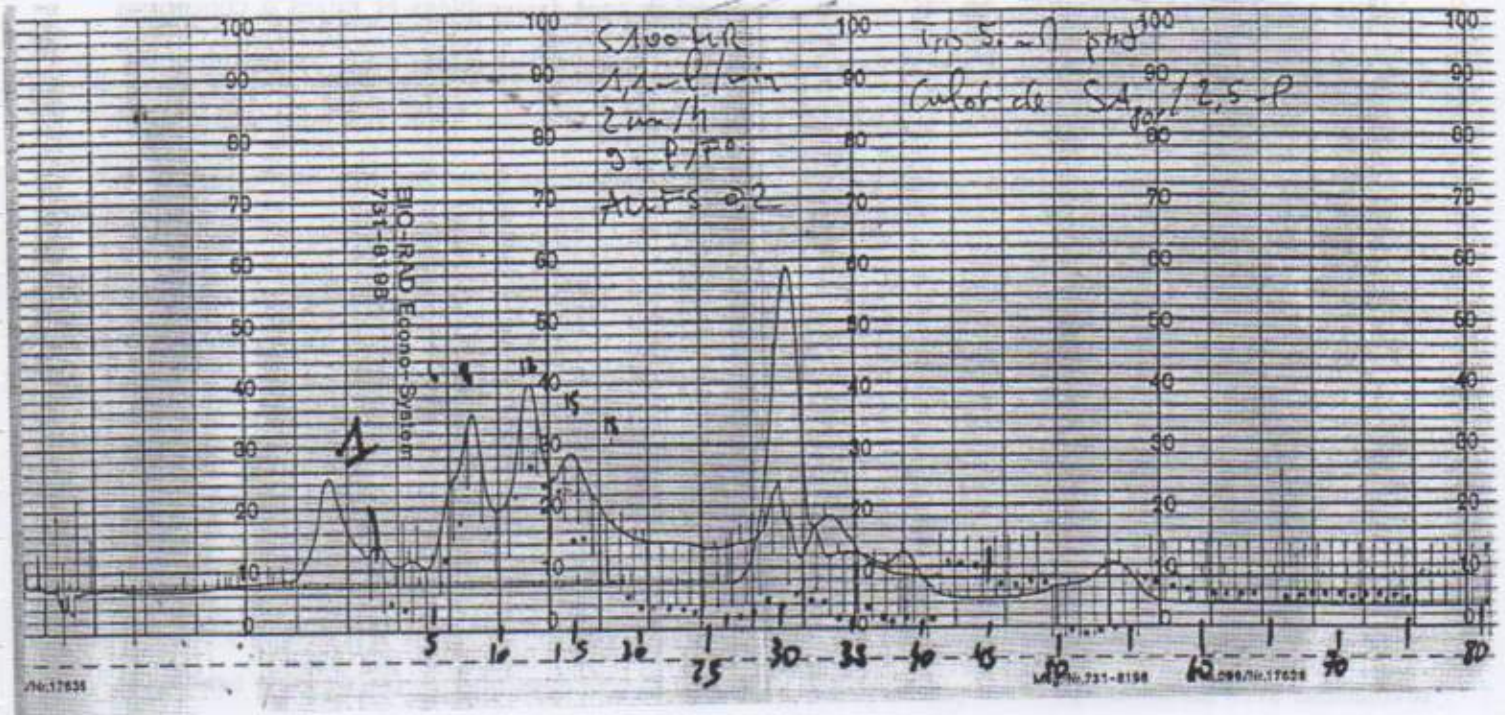


Figure 51 : Allure de la courbe obtenue en sortie de filtration sur gel des fractions 20-30.

Après une analyse sur gel SDS page 15%, il s'avère que la trx h se trouve dans les fraction 14 à 18 et relativement pure (voir Figure 52).

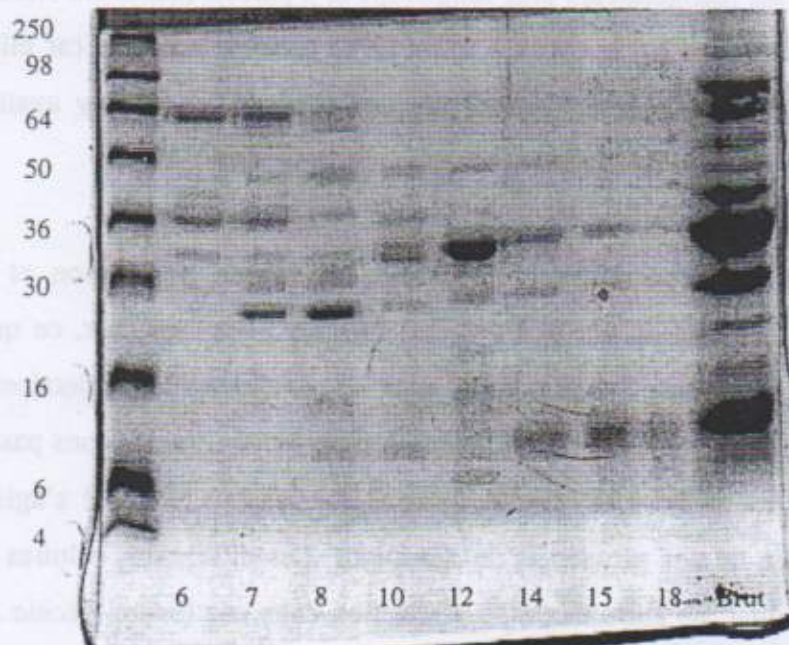


Figure 52 : Gel des fractions contenant la trx h en sortie de la S100HR

L'obtention de la protéine :

Les fractions 14 à 18, contenant la protéine recherchée sont rassemblées et mises à concentrer d'abord sur centriprep (membrane YM3) puis sur centricon jusqu'à l'obtention d'un volume de 1 ml. Nous avons alors ensuite modifié le tampon de la solution en ajoutant 1 ml de Phosphate 20 mM $\text{pH}=6.4$ puis en reconcentrant de nouveau à 1 ml, ceci 10 fois de suite pour éliminer au maximum le TRIS présent dans la solution.

c) Les problèmes majeurs rencontrés :

pb majeur : production
ou purification ?

i) La détection de la protéine :

Etant donné le faible niveau de concentration protéique en *trx h*, nous avons dû effectuer systématiquement une analyse par Western Blott après chaque électrophorèse sur gel. La détection de la *trx h* a donc été effectuée grâce à des anticorps spécifiques. Cette détection s'est avérée délicate car la *trx h* ne se colore pour ainsi dire pas au bleu de coomassie sur les gels (électrophorèse). Les révélations avec les anticorps n'ont pas toujours permis de savoir dans quelles fractions se trouvait la protéine en sortie des colonnes de séparation. Ainsi il est arrivé que l'on choisisse de mauvaises fractions parce que nous laissons les Western Blott se colorer un peu trop longtemps, ceux-ci faisant alors apparaître d'autres protéines qui ne correspondaient malheureusement plus à la nôtre. Il s'est avéré plus tard que les anticorps dont nous disposions interagissaient également plus ou moins avec la thiorédoxine d'*E. Coli*, créant bien sûr des troubles au niveau des analyses. De cette façon, la protéine a pu se dégrader dans les réfrigérateurs plus longtemps et a parfois même été entièrement perdue. Un autre problème de détection a été la quantification de la protéine produite car elle était invisible ou presque sur les gels, et lorsque la coloration aux anticorps se faisait, il y avait une réaction trop violente pour donner d'autres détails que la présence ou non de la *trx h*.

ii) La reproductibilité des expériences :

7 Nous avons rencontré de nombreux problèmes au niveau préparation et purification de la protéine. En effet, il semble bien que la *trx h* soit difficile à sur-exprimer, ce qui est un problème majeur. Plusieurs tentatives ont été effectuées pour en améliorer la production mais sans succès évident jusqu'à ce jour. De plus, dès le début de la purification nous ne savions pas si la *trx h* était bel et bien présente dans les cellules en grande quantité ou pas, ou bien s'il s'agissait d'une autre macromolécule (à cause de nos problèmes de détection). Les différentes cultures n'ayant pas fourni deux fois le même résultat, un autre mode de production dans une levure, *Pichia Pastoris* [Cregg, et al., 1993], est à l'étude dans le but d'améliorer les rendements de production et leur reproductibilité.

d) Les tests négatifs réalisés :

i) L'emploi d'une colonne spécifique :

Nous avons tenter d'améliorer et de purifier au mieux les fractions impures en employant une colonne spécifique pour fixer le site actif de protéine -S-S- comme les thiorédoxines. (colonne de chez de la société INVITROGEN), comportant de l'arsenic (Figure 53).

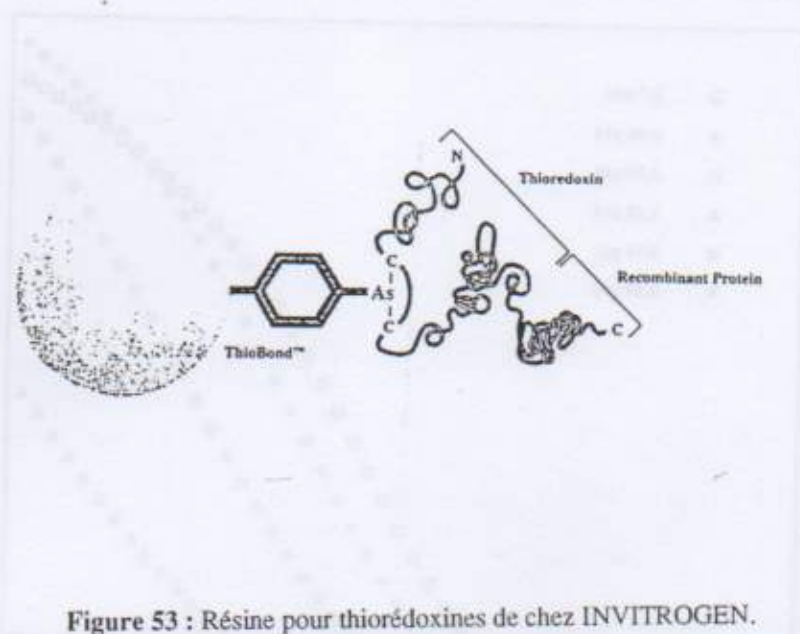


Figure 53 : Résine pour thiorédoxines de chez INVITROGEN.

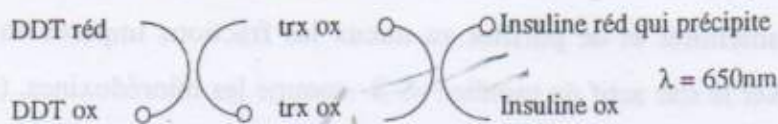
Les thiorédoxines sont censées s'accrocher dans cette colonne spécifique pour pouvoir être séparées des autres protéines. Dans notre cas, la trx h ne s'est jamais accrochée, elle est toujours sortie dès les premiers lavages de la colonne, le protocole d'activation de la résine ayant pourtant été suivi à la lettre.

ii) L'ajout d'une étape de HPLC :

Nous avons également essayé une colonne d'échange d'ions en HPLC (colonne Q8HR de chez Pharmacia) mais sans réel succès. La plupart du temps, la protéine était en trop faible quantité pour que l'on puisse en faire quoique ce soit en RMN. En effet, à chaque étape de purification, le rendement n'est pas de 100% et les pertes s'accumulent.

2. Tests d'activité biologique :

L'activité de la protéine obtenue a été testé in vitro en présence de DTT sur la réduction de l'insuline. La décroissance de l'absorption à $\lambda=650\text{nm}$ de l'insuline oxydée est mesurée en fonction du temps pour différentes concentrations en trx h. La même activité est retrouvée aussi pour la trx h isolée que pour la trx h native mélangée avec les autres protéines du blé. (Figure 54).



DTT reduction of Insulin mediated by TRX h expressed in *E. coli*.

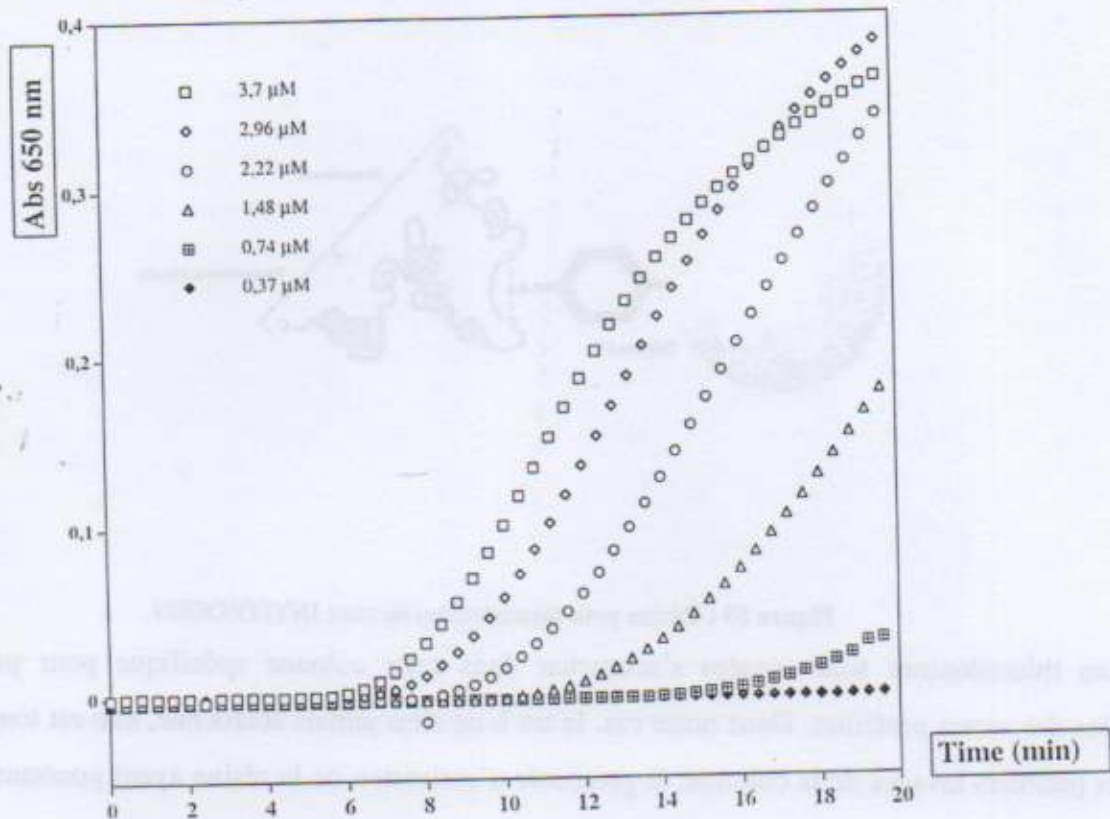


Figure 54

3. Analyse par spectrométrie de masse :

Une analyse par spectrométrie de masse a finalement pu être réalisée sur la protéine purifiée dans le but de vérifier sa masse moléculaire et éventuellement la présence d'impuretés de masses différentes. L'échantillon a été préparé en éliminant toute traces de sels (de phosphate, de sodium ou de potassium) qui peuvent perturber le détecteur de l'appareil. Pour ce faire, une précipitation à l'acide trichloro-acétique 20% (Cl_3COOH ou TCA) a été effectuée. Le précipité a été repris en solution (100 μl) puis analysé par spectrométrie de masse.

Un pic de masse moléculaire voisin de 12 kDa a été détecté.

a) Discussion :

La masse moléculaire de la trx h est normalement de 13.5 kDa. Nous obtenons une masse inférieure ce qui indique que soit la protéine étudiée n'est pas la bonne, soit elle a subi une protéolyse dans une des étapes de préparation qui précèdent (culture ou purification). Suite aux résultats obtenus

en RMN (voir plus loin), nous pouvons affirmer que la protéine étudiée est bien une thiorédoxine dont une partie de la séquence correspond à la trx h escomptée. Ainsi la spectrométrie de masse, associée à la RMN, nous indique la perte d'une partie de la chaîne d'acides aminés en partie N-terminale ou C-terminale. Des investigations plus approfondies ont été effectuées par les biologistes s'occupant de la préparation du fragment d'ADN intégré dans la bactérie. Il s'est avéré qu'un site de coupure protéolytique existe dans *E.Coli* pour le fragment d'ADN intégré au sein de la bactérie et codant pour la trx h de *Triticum Aestivum*. Ce site correspond à une coupure dans la partie N-Terminale au niveau des trois Alanines. De plus la masse détectée correspond elle-aussi à une disparition des 13 premiers acides aminés en partie N-Terminale.

4. conclusion :

Nous avons obtenu une méthode de purification qui semble fonctionner pour la thiorédoxine h, néanmoins la protéine obtenue n'est pas complète, il manque la partie N-Terminale. Les problèmes rencontrés s'expliquent dans la mesure où nous avons une faible expression de la protéine. Ceci rend donc très difficile la tâche de la préparation de l'échantillon pour une expérience de RMN où la concentration protéique doit être de 1mM au moins. La fragilité du peptide N-Terminale peut également s'expliquer par le fait qu'il ne serait pas vraiment structuré, la prédiction de structures secondaires ne donne en effet pas de résultats clairs à ce sujet. Ainsi ces problèmes de sous-expression de la protéine, de non structuration probable de la partie N-Terminale, de détection qui laisse aux protéases d'*E.Coli* le temps d'agir sur les faibles quantités de protéine produite rendent cette étude très délicate. Il serait peut-être judicieux d'employer un autre moyen de détection pour repérer la thiorédoxine dans les fractions lors des étapes de purification. Et il faut obtenir une meilleure production pour avancer dans la recherche structurale de la trx h de *Triticum Aestivum* d'autant plus qu'il sera vraisemblablement nécessaire de marquer la protéine à l'azote ^{15}N pour une étude poussée en RMN, étant donné le nombre important de résidus dans cette protéine.

F. Etude par résonance magnétique nucléaire (RMN) :

1. Introduction :

Les différentes étapes classiques de l'étude structurale de protéines par RMN¹¹ ont été démarrées dès l'obtention d'un échantillon. Cet échantillon a été préparé pour une étude en solution dans l'eau puis une étude en solution dans D_2O .

2. Partie expérimentale :

a) Matériels :

¹¹ Une étude par diffraction des rayons X a également été envisagée mais l'obtention de cristaux de taille suffisante n'a pas été possible, par ailleurs la difficulté de production nous a obligé à choisir une méthode d'étude plutôt qu'une autre.

Nous avons employé un spectromètre RMN AMX 600MHz Bruker pour effectuer toutes les expériences permettant d'entamer l'attribution de la trx h de *Triticum Aestivum*. Le pH choisi pour les expériences est de 6.4, nous ne pouvons pas descendre beaucoup plus bas car le point isoélectrique de la trx h est voisin de 5.5, nous risquerions alors de voir précipiter nos échantillons.

b) Travail dans D₂O (eau lourde) :

L'avantage des expériences de RMN dans D₂O vient simplement du fait qu'il n'y a plus, ou pratiquement plus d'eau dans la solution étudiée. Ainsi, tous les signaux qui ont le même déplacement chimique que celui de l'eau ne sont plus masqués et apparaissent clairement. Un autre avantage de l'eau lourde est lié à l'échange de certains protons (les hydrogènes amides entre autre) avec le solvant, ces protons disparaissent alors du spectre RMN qui est ainsi considérablement clarifié. Cela permet de détecter plus facilement certains systèmes de spins tels que les noyaux aromatiques ou certains résidus fortement impliqués dans la structure via leur hydrogène amide qui alors ne s'échange pas ou très lentement avec l'eau.

i) COSY :

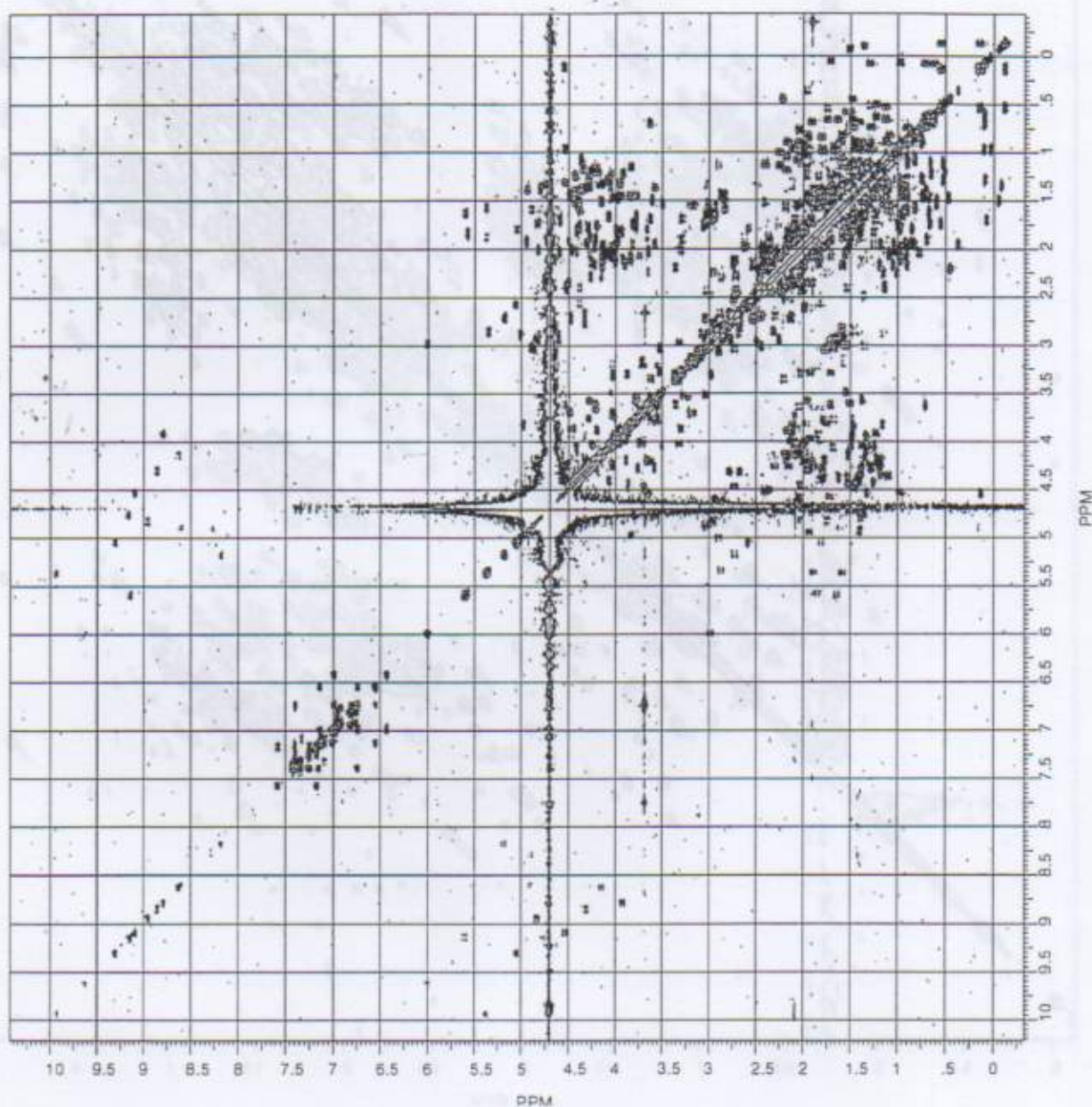


Figure 55 : COSY, trx h, 600MHz, pH 6.4, cc~0.7mM, 295K, D₂O

On remarque que certains résidus n'échangent pas leur hydrogènes amides avec le solvant, ceci a été vérifié sur une durée de plus de 6 mois. Ces résidus sont donc vraisemblablement enfouis dans le cœur hydrophobe de la protéine. (Figure 55 et Figure 56).

ii) TOCSY :

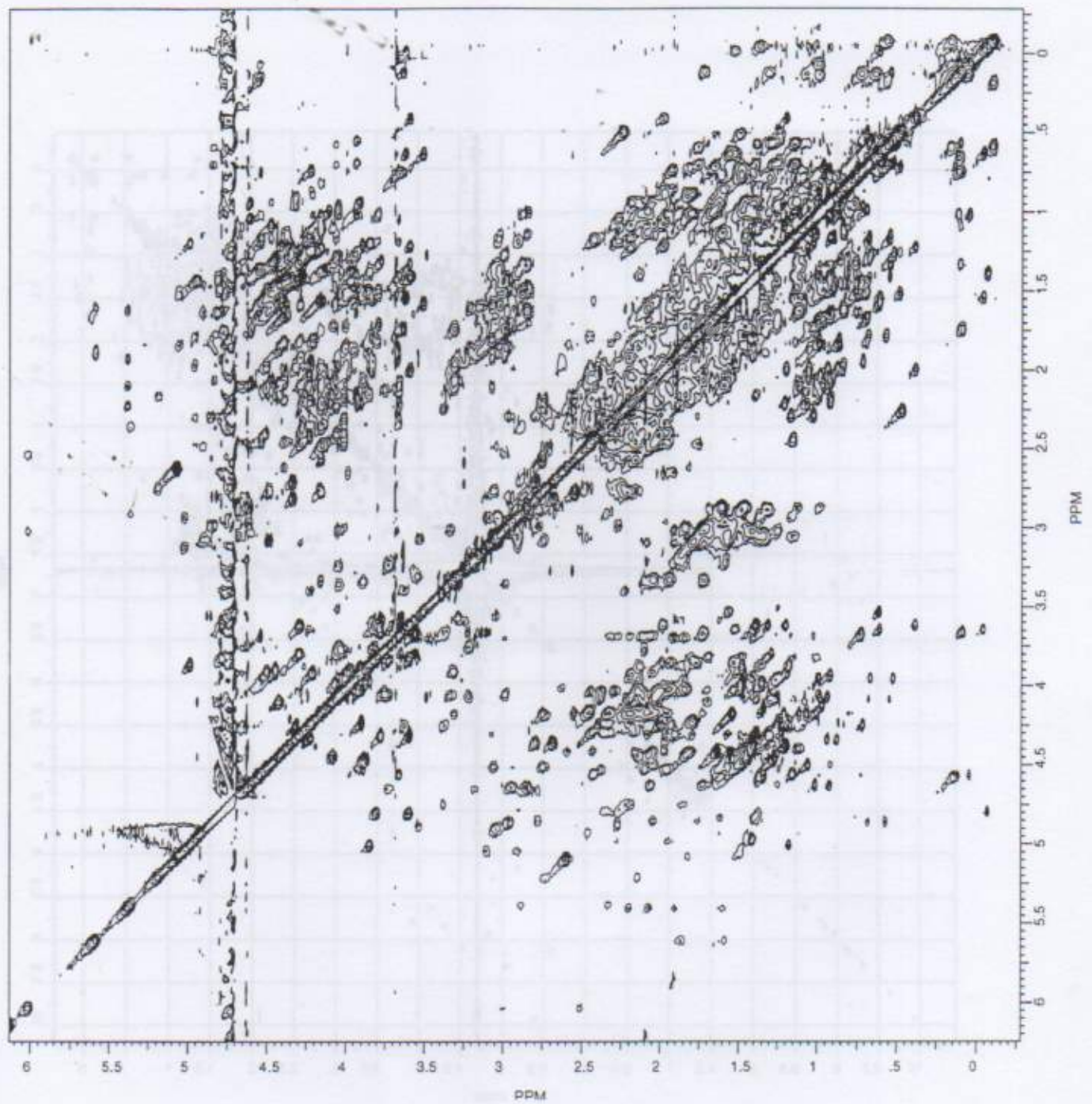


Figure 56 : TOCSY, trx h, 600MHz, pH 6.4, cc=0.7mM, 295K, 65ms, D₂O.

c) Travail dans H₂O :

i) TOCSY :

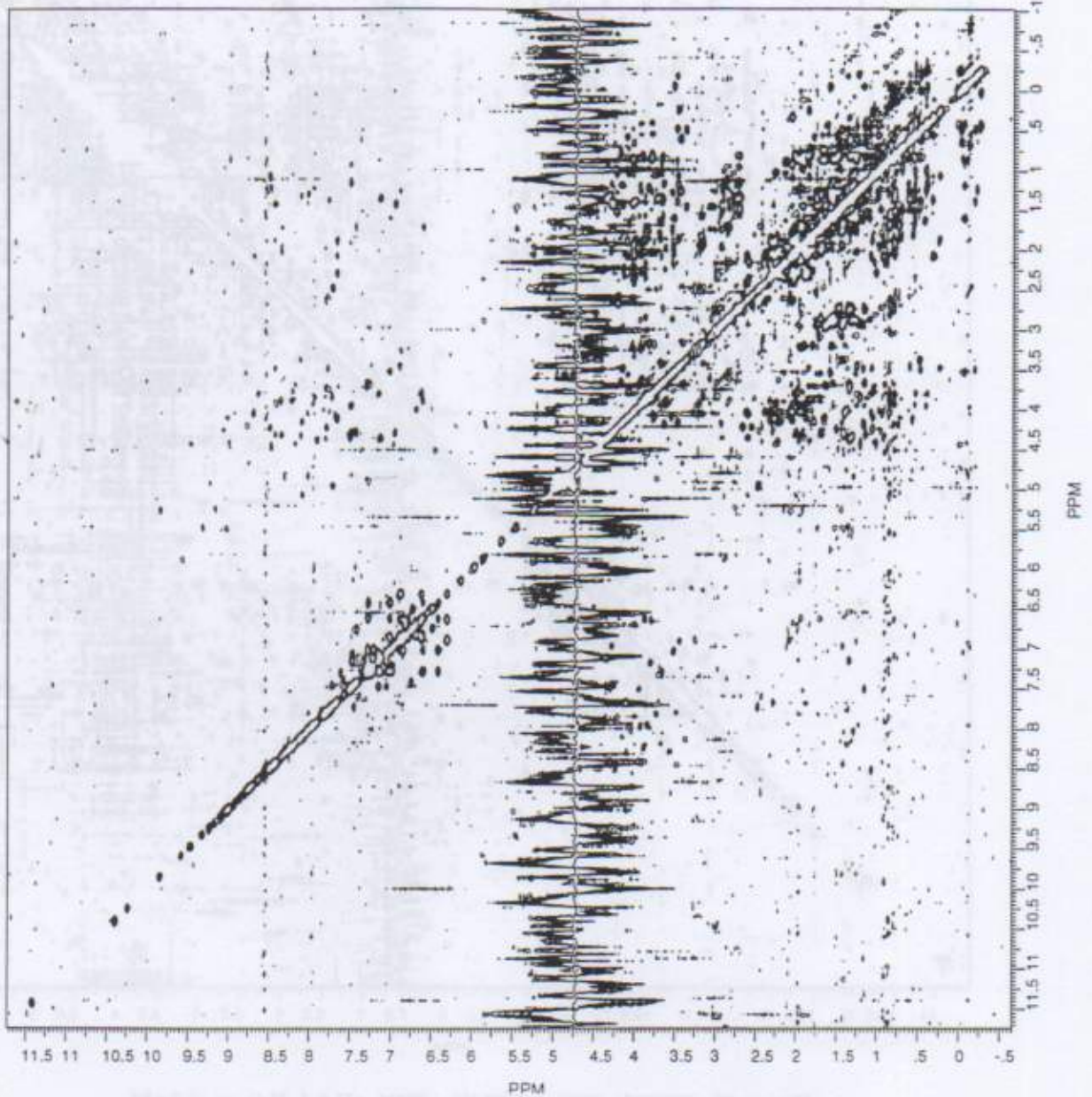


Figure 57 : TOCSY, 65ms, trx h, 600MHz, pH 6.4, cc~0.7mM, 295K, H₂O

On peut constater (Figure 57) la pauvreté du signal par rapport aux spectres obtenus page 94

ii) NOESY :

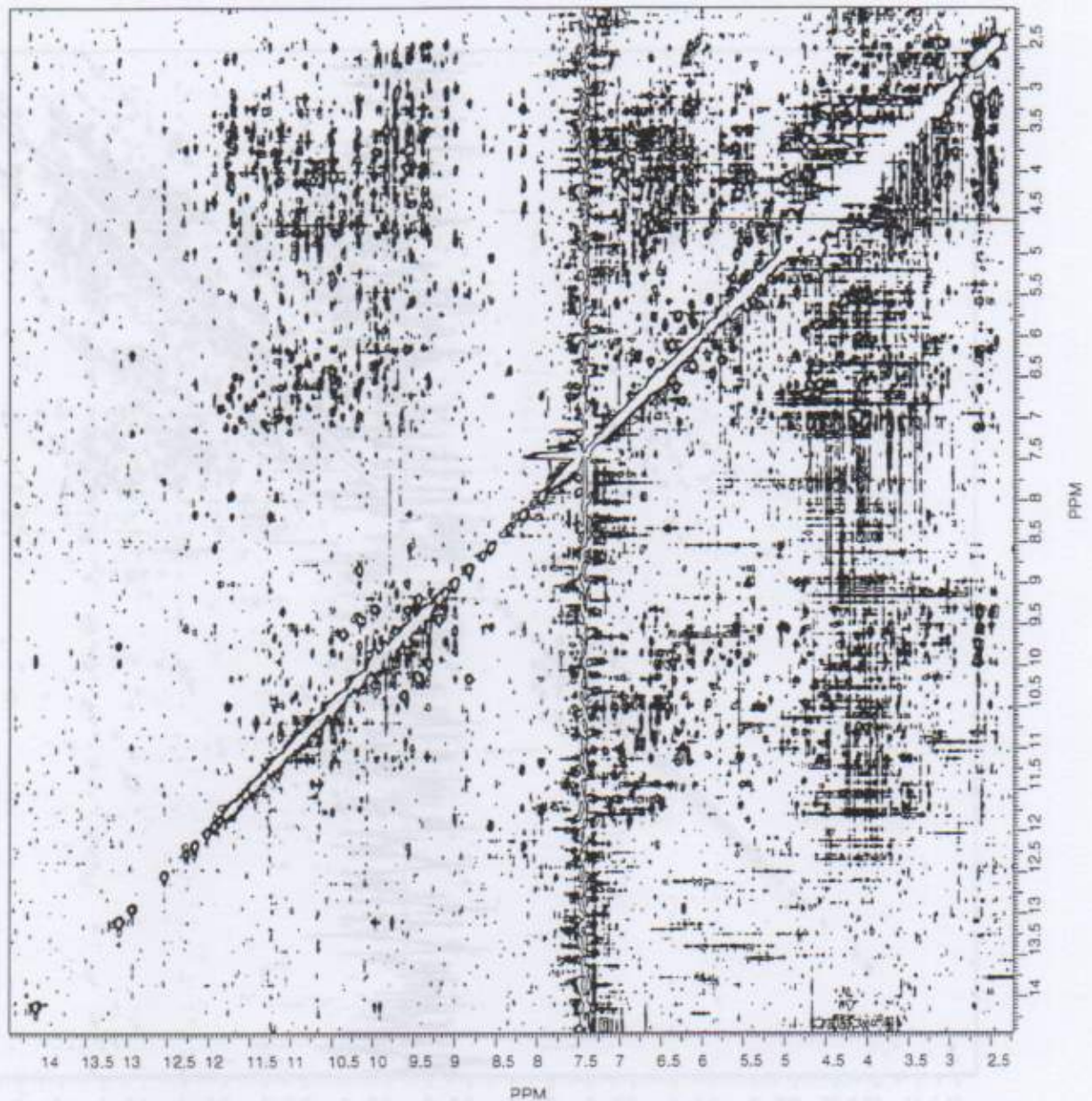
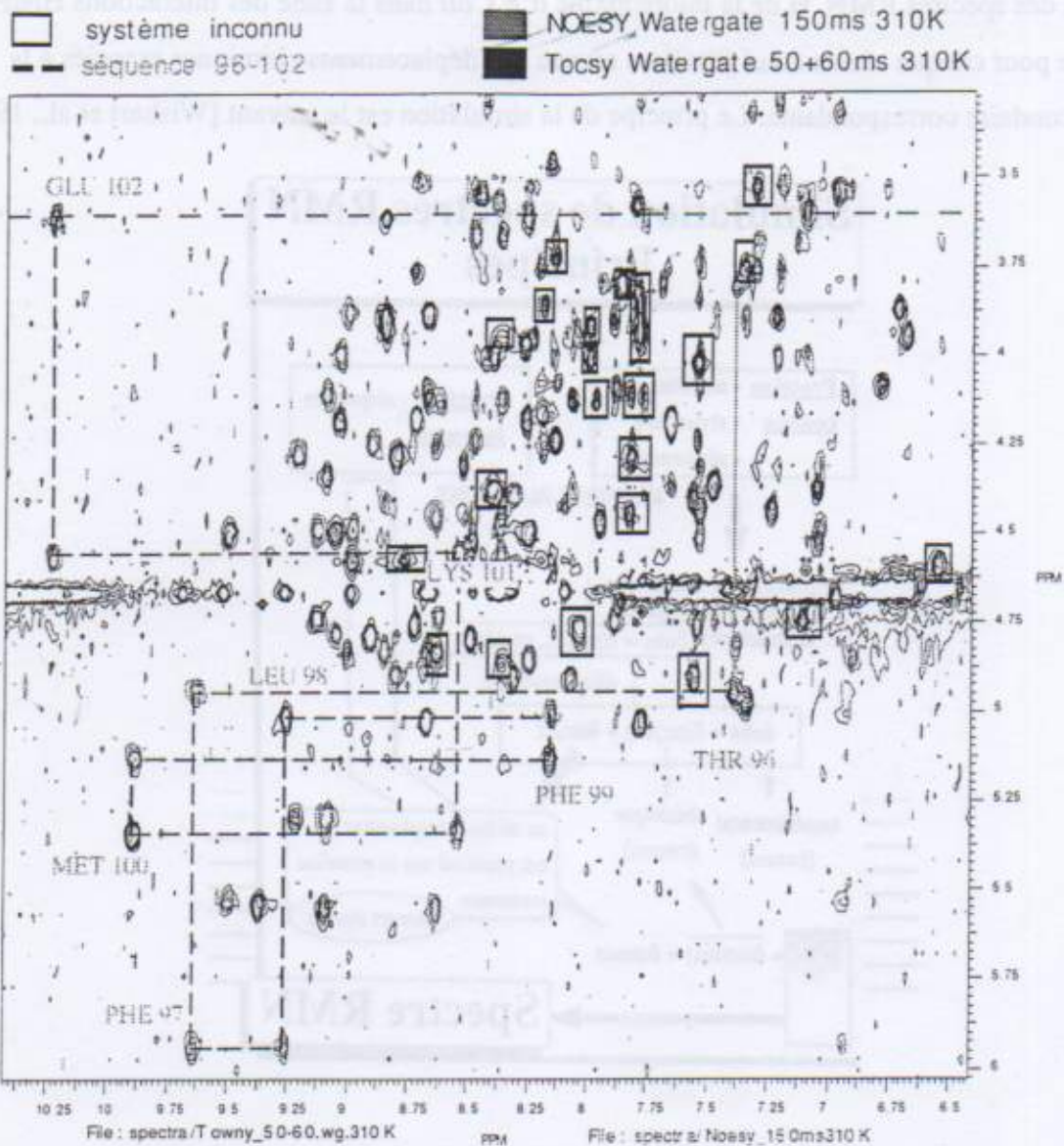


Figure 58 : NOESY, 120ms, 600MHz, 295K, pH 6.4, H₂O, cc=0.7mM.

d) Résultats :

i) Attribution obtenue :

Nous avons réussi à repérer l'enchaînement des résidus 96 à 102 (numérotation dans la séquence complète de trx h *Triticum Aestivum*). Au total, seulement une cinquantaine de systèmes de spins ont pu être relevés.



(voir en annexe l'attribution en cours de *Triticum Aestivum*)

ii) La reconstruction de structure :

La reconstruction de la structure est pour le moment impossible étant donné l'avancement des travaux. Il quantité de signal dans les spectres RMN est bien trop pauvre pour détecter toutes les corrélations, il faut donc des échantillons plus concentrés en trx h.

3. Simulation des spectres de RMN de *Triticum Aestivum* :

Dans le but d'attribuer un peu plus facilement les spectres obtenus et surtout de vérifier les attributions trouvées, nous avons simulé une partie intéressante du spectre RMN : la zone des corrélations HN/H α .

a) Méthode :

Une analyse des spectres RMN ^1H de la thiorédoxine d'*E.Coli* dans la zone des interactions HN/H α a été effectuée pour chaque résidu (sauf PRO) au niveau des déplacements chimiques associés à la structure secondaire correspondante. Le principe de la simulation est le suivant [Wishart et al., 1991] :

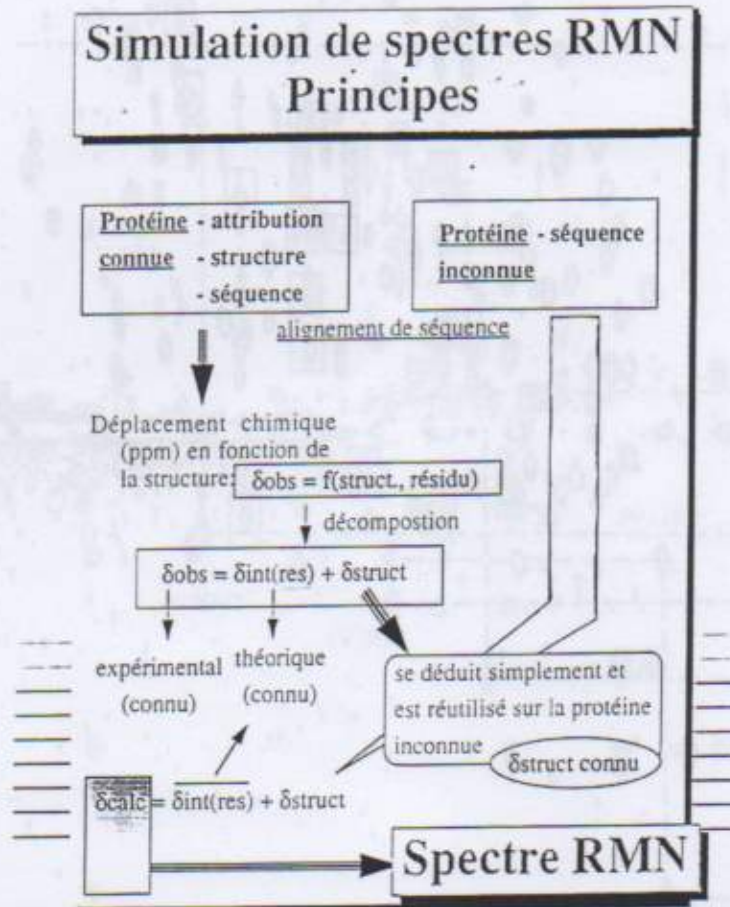


Figure 59 : Principes de la simulation de spectres RMN appliquée

Nous transformons le déplacement chimique observé expérimentalement en la somme d'un déplacement chimique intrinsèque au résidu considéré et d'un déplacement chimique dû à la structure secondaire dans laquelle il est impliqué. Cette valeur du déplacement chimique dû à la structure secondaire est appliquée au résidu censé occuper la même position dans l'autre protéine, associé à la valeur du déplacement chimique intrinsèque de ce nouveau résidu. (voir Figure 59).

b) Résultats de la simulation du spectre RMN dans la région HN/HA :

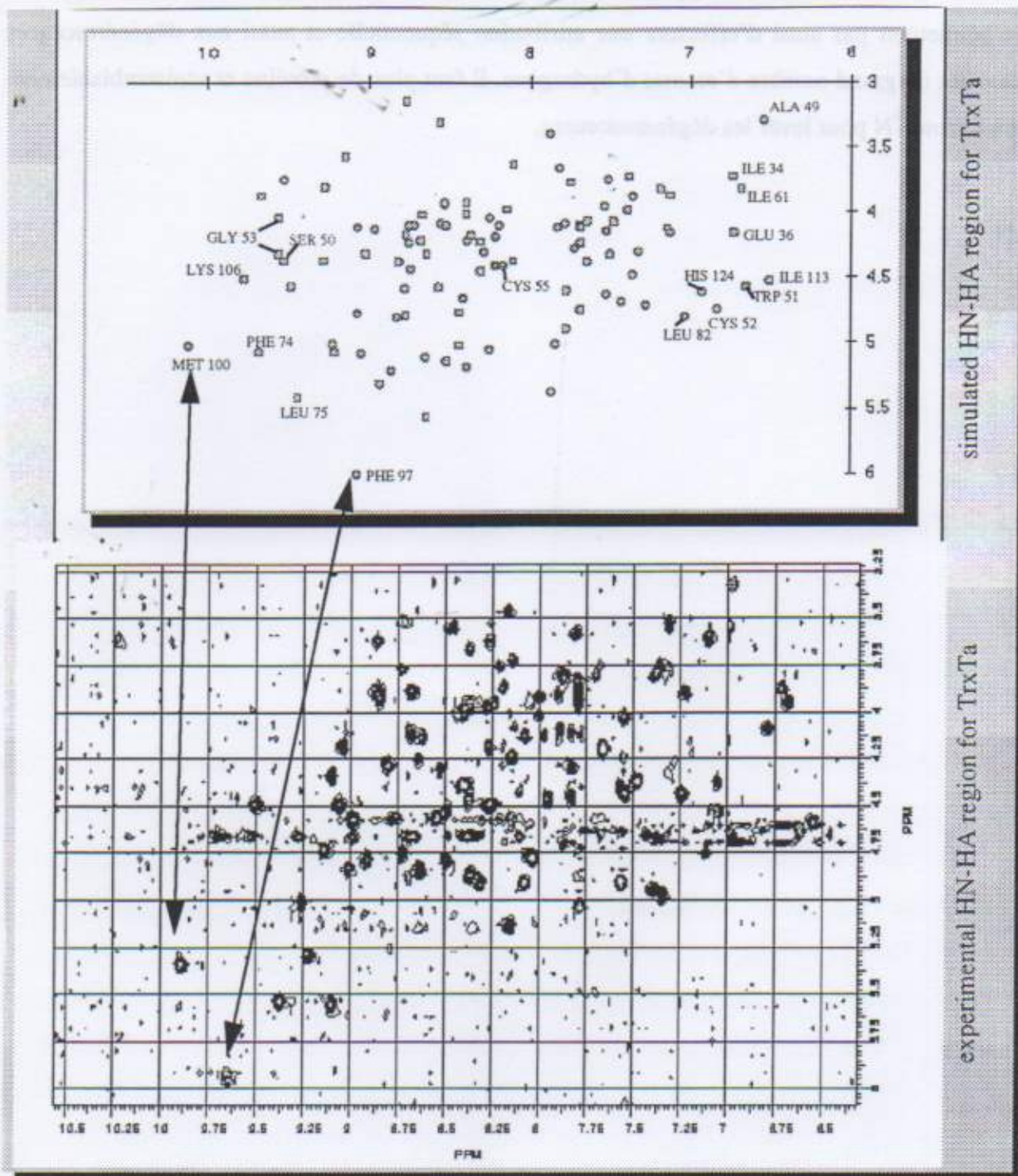


Figure 60 : Simulation obtenue pour la région HN/HA comparée à la même zone dans le spectre RMN

c) Discussion :

L'allure générale du spectre simulé est assez différente du spectre expérimental (dans la zone HN/H α). Toutefois, certaines corrélations sont retrouvées à ± 1 ppm près dans les deux dimensions spectrales (PHE97, MET100) et correspondent bien à celles trouvées expérimentalement, mais ce n'est pas systématique (voir Figure 61). Le fait d'avoir une erreur sur le déplacement chimique simulé par rapport au déplacement chimique expérimental ne permet que d'aider à détecter les résidus très impliqués dans les structures secondaires et ayant des taches de corrélation très isolées. Les difficultés

rencontrées, malgré cette aide assez appréciable de la simulation, sont restées associées au manque de signal ne permettant pas ainsi d'effectuer une attribution séquentielle et aussi aux dégénérescences observées entre un grand nombre d'atomes d'hydrogène. Il faut plus de protéine et vraisemblablement marquée à l'azote ^{15}N pour lever les dégénérescences.



Figure 10 : Spectres RMN de la thiorédoxine h de *Triticum Aestivum* marquée à l'azote ^{15}N .

CONCLUSION

L'étude structurale de la thiorédoxine h de *Triticum Aestivum* marquée à l'azote ^{15}N a permis d'obtenir des données RMN de haute résolution. Les données RMN ont été comparées avec les données de la structure cristalline de la thiorédoxine h de *Triticum Aestivum* (PDB: 1TTH) et ont permis de valider la structure cristalline. Les données RMN ont également permis d'identifier des sites de liaison pour des ligands et de caractériser des sites de liaison pour des ligands. Les données RMN ont été utilisées pour la modélisation de la structure de la thiorédoxine h de *Triticum Aestivum* et ont permis de valider la structure cristalline.

CONCLUSION GENERALE

Cette thèse m'a permis de mieux appréhender les différentes étapes d'un travail de recherche mais aussi de mieux comprendre le fonctionnement d'un laboratoire de recherche.

Une partie du travail de recherche introduit entre autre l'importance des thiorédoxines comme protéine de régulation de nombreux systèmes vivants. Ce sont des protéines qui sont à l'étude depuis les années soixante et dont l'intérêt que lui porte la recherche internationale ne cesse de croître. Les résultats présentés tendent à expliquer comment évolue la structure des thiorédoxines au moment de son interaction avec sa cible.

L'ensemble des travaux présentés dans cette thèse rendent compte de l'intérêt de la résonance magnétique nucléaire dans la recherche de structure des protéines. Ils font part des différentes étapes inhérentes à la détermination de structures par résonance magnétique nucléaire, de la biochimie à la modélisation des données obtenues en passant par la description de l'étape de géométrie de distance. Cette dernière méthode est par ailleurs bien détaillée montrant les performances actuelles et à venir de ce type d'approche pour l'obtention de structures par résonance magnétique nucléaire sans passer par l'attribution des spectres.

Actuellement, les différents groupes de recherche parviennent de plus en plus à marquer les macromolécules biologiques avec des hétéronoyaux (^{15}N , ^{13}C , etc.), ce qui permet de voir se développer dans un avenir proche de nouvelles méthodes de détermination de structures automatisées ou semi-automatisées utilisant les expériences de résonance magnétique nucléaire hétéronucléaire. C'est une technique de pointe en pleine croissance et on peut dire que de nos jours, l'informatique associée à la biologie (bio-informatique) est en pleine extension et nous promet de nouvelles découvertes scientifiques majeures (Li, K.B., 1996, *J Chem Inform Comput Sci*, 36 :3, 585-593) (Bernstein, R. et al., 1993, *J Magn Resonance Ser B*, 102 :1, 116-119) (Eads, C.D. et al., 1993, *J Magn Resonance Ser B*, 102 :3, 310-313) (Kleywegt, G.J., 1993, *J Magn Resonance Ser B*, 102 :2, 166-176) (Spraul M. et al., 1994, *J of Pharmaceutical and Biomedical Analysis*, 12 :10, 1215-1225). (Drusi A. et al., 1995, *Eur J Biochem*, 229 :2, 494-502).

ANNEXES

Chapitre VIII. Annexes:

A. Les différents types de systèmes de spins :

Insérer les deux pages décrivant les COSY, TOCSY, des systèmes de spins.

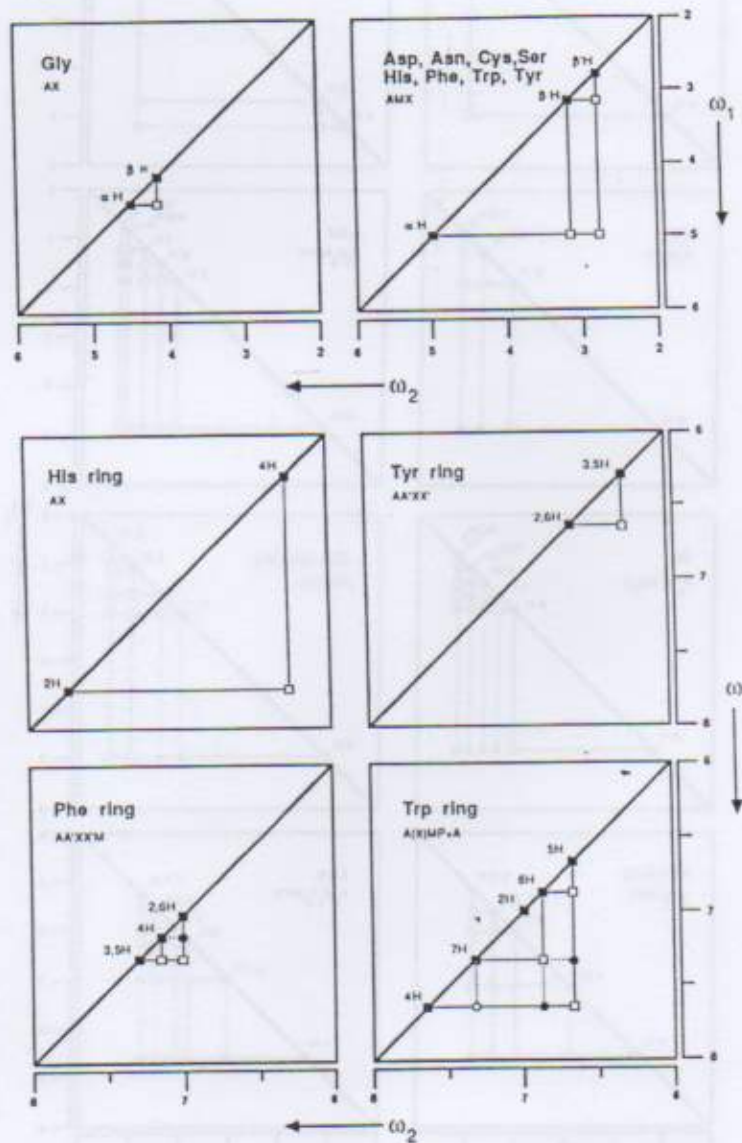


Fig. 2B.

relayed-COSY connectivities (●), and double-relayed-COSY connectivities (○). TOCSY or HOHAHA experiments will show all of these cross-peaks plus the ones indicated with crosses. Only direct COSY connectivities are joined with solid lines. The most common chemical shifts are indicated. (Adapted from Wider *et al.*⁴)

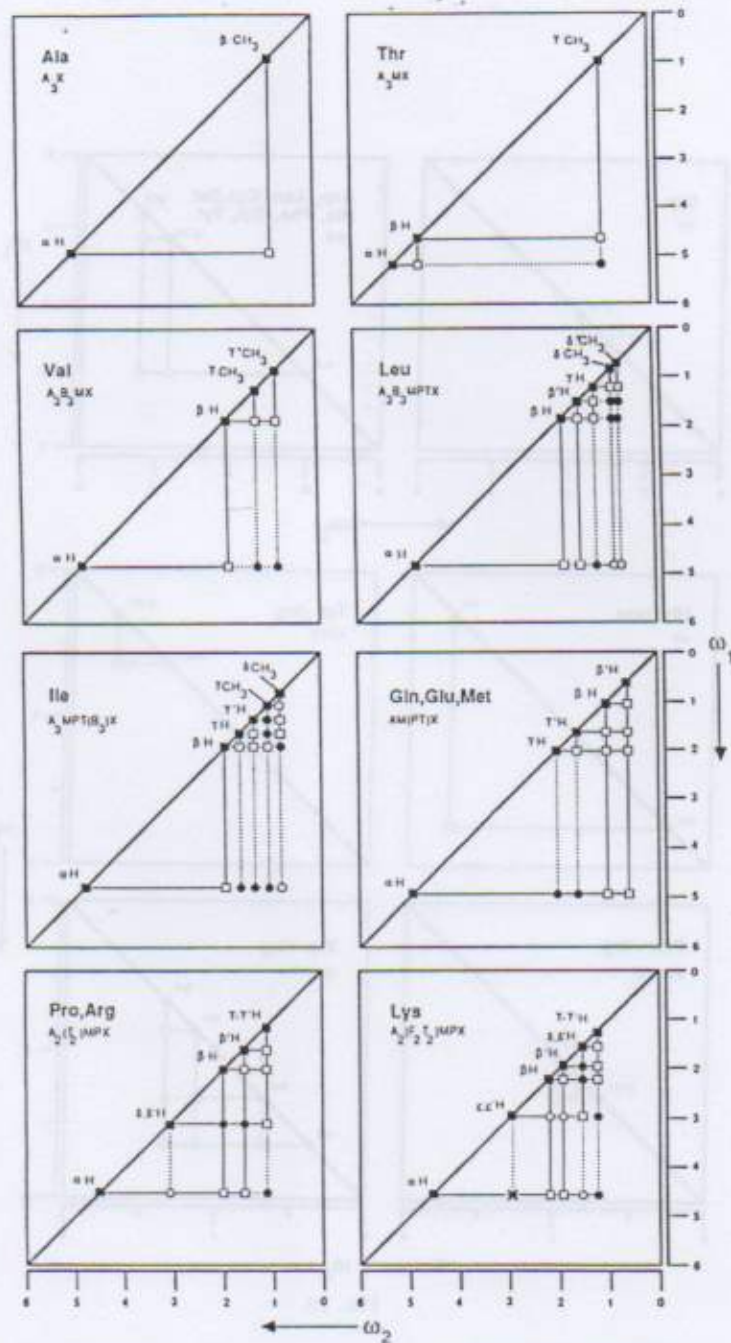


FIG. 2. Connectivity diagrams (A and B) for all the nonexchangeable protons of the common amino acid residues. Shown are diagonal peaks (■), COSY connectivities (□),

1. Les installations de traitement des eaux usées de la ville de ...

1.1. Les installations de traitement des eaux usées de la ville de ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

1.2. Les installations de traitement des eaux usées de la ville de ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

Le traitement des eaux usées de la ville de ... est assuré par une station d'épuration ...

- List of technical specifications or data points related to the water treatment installations.

Additional notes or references at the bottom of the page.

B. Les algorithmes de géométrie de distance en Matlab :

1. Les inégalités du triangle :

```
% Propagation de l'inegalite du triangle sur les matrices de contraintes
% de distance U(pper) et L(ower).
% fournit en sortie les contraintes dans U & L
%
function [U , L] = triangle(Upper , Lower)

rayon = 500;
contact = 0.5;

U = full(Upper);
U = (U==0)*rayon + U;
L = full(Lower);
L = (L==0)*contact + L;

dim = max(size(U));
for k = 1:dim;
    for i = 1:dim-1;
        j = i+1;
        U(i,j:dim) = min([U(i,j:dim)' (U(i,k) + U(k,j:dim))']));
        U(j:dim,i) = U(i,j:dim)';

        L(i,j:dim) = max([L(i,j:dim)' (L(i,k) - U(k,j:dim))']));
        L(i,j:dim) = max([L(i,j:dim)' (L(j:dim,k) - U(k,i))']));
        L(j:dim,i) = L(i,j:dim)';
    end;
end;
```

2. Les inégalités du tétrangle :

```
% Propagation de l'inegalite du tetriangle sur les matrices de contraintes
% de distance U(pper) et L(ower).
% fournit en sortie les contraintes dans U & L
%
function [U , L] = tetriangle3(matA, matB)

rayon = 500;
rayon2 = 250000;
contact = 0.5;
contact2 = 0.25;

U = full(matA);
U = (U==0)*rayon + U;
L = full(matB);
L = (L==0)*contact + L;

dim = max(size(U));

for z=1:dim % ajout du 10/02/94
    U(z,z)=0.0; % ajout du 10/02/94
    L(z,z)=0.0; % ajout du 10/02/94
end;

%(U,L) = triangle2(U,L);

for i=1:dim-1;
    for j=i+1:dim;
        for k=1:dim-1;
            if (k==i) & (k==j)

                P = ones(1,dim);
                P(1,1:k) = zeros(1,k);
                P(1,i) = 0;
                P(1,j) = 0;

                for zz = 1:dim; %
                    if (P(1,zz)==1); %
                        if (L(k,zz) == U(k,zz)); %
                            P(1,zz)=0; % Pb L(k,p) = U(k,p)
                        end; %
                    end; %
                end; %

            if (max(P') == 0)

                [U1,L1] = calmeng3(L(1,j),U(1,k),U(1,P),U(j,k),U(j,P));
                [U2,L2] = calmeng3(U(1,j),L(1,k),L(1,P),U(j,k),U(j,P));
                [U3,L3] = calmeng3(U(1,j),U(1,k),U(1,P),L(j,k),L(j,P));
                [U4,L4] = calmeng3(U(1,j),U(1,k),L(1,P),L(j,k),U(j,P));
                [U5,L5] = calmeng3(U(1,j),L(1,k),U(1,P),U(j,k),L(j,P));
                [U6,L6] = calmeng3(L(1,j),L(1,k),U(1,P),L(j,k),U(j,P));
                [U7,L7] = calmeng3(L(1,j),U(1,k),L(1,P),U(j,k),L(j,P));

                Um= max([U1;U2;U3;U4;U5;U6;U7]);
                Lm= min([L1;L2;L3;L4;L5;L6;L7]);
            end;
        end;
    end;
end;
```

```

Um= sqrt(Um .* (Um>0) );
Lm= sqrt(Lm .* (Lm>0) );

U(k,P) = min (U(k,P), Um);
U(P,k) = U(k,P)';
L(k,P) = max (L(k,P), Lm);
L(P,k) = L(k,P)';

end;
end;
end;
end;
end;

```

3. La métrisation :

```

% Métrisation : Choix aleatoire de distances entre la borne inf. et la
% borne sup. (métrisation partielle)
%
% "Knowing the distances from four atoms to all others is equivalent to
% knowing the distances from all atoms to all others."
% 'Four-atoms Metrization'
%
% Donne en sortie une matrice de distance D
%

```

```
function D = met(type_met,methode,U,L,p_cent)
```

```

dim = max(size(U));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if (type_met == 1)
    for i = 1:dim;
        for j = i+1:dim;
            U(i,j) = L(i,j) + rand(1)*(U(i,j)-L(i,j));
            U(j,i) = U(i,j); L(i,j) = U(i,j); L(j,i) = L(i,j);
            if (methode == 4) [U , L] = tetriangle3(U,L);end;
            if (methode == 3) [U , L] = subtria(U,L,i,j);end;
        end;
    end;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
else
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% generateur aleatoire de parcours de i et j
%
A = alea(dim);
compteur = ((dim-1)*dim/2); % cas metrisation totale
if (type_met == 3), compteur = fix(p_cent*(dim-1)*dim/200), end; %partielle
%
% metrisation aleatoire totale(2) ou partielle(3)
%
for I = 1:compteur;
    U(A(1,I),A(2,I)) = L(A(1,I),A(2,I)) + rand(1)*(U(A(1,I),A(2,I))-L(A(1,I),A(2,I)));
    U(A(2,I),A(1,I)) = U(A(1,I),A(2,I)); L(A(1,I),A(2,I)) = U(A(1,I),A(2,I));
    L(A(2,I),A(1,I)) = L(A(1,I),A(2,I));
    if (methode == 4) [U , L] = tetriangle3(U,L);end;
    if (methode == 3) [U , L] = subtria(U,L,A(1,I),A(2,I));end;
end;
end;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
D = U;
for i = 1:dim
    D(i,i) = 0.0;
end;

```

a) alea.m :

```

% Donne une matrice d'indice i,j nommée A
% de dimension (2 lignes, n colonnes)
%

```

```
function A = alea(n)
```

```

%
% generateur aleatoire de parcours de i et j
%
A = [];
for i = 1:n;
    x = ones(1,n)*i;
    A = [A {x(1:(n-i))' (i+1:n)'}];
end;
sig_perm = randperm((n-1)*n/2);

```

```
A = A(:,sig_perm(:));
```

4. Le plongement :

```
% fonction coor = embed(D);
% fournit une matrice 3xN des coordonnees a partir d'une matrice de
% distance D NxN ;
% etape "d'embedding" ref: Acta Cryst. (1978). A34,282-284
%
function coor = embed(D);

global G;
N= max(size(D));
G=zeros(N);
DD= D.*D; DD1=zeros(N,1);
DD2=0.0;
coor=zeros(N,3);

for j=1:N ;
    DD1(:,j) = DD1(:,j) + DD(:,j) ;
end;
for j=2:N;
    for k=1:j-1;
        DD2= DD2 + DD(j,k);
    end;
end;
DD1=DD1/N;
DD2=DD2/(N*N);
DD1(:,1) = DD1-DD2;

for i=1:N
    for j=1:N
        G(i,j) = 0.5 * (DD1(i,1)+DD1(j,1)-DD(i,j));
    end;
end;

sum(sum(G))

[EigVectors,EigValues] = eig(G);

% tri des valeurs propres pour recuperer les plus grandes
%
% etude sur les valeurs propres
[toto,index]=sort(abs(diag(EigValues))); % ligne a ajouter pour avoir les
VP=diag(EigValues(index,index)); % valeurs propres trieles! <>

somme_des_carres = sum(VP(:,1).^2);
pourcentage_3prem = sum(VP(N-2:N,1).^2)*100/somme_des_carres
pcent_de_la_1 = VP(N,1).^2*100/somme_des_carres
pcent_de_la_2 = VP(N-1,1).^2*100/somme_des_carres
pcent_de_la_3 = VP(N-2,1).^2*100/somme_des_carres

[EigValues,index] = sort(diag(EigValues));
coor(:,1) = sqrt(abs(EigValues(N))) * EigVectors(:,index(N));
coor(:,2) = sqrt(abs(EigValues(N-1))) * EigVectors(:,index(N-1));
coor(:,3) = sqrt(abs(EigValues(N-2))) * EigVectors(:,index(N-2));
```

5. La minimisation :

a) Algorithme de majoration :

```
% main macro
% The majorization algorithm
% J.of Computational chemistry, Vol14, (1),p.114-120(1993)
%
function Xout = maj(X0,Upper,Lower);
D0 = dist(X0); % calcul de la matrice de distance associee a X0
N = max(size(D0));
% calcul de la matrice des "dissimilarities"
Delta0 = (D0<Lower).*Lower + (D0>Upper).*Upper + ((D0>=Lower).*(D0<=Upper)).*D0;

%----- initialise -----
% calcul du poids donne a chaque contrainte
W = 1./(1+10*(Upper-Lower));
% construction de V
V = (-W).*(eye(N)==0);Vdiag=zeros(N);
for i=1:N
    Vdiag(i,i) = sum(W(i,:)) - W(i,i);
end;
V = V + Vdiag;
Vi=inv(V);
%-----
sig0 = 0.5*sum(sum(W.*(D0-Delta0).*(D0-Delta0))) % fonction a minimiser
% etat initial (sig0)
```

```

Xk = majalg(X0,Vi,W,Lower,Upper);
Dk=dist(Xk);
Deltak = (Dk<Lower).*Lower + (Dk>Upper).*Upper + ((Dk>=Lower).*(Dk<=Upper)).*Dk;
sigk = 0.5*sum(sum(W.*((Dk-Deltak).*(Dk-Deltak))));
sigmin=sig0;X=X0;
erreur = 1.0e-5;x=1;
while (abs(sig0-sigk) >erreur);
    sig0=sigk;X=Xk;
    Xk= majalg(X0,Vi,W,Lower,Upper);
    Dk=dist(Xk);
    Deltak = (Dk<Lower).*Lower + (Dk>Upper).*Upper + ((Dk>=Lower).*(Dk<=Upper)).*Dk;
    sigk = 0.5*sum(sum(W.*((Dk-Deltak).*(Dk-Deltak))));
    if (sigk<sigmin) sigmin=sigk;X=Xk;p=x+1;P=ones(10,1)*p;end;
    x=x+1;
    y = [y;sigk];
    if (x>2000) sig0=sigk;end;
end;
sigmin
Xout = X;

```

----- end main ----- begin sub routine

```

% The majorization algorithm
% J.of Computational chemistry, Vol14, (1),p.114-120(1993)
%
function Y = majalg(X,Vi,W,Lower,Upper);
DD=dist(X);
N=length(X);
% calcul de la matrice des "dissimilarities"
DELTA = (DD<Lower).*Lower + (DD>Upper).*Upper + ((DD>=Lower).*(DD<=Upper)).*DD;
% construction de Bk(X)
Bk = (eye(N)==0).*((DD>0).*(-W).*DELTA./(DD+(DD==0)*1));Bkdiag = zeros(N);
for i=1:N
    Bkdiag(i,i) = - sum(Bk(i,:)) + Bk(i,i);
end;
Bk = Bk + Bkdiag;
Y = Vi*Bk*X;

```

b) Monte-carlo Métropolis :

```

% Monte-Carlo Metropolis
%
% [Xout,sig0,sig1] = metro(X0,Upper,Lower);

function [Xout,sig0,sig1] = metro(X0,Upper,Lower);

DO = dist(X0); % calcul de la matrice de distance associee a X0
N = max(size(DO)); % recupere la dimension de la matrice carree
pas = 0.25;
% calcul de la matrice des "dissimilarities"
Delta0 = (DO<Lower).*Lower + (DO>Upper).*Upper + ((DO>=Lower).*(DO<=Upper)).*DO;

% calcul du poids donne a chaque contrainte
W = 1./(1+10*(Upper-Lower));

sig0 = 0.5*sum(sum(W.*((DO-Delta0).*(DO-Delta0)))); % fonction a minimiser
% fonction relative aux ecarts par rapport aux contraintes

i = fix(rand(1)*N)+1;
j = fix(rand(1)*3)+1; % modification aleatoire (choix de Xij)

X1= X0; X1(i,j) = X1(i,j) + (rand(1)*2-1)*pas*X1(i,j);

D1 = dist(X1);
Deltal = (D1<Lower).*Lower + (D1>Upper).*Upper + ((D1>=Lower).*(D1<=Upper)).*D1;
sig1 = 0.5*sum(sum(W.*((D1-Deltal).*(D1-Deltal))));

dE = sig1-sig0;
kT = 0.1;
if (dE<0);
    Xout = X1;
else;
    P = exp(-dE/kT);
    x = rand(1);
    if (x<P);
        Xout=X1;
    else
        Xout=X0;
    end;
end;
end;

```

6. Autres :

a) subtria.m :

```

% Propagation de l'inegalite du triangle sur les matrices de contraintes
% de distance U(pper) et L(ower).
% fournit en sortie les contraintes dans U & L
%
function [U , L] = subtria(Upper , Lower, a, b)

```



```

rayon = 500;
contact = 0.5;

U = full(Upper);
U = (U==0)*rayon + U;
L = full(Lower);
L = (L==0)*contact + L;

dim = max(size(U));

for k = 1:dim;
    U(a,(a+1):dim) = min([U(a,(a+1):dim)' (U(a,k) + U(k,(a+1):dim))]);
    U((a+1):dim,a) = U(a,(a+1):dim)';

    L(a,(a+1):dim) = max([L(a,(a+1):dim)' (L(a,k) - U(k,(a+1):dim))]);
    L(a,(a+1):dim) = max([L(a,(a+1):dim)' (L((a+1):dim,k) - U(k,a))]);
    L((a+1):dim,a) = L(a,(a+1):dim)';
end;

for k = 1:dim;
    U(1:(b-1),b) = min([U(1:(b-1),b) (U(1:(b-1),k) + U(k,b))]);
    U(b,1:(b-1)) = U(1:(b-1),b)';

    L(1:(b-1),b) = max([L(1:(b-1),b) (L(1:(b-1),k) - U(k,b))]);
    L(1:(b-1),b) = max([L(1:(b-1),b) (L(b,k) - U(k,1:(b-1)))]);
    L(b,1:(b-1)) = L(1:(b-1),b)';
end;

```

b) calmeng.m :

% Calcul de l'extension de Laplace du Determinant de Caley-Menger pour
% la determination des inegalite du

% TETRANGLE

% - en entree 5 distances d12 d13 d14 d23 d24

% - en sortie L34 et D34

%

function [U34,L34] = calmeng(d12,d13,d14,d23,d24)

rayon2 = 250000;

contact2 = 0.25;

D12= d12*d12;

D13= d13*d13;

D14= d14*d14;

D23= d23*d23;

D24= d24*d24;

```

%A34= -1/8 * det([0 1 1 ;
%             1 0 D12 ; % Symetrique
%             1 D12 0 ]);

```

A34 = - D12/4;

```

%B34= -1/4 * det([0 1 1 1 ;
%             1 0 D12 D14 ;
%             1 D12 0 D24 ; % Asymetrique
%             1 D13 D23 0 ]);

```

B34 = (1/4)*(-D12*D12+D12*(D13+D14+D23+D24)+(D13-D23)*(D24-D14));

```

%C34= 1/8 * det([0 1 1 1 1 ;
%             1 0 D12 D13 D14 ;
%             1 D12 0 D23 D24 ; % Symetrique
%             1 D13 D23 0 0 ;
%             1 D14 D24 0 0 ]);

```

v1= D23-D24;

v2= D14-D13;

C34= (1/4)*((D13*D24-D14*D23)*(v1+v2)+D12*v1*v2);

Delta= B34*B34 - 4*A34*C34;

denom= 2*A34;

Dzero= -B34/denom;

if (Delta >= 0)

RDelta = sqrt(Delta);

L34 = Dzero + RDelta/denom;

U34 = Dzero - RDelta/denom;

else

L34=contact2;

U34=rayon2;

end;

c) energie violation2.m :

% calcul la valeur de la fonction d'energie associe a l'ecart
% par rapport aux contraintes fixees (contraintes de distance)

%

```

function sig = energie_violation2(D0,Upper,Lower);

% calcul de la matrice des "dissimilarities"
Delta0 = (D0<Lower).*Lower + (D0>Upper).*Upper + ((D0>=Lower).*(D0<=Upper)).*D0;
% calcul du poids donne a chaque contrainte
W = 1./(1+10*(Upper-Lower));
sig = 0.5*sum(sum(W.*(D0-Delta0).*(D0-Delta0))); % fonction a minimiser

```

d) megamax.m :

```

% calcul le max de (a-b) * (c-d) avec
% respectivement La<=a<=Ua, ..., Ld<=d<=Ud
% Dmax = megamax(La,Ua,Lb,Ub,Lc,Uc,Ld,Ud)
%
function Dmax = megamax(La,Ua,Lb,Ub,Lc,Uc,Ld,Ud)

Dmax = (Ua-Lb)*(Uc-Ld);
Dmax = max(max(max(Dmax, (Ua-Lb)*(Lc-Ud)), (La-Ub)*(Lc-Ud)), (La-Ub)*(Uc-Ld));

```

e) bornes.m :

```

% Calcul les valeurs superieures et inferieures LA,UA,LB,UB,LC et UC
% des coefficients du polynome obtenu en developpant l'extension de
% Laplace du determinant de Caley-Menger
% [G.M.Crippen & T.F. Havel, D.G. and Mol.Conformation, p 256]
%
% Modification de la propagation des inegalites du tetrangle;
%
% [Bsup,Binf] = bornes(L12,U12,L13,U13,L14,U14,L23,U23,L24,U24);
%

```

```

function [Bsup,Binf] = Bornes(L12,U12,L13,U13,L14,U14,L23,U23,L24,U24)
L12=L12*L12;
U12=U12*U12;
L13=L13*L13;
U13=U13*U13;
L14=L14*L14;
U14=U14*U14;
L23=L23*L23;
U23=U23*U23;
L24=L24*L24;
U24=U24*U24;

LA= -U12*U12/4;
UA= -L12*L12/4; %Bornes de A34 (c.f. reference)

cst1= megamax(L13,U13,L23,U23,L24,U24,L14,U14);
UB= (1/4)*(-L12*L12+U12*(U13+U14+U23+U24)+cst1);
cst2= millimn(L13,U13,L23,U23,L24,U24,L14,U14);
LB= (1/4)*(-U12*U12+L12*(L13+L14+L23+L24)+cst2); %Bornes de B34

cst1= megamax(L13*L24,U13*U24,L14*L23,U14*U23,L23+L14,U23+U14,L24+L13,U24+U13);
cst2= megamax(L23,U23,L24,U24,L14,U14,L13,U13);
UC= (1/4)*(cst1+U12*cst2);
cst1= millimn(L13*L24,U13*U24,L14*L23,U14*U23,L23+L14,U23+U14,L24+L13,U24+U13);
cst2= millimn(L23,U23,L24,U24,L14,U14,L13,U13);
LC= (1/4)*(cst1+L12*cst2); %Bornes de C34

%URdis= sqrt(max(UB*UB, LB*LB)-4*UC*LA);
if (min(UB*UB, LB*LB)-4*UA*LC >= 0)
    LRdis= sqrt(min(UB*UB, LB*LB)-4*UA*LC);
    Bsup= (-UB+LRdis)/(2*LA);
    Binf= (-LB-LRdis)/(2*UA);
    if (Binf > Bsup)
        temp = Bsup;
        Bsup = Binf;
        Binf = temp;
    end;
else
    Bsup = 300000;
    Binf = 0.00001;
end;

```

C. Les algorithmes de géométrie de distance en FORTRAN :

1. Le fichier makefile :

```

LAPACK = /d4/LAPACK/lapack.a
BLAS   = /d4/LAPACK/blas.a
LINK   = f77 -g +U77
FORT # = f77 -c +U77 -g
# Options: -c compile
#          -O Optimise
#          -g debug(incompatible avec -O)
#          +U77 link /usr/lib/libU77.a pour etime()

OBJECT = tetra.o triangle.o dg.o random.o alea.o met.o subtria.o subtetra.o \
        embed.o majalg.o triangle_4.o alea_simpl.o

dg : $(OBJECT)
    $(LINK) -o dg $(OBJECT) $(LAPACK) $(BLAS)
dbinv : inversmat.o
    $(LINK) -o dbinv inversmat.o $(LAPACK) $(BLAS)

dg.o : dg.f
    $(FORT) dg.f
triangle.o : triangle.f
    $(FORT) triangle.f
subtria.o : subtria.f
    $(FORT) subtria.f
triangle_4.o : triangle_4.f
    $(FORT) triangle_4.f
tetra.o : tetra.f
    $(FORT) tetra.f
subtetra.o : subtetra.f
    $(FORT) subtetra.f
random.o : random.f
    $(FORT) random.f
alea.o : alea.f
    $(FORT) alea.f
alea_simpl.o : alea_simpl.f
    $(FORT) alea_simpl.f
met.o : met.f
    $(FORT) met.f
embed.o : embed.f
    $(FORT) embed.f
majalg.o : majalg.f
    $(FORT) majalg.f
inversmat.o : inversmat.f
    $(FORT) inversmat.f

clean :
    rm -f *.o

```

2. Le fichier principal dg.f :

```

C 7 MAIN dg.f
C Lecture des contraintes de distances et remplissage des
C tableaux U(pper) et L(ower)
C
C -----
C |          PROCEDURE DE SAUVEGARDE ...
C | SAUVEGARDE la Matrice des metrisations successives apres
C | chaque etape de metrisation
C |-----
C Subroutine save_step(M,Leading,N,nbre)

implicit none

integer Leading,N,nbre,x1,x2,ios
real M(Leading,nbre*Leading)

C
C --> status qui ecrase le fichier dg.tmp s'il existait
C
open (unit=8,err=7,file='dg.tmp',status='UNKNOWN')
do x1=1,N
    write(8,fmt=120, iostat= ios, err= 7) (M(x1,x2),x2=1,N*nbre)
enddo
close(8)
Return

7 write(*,*) '**** erreur d'écriture ****'
120 format(20000 F7.2)
End

C =====
C MAIN PROGRAM
C =====
C
PROGRAM dg

implicit none

```

```

character*80 Filename,alpha
integer N,ios,i,j,Nmax,rayon,compte,nbmet,ffst
real time,etime,time_array(2) ! servent pour mesurer temps CPU
real d,dplus,dmoins,contact
parameter (Nmax=200)
parameter (rayon=500,contact=0.5)
parameter (nbmet=5)

real U(Nmax,Nmax),L(Nmax,Nmax),Upp(Nmax,Nmax),Low(Nmax,Nmax)
real A(Nmax,Nmax),B(Nmax,Nmax),Vinv(Nmax,Nmax),W(Nmax,Nmax)
real Dmet(Nmax,nbmet*Nmax),XofA(Nmax,3),XX(Nmax,3)
real Energie,Egie

external etime,Energie

C common pour verifier s'il s'agit d'un premier passage.
COMMON /FIRST/ ffst

C Initialisations
C
ffst = 0 ! sert pour la sousroutine maj
open(unit=9,err=7,file='Xembed.out',status='UNKNOWN')
close(9) ! cree fichier de sauvegarde des coordonnees calculees

C LECTURE DES CONTRAINTES DE DISTANCE
C demande nom du fichier de contraintes

C Filename='bpti_hnha.datp' ! dist. HN/HA <5A +/- 0.5A
C Filename='bpti_hnha_4A.datp' ! dist. HN/HA <4A +/- 0.5A
C Filename='bpti_hnha_3A.datp' ! dist. HN/HA <4A +/- 0.5A
C ,Filepame='helice.dat'
Filename='cube.dat'
write(*,*) Filename

C Initialise les tableaux U et L

U(Nmax,Nmax)=0
L(Nmax,Nmax)=0
do i=1,Nmax-1
  U(i,1) = 0
  L(i,1) = 0
  do j=i+1,Nmax
    U(i,j) = rayon
    U(j,i) = rayon
    L(i,j) = contact
    L(j,i) = contact
  enddo
enddo

C Lit les contraintes et modifie U,L
C distance = d +/- dplus/dmoins

open (unit=7,err=5,file=Filename,status='OLD')
read (7,fmt=*,iostat= ios, err= 6) N
do while (ios.EQ.0)
  read (7,fmt=*,iostat= ios, err= 6) i,j,d,dmoins,dplus
  U(i,j) = d+dplus
  U(j,i) = U(i,j)
  L(i,j) = d-dmoins
  L(j,i) = L(i,j)
enddo
close(7)

do i=1,N-1
  do j=i+1,N ! recuperation des contraintes
    ! initiales dans Upp & Low
    Upp(i,j) = U(i,j) ! (necessaires pour l'amelioration
    Upp(j,i) = Upp(i,j) ! de structures)
    Low(i,j) = L(i,j)
    Low(j,i) = Low(i,j)
  enddo
  Upp(i,i) = 0.
  Low(i,i) = 0.
enddo
Upp(N,N) = 0
Low(N,N) = 0

write(*,*) 'Nombre de points: ',N
-----
C Sauvegarde Upper-Lower Matrices
C
open (unit=9,err=7,file='bpti_upper.dat',status='UNKNOWN')
open (unit=10,err=7,file='bpti_lower.dat',status='UNKNOWN')
do i=1,N
  write(9,100) (Upp(i,j),j=1,N)
  write(10,100) (Low(i,j),j=1,N)
enddo
close(9)

```

```

c      close(10)
c      STOP

C -----
C Propagation des inegalites du triangle
C
      CALL triangle(Nmax,U,L,N)
C -----
C Propagation des inegalites du tetrangle
C
      CALL tetra(Nmax,U,L,N)
C -----
c      do compte=1,nhmet
C remplissage de A et B avec U et L (reinitialisation)
      do i=1,N
        do j=1,N
          A(i,j)=U(i,j)
          B(i,j)=L(i,j)
        enddo
      enddo

C
C      METRISATION par les ineg. du triangle 'trian'
C      ou du tetrangle 'tetra'
C
      CALL met(4,'tetra',Nmax,N,A,B,100)

C met.doc:
C      Call met(p,'trian',Nmax,N,A,B,p_cent)
C      trian : propagation des inegalites du triangle
C      tetra : propagation des inegalites du tetrangle
C      p>=4 : metrise / p atomes (metrisation partielle aleatoire)
C      p=1 : metrisation totale ordonnee
C      p=2 : metrisation totale aleatoire
C      p=3 : metrisation partielle aleatoire a p_cent%
C      p_cent pris en compte uniquement si p=3.

      do i=1,N
        do j=1,N
          Dmet(i,(j+(compte-1)*N))=A(i,j)
        enddo
      enddo

C      goto 1000
C =====
C
C --- TESTS SUR EMBEDDING effectuee sur 5 metrisations avec cube.dat
C      compare aux resultats de matlab
C -> resultats similaires sauf orientation de l'espace
C      (sens direct ou indirect)
C
      write(*,*) 'embedding num.',compte
      Call embed(A,Nmax,N,XofA,5*N,8*N)
      write(*,*) 'initialise Vinv et W ...'
C calcul de l'inverse ...
      Call init_VW(Vinv,W,Upp,Low,Nmax,N)
      write(*,*) 'inverse O.K., calcule energie'
      Egie = Energie(XofA,U,L,Nmax,N,Vinv,W)
      write(*,*) 'Energie de la structure projetee: ',Egie
C =====
C --- AMELIORATION DE STRUCTURE ---
C raffinement de structure apres embedding: minimisation des violations
C [algorithme de majorisation]
C
      write(*,*) 'Raffinement'
      Call maj(XX,XofA,Upp,Low,Nmax,N) ! Raffinement
      write(*,*) 'Sauvegarde...'
      open(unit=9,err=7,file='Xembed.out',status='OLD')

c positionnement en fin de fichier:
      read(9,fmt=*,iostat=ios,err=6) alpha
      do while (ios.eq.0)
        read(9,fmt=*,iostat=ios,err=6) alpha
      enddo

c sauvegarde des coordonnees minimisees:
      do i=1,N
        write(9,160) (XX(i,j),j=1,3)
      enddo
      write(9,*) ! passe une ligne

      close(9)

C
C ----- sauvegarde des resultats -----
C
1000 Call save_step(Dmet,Nmax,N,nhmet)

C ----- decompote avancee du calcul -----
      open (unit=9,err=7,file='dga.tmp',status='UNKNOWN')
      write(9,*) '*** avancee de dg ***'
      write(9,130) (100*compte/nhmet),' % effectues'

```

```

      close(9)
    enddo
C
C Determination de la duree d'execution (etime)
C
      time = etime(time_array)
      write(*,*) 'real: ',time
      write(*,*) 'user: ',time_array(1)
      write(*,*) 'sys : ',time_array(2)
4 STOP

```

C TRAITEMENT DES ERREURS

```

5 write(*,*) '**** erreur a l'ouverture ****'
  stop
6 write(*,*) '**** erreur de lecture ****'
  stop
7 write(*,*) '**** erreur d'écriture ****'
  stop

```

C FORMATS

```

100 format(200 (F8.3, ' '))
110 format(18 F6.2)
120 format(1800 F6.2)
130 format(F6.2,A12)
150 format(200 F7.2)
160 format(3 F15.5)
END

```

3. Les inegalités du triangle :

```

C 7 file triangle.f
C Propagation de l'inegalite du triangle sur les matrices de contraintes
C de distance U(pper) et L(ower)
C

```

```

SUBROUTINE triangle(Leading,U,L,N)

```

```

  implicit none

```

```

  integer Leading,N,i,j,k
  real U(Leading,Leading) , L(Leading,Leading)

```

```

  do k=1,N
    do i=1,N-1
      do j=i+1,N
        U(i,j) = min(U(i,j),U(i,k)+U(k,j))
        U(j,i) = U(i,j)
        L(i,j) = max(L(i,k)-U(k,j),L(j,k)-U(k,i),L(i,j))
        L(j,i) = L(i,j)
      enddo
    enddo
  enddo
  return
END

```

4. Les inegalités du tétrangle :

```

C 7 file tetra.f
C Propagation de l'inegalite du tetrangle sur les matrices de contraintes
C de distance U(pper) et L(ower)
C

```

```

SUBROUTINE calmeng(a,b,d12,d13,d14,d23,d24)

```

```

  implicit none

```

```

  real a,b,d12,d13,d14,d23,d24,A34,B34,C34
  real dd12,dd13,dd14,dd23,dd24,v1,v2
  real rayon2,contact2,delta,Rdelta,denom,Dzero

```

```

C definition du max et du min des fourchettes de distance^2
  parameter(rayon2=250000,contact2=0.25)

```

```

C calcul des distances au carre

```

```

  dd12=d12**2
  dd13=d13**2
  dd14=d14**2
  dd23=d23**2
  dd24=d24**2
  v1=dd23-dd24
  v2=dd14-dd13

```

```

C 1er determinant de Caley-Menger

```

```

  A34 = -dd12*0.25

```

```

C 2ieme determinant de Caley-Menger

```

```

  B34 = 0.25*(-dd12*dd12 + dd12*(dd13+dd14+dd23+dd24)+
  + (dd13-dd23)*(dd24-dd14))

```

```

C 3ieme determinant de Caley-Menger

```

```

  C34 = 0.25*((dd13+dd24-dd14*dd23)*(v1+v2)+dd12*v1*v2)

```

```

C determination des racines de  $A34X^2+B34X+C34=0$ 
  Delta= B34*B34-4*A34*C34
  denom = 2*A34
  Dzero = -B34/denom

  if ((Delta.GE.0).and.(denom.LT.0)) then
    Rdelta = sqrt(Delta)/denom
    b = Dzero+Rdelta
    a = Dzero-Rdelta
  else

C s'il n'y a pas de racines => soit fourchette maximale
C   b = contact2
C   a = rayon2
C s'il n'y a pas de racines => soit pas de changements (tout en commentaire)
  end if
  Return
END

SUBROUTINE tetra(Leading,U,L,N)
  implicit none

  integer Leading,N,i,j,k,p
  real U(Leading,Leading) , L(Leading,Leading)
  real Um,Im
  real u1,u2,u3,u4,u5,u6,u7,l1,l2,l3,l4,l5,l6,l7

C propagation des inegalites du triangle
  CALL triangle(Leading,U,L,N)

C propagation des inegalites du tetrangle
  do i=1,N-1
    do j=i+1,N
      do k=1,N-1
        do p=k+1,N

C ----- boucle sur p -----

C tests sur i,j,k,p pour avoir des quadruplets distincts
C et si L(k,p) = U(k,p) alors on ne fait rien
          if ((k.NE.j).AND.(k.NE.i).AND.(p.NE.j).AND.(p.NE.i).AND.
             & (L(k,p).LT.U(k,p))) then

C calcul des 7 determinants de Caley-Menger relatifs au tetrangle
C Bull. of Mathematical Biology, Vol 51, (1),p173-194, (1989)
            CALL calmeng(u1,l1,L(i,j),U(i,k),U(i,p),U(j,k),U(j,p))
            CALL calmeng(u2,l2,U(i,j),L(i,k),L(i,p),U(j,k),U(j,p))
            CALL calmeng(u3,l3,U(i,j),U(i,k),U(i,p),L(j,k),L(j,p))
            CALL calmeng(u4,l4,U(i,j),U(i,k),L(i,p),L(j,k),U(j,p))
            CALL calmeng(u5,l5,U(i,j),L(i,k),U(i,p),U(j,k),L(j,p))
            CALL calmeng(u6,l6,L(i,j),L(i,k),U(i,p),L(j,k),U(j,p))
            CALL calmeng(u7,l7,L(i,j),U(i,k),L(i,p),U(j,k),L(j,p))

C recuperation des bornes calculees (Um > Im)
            Um= max(u1,u2,u3,u4,u5,u6,u7)
            Im= min(l1,l2,l3,l4,l5,l6,l7)

C Tests pour assurer une geometrie consistante
C symetrisation des matrices U et L
            if ((Im.GE.0).and.(sqrt(Im).LE.U(k,p))
               & .and.(sqrt(Um).GE.L(k,p))) then
              U(k,p)= min(U(k,p),sqrt(Um))
              U(p,k)= U(k,p)
              L(k,p)= max(L(k,p),sqrt(Im))
              L(p,k)= L(k,p)
            end if
          end if
        end do
      end do
    end do
  end do

C Propagation des inegalites du triangle...
C Cas ou certains U(k,p) n'ont pu etre traites par
C la propagation des inegalites du tetrangle
C (NB: test : Cette position a donne les meilleurs
C au niveau des violations sur les resultats)
  CALL Triangle(Leading,U,L,N)

  return
END
5. La métrisation :
C 7... file met.f

```

```

C Metrisation : Choix aleatoire de distances entre la borne inf. et la
C borne sup. (metrisation partielle ou totale)
C
C "Knowing the distances from four atoms to all others is equivalent to
C knowing the distances from all atoms to all others."
C 'Four-atoms Metrization' -> a faire
C
C Donne en sortie une matrice de distance D
C
C utilisation: D = met(Type_met,methode,Leading,N,U,L,p_cent)
C
C SubRoutine met(Type_met,methode,Leading,N,U,L,p_cent)
C
C implicit none
C
C Character(5) methode
C integer Type_met,p_cent,Leading,N,i,j,Pos(2,(Leading-1)*Leading/2)
C integer graine,co,X(Type_met)
C real U(Leading,Leading),L(Leading,Leading),random
C
C external random
C
C parameter(graine=2047)
C
C @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
C Metrisation Ordonnee totale
C
C if (Type_met.EQ.1) then
C   do i = 1,N-1
C     do j = i+1,N
C       ----- stocke avance du calcul -----
C       open (unit=9,file='met.tmp',status='UNKNOWN')
C       write(9,*) '*** avancees de met ***'
C       write(9,*) i, ' ',j
C       close(9)
C       -----
C       U(i,j) = L(i,j) + random(graine)*(U(i,j)-L(i,j))
C       U(j,i) = U(i,j)
C       L(i,j) = U(i,j)
C       L(j,i) = L(i,j)
C       if (methode.EQ.'trian') Call triangle(Leading,U,L,N)
C       if (methode.EQ.'tetra') Call tetra(Leading,U,L,N)
C     enddo
C   enddo
C else
C
C Type_met = 2,3,4 et + <=N
C @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
C Generateur de parcours aleatoire de i et j
C Les differents couples (i,j) sont stockes dans Pos
C
C   Call Alea(Leading,N,Pos,graine)
C   co = ((N-1)*N/2)
C   if (Type_met.EQ.3) co = int(p_cent*(N-1)*N/200)
C
C Essais de metrisation /p atomes [NE SERT QUE SI Type_met >=4 ]
C (est-il preferable de changer a chaque fois le groupe de p atomes?)
C Construction d'un tableau X contenant Type_met indices d'atomes distincts
C (sert pour les methodes du type 'Four-atoms Metrization')
C
C   Call Alea_smp(Leading,N,Type_met,X,graine) ! change les p atomes
C
C
C Metrisation aleatoire totale(2) ou partielle(3) a p_cent%
C
C   do i = 1,co
C     ----- stocke avance du calcul -----
C     open (unit=9,file='met.tmp',status='UNKNOWN')
C     write(9,*) '*** avancees de met ***'
C     write(9,*) (i*100/co),' % effectuees'
C     close(9)
C     -----
C     U(Pos(1,i),Pos(2,i)) = L(Pos(1,i),Pos(2,i)) +
C     + random(graine)*(U(Pos(1,i),Pos(2,i))-L(Pos(1,i),Pos(2,i)))
C     U(Pos(2,i),Pos(1,i)) = U(Pos(1,i),Pos(2,i))
C     L(Pos(1,i),Pos(2,i)) = U(Pos(1,i),Pos(2,i))
C     L(Pos(2,i),Pos(1,i)) = L(Pos(1,i),Pos(2,i))
C
C     if (Type_met.LT.4) then !Type_met = 2,3
C
C Repropager les inegalites sur la totalite des matrices U et L
C (2 lignes ci-dessous) serait inutile:
C   if (methode.EQ.'trian') Call triangle(Leading,U,L,N)
C   if (methode.EQ.'tetra') Call tetra(Leading,U,L,N)
C Les inegalites ne sont repropagees que sur la ligne et la colonne
C dont la valeur vient d'etre fixee ( Pos(1,i),Pos(2,i) )
C d'ou: (gain serait considerable, on gagne 1 ordre N)
C Pb: essais peu concluants avec le fichier helice.dat

```



```

C
C      if (methode.EQ.'trian')
C      #      Call subtria(Leading,U,L,N,Pos(1,i),Pos(2,i))
C
C      if (methode.EQ.'tetra')
C      #      Call subtetra(Leading,U,L,N,Pos(1,i),Pos(2,i))
C
C      #*      else      !Type_met sert maintenant de nombre d'atomes par
C                      rapport auxquels on propage les inegalites
C
C Metrisation partielle: "four atoms metrization"
C voir "p atoms metrization"
C
C      if (methode.EQ.'trian')
C      #      Call triangle_4(Leading,U,L,N,X,Type_met)
C      if (methode.EQ.'tetra')
C      #      Call tetra_4(Leading,U,L,N,X,Type_met)
C      end if
C
C      enddo
C
C      end if
C
C @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
C Annulation de tous les elements diagonaux
C et remplissage de la matrice de distance (met)
C do i = 1,N
C   U(i,i) = 0.0
C   L(i,i) = 0.0
C enddo
C
C Return
C End

```

6. Le plongement :

```

C 7 file embed.f
C etape "d'embedding" ref: Acta Cryst. (1978). A34,282-284
C passage d'une matrice de distance Dij NxN a une matrice
C de coordonnees 3xN
C
C Subroutine embed(Dist,Leading,N,Xout,Liw,Lk)
C
C implicit none
C
C integer i,j,Leading,N
C real Xout(Leading,3),DD2,DD1(N)
C real Dist(Leading,Leading),Dist2(N,N)
C
C external evx
C -----> lignes pour sbroutine sseyevx
C integer Liw,Lk
C integer NBEV,Iw(Liw),Ifail(N),info ! c.f. <Call sseyevx(...)>
C real Wval(N),Zmat(N,N),Wk(Lk) !
C sseyevx fin de declaration <-----
C
C calcul la matrice des distances au carre
C
C do j=1,N
C   DD1(j) = 0.0
C   do i=1,N
C     Dist2(i,j) = Dist(i,j)*Dist(i,j)
C   enddo
C enddo
C
C calcule la somme pour i<j des Dist2(j,i) divisee par N^2
C
C do j=2,N
C   do i=1,j-1
C     DD2 = DD2 + Dist2(j,i)
C   enddo
C enddo
C DD2 = DD2 / (N*N)
C
C construit le vecteur DD1 :
C DD1(i) = somme sur j de Dist2(i,j) divisee par N - DD2
C [distance de i au centre de masse]
C
C do i=1,N
C   do j=1,N
C     DD1(i) = DD1(i) + Dist2(i,j)
C   enddo
C   DD1(i) = (DD1(i)/N) - DD2
C enddo
C
C Reconstruction de la matrice metrique en utilisant le
C centre de masse O.
C Dist2ij = 1/2 (diO^2 + djO^2 - dij^2)
C

```

```

do i=1,N
  do j=1,N
    Dist2(i,j) = 0.5 * (DD1(i)+DD1(j)-Dist2(i,j))
  enddo
enddo

C Calcul des valeurs propres et vecteurs propres de Dist2
C [utilisation de la bibliothèque mathématique LAPACK]
C sous-routine sseyevx, Real dble prec. sym. eig. values & vectors
C #
  Call sseyevx('V', 'A', 'U', N, Dist2, N, 0.0, 0.0, 0.0, 0.0, 0.0,
    $      0., NBEV, Wval, Zmat, N, Wk, Lk, Iw, Ifail, info)

C 'V' option pour calcul eig.val. et vectors
C 'A' recherche de toutes les eig.val.
C 'U' indique de travailler avec le triangle sup.
C      (indifferent ici car symétrique)
C N nbre de rangs et col. dans Dist2
C Dist2 matrice a diagonaliser
C N Leading dimension
C 0.0 4 valeurs indifferentes vu l'option 'A'
C 0. relatif a la convergence dans le calcul des eig.val.
C      (Le choix effectue correspond a une precision
C      appropriée pour la majorite des calculs;
C      c.f. Manuel Lapack sur SSYEVX)
C NBEV Nombre de Valeurs trouvees
C Wval Les eig.val. stockees par ordre croissant
C Zmat Les eig.vectors respectifs
C N leading dimension de Zmat
C Wk tableau utiliser pour obtenir une performance optimale
C Lk entier correspondant a la taille de Wk ( >=8*N )
C Liw entier correspondant a la taille de Iw ( 5*N )
C Iw espace de travail
C Ifail Stocke les indices des eig.vect. qui n'ont pas converge
C info =0 : calcul reussi
C      <0 : info = -i, le i ieme argument a une valeur illegale
C      >0 : info = +i, le i eig.vect. n'a pas converge.
C

```

```

C Calcul des coordonnees X Y Z (stockee dans Xout(N,3))
C

```

```

do i=1,N
  Xout(i,1)=sqrt(abs(Wval(N)))*Zmat(i,N)
  Xout(i,2)=sqrt(abs(Wval(N-1)))*Zmat(i,N-1)
  Xout(i,3)=sqrt(abs(Wval(N-2)))*Zmat(i,N-2)
enddo

```

```

Return
End

```

7. La minimisation :

```

C The majorization algorithm (file majalg.f)
C J.of Computational chemistry, Vol14, (1),p.114-120(1993)
C
C 7

```

```

C - - - - - sub routine init_VW - - - - -
C      INITIALISE V (et calcule son inverse)
C      INITIALISE W (poids donne a chaque ecart)
C - - - - -
subroutine init_VW(Vinv,W,Upp,Low,Leading,N)

  implicit none

  integer Leading,i,j,info,k,N
  real W(Leading,Leading),Upp(Leading,Leading)
  real Wk(Leading),s,Low(Leading,Leading),Vinv(Leading,Leading)
  integer ipiv(N) ! contient detail des permutations

  do i=1,N-1
C construction de la diagonale de W
    W(i,i) = 1.0

    do j=i+1,N

C construction de W (hors diagonale)
      W(i,j) = 1/(1+(10*(Upp(i,j)-Low(i,j))))
      W(j,i) = W(i,j)

C construction de V (hors diagonale); ici Vinv = V
      Vinv(i,j) = -W(i,j) ! sera inverse + tard
      Vinv(j,i) = Vinv(i,j)
    enddo
  enddo

C ajout de W(N,N)
  W(N,N) = 1.0

C construction de la diagonale de V; ici Vinv = V
  do i = 1,N
    s=0

```

```

do k=1,N
  s = s+W(i,k)
enddo
Vinv(i,i) = s - W(i,i)      ! sera inverse + tard
enddo

C calcul de l'inverse de V: Vinv
C l'inverse est stocke dans la matrice en entree
C [c.f. + de details ds bibliotheque LAPACK]
C
  write(*,*) 'calcul inverse de V'
c sauvegarde des donnees et inversion avec matlab...
c
c   open(unit=13,FILE='Vinv.dat',STATUS='UNKNOWN')
c   do i=1,N
c     write(13,10) (Vinv(i,j),j=1,N)
c   enddo
c La matrice Vinv.dat est inversible avec matlab sans aucun probleme!!!
c
c   close(13)
  Call msytrf('U',N,Vinv,Leading,ipiv,Wk,Leading,info)
  Call msytri('U',N,Vinv,Leading,ipiv,Wk,info)
do i=1,N
  do j=1,i
    Vinv(i,j) = Vinv(j,i)
  enddo
enddo
write(*,*) 'O.K. inverse calcule'
C
  STOP

10 Format(2000 F15.6)
  Return
end

-----
C dist: calcul la matrice de distance associee a Xin
C
  subroutines dist(Xin,Leading,N,Dout)

  implicit none

  integer Leading,i,j,N
  real Xin(Leading,3)
  real Dout(Leading,Leading)

  do i=1,N-1
    do j=i+1,N
      Dout(i,j) = sqrt((Xin(j,1)-Xin(i,1))**2 +
+ (Xin(j,2)-Xin(i,2))**2 + (Xin(j,3)-Xin(i,3))**2)
      Dout(j,i) = Dout(i,j)
    enddo
    Dout(i,i) = 0.0
  enddo
  Dout(N,N) = 0.0

  Return
End

-----
C Calcul de la valeur de la somme des ecart (-Energie)
C
  Real function Energie(Xin,Upp,Low,Lead,N,Vinv,W)

  implicit none

  integer i,j,N,Lead
  real Upp(Lead,Lead),Low(Lead,Lead),Vinv(Lead,Lead)
  real W(Lead,Lead),DD(Lead,Lead),diss(Lead,Lead)
  real Xin(Lead,3)

C calcul de la matrice de distance associee a Xin
  Call dist(Xin,Lead,N,DD)

C mise a zero de l'Energie
  Energie=0.0

C boucle sur i et j (i<j)
  do i=1,N-1
    do j=i+1,N

C calcul de la matrice des "dissimilarities"
C travail effectue dans le triangle superieur (i<j)
C [cette matrice ne contient pas de violation de Upp ou Low]
C
      if (DD(i,j).LT.Low(i,j)) diss(i,j)=Low(i,j)
      if (DD(i,j).GT.Upp(i,j)) diss(i,j)=Upp(i,j)
      if ((DD(i,j).GE.Low(i,j)).AND.(DD(i,j).LE.Upp(i,j)))
$         diss(i,j) = DD(i,j)

C calcul de l'Energie
C E = (1/2) * Somme(Wij(Dij-Dissij)^2)
C comme on est dans le triangle superieur, le facteur (1/2) n'est

```

```

C pas a mettre, on a directement l'Energie
  Energie = Energie + W(i,j)*(DD(i,j)-diss(i,j)**2)
  enddo
enddo
Return
End
-----
C
C SubRoutine majalg(Yout,Xin,Vinv,W,Low,Upp,Leading,N)

implicit none

integer Leading,N,i,j,k,p
real Yout(Leading,3),Xin(Leading,3),s,Sk
real Vinv(Leading,Leading),W(Leading,Leading)
real Low(Leading,Leading),Upp(Leading,Leading)
real DD(Leading,Leading),diss(Leading,Leading)
real B(Leading,Leading)

C calcul de la matrice de distance associee a Xin
Call dist(Xin,Leading,N,DD)

do i=1,N-1
  do j=i+1,N

C calcul de la matrice des "dissimilarities"
C travail effectue dans le triangle superieur (i<j)
C [cette matrice ne contient pas de violation de Upp ou Low]
C
  if (DD(i,j).LT.Low(i,j)) diss(i,j)=Low(i,j)
  if (DD(i,j).GT.Upp(i,j)) diss(i,j)=Upp(i,j)
  if ((DD(i,j).GE.Low(i,j)).AND.(DD(i,j).LE.Upp(i,j)))
    $   diss(i,j) = DD(i,j)

C construction de la "B box" hors diagonale
C
  if (DD(i,j).EQ.0) then
    B(i,j) = 0
  else
    B(i,j) = -W(i,j)*diss(i,j)/DD(i,j)   !### Bij
  end if
  B(j,i)=B(i,j)
enddo
enddo

C construction de la diagonale de la "B box"
C
do i=1,N
  s=0
  do k=1,N
    s = s-B(i,k)   !### Bij = -scome(Bik) k#i
  enddo
  B(i,i) = s + B(i,i)
enddo

C Yout = Vinv*B*Xin; !##### CALCUL DE Yout
C # produit de 3 matrices
do i=1,N
  do j=1,3
    Yout(i,j) = 0.0
    do p=1,N
      Sk = 0.0
      do k=1,N
        Sk = Sk + B(p,k)*Xin(k,j)
      enddo
      Yout(i,j) = Yout(i,j) + Vinv(i,p)*Sk
    enddo
  enddo
enddo

Return
End
-----
C----- Main : sb-routine maj -----
C
SubRoutine maj(Xout,Xin,Upp,Low,Lead,N)

implicit none

real Xin(Lead,3),Xout(Lead,3),Xmin(Lead,3)
integer Lead,N,nstep,step,erreur,Energie,i,j,ffst
real sig0,sig1,W(Lead,Lead),DO(Lead,Lead)
real Upp(Lead,Lead),Low(Lead,Lead),Vinv(Lead,Lead)
parameter(erreur = 1.0E-05,nstep=2000)

COMMON /FIRST/ ffst
c calcul des matrices W et Vinv [inverse(V)]
c W matrice des poids donnees aux ecart par rapp. aux contraintes

```

```

c Vinv: "Moore-Penrose inverse of V"
  if (ffst.NE.1) Call init_VW(Vinv,W,Upp,Low,Lead,N)

c Calcul de l'Energie initiale
  sig0 = Energie(Xin,Upp,Low,Lead,N,Vinv,W)

c leq,pas et Calcul de son Energie
  Call majalg(Xout,Xin,Vinv,W,Low,Upp,Lead,N)
  sig1 = Energie(Xout,Upp,Low,Lead,N,Vinv,W)
  step = 1

c etapes de minimisation (limite a nstep etapes)
  do while (abs(sig0-sig1) .GT. erreur)
    sig0= sig1
    do i=1,N
      do j=1,3
        Xin(i,j) = Xout(i,j)
      enddo
    enddo

    Call majalg(Xout,Xin,Vinv,W,Low,Upp,Lead,N)
    sig1 = Energie(Xout,Upp,Low,Lead,N,Vinv,W)

    if (sig1.LT.sig0) then ! sauvegarde du meilleur Xout
      do i=1,N
        do j=1,3
          Xmin(i,j) = Xout(i,j)
        enddo
      enddo
    end if

    step = step+1
    if (step.GE.nstep) sig0=sig1 ! limite a nstep pas de calcul
  enddo

C recuperation de la meilleure structure
  do i=1,N
    do j=1,3
      Xout(i,j) = Xmin(i,j)
    enddo
  enddo

  ffst = 1 ! =0 lors du premier passage dans la subroutine maj
           ! -> permet de ne calculer qu'une seule fois l'inverse
           ! de V

  Return
  End
-----End Of File

```

8. L'inversion de matrice :

```

C calcul de l'inverse d'une matrice
C file inversmat.f
  subroutine inver(Mat,Lead,N)

  implicit none

  integer Lead,info,N,i,j
  integer ipiv(N)
  real Wk(Lead),Mat(Lead,Lead)
  character*1 Uplo

c inverse calcule via la bibliotheque LAPACK
  Uplo = 'U'
  Call sgetrf(Uplo,N,Mat,Lead,ipiv,Wk,Lead,info)
  if (info.EQ.0) write(*,*) 'Successfull triangulation'
  Call sgetri('U',N,Mat,Lead,ipiv,Wk,info)
  if (info.EQ.0) write(*,*) 'Successfull inversion'
  do i=1,N
    do j=1,i
      Mat(i,j) = Mat(j,i)
    enddo
  enddo
  Return
  End

C ----- MAIN -----
PROGRAM dbinv

  implicit none

  integer Leading,N,i,j,k,ilaenv
  parameter ( Leading = 200 )
  real Matrice(Leading,Leading),Ml(Leading,Leading)
  real Produit(Leading,Leading)
  parameter ( N = 4 )
  external ilaenv

  do i=1,N
    do j=i,N

```

```

Matrice(i,j) = j
Matrice(j,i) = Matrice(i,j)
M1(i,j) = j
M1(j,i) = Matrice(i,j)
enddo
enddo

* do i=1,N
  write(*,*) (Matrice(i,j),j=1,N)
enddo
write(*,*)

Call inver(Matrice,Leading,N)

do i=1,N
  write(*,*) (Matrice(i,j),j=1,N)
enddo
write(*,*)

do i=1,N
  do j=1,N
    Produit(i,j) = 0
    do k=1,N
      Produit(i,j) = Produit(i,j) + Matrice(i,k) * M1(k,j)
    enddo
  enddo
enddo

do i=1,N
  write(*,*) (Produit(i,j),j=1,N)
enddo

Stop *
End

```

9. Les tirages aléatoires :

a) Le fichier random.f :

```

C This file is a part of the GIFA program
C This software has been developed by the NMR Group in GIF/Yvette.
C
C Authors :      M.A.Delsuc
C              MMR Laboratory,
C              ICSN - CNRS
C              91190 GIF/YVETTE  FRANCE
C
C This software cannot be used unless the user have been fully
C licensed to do so form the above laboratory.
C This file cannot be copied, duplicated or used in any other program,
C without written permission from the authors.
C
C function random(idum)
C
C return a uniform random deviate between 0 and 1
C form Num. Recipes p197
C
C implicit none
C real random,rm
C integer ir(97),iff,iy,j,idum,m,ia,ic
C parameter (m=714025,ia=1366,ic=150889,rm=1.0/m)
C save
C save ir,iff,iy,j
C data iff /0/
C
C initialize if idum is neg, or if it is the first time
C
C if (idum.lt.0.or.iff.eq.0) then
C   iff=1
C   idum=mod(ic-idum,m)
C   do j=1,97
C     idum=mod(ia*idum+ic,m)
C     ir(j)=idum
C   enddo
C   idum=mod(ia*idum+ic,m)
C   iy=idum
C endif
C j=1+(97*iy)*rm
C if (j.gt.97.or.j.lt.1)pause
C iy=ir(j)
C random=iy*rm
C idum=mod(ia*idum+ic,m)
C ir(j)=idum
C return
C end
C
C .....
C function gasdev(idum)
C
C this function returns a normally distributed deviate with 0 mean and unit
C variance, usins RANDOM(IDUM)
C form Num. Recipes p203

```

```

implicit none
real gasdev,v1,v2,r,fac,gset,random
external random
integer idum,iset
save
c save iset,gset
data iset/0/
if (iset.eq.0) then
1 v1=2.0*random(idum)-1.0
  v2=2.0*random(idum)-1.0
  R=v1**2+v2**2
  if (R.ge.1.0) goto 1
  fac=sqrt(-2.0*log(r)/r)
  gset=v1*fac
  gasdev=v2*fac
  iset=1
else
  gasdev=gset
  iset=0
endif
return
end

```

b) Deuxième fichier :

```

C 7 file alea.f
C Remplissage d'un tableau ATab(2,N*(N-1)/2) contenant un parcours
C aleatoire des couples (i,j) avec i=1:N & j=1:N
C

```

```
SubRoutine Alea(Nmax,N,Atab,graine)
```

```
implicit none
```

```
integer Nmax, N, co, i, j, xa, Id, dim, graine
real random
```

```
external random
```

```

C Declaration d'un tableau pour les indices ordonnées
integer Ord(2, (Nmax*(Nmax-1)/2)), ATab(2, (Nmax*(Nmax-1)/2))

```

```
dim = N*(N-1)/2
Id = dim
co = 0
```

```

do i=1,N-1
  do j=i+1,N
    co = co + 1
    C remplissage du tableau ordonne
    Ord(1,co) = i
    Ord(2,co) = j
  enddo
enddo

```

```

do i = 1,dim
  C valeur choisie entre 1 et Id pseudo-aleatoire
  xa = int(Id * random(graine) + 1)
  C remplissage du tableau desordonne
  ATab(1,i) = Ord(1,xa)
  ATab(2,i) = Ord(2,xa)
  if (xa.NE.Id) then
    do j = xa,Id-1
      Ord(1,j) = Ord(1,j+1)
      Ord(2,j) = Ord(2,j+1)
    enddo
  end if
  Id = Id - 1
enddo
Return
end

```

c) Troisième fichier :

```

C 7 file alea_simpl.f
C Remplissage d'un tableau ATab(N) contenant un parcours
C aleatoire des de l'indice i avec i=1:N
C

```

```
SubRoutine Alea_simpl(Nmax,N,p,Atab,graine)
```

```
implicit none
```

```
integer Nmax, N, i, j, xa, Id, graine, p
real random
```

```
external random
```

```

C   Declaration d'un tableau pour les indices ordonnees
integer Ord(Nmax), Atab(Nmax)

Id = N

do i=1,N
C   #1      remplissage du tableau ordonne
  Ord(i) = 1
enddo

do i = 1,p
C           valeur choisie entre 1 et Id pseudo-aleatoire
  xa = int(Id * randomd(graine) + 1)
C           remplissage du tableau desordonne
  Atab(i) = Ord(xa)
  Atab(i) = Ord(xa)
  if (xa.NE.Id) then
    do j = xa,Id-1
      Ord(j) = Ord(j+1)
      Ord(j) = Ord(j+1)
    enddo
  end if
  Id = Id - 1
enddo
Return
end

```

10, Autres parties :

a) triangle 4.f :

```

C   7 file triangle_4.f
C   Propagation de l'inegalite du triangle sur les matrices de contraintes
C   de distance U(pper) et L(ower), par rapport a p atomes
C

```

```

SUBROUTINE triangle_4(Leading,U,L,N,X,p)

implicit none

integer Leading,N,i,j,k,p
integer X(p) ! contient les indices des p atomes choisis (aleatoire)
real U(Leading,Leading),L(Leading,Leading)

do k=1,p
do i=1,N-1      ! => metrisation devient en p*N^4 (N^5 maxi)
  do j=i+1,N    ! j ... = p atomes choisis
    U(i,j) = min(U(i,j),U(i,X(k))+U(X(k),j))
    U(j,i) = U(i,j)
    L(i,j) = max(L(i,X(k))-U(X(k),j),L(j,X(k))-U(X(k),i),L(i,j))
    L(j,i) = L(i,j)
  enddo
enddo

return
END

```

b) subtria.f :

```

C   7 file subtria.f
C   Propagation partielle de l'inegalite du triangle sur les matrices
C   de contraintes de distance U(pper) et L(ower)
C   -> sert pour la metrisation
C

```

```

SUBROUTINE subtria(Leading,U,L,N,a,b)

implicit none

integer Leading,N,i,j,k,a,b
real U(Leading,Leading) , L(Leading,Leading)

do k=1,N
  i=a
  do j=a+1,N
    U(i,j) = min(U(i,j),U(i,k)+U(k,j))
    U(j,i) = U(i,j)
    L(i,j) = max(L(i,k)-U(k,j),L(j,k)-U(k,i),L(i,j))
    L(j,i) = L(i,j)
  enddo
  j=b
  do i=1,b-1
    U(i,j) = min(U(i,j),U(i,k)+U(k,j))
    U(j,i) = U(i,j)
    L(i,j) = max(L(i,k)-U(k,j),L(j,k)-U(k,i),L(i,j))
    L(j,i) = L(i,j)
  enddo
enddo
return

```



```

END
c) subtetra.f
C 7 file subtetra.f
C Propagation de l'inegalite du tetraangle sur les matrices de contraintes
C de distance U(ppper) et L(ower)
C
SUBROUTINE subtetra(Leading,U,L,N,a,b)
implicit none

integer Leading,N,i,j,k,p,a,b
real U(Leading,Leading) , L(Leading,Leading)
real Um,Im
real u1,u2,u3,u4,u5,u6,u7,l1,l2,l3,l4,l5,l6,l7
external calmeng

C propagation des inegalites du triangle
CALL subtria(Leading,U,L,N,a,b)

C propagation des inegalites du tetraangle
do i=1,N-1
do j=i+1,N

C ----- boucle sur p -----
k=a
do p=a+1,N

C tests sur i,j,k,p pour avoir des quadruplets distincts
C et si L(k,p) = U(k,p) alors on ne fait rien
C
if ((k.NE.j).AND.(k.NE.i).AND.(p.NE.j).AND.(p.NE.i).AND.
& .(p.NE.b).AND.(L(k,p).LT.U(k,p))) then

C calcul des 7 determinants de Caley-Menger relatifs au tetraangle
C Bull. of Mathematical Biology, Vol 51, (1),p173-194, (1989)
C
CALL calmeng(u1,l1,L(i,j),U(i,k),U(i,p),U(j,k),U(j,p))
CALL calmeng(u2,l2,U(i,j),L(i,k),L(i,p),U(j,k),U(j,p))
CALL calmeng(u3,l3,U(i,j),U(i,k),U(i,p),L(j,k),L(j,p))
CALL calmeng(u4,l4,U(i,j),U(i,k),L(i,p),L(j,k),U(j,p))
CALL calmeng(u5,l5,U(i,j),L(i,k),U(i,p),U(j,k),L(j,p))
CALL calmeng(u6,l6,L(i,j),L(i,k),U(i,p),L(j,k),U(j,p))
CALL calmeng(u7,l7,L(i,j),U(i,k),L(i,p),U(j,k),L(j,p))

C recuperation des bornes calculees (Um > Lm)
Um= max(u1,u2,u3,u4,u5,u6,u7)
Lm= min(l1,l2,l3,l4,l5,l6,l7)

C Tests pour assurer une geometrie consistente
C symetrisation des matrices U et L
if ((Im.GE.0).and.(sqrt(Im).LE.U(k,p))
& .and.(sqrt(Um).GE.L(k,p))) then
U(k,p)= min(U(k,p),sqrt(Um))
U(p,k)= U(k,p)
L(k,p)= max(L(k,p),sqrt(Im))
L(p,k)= L(k,p)
end if
end if

C ----- fin de boucle sur p -----
enddo

C ----- boucle sur k -----
p=b
do k=1,p-1

C tests sur i,j,k,p pour avoir des quadruplets distincts
C et si L(k,p) = U(k,p) alors on ne fait rien
C
if ((k.NE.j).AND.(k.NE.i).AND.(p.NE.j).AND.(p.NE.i).AND.
& (k.NE.a).AND.(L(k,p).LT.U(k,p))) then

C calcul des 7 determinants de Caley-Menger relatifs au tetraangle
C Bull. of Mathematical Biology, Vol 51, (1),p173-194, (1989)
C
CALL calmeng(u1,l1,L(i,j),U(i,k),U(i,p),U(j,k),U(j,p))
CALL calmeng(u2,l2,U(i,j),L(i,k),L(i,p),U(j,k),U(j,p))
CALL calmeng(u3,l3,U(i,j),U(i,k),U(i,p),L(j,k),L(j,p))
CALL calmeng(u4,l4,U(i,j),U(i,k),L(i,p),L(j,k),U(j,p))
CALL calmeng(u5,l5,U(i,j),L(i,k),U(i,p),U(j,k),L(j,p))
CALL calmeng(u6,l6,L(i,j),L(i,k),U(i,p),L(j,k),U(j,p))
CALL calmeng(u7,l7,L(i,j),U(i,k),L(i,p),U(j,k),L(j,p))

C recuperation des bornes calculees (Um > Lm)
Um= max(u1,u2,u3,u4,u5,u6,u7)
Lm= min(l1,l2,l3,l4,l5,l6,l7)

C Tests pour assurer une geometrie consistente
C symetrisation des matrices U et L
if ((Im.GE.0).and.(sqrt(Im).LE.U(k,p))

```

```

      .and. (sqrt(Um).GE.L(k,p)) then
    U(k,p)= min(U(k,p),sqrt(Um))
    U(p,k)= U(k,p)
    L(k,p)= max(L(k,p),sqrt(Lm))
    L(p,k)= L(k,p)
  end if
end if
C ----- fin de boucle sur k -----
  enddo
enddo
enddo

C Propagation des inegalites du triangle...
C Cas ou certains U(k,p) n'ont pu etre traites par
C la propagation des inegalites du tetraedre
C (NB: test : Cette position a donne les meilleurs
C au niveau des violations sur les resultats)
C
  CALL subtria(Leading,U,L,N,a,b)

  return
END
```

1987-1988
1988-1989
1989-1990
1990-1991
1991-1992

1992-1993
1993-1994
1994-1995
1995-1996
1996-1997

1997-1998
1998-1999
1999-2000
2000-2001
2001-2002

2002-2003
2003-2004
2004-2005
2005-2006
2006-2007

2007-2008
2008-2009
2009-2010
2010-2011
2011-2012

2012-2013
2013-2014
2014-2015
2015-2016
2016-2017

2017-2018
2018-2019
2019-2020
2020-2021
2021-2022

2022-2023
2023-2024
2024-2025
2025-2026
2026-2027

2027-2028
2028-2029
2029-2030
2030-2031
2031-2032

2032-2033
2033-2034
2034-2035
2035-2036
2036-2037

2037-2038
2038-2039
2039-2040
2040-2041
2041-2042

2042-2043
2043-2044
2044-2045
2045-2046
2046-2047

2047-2048
2048-2049
2049-2050
2050-2051
2051-2052

2052-2053
2053-2054
2054-2055
2055-2056
2056-2057

2057-2058
2058-2059
2059-2060
2060-2061
2061-2062

2062-2063
2063-2064
2064-2065
2065-2066
2066-2067

2067-2068
2068-2069
2069-2070
2070-2071
2071-2072

2072-2073
2073-2074
2074-2075
2075-2076
2076-2077

2077-2078
2078-2079
2079-2080
2080-2081
2081-2082

2082-2083
2083-2084
2084-2085
2085-2086
2086-2087

2087-2088
2088-2089
2089-2090
2090-2091
2091-2092

2092-2093
2093-2094
2094-2095
2095-2096
2096-2097

D. Aspects généraux en biologie :

1. La préparation des macromolécules biologiques :

Etudier une protéine, c'est aussi savoir la produire, la purifier et savoir l'analyser ; entre autre connaître l'enchaînement des acides aminés et son activité est un atout. La purification se fait donc en utilisant les propriétés spécifiques à chaque protéine du milieu au sein duquel se trouve la protéine recherchée. On peut jouer sur leur solubilité, leur charge, leur taille ou d'autres caractéristiques intrinsèques.

a) Les outils courants pour la purification des protéines :

i) Les méthodes dénaturantes :

Ce sont des méthodes qui détruisent les interactions qui maintiennent la structure des protéines, celles-ci ayant alors tendance à s'agglomérer et à précipiter. Ce sont en général des méthodes avec lesquelles il faut être prudent car la dénaturation est irréversible et, dans ce cas, la purification s'arrête là. Il est donc préférable de tester ces méthodes sur des aliquotes de ce que l'on désire purifier.

La température :

C'est bien souvent un paramètre physique employé pour séparer les protéines. En effet une dénaturation thermique détruit des interactions non covalentes, celles qui maintiennent la structure tridimensionnelle des protéines. Les parties insolubles de la protéine se trouvent alors exposées au solvant et cela provoque leur précipitation. Cette méthode est couramment utilisée pour séparer des protéines. Lorsque nous avons employé cette technique pour la purification, le contrôle de la température a été suivi de façon attentive pour éviter les désagréments d'une précipitation. (Voir Le choc thermique page 110)

La solubilité :

Il existe des sels comme $(\text{NH}_4)_2\text{SO}_4$ qui sont capables de fixer des molécules d'eau en plus en moins grande quantité. Ainsi une protéine qui nécessite plus de molécules d'eau qu'une autre pour rester soluble va précipiter plus tôt. Il est donc possible de les séparer de cette façon par précipitation. Cette méthode, tout comme la précédente, est relativement efficace. Néanmoins il n'est pas rare d'observer la protéine qui devait rester soluble dans le précipité, simplement parce qu'elle y a été entraînée par d'autres. Ceci a été couramment observé lors de la préparation de nos échantillons (voir page 110).

La charge :

La charge d'une protéine dépend de son point isoélectrique pI, il s'agit du pH pour lequel la protéine n'est globalement pas chargée, et en général, c'est le pH auquel la protéine précipite puisqu'à ce moment les protéines non chargées peuvent s'agglomérer et donc devenir insolubles. Cette propriété peut être employée pour séparer des protéines en modifiant simplement le pH de la solution. Si on compare 2 protéines P1 et P2 constituées des mêmes acides aminés, elles n'ont pas forcément le même point isoélectrique. Cette différence est due aux groupements ionisables et à la différence dans leur séquence, les groupements ionisables ont un pKa différent suivant leur environnement spatial dans la structure. Ainsi le pI de deux protéines, même constituées des mêmes acides aminés, peut permettre de les séparer.

L'électrophorèse :

Elle est basée sur la différence de mobilité des protéines sous l'influence d'un champ électrique. La mobilité va dépendre du pH, de la différence de potentiel et d'un coefficient de friction lié à la forme de la molécule.

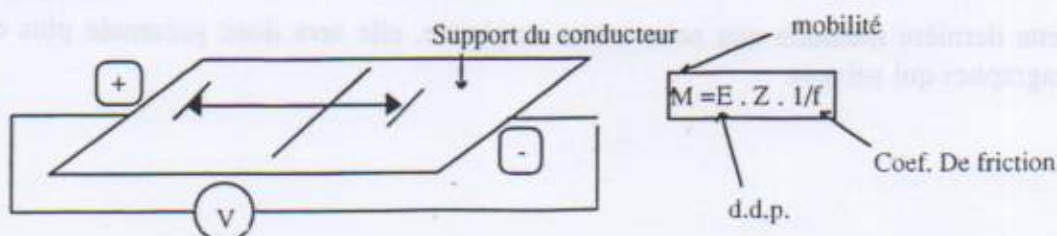


Figure 61 : Représentation schématique de l'électrophorèse.

Le support conducteur est un polymère imprégné d'un tampon (gel d'agarose ou de polyacrylamide - voir page 135). Il est possible de séparer les molécules en fonction de leur poids moléculaire en utilisant cette méthode. Le problème de cette technique est qu'elle plutôt analytique, elle permet une séparation très fine mais sur de toutes petites quantités. C'est une méthode de routine pour identifier des protéines de masses moléculaires différentes dans un même échantillon.

Une variante de ce principe (l'électrofocalisation) consiste à avoir un gradient de pH sur le support (obtenu grâce à des ampholytes), chaque molécule migre alors jusqu'à ce qu'elle atteigne le pH correspondant à son pI.

ii) Les méthodes non dénaturantes :

La chromatographie d'échange d'ions :

On utilise une colonne dont le support baigne dans une phase mobile (le tampon) et on étudie la migration d'une substance dans un tel système. Le support est généralement un polymère où sont greffés des groupement ionisables (polydextrane, polystyrène, polyacrylamide, cellulose). Il en existe deux types : les supports chargés positivement et ceux chargés négativement respectivement échangeurs d'anions et de cations. A un pH donné, certaines protéines en solution seront chargées positivement ou négativement suivant leur pI et seront alors retenues ou non dans la colonne. Pour récupérer progressivement les protéines en interaction avec la colonne, on fait varier progressivement soit le pH de la solution d'éluion ou soit sa force ionique avec un gradient de sel (typiquement NaCl ou KCl). L'intérêt de cette méthode est qu'elle est quantitative et permet une préparation à grande échelle des protéines.

La chromatographie d'exclusion ou filtration sur gel :

Dans cette méthode, on sépare les macromolécules suivant leur taille. On dispose d'une colonne et d'un support constitué d'un tamis moléculaire. Ce dernier est formé de petites billes extrêmement poreuse laissant passer à travers elles des molécules de taille inférieure à celle des pores. Les petites molécules qui circulent dans tous les interstices parcourent un chemin plus long à travers les billes que les molécules plus grosses qui sont plus vite éluées en bas de la colonne.

2. La biochimie structurale ou l'étude de structure des molécules biologiques :

En biochimie structurale, le choix d'une méthode plutôt qu'une autre pour déterminer une structure se fait en fonction de la taille de la molécule que l'on désire étudier, mais aussi en fonction des diverses étapes limitantes rencontrées au cours de la préparation et l'obtention d'un échantillon de cette molécule à l'état pur.

Il faut savoir qu'il existe différentes outils pour l'étude structurale des molécules biologiques, ne fournissant pas les mêmes informations :

- Le dichroïsme circulaire. [Singer, 1985]
- La spectrométrie de masse. [Siuzdak, 1994]
- la diffusion quasi-élastique de la lumière. [Spurej, 1992]
- l'anisotropie de fluorescence. [Castanho, 1992]

Méthodes donnant des informations sur la position spatiale des atomes :

- la microscopie électronique ou de champ local. [Le Cam, 1994]
- La cristallographie par diffraction des rayons X. [Wagner, 1995]
- la résonance magnétique nucléaire (RMN). [Hill, 1996]

C'est cette dernière méthode que nous avons employée, elle sera donc présentée plus en détails dans les paragraphes qui suivent.

E. Le gel d'électrophorèse SDS Page :**Le gel de Séparation :**

pour 4 gels (30 ml)	10%	12.5%	15%	17.5%	20%
Sol A ; Acryl Bis 30 :0.8	10 ml	12.5 ml	15 ml	17.5 ml	20 ml
Sol B Tris pH=8	7.5 ml	7.5 ml	7.5 ml	7.5 ml	7.5 ml
H ₂ O	10.4 ml	7.9 ml	5.4 ml	2.9 ml	0.4ml
Sol C 10% SDS	0.6 ml				
Teemed	20 µl ou plus				
persulfate 10mg/ml	1.5 ml				

Le gel de concentration :

pour 4 gels (30 ml)	10%	12.5%	15%	17.5%	20%
Sol A ; Acryl Bis 30 :0.8	2.5 ml				
Sol D Tris pH=6.8	3.75 ml				
H ₂ O	7.85 ml				
Sol C 10% SDS	0.15 ml				
Teemed	10 µl ou plus				
persulfate 10mg/ml	0.75 ml				

La solution A :

- Acrylamide 300g
- bis acrylamide 8g
- compléter à 1 litre avec de l'eau milliQ.

La solution C :

- 5g de SDS
- Ajuster à 50 ml d'eau.

La solution B :

- Tris 90.8g
- Ajouter 350ml d'eau.
- Ajuster le pH à 8.8 avec de l'HCl 6N
- compléter à 500 ml

La solution D :

- Tris 6.05g
- Ajouter 90ml d'eau
- Ajuster le pH à 6.8 avec HCl 6N
- Compléter à 100ml

Le tampon de l'étape d'électrophorèse : (étape de migration de environ 1h20min)

- Tris 120g
- Acid Boric 92.2g
- SDS 20g
- Ajuster à 2 litres

Le Bleu SDS Page : (utilisé pour les aliquotes déposées sur le gel)

- SDS 1%
- Sucrose 40%
- β mercapto éthanol (500ml pour 50ml)
- Quelques gouttes de Bleu de Bromophénol.

La solution de coloration des gels :

- Méthanol 400ml/Ac.acétique 100ml/H₂O 500ml
- Bleu de Coomassie R ou G 2.75g
- Bleu de Bromophénol

La solution de décoloration :

- éthanol 400ml
- Ac.acétique 100ml
- H₂O 1500ml

F. Le « Western Blotting Grade » :

Le Protocole de chez BioRad* :

Les solutions de base :

- PBS Tween20 0.05% ; 500ml PBS/0,25ml Tween20
- Anticorps primaire : PBS-Tween20 de base/anticorps primaire (1000:1)
- Anticorps secondaire : PBS-Tween20 de base/anticorps secondaire (500:1)
- Révélation : Chloronaphtol- H_2O_2
4-Chloro-1-naphtol 30mg/MeOH (à 0°C) 5ml/PDB Tween20 de base 50ml/ H_2O_2 30 μ l.

Après migration par électrophorèse, le gel (SDS page) est mis à transférer sur une feuille de nitrocellulose pendant 1 heure sous 300V. Les protéines du gel migrent vers la membrane et s'y fixent, sauf si la migration est trop longue, dans ce cas elles passent à travers. Après transfert et rinçage dans du tris 100mM pH=8, le gel est mis à colorer dans la solution de coloration et la membrane est mise en contact avec l'anticorps I^{me} durant 1 heure. Après un rinçage de 15min, la membrane est mise cette fois en contact avec l'anticorps II^{me} pendant 1 heure. Enfin, après un autre rinçage, toujours avec du Tris 100mM pH=8, la révélation est effectuée avec la solution ci-dessus. Une fois terminée, elle est stoppée après quelques secondes (c'est très rapide), par un lavage sous de l'eau courante.

* Bio-Rad Laboratories, 1414 Harbour Way South, Richmond, CA 94804 - phone (415) 232-7000



G. L'attribution de P40A :

Modified GIFA Output of the spin_list file.

#PPM	Nom	Système	Spin #	résidu	n°résidu
3.081	HB	13	49	HIS	6
4.859	HA	13	48	HIS	6
8.838	HN	13	47	HIS	6
-0.209	HG1	12	45	VAL	16
-0.683	HG2	12	46	VAL	16
0.650	HB	12	44	VAL	16
3.703	HA	12	43	VAL	16
7.518	HN	12	42	VAL	16
1.302	HG2	55	236	LYS	18
1.391	HG1	55	235	LYS	18
1.569	HD2	55	234	LYS	18
1.628	HD1	55	233	LYS	18
1.776	HB2	55	232	LYS	18
2.043	HB1	55	231	LYS	18
2.903	HE1	55	230	LYS	18
4.326	HA	55	229	LYS	18
7.310	HN	55	228	LYS	18
1.213	HB	14	52	ALA	19
4.148	HA	14	51	ALA	19
6.569	HN	14	50	ALA	19
2.347	HB2	15	56	ASP	20
2.606	HB1	15	55	ASP	20
4.652	HA	15	54	ASP	20
8.416	HN	15	53	ASP	20
1.361	HB	16	59	ALA	22
5.126	HA	16	58	ALA	22
8.343	HN	16	57	ALA	22
0.828	MET	25	100	LEU	24
0.917	MET	25	99	LEU	24
0.976	HB	24	94	ALA	24
1.213	HG	25	98	LEU	24
1.954	HB	25	97	LEU	24
5.126	HA	25	96	LEU	24
5.334	HA	24	93	ALA	24
9.129	HN	25	95	LEU	24
9.187	HN	24	92	ALA	24
0.176	HG2	3	12	VAL	25
0.739	HG1	3	11	VAL	25
2.162	HB	3	10	VAL	25
4.415	HA	3	9	VAL	25
9.725	HN	3	8	VAL	25
2.162	HB2	37	154	ASP	26
2.814	HB1	37	153	ASP	26
5.126	HA	37	152	ASP	26
8.823	HN	37	151	ASP	26
2.695	HB2	19	73	PHE	27
3.259	HB1	19	72	PHE	27
5.245	HA	19	71	PHE	27
8.692	HN	19	70	PHE	27
0.798	HD2	28	114	LEU	28
1.095	HD1	28	113	LEU	28
1.510	HB2	28	112	LEU	28
2.310	HB1	28	111	LEU	28
2.932	HB2	56	240	TRP	28
3.021	HB1	56	239	TRP	28
3.822	HA	28	110	LEU	28
5.067	HA	56	238	TRP	28
6.932	HN	28	109	LEU	28
8.634	HN	56	237	TRP	28
0.324	HB	29	117	ALA	29
3.259	HA	29	116	ALA	29
6.452	HN	29	115	ALA	29
2.043	HB	8	30	GLU	30
2.310	HD1	8	29	GLU	30
4.118	HA	8	28	GLU	30
9.449	HN	8	27	GLU	30
3.140	HB2	26	104	TRP	31
3.614	HB1	26	103	TRP	31
4.533	HA	26	102	TRP	31
6.627	HN	26	101	TRP	31
2.873	HB2	61	259	CYS	32
3.021	HB1	61	258	CYS	32
4.652	HA	61	257	CYS	32
6.772	HN	61	256	CYS	32

2.755	HB	39	161	ASP	47
4.504	HA	39	160	ASP	47
-7.180	HN	39	159	ASP	47
3.703	HA2	11	41	GLY	51
4.326	HA1	11	40	GLY	51
9.187	HN	11	39	GLY	51
0.739	HG2	4	17	VAL	55
0.798	HG1	4	16	VAL	55
1.984	HB	4	15	VAL	55
4.474	HA	4	14	VAL	55
9.754	HN	4	13	VAL	55
2.666	HB2	27	108	ASN	63
2.725	HB1	27	107	ASN	63
5.126	HA	27	106	ASN	63
7.310	HN	27	105	ASN	63
10.200	HN	1	1	GLY	65
3.614	HA2	1	2	GLY	65
3.911	HA1	1	3	GLY	65
0.383	HG	30	120	THR	66
3.703	other	30	119	THR	66
7.820	HN	30	118	THR	66
1.272	MET	5	20	ALA	67
3.822	HA	5	19	ALA	67
9.623	HN	5	18	ALA	67
2.132	HB2	33	131	TYR	70
3.377	HB1	33	130	TYR	70
4.385	HA	33	129	TYR	70
7.252	HN	33	128	TYR	70
1.065	HG	38	158	THR	77
3.792	HB	38	157	THR	77
4.830	HA	38	156	THR	77
7.660	HN	38	155	THR	77
0.680	MET	17	64	LEU	78
0.798	MET	17	63	LEU	78
1.628	HB	17	62	LEU	78
5.660	HA	17	61	LEU	78
9.092	HN	17	60	LEU	78
0.768	MET	18	69	LEU	79
1.272	HG	18	68	LEU	79
1.806	HB	18	67	LEU	79
5.037	HA	18	66	LEU	79
8.998	HN	18	65	LEU	79
0.442	HD2	36	150	LEU	80
0.650	HD1	36	149	LEU	80
0.976	HG2	36	148	LEU	80
1.243	HG1	36	147	LEU	80
1.895	HB	36	146	LEU	80
5.245	HA	36	145	LEU	80
8.823	HN	36	144	LEU	80
2.784	HB2	2	7	PHE	81
2.873	HB1	2	6	PHE	81
5.126	HA	2	5	PHE	81
9.885	HN	2	4	PHE	81
2.784	HB2	9	34	ASN	83
3.051	HB1	9	33	ASN	83
4.533	HA	9	32	ASN	83
9.332	HN	9	31	ASN	83
3.703	HA2	7	26	GLY	84
4.326	HA1	7	25	GLY	84
9.507	HN	7	24	GLY	84
1.895	HG	20	78	GLU	85
2.043	HB2	20	77	GLU	85
2.280	HB1	20	76	GLU	85
4.859	HA	20	75	GLU	85
7.787	HN	20	74	GLU	85
0.516	MET	52	216	VAL	86
1.806	HB	52	215	VAL	86
3.051	HA	52	214	VAL	86
8.750	HN	52	213	VAL	86
1.272	MET	6	23	ALA	87
4.444	HA	6	22	ALA	87
9.565	HN	6	21	ALA	87
1.006	HB2	21	82	THR	89
3.762	HB1	21	81	THR	89
5.037	HA	21	80	THR	89
8.460	HN	21	79	THR	89
0.531	HG	35	143	LYS	90
0.650	HB2	35	142	LYS	90
1.154	HB1	35	141	LYS	90
1.361	HD2	35	140	LYS	90
1.450	HD1	35	139	LYS	90
4.474	HA	35	138	LYS	90

8.903	HN	35	137	LYS	90
4.504	HA	57	242	VAL	91
8.590	HN	57	241	VAL	91
3.585	HA2	31	123	GLY	92
4.326	HA1	31	122	GLY	92
7.950	HN	31	121	GLY	92
3.733	HA2	22	85	GLY	97
3.881	HA1	22	84	GLY	97
8.736	HN	22	83	GLY	97
0.027	unk	23	91	LEU	103
0.531	MET	23	90	LEU	103
0.857	MET	23	89	LEU	103
1.776	HB	23	88	LEU	103
3.436	HA	23	87	LEU	103
8.503	HN	23	86	LEU	103
2.577	HB2	10	38	ASP	104
2.666	HB1	10	37	ASP	104
4.296	HA	10	36	ASP	104
9.129	HN	10	35	ASP	104
<hr/>					
3.199	HB2	32	127		
3.288	HB1	32	126		
4.385	HA	32	125		
7.601	HN	32	124		
<hr/>					
0.620	HD1	34	136		
1.006	HG	34	135		
1.687	HB	34	134		
5.067	HA	34	133		
8.943	HN	34	132		
<hr/>					
1.302	*MET*	40	164		
4.029	HA	40	163		
7.427	HN	40	162		
<hr/>					
0.413	MET	41	170		
0.620	HG2	41	169		
0.768	HG1	41	168		
1.628	HB	41	167		
4.266	HA	41	166		
7.441	HN	41	165		
<hr/>					
1.391	MET	42	173		
4.148	HA	42	172		
7.514	HN	42	171		
<hr/>					
2.606	HB2	43	177		
2.666	HB1	43	176		
4.266	HA	43	175		
7.514	HN	43	174		
<hr/>					
0.798	MET	44	182		
1.569	HB2	44	181		
1.687	HB1	44	180		
4.385	HA	44	179		
7.732	HN	44	178		
<hr/>					
3.762	HA2	45	185		
4.177	HA1	45	184		
7.572	HN	45	183		
<hr/>					
3.703	HA2	46	188		
3.792	HA1	46	187		
7.470	HN	46	186		
<hr/>					
1.865	HG	47	193		
2.251	HB2	47	192		
2.399	HB1	47	191		
4.089	HA	47	190		
7.630	HN	47	189		
<hr/>					
1.302	MET	48	196		
3.940	HA	48	195		
8.067	HN	48	194		
<hr/>					
1.480	other	49	202		
1.717	HG1	49	201		
1.954	HB2	49	200		
2.043	HB1	49	199		
4.029	HA	49	198		
8.052	HN	49	197		
<hr/>					
1.154	MET	50	205		
4.059	HA	50	204		
8.081	HN	50	203		
<hr/>					
1.391	MET	51	212		
1.569	HG2	51	211		
1.628	HG1	51	210		
1.836	HB2	51	209		
1.925	HB1	51	208		
4.563	HA	51	207		
8.110	HN	51	206		
<hr/>					
1.480	HG2	53	221		
1.539	HG1	53	220		
1.836	HB	53	219		
3.081	HA	53	218		

7.107	HN	53	217
2.191	HG2	54	227
2.310	HG1	54	226
2.429	HB2	54	225
2.488	HB1	54	224
4.266	HA	54	223
7.092	HN	54	222
<hr/>			
2.577	HB2	58	246
2.636	HB1	58	245
4.504	HA	58	244
8.154	HN	58	243
<hr/>			
0.027	HG2	59	251
0.531	HG1	59	250
1.480	HB	59	249
3.733	HA	59	248
7.165	HN	59	247
<hr/>			
3.318	HB	60	255
4.118	HA	60	254
4.415	HA	60	253
8.140	HN	60	252
<hr/>			
2.102	HB2	62	263
2.606	HB1	62	262
3.911	HA	62	261
8.983	HN	62	260

H. Publication sur les effets de la mutation P40A dans *Escherichia Coli* :

STRUCTURAL AND FUNCTIONAL ROLES OF A CONSERVED PROLINE RESIDUE IN THE $\alpha 2$ HELIX OF *E. coli* THIOREDOXIN

(Cette publication a été soumise à une deuxième lecture en vue de sa parution dans BIOCHEMISTRY)

Frédéric de LAMOTTE-GUERY[‡]°, Christian PRUVOST[§], Philippe MINARD[‡], Marc-André DELSUC[‡], Myroslawa MIGINIAC-MASLOW[‡], Jean-Marie SCHMITTER[⊥], Mariana STEIN[‡], and Paulette DECOTTIGNIES^{†*}

‡ Institut de Recherche sur les Plantes, Université PARIS XI, 91 405 Orsay, FRANCE.

° Present address: Unité de Biochimie et Biologie Moléculaire des Céréales, INRA, 34060 Montpellier, FRANCE

§ Centre de Biochimie Structurale, Faculté Pharmacie, 34060 Montpellier, FRANCE

|| Laboratoire d'Enzymologie Physicochimique Moléculaire, Université PARIS XI, 91405 Orsay, FRANCE.

⊥ Laboratoire de Biochimie, Ecole Polytechnique, 91128 Palaiseau, FRANCE

*Corresponding author:

P. DECOTTIGNIES, Institut de Recherche sur les Plantes, Bat. 630, Université PARIS XI, 91 405 Orsay Cedex, FRANCE.

Running Title: Pro40 substitution in thioredoxin

Abbreviations

DTT: Dithiothreitol

DTNB: 5,5'-Dithio-bis(-2-nitro-benzoic acid)

Gdn-HCl : guanidine hydrochloride

NADP-MDH: NADP - dependent malate dehydrogenase

P40A: thioredoxin with Pro 40 replaced by Alanine

rms: root mean square

SFE: Solvation Free Energy

WT: wild type

ABSTRACT

Proline 40 in *E. coli* thioredoxin is located close to the redox active site (Cys 32- Cys 35) within the $\alpha 2$ helix that it binds. The high conservation of this residue among thioredoxins suggests that it could play an important role in the structure and/or the function of this protein. We have substituted Pro 40 by Ala by using site directed mutagenesis and expressed the mutant P40A in *E. coli*. The effects of the mutation on the biophysical and biological properties of thioredoxin have been analyzed and compared to molecular dynamic simulations. Modeling predicted that the replacement of Pro40 by Ala induces a displacement of the active site which exposes Trp31 to the solvent and opens a cleft located between helices $\alpha 2$ and $\alpha 3$. The SFE calculation also indicated that P40A was more hydrophobic as W31 became more accessible. These predictions were totally in agreement with the experimental results. The mutant P40A exhibited chromatographic behavior and fluorescence properties very different from the ones of the wild-type (WT) protein. The determination of the free energy of unfolding of P40A shows that the mutant was globally destabilized. However, the effect of the mutation on the transition curve was highly unusual as the midpoint of the unfolding transition increased, indicating that some local structures were actually stabilized by the mutation. Neither the ability of the protein to reduce a chloroplastic enzyme nor its reactivity with the bacterial reductase decreased. The only functional difference was the higher stability of P40A in light activation of NADP-malate dehydrogenase under air, which suggests that the mutant was less rapidly re-oxidized than WT. It is proposed that the global mode of relaxation induced by the mutation might reflect the natural movement of thioredoxin when it interacts with its target.

INTRODUCTION

Thioredoxins are small molecular mass (12 kDa) heat stable proteins containing a readily reducible disulfide bridge with the amino acid sequence Cys-Gly-Pro-Cys. Through their isolation and characterization from numerous species, thioredoxins have been found to play a role in physiological processes as diverse as blood clotting (Blombäck *et al.*, 1986), free radical defence in human epidermis melanocytes (Schallreuter *et al.*, 1989), assimilation of nitrogen (Schallreuter *et al.*, 1989) and regulation of photosynthesis (Buchanan, 1980).

At the molecular level, thioredoxins serve as redox carriers in a wide variety of reactions, such as ribonucleotide reduction (Laurent *et al.*, 1964) or methionine sulfoxide reduction (Gonzalez Porque *et al.*, 1980). Their function as general protein disulfide reductases has been well demonstrated in the cases of insulin (Holmgren, 1979), choriogonadotropin, (Holmgren & Bränden, 1989) and several chloroplast enzymes including NADP-malate dehydrogenase (NADP-MDH) (Buchanan, 1980), phosphoribulokinase (Porter *et al.*, 1988) and fructose-1,6-bisphosphatase (Schürmann & Wolosiuk 1978). In addition to their redox function, thioredoxins have been shown to play a structural role in several systems. *E. coli* thioredoxin is required for the replication of T7 bacteriophage, as a subunit of the T7 DNA polymerase complex (Tabor *et al.*, 1987), and is essential for the assembly of the filamentous phages M13 and f1 (Russel & Model, 1985). In these two cases, mutagenesis experiments have shown that the disulfide bridge of thioredoxin is not required (Russel & Model, 1986) while it is absolutely necessary for chloroplast enzyme activation (Jacquot & Decottignies, 1986). All thioredoxins present a common fold consisting of a central 5-stranded β sheet with two parallel junctions and two anti-parallel junctions. This central sheet is surrounded by 4 α helices. The active site is located in the loop which connects $\alpha 2$ helix with the 2nd strand of the β sheet (Katti *et al.*, 1990). The structure of reduced *E. coli* thioredoxin has been elucidated by NMR (Dyson *et al.*, 1989). Reduced and oxidized thioredoxins have identical secondary structures and tertiary folds.

Many of the highly conserved residues have been extensively studied by several groups. The role of the residues surrounding the active site (Cys 32- Cys 35), Lys 36 as well as Trp 31 and Pro 34 has been investigated (Fuchs & Eklund; Holmgren). On the other hand, Asp 26, buried in the core of the protein has been scrutinized by (Woodward and sturtevand). Among the conserved residues, Pro 40 is present in all thioredoxins sequenced so far, except a chloroplast-encoded thioredoxin present in a red alga (Reynolds *et al.*, 1994) and the two h-type thioredoxins from *N. tabacum* (Brugidou *et al.* 1993). This residue shows the remarkable feature of being located in the $\alpha 2$ helix, disrupting the canonical secondary structure. This induces a bent into the helix which bears the active site and connects the two domains of the protein.

In the present study, we investigate the role of this conserved Proline, both biochemically and structurally. The P40A mutant is compared to the wild type protein using a wide range of enzymatical and physicochemical tests. This results are interpreted in the light of a molecular dynamic simulation.

MATERIALS AND METHODS

Enzymes. NADPH thioredoxin reductase and thioredoxin from *E. coli* were purchased from IMCO (Stockholm, Sweden). NADP-malate dehydrogenase was purified according to Decottignies *et al.* (1988). Ferredoxin (Mayhew, 1971) and ferredoxin-thioredoxin reductase (Droux *et al.*, 1987) were purified as described.

Bacterial strain and vectors. TG2 (supE hsdR Δ 5 thi Δ (lac-proAB) Δ (srl-recA)306 :: Tn10 (tet^r); F' [traD36 pro AB⁺ lacI^q lacZ Δ Em15]) was used for M13 (Pharmacia) and BlueScript KS II plasmid (Stratagene) preparation, as well as for thioredoxin overexpression. Bacteria were grown in Luria-Bertani medium with ampicillin at a final concentration of 50 μ g/ml when necessary.

Construction of mutations. The *KpnI/PstI* fragment containing the *trxA* gene from *E. coli* strain K38 and its regulatory sequences was inserted into M13 mp19. Single stranded DNA was used as a template for the site directed mutagenesis (Kunkel, 1985). A 21-bases deoxyoligonucleotide was synthesized complementary to the coding strand with the following sequence (mismatch underlined): ATC CAG AAT GCG GGG GAT CAT. The replicative form DNA of one positive clone was used to transfer the *trxA* gene into pFP1 (de Lamotte-Guéry *et al.*, 1991).

Production and purification of wild type and mutant thioredoxins (de Lamotte-Guéry *et al.*, 1991). Mutant and wild type thioredoxins were expressed and purified as described previously using the pFP1 overexpression plasmid. Thioredoxin P40A was further purified on a TSK phenyl 5PW column (0.75 cm x 7.5 cm; LKB) by using a 40 min linear decreasing gradient of ammonium sulphate (0.9 - 0 M) in 0.1M sodium phosphate pH 7.2. The flow rate was 1 ml. min⁻¹. Thioredoxin concentrations were determined from the absorbance at 280 nm, using 12650 M⁻¹ cm⁻¹ as molar absorption coefficient. Purity of the preparations was determined by SDS-PAGE and native gels. Commercial wild type thioredoxin was used as a standard.

Control of the mutation.

1) Tryptic mapping and sequencing. Thioredoxins (8 nmol) were reduced with 3.5 mM DTT for 30 min at 37°C and carboxymethylated with 5 mM iodo [1-¹⁴C] acetic acid for 15 min in the dark. The reaction was stopped with 30 mM 2-mercaptoethanol and the reagents in excess were removed using a Centricon 10 microconcentrator (Amicon). In the same process the phosphate buffer was replaced by 0.1M NH₄HCO₃, pH 8.3. ¹⁴C-carboxymethylated thioredoxins were cleaved with trypsin (1/40, w/w) for 5h at 37°C. The digest was purified by reverse-phase liquid chromatography on a C4 column (0.46 cm x 15 cm; Vydac) with a 90 min-linear gradient over 0-70% CH₃CN in 0.1% trifluoroacetic. The peptide containing the mutated amino acid was sequenced using an Applied Biosystems model 470 A sequencer.

2) Electrospray Mass spectrometry. Thioredoxins were analyzed with a VG BioQ mass spectrometer, using a 20 pmol/ml protein solution in 1/5/5/ formic acid/methanol/ water.

Reverse-Phase Liquid Chromatography. Thioredoxins were chromatographed on a C4 column (0.46 cm x 15 cm, Vydac). Elution was performed with a 30 min linear gradient over 42 - 56 % CH₃CN in 0.1% trifluoroacetic acid, at a flow rate of 1 ml. min⁻¹.

Fluorescence measurements. Spectra were recorded using a Perkin Elmer MPF-44B fluorimeter at 20°C in a 1 cm path length cell in a total volume of 1 ml containing 10 mM phosphate buffer pH 7.2 and 7 μ M thioredoxin. The excitation wavelength was 280 nm and emission was recorded from 280 to 400 nm. Reduction was performed with 5 mM DTT for 10 min at 20°C.

Circular dichroism. Spectra were recorded from 185 to 250 nm using a Jobin-Yvon Mark IV dichrograph at 20 °C. Measurements were made in a 0.2 cm path length cell in a total volume of 0.7 ml of 10 mM phosphate buffer pH 7.2 containing 7 μ M thioredoxin. Reduction was performed with 5 mM DTT for 10 min at 20°C. Data were analyzed using the method of Chang *et al.* (1978).

Denaturation and renaturation experiments. Thioredoxin samples (5 μ M) in 25 mM phosphate buffer, pH 7.2, were equilibrated overnight at 25°C in the presence of variable concentrations of Gdn-HCl (0 to 6 M). Fluorescence and circular dichroism spectra were recorded as described above. For renaturation experiments, oxidized samples were denatured with 6 M Gdn-HCl overnight, then diluted with buffer at various concentrations.

Enzyme assays.

1) Reduction by thioredoxin reductase (Slaby & Holmgren, 1975).

Thioredoxins were tested as substrates of *E. coli* NADPH-thioredoxin reductase in a final volume of 1 ml containing 100 mM phosphate buffer pH 7.1, 10 mM EDTA, 150 μ M NADPH, 500 μ M 5,5'-Dithio-bis(-2-nitrobenzoic acid) and variable amounts (0.3-12 μ M) of thioredoxin. The reaction was started by adding 10 nM of thioredoxin reductase and was monitored by the absorbance change at 412 nm. The activity was expressed as μ moles of thioredoxin reduced per min, using 13600 M⁻¹ cm⁻¹ as the molar absorption coefficient of DTNB.

2) DTT dependent activation of corn leaf NADP-MDH. (Jacquot *et al.* 1981)

The enzyme was activated in 30 μ l of medium containing 100 mM Tris-HCl pH 7.9, 10 mM DTT, 0.6 μ M NADP-MDH and thioredoxin (0.6 to 22 μ M). After incubating for 15 min at 20°C, an aliquot of 10 μ l was used to determine the activity of NADP-MDH at 30°C in a reaction medium (1 ml) containing 100 mM Tris-HCl buffer pH 7.9, 0.15 mM NADPH and 0.7 mM oxaloacetic acid. NADPH oxidation was monitored as a decrease in absorbance at 340 nm.

3) Light activation of corn leaf NADP-MDH (Droux *et al.*, 1987)

The enzyme was activated in 40 μ l of mixture containing 100 mM Tris-HCl pH 7.9, pea leaf thylakoids (equivalent to 17 μ g of chlorophyll), 12 μ M spinach ferredoxin, 2 μ M ferredoxin-thioredoxin-reductase, 0.5 μ M NADP-MDH and variable amounts (1 to 20 μ M) of thioredoxin, at 25°C under illumination (400 μ mol quanta/m²/sec). At intervals, 30 μ l aliquots were withdrawn to measure the activity of NADP-MDH as described above.

Modeled structures of P40A and WT thioredoxins

The X-ray structure of thioredoxin (Kalti *et al.*, 1990) was the starting point of all the simulations, which were performed on the reduced form of native and mutant proteins. The P40A substitution was built by mutating directly the Proline 40 into an Alanine. Both structures were minimized before further use. The solvent was incorporated as a 25 Å sphere around the solute centre-of-mass, comprising 1540 water molecules. The hydrogen charge in water was set to 0.417 C. The solute-solvent ensemble was equilibrated by alternating steps of energy minimization and 10 K dynamics. The minimization was first applied to the water molecules, then extended progressively to the solute hydrogens, side-chains and finally to the protein backbone. A 400 ps dynamics in water was performed at 300 K with a 200 fs contact time and a 1 fs step. The simulation was started at 10 K and the system was allowed to reach the bath temperature in 20 ps. A distance dependent dielectric function, $\epsilon = r_{ij}$, and a cutoff distance of 10 Å for the non-bonded interactions were used. Results were analysed on the last 100 ps after stabilization of the total energy and rms deviation from the starting structure.

The simulations were performed using the program Amber 4.0. The programs InsightII and Access (Eisenberg & McLachan, 1986). were used for comparison and first evaluation of the structures, and for the Solvation Free Energy (SFE) estimation, the solvent accessibility surface calculations. The helix angle was approximated by the angle between two lines, each of them positioned in the direction of the helix axis on each side of the helix. All the calculations were performed on a Hewlett Packard HP735 under OS HPUX 9.01.

RESULTS

Purification of WT and P40A thioredoxins and control of the mutation

After sequencing the *trxA* gene and subcloning in pFP1, *E. coli* WT and P40A thioredoxins were purified by heat treatment, ammonium sulfate precipitation, gel filtration and anion-exchange chromatography. Such a procedure allowed us to obtain WT thioredoxin in a pure form, but P40A required to be further purified by hydrophobic interaction chromatography, because it contained a small amount of the endogenous WT protein. Thioredoxin activity was monitored during purification by using the DTT-dependent NADP-MDH activation test. After these steps, a single band was observed on SDS-PAGE and on native gels. The preparation was analyzed by protein sequencing which revealed the sequence: Ser-Asp-Lys-Ile-Ile-His-Leu-Thr-Asp-Asp... which is identical to the N-terminal sequence of WT protein.

The replacement of a Proline residue by an Alanine does not change the number of sites cleaved by trypsin. Thus, the tryptic patterns were expected to be identical or almost identical for WT and P40A. Only one peptide had a slightly different retention time when the two patterns were compared. This peptide had been previously shown by mass spectrometry to correspond to the stretch M37 to K52. When sequenced, the fragment derived from P40A revealed the expected mutation, i. e. ³⁷MIAAILDE...

P40A and WT were also analyzed by Electrospray Mass Spectrometry. The measured mass of WT was 11674.8 \pm 0.7 da which has to be compared to the calculated mass: 11675.4. The measured mass of P40A was 11648.9 \pm 0.5 da (calculated mass: 11649.4 Da). Therefore, the difference of mass was in good agreement with the performed mutation and we could conclude that the mutant protein contained only one modification, the replacement of Proline 40 by an alanine residue. In addition, these data confirmed that the purified P40A preparation was no longer contaminated by the WT protein.

Physicochemical characterization of P40A

The observation that P40A was easily separated from contaminating WT thioredoxin on a hydrophobic interaction column during the purification prompted us to investigate the chromatographic behavior of WT and P40A more thoroughly. A mixture of the two purified proteins could be also separated on a C4 reverse phase column. When reduced, both proteins eluted earlier from this column as already observed (Wetterauer *et al.*, 1991) but were separated as efficiently as the oxidized forms.

Fluorescence emission spectra of oxidized and reduced WT and P40A thioredoxins were recorded (λ_{ex} = 280 nm) (Fig.2). Fluorescence of oxidized WT was low, as expected since fluorescence is usually quenched in the WT in those conditions (Holmgren, 1972). A 3-fold increase was observed after reduction, as already reported. On the other hand, the fluorescence emission of oxidized P40A was significantly higher than the fluorescence of WT protein, revealing that the conformation is less quenched in the mutant protein. When P40A was reduced, fluorescence also increased, but

to a lesser extent (2-fold).

No difference between wild type and mutated thioredoxin CD spectra was observed. This suggests that the amino acid substitution did not alter the overall secondary structure of the protein.

Both overproduced thioredoxins were submitted to irreversible inactivation assay by heat at 95°C, since *E. coli* WT thioredoxin is known to be remarkably heat stable (Ladbury *et al.*, 1993). Aliquots were taken at different times up to two hours and tested for their ability to activate the NADP-MDH in the presence of DTT. No difference was observed as both proteins exhibited remarkable thermostability, activating the target enzyme with a constant efficiency.

The unfolding of oxidized P40A and WT thioredoxins induced by Gdn-HCl was followed by the change in Tryptophan fluorescence emission and CD ellipticity. Addition of Gdn-HCl resulted in a cooperative and reversible unfolding of both thioredoxins. Similar results were obtained when denaturation was monitored by CD at 220 nm. As shown in table 1, the midpoint of the transition curves observed with P40A significantly increased compared to the WT protein. However, this effect was associated with a decrease in the cooperativity of the folding transition. Therefore, oxidized P40A appears to be less stable by 2.9 kcal/mol than the WT protein in spite of the unusual increase of the C_m .

Enzymatic characterization of P40A

Several tests were used to determine if the mutation modified the activity of the protein. Thioredoxins were reduced either by chemical (DTT) or protein (NADPH-thioredoxin reductase or ferredoxin-thioredoxin-reductase) reductants.

Oxidized WT and P40A were tested as substrates for *E. coli* NADP-thioredoxin reductase. The values of K_m and k_{cat} were almost identical for the two proteins (Table 2). Thus the mutation did not significantly alter the ability of the protein to be reduced by its specific reductase.

Since *E. coli* thioredoxin was shown to be as efficient as the thioredoxin m from spinach chloroplasts in the DTT-dependent activation of corn NADP-MDH (Decottignies *et al.*, 1990), the activity of NADP-MDH was monitored at various concentrations of WT and P40A thioredoxins. The concentration required to obtain an activation rate of NADP-MDH equivalent to half of V_{max} ($S_{0.5}$) was identical for the two proteins and catalytic rates were similar (Table 2).

The *in vitro* light activation system mimics the chloroplastic process of activation of assimilatory enzymes under light. It requires the reduction of thioredoxins by the ferredoxin-thioredoxin reductase prior to their action on the target enzymes. When NADP-MDH was activated this way, identical concentrations of WT and P40A thioredoxins were required to reach the maximal activity. However the behavior of the system at non-saturating concentration of thioredoxin was different under air and under argon. Under air, the activity of NADP-MDH dramatically decreased after 10 min in the presence of WT protein (Fig.4). This decrease was not observed when the experiment was performed under argon. When P40A was tested in this system, the activity of MDH was stable for at least 20 min under air (Fig.4), and no difference was observed under argon.

Modeled Structure analysis of P40A and WT thioredoxin

Comparison of minimized forms

Refined P40A and WT thioredoxin structures, minimized in spheres of water, were analyzed. The total energy during refinement reached a minimum point of -5750 KCal/mol for the WT and -5780 KCal/mol for the P40A with an average root mean square deviation of 0.01 Å for both structures.

Analysis and comparison of trajectories, structure evolution

The molecular dynamics protocol was applied on both WT and mutant structures. Considering the overall trajectories, the global structure remained conserved between the P40A and WT proteins (Fig. 5). The two structures kept their overall fold, with the conservation of the global shape and the secondary structures, except for the small N-terminal $\beta 1$ strand, which appeared to be slightly more flexible in the mutant protein. The other β strands were very well conserved with a backbone rms of 1.54 Å. All the helices remained highly structured and conserved during the simulations. $\alpha 1$ helix was less conserved during the dynamic of the mutant protein (rms= 1.38 Å) because of a small orientation difference due to the shift of the $\beta 1$ strand. A striking effect of the substitution was observed in helix $\alpha 2$ which became more straight in P40A (rms= 0.91 Å), shifting the active site and pulling on the entire structure. $\alpha 3$ and $\alpha 4$ helices did not substantially change (rms= 0.66 Å and 0.88 Å) but the C-terminal helix moved a little along its axis with a very well conserved geometry.

Some differences in the $\alpha 2$ helix could be seen around the location of the Pro40 mutation. The hydrogen bond network was more important in the $\alpha 2$ helix of the mutant, allowing this helix to approach a more regular structure in the simulations. The angle in this helix was 59° in WT and 13° in P40A (Fig.6).

The mutation did not modify the structure of the active site, for which internal geometry was very well conserved (rms_{res31,35}= 0.93 Å and rms_{res32,35}= 0.27 Å on the backbone), but moved from its original position by about 6 Å (Fig.7). The shift of the active site entailed very different contacts in the two structures in the vicinity of the active site.

The most striking differences concerned W31. The hydrophobic bridge binding W31 and the loop between $\beta 2$ and $\alpha 3$ was suppressed in P40A, deleting the contacts with I5, P30, I60, E62, D63, P64 and I72. This modification induced many changes in hydrogen bonds, for instance: in the mutant thioredoxin, W31-HNE:::I60-O disappeared, W28-HNE:::K57-O was weakened and W28-HN:::K57-O appeared (Fig.8). Those changes drastically modified the mobility of W31. This residue appeared to be very static in the WT structure, stuck to a hydrophobic patch constituted of I60, P64, I72 whereas, in P40A, it was detached from the molecule surface and free to move (Fig.9). Whereas the mean distance between the two indol rings varied only from 9.64 Å to 10.54 Å between the two forms, the mean angle between the planes drastically changed from 16° to 94°. We calculated the solvent access surface for the tryptophan residues during the last 100 ps. The mean solvent access surface increased from 45.5 Å² to 74.4 Å² for W28 and from 112.7 Å² to 233.0 Å² for W31. The mean SFE was calculated also on the last 100 ps, showing a more hydrophilic surface for WT with a difference of 2.9 Kcal/mol. The distances between the indol rings and the disulfide bond were measured to 7.95 Å for W31 and 7.08 Å for W28 in the wild type form, and 9.99 Å for W31 and 7.26 Å for W28 in the mutant.

DISCUSSION

We modified a very well conserved residue, Proline 40, implicated in the structuration of the active site (C32-C35). This proline which induces a kink in the $\alpha 2$ helix, was mutated into an Alanine. The effects of this mutation were monitored by a wide array of modeling, biological and biophysical tests. Whereas modeling hints to some structural changes, biological tests reveal only one subtle modification. The biophysical analyses however, confirm the effect observed in the simulation.

The overall structure of thioredoxin is conserved in P40A as predicted by modeling and experimentally confirmed by CD data. The modeling predicts that the kink in $\alpha 2$ helix induced by the presence of the proline residue is suppressed in P40A and that the helix adopts a more regular structure. The active site is predicted to undergo a 6 Å shift without any modification of its intimate geometry. This displacement exposes W31 of the active site to the solvent and opens up a cleft located between helices $\alpha 2$ and $\alpha 3$. Such a change induces the deletion of various hydrogen bonds engaged with W31. However, the residues which are considered to be in interaction with target proteins (residues 33-34, 75-76, 91-93) (Eklund *et al*, 1984) are not modified according to the modeling data, and their hydrogen bonds are not affected by the mutation. This can explain that biological activities of P40A are identical to the ones of WT thioredoxin : neither the ability of the protein to reduce a chloroplastic enzyme nor its reactivity with the bacterial reductase are decreased.

The calculation of Solvation Free Energy indicates that the mutant is more hydrophobic as W31 becomes more accessible.

As demonstrated earlier (Krause 91; Slaby 96), the major contribution to fluorescence of the reduced *E. Coli* thioredoxin comes from W28, whereas the persisting fluorescence displayed by the oxidized form is probably due to W31, W28 being strongly quenched by W31 and the disulfide bridge. In the present study, both redox states of P40A exhibit a stronger fluorescence than their wild type counterparts. This effect can be rationalized in the light of the modeling of P40A : i) the angle between the two tryptophan planes is modified so as to bring the two moieties orthogonal. This could explain the stonger fluorescence of P40A reduced form as W31 being less efficient in quenching W28. ii) the distance between W31 and the cystine increases by 2 Å, allowing W31 to participate to the increased fluorescence of the oxidized form.

WT thioredoxin is more stable upon denaturation by Gdn-HCl than P40A. However, the effect of the mutation on the transition curve is highly unusual. In almost all known examples, destabilizing mutations are observed to decrease either the transition midpoint, or less frequently, the cooperativity of the unfolding transition. The remarkable feature of P40A thioredoxin is that the transition midpoint increased, indicating that some local structures are actually stabilized by the mutation. But the overall structure is globally destabilized as reflected by the decrease of the m and ΔG values. This might result of several changes, but the main contribution to the destabilization could be the exposure of W31 to the solvent and the unmasking of its natural pocket. The magnitude of this effect (2.9 kcal/mole) is similar to the one observed in several thioredoxin with single mutations in the hydrophobic core (Hellinga *et al*, 1992).

In addition, the chromatographic behavior of P40A on a reverse-phase column can be related to its decreased thermodynamic stability. Although this chromatography is considered to be performed in «denaturing» conditions (organic solvents and low pH), it has already been observed that tertiary structure can be more or less retained, depending on the stability of the protein (Kunitani *et al*, 1986). It can be proposed that P40A, which is less stable than WT, is much more unfolded on the reverse-phase column and thus more tightly retained by hydrophobic interactions.

The only observed functional difference between the two proteins is the higher stability of the mutant in light activation of NADP-malate dehydrogenase, under air. In this system, there is an equilibrium between reduction by illuminated photosynthetic membranes and re-oxidation by air. The decrease in activity observed for WT is usually attributed to the fact that oxidation takes over reduction after 10 min (Decottignies *et al*, 1990) because the efficiency of

the electron transfer is decreasing and oxygen concentration of the medium is increasing. As P40A exhibits a very stable level of NADP- MDH activation, this suggests that it is less rapidly re-oxidized than WT.

The behavior of the mutant P40A is convergent with previously reported results. It has already been shown (Hellinga *et al.*, 1992) that structurally mutant thioredoxins exhibited perturbed thermodynamic properties while being not affected in their biological activity. Those mutations, located in the hydrophobic core of the protein, were assayed for *in-vivo* assembly of filamentous bacteriophage for which the redox activity is not required. The authors concluded that, in thioredoxins, the link between stability and function was not as strong as found in other systems. More recently, the replacement of Trp 28 in *E. coli* was described to induce effects only on thermodynamic stability and not on functional properties (Slaby *et al.*, 1996).

In addition, our results should be connected with observations made by Qin *et al.* (1995-1996) on the interaction of human thioredoxin with a target peptide from Ref-1 and NF κ B. These authors describe that a similar cleft opens when the interaction takes place, and several hydrophobic interactions occur between the peptide residues and the residues of thioredoxin exposed by this opening. In these complexes, W31 appears to be detached from its hydrophobic patch and to be engaged in several contacts with the target peptides. Even if initiated by the release of the structural constraints caused by the presence of a Proline within α 2 helix, the global mode of relaxation observed in P40A might, in some ways, reflect its natural movement as observed by Qin *et al.* (1995-1996) This hypothesis is supported by the unaltered biological activity of our mutant, the effect of the P40A mutation studied here being to favor the conformation that thioredoxin naturally adopts when in interaction with its target.

Acknowledgments

The authors would like to thank Alain Van Dorsselaer for help in mass spectrometry, Markku Saarinen and Thérèse Malliavin for helpful discussion, and the Centre National de Transfusion Sanguine for the use of their dichrograph.

REFERENCES

- Blombäck, B., L. Adamson, Hessel, B. Holmgren, A. Hogg, D. & Procyk, R. (1986) in *Thioredoxin and glutaredoxin systems: structure and function* pp 357-367. Raven Press, New York.
- Brugidou, C, Marty, I., Chartier, Y. & Meyer, Y. (1993) *Mol. Gen. Genet.* 238, 285-293.
- Buchanan, B. B. (1980) *Ann. Rev. Plant. Physiol.* 31, 341-374.
- Chang, T. C., Wu, C. S. C., & Yang, J.T. (1978) *Anal. Biochem.* 91, 13-31.
- Decottignies, P., Schmitter J-M., Jacquot, J-P., Dutka, S., Picaud, A. & Gadal, P. (1990) *Arch. Biochem. Biophys.* 280, 112-121.
- Decottignies, P., Schmitter, J-M., Miginiac-Maslow, M., Le Maréchal P., Jacquot, J-P, & Gadal, P. (1988) *J.Biol. Chem.* 263, 11780-11785.
- Droux, M., Miginiac-Maslow, M., Jacquot, J-P, Gadal, P., Crawford, N.A., Kosower, N.S., & Buchanan, B.B. (1987) *Arch. Biochem. Biophys.* 256, 372-380.
- Dyson, H.J., Holmgren, A., & Wright, P.E. (1989) *Biochemistry* 28, 7074-7087.
- Eisenberg, D. & McLachan, A. D. (1986) *Nature* 319, 199-203.
- Eklund, H., Cambillau, C., Sjöberg, B-M., Holmgren, A., Jörnvall, H., Höög, J-O., & Bränden, C-I. (1984), *EMBO Journal* 3, 1443-1449.
- Gonzalez Porque, P., Baldensten, A., & Reichard, P. (1970) *J.Biol. Chem.* 245, 2371-2374.
- Hellinga, H.W., Wynn, R., & Richards, F.M. (1992) *Biochemistry* 31, 11203-11209.
- Holmgren, A. (1972) *J.Biol. Chem.* 247, 1991-1998.
- Holmgren, A. (1979) *J.Biol. Chem.* 254, 9627-9632.
- Holmgren, A., and Bränden, C. I. (1989) *Nature* 342, 248-251.
- Jacquot, J.P., Buchanan, B.B., Martin, F. & Vidal, J. (1981) *Plant Physiol.* 68, 300-304.
- Jacquot, J. P., & Decottignies P. (1986) *FEBS Letters* 209, 87-91.
- Katti, S.K., Lemaster, D.M., & Eklund, H. (1990) *J. Mol. Biol.* 212 167-184.
- Krause et al., (1991) *J Biol Chem.* 266:15, 9494-500.
- Kunitani, M., Johnson, D., & Snyder, L.R. (1986) *J. of Chromato.* 371, 313-333.
- Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci.* 82, 488-492.
- Ladbury, J.E., Wynn, R., Hellinga, H.W & Sturtevant, J.M. (1993) *Biochemistry* 32, 7526-7530.
- Laurent, T. C., Moore, E.C., & Reichard, P. (1964) *J. Biol. Chem.* 239, 3436-3441.
- de Lamotte-Guéry, F., Miginiac-Maslow M., Decottignies P., Stein M., Minard P., & Jacquot, J.P. (1991) *Eur. J. Biochem.* 196, 287-294.
- Le Maréchal P., Hoang, B.M.C., Schmitter, J.M., Van Dorsselaer A. & Decottignies, P. (1992) *Eur. J. Biochem.* 210.

421-429.

Mayhew, S.G. (1971) *Anal. Biochem.* 42, 191-194.

Porter, M. A., C. D. Stringer, & Hartman, F.C. (1988) *J. Biol. Chem.*, 263, 123-129.

Qin, J., Clore, G.M., Poindexter Kennedy, W.M., Huth, J.R. & Gronenborn, A. (1995) *Structure*, 3, 289-297.

Reynolds, A.E., Chesnick, J.M., Woolford, J., & Cattolico, R.A. (1994) *Plant Mol. Biol.* 25, 13-21.

Russel, M., & Model, P. (1985) *Proc. Natl. Acad. Sci.* 82, 29-33.

Russel, M., & Model, P. (1986) *J. Biol. Chem.* 261, 14997-15005.

Schallreuter, K. U., Pittelkow, M. R., & Wood, J.M. (1989) *Biochem. Biophys. Res. Comm.* 162, 1311-1316.

Schürmann, P. & Wolosiuk, R.A. (1978) *Biochem. Biophys. Acta.* 522, 130-138.

Slaby, I. & Holmgren, A. (1975) *J. Biol. Chem.* 250, 1340-1347.

Slaby, I., Cerna, V., Jeng, M.F., Dyson, H.J., & Holmgren, A. (1996) *J. Biol. Chem.* 271, 3091-3096.

Tabor, S., Hubert, H. E., & Richardson, C.C. (1987) *J. Biol. Chem.* 262, 16212

Wetterauer, B., Véron, M., Miginiac-Maslow, M., Decottignies, P., & Jacquot J.P. (1992) *Eur. J. Biochem.* 209, 643-649.

Table 1: Guanidine hydrochloride denaturation parameters for WT and P40A thioredoxins

	(Gdn-HCl) _m (M)	ΔG° (keal/mol)	m (kcal/mol/M)
WT	2.4	8.9	3.7
P40A	2.8	6.0	2.2

ΔG° is the free energy in the absence of denaturant, (Gdn-HCl)_m is the denaturant concentration at the midpoint of the unfolding transition and m the dependence of ΔG on denaturant concentration. All the values are derived from data presented in Fig. 3.

Table 2 : Kinetic constants of WT and P40A thioredoxins

	NADPH-Thioredoxin Reductase			NADP-MDH activity	
	K _m (μM)	k _{cat} (s ⁻¹)	k _{cat} /K _m	S _{0,5} (μM)	k _{cat} (s ⁻¹)
WT	3.2	12.5	3.9	5	270
P40A	1.7	10.5	6.2	4.5	230

The activity of NADPH-Thioredoxin Reductase was tested as described in Material and Methods. NADP-MDH activity was tested as described in Material and Methods after DTT-dependent activation.

FIGURE LEGENDS

Fig.2: Fluorescence emission spectra of WT and P40A thioredoxins, in oxidized and reduced forms, at pH 7.2. The samples were excited at 280 nm and the emission was recorded from 290 to 400 nm. Reduction was performed with 5 mM DTT for 10 min.

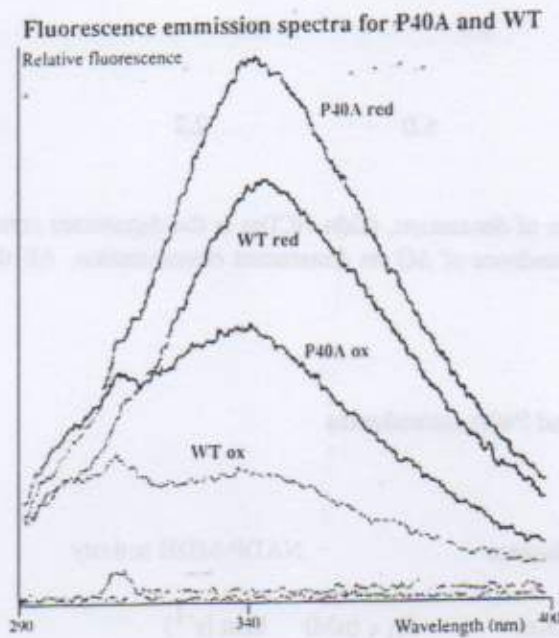


Fig.4: Thioredoxin-mediated light activation of NADP-MDH, at 25°C and under air. The activation was performed as described in Material and Methods, at saturating thioredoxin concentration (20 μ M).

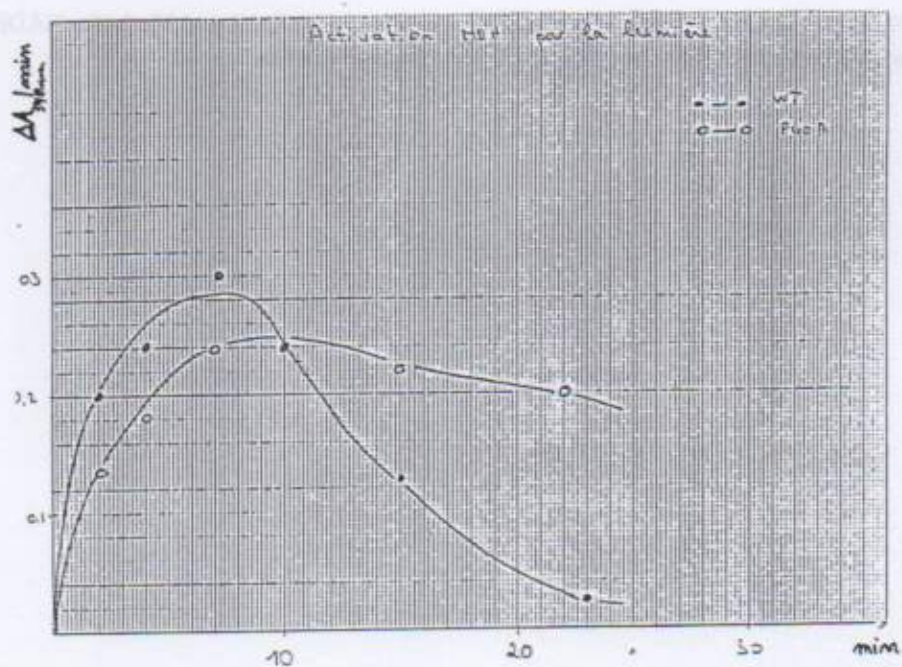


Fig.6: Representation of the average structures and the $\alpha 2$ helix shape in P40A with Ala in position 40 and in WT thioredoxin with Pro in position 40 after superimposition of the β sheets (rms=1.54 Å).

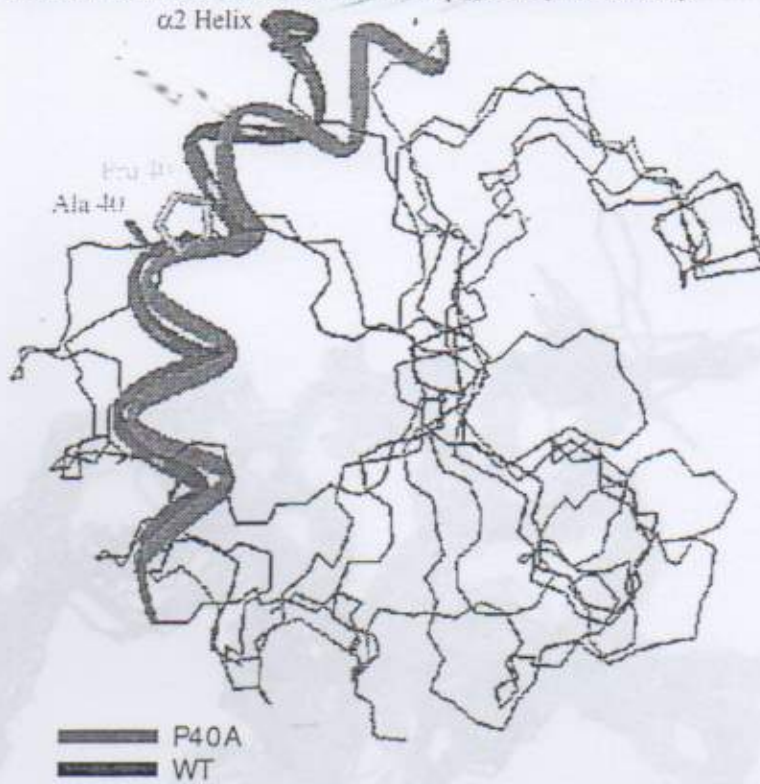


Fig.7: Overview of the superimposition of the active site in P40A (red) and WT (black) thioredoxins.

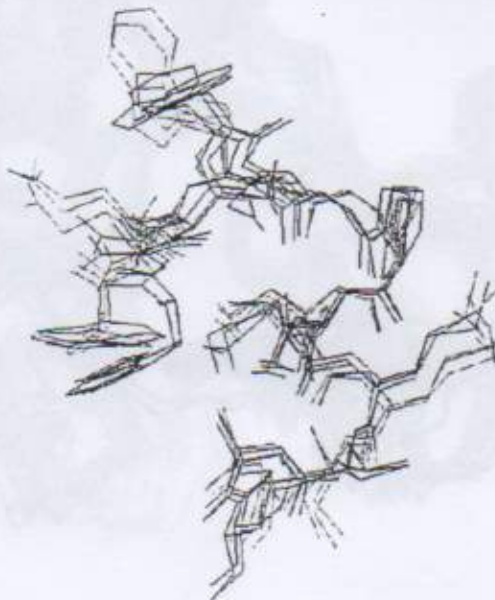


Fig.9: Superimposition of 10 structures of P40A (red) and WT (black) thioredoxins during the last 100 ps of the dynamic simulations, showing the difference of accessibility for Trp31.



I. Attribution de *Triticum Aestivum* :

attribution dans GIFA													
δppm		δppm		système de spins									
F1	F2	corrélation											
0,063	0,977	Hb-M2	1	5,345	2,164	Hb1-Ha	7	2,130	7,666	HN-Hb1	12		
-0,024	9,062	HN-M2	1	1,257	4,263	unk-unk	7	2,364	7,666	HN-Hb2	12		
0,063	9,062	HN-M1	1	1,543	5,345	Ha-Hd2	7	4,201	7,666	HN-Ha	12		
0,995	9,062	HN-Hb	1	1,873	5,345	Ha-Hd1	7	7,666	7,666	HN-HN	12		
4,488	9,062	HN-Ha	1	2,030	5,345	Ha-Hb2	7	3,619	3,619	Ha-Ha	13		
4,488	9,062	HN-Ha	1	2,164	5,345	Ha-Hb1	7	10,224	3,623	Ha-HN	13		
2,218	10,502	unk-unk	1	9,898	5,345	Ha-HN	7	1,752	4,176	unk-unk	13		
4,285	0,831	Ha-M1	2	1,543	9,898	HN-Hd2	7	4,867	8,364	unk-unk	13		
4,285	0,904	Ha-M2	2	5,345	9,898	HN-Ha	7	3,619	10,240	HN-Ha	13		
4,285	8,801	HN-Ha	2	9,898	9,898	HN-HN	7	1,576	1,055	Hb1-Hb2	14		
0,820	8,815	HN-M1	2	1,324	-0,104	H1-H2	8	4,334	1,055	Hb1-Ha	14		
0,878	8,815	HN-M2	2	4,750	-0,104	H1-Ha	8	1,055	1,576	Hb2-Hb1	14		
1,926	8,815	HN-Hb	2	9,136	-0,104	H1-HN	8	4,334	1,576	Hb2-Ha	14		
0,874	0,874	MET-MET	3	-0,104	1,324	H2-H1	8	1,140	4,176	unk-unk	14		
2,108	0,874	MET-HB	3	9,136	1,324	H2-HN	8	1,055	4,334	Ha-Hb1	14		
3,884	0,874	MET-HA	3	1,519	4,249	unk-unk	8	1,576	4,334	Ha-Hb2	14		
6,726	0,874	MET-HN	3	-0,104	9,136	HN-H1	8	4,896	8,321	unk-unk	14		
0,874	2,108	HB-MET	3	1,324	9,136	HN-H2	8	1,053	9,092	HN-Hb1	14		
3,884	2,108	HB-HA	3	4,750	9,136	HN-Ha	8	1,606	9,092	HN-Hb2	14		
6,726	2,108	HB-HN	3	9,136	9,136	HN-HN	8	4,334	9,106	HN-Ha	14		
0,874	3,884	HA-MET	3	1,431	0,977	M2-M1	9	9,106	9,106	HN-HN	14		
2,108	3,884	HA-HB	3	9,263	0,991	M2-HN	9	2,925	2,250	Hb1-Hb2	15		
6,726	3,884	HA-HN	3	9,263	1,427	M1-HN	9	4,828	2,250	Hb1-Ha	15		
2,742	4,307	unk-unk	3	0,966	1,456	M1-M2	9	9,015	2,250	Hb1-HN	15		
0,874	6,726	HN-MET	3	9,263	1,631	Hd-HN	9	2,250	2,925	Hb2-Hb1	15		
2,108	6,726	HN-HB	3	9,263	1,762	Hb-HN	9	4,828	2,925	Hb2-Ha	15		
3,884	6,726	HN-HA	3	5,034	5,001	Ha-Ha	9	9,015	2,925	Hb2-HN	15		
2,962	2,437	Hb2-Hb1	4	9,264	5,034	Ha-HN	9	2,683	4,132	unk-unk	15		
5,953	2,437	Hb2-Ha	4	0,978	9,264	HN-M2	9	2,250	9,015	HN-Hb1	15		
2,437	2,962	Hb1-Hb2	4	1,419	9,264	HN-M1	9	2,925	9,015	HN-Hb2	15		
5,953	2,962	Hb1-Ha	4	1,758	9,264	HN-Hb	9	4,828	9,015	HN-Ha	15		
9,643	2,962	Hb1-HN	4	5,034	9,264	HN-Ha	9	9,015	9,015	HN-HN	15		
2,654	4,307	unk-unk	4	9,264	9,264	HN-HN	9	3,954	1,786	Hb-Ha	16		
2,437	5,953	Ha-Hb2	4	1,606	9,266	HN-Hd	9	6,695	1,786	Hb-HN	16		
2,962	5,953	Ha-Hb1	4	2,814	2,237	Hb2-Hb1	10	1,786	3,954	Ha-Hb	16		
5,953	5,953	Ha-Ha	4	5,305	2,237	Hb2-Ha	10	6,695	3,954	Ha-HN	16		
9,643	5,953	Ha-HN	4	2,237	2,814	Hb1-Hb2	10	1,786	6,695	HN-Hb	16		
2,962	9,643	HN-Hb1	4	5,305	2,814	Hb1-Ha	10	3,954	6,695	HN-Ha	16		
5,953	9,643	HN-Ha	4	2,237	5,305	Ha-Hb2	10	4,375	1,773	H-Ha	17		
9,643	9,643	HN-HN	4	2,814	5,305	Ha-Hb1	10	7,485	1,773	H-HN	17		
2,358	1,093	other-HB	5	5,305	5,305	Ha-Ha	10	4,375	1,909	Hb2-Ha	17		
9,502	1,093	other-HN	5	9,216	5,305	Ha-HN	10	7,485	1,909	Hb2-HN	17		
1,093	2,358	HB-other	5	2,237	9,216	HN-Hb2	10	4,375	2,098	Hb1-Ha	17		
9,502	2,358	HB-HN	5	2,814	9,216	HN-Hb1	10	7,485	2,098	Hb1-HN	17		
3,557	4,263	unk-unk	5	5,305	9,216	HN-Ha	10	1,773	4,375	Ha-H	17		
1,093	9,502	HN-other	5	9,216	9,216	HN-HN	10	7,485	4,375	Ha-HN	17		
2,358	9,502	HN-HB	5	1,880	0,673	Hb2-Hb1	11	1,773	7,485	HN-H	17		
4,488	9,502	HN-HA	5	4,201	0,673	Hb2-Ha	11	1,773	7,485	HN-H	17		
9,502	9,502	HN-HN	5	9,053	0,673	Hb2-HN	11	1,773	7,485	HN-H	17		
5,539	1,524	Hb2-Ha	6	0,673	1,880	Hb1-Hb2	11	4,375	7,485	HN-Ha	17		
5,539	1,815	Hb1-Ha	6	9,053	1,880	Hb1-HN	11	7,485	7,485	HN-HN	17		
1,524	5,539	Ha-Hb2	6	0,673	4,201	Ha-Hb2	11	4,780	8,044	unk-unk	17		
1,781	5,539	Ha-Hb1	6	9,053	4,201	Ha-HN	11	4,099	1,928	Hb-Ha	18		
5,539	5,539	Ha-Ha	6	0,673	9,053	HN-Hb2	11	6,791	1,928	Hb-HN	18		
1,524	9,370	HN-Hb2	6	1,880	9,053	HN-Hb1	11	1,928	4,099	Ha-Hb	18		
1,781	9,370	HN-Hb1	6	4,201	9,053	HN-Ha	11	6,791	4,099	Ha-HN	18		
5,539	9,370	HN-Ha	6	4,201	1,960	H-Ha	12	1,928	6,791	HN-Hb	18		
9,370	9,370	HN-HN	6	7,666	1,960	H-HN	12	4,099	6,791	HN-Ha	18		
2,164	1,543	Hd2-Hd1	7	4,201	2,169	Hb1-Ha	12	3,539	1,342	Hb-Ha	19		
5,345	1,543	Hd2-Ha	7	7,666	2,169	Hb1-HN	12	6,991	1,342	Hb-HN	19		
9,898	1,543	Hd2-HN	7	4,201	2,364	Hb2-Ha	12	1,342	3,539	Ha-Hb	19		
5,345	1,873	Hd1-Ha	7	7,666	2,364	Hb2-HN	12	6,991	3,539	Ha-HN	19		
5,345	2,030	Hb2-Ha	7	1,960	4,201	Ha-H	12	0,937	4,132	unk-unk	19		
9,898	2,030	Hb2-HN	7	2,130	4,201	Ha-Hb1	12	1,342	6,991	HN-Hb	19		
1,543	2,164	Hd1-Hd2	7	2,364	4,201	Ha-Hb2	12	3,539	6,991	HN-Ha	19		
				4,201	4,201	Ha-Ha	12	4,896	7,564	unk-unk	19		
				7,666	4,201	Ha-HN	12	3,295	2,171	Hb-Ha	20		
				1,960	7,666	HN-H	12	6,986	2,171	Hb-HN	20		

2.171	3.330	Ha-Hb	20
6.986	3.330	Ha-HN	20
3.266	4.118	unk-unk	20
2.171	6.986	HN-Hb	20
3.295	6.986	HN-Ha	20
4.379	1.502	Hb-Ha	21
7.055*	1.502	Hb-HN	21
2.712	4.118	unk-unk	21
1.502	4.379	Ha-Hb	21
1.502	7.055	HN-Hb	21
4.379	7.055	HN-Ha	21
4.518	0.110	Hb-Ha	22
7.054	0.110	Hb-HN	22
2.509	4.118	unk-unk	22
0.110	7.054	HN-Hb	22
4.518	7.054	HN-Ha	22
4.750	7.114	unk-unk	22
2.201	0.422	Hg?-Hb?	23
7.025	0.422	Hg?-HN	23
0.422	2.201	Hb?-Hg?	23
7.025	2.201	Hb?-HN	23
2.392	4.118	unk-unk	23
0.422	7.025	HN-Hg?	23
2.201	7.025	HN-Hb?	23
2.062	0.931	Hb2-Hb1	24
3.908	0.931	Hb2-Ha	24
0.931	2.062	Hb1-Hb2	24
3.908	2.062	Hb1-Ha	24
7.237	2.062	Hb1-HN	24
0.931	3.908	Ha-Hb2	24
2.062	3.908	Ha-Hb1	24
7.237	3.908	Ha-HN	24
2.305	4.118	unk-unk	24
0.931	7.237	HN-Hb2	24
2.062	7.237	HN-Hb1	24
3.908	7.237	HN-Ha	24
4.438	1.422	Hb-Ha	25
7.245	1.422	Hb-HN	25
2.014	4.118	unk-unk	25
1.422	4.438	Ha-Hb	25
1.422	7.245	HN-Hb	25
4.438	7.245	HN-Ha	25
5.129	2.111	Hb1-Ha	26
8.157	2.111	Hb1-HN	26
2.712	2.125	Hb2-Hb1	26
5.135	2.688	Hb2-Ha	26
8.159	2.688	Hb2-HN	26
2.101	2.722	Hb1-Hb2	26
2.688	5.135	Ha-Hb2	26
8.159	5.135	Ha-HN	26
2.072	5.136	Ha-Hb1	26
2.101	8.146	HN-Hb1	26
2.688	8.159	HN-Hb2	26
5.135	8.159	HN-Ha	26
5.045	2.552	Hb-Ha	27
7.791	2.552	Hb-HN	27
1.664	4.118	unk-unk	27
2.552	5.045	Ha-Hb	27
7.791	5.045	Ha-HN	27
2.552	7.791	HN-Hb	27
5.045	7.791	HN-Ha	27
4.319	1.157	Hb-Ha	28
8.530	1.157	Hb-HN	28
1.157	4.319	Ha-Hb	28
8.530	4.319	Ha-HN	28
1.157	8.530	HN-Hb	28
4.319	8.530	HN-Ha	28
1.672	0.040	M2-M1	29
3.608	0.040	M2-Ha	29
0.040	1.672	M1-M2	29
3.608	1.672	M1-Ha	29
7.103	1.672	M1-HN	29

0.040	3.608	Ha-M2	29
1.672	3.608	Ha-M1	29
7.103	3.608	Ha-HN	29
1.024	4.118	unk-unk	29
0.040	7.103	HN-M2	29
1.672	7.103	HN-M1	29
3.608	7.103	HN-Ha	29
0.633	0.515	M1-M2	30
1.710	0.515	M1-Hb	30
4.780	0.515	M1-Ha	30
8.916	0.515	M1-HN	30
0.633	0.633	M2-M2	30
1.710	0.633	M2-Hb	30
4.780	0.633	M2-Ha	30
8.916	0.633	M2-HN	30
0.633	1.710	Hb-M2	30
0.633	1.710	Hb-M1	30
4.780	1.710	Hb-Ha	30
8.916	1.710	Hb-HN	30
2.276	4.103	unk-unk	30
0.633	8.916	HN-M2	30
0.633	8.916	HN-M1	30
1.710	8.916	HN-Hb	30
4.780	8.916	HN-Ha	30
0.500	8.917	HN-M1	30
2.017	0.882	M2-Hb	31
4.576	0.882	M2-Ha	31
8.995	0.882	M2-HN	31
2.017	0.974	M1-Hb	31
4.576	0.974	M1-Ha	31
8.995	0.974	M1-HN	31
0.882	2.017	Hb-M2	31
0.882	2.017	Hb-M1	31
4.576	2.017	Hb-Ha	31
8.995	2.017	Hb-HN	31
0.995	2.024	Hb-M1	31
2.159	4.103	unk-unk	31
0.878	8.990	HN-M2	31
0.966	8.990	HN-M1	31
2.017	8.995	HN-Hb	31
4.576	8.995	HN-Ha	31
4.794	2.733	Hb1-Ha	32
8.500	2.733	Hb1-HN	32
4.794	2.922	Hb2-Ha	32
8.500	2.922	Hb2-HN	32
2.733	8.500	HN-Hb1	32
2.922	8.500	HN-Hb2	32
4.794	8.500	HN-Ha	32
4.576	8.786	unk-unk	32
3.067	2.844	Hb1-Hb2	33
4.983	2.844	Hb1-Ha	33
7.345	2.844	Hb1-HN	33
2.844	3.067	Hb2-Hb1	33
4.983	3.067	Hb2-Ha	33
7.345	3.067	Hb2-HN	33
2.844	7.345	HN-Hb1	33
3.067	7.345	HN-Hb2	33
4.983	7.345	HN-Ha	33
0.560	-0.081	Hg2-Hg1	34
1.431	-0.081	Hg2-Hb	34
3.586	-0.081	Hg2-Ha	34
-0.112	0.560	Hg1-Hg2	34
1.431	0.560	Hg1-Hb	34
3.586	0.560	Hg1-Ha	34
7.810	0.560	Hg1-HN	34
-0.112	1.469	Hb-Hg2	34
0.560	1.469	Hb-Hg1	34
3.586	1.469	Hb-Ha	34
7.810	1.469	Hb-HN	34
-0.112	3.586	Ha-Hg2	34
0.560	3.586	Ha-Hg1	34
1.431	3.586	Ha-Hb	34

7.810	3.586	Ha-HN	34
-0.112	7.810	HN-Hg2	34
0.560	7.810	HN-Hg1	34
1.431	7.810	HN-Hb	34
3.586	7.810	HN-Ha	34
1.512	0.557	Hb-Hb2	35
3.471	0.557	Hb-Ha	35
8.174	0.557	Hb-HN	35
0.557	1.512	Hb2-Hb	35
3.471	1.512	Hb2-Ha	35
8.174	1.512	Hb2-HN	35
0.557	3.471	Ha-Hb	35
1.512	3.471	Ha-Hb2	35
8.174	3.471	Ha-HN	35
1.461	4.074	unk-unk	35
0.557	8.174	HN-Hb	35
1.512	8.174	HN-Hb2	35
3.471	8.174	HN-Ha	35
3.917	1.295	Me-Ha	36
8.004	1.295	Me-HN	36
1.295	3.917	Ha-Me	36
8.004	3.917	Ha-HN	36
1.315	4.045	unk-unk	36
1.295	8.004	HN-Me	36
3.917	8.004	HN-Ha	36
1.145	0.330	M2-M1	37
1.927	0.330	M2-Hb	37
3.561	0.330	M2-Ha	37
0.330	1.145	M1-M2	37
1.927	1.145	M1-Hb	37
3.561	1.145	M1-Ha	37
0.330	1.927	Hb-M2	37
1.145	1.927	Hb-M1	37
3.561	1.927	Hb-Ha	37
0.330	3.561	Ha-M2	37
1.145	3.561	Ha-M1	37
1.927	3.561	Ha-Hb	37
8.461	3.561	Ha-HN	37
2.975	4.031	unk-unk	37
0.330	8.461	HN-M2	37
1.145	8.461	HN-M1	37
1.927	8.461	HN-Hb	37
3.561	8.461	HN-Ha	37
8.461	8.461	HN-HN	37
0.623	0.462	M1-M2	38
1.178	0.462	M1-Hb	38
3.894	0.462	M1-Ha	38
8.687	0.462	M1-HN	38
0.462	0.623	M2-M1	38
1.178	0.623	M2-Hb	38
3.894	0.623	M2-Ha	38
8.687	0.623	M2-HN	38
0.462	1.178	Hb-M1	38
0.623	1.178	Hb-M2	38
3.894	1.178	Hb-Ha	38
8.687	1.178	Hb-HN	38
0.462	3.894	Ha-M1	38
0.623	3.894	Ha-M2	38
1.178	3.894	Ha-Hb	38
8.687	3.894	Ha-HN	38
2.363	4.031	unk-unk	38
0.462	8.687	HN-M1	38
0.623	8.687	HN-M2	38
1.178	8.687	HN-Hb	38
3.894	8.687	HN-Ha	38
1.962	0.860	M1-Hb	39
4.005	0.860	M1-Ha	39
8.412	0.860	M1-HN	39
1.962	0.927	M2-Hb	39
4.005	0.927	M2-Ha	39
8.412	0.927	M2-HN	39
0.927	1.962	Hb-M2	39

0,927	1,962	Hb-M2	39
4,005	1,962	Hb-Ha	39
0,927	4,005	Ha-M2	39
0,927	4,005	Ha-M2	39
1,962	4,005	Ha-Hb	39
8,412	4,005	Ha-HN	39
2,159	4,031	unk-unk	39
0,927	8,412	HN-M2	39
0,927	8,412	HN-M2	39
1,962	8,412	HN-Hb	39
4,005	8,412	HN-Ha	39
1,155	1,009	M1-M2	40
2,150	1,009	M1-Hb	40
4,119	1,009	M1-Ha	40
8,677	1,009	M1-HN	40
1,009	1,155	M2-M1	40
2,150	1,155	M2-Hb	40
4,119	1,155	M2-Ha	40
8,677	1,155	M2-HN	40
1,009	2,150	Hb-M1	40
1,155	2,150	Hb-M2	40
4,119	2,150	Hb-Ha	40
8,677	2,150	Hb-HN	40
1,461	4,031	unk-unk	40
1,009	4,119	Ha-M1	40
1,155	4,119	Ha-M2	40
2,150	4,119	Ha-Hb	40
8,677	4,119	Ha-HN	40
1,009	8,677	HN-M1	40
1,155	8,677	HN-M2	40
2,150	8,677	HN-Hb	40
4,119	8,677	HN-Ha	40
4,198	3,626	Ha1-Ha2	41
8,272	3,626	Ha1-HN	41
1,344	4,031	unk-unk	41
3,626	4,198	Ha2-Ha1	41
8,272	4,198	Ha2-HN	41
3,626	8,272	HN-Ha1	41
4,198	8,272	HN-Ha2	41
5,566	1,209	Hb1-HN	42
5,566	1,631	Hb2-HN	42
1,228	4,031	unk-unk	42
1,169	5,565	Ha-Hb1	42
1,606	5,565	Ha-Hb2	42
4,372	8,408	unk-unk	42
1,169	9,101	HN-Hb1	42
1,606	9,101	HN-Hb2	42
5,565	9,101	HN-Ha	42
8,152	3,463	Ha1-HN	43
4,246	3,734	Ha1-Ha2	43
3,295	4,016	unk-unk	43
3,734	4,246	Ha2-Ha1	43
8,152	4,246	Ha2-HN	43
3,734	8,152	HN-Ha1	43
4,246	8,152	HN-Ha2	43
1,353	1,353	M-M	44
4,911	1,353	M-Ha	44
8,073	1,353	M-HN	44
3,149	4,016	unk-unk	44
1,353	8,073	HN-M	44
4,911	8,073	HN-Ha	44
3,964	8,393	unk-unk	44
4,114	1,495	M-Ha	45
8,082	1,495	M-HN	45
2,072	4,016	unk-unk	45
1,495	4,114	Ha-M	45
8,082	4,114	Ha-HN	45
1,495	8,082	HN-M	45
4,114	8,082	HN-Ha	45
1,490	8,423	unk-unk	45
8,857	3,914	Ha-HN	46
8,857	3,914	Ha-HN	46

1,985	4,016	unk-unk	46
3,914	8,857	HN-Ha	46
8,857	3,623	Ha-HN	47
3,964	8,364	unk-unk	47
3,623	8,857	HN-Ha	47
4,459	1,408	Hb2-Ha	48
4,459	1,524	Hb1-Ha	48
1,408	7,957	HN-Hb2	48
1,408	7,957	HN-Hb2	48
4,459	7,957	HN-Ha	48
3,964	8,277	unk-unk	48
2,975	8,379	unk-unk	48
4,434	1,391	M-Ha	49
1,461	4,001	unk-unk	49
1,391	4,434	Ha-M	49
1,391	8,103	HN-M	49
4,434	8,103	HN-Ha	49
2,421	8,364	unk-unk	49
4,401	1,225	M-Ha	50
7,572	1,225	M-HN	50
1,140	4,001	unk-unk	50
1,225	7,572	HN-M	50
4,401	7,572	HN-Ha	50
1,373	8,350	unk-unk	50
3,532	0,695	M-Hb	51
4,503	0,695	M-Ha	51
0,695	3,532	Hb-M	51
4,503	3,532	Hb-Ha	51
2,392	3,987	unk-unk	51
0,695	8,263	HN-M	51
3,532	8,263	HN-Hb	51
4,503	8,263	HN-Ha	51
1,723	8,321	unk-unk	51
4,576	1,560	HB2-HA	52
8,542	1,560	HB2-HN	52
4,576	1,676	HB1-HA	52
8,542	1,676	HB1-HN	52
2,334	3,987	unk-unk	52
3,935	8,233	unk-unk	52
1,897	8,306	unk-unk	52
1,560	8,542	HN-HB2	52
1,560	8,542	HN-HB2	52
4,576	8,542	HN-HA	52
1,693	8,553	HN-Hb2	52
3,807	1,131	Hg-Hb	53
4,960	1,131	Hg-Ha	53
1,131	3,807	Hb-Hg	53
4,960	3,807	Hb-Ha	53
7,392	3,807	Hb-HN	53
2,101	3,987	unk-unk	53
7,392	4,960	Ha-HN	53
1,131	7,392	HN-Hg	53
3,807	7,392	HN-Hb	53
4,960	7,392	HN-Ha	53
3,761	8,219	unk-unk	53
1,868	8,233	unk-unk	53
0,950	0,950	M2-M2	54
0,950	1,062	M1-M2	54
2,041	1,062	M1-Hb	54
4,112	1,062	M1-Ha	54
0,950	2,041	Hb-M2	54
4,112	2,041	Hb-Ha	54
1,956	3,987	unk-unk	54
0,950	4,112	Ha-M2	54
0,950	4,112	Ha-M2	54
2,041	4,112	Ha-Hb	54
0,950	7,895	HN-M2	54
0,950	7,895	HN-M2	54
2,041	7,895	HN-Hb	54
4,112	7,895	HN-Ha	54
1,024	7,899	HN-M1	54
3,877	8,190	unk-unk	54

3,776	3,506	Hb1-Hb2	55
8,732	3,506	Hb1-HN	55
4,750	3,507	Ha-Hb2	55
4,750	3,769	Ha-Hb1	55
3,506	3,776	Hb2-Hb1	55
8,732	3,776	Hb2-HN	55
3,702	3,972	unk-unk	55
4,750	8,728	HN-Ha	55
3,506	8,732	HN-Hb1	55
3,776	8,732	HN-Hb2	55
5,645	1,117	H-Ha	56
3,237	3,972	unk-unk	56
1,117	5,645	Ha-H	56
2,392	8,146	unk-unk	56
4,259	1,377	Hb1-Ha	57
4,259	1,757	Hb2-Ha	57
2,247	3,972	unk-unk	57
1,377	4,259	Ha-Hb1	57
1,757	4,259	Ha-Hb2	57
2,334	8,146	unk-unk	57
1,377	8,641	HN-Hb1	57
1,757	8,641	HN-Hb2	57
4,259	8,641	HN-Ha	57
4,838	1,575	M-Ha	58
1,897	3,972	unk-unk	58
2,276	8,146	unk-unk	58
1,575	8,641	HN-M	58
4,838	8,641	HN-Ha	58
4,143	2,048	M-Ha	59
8,625	2,048	M-HN	59
1,140	3,972	unk-unk	59
2,048	4,143	Ha-M	59
8,625	4,143	Ha-HN	59
2,048	8,625	HN-M	59
4,143	8,625	HN-Ha	59
4,576	2,806	Hb1-Ha	60
8,782	2,806	Hb1-HN	60
4,576	2,913	Hb2-Ha	60
8,782	2,913	Hb2-HN	60
3,091	3,958	unk-unk	60
2,913	8,782	HN-Hb2	60
2,913	8,782	HN-Hb2	60
4,576	8,782	HN-Ha	60
4,459	3,679	Ha1-Ha2	61
8,379	3,679	Ha1-HN	61
2,945	3,958	unk-unk	61
4,023	8,001	unk-unk	61
3,679	8,379	HN-Ha1	61
4,459	8,379	HN-Ha2	61
1,286	1,296	MET-MET	62
4,518	1,296	MET-HA	62
8,477	1,311	MET-HN	62
2,421	2,421	hb2	62
4,866	2,421	unk	62
8,378	2,421	hn	62
2,945	2,945	unk	62
8,378	2,945	hn	62
2,596	3,958	unk-unk	62
2,421	4,866	unk	62
2,945	4,866	unk	62
4,866	4,866	unk	62
8,378	4,866	unk	62
8,378	8,378	hn	62
1,286	8,495	HN-MET	62
4,518	8,495	HN-HA	62
8,477	8,510	HN-HN	62
4,081	0,918	MET-HA	63
0,907	4,089	HA-MET	63
8,128	4,103	HA-HN	63
4,139	7,972	unk-unk	63
0,937	8,132	HN-MET	63
4,081	8,132	HN-HA	63

7,545	1,006	HB2-HN	64
1,985	1,020	HB2-HB1	64
4,430	1,020	HB2-HA	64
0,995	1,995	HB1-HB2	64

4,459	1,995	HB1-HA	64
7,545	1,995	HB1-HN	64
1,635	3,958	unk-unk	64
0,995	7,535	HN-HB2	64

1,985	7,535	HN-HB1	64
4,430	7,550	HN-HA	64
1,577	8,030	unk-unk	64

1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			

1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			

1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			

Chapitre IX. Références:

- Abragam, A. (1961) Principles of nuclear magnetism Oxford University Press, London
- Aota, M., Matsuda, K., Isowa, N., Wada, H., Yodoi, J.J. and Ban, T. (1996) *J Cardiovasc Pharmacol*, 27, 727-732.
- Beckman, R.A., Litwin, S. and Wand, A.J. (1993) *J Biomol Nmr*, 3, 675-700.
- Biguet, C., Wakasugi, N., Mishal, Z., Holmgren, A., Chouaib, S., Tursz, T. and Wakasugi, H. (1994) *J Biol Chem*, 269, 28865-28870.
- Biosym (1993)
- Bloch, F., Hansen, W.W. and Packard, M. (1946) *Phys. Rev.*, 69, 127.
- Blombäck, B., Adamson, L., Hessel, B., Homlgren, A., Hogg, D. and Procyk, R. (1986) Raven Press, New York
- Blumenthal, L. (1953) Theory and applications of distance geometry Cambridge Univ. Press, Cambridge
- Brugidou, C., Marty, I., Chartier, Y. and Meyer, Y. (1993) *Mol Gen Genet*, 238, 285-293.
- Brünger, A.T. (1988) X-PLOR 3.1 a system for X-Ray Crystallography and NMR Yale University, New Haven CT 06511 USA
- Buchanan, B.B. (1991) *Arch Biochem Biophys*, 288, 1-9.
- Buchanan, B.B., Schurmann, P., Decottignies, P. and Lozano, R.M. (1994a) *Arch Biochem Biophys*, 314, 257-260.
- Buchanan, B.B., Schürmann, P., Decottignies, P. and Lozano, R.M. (1994b) *Archives of Biochemistry and Biophysics*, 314, 257-260.
- Cai, M.L., Liu, J.H., Gong, Y.X. and Krishnamoorthi, R. (1995) *J Magn Resonance Ser B*, 107, 172-178.
- Castanho, M.A. and Prieto, M.J. (1992) *Eur J Biochem*, 207, 125-134.
- Cayley, A. (1841) *J. Cambridge Math.*, II, 267-271.
- Chergui, J. (1993) Bibliothèque LAPACK (Manuel LAPACK) CIRCE, CNRS,
- Constantine, K.L., Friedrichs, M.S. and Mueller, L. (1994) *J Magn Resonance Ser B*, 104, 62-68.
- Cregg, J.M., Vedvick, T.S. and Raschke, W.C. (1993) *Bio/Technology*, 11, 905-909.
- Crippen, G.M. and Havel, T.F. (1988) Distance Geometry and Molecular Conformation John Wiley & Sons Inc., New York
- Darling, A.J. (1988) *J Gen Virol*, 69, 515-523.
- Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) Atlas of Protein Sequence and Structure, 5, 345-
- de Lamotte-Guery, F., Miginiac-Maslow, M., Decottignies, P., Stein, M., Minard, P. and Jacquot, J.P. (1991) *Eur J Biochem*, 196, 287-294.
- Delsuc, M.-A. (1996)
- Dyson, H.J., Holmgren, A. and Wright, P.E. (1988) *FEBS Lett*, 228, 254-258.
- Dyson, H.J., Holmgren, A. and Wright, P.E. (1989) *Biochemistry*, 28, 7074-7087.
- Dyson, H.J., Jeng, M.F., Model, P. and Holmgren, A. (1994) *FEBS Lett*, 339, 11-17.
- Easthope, P.L. and Havel, T.F. (1989) *Bulletin of Mathematical Biology*, 51, 173-194.
- Eisenberg, D. and McLachan, A.D. (1986) *Nature*, 319, 199-203.
- Eklund, H., Cambillau, C., Sjöberg, B.M., Holmgren, A., Jornvall, H., Hoog, J.O. and Branden, C.I. (1984) *Embo J*, 3, 1443-1449.
- Emsley, J., Fenney, J. and Sutcliffe, L. (1966-73) Progress in NMR spectroscopy Pergamon, London
- Forman-Kay, J.D., Clore, G.M., Stahl, S.J. and Gronenborn, A.M. (1992) *J Biomol NMR*, 2, 431-445.

- Forman-Kay, J.D., Clore, G.M., Wingfield, P.T. and Gronenborn, A. (1991) *Biochemistry*, 30, 2685-2698.
- Fredman, M.L. and Tarjan, R.E. (1987) *Journal of the Association for Computing Machinery*, 34, 596-615.
- Geourjon, C. and Deleage, G. (1993) *CABIOS*, 9, 87-91.
- Geourjon, C. and Deleage, G. (1994) *Protein Engineering*, 7, 157-164.
- Gibrat (1987) *J. Mol. Biol.*, 198, 425-443.
- Glunt, W., Hayden, T.L. and Raydan, M. (1993) *Journal of Computational Chemistry*, 14, 114-120.
- Gray, M.W. (1996) *Nature*, 383, 299-300.
- Havel, T.F. (1991) *Proq. Biophys. Molecular Biology*, 56, 43-78.
- Hellinga, H.W., Wynn, R. and Richards, F.M. (1992) *Biochemistry*, 31, 11203-11209.
- Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1991)
- Hill, J.M., Alewood, P.F. and Craik, D.J. (1996) *Biochemistry*, 35, 8824-8835.
- Hockney, R.W. and Eastwood, J.M. (1981) *Computer simulations using particules* McGraw-Hill, New York
- Holmgren, A. (1972) *The Journal of Biological Chemistry*, 247, 1992-1998.
- Holmgren, A. (1985) *Ann. Rev. Biochem.*, 54, 237-271.
- Holmgren, A. and Roberts, G. (1976) *FEBS Lett*, 71, 261-265.
- Hopper, S. and Iurlano, D. (1983) *J Biol Chem*, 258, 13453-13457.
- Issakidis, E., Decottignies, P. and Miginiac-Maslow, M. (1993) *FEBS Lett*, 321, 55-58.
- Jacquot, J.P., Gadal, P., Nishizawa, A.N., Yee, B.C., Crawford, N.A. and Buchanan, B.B. (1984) *Arch Biochem Biophys*, 228, 170-178.
- Jeng, M.F., Campbell, A.P., Begley, T., Holmgren, A., Case, D.A., Wright, P.E. and Dyson, H.J. (1994) *Structure*, 2, 853-868.
- Jeng, M.F. and Dyson, H.J. (1996) *Biochemistry*, 35, 1-6.
- Kallis, G.-B. and Holmgren, A. (1980) *The Journal of Biological Chemistry*, 255, 10261-10265.
- Katti, S.K., LeMaster, D.M. and Eklund, H. (1990) *Biochemistry*, 29, 167-184.
- Kessler, H., Gehrke, M. and Griesinger, C. (1988) *Ang. Chem.*, 27, 490-536.
- Knut, L., A., F.J. and Clare, W. (1991) *Biochemistry*, 30, 7603-7609.
- Kraulis, P.J. (1994) *J Mol Biol*, 243:4, 696-718.
- Ladbury, J.E., Kishore, N., Hellinga, H.W., Wynn, R. and Sturtevant, J.M. (1994) *Biochemistry*, 33, 3688-3692.
- Ladbury, J.E., Wynn, R., Hellinga, H.W. and Sturtevant, J.M. (1993) *Biochemistry*, 32, 7526-7530.
- Lancelin, J.M., Stein, M. and Jacquot, J.P. (1993) *J Biochem (Tokyo)*, 114, 421-431.
- Laskowski, R.A., MacArthur, M., Moss, D.S. and Thornton, J.M. (1993) *Appl. Cryst.*, 26, 283-291.
- Lattman, E.E. (1994) *Proteins: Structure, Function and Genetics*, 18, 103-106.
- Le Cam, E., Frechon, D., Barray, M., Fourcade, A. and Delain, E. (1994) *Proc Natl Acad Sci U S A*, 91, 11816-11820.
- Le Marechal, P., Hoang, B.M., Schmitter, J.M., Van Dorsselaer, A. and Decottignies, P. (1992) *Eur J Biochem*, 210, 421-429.
- Levin (1986) *Febs Lett.*, 205, 303-308.
- Liu, Y., Zhao, D., Altman, R. and Jardetzky, O. (1992) *Journal of Biomolecular NMR*, 2, 373-388.
- Lu, Z.J., Diblasiosmith, E.A., Grant, K.L., Warne, N.W., Lavallie, E.R., Collinsracie, L.A., Follettie, M.T., Williams, M.J. and McCoy, J.M. (1996) *J Biol Chem*, 271, 5059-5065.
- Lunn, C.A. and Pigiet, V.P. (1982) *J Biol Chem*, 257, 11424-11430.
- Masutani, H., Nakamura, H., Ueda, Y., Kitaoka, Y., Kawabe, T., Iwata, S., Mitsui, A. and Yodoi, J. (1992) *Adv Exp Med Biol*, 319, 265-274.

- Matthews, J.R., Wakasugi, N., Virelizier, J.L., Yodoi, J. and Hay, R.T. (1992) *Nucleic Acids Res.*, 20, 3821-3830.
- Menger, K. (1928) *Ann. Math.*, 100, 75-163.
- Menger, K. (1931) *Am. J. Math.*, 53, 721-745.
- Mirandavizuete, A., Martinezgalisteo, E., Aslund, F., Lopezbarea, J., Pueyo, C. and Holmgren, A. (1994) *J Biol Chem*, 269, 16631-16637.
- Nagayama, K., Wüthrich, K., Bachmann, P. and Ernst, R.R. (1977) *Biochem. Biophys. Res Comm.*, 78, 99-105.
- Nikkola, M., Gleason, F.K., Fuchs, J.A. and Eklund, H. (1993) *Biochemistry*, 32, 5093-5098.
- Nisbet, E.G. and Fowler, C.M.R. (1996) *Nature*, 382, 404-405.
- Nygren, H., Rozell, B., Holmgren, A. and Hansson, H.A. (1981) *FEBS Lett*, 133, 145-150.
- Oppenheimer, N.J. and James, T.L. (1989) *Nuclear Magnetic Resonance, Structure and Mechanism* Harcourt Brace Jovanovich, San Diego
- Oshiro, C.M. and Kuntz, I.D. (1993) *Biopolymers*, 33, 107-115.
- Pardi, A., Billeter, M. and Wüthrich, K. (1984) *J. Mol. Biol.*, 180, 741-751.
- Pastore, A.e.a. (1991) *Proteins: Structure, function and genetics*, 10, 22-32.
- Pearlman, D.A., Case, D.A., Caldwell, J.C., Seibel, G.L., Singh, U.c., Paul, P.W. and Peter, A.K. (1991).
- Pla, A. and Lopez-Gorge, J. (1981) *Biochim Biophys Acta*, 636, 113-118.
- Polshakov, V.I., Frenkiel, T.A., Birdsall, B., Soteriou, A. and Feeney, J. (1995) *J Magn Resonance Ser B*, 108, 31-43.
- PRUVOST, C. (1992)
- Purisima, E.O. and Scheraga, H.A. (1986) *Proc. Natl. Acad. Sci.*, 83, 2782-2786.
- Qin, J., Clore, G.M. and Gronenborn, A.M. (1996a) *Biochemistry*, 35, 7-13.
- Qin, J., Clore, G.M., Kennedy, W.M., Huth, J.R. and Gronenborn, A.M. (1995) *Structure*, 3, 289-297.
- Qin, J., Clore, G.M., Kennedy, W.P., Kuszewski, J. and Gronenborn, A.M. (1996b) *Structure*, 4, 613-620.
- Raj, N., Morley, S.D. and Jackson, D.E. (1994) *Theochem-J Mol Struct*, 114, 175-190.
- Reutimann, H., Straub, B., Luisi, P.L. and Holmgren, A. (1981) *J Biol Chem*, 256, 6796-6803.
- Reynolds, A.E., Chesnick, J.M., Woolford, J. and Cattolico, R.A. (1994) *Plant Mol Biol*, 25, 13-21.
- Salamon, Z., Tollin, G., Hirasawa, M., Gardetsalvi, L., Strittetter, A.L., Knaff, D.B. and Schurmann, P. (1995) *Bba-Bioenergetics*, 1230, 114-118.
- Sali, A. (1994) *Modeller (A protein structure modelling program)* The Rockefeller University, New York
- Schallreuter, K.U., Wood, J.M., Farwell, D.W., Moore, J. and Edwards, H.G.M. (1996) *J Invest Dermatol*, 106, 583-586.
- Schenk, H., Klein, M., Erdbrugger, W., Droge, W. and Schulze-Osthoff, K. (1994) *Proc Natl Acad Sci U S A*, 91, 1672-1676.
- Schoenberg, I.J. (1935) *Annals Math*, 36, 724-732.
- Schulzeosthoff, K., Schenk, H. and Droge, W. (1995) In *Biothiols, Pt B Vol.252* (Ed. Packer, L.) Academic Press Inc, 525 B Street, Suite 1900, San Diego, CA 92101-4495, pp. 253-264
- Schurmann, P. *Methods Enzymol* 1995 ;252:274-283,
- Singer, T.P. and Johnson, M.K. (1985) *FEBS Lett*, 190, 189-198.
- Sippl, M.J. and Scheraga, H.A. (1985) *Proc. Natl. Acad. Sci.*, 84, 2197-2201.
- Sippl, M.J. and Scheraga, H.A. (1986) *Proc. Natl. Acad. Sci.*, 83, 2283-2287.
- Siuzdak, G. (1994) *Proc Natl Acad Sci U S A*, 91, 11290-11297.

- Slaby, I., Cerna, V., Jeng, M.F., Dyson, H.J. and Holmgren, A. (1996) *J Biol Chem*, 271, 3091-3096.
- Spurej, E., Glatter, O. and Pfeiler, G. (1992) *Experientia*, 48, 71-79.
- Stone, M.J., Chandrasekhar, K., Holmgren, A., Wright, P.E. and Dyson, H.J. (1993) *Biochemistry*, 32, 426-435.
- Stryer, L., Holmgren, A. and Reichard, P. (1967) *Biochemistry*, 6, 1016-1020.
- Udvardy, J., Borbely, G., Juhasz, A. and Farkas, G.L. (1984) *J Bacteriol*, 157, 681-683.
- Vargas, C., Wu, G., Davies, A.E. and Downie, J.A. (1994) *J Bacteriol*, 176, 4117-4123.
- Wagner, U.G., Muller, N., Schmitzberger, W., Falk, H. and Kratky, C. (1995) *J. Mol. Biol.*, 247, 326-337.
- Wilson, N.A., Barbar, E., Fuchs, J.A. and Woodward, C. (1995) *Biochemistry*, 34, 8931-8939.
- Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, 222, 311-333.
- Wong, J.H., Kobrehel, K. and Buchanan, B.B. (1995) In *Biothiols*, Pt B Vol.252 (Ed. Packer, L.) Academic Press Inc, 525 B Street, Suite 1900, San Diego, CA 92101-4495, pp. 228-240
- Yasukawa, T., Kaneiishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T. and Ishii, S. (1995) *J Biol Chem*, 270, 25328-25331.



[Faint, illegible text or bleed-through from the reverse side of the page]

Vu et permis d'imprimer

Montpellier, le

Le président de l'Université MONTPELLIER I

YVES LOUBATIERES
