



THESIS / THÈSE

MASTER EN SCIENCES BIOLOGIQUES

Réalisation d'un modèle tridimensionnel de l'alcool déshydrogénase I de la levure *Saccharomyces cerevisiae*

Cronet, Philippe

Award date:
1990

Awarding institution:
Universite de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1990



FACULTÉS UNIVERSITAIRES N.D. DE LA PAIX
NAMUR
FACULTÉ DES SCIENCES

Réalisation d'un modèle tridimensionnel de l'alcool
déshydrogénase I de la levure *Saccharomyces*
cerevisiae.

Mémoire présenté pour l'obtention du grade
de Licencié en Sciences
biologiques
par

Philippe Cronet

Facultés Universitaires Notre-dame de la Paix
FACULTE DES SCIENCES
Rue de Bruxelles, 61 -5000 Namur
Tél 081/ 72 41 11 - Telex 59222 Facnam-b - Telefax 081/23 03 91

**"Réalisation d'un modèle tridimensionnel de l'alcool
déshydrogénase I de la levure *Saccharomyces
cerevisiae*."**

CRONET Philippe

Résumé

Un modèle tridimensionnel de l'alcool déshydrogénase I de *Saccharomyces cerevisiae* est établi à partir de la structure tridimensionnelle de l'alcool déshydrogénase de foie de cheval et selon l'alignement des structures primaires proposé par Jörnvall *et al* (J. M. Biol. 253: 8414-8419, 1978). L'énergie potentielle de la structure est minimisée de manière à obtenir une conformation stable représentative d'un état donné de la protéine. Le modèle obtenu permet la proposition d'hypothèses quant à la structure quaternaire et les propriétés catalytiques de l'enzyme. Il servira également de base à de futures études fondamentales de la structure tridimensionnelle de l'enzyme ainsi qu'à des applications dans le domaine de l'ingénierie enzymatique.

Mémoire de licence en Sciences biologiques

Juin 1990

Promoteur: E. Feytmans

Avant de clôturer ce travail (ou presque...), j'exprime toute ma gratitude à Mr le Professeur Feytmans pour son accueil chaleureux et sa disponibilité tout au long de cette année académique. La compétence dont il fait preuve dans de nombreux domaines m'a été très précieuse à maintes reprises. Je lui suis également très reconnaissant de m'avoir permis de découvrir un domaine encore peu connu de la biologie moléculaire et pour les connaissances que j'ai acquises lors de la réalisation de ce mémoire. Merci enfin de m'avoir donné l'envie de poursuivre ma formation de biologiste.

Je voudrais également remercier Carla pour ses précieux conseils. Son dynamisme, son organisation et sa collaboration aussi efficace qu'agréable m'ont souvent permis de progresser plus rapidement. Merci à mes parents et à Ingrid pour leur soutien durant ces quatre années d'études.

Je remercie également Mr J. M. André, Mr Jean Delcour, Dominique Delforge, Eric Depiereux, Pierre Hayen, Mr L. Hevesi, Etienne Magonet, Melle D. Prozzi, Mr D. Vercauteren pour leur contribution à ce travail.

Abbréviations utilisées

aa	: acide aminé
C α	: carbone alpha
C β	: carbone bêta
G	: énergie libre
S	: entropie
H	: enthalpie
ϕ	: angle de torsion phi
ψ	: angle de torsion psi
ω	: angle de torsion oméga
VDW	: Van der Waals
Å	: Angström (10 ⁻¹⁰ m)
ΔG_{prot}	: Variation d'énergie libre de la chaîne polypeptidique lors du repliement de la protéine
ΔG_{trans}	: variation d'énergie libre lors du transfert d'un aa depuis un milieu hydrophile vers un milieu hydrophobe
ΔG_{solv}	: Variation d'énergie libre du solvant lors du repliement de la protéine
PIR	: Protein Information Ressources
PDB	: Protein Data Bank
EMBL	: European Molecular Biology Laboratory
RuBP carboxylase	: Ribulose Bi-phosphate Carboxylase
μ_s	: moment hydrophobique
Ep	: énergie potentielle
F	: force
m	: masse
a	: accélération
NAD ⁺	: Nicotinamide Adénine Dinucléotide oxydé
NADH	: Nicotinamide Adénine Dinucléotide réduit
YADH	: Alcool déshydrogénase I de <i>Saccharomyces cerevisiae</i>
HADH	: Alcool déshydrogénase de foie de cheval.
Zn ⁺⁺	: Ion zinc 2 * positif
CVFF	: Consistent Valence Force Field
AMBER	: Assisted model Building With Energy Refinement.

TABLE DES MATIERES

page

I. ANALYSE BIBLIOGRAPHIQUE:	1
I.1 AVANT PROPOS:	1
I.2 CONFORMATION DES PROTEINES:	3
I.2.1 LA STRUCTURE PRIMAIRE:	3
I.2.2 REPLIEMENT DES CHAINES POLYPEPTIDIQUES:	3
I.2.2.1 Les interactions covalentes (Biosym-Discover,89)	4
a)Un angle de valence	4
b)Un angle de torsion	4
c)Un angle "Out-of-plane"	5
I.2.2.2 Les interactions non covalentes	5
a)Les interactions électrostatiques:	5
b)Les ponts hydrogène	5
c)Les interactions de Van der Waals (VDW):	6
d)L'hydrophobicité:	6
I.2.2.3 Comment varie l'énergie libre au cours du repliement?	7
I.2.2.4 Structure native: minimum local ou global?	8
I.2.3 DIFFERENTS NIVEAUX DE REPLIEMENT:	9
I.2.3.1 Les structures secondaires:	9
a) Les hélices α :	9
b) Les plans β :	10
c)Les turns:	10
I.2.3.2 La structure tertiaire	11
I.2.3.3 Certaines protéines sont polymériques.	11
I.2.4 CONCLUSIONS	13
I.3 ETUDE ET MODELISATION DES PROTEINES:	14
I.3.1 LA CRISTALLOGRAPHIE AUX RAYONS X:	14
I.3.2 ANALYSES DE STRUCTURES TRIDIMENSIONNELLES:	14
I.3.2.1 Les graphes de Ramachandran	15
I.3.2.2 Distribution des résidus au sein de structures secondaires	15
I.3.2.3 "C α - Contact maps"	15

I.3.2.4 Profil d'hydrophobicité:	16
I.3.2.5 Les moments hydrophobiques:	16
I.2.3.6 Surface externe de la protéine	17
I.3.3 PREDICTION DE STRUCTURES TRIDIMENSIONNELLES:	17
I.3.4. MODELISATIONS DE STRUCTURES TRIDIMENSIONNELLES	18
I.3.4.1 Le Force field	19
I.3.4.2 La minimisation d'énergie:	20
a)"Steepest descent":	22
b)Le "conjugate gradient":	23
c)Autres méthodes:	23
d)Limites d'application	23
I.3.4.3 Méthode de Monte Carlo:	23
I.3.4.4 Utilisation de la dynamique moléculaire:	24
I.3.5 MODELISATION PAR SIMILARITE DE STRUCTURE AVEC DES PROTEINES ANALYSEES PAR CRISTALLOGRAPHIE:	26
I.3.5.1 Les alignements de séquences:	26
a)Les alignements simples:	27
i-Les "dot plots":	28
ii- Méthode de Wilbur et Lipman	28
b)Les alignements multiples	29
i-La méthode de Needleman&Wunsch	29
ii La méthode de Johnson et Doolittle	29
iii- La méthode de Gribskov et al	29
iv-Méthode de E. Depiereux et E. Feytmans	30
I.3.5.2 Les alignements de structures tridimensionnelles:	30
I.3.5.3 Construction du modèle:	30
I.4 L'ALCOOL DESHYDROGENASE	31
I.4.1 PLACE DANS LE METABOLISME:	31
I.4.2 L'ALCOOL DESHYDROGENASE DE <i>Saccharomyces cerevisiae</i> :	31

I.4.2.1 L'YADH I:	32
a)Structure primaire:	32
b)Mécanisme réactionel:	32
c)Spécificité de l'enzyme:	33
I.4.3 L'ALCOOL DESHYDROGENASE DE FOIE DE CHEVAL:	33
I.4.3.1 L'apoenzyme	33
a)Le domaine catalytique:	33
b)Le domaine de fixation du coenzyme:	34
I.4.3.2 L'holoenzyme:	34
a)Changements induits par la fixation du coenzyme:	34
b)Etude de la fixation du coenzyme :	34
c)Fixation du substrat:	35
d)Mécanisme réactionnel	35
e)Stéréospécificité de la réaction:	35
I.5 IDEE GENERALE DU TRAVAIL:	36
I.5.1 MECANISME CATALYTIQUE:	36
I.5.2 PROPRIETES ENZYMATIQUES:	36

II MATERIEL ET METHODES:	37
---------------------------------	-----------

II.1 SUPPORT INFORMATIQUE:	37
II.2 BANQUES DE DONNEES:	37
II.3 LOGICIELS DE MODELISATION:	37
II.3.1 INSIGHT (VERSION II):	37
II.3.2 MOLEDT:	38
II.3.3 DISCOVER	38
II.3.3.1 Le Force Field:	38
II.4 ALIGNEMENTS DE SEQUENCES:	39

II.5	MODELISATION DE L'APOENZYME DE LEVURE	39
II.6	MODELISATION DES MOLECULES ANNEXES:	40
II.6.1	LE COENZYME:	40
II.6.2	LE SUBSTRAT:	40
II.6.3	L'EAU CONSTITUTIVE:	41
II.7	CONSTRUCTION DE L'HOLOENZYME:	41
II.8	LA MINIMISATION D'ENERGIE:	42
II.9	ETAPES DE LA MINIMISATION:	45
II.10	MINIMISATION DE L'HADH:	47
II.11	VALIDATION DU MODELE:	48
II.11.1	CALCUL DU COMPACTAGE DES RESIDUS	48
II.11.2	CALCUL DES MOMENTS HYDROPHOBQUES	48

III	RESULTATS ET DISCUSSIONS:	49
------------	----------------------------------	----

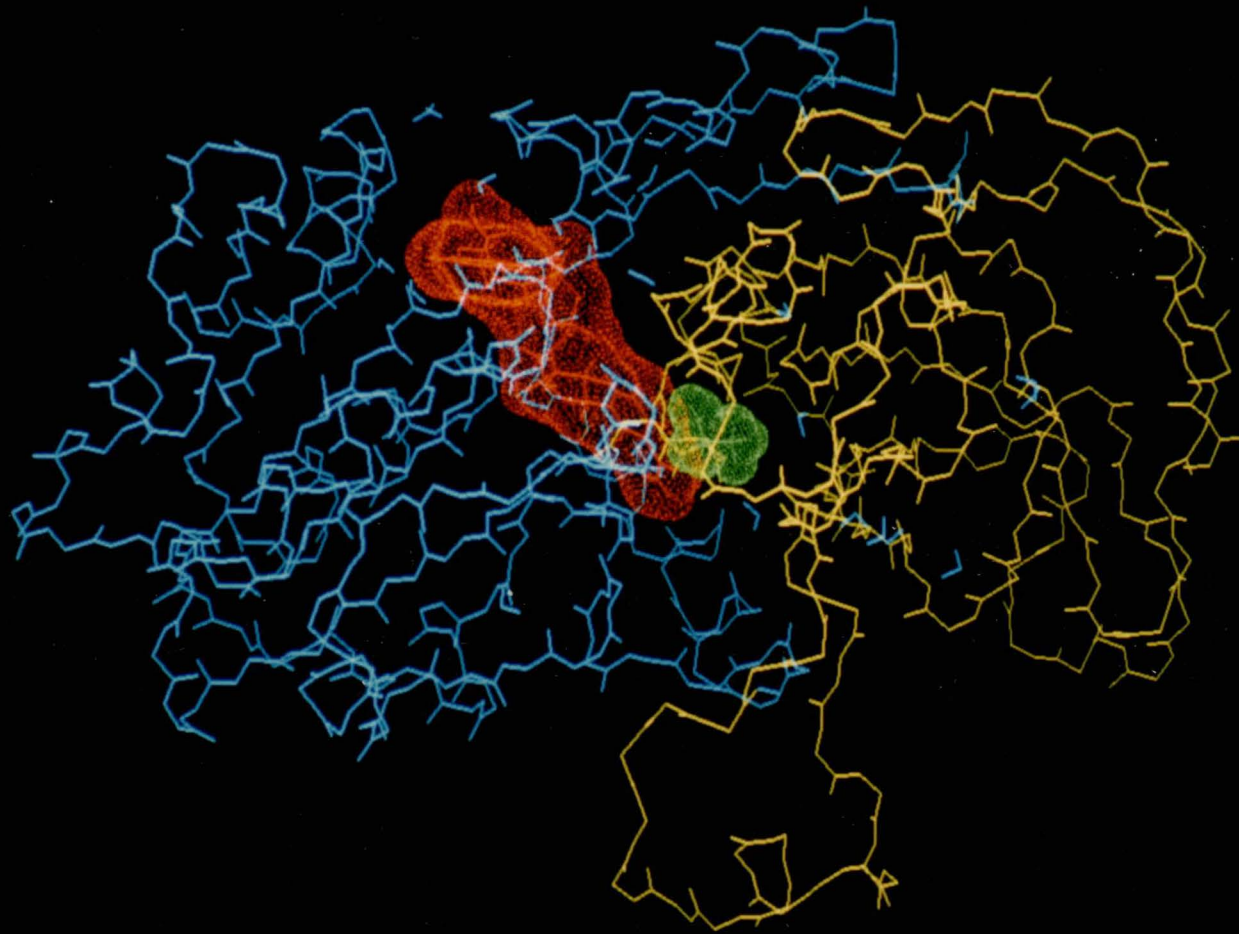
<u>III.1</u>	<u>LES ALIGNEMENTS:</u>	49
III.1.1	<u>L'ALIGNEMENT DE JÖRNVALL:</u>	49
III.1.2	<u>AUTRES ALIGNEMENTS:</u>	50
<u>III.2</u>	<u>MODELE AVANT MINIMISATION:</u>	
<u>III.3</u>	<u>RELAXATION DE LA MOLECULE:</u>	51
<u>III.4</u>	<u>MINIMISATION DES LIAISONS COVALENTES:</u>	52
<u>III.5</u>	<u>VERIFICATION DES ANGLES DE TORSION:</u>	54
<u>III.6</u>	<u>LIBERATION PROGRESSIVE DES ATOMES FIXES:</u>	56
<u>III.7</u>	<u>MINIMISATION DU SITE ACTIF:</u>	58
<u>III.8</u>	<u>MINIMISATION DE L'HADH:</u>	58

<u>III.9 VALIDATION DU MODELE:</u>	60
<u>III.9.1.CORRELATION ENTRE LES VECTEURS RADIAUX D'HYDROPHOBICITE:</u>	60
<u>III.9.2 VERIFICATION DU COMPACTAGE:</u>	61
<u>III.10 EXPLOITATION DU MODELE:</u>	61
<u>III.10 1. COMPARAISON DES SITES ACTIFS:</u>	61
<u>III.10. 2 MODELISATION DU PEPTIDE OBTENU PAR LYSE TRYPSINIQUE:</u>	63
<u>III.10. 3. STRUCTURE QUATERNAIRE:</u>	63

IV PERSPECTIVES ET CONCLUSIONS:	64
--	----

<u>IV. I CONCLUSIONS GENERALES:</u>	65
--	----

DOMAINE DE FIXATION DU COENZYME



DOMAINE CATALYTIQUE

I. ANALYSE BIBLIOGRAPHIQUE:

I.1 Avant propos:

Les protéines ont toujours fasciné les chercheurs tant par leur diversité que par leurs propriétés essentielles à toute forme de vie.

Bien qu'il soit connu depuis longtemps qu'elles sont toutes construites sur base des vingt mêmes acides aminés, ce n'est que beaucoup plus tard que l'on a appréhendé l'importance de leur structure tridimensionnelle. On peut classer les protéines en trois grandes familles, les protéines structurelles (essentiellement des protéines fibrillaires), les protéines du métabolisme (enzymes, protéines transporteuses) et les protéines du système immunitaire (immunoglobulines, complexe majeur d'histocompatibilité).

La disposition spatiale relative de leurs divers composants leur permet d'organiser une structure réalisant une fonction spécifique.

De par leur architecture certaines protéines peuvent, après une purification rigoureuse, s'organiser en un réseau cristallin. Cette propriété a rendu possible l'étude de leur structure spatiale par l'application des techniques de cristallographie aux rayons X et de résonance magnétique nucléaire (NMR).

Les gros désavantages de ce type d'étude sont leur lenteur, leur coût, et l'aspect figé du modèle proposé. En effet, quelque deux cents protéines ont seulement pu être cristallisées, c'est à dire à peine un dixième des protéines connues actuellement. De plus, elles subissent souvent d'importantes modifications de structure en cours de réaction; à moins de ne pouvoir la cristalliser avec et sans ses molécules associées, une bonne part d'information reste souvent inaccessible.

Et pourtant, quels formidables outils et quelles possibilités offrirait les protéines si l'on pouvait les maîtriser, les modifier en vue de les utiliser à des tâches prédéfinies!

Alors que cela relevait de la science fiction il y a quelques années seulement, les récents développements des techniques de modélisation moléculaire assistée par ordinateur permettent la conception de modèles tridimensionnels de protéines non cristallisées. Même si les conditions d'application de ces techniques sont encore très contraignantes, il devient possible d'anticiper les changements conformationnels engendrés par des mutations ponctuelles de la chaîne polypeptidique. De même, ces modèles rendent de grands services dans l'étude d'interactions inter-moléculaires et dans la conception de substances thérapeutiques spécifiques.

L'objectif de ce mémoire est d'obtenir un modèle tridimensionnel de L'ALCOOL DESHYDROGENASE I (YADH) de *Saccharomyces cerevisiae*.

Il permettra de mieux comprendre les mécanismes responsables de la catalyse et servira de "banc d'essais" pour de futurs travaux de génie biochimique et génétique réalisé sur cette enzyme.

Cette étude s'inscrit donc dans une perspective de complément aux études réalisées sur cette enzyme, mais également d'introduction aux techniques "d'ingénierie des protéines". Les potentialités de la modélisation moléculaire assistée par ordinateur sont énormes et elles ne cesseront de croître dans les années à venir; j'espère pouvoir vous en convaincre à travers ce travail.

Amino acid or residue thereof	Three-letter symbol	One letter symbol	Mnemonic help for one-letter symbol	Relative abundance in <i>E. coli</i> proteins (19) (%)	M.W. of residue at pH7.0 (daltons)	pK value of side chain (19)	ΔG values for transfer of side chain from water to ethanol at 25°C (16) (kcal/mol)
Alanine	Ala	A	<u>A</u> lanine	13.0	71		-0.5
Glutamate	Glu	E	glu <u>E</u> tamic acid		128	4.3	
Glutamine	Gln	Q	<u>Q</u> -tamine	10.8	128		
Aspartate	Asp	D	aspar <u>D</u> ic acid		114	3.9	
Asparagine	Asn	N	asparagi <u>N</u> e	9.9	114		
Leucine	Leu	L	<u>L</u> eucine	7.8	113		-1.8
Glycine	Gly	G	<u>G</u> lycine	7.8	57		
Lysine	Lys	K	before <u>L</u>	7.0	129	10.5	
Serine	Ser	S	<u>S</u> erine	6.0	87		+0.3
Valine	Val	V	<u>V</u> aline	6.0	99		-1.5
Arginine	Arg	R	a <u>R</u> ginine	5.3	157	12.5	
Threonine	Thr	T	<u>T</u> hreonine	4.6	101		-0.4
Proline	Pro	P	<u>P</u> roline	4.6	97		
Isoleucine	Ile	I	<u>I</u> soleucine	4.4	113		
Methionine	Met	M	<u>M</u> ethionine	3.8	131		-1.3
Phenylalanine	Phe	F	<u>F</u> enylalanine	3.3	147		-2.5
Tyrosine	Tyr	Y	t <u>Y</u> rosine	2.2	163	10.1	-2.3
Cysteine	Cys	C	<u>C</u> ysteine	1.8	103		
Tryptophan	Trp	W	t <u>W</u> o rings	1.0	186		-3.4
Histidine	His	H	<u>H</u> istidine	0.7	137	6.0	-0.5

Weighted mean 108.7

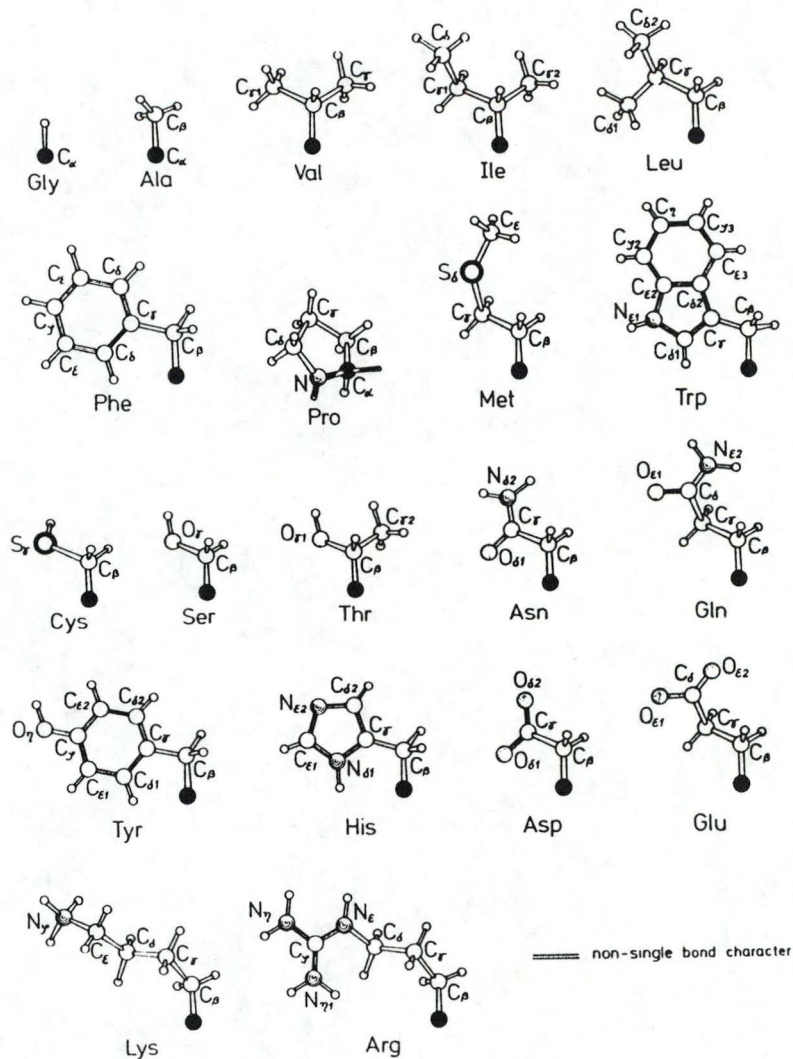


Table 1:

Représentation et propriétés des 20 acides aminés standards. (Schulz, 85)

I.2 CONFORMATION DES PROTEINES:

La compréhension des aspects fondamentaux de l'architecture des protéines est nécessaire pour envisager leur modélisation et leur transformation.

Au sein de leur apparente diversité de structure, on distingue communément quatre niveaux d'organisation: la structure primaire, secondaire, tertiaire et quaternaire (Linderström, 1959). Chacun de ces niveaux répond à un ensemble de paramètres précis.

I.2.1 LA STRUCTURE PRIMAIRE:

La structure primaire ou séquence en acides aminés (aa), est le niveau d'organisation le plus fondamental. La nature mais aussi l'ordre des acides aminés vont déterminer entièrement la configuration tridimensionnelle de la molécule (Jaenicke, 1987). Dans la nature, tous les aa sont sous la forme stéréoisomérique L. Selon la convention en vigueur, la structure primaire se lit depuis l'aa dit "N-terminal" ou "amino-terminal" (groupement NH₂ libre) jusqu'à l'aa "C-terminal" ou "Carboxy-terminal" (groupement acide libre). Le carbone central, portant la chaîne latérale, est appelé: "carbone α " (C α). La configuration et les propriétés des chaînes latérales sont responsables des propriétés des acides aminés. Ces propriétés sont reprises à la table (1).

La structure primaire d'une protéine est déterminée selon la méthode de Sanger (Sanger, 1951)

I.2.2 REPLIEMENT DES CHAINES POLYPEPTIDIQUES:

Dès le début de sa synthèse, la protéine va subir d'important changements. Elle va très rapidement adopter une configuration spatiale lui conférant des caractéristiques particulières. Ces modifications constituent l'étape de "folding" ou "repliement" de la protéine (Jaenicke, 1984). Les mécanismes par lesquels une protéine acquiert sa structure native sont encore mal connus. L'enjeu que représente la compréhension de tels phénomènes est important; il serait dès lors possible de prédire la conformation active d'une chaîne polypeptidique à partir de sa structure primaire; mais nous n'en sommes pas encore là, loin s'en faut!

Comment un polypeptide passe-t-il d'une conformation désordonnée à son état natif? Des expériences de dénaturation/renaturation semblent montrer que le processus de repliement est spontané (Anfinsen, 1961). Il s'agirait donc d'une transition entre un état d'énergie libre élevée (désordonné ou "random") et un état de plus faible énergie libre (natif) (Schulz, 1985: p37)

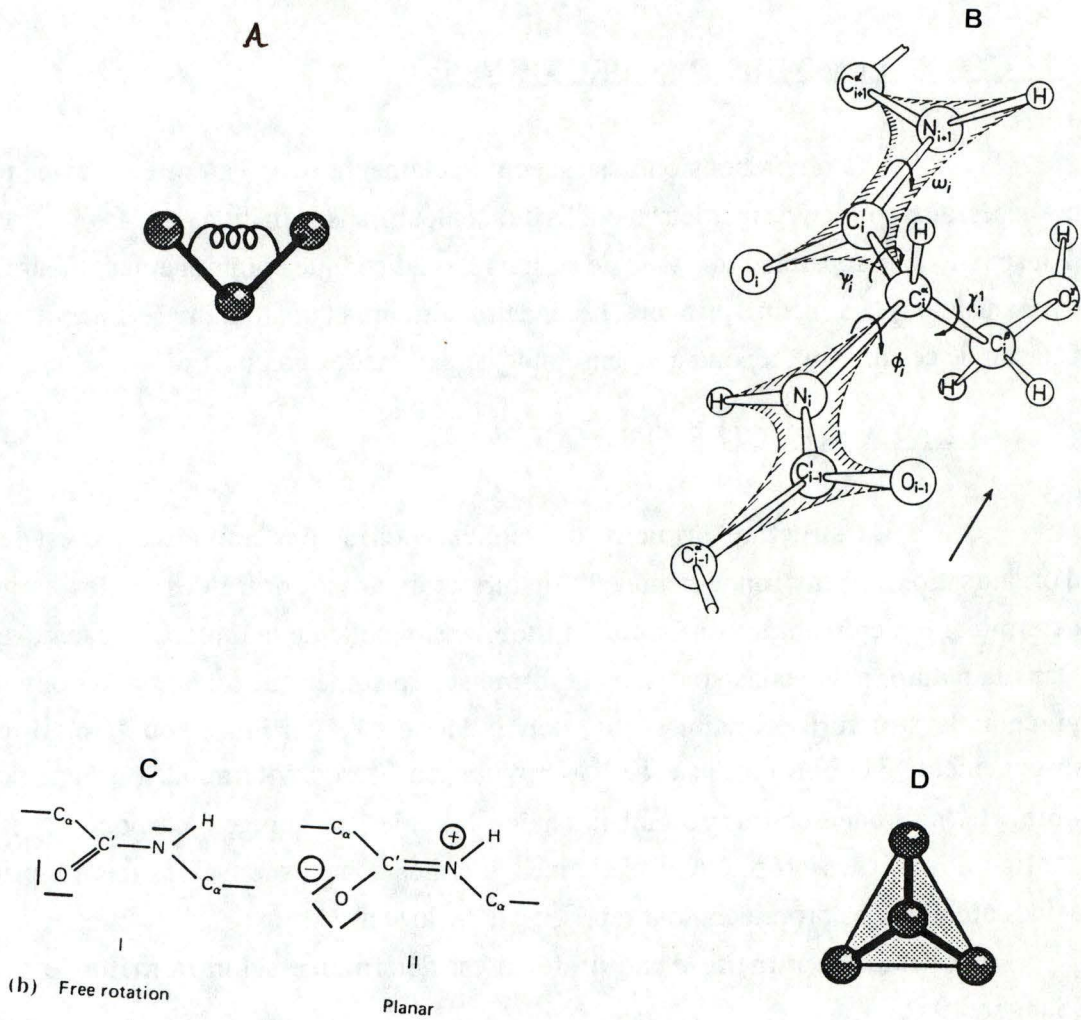


Figure 1:

Représentation schématique des différents angles d'une chaîne polypeptidique:

- a) a) Angle de valence (Discover, 1989).
- b) b) Angles de torsion (Schulz, 1985)
- c) c) La résonance du lien peptidique rigidifie l'angle ω (Schulz, 1985)
- d) d) Angle "out of plane" (Discover, 1989)

L'énergie libre de Gibbs (G) est fonction de deux composantes: l'enthalpie (H) qui reflète l'énergie des liaisons au sein de la protéine et l'entropie (S) qui représente la probabilité de trouver la chaîne d'aa dans telle ou telle conformation, ou encore, la mesure du désordre de la chaîne. L'énergie libre est également fonction de la température, intervenant comme facteur du terme entropique. Une réaction est spontanée si la variation d'énergie libre au cours de cette réaction est négative.

$$\Delta G = \Delta H - \Delta T^{\circ} \cdot S \quad \text{avec } \Delta H = H_{\text{natif}} - H_{\text{random}}$$

et $\Delta S = S_{\text{natif}} - S_{\text{random}}$

La variation d'enthalpie découle des modifications des interactions entre les atomes de la chaîne polypeptidique lors de son repliement. La variation d'entropie résulte principalement de la réorganisation du solvant autour de la protéine repliée (Schulz, 1985: p39).

Quelles sont les différentes interactions intervenant dans le processus de repliement?

1.2.2.1 Les interactions covalentes (Biosym-Discover,89)

Les interactions covalentes interviennent entre deux atomes liés de manière covalente. L'énergie augmente rapidement dès que les atomes s'éloignent de la distance optimum, c-à-d du rayon covalent des atomes en présence. C'est le type d'interaction qui possède les énergies les plus élevées. D'autres types d'interactions, découlant de la distribution électronique des liaisons covalentes, sont à envisager

a) Un angle de valence (fig 1) est l'angle formé par deux liaisons covalentes successives.

b) Un angle de torsion (fig 1); trois angles de torsions existent sur la chaîne principale d'un dipeptide.

L'angle de torsion autour de l'axe N- C α est appelé angle ϕ celui autour de la liaison C α -C est l'angle ψ et celui autour de l'axe C - N (de l'aa suivant) est l'angle " ω " (fig 1). Les angles de torsions de la chaîne latérale sont appelés χ_n . Ces angles se mesurent de 0° à +180° et de 0° à -180°. ω adopte spontanément une valeur de +180° (dite "Trans") relativement constante. Sa rigidité découle de la liaison "double partielle", engendrée par la résonance entre le C=O et le doublet de l'azote du lien peptidique. Une configuration "Cis" (0°) n'est pas favorable d'un point de vue énergétique car les chaînes latérales des résidus adjacents sont trop proches l'une de l'autre.

c) Un angle "Out-of-plane" fig(1) est l'angle formé entre le plan de base d'un tétraèdre et la direction de la liaison entre un des atomes du plan et l'atome en dehors de ce plan.

I.2.2.2 Les interactions non covalentes

Les interactions non covalentes sont d'une grande importance pour la formation et la stabilisation de la structure spatiale d'une protéine. Au cours du repliement, des atomes qui étaient isolés interagissent avec d'autres atomes, des charges apparaissent sur certains groupements et un réseau complexe d'influences va s'établir, stabilisant la chaîne polypeptidique dans une conformation donnée. Le rôle du solvant est également fondamental pour la stabilité conformationnelle de la protéine. Preuve en est qu'elle perd toute activité lorsqu'elle est transférée dans un milieu apolaire. On parle alors de dénaturation, qui consiste en une rupture du réseau des liaisons non covalentes intra (au sein de la protéine et du solvant) et inter (entre la protéine et le solvant)- moléculaire.

a) Les interactions électrostatiques:

Les interactions électrostatiques interviennent entre deux atomes chargés et non liés de manière covalente(ex: deux ions). L'énergie est fonction des charges, de la distance les séparant et de la constante diélectrique du milieu. Cette énergie n'est jamais nulle même à grande distance

b) Les ponts hydrogène:

Un proton peut être partagé entre un acide potentiel et une base potentielle et former un pont hydrogène. Ces interactions se créent donc entre des groupements polaires. La longueur d'un pont hydrogène varie de 2,5 Å ($2,5 \cdot 10^{-10}$ m) entre deux groupements hydroxyl jusqu'à 3,7 Å entre un sulfidryl et un groupement amide. Les ponts hydrogène sont très importants dans la stabilisation de la structure spatiale des protéines. Ils tendent également à être plus stables lorsque le donneur et l'accepteur sont alignés sur un même axe (Schulz, 1985).



Residue	Nozaki	Wolfenden	Eisenberg
Ala	0.5	1.9	0.25
Cys	—	-1.2	0.04
Asp	—	-10.9	-0.72
Glu	—	-10.2	-0.62
Phe	2.5	-0.8	0.61
Gly	0.0	2.4	0.16
His	0.5	-10.2	-0.40
Ile	2.6	2.2	0.73
Lys	—	-9.5	-1.10
Leu	1.8	2.3	0.53
Met	1.3	-1.5	0.26
Asn	—	-9.7	-0.64
Pro	—	—	-0.07
Gln	—	-9.4	-0.69
Arg	—	—	-1.80
Ser	-0.3	-5.1	-0.26
Thr	0.4	-4.9	-0.18
Val	1.5	2.0	0.54
Trp	3.4	-5.9	0.37
Tyr	2.3	-6.1	0.02

Table 2:

Echelles d'hydrophobicité des 20 acides aminés standards selon différents auteurs (Von-Heijne, 1989)

c) Les interactions de Van der Waals (VDW):

Des hétérogénéités temporaires apparaissent dans la distribution électronique de certains atomes de la protéine ce qui induit la formation de dipôles momentanés. Ces dipôles engendrent une interaction entre des atomes pourtant globalement neutres. Ces interactions sont assez faibles mais sont essentielles à la formation du réseau assurant la stabilité de la structure tridimensionnelle d'une protéine.

La distance entre deux atomes non covalents résulte d'un équilibre entre, d'une part l'attraction entre les dipôles temporaires et d'autre part, la répulsion entre les orbitales électroniques des partenaires. La distance optimale pour ce type d'interaction entre deux atomes est appelée: RAYON DE VAN DER WAALS. Chaque paire d'atome possède son rayon de VDW particulier. On attribue également un rayon de VDW moyen à chaque atome. Ils varient entre 3Å et 4 Å. En deçà de cette distance optimale, une forte répulsion existe entre les partenaires alors qu'au delà, l'attraction tend à les rapprocher pour autant que la distance les séparant ne soit pas trop importante. Plus la distance augmente, plus l'énergie de l'interaction tend vers zéro.

d) L'hydrophobicité:

Dans une protéine, les résidus hydrophobes (apolaires) en contact avec le solvant polaire, structurent le réseau de ponts hydrogène existant au sein de ce dernier et provoquent une diminution de son entropie. Plus le nombre de résidus hydrophobes en contact avec le solvant est important, plus l'entropie du solvant est faible. Or, les phénomènes spontanés sont caractérisés par une augmentation de l'entropie (second principe de la thermodynamique). Le processus de repliement de la chaîne polypeptidique résulterait donc en partie, de la tendance naturelle du solvant à augmenter son entropie en minimisant sa surface de contact avec des résidus hydrophobes (Schulz, 1985).

Il n'existe pas d'équation décrivant directement ce phénomène empirique, mais il peut être quantifié en mesurant, à une concentration donnée, l'énergie libre de transfert de l'aa entre une phase aqueuse et une phase organique (Nozaki et Tanford, 1971), ou encore par la tension superficielle du film formé par les résidus hydrophobes à l'interface air-solvant (Wolfenden et al, 1979). Un aperçu de ces différentes échelles ainsi que d'une échelle consensus (Eisenberg, 1984) est donné à la table (2).

L'échelle obtenue montre que l'hydrophobicité est proportionnelle à la surface accessible au solvant (voir §I.2.3.6 pour déf.) de la chaîne latérale de l'aa (Chothia, 1974).

I.2.2.3 Comment varie l'énergie libre au cours du repliement?

Compte tenu des différentes contributions à l'énergie libre d'une protéine, le passage de l'état "random" à l'état "natif" peut être décrit par l'équation suivante (Schulz, 1985: p38):

$$\Delta G_{\text{total}} = \Delta H_{\text{prot}} - T^{\circ} \Delta S_{\text{prot}} + \Delta H_{\text{solv}} - T^{\circ} \Delta S_{\text{solv}}$$

avec Δ_{prot} = interactions à l'intérieur de la protéine + les interactions protéine-solvant.

Δ_{solv} = interactions au sein du solvant

En solution aqueuse, les ΔH_{prot} sont positifs et les ΔS_{prot} sont négatifs ce qui tend à favoriser l'état désordonné de la chaîne polypeptidique. Ceci peut-être expliqué par le fait que lors du repliement, les résidus enfouis au centre de la protéine sont moins influencés par les dipôles des molécules d'eau et sont moins polarisés. Il existe donc moins d'interactions de VDW entre les résidus de la chaîne polypeptidique dans l'état natif que dans l'état désordonné. La perte de liaisons est un phénomène endergonique (prise de chaleur dans le milieu) et entraîne une augmentation de l'enthalpie de la chaîne polypeptidique lors de son repliement (ΔH_{prot} positif). Par ailleurs, l'ordre de la chaîne polypeptidique étant plus important dans l'état natif, le ΔS_{prot} est négatif. Le ΔG_{prot} est donc positif ce qui est caractéristique d'un phénomène non spontané.

Le solvant, quant à lui, présente des variations d'enthalpie et d'entropie tendant à favoriser l'état natif de la chaîne polypeptidique. L'organisation du solvant autour de la protéine repliée est un phénomène exergonique (libération de chaleur) induisant une diminution d'enthalpie ainsi qu'une diminution de son entropie (la surface de contact avec des résidus hydrophobes est plus faible). Le ΔG_{solv} résultant est largement négatif et donc favorable.

Le ΔG_{total} résultant de ces deux phénomènes est légèrement négatif. L'acquisition de la structure native est donc spontanée. Cependant, le caractère négatif peu prononcé peut expliquer la relative facilité qu'ont les protéines à se dénaturer quand les conditions standards (p.ex. de T°) ne sont plus respectées.

D'un point de vue expérimental il est difficile de quantifier ces valeurs. Par contre il est relativement aisé de déterminer la variation d'énergie libre de transfert (ΔG_{trans}) d'un aa hydrophobe passant d'une phase aqueuse à une phase organique (Creighton, 1989: chap14).

Helix		
Residue	C&F ^a	L ^b
Glu	1.51	1.44
Met	1.45	1.47
Ala	1.42	1.29
Leu	1.21	1.30
Lys	1.16	1.23
Phe	1.13	1.07
Gln	1.11	1.27
Trp	1.08	0.99
Ile	1.08	0.97
Val	1.06	0.91
Asp	1.01	1.04
His	1.00	1.22
Arg	0.98	0.96
Thr	0.83	0.82
Ser	0.77	0.82
Cys	0.70	1.11
Tyr	0.69	0.72
Asn	0.67	0.90
Pro	0.57	0.52
Gly	0.57	0.56

Table 3

Indice de propension des différents résidus à former les hélices α selon Chou et Fasman (1974) et selon Levitt (1978)

Plus ces indices sont élevés, plus la propension est élevée (Von-Heijne, 1989)

$$\Delta G_{\text{trans}} = RT \ln C_{\text{org}} / C_{\text{eau}}$$

où $C_{\text{}}$ est la concentration de l'aa
dans la phase concernée (mole/L).

R est la constante des Gaz parfaits

T est la Température absolue (K°)

Sur base de ces valeurs, on peut montrer que:

$$\Delta G_{\text{trans}} = \Delta H_{\text{trans}} - T^{\circ} \cdot \Delta S_{\text{trans}}$$

et $\Delta H_{\text{trans}} = \Delta H_{\text{prot}} + \Delta H_{\text{solv}}$

$$\Delta S_{\text{trans}} = \Delta S_{\text{sol}}$$

I.2.2.4 Structure native: minimum local ou global?

Au vu du nombre d'interactions présentes au sein d'une protéine, on peut aisément concevoir la complexité de sa fonction énergie libre. En réalité, celle-ci présente un très grand nombre de minima reflétant chacun une configuration potentielle métastable.

Quel est l'état énergétique correspondant à la configuration native de la protéine?

Deux théories s'opposent (Schulz, 1985: p153). La première (hypothèse du contrôle thermodynamique) suppose que la protéine native correspond au minimum global d'énergie. Cette hypothèse se verrait confirmée par les expériences de dénaturation/renaturation (Anfinsen, 1959) dont les produits de renaturation retrouveraient tous la même activité et donc la même structure tridimensionnelle. La seconde hypothèse (hypothèse du contrôle cinétique) avance que le repliement d'une protéine est trop rapide dans la plupart des cas, pour qu'elle ait eu le temps de retrouver la configuration la plus stable parmi toutes celles existant. L'acquisition de la structure tertiaire devrait donc s'effectuer selon un ordre précis ne conduisant pas nécessairement au minimum global.

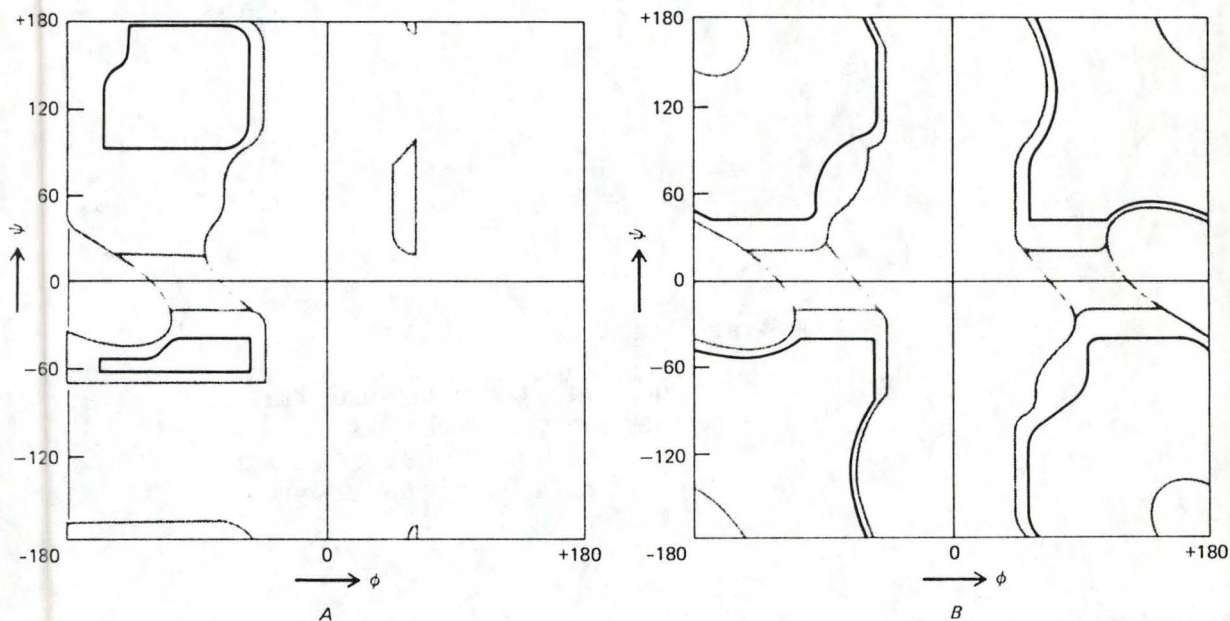


Figure 2

Graphe de Ramachandran des angles ϕ et ψ pour l'Ala en A et pour la Gly en B. Les conformations les plus stables appartiennent aux plages entourées d'un trait noir. Les conformations moins stables mais parfois rencontrées sont délimitées par les traits gris. (Creighton, 1984)

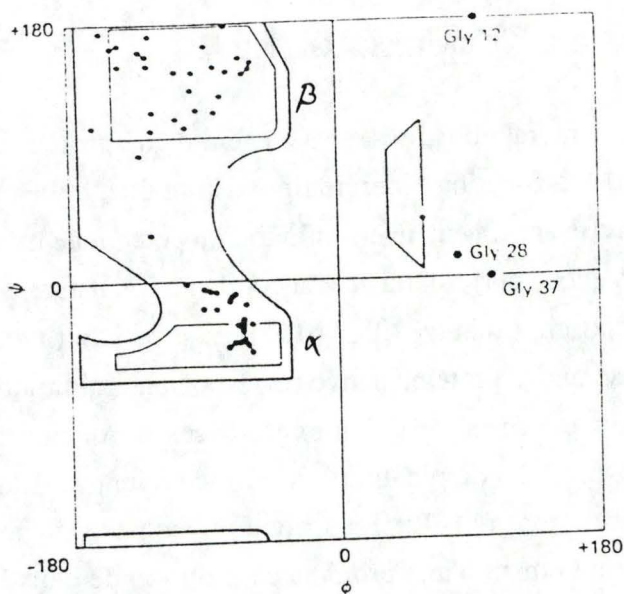


Figure 3

Graphe de Ramachandran des conformations des résidus appartenant aux hélices α (α) et aux plans β (β) (Creighton, 1984)

I.2.3 DIFFERENTS NIVEAUX DE REPLIEMENT:

I.2.3.1 Les structures secondaires:

On retrouve principalement trois types de structures secondaires stables dans la nature: les hélices α , les plans β , et les "turns" et les "loops". Les structures secondaires sont essentiellement déterminées par les valeurs adoptées les différents angles de torsion de la chaîne polypeptidique. En particulier les valeurs de ϕ et ψ sont très conservées au sein d'une même structure.

Ramachandran *et al.* (1963) ont corrélié les angles ψ en fonction des angles ϕ (§ I.3.2.1). Deux zones distinctes se démarquent, une propre aux hélices α et l'autre aux plans β . On peut également remarquer que les aa ne présentent pas tous les mêmes exigences conformationnelles; L'alanine et la glycine, à cause de leur chaîne latérale particulière, tolèrent des angles ϕ positifs (fig 2). Les zones propres à la valine et à l'isoleucine sont, au contraire, très restreintes. Il ne faut pourtant pas oublier que nous considérons ici des fréquences d'apparition de telle ou telle conformation pour chaque résidu dans une situation donnée. Des exceptions existent et peuvent jouer un rôle important dans la structure des protéines concernées.

a) Les hélices α :

La chaîne principale de la protéine se dispose de manière hélicoïdale. La majorité des hélices naturelles sont construites selon un pas droit. Leur diamètre est en moyenne de 2,3 Å, un tour complet est constitué de 3,6 résidus avec un déplacement vertical de 1,5 Å/résidu. La conformation des résidus constituant l'hélice n'est pas quelconque. Un graphe de Ramachandran montre que les angles ϕ et ψ sont tous deux négatifs (fig 3).

Ces structures sont stabilisées par des ponts hydrogène entre l'amine du résidu i et le carboxyl du résidu $i+3$. L'ensemble des dipôles présents dans les liens peptidiques et dans ces ponts hydrogène donnent un caractère polaire global à l'hélice appelé " moment dipolaire". Ce moment dipolaire est représenté sous forme d'un vecteur. Des hélices adjacentes dans une protéine peuvent s'influencer mutuellement via leur moment dipolaire (Schulz, 1985:p31). Tous les acides aminés ne sont pas favorables à la formation d'hélices α . CHOU et FASMAN (1974) et Levitt (1978) les ont classé selon leur fréquence d'apparition au sein d'hélices dans des protéines de structure tridimensionnelle connue. Plus la fréquence est élevée, plus sa propension à former une hélice ($P_i\alpha$) est grande (table 3). Par exemple, la proline ne pouvant former de pont hydrogène et possédant un angle ϕ de -80° , imprime une courbure aux hélices. La glycine les déstabilise car sa conformation est trop "libre".

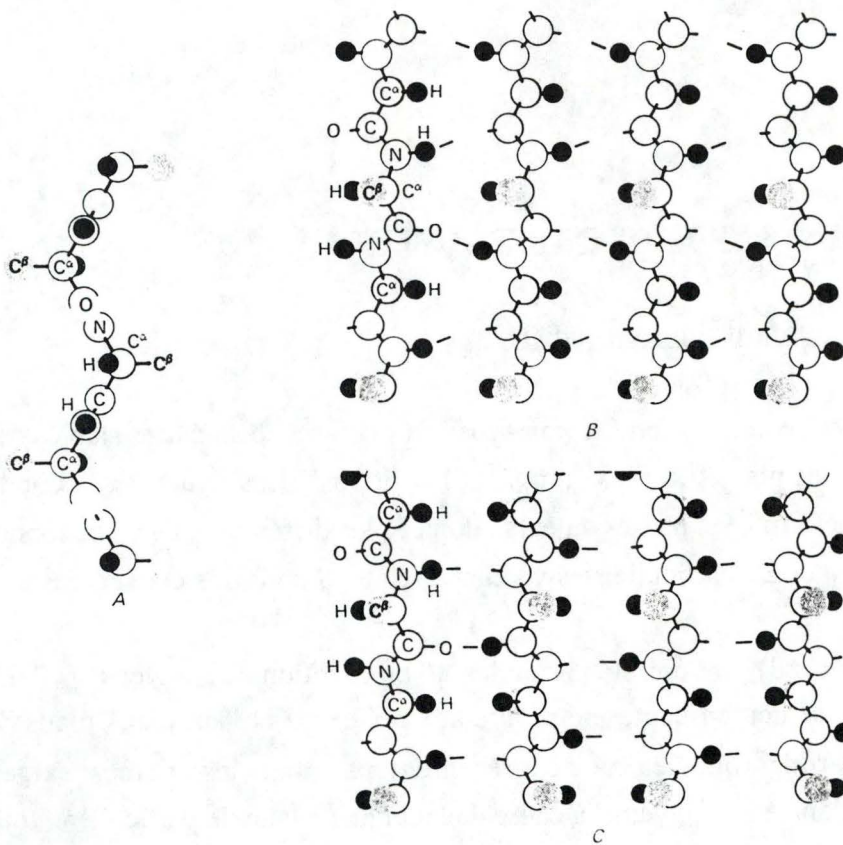


Figure 4

Représentation d'une chaîne d'un plan β (A), d'un plan β parallèle (B) et d'un plan β antiparallèle (C) (Creighton, 1984).

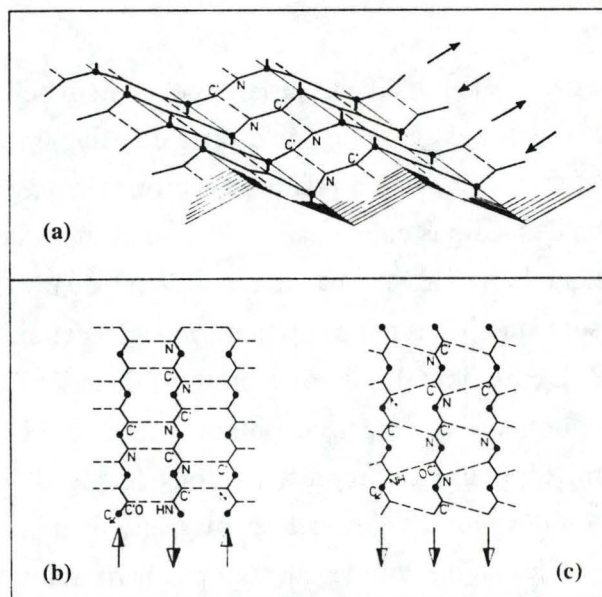


Figure 5

Vue d'un plan β antiparallèle en perspective (A). Illustration des ponts hydrogène d'un plan β antiparallèle (B) et des ponts hydrogène d'un plan β parallèle (C). Les ponts hydrogènes sont représentés par des pointillés, les C α par des points, et le sens des chaîne polypeptidique par des flèches (Schulz, 1985).

b) Les plans β :

Les plans β sont constitués de chaînes d'aa étendues et accolées les unes aux autres. L'alignement de chaînes côte à côte donne à cet ensemble un aspect de feuillet. De même que pour les hélices, les angles ϕ et ψ répondent à des valeurs particulières (respectivement $+90^\circ$ et $+120^\circ$) (fig 3). Chou et Fasman ont également réalisé une analyse des résidus constituant les plans β de protéines de structure tridimensionnelle connue (voir table 4 § I.3.2.2). Une fois de plus, les propriétés particulières de la proline lui confèrent un mauvais indice $P_i\beta$. La plupart des plans présentent une conformation légèrement tordue, un peu comme si ils tendaient à adopter une conformation hélicoïdale. Il s'agirait d'un état énergétiquement plus favorable.

On rencontre principalement deux types de plans β : les parallèles et les anti-parallèles (fig 4). La largeur des plans est d'environ 5\AA par chaîne accolée et leur longueur est en moyenne de 20\AA (6 résidus).

Les chaînes des plans β parallèles sont accolées depuis N-terminal jusque C-terminal dans une même direction. Les groupes carbonyles d'une chaîne forment des ponts hydrogène avec les groupes amines de l'autre chaîne (fig 5). Ils sont légèrement étirés . Les résidus d'une même chaîne se retrouvent alternativement au dessus et en dessous du plan, ce qui évite les encombrements stériques.

Les chaînes des plans β antiparallèles sont accolées selon des orientations opposées. Cette configuration est plus favorable d'un point de vue énergétique que la parallèle, car les ponts hydrogène sont linéaires. Ils sont donc plus couramment rencontrés.

c) Les turns:

Ces structures constituent souvent un lien entre les hélices et/ou les plans. Plusieurs types de turns existent et sont présentés à la figure (6) (Birktof *et al*, 1972). La distribution des résidus n'est pas aléatoire et dépend du type de turn. par exemple, la proline ne peut pas se retrouver en position 3 dans le type I, le type II nécessite une glycine en position 2 et 3,... . Les turns et les loops (plus longs) sont responsables d'un troisième niveau d'organisation spatiale des protéines: **les super-structures secondaires**. Ces super- structures sont constituées de stuctures secondaires assemblées selon des motifs particuliers .

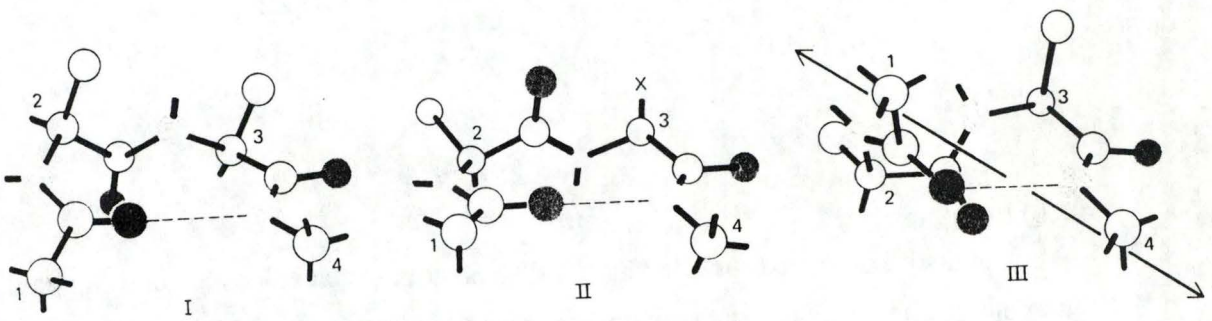


Figure 6

Illustration des Turns les plus souvent observés dans les protéines. Les pointillés représentent les ponts hydrogènes. Les sphères blanches sont les atomes de carbones, les sphères grises représentent les azotes et les sphères noires représentent les oxygènes. (Creighton, 1984)

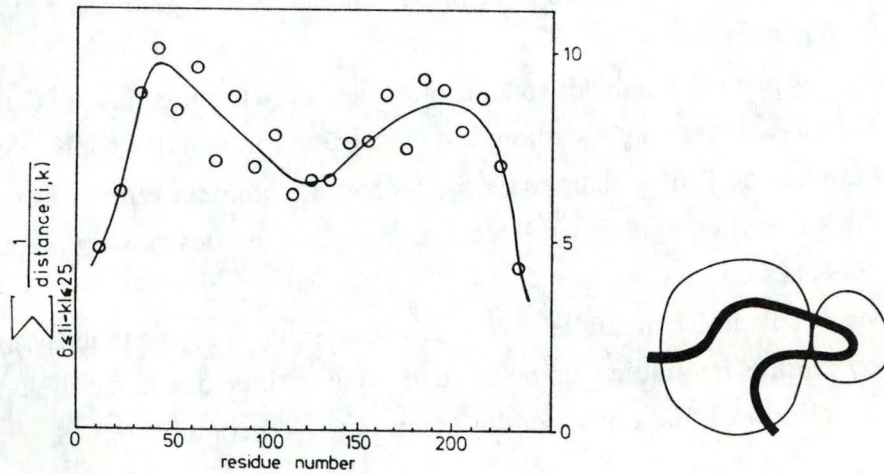


Figure 7

Mesure du compactage des résidus dans une chaîne polypeptidique. Le compactage est estimé à partir d'une fonction inversement proportionnelle aux distances inter-résidu. Les deux domaines apparaissent clairement (résidus mieux compactés) (Schulz, 1985).

I.2.3.2 La structure tertiaire

Si on peut représenter grossièrement une protéine globulaire comme une sphère, la distribution des résidus à l'intérieur même de la sphère est souvent hétérogène. Des zones de densité plus élevée apparaissent, constituant les domaines de la protéine (fig 7) p84, Schulz, 1985). Ces domaines sont des ensembles de résidus organisés selon une architecture particulière. Certains participent par exemple à la fixation du coenzyme ou du substrat,..., et sont des domaines fonctionnels (catalytique,...). Quand aucune fonction particulière ne leur est reconnue, on parle de domaines structuraux (Schulz, 1985: p84). Une illustration des domaines de l'alcool déshydrogénase de levure (modèle) est donnée en première page de ce travail .

Nous venons de décrire les différentes étapes par lesquelles une chaîne polypeptidique se replie dans l'espace pour constituer un monomère. Au sein d'un monomère, les atomes occupent au maximum le volume intérieur (compactage). Le rapport entre le volume total de la protéine et le volume de l'ensemble des atomes est égal à 0,76. Par comparaison, ce rapport calculé pour un ensemble de billes de verre de taille unique est de 0,74. La compressibilité des protéines est d'ailleurs très faible. Ce rapport est important et peut être un des critères de validation de modèles tridimensionnels (Schulz, 1977).

I.2.3.3 Certaines protéines sont polymériques.

Ce dernier niveau d'organisation est connu sous le nom de structure quaternaire. Chaque monomère est appelé "sous-unité". Le nombre de sous-unités est très variable (Fasman, 1989b: p158-184) et va de 2 à 400. Le record est détenu par le virus de la mosaïque du tabac avec 2130 sous-unités, encore qu'il soit difficile de le considérer comme une protéine! Certaines protéines peuvent exister sous plusieurs formes fonctionnelles, telle la Ribulose Biphosphate Carboxylase (RuBP Carboxylase) que l'on rencontre sous forme de 12 ou 24 monomères. L'assemblage des sous-unités peut être soit symétrique, soit asymétrique. Les assemblages non symétriques sont constitués d'oligomères dont les sous-unités sont différentes, comme dans les ribosomes par exemple. Les assemblages symétriques se retrouvent dans la plupart des protéines.

Selon une analyse récente (Miller, 1989), l'assemblage des oligomères répondrait à des caractéristiques structurales précises:

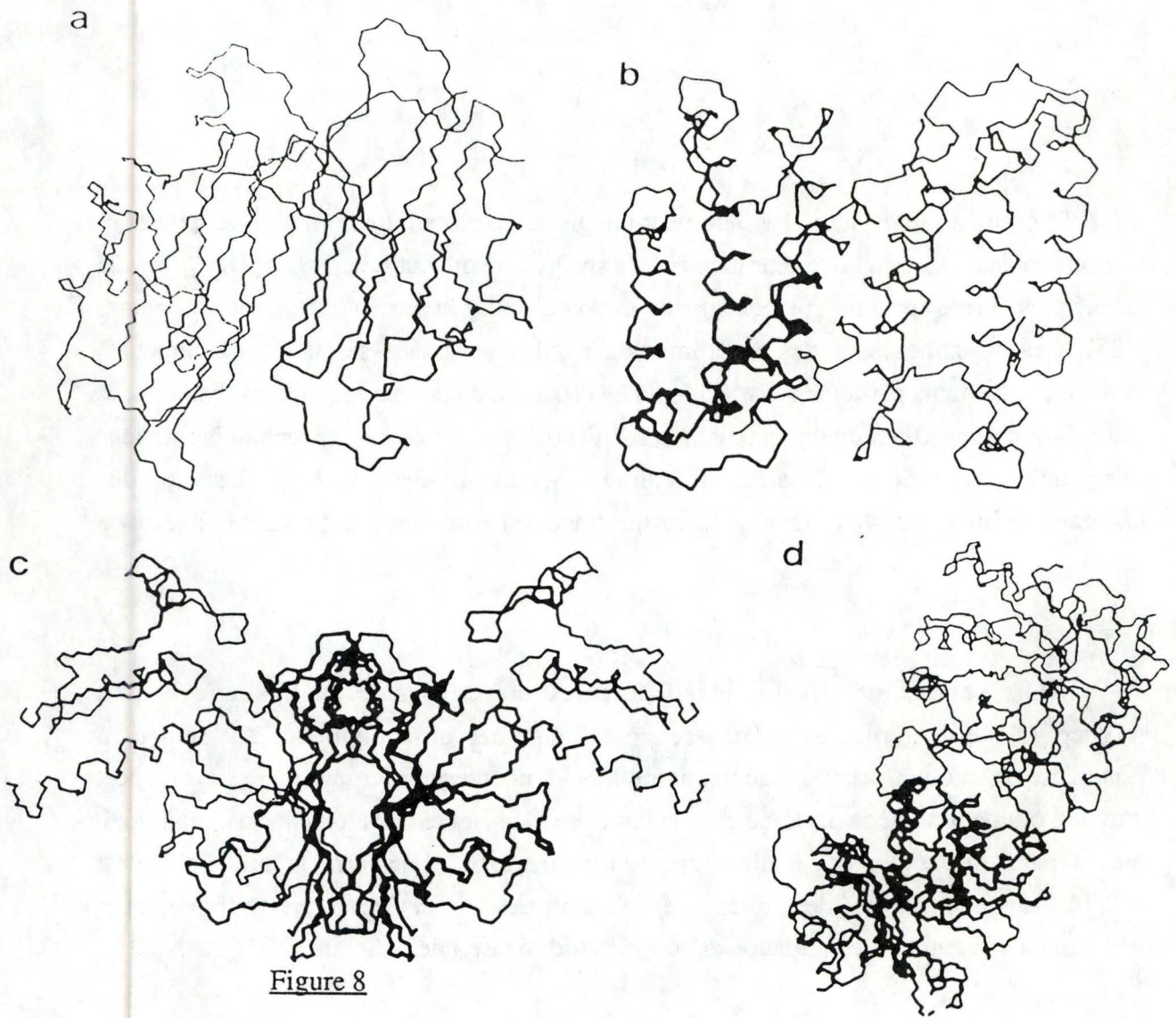


Figure 8

Quelques structures secondaires des interfaces rencontrées dans les protéines suivantes: préalbumin - plan β - (a), cytochrome c- hélices, hélices - (b), glycéraldéhyde 3P - plans β , plans β - (c) et triose phosphate isomérase - loop, loop-(d).(Miller, 1989).

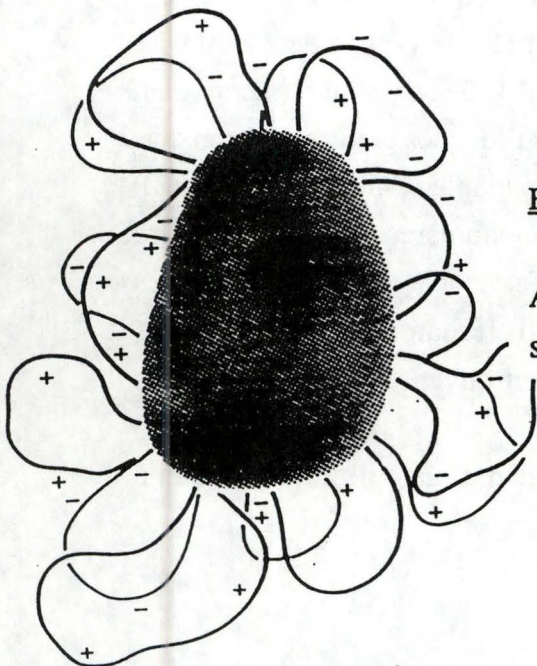


Figure 9

Aspect général d'une zone interfaciale où des résidus hydrophobes (centre) sont entourés de résidus chargés (Miller, 1989)

- Les oligomères composés d'un nombre pair de monomères seraient organisés en unités dimériques. Les interactions entre dimères étant moins prononcées qu'entre monomères, un canal central est souvent présent dans les tétramères.
- Les zones interfaciales sont complémentaires. Si l'une présente une protrusion, Celle du monomère en vis à vis comportera une invagination.
- Quatre motifs se retrouvent sur la plupart des interfaces et plusieurs de ces motifs peuvent être présents sur un même oligomère:
 - Les plans β antiparallèles formés par des chaînes appartenant à différents monomères (alcool déshydrogénase, glutathion peroxydase, catalase,...). Les assemblages d'hélices α en une structure tertiaire (glutathion peroxydase, lactate déshydrogénase,...).(fig 8)
 - Les assemblages de plans β en une structure tertiaire.
 - Les associations de loops possédant des résidus chargés
- La surface d'association doit présenter des zones à caractère hydrophobe associées à des zones comprenant des résidus chargés permettant la formation de ponts hydrogène (fig 9).

Lors de la formation d'un oligomère, les résidus hydrophobes superficiels sont isolés du milieu aqueux. La stabilité de la structure quaternaire est assurée par les interactions de Van der Waals entre les zones hydrophobes et par les ponts hydrogène entre les résidus complémentaires chargés.

L'association de monomères présente certains avantages énergétiques et évolutifs (Schulz,1985:p100). Elle permet l'isolement de résidus hydrophobes superficiels par rapport au milieu aqueux. Ce phénomène est favorable d'un point de vue énergétique comme nous l'avons vu dans un paragraphe précédent. La segmentation d'une protéine active en plusieurs sous- unités identiques permet une réduction de la longueur de DNA nécessaire à l'encodage du polypeptide . Les risques d'erreurs de traduction/transcription ainsi que l'ampleur de ces erreurs sont ainsi limités par la même occasion. Enfin, les protéines oligomériques sont beaucoup plus souples d'un point de vue évolutif que de grandes chaînes polypeptidiques uniques.

I.2.4 CONCLUSIONS

Les phénomènes régissant la structure tridimensionnelle de molécules aussi complexes que les protéines sont encore relativement mal connus. Deux grandes voies de recherche se dégagent. La première concerne l'étude des relations entre la structure primaire et la structure tertiaire des protéines. Les données qu'apportent les structures protéiques analysées par cristallographie constituent une base solide pour la recherche de corrélations entre ces niveaux d'organisation.

La seconde concerne la compréhension de la cinétique et de l'aspect énergétique du repliement des chaînes polypeptidiques. Grâce aux progrès réalisés dans ces domaines, on commence à mieux appréhender certains principes du repliement des protéines. Il devient possible, dans des limites précises, d'envisager la prédiction de structures tridimensionnelles de petites molécules. Les difficultés d'établissement de modèles valables résident dans la complexité structurelle de telles édifices moléculaires, inhérente à leurs fonctions hautement spécialisées.

Dans le chapitre qui suit, nous allons tenter de donner un aperçu de ces méthodes et de leurs champs d'application.

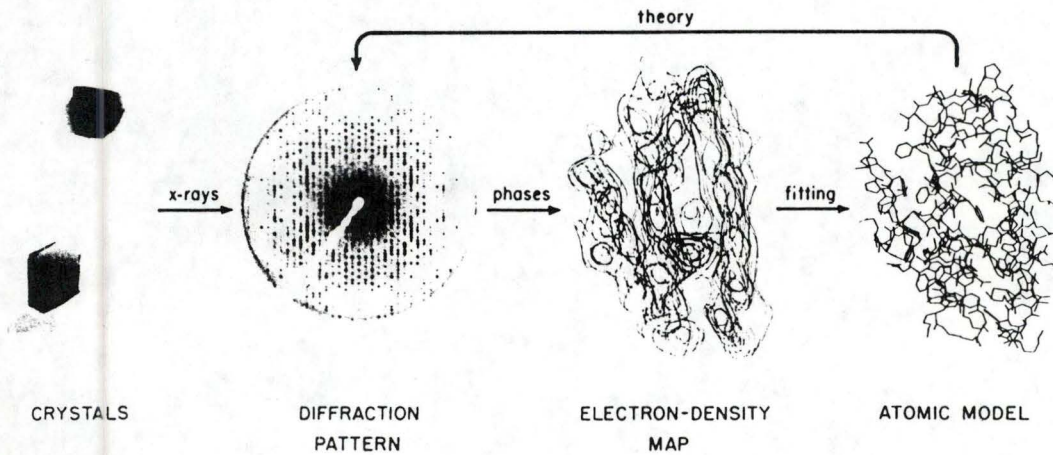
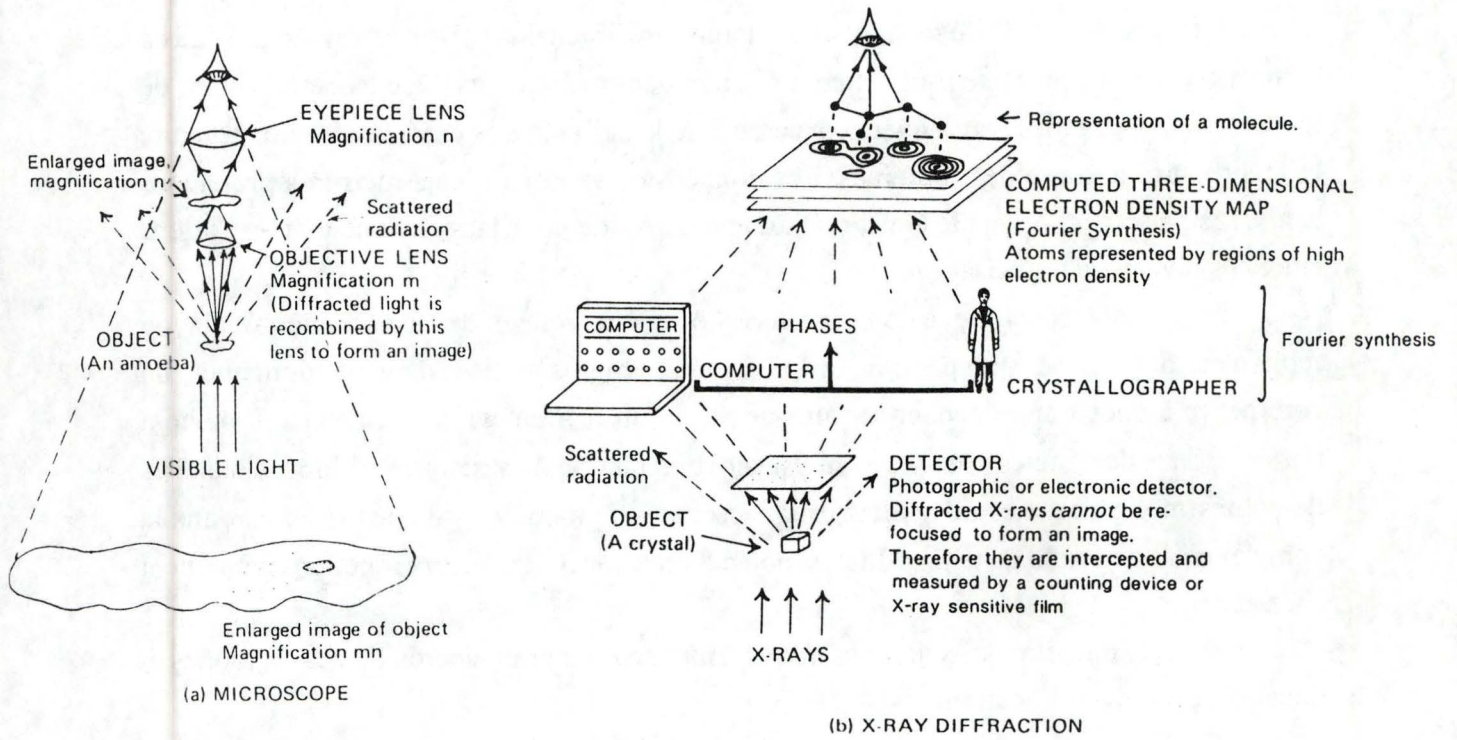


Figure 10

Au dessus: Analogie entre le microscope optique (a) et la cristallographie par diffraction des rayons X (b) (Pickworth, 1985).

En dessous: Illustration des différentes opérations d'analyse d'une structure tridimensionnelle par diffraction aux rayons X (Hendrickson, 1987).

I.3 ETUDE ET MODELISATION DES PROTEINES:

I.3.1 LA CRISTALLOGRAPHIE AUX RAYONS X:

L'observation de fines structures au microscope nécessite une très bonne résolution. Or, le pouvoir de résolution d'un microscope est fonction de la longueur d'onde de la lumière utilisée. La résolution nécessaire à l'observation de la structure atomique d'une molécule doit être de l'ordre de l'Angström (10^{-10} m), les seuls rayonnements électromagnétiques de longueur d'onde appropriée sont les rayons X.

Soit un faisceau d'ondes (lumineuses dans ce cas) passant au travers d'une ouverture de taille similaire à sa longueur d'onde. En vertu du principe de Huygens, chaque point du front d'onde (c-à-d l'ensemble des points en phase sur chaque onde du faisceau) est le siège de l'émission d'une onde dite "secondaire". Ces ondes secondaires interfèrent entre elles, ainsi qu'avec l'onde primaire, pour produire une figure de diffraction que nous observons sur un écran placé au delà de l'objet (fig 10). Dans un microscope optique ou électronique, les lentilles focalisent les rayons diffractés et reconstituent une image de l'objet. Il n'existe malheureusement aucune lentille capable de focaliser des rayons X. On recourt alors à une simulation de focalisation par traitement mathématique (la synthèse de Fourier) en vue d'obtenir une image interprétable de la structure atomique du cristal (Pickworth, 1985).

Pour analyser une protéine selon cette technique, il faut avant tout obtenir des cristaux suffisamment grand, ce qui n'est pas toujours possible. Enfin, l'interprétation des figures de diffraction est une tâche de spécialistes.

A cause de ces handicaps, un retard considérable s'est accumulé dans la connaissance des structures tridimensionnelles par rapport à celle des séquences protéiques.

I.3.2 ANALYSES DE STRUCTURES TRIDIMENSIONNELLES:

Le développement de prédictions de structures tridimensionnelles n'aurait pu se réaliser sans l'accumulation des données cristallographiques ni sans les techniques de génie génétique permettant d'étudier les effets d'une mutation particulière. De plus, des banques de données informatisées ont été développées regroupant l'ensemble des informations concernant les structures cristallographiques des protéines analysées à ce jour: la "Protein Data Bank" (PDB; Brookhaven, Cambridge; USA), les séquences des protéines connues: "Protein Information Ressources"(PIR; National Biochemical Research Foundation, Washington) ou encore leur séquence nucléotidique: "European Molecular Biology Laboratory Data Bank" (EMBL; Heidelberg). Des recherches simultanées dans plusieurs de ces banques sont souvent possibles et permettent de recouper, de compléter ou de comparer les informations reçues.

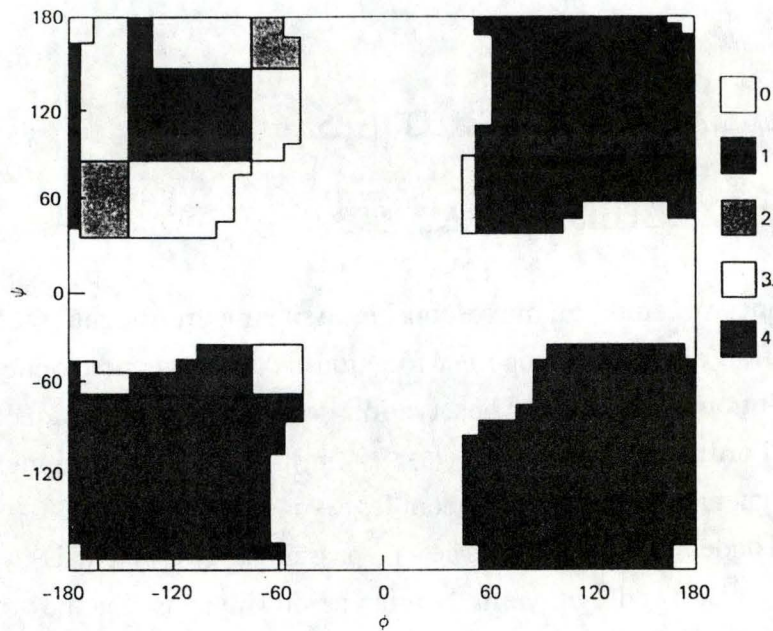


Figure 11

Graphique de Ramachandran des différentes conformations admissibles pour les acides aminés suivant: Gly de 1 à 4, Ala de 2 à 4, les aa à chaînes latérales plus longues: 3 et 4 sauf Val et Ile qui sont restreintes à la zone 4 (Creighton, 1984).

Helix			Sheet			Turn		
Residue	C&F ^a	L ^b	Residue	C&F ^a	L ^b	Residue	C&F ^a	L ^b
Glu	1.51	1.44	Val	1.70	1.49	Asn	1.56	1.28
Met	1.45	1.47	Ile	1.60	1.45	Gly	1.56	1.64
Ala	1.42	1.29	Tyr	1.47	1.25	Pro	1.52	1.91
Leu	1.21	1.30	Phe	1.38	1.32	Asp	1.46	1.41
Lys	1.16	1.23	Trp	1.37	1.14	Ser	1.43	1.32
Phe	1.13	1.07	Leu	1.30	1.02	Cys	1.19	0.81
Gln	1.11	1.27	Cys	1.19	0.74	Tyr	1.14	1.05
Trp	1.08	0.99	Thr	1.19	1.21	Lys	1.01	0.96
Ile	1.08	0.97	Gln	1.10	0.80	Gln	0.98	0.98
Val	1.06	0.91	Met	1.05	0.97	Thr	0.96	1.04
Asp	1.01	1.04	Arg	0.93	0.99	Trp	0.96	0.76
His	1.00	1.22	Asn	0.89	0.76	Arg	0.95	0.88
Arg	0.98	0.96	His	0.87	1.08	His	0.95	0.68
Thr	0.83	0.82	Ala	0.83	0.90	Glu	0.74	0.99
Ser	0.77	0.82	Ser	0.75	0.95	Ala	0.66	0.77
Cys	0.70	1.11	Gly	0.75	0.92	Met	0.60	0.41
Tyr	0.69	0.72	Lys	0.74	0.77	Phe	0.60	0.59
Asn	0.67	0.90	Pro	0.55	0.64	Leu	0.59	0.58
Pro	0.57	0.52	Asp	0.54	0.72	Val	0.50	0.47
Gly	0.57	0.56	Glu	0.37	0.75	Ile	0.47	0.51

Table 4

Propension des différents aa à former les différentes structures secondaires selon Chou et Fasman (1974) et Levitt (1978) (Von Heijne, 1989).

La structure d'une protéine peut être caractérisée par un ensemble de critères tels que: conformations des angles, positions relatives des aa , fréquence d'apparition d'un résidu particulier au sein d'une structure secondaire,

Ces analyses sont très utiles car elles permettent la mise en évidence de corrélations entre la séquence et la conformation des chaînes polypeptidiques. Elles permettent également la validation de modèles moléculaires. La liste qui est présentée ici n'est pas exhaustive mais donne un aperçu des approches couramment utilisées.

I.3.2.1 Les graphes de Ramachandran

Les valeurs des angles ψ en fonction des angles ϕ sont reportées dans un graphe. Ces valeurs sont celles relevées parmi un grand nombre de protéines. Les zones regroupant les valeurs les plus fréquemment rencontrées sont donc supposées correspondre aux conformations de moindre énergie libre (fig 11).

I.3.2.2 Distribution des résidus au sein de structures secondaires

La fréquence d'apparition de chaque résidu dans les différentes structures secondaires permet d'évaluer leur importance dans la stabilité d'une conformation particulière (Oxender et fox, 1988 p155)

Chou et Fasman (1974) et Levitt (1978) ont classé les aa selon leur caractère stabilisant ou déstabilisant pour chaque structure secondaire (table 4).

I.3.2.3 "C α - Contact maps"

Les deux séquences sont alignées selon l'abscisse et l'ordonnée. Des points sont placés aux coordonnées des résidus dont les C α sont suffisamment proches (distance appréciée par l'auteur). La présence de bandes sur la diagonale principale est caractéristique des hélices α . Les plans β parallèles donnent des alignements de points parallèles à la diagonale, tandis que les plans antiparallèles donnent des alignements perpendiculaires à la diagonale. (fig 12).

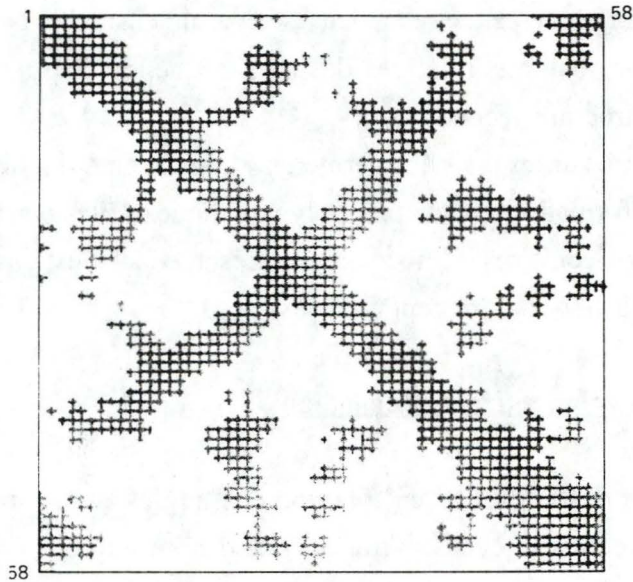


Figure 12

"C α contact map" de la structure native de l'inhibiteur de la trypsine pancréatique. Un point est placé lorsque la distance entre deux carbones α est inférieure à 10 Å.

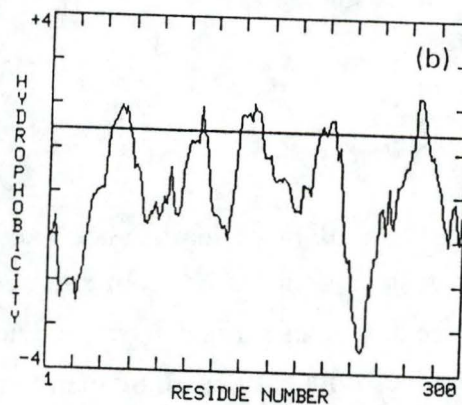


Figure 13

Profil d'hydrophobicité d'une protéine globulaire (L- arabinose binding protein). Les valeurs d' hydrophobicité sont calculées sur une fenêtre de 19aa (Von Heijne, 1989).

I.3.2.4 Profil d'hydrophobicité:

Les profils d'hydrophobicité peuvent mettre en évidence des zones d'association intermoléculaire, des hélices α transmembranaires, ... (Oxender et Fox ,1988 p176). Dans cette optique, on reporte en ordonnée l'hydrophobicité moyenne d'un groupe d'aa de longueur donnée (ou "fenêtre") que l'on déplace le long de la séquence peptidique placée en abscisse.

Le choix de la taille de la fenêtre est fonction du type de structure que l'on veut mettre en évidence. Une fenêtre de 7 aa permet de détecter des sites antigéniques potentiels, alors que des segments intra-membranaires apparaîtront mieux avec une fenêtre d'une vingtaine d'aa (Von Heijne, 1989: p103) (fig13).

I.3.2.5 Les moments hydrophobiques:

Le moment hydrophobique d'un aa peut être mesuré de plusieurs manières différentes (Eisenberg, 1982 et 1984). Quand la structure tridimensionnelle de la protéine est connue, c-à-d quand les coordonnées des atomes sont déterminées, la valeur du moment hydrophobique " μ_s " d'une structure secondaire est donnée par la relation:

$$\mu_s = \sum_{n=1}^N H_n s_n$$

où H_n est l'hydrophobicité du $n^{\text{ième}}$ résidu et s_n un vecteur unitaire pointant dans la direction du centre géométrique de la chaîne latérale avec comme origine, le $C\alpha$ de ce résidu. Le moment hydrophobique de la structure secondaire est donc un vecteur perpendiculaire à l'axe longitudinal de cette structure et de grandeur égale à la somme des moments hydrophobes des résidus qui la constitue.

On peut également évaluer le moment hydrophobique d'une structure secondaire par la relation suivante:

$$\mu = \left\{ \left[\sum H_n \sin(\delta_n) \right]^2 + \left[\sum H_n \cos(\delta_n) \right]^2 \right\}^{\frac{1}{2}}$$

Où δ représente l'angle de rotation formé par les chaînes latérales de deux résidus contigus par rapport à l'axe de la structure. Une hélice α forme un angle d'environ 100° et un plan β un angle d'environ 180° .

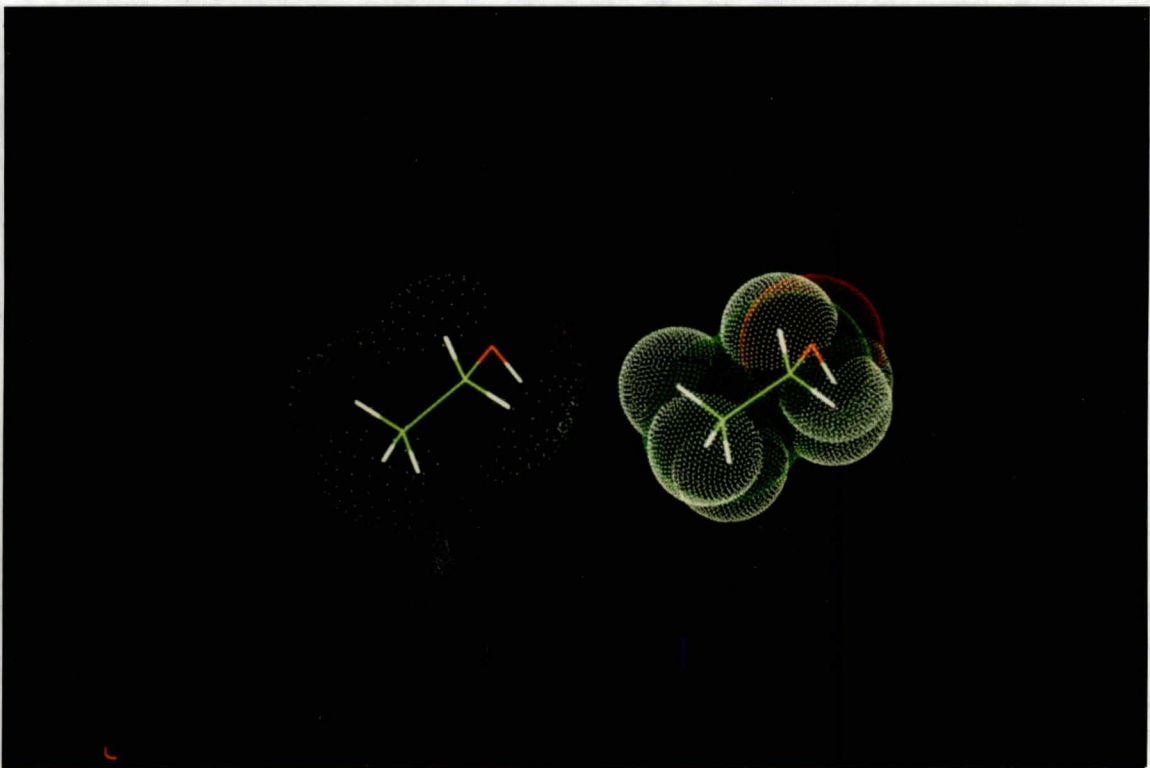


Figure 14:
Illustration des surfaces de Van der Waals (à gauche) et de Connolly (à droite) de l'éthanol.

Quel que soit le mode d'évaluation, des valeurs élevées de μ_s correspondent à l'apparition périodique de résidus hydrophobes; autrement dit, elles traduisent l'amphiphilicité du segment considéré. L'analyse des conformations en terme de moments hydrophobiques permet de mieux cerner le rôle de l'hydrophobicité dans le repliement des chaînes polypeptidiques ou dans la formation de protéines multimériques (Rose *et al*, 1989).

Les moments hydrophobiques des diverses structures secondaires tendent à présenter des valeurs opposées au sein d'une même protéine; ceci peut constituer un critère d'évaluation d'un modèle tridimensionnel (Rose *et al*, 1989).

I.2.3.6 Surface externe de la protéine

Les protéines globulaires ont une surface externe grossièrement sphérique. On définit principalement cette surface de deux manières différentes: la surface accessible et la surface de Van der Waals.

La surface de Van der waals est définie par l'ensemble des sphères de Van der Waals des atomes extérieurs de la molécule (fig 14).

La surface accessible au solvant ou surface de Connolly (Connolly, 1985) (fig14) est définie comme étant la surface décrite par le centre d'une sphère (molécule du solvant) parcourant la surface de Van der Waals de la protéine.

I.3.3 PREDICTION DE STRUCTURES TRIDIMENSIONNELLES:

Le développement d'une méthode de prédiction de structures protéiques doit tenir compte de l'ensemble des interactions présentes au sein de la molécule mais aussi de celles exercées par le milieu dans lequel elle se trouve. Or ces interactions sont encore mal connues. La mise au point d'une méthode de prédiction universelle et fiable constitue un des domaines de recherche en biologie moléculaire pour les années à venir.

A l'heure actuelle, la plupart des prédictions se concentrent sur les structures secondaires; hélices α , plans β et loops. Deux tendances se dégagent parmi les quelque 20 méthodes proposées actuellement.

La première repose sur des considérations théoriques et fait largement appel aux notions physicochimiques développées dans le chapitre précédent. Vasques et Sheraga (1985) basent leurs évaluations sur des calculs d'énergie libre, Ptitsyn et Finkelstein (1989) sur l'hydrophobicité, Lim (1974) sur un ensemble de principes d'assemblage de structures secondaires.

Dans la seconde, les prédictions sont basées sur des constatations se dégageant de l'analyse de structures cristallographiques. La méthode la plus connue est celle proposée par Chou et Fasman en 1974 (Chou et Fasman, 1974). Selon cette méthode, les aa d'une protéine de structure inconnue sont supposés adopter la conformation dans laquelle ils sont le plus souvent observés dans d'autres protéines dont la structure est connue. Une autre méthode (Garnier 1978) considère que la conformation adoptée par chaque résidu est fonction des 8 aa de part et d'autre de ce résidu. Une valeur est attribuée à chacun des 16 résidus en fonction du résidu central et de la structure secondaire potentielle. La conformation la plus probable est celle dont la somme des valeurs recueillies sur la "fenêtre" des 17 aa est la plus élevée.

Malgré les taux de fiabilité élevés annoncés par les auteurs, toutes ces méthodes donnent des résultats assez peu fiables pour des protéines autres que celles ayant servi à la détermination des paramètres de prédiction (50% à 60% de prédiction correcte) (Von Heijne, 1989: p92) Il semble donc qu'il faille envisager la prédiction sous un autre jour et des recherches actives sont actuellement menées en ce sens par de nombreux chercheurs.

I.3.4. MODELISATIONS DE STRUCTURES TRIDIMENSIONNELLES

Nous avons vu dans un chapitre précédent qu'une protéine- dans sa forme native- tend à adopter une conformation stable, correspondant probablement à un minimum de sa fonction énergie libre. Rappelons qu'une controverse existe quant à savoir si la structure native d'une protéine correspond au minimum global ou à un minimum local de sa fonction énergie libre (§I.2.2.4). Des techniques de simulation du repliement des protéines ont été mises au point dans le but de tester ces hypohèses. Malheureusement, plusieurs problèmes de taille (au sens propre et au sens figuré...) limitent considérablement les investigations dans ce domaine.

Le concept d'énergie libre est donc un bon critère de caractérisation de la structure tridimensionnelle d'une protéine.

$$\Delta G = -RT \text{Ln} [e^{-\Delta v / RT}]_a$$

où a= nbre d'atomes du
système

$$\Delta v = \Delta E \text{ potentielle}$$

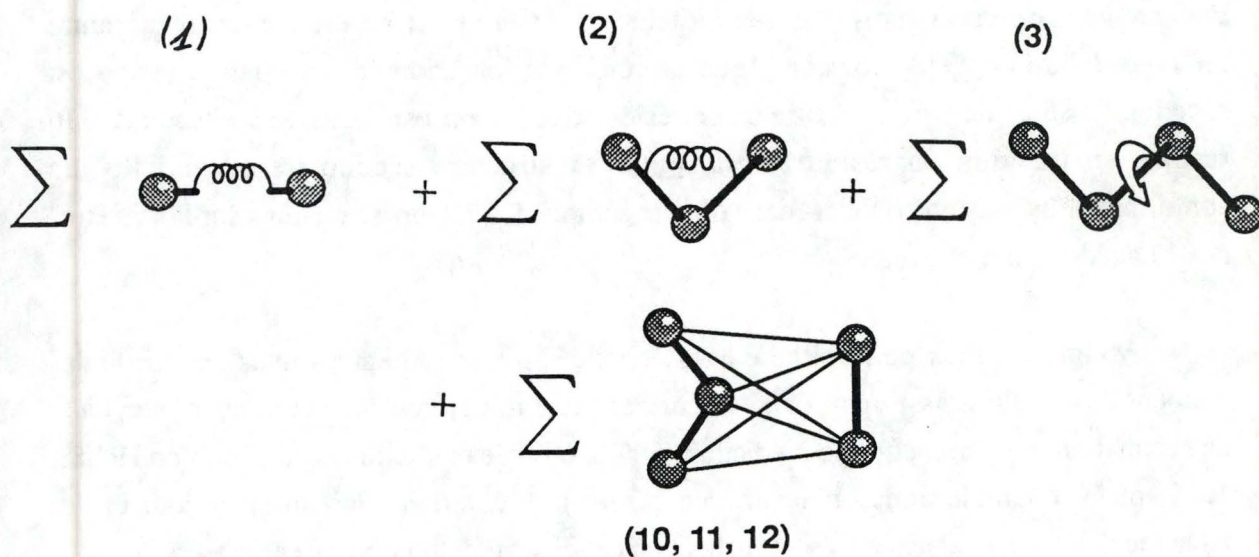


Figure 15

Illustration schématique des interactions covalentes (1); des interactions dans les angles de valence (2); des interactions dans les angles de torsion (3) et des interactions non covalentes comprenant les interactions de VDW, par ponts hydrogène et électrostatiques (10, 11 et 12) (Discover, 1989)

La variation d'énergie libre d'un système peut donc être calculée à partir de sa variation d'énergie potentielle. Idéalement, l'énergie potentielle (E_p) d'un système est obtenue à partir de fonctions quantiques très complexes et qui ne sont applicables à des systèmes aussi importants que les protéines (Mc Cammon, 1984). Il fallait donc trouver d'autres fonctions permettant d'évaluer, avec le plus de précision possible, l'état énergétique d'un système aussi important qu'une protéine. C'est pourquoi ont été développées depuis quelques années des fonctions empiriques dont le traitement mathématique est beaucoup moins laborieux que les précédentes. L'ensemble de ces fonctions ainsi que les valeurs attribuées à leurs paramètres sont regroupées sous le terme générique de "force field".

I.3.4.1 Les "force fields"

Le force field le plus communément utilisé décrit l'énergie potentielle d'une molécule de la manière suivante (Mc Cammon, 1988: p40):

$$E_p = \frac{1}{2} \sum_{\text{bond}} K_b (b - b_0)^2 + \frac{1}{2} \sum_{\text{valence}} K_b (\theta - \theta_0) + \frac{1}{2} \sum_{\text{torsions}} K_\psi [1 + \cos(n_\psi - \delta)]$$

$$+ \sum_{\text{VDW}} \left(\frac{A}{r^{12}} - \frac{B}{r^6} \right) + \sum_{\text{Electrostatique}} \left(\frac{Q_1 Q_2}{D \cdot r} \right)$$

Les différentes interactions reprisent dans cette fonction sont illustrées à la fig(15).

Les liens covalents :

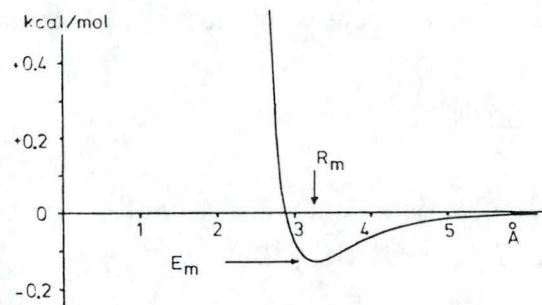
Les liens covalents sont décrits selon une fonction de type quadratique ou K_b est une caractéristique de la paire atomique et b_0 la distance d'équilibre.

Les angles de valence:

Les paramètres K et θ sont respectivement les analogues des constantes décrite ci-dessus.

Les angles de torsion:

La fonction énergie qui leur est associée est périodique avec des minima caractéristiques dont nous avons parlés dans un chapitre précédent. ψ représente la valeur de l'angle de torsion considéré et δ l' angle de torsion dont l'énergie potentielle est minimale. La contribution de ces énergies à l'énergie globale est assez élevée (en seconde position, par ordre décroissant après les liaisons covalentes)



Lennard-Jones 6-12 potential for dispersion forces and electron repulsion ($R_m = 3.24$ Å, $E_m = -0.13$ kcal/mol). The two-parameter formula is given below in its computationally practical form (parameters A , B) and in its normalized form (parameters E_m , R_m). Because of repulsion between electronic shells and attraction by dispersion forces, A and B are positive, and there exists a relative minimum. With smaller B and larger A , E_m becomes smaller and the corresponding atomic distance R_m larger. The potential wall is asymmetric. Repulsion balances attraction when the atomic distance is reduced to $0.89 \cdot R_m$. On the other side the attraction energy is still one-sixth of E_m at a distance of $1.5 \cdot R_m$.

Figure 16

Allure de la variation de l'énergie de VDW en fonction de la distance interatomique (Schulz, 1984)

Les interactions de Van Der Waals:

Elles comprennent deux termes; le terme en r^{12} correspond aux répulsions entre les orbitales électroniques des partenaires r étant la distance interatomique. Le terme en r^6 correspond aux attractions entre dipôles temporaires apparaissant sur les partenaires. Les constantes A et B sont caractéristiques de la paire atomique considérée. L'allure de la fonction énergie potentielle est donnée à la figure (16). Les interactions par ponts hydrogène sont également reprises par cette fonction

Les interactions électrostatiques:

Elles interviennent entre deux atomes chargés. Q est la charge nette des deux atomes et D la constante diélectrique du milieu (80 dans l'eau).

Toute la problématique du développement des "force fields" réside dans la reproduction fidèle de l'ensemble des interactions intra-moléculaires et donc dans la meilleure estimation possible l'énergie du système. Dans cette optique, certains force fields comprennent des termes supplémentaires, les "cross terms", décrivant les interactions entre les différents types de liaisons envisagées ci-dessus (voir Matériel et Méthodes § II.3.3.1).

I.3.4.2 La minimisation d'énergie:

La première idée à été de minimiser l'énergie potentielle de la chaîne polypeptidique d'une protéine se trouvant dans une configuration aléatoire de manière à atteindre le minimum correspondant à sa conformation native. Du point de vue de la simulation, il s'agit de modifier les valeurs initiales des coordonnées atomiques de manière à ce que l'énergie potentielle de la conformation obtenue soit minimum.

Selon l'approche analytique, les valeurs des paramètres correspondant au minimum d'une fonction sont celles annulant la dérivée première de cette fonction. Il est malheureusement impossible de trouver des solutions analytiques pour une fonction telle que celle décrivant l'énergie potentielle d'une molécule.

L'autre approche de ce problème consiste à estimer ces valeurs par des méthodes numériques. Le principe général de ces méthodes est de modifier progressivement les coordonnées atomiques de manière à ce que l'énergie globale de la molécule tende à diminuer le plus rapidement possible (Chou et Fasman, 1989 p321-325).

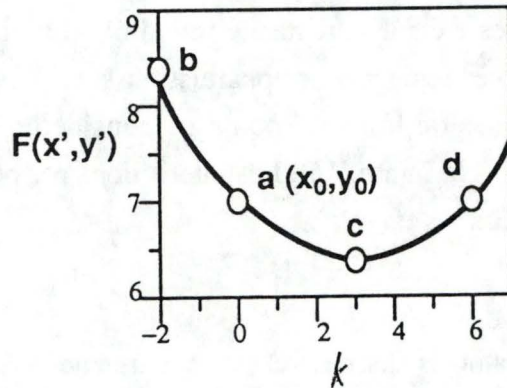


Figure 17

Evolution de l'Ep ($F(x', y')$) en fonction de k . Le minimum de la fonction se situe en (c). Un déplacement dans la direction du gradient positif à partir de (a) mène vers (b); un déplacement selon le gradient négatif mène vers (c). Si k est trop important, l'énergie potentielle réaugmente (Discover, 1989)

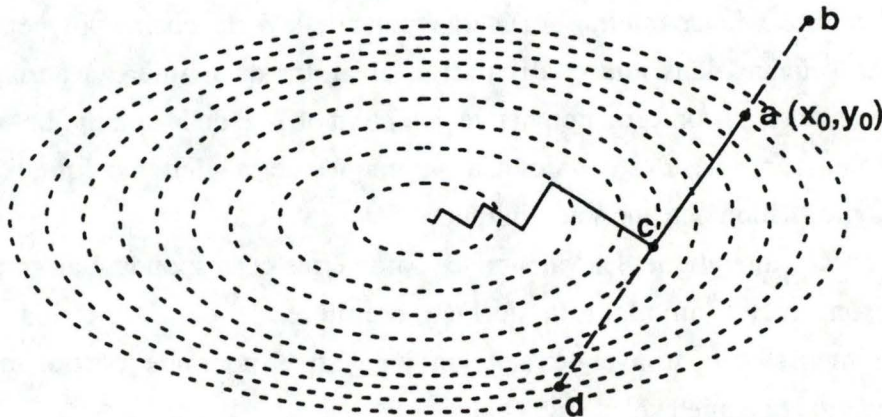


Figure 18

Recherche du minimum de la fonction $E_p = x^2 + 5y^2$ par la méthode de "Steepest descent". Le trajet entre (b), (a), (c) et (d) est celui illustré à la fig 19. Les lignes en pointillés sont les lignes de potentiel constant (Discover, 1989)

Exemple:

Soit la fonction $E_p(x,y) = x^2 + 5y^2$ Fig()

Partant d'un point donné (x_0, y_0) , la dérivée de la fonction en ce point est un vecteur de coordonnées $(dE_p/dx, dE_p/dy)$.

On peut donc se déplacer le long de ce vecteur jusqu'à un autre point (x', y') où la dérivée est recalculée et ainsi de suite.

$$x' = x_0 + k \cdot dE_p/dx$$

$$y' = y_0 + k \cdot dE_p/dy$$

Un graphe de $E_p(x', y') = F(k)$ (fig 17) montre bien qu'un déplacement selon la direction du gradient positif tend à générer une conformation d'énergie supérieure. On choisit donc un déplacement dans la direction du gradient négatif. Les gradients étant tous perpendiculaires les uns aux autres, on progresse par oscillations successives (fig 18) vers le minimum de la fonction.

L'application de ce principe à la fonction énergie d'une protéine nécessite l'intervention de puissants ordinateurs car le nombre d'équations à résoudre est énorme.

rappel:

$$E_p = \frac{1}{2} \sum_{\text{bond}} K_b (b - b_0)^2 + \frac{1}{2} \sum_{\text{valence}} K_b (\theta - \theta_0) + \frac{1}{2} \sum_{\text{torsions}} K_\psi [1 + \cos(n_\psi - \delta)]$$
$$+ \sum_{\text{VDW}} \left(\frac{A}{r^{12}} - \frac{B}{r^6} \right) + \sum_{\text{Electrostatique}} \left(\frac{Q_1 \cdot Q_2}{D \cdot r} \right)$$

Il faut donc, dériver partiellement cette équation par rapport à chaque variable intervenant dans les différents termes et calculer le déplacement à appliquer à chaque atome en vue de minimiser globalement l'énergie de la molécule. Plusieurs techniques basées sur ce principe ont été développées (Discover, 1989):

a) "Steepest Descent":

Si l'amplitude du déplacement atomique est trop importante le long du gradient négatif, elle risque de générer une conformation d'énergie plus élevée.

La méthode de steepest descent évalue le déplacement de la manière suivante:

- La direction suivie correspond au gradient négatif de l'Ep. De cette manière, on est sûr que l'énergie diminue. Pour rappel, $-dE_p/dx,y,z = F(x,y,z)$; chaque atome est donc déplacé selon la direction de la force auquel il est soumis, ce qui peut être formulé de la manière suivante:

$$(x',y') = (x_0,y_0) + k \cdot dE_p/d(x,y)$$

où (x',y') sont les coordonnées atomique après déplacement.

(x_0,y_0) sont les coordonnées atomiques initiales.

k est une constante propre à chaque atome.

- Une certaine valeur est attribuée à k
- Les atomes sont déplacés en fonction de k
- L'énergie du système est recalculée; si elle diminue, k est augmenté jusqu'à ce que le déplacement de l'atome provoque un accroissement de son énergie.
- le nouveau gradient est calculé à partir de la position atomique précédant l'augmentation de l'énergie.

Ceci a pour effet de gommer des oscillations trop importantes autour des minima d'énergie potentielle successifs.

L'inconvénient majeur de cette méthode est que le gradient diminue lorsque l'énergie se rapproche du minimum. Quand il est proche de 0, la convergence vers le minimum est également presque nulle. On utilise donc la méthode de steepest descent en début de minimisation pour des structures mal définies dont le gradient est important.

b) Le "conjugate gradient":

Pour éviter les oscillations propres au steepest descent, la technique du conjugate gradient utilise une fonction du gradient comme direction de déplacement le long de la fonction énergie potentielle de chaque interaction. Ceci a pour effet d'augmenter la vitesse de convergence vers le minimum de la fonction, surtout lorsque le gradient diminue en fin de minimisation.

c) Autres méthodes:

Les méthodes d'ajustements linéaires présentées ci-dessus sont parmi les plus simples à mettre en oeuvre et sont donc facilement applicables à des systèmes aussi importants que les protéines. D'autres méthodes, beaucoup plus complexes, existent mais sont d'utilisation moins courante dans le domaine la modélisation moléculaire.

d) Limites d'application

La minimisation de l'énergie d'une configuration particulière provoque le déplacement des atomes de manière à ce que l'énergie diminue le plus rapidement possible. Or, la fonction énergie potentielle d'une protéine présente des minima locaux, chacun étant représentatif d'une conformation métastable. Des barrières énergétiques souvent importantes existent entre ces minima. La minimisation ne permettant pas à l'énergie d'augmenter, elle a donc toutes les chances d'aboutir à un minimum local de la fonction ne correspondant pas à la conformation native de la protéine.

Pour tenter de remédier à ce problème, des techniques de recherches systématiques ont été mises au point. L'idée est de considérer un grand nombre de conformations possibles pour une chaîne polypeptidique donnée et de sélectionner celle dont l'énergie est la plus faible. Deux méthodes ont été développées dans cette optique: la simulation dite de MONTE CARLO et la DYNAMIQUE MOLECULAIRE.

I.3.4.3 La méthode de Monte Carlo:

Le principe de cette méthode est de générer au hasard différentes structures tridimensionnelles en modifiant localement et de manière pseudo-aléatoire, les coordonnées du système (Mc Cammon, 1988: p38) . Si la conformation obtenue possède une énergie libre

plus faible, elle est retenue et servira de base à la simulation suivante. Si son énergie est plus élevée, elle n'est retenue que si sa probabilité d'occurrence est acceptable (Frühbeis et al , 1987). On dispose alors d'un grand nombre de conformations parmi lesquelles on peut repérer celle correspondant au minimum global d'énergie.

I.3.4.4 La dynamique moléculaire:

Toutes les molécules se trouvant dans un environnement dont la température est supérieure au 0° absolu sont soumises au mouvement brownien. L'agitation des atomes est fonction de la température du milieu et donc de la quantité d'énergie (chaleur) qui leur est communiquée.

Si la température est assez élevée la molécule- une protéine en l'occurrence - peut surmonter certaines barrières énergétiques et changer de conformation.

Il est possible de simuler ce phénomène en considérant chaque atome d'une molécule comme un corps de masse donnée. Sa direction de mouvement est déterminée par la force résultante auquel il est soumis (Discover, 1989).

La force - $F(x,y,z)$ - exercée sur chaque atome est donnée par la dérivée négative de son énergie potentielle par rapport à sa position dans le système:

$$F = -dE_p / dr = m.a$$

où r = coordonnées cartésiennes de
l'atome

m = masse de l'atome

a = accélération

Dans la réalité, le déplacement des atomes est continuellement réajusté puisque la moindre variation de leurs coordonnées change leur E_p et donc la direction, le sens et l'amplitude du vecteur force.

Lors d'une simulation, connaissant l'accélération subie par la particule ainsi que sa vitesse initiale, on calcule l'espace parcouru par chaque atome sur un intervalle de temps très court ("méthode de Verlet" Mc Cammon, 1988:p173). Une fois le déplacement effectué, on réévalue l'énergie potentielle du système, la nouvelle direction de déplacement et le processus se poursuit. La taille de l'intervalle doit être adaptée à la température du milieu car plus la T° augmente, plus la vitesse des atomes est élevée et plus le déplacement par unité de temps est important. Des déplacements trop importants génèrent des interactions interatomiques non représentatives du phénomène réel. A la température de 300° K, la direction du déplacement des atomes doit être recalculée toutes les femtoseconde (10^{-15} sec) au maximum.

Pour explorer différentes conformations il faut permettre aux atomes de franchir les barrières énergétiques existant entre les différents états métastables de la protéine. La dynamique est réalisée de manière à ce que l'énergie totale du système (E potentielle +E cinétique) soit supérieure à n'importe quelle barrière énergétique et qu'elle reste constante durant toute la simulation. La relation entre la vitesse des atomes de la protéine et la température du milieu est donnée par l'équation suivante (Mc Cammon, 1988: p 64):

$$3k_b \cdot T = \sum_{i=1}^N m_i \cdot v_i \cdot v_i / N$$

ou k_b est la constante de Boltzmann, m la masse de l'atome, v sa vitesse et N le nombre total d'atomes de la molécule. A la fin de chaque déplacement, l'énergie totale de la protéine est réajustée par modification de la T° en fonction de la variation de son énergie potentielle. On peut donc suivre l'évolution de la T° ou de l'énergie potentielle au cours du temps et repérer les minima d'énergie potentielle et/ou les maxima d'énergie cinétique (T°).

Limites d'application:

On pourrait penser, au vue de ces méthodes, que l'on détient un moyen infaillible de détecter le minimum correspondant à la structure native de la protéine! Un obstacle majeur rend malheureusement ce projet totalement irréaliste; le nombre de conformations possibles pour une protéine de taille moyenne est gigantesque et il est impensable de pouvoir les passer systématiquement en revue.

Prenons l'exemple d'une protéine adoptant sa configuration native en une milliseconde. En admettant que le calcul nécessaire à la simulation d'une femtoseconde du repliement soit achevé en une seconde de calcul, la simulation complète du repliement de la protéine durerait 10^{12} sec, c-à-d environs 30.000 ans (à quelques secondes près).

En pratique, les ordinateurs les plus performants permettent des simulations sur des périodes d'environ 0.1 nanoseconde pour des systèmes de quelques milliers d'atomes c-à-d une période inférieure de plusieurs ordres de grandeur à l'échelle de temps caractéristique des phénomènes biologiques.

La question du minimum correspondant à la structure native d'une protéine n'est donc toujours pas résolu. Pour remédier à ces difficultés, une autre approche de la modélisation à été développée:

I.3.5 "MODELISATION PAR SIMILARITE DE STRUCTURE AVEC DES PROTEINES ANALYSEES PAR CRISTALLOGRAPHIE:

Le nombre croissant de séquences connues, et les tentatives infructueuses de prédictions de structures *ab-initio* ont incité les chercheurs à explorer de nouvelles méthodes de modélisation.

L'étude parallèle des caractéristiques structurales et séquentielles de protéines homologues a fait apparaître que si l'évolution affecte profondément et rapidement les séquences peptidiques, les motifs structuraux qui en découlent sont, globalement, remarquablement conservés (Blundell *et al*, 1987). Inversement, des analyses de structures tridimensionnelles de protéines homologues montrent que des similarités de structures primaires révèlent souvent des zones de la protéine dont la conformation est semblable.

On pourrait donc, à priori, modéliser l'aspect d'une protéine inconnue à condition que sa séquence soit proche de celle d'une autre protéine dont la structure est connue. Cette approche permettrait en outre de surmonter le principal obstacle à la prédiction *ab-initio*, à savoir la multiplicité des conformations possibles pour une séquence donnée. Pour envisager ce type de modélisation, il faut d'abord un moyen d'évaluer la ressemblance entre la séquence d'une protéine inconnue et une protéine déjà cristallisée.

I.3.5.1 Les alignements de séquences:

:

Commençons par définir quelques termes:

- Deux aa sont identiques si il s'agit des mêmes résidus dans les deux séquences.
- Deux aa sont similaires si ils possèdent des propriétés physicochimiques (hydrophobicité,...) semblables.
- Deux aa sont homologues si ils apparaissent dans deux protéines ayant des origines communes.

Le but d'un alignement est donc d'obtenir un maximum de correspondances (similarités, identités, homologie) entre la séquence d'une protéine connue et la séquence considérée.

	A	B	C	D	E	F
A	1	0	0	0	0	0
B	0	1	0	0	0	0
C	0	0	1	0	0	0
D	0	0	0	1	0	0
D	0	0	0	1	0	0
E	0	0	0	0	1	0

Alignement 1:

A B C D _ E F
A B C D D E

Alignement 2:

A B C _ D E F
A B C D D E

Fig:19

Deux fragment de séquences fictives sont alignées par la méthode de Needleman & Wunsch. Dans ce cas, deux alignements sont possibles et sont illustrés en dessous du tableau. D'autres alignements seraient réalisables mais présenteraient un taux d'identité plus faible.

Plusieurs types d'alignements ont été développés depuis une vingtaine d'années. Toutes ces méthodes sont informatisées, ce qui permet l'intégration d'un très grand nombre de critères ainsi que le traitement d'un maximum d'informations.

On distingue principalement les alignements de séquences, soit simples (deux séquences), soit multiples (plus de deux séquences) et les alignements de structures tridimensionnelles.

a) Les alignements simples:

La plupart des alignements dérivent de la méthode proposée par Needleman et Wunsch en 1970.

Principe:

Les séquences à aligner, respectivement de longueurs $i=1\dots n$ et $j=1\dots m$, sont disposées sur deux axes perpendiculaires X et Y. Chaque résidu d'une séquence est comparé à tous les résidus de l'autre séquence.

On obtient une matrice (m,n) dont les éléments correspondant à une identité sont "1" et les autres "0".

L'alignement incluant le plus grand nombre d'identités, correspond au déplacement à travers la matrice dont la somme des éléments rencontrés est la plus élevée (fig 19).

Technique:

Les alignements sont construits selon des tracés ayant pour origine l'élément (m,n) et se poursuivant d'identité en identité de manière à ce qu'un même aa ne soit apparié qu'à un seul aa de l'autre séquence. Les morceaux de tracés n'appartenant pas à la diagonale principale impliquent la présence de délétions ou d'insertions dans une des séquences. Un score global est attribué à chaque alignement sur base du nombre d'identités de la taille des délétions/insertions. La méthode la plus simple est d'additionner les identités et de retirer ensuite une valeur (pondération) proportionnelle au nombre et à la taille des dél./ins. .

Le meilleur alignement est celui dont le score global est le plus élevé. L'utilisateur de cette méthode ne doit jamais oublier que l'alignement obtenu est fonction du type de pondération utilisée. On peut, par exemple, donner plus d'importance à la taille qu'au nombre de délétions/insertions.

	Gly	Ala	Val	Leu	Ile	Met	Cys	Ser	Thr	Asn	Gln	Asp	Glu	Lys	Arg	His	Phe	Tyr	Trp	Pro
Gly		11	38				16	16				21	21		43					13
Ala	58		41					35	39			23	26							36
Val	10	37		48	18	12						20	23				16			
Leu	2	10	30		16	21		13			16				22	11	41		9	26
Ile		7	66	25		25		15	17	18							16			
Met	1	3	8	21	6				11					19	7					
Cys	1	3	3		2			21							10		12	15	16	
Ser	45	77	4	3	2	2	12		41	15					30		13	15	8	24
Thr	5	59	19	5	13	3	1	70		17				24	12					28
Asn	16	11	1	1	1			43	17			19		49		13		18		
Gln	3	9	3	8	1	2		5	1	5			23	24	13	26				15
Asp	16	15	2		1			10	6	53	8		49			15		20		
Glu	11	27	4	2	1	1		9	3	9	42	83		30						
Lys	6	6	2	4	4	9		17	20	32	15		10		20					
Arg	1	3	2	2	3	2	1	14	2	2	12	9		48		8			13	20
His	1	2	3	4			1	3	1	23	21	4	2	2	10			13		10
Phe	2	2	1	17	9	2		4	1	1					1	2		16		
Tyr		2	2	2	1		3	2	2	4			1	1		4	26			
Trp				1				2							3		1	1		
Pro	5	35	5	4	1		1	27	7	3	9	1	4	4	7	5	1			

Observed Number of Replacements

Figure 3-9

Relative frequencies of amino acid replacements observed in a total of 1572 examples between closely related proteins (bottom, left) and those expected for random single-site mutations (top, right), calculated from the genetic code and the frequencies of occurrence of the different amino acid residues. The greatest discrepancies are shown in bold type. The observed replacements between chemically similar amino acids are generally much more frequent than expected for random mutations. (Observed replacements compiled by M. O. Dayhoff (ed.), *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, Washington, D.C., National Biomedical Research Foundation, 1978.)

Table 5

Fréquence relative de remplacement d'un aa par un autre au sein des 1552 protéines homologues (Dayhoff, 1978).

Principales variantes de cette méthode:

Plutôt que de considérer les identités seules, on peut attribuer une valeur à chaque paire d'aa en fonction du degré de similarité entre les résidus. Ces valeurs peuvent être calculées sur base de la distance génétique les séparant, à partir de leur fréquence de remplacement mutuel au sein de protéines homologues (homologie) (Dayhoff, 1972) (table 5) ou encore à partir de leurs hydrophobicité respectives (similarité) (Eisenberg, 1983). Des pénalités plus élevées peuvent être attribuées à des délétions/insertions se trouvant dans des structures secondaires importantes (Fitch et Smith, 1983)

i- Les "dot plots":

Il s'agit d'une méthode de N&W où les alignements sont appréciés de manière visuelle (Maizel et Lenk 1981).

La valeur propre à chaque couple d'aa identiques est remplacée par un point. Les segments diagonaux plus contrastés traduisent la présence de zones présentant une identité élevée (fig 20)

ii- Méthode de Wilbur et Lipman (1983)

L'alignement de très longues séquences peut générer un grand nombre d'alignements possibles parmi lesquels il faut sélectionner celui présentant la similarité la plus élevée.

Les séquences à aligner sont décomposées en groupes de k résidus contigus. Un tableau contenant une liste des différents groupes, ainsi que leurs positions respectives dans chaque séquence, est établi.

Chaque groupe d'une séquence est comparé avec l'ensemble des groupes de l'autre séquence. Les différences entre les positions relatives sont évaluées pour les groupes similaires apparaissant dans les deux séquences. Les groupes d'une séquence possédant la même distance relative par rapport aux groupes qui leur sont similaires dans l'autre séquence, peuvent être alignés sans délétions ni insertions. Plus le nombre de groupes similaires présentant la même distance relative est élevé, plus ces groupes ont de chances d'appartenir à l'alignement présentant le degré de similarité est le plus élevé. Certains alignements partiels peuvent être sélectionnés lorsque leur taux de similarité est significativement supérieur à celui de la moyenne de l'ensemble des alignements partiels. L'alignement global est réalisé sur base de cette sélection.

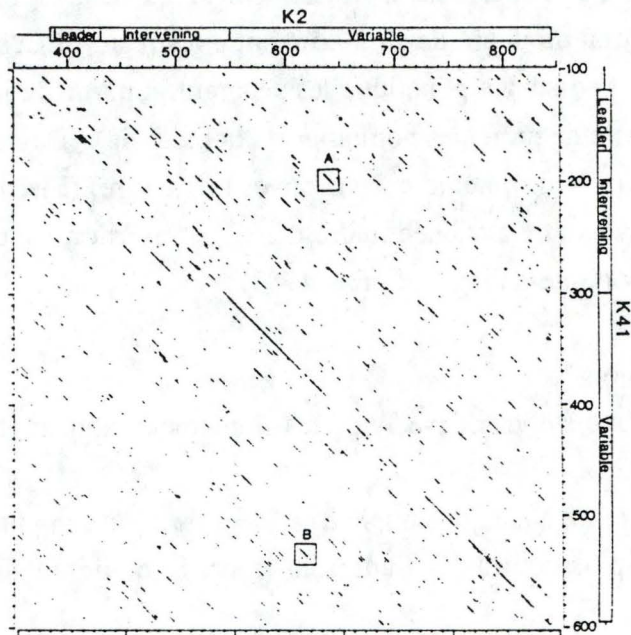


Figure 20

Matrice graphique (ou Dot Plot) de deux séquences nucléotidiques. Les similarités sont évaluées sur base d'une fenêtre de 9 bases (Maizel & Lenk, 81)

b) Les alignements multiples:

On peut facilement comprendre que plus le nombre de séquences comparées est important, mieux les résidus conservés, et donc importants par rapport à la fonction de la protéine sont mis en évidence.

De même que pour les alignements simples, beaucoup de méthodes ont été mises au point. Il est difficile de les détailler toutes dans le cadre de ce travail. Nous allons plutôt tenter de mieux cerner les problèmes des alignements multiples à travers quelques exemples concrets

i- La méthode de Needleman & Wunsch peut être extrapolée à l'alignement de trois séquences (Murata, 1985). Les séquences sont disposées selon trois axes orthogonaux; on obtient ainsi une matrice à trois dimensions sur laquelle on effectue les mêmes opérations que pour l'alignement de deux séquences. On peut également envisager l'alignement de plus de trois séquences mais le temps calcul nécessaire pour trouver le meilleur alignement à travers un espace à N dimensions devient rapidement trop important. Par exemple, 3 séquences de 100 aa chacune peuvent être alignées en une centaine d'heures de calcul sur un ordinateur tel qu'un VAX 11/780; Le traitement de 6 séquences de même taille demanderait quelques 20.000 années de calcul! (Subbiah et Harrison, 1989)

Une série d'options ont été développées pour tenter de remédier à ce problème:

ii- La méthode de Johnson et Doolittle (1986)

Plutôt que d'aligner les N séquences dans leur intégralité, elles sont décomposées et alignées par groupe de k résidus à la fois. Le nombre de résidus alignés simultanément (k) étant relativement restreint, le temps calcul reste limité. L'alignement obtenu n'est malheureusement qu'une approximation de l'alignement optimum.

iii- La méthode de Gribskov et al (1987):

Les alignements sont réalisés sur plusieurs séquences considérées deux à deux ce qui permet la sélection d'une séquence consensus. Le consensus est constitué d'une suite de résidus, chacun étant représentatif de l'ensemble des résidus alignés sur sa position dans la séquence. Le résidu le plus représentatif d'une position est choisi sur base de sa fréquence d'apparition à cette position parmi toutes les séquences alignées. Cette fréquence est pondérée par le score qui lui est attribué selon Dayhoff (1972) par rapport à chacun des résidus qu'il représente. Une fois constitué, le consensus sert de base à un nouvel alignement et ainsi de suite pour l'ensemble des séquences.

Le problème de cette technique est qu'une erreur peut se transmettre d'un alignement à l'autre et donner un consensus inexact. De plus, le résultat final est fortement influencé par l'ordre des alignements successifs.

iv-Méthode de E. Depiereux et E. Feytmans (1989)

Cette approche permet la réalisation d'alignements simultanés. Les séquences à analyser sont assemblées bout à bout en une longue suite de résidus. Un ensemble de critères de similarité sont définis par l'utilisateur. Une fenêtre de taille donnée est déplacée le long de la chaîne polypeptidique de manière à détecter des zones de similarité entre les séquences. Seules sont conservées les zones formant des "groupes de similarité" c-à-d que chaque zone retenue doit être similaire à l'ensemble des zones qui lui correspondent dans les autres séquences. Ceci évite la situation dans laquelle trois protéines A,B,C sont alignées par rapport à une zone de leur séquence avec A similaire à B, B similaire à C mais A différent C. L'alignement est ensuite réalisé selon une méthode faisant appel à l'analyse multivariée.

I.3.5.2 Les alignements de structures tridimensionnelles:

L'alignement est réalisé sur base des coordonnées atomiques de deux protéines (Greer: 80, 81). On réalise une superposition des deux protéines de manière à minimiser l'écart type des distances entre les C α des résidus identiques de manière à convertir les coordonnées des atomes des deux protéines dans le même système de mesure.

Les distances entre les C α des deux protéines sont mesurées. Tous les C α séparés de moins de 1Å sont considérés comme alignés. En général les zones alignées correspondent à des segments dont les séquences sont très conservées.

I.3.5.3 Construction du modèle:

Disposant d'un alignement optimum entre deux séquences, on procède alors au remplacement des résidus de la protéine de référence par ceux de la protéine inconnue. De même Les délétions et les insertions nécessaires sont introduites. Il va de soit que ces modifications de séquences, spécialement les délétions et les insertions, provoquent l'apparition de tensions entre certains atomes de la protéine et déstabilisent sa structure tridimensionnelle. Il faut donc que la structure s'adapte à ces modifications en vue de rééquilibrer l'ensemble des interactions dont elle est le siège et donc de minimiser son énergie. La simulation de ces modifications se réalise grâce aux méthodes décrites dans le § I.3.4.2.

De même, la méthode de dynamique moléculaire peut être appliquée au modèle dans le but d'explorer les configurations correspondant aux minima proches de celui atteint par la minimisation. Ces deux méthodes sont souvent employées conjointement dans la réalisation de modèles de structures moléculaires.

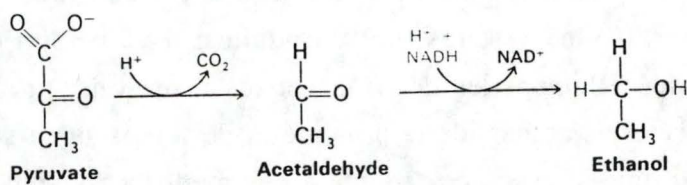
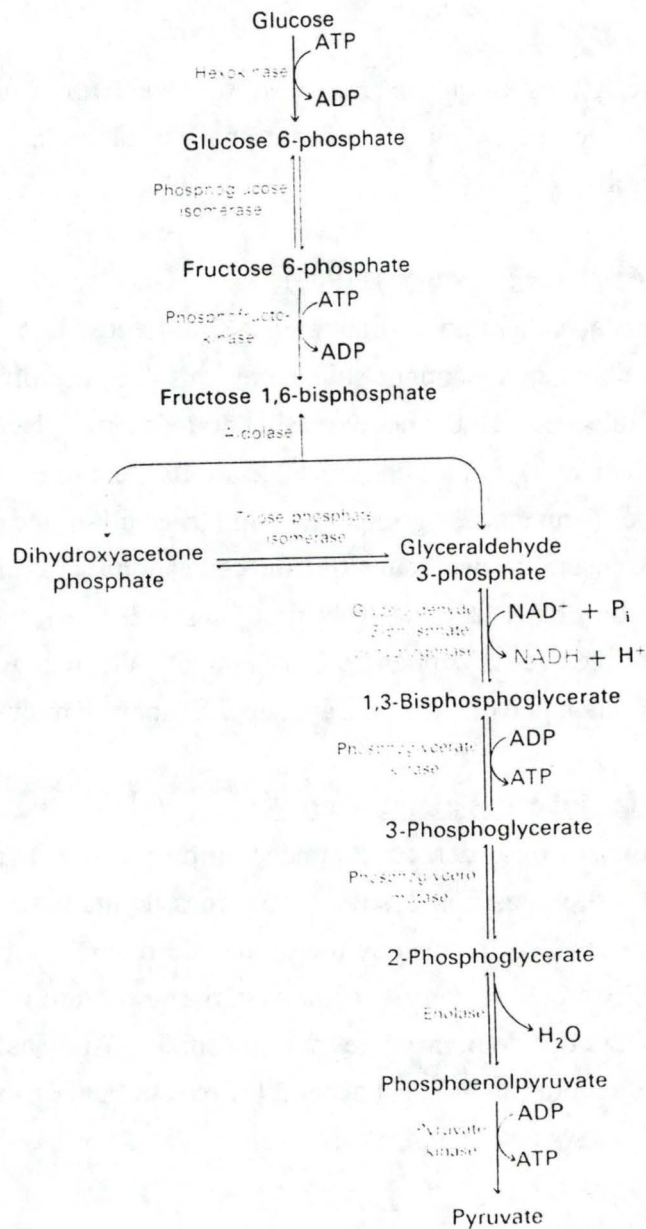


Figure 21

Glycolyse et fermentation alcoolique (Styer, 1988)

I.4 L'ALCOOL DESHYDROGENASE

L'alcool déshydrogénase appartient au groupe des oxydoréductases. Elle oxyde les alcools en aldéhydes et vice versa avec le nicotinamide adénine dinucléotide -forme oxydée (NAD⁺)- comme coenzyme et l'ion Zn²⁺ comme cofacteur.

Rappel:

L'enzyme sans son coenzyme est appelée "Apoenzyme". L'enzyme avec son coenzyme est appelée "Holoenzyme".
le substrat et le coenzyme sont appelés " les cosubstrats"

I.4.1 PLACE DANS LE METABOLISME:

Voir figure (21)

Lors de la glycolyse, la transformation du glucose en pyruvate induit la réduction de NAD⁺ en NADH (nicot. adén. dinucl. forme réduite) . En conditions aérobies, la chaîne respiratoire de la mitochondrie régénère le NAD⁺ et la glycolyse peut se poursuivre. Chez la levure en conditions anaérobies, le pyruvate est transformé en acétaldéhyde (par la *pyruvate décarboxylase*). *L'alcool déshydrogénase* réduit l'acétaldéhyde en éthanol et réoxyde le NADH en NAD⁺.

I.4.2 L'ALCOOL DESHYDROGENASE DE *Saccharomyces cerevisiae*:

L'alcool déshydrogénase de levure (E.C.1.1.1.1: alcool: NAD⁺ oxydoréductase) recouvre en fait quatre isoenzymes. La plus récemment identifiée est l'alcool déshydrogénase IV (YADH IV), isolée en 1986, dont la fonction est encore inconnue (Williamson et Paquin, 1987). L'YADH III est mitochondriale et plus thermostable que les autres formes de l'enzyme (Bille, 1987). Les deux premières isoenzymes sont les mieux caractérisées; elles représentent environ 1% des protéines cytoplasmiques de la cellule.

Bien qu'elle soit réversible, l'YADH I participe essentiellement aux réactions de fermentation. L'YADH II quant à elle, participe à la respiration par l'oxydation de l'éthanol et réduction de NAD⁺(Schimpfessel, 1968). L'étude cinétique des isoenzymes montre par ailleurs que la constante de Michaelis (Km) de L'YADH II en présence d'éthanol est 10 fois plus faible que celui de l'YADH I(Plapp *et al*, 1987).

1 5 10 15
 Acyl-Ser-Ile-Pro-Glu-Thr-Gln-Lys-Gly-Val-Ile-Phe-Tyr-Glu-Ser-His-Gly-Lys-Leu-
 20 25 30 35
 -Glu-Tyr-Lys-Asp-Ile-Pro-Val-Pro-Lys-Pro-Lys-Ala-Asn-Glu-Leu-Leu-Ile-Asn-Val-
 40 45 50 55
 -Lys-Tyr-Ser-Gly-Val-Cys-His-Thr-Asp-Leu-His-Ala-Trp-His-Gly-Asp-Trp-Pro-Leu-
 60 65 70 75
 -Pro-Thr-Lys-Leu-Pro-Leu-Val-Gly-Gly-His-Glu-Gly-Ala-Gly-Val-Val-Val-Gly-Met-
 80 85 90
 -Gly-Glu-Asn-Val-Lys-Gly-Trp-Lys-Ile-Gly-Asp-Tyr-Ala-Gly-Ile-Lys-Trp-Leu-Asn-
 95 100 105 110
 -Gly-Ser-Cys-Met-Ala-Cys-Glu-Tyr-Cys-Glu-Leu-Gly-Asn-Glu-Ser-Asn-Cys-Pro-His-
 115 120 125 130
 -Ala-Asp-Leu-Ser-Gly-Tyr-Thr-His-Asp-Gly-Ser-Phe-Gln-Gln-Tyr-Ala-Thr-Ala-Asp-
 135 140 145 150
 -Ala-Val-Gln-Ala-Ala-His-Ile-Pro-Gln-Gly-Thr-Asp-Leu-Ala-Glu-Val-Ala-Pro-Val-
 155 160 165 170
 -Leu-Cys-Ala-Gly-Ile-Thr-Val-Tyr-Lys-Ala-Leu-Lys-Ser-Ala-Asn-Leu-Met-Ala-Gly-
 175 180 185
 -His-Trp/Val-Ala-Ile-Ser-Gly-Ala-Ala-Gly-Gly-Leu-Gly-Ser-Leu-Ala-Val-Gln-Tyr-
 190 195 200 205
 -Ala-Lys-Ala-Met-Gly-Tyr-Arg-Val-Leu-Gly-Ile-Asp-Gly-Gly-Glu-Gly-Lys-Glu-Glu-
 210 215 220 225
 -Leu-Phe-Arg-Ser-Ile-Gly-Gly-Glu-Val-Phe-Ile-Asp-Phe-Thr-Lys-Glu-Lys-Asp-Ile-
 230 235 240 245
 -Val-Gly-Ala-Val-Leu-Lys-Ala-Thr-Asn-Gly-Gly-Ala-His-Gly-Val-Ile-Asn-Val-Ser-
 250 255 260 265
 -Val-Ser-Glu-Ala-Ala-Ile-Glu-Ala-Ser-Thr-Arg-Tyr-Val-Arg-Ala-Asn-Gly-Thr-Thr-
 270 275 280
 -Val-Leu-Val-Gly-Met-Pro-Ala-Gly-Ala-Lys-Cys-Cys-Ser-Asp-Val-Phe-Asn-Gln-Val-
 285 (290) (295) (300)
 -Val-Lys/Ser-Ile-Ser-Ile-Val-Gly-Ser-Tyr-Val-Gly-Asn-Arg-Ala-Asp-Thr-Arg-Glu-
 (305) (310) (315) (320)
 -Ala-Leu-Asp-Phe-Phe-Ala-Arg-Gly-Leu-Ile-Lys-Ser-Pro-Ile-Lys-Val-Val-Gly-Leu-
 (325) (330) (335) (340)
 -Ser-Thr-Leu-Pro-Glu-Ile-Tyr-Glu-Lys-Met-Glu-Lys-Gly-Gln-Val-Val-Gly-Arg-Tyr-
 (345)
 -Val-Val-Asp-Thr-Ser-Lys.

Figure 22

Structure primaire de l'alcool déshydrogénase de *Saccharomyces cerevisiae* (Jörnvall, 77a).

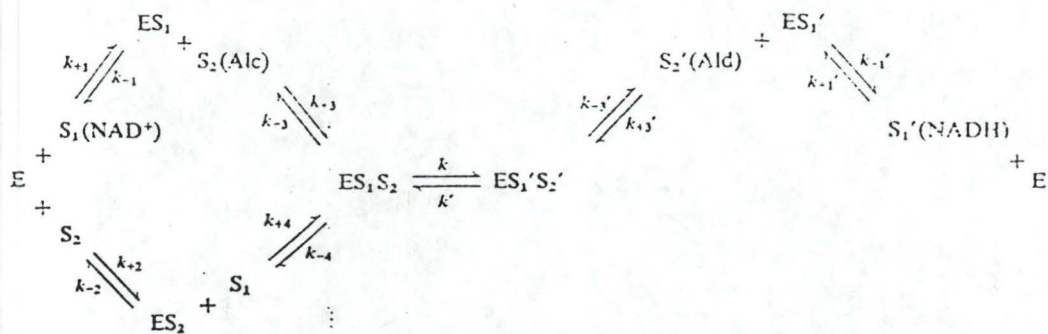


Figure 23

Mécanisme réactionnel ("Compulsory ordered") de l'alcool déshydrogénase de *Saccharomyces cerevisiae* proposé par Dickenson et Dickinson (1977).

Mécanisme Ordonné Bi Bi :

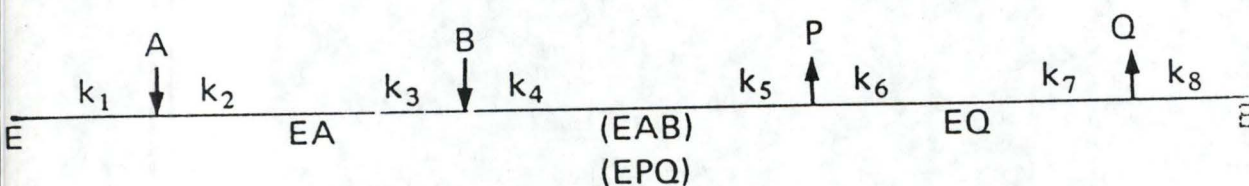


Figure 24

Mécanisme réactionnel (Ordonné Bi-Bi) proposé par Ganzhorn *et al* (1987) pour l'alcool déshydrogénase de *Saccharomyces cerevisiae*

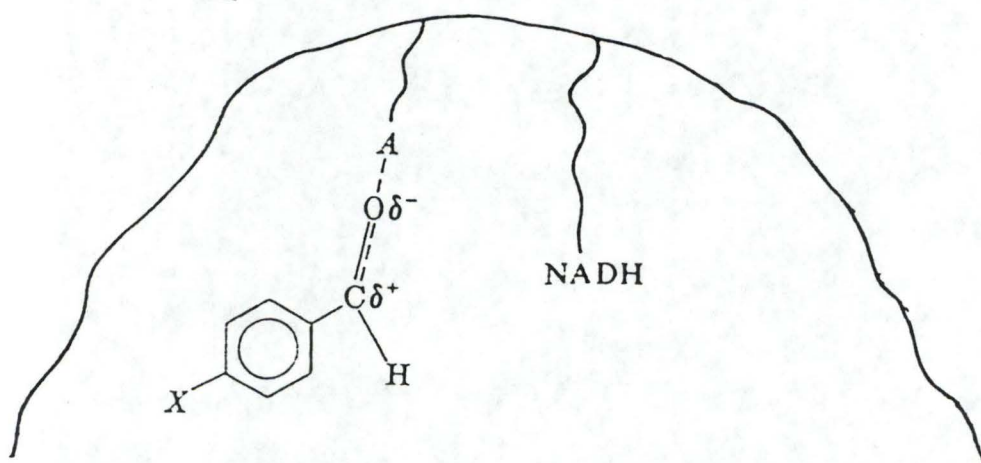


Figure 25

Polarisation exercée par l'ion zinc (A) sur le groupement aldéhydique d'un substrat (cyclohexanol) de l'YADH (Klinman, 1972).

I.4.2.1 L'YADH I:

Cette enzyme de masse moléculaire (MM) de 145 Kilodaltons (Kd) est un tétramère dont les sous-unités sont toutes identiques (Bünher et Sund, 1969) . Sur les quatre monomères présents, seuls deux seraient actifs (Leskovac, 1975). Ce résultat est pourtant contesté par d'autres auteurs ayant relevé quatre sites actifs fonctionnels (Temler et Kägi, 1973). L'enzyme contient au moins un atome de Zn^{++} par monomère (Kaptein, 1989). Le dosage des cystéines libre dans l'enzyme active semble pourtant indiquer la présence d'un seul ion Zinc par monomère (E. Magonet, 1989, non publié) Un seul est essentiel pour l'activité de l'enzyme. L'autre ne semble pas avoir d'influence sur l'activité, ce qui expliquerait qu'il soit parfois absent. L'YADHI possède 8 groupes thiol libres. Branden et al (1975) ont montré que deux cystéines sont indispensables à l'activité catalytique et que leur oxydation libère un ion Zinc.

a) structure primaire:

La structure primaire a été déterminée en 1976 par Jörnvall (Jörnvall, 1977a) fig (22). Chaque monomère est constitué de 347 résidus. La distribution de ces résidus n'est pas homogène; les Arg se situent toutes dans la partie C-terminale alors que les Trp sont situées dans la partie N-terminale. Ces hétérogénéités se rencontrent également pour les Phe, His, Pro, Cys.

b) Mécanisme réactionnel:

La réduction de l'acétaldéhyde en éthanol est décrite comme un "Compulsory ordered mechanism" (fig 23) (Dickenson et Dickinson, 1977). Selon cette hypothèse, le substrat et le NADH se fixent selon un ordre quelconque pour former un complexe ternaire. La réduction de l'acétaldéhyde et l'oxydation du NADH se réalisent dans le complexe ternaire. La dissociation des produits libère d'abord l'éthanol puis le NAD^+ . Cette dernière constitue l'étape limitante du mécanisme catalytique. Ganzhorn *et al* (1987) quant à eux, proposent un mécanisme "ordered bi bi" (fig 24) dont la seule différence par rapport au mécanisme précédent réside dans la fixation ordonnée du NADH puis de l'acétaldéhyde.

La réduction de l'éthanol nécessite à la fois le transfert d'un ion hydruure depuis le NADH vers le carbone de l'aldéhyde et d'un proton depuis le milieu (solvant) sur l'oxygène de l'alcool. L'ion zinc faciliterait ces réactions en polarisant la liaison $C=O$ de l'acétaldéhyde pour donner un alkoxyde (Klinman, 1972) (fig 25). Des études d'inhibitions spécifiques ont montré qu'une seule histidine est essentielle à l'activité enzymatique (Leskovac, 1975); elle participerait dès lors, au transfert de proton sur l'oxygène de l'alkoxyde.

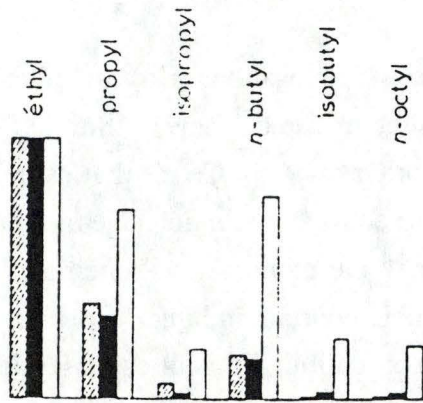


Figure 26

Spectre d'activité de l'alcool déshydrogénase I (en hachuré), II (en gras) et III (en clair) vis à vis de différents substrats (Scimpfessel, 1968)

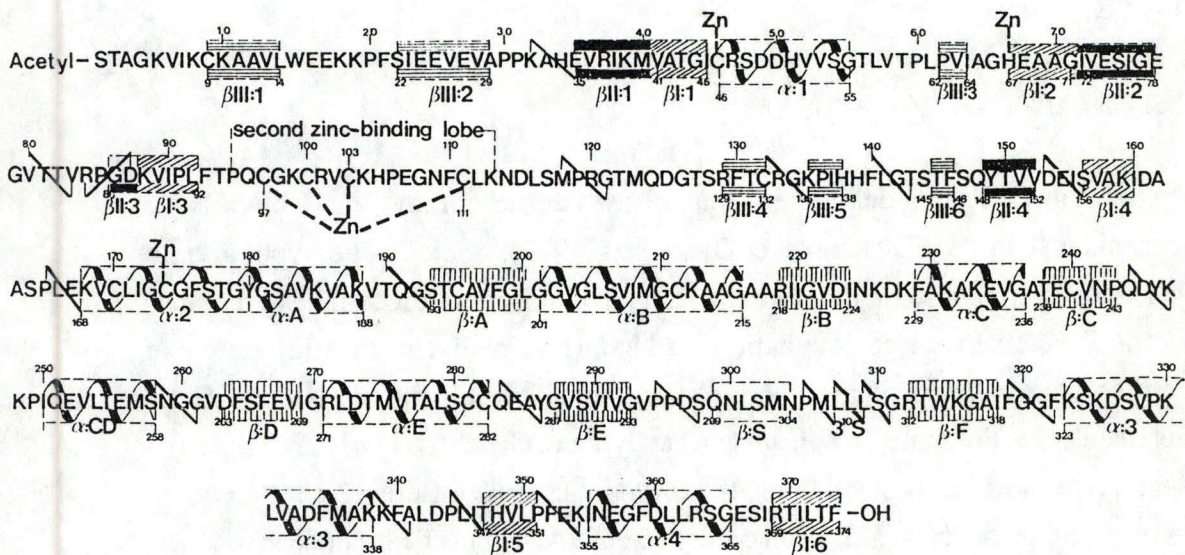


Figure 27

Structure secondaire de l'alcool déshydrogénase de foie de cheval (Jörnvall, 1976).

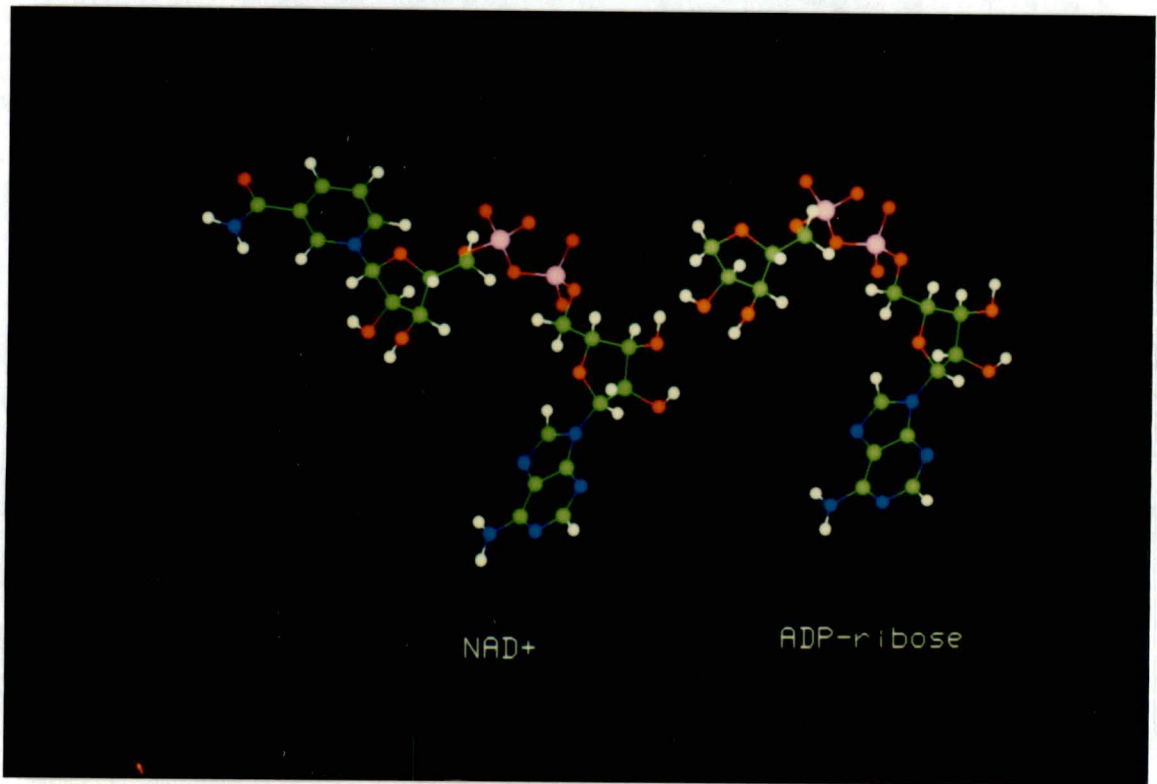


Figure 28

Comparaison entre la structure du NAD^+ (à gauche) et l'ADP-ribose (à droite).

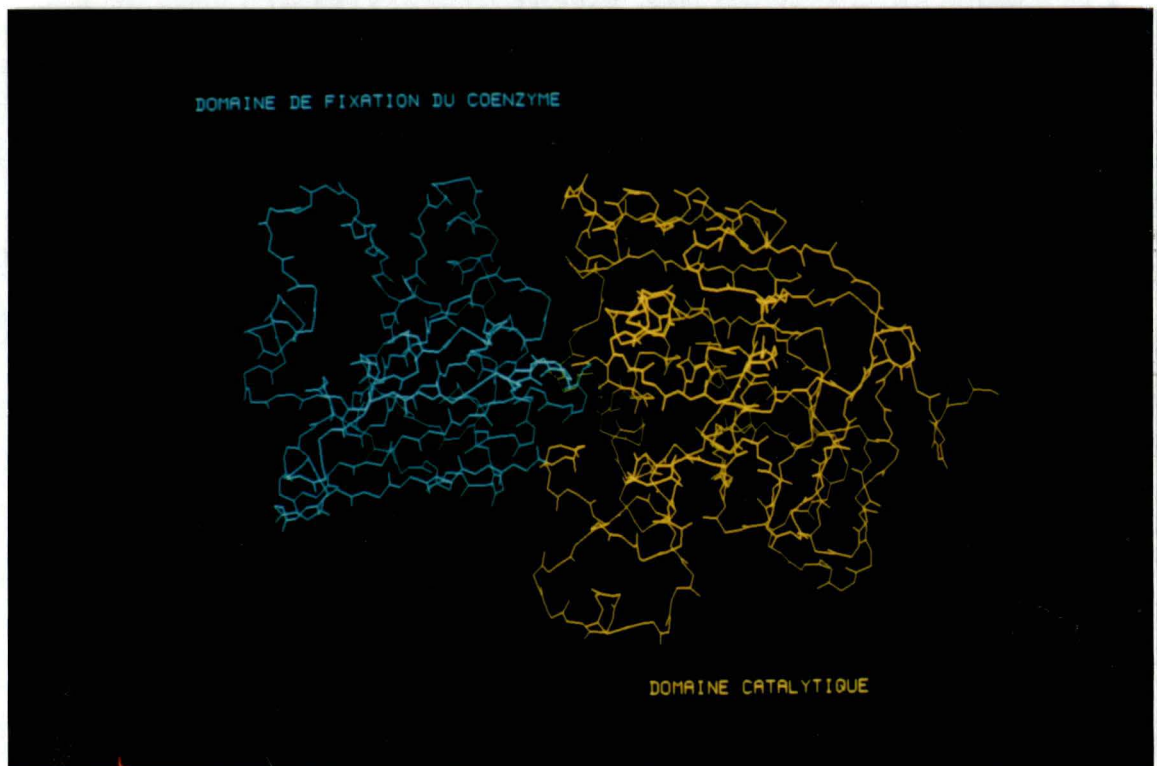


Figure 29

Illustration des domaines de l'alcool déshydrogénase de foie de cheval.

c) Spécificité de l'enzyme:

L'enzyme de levure est, d'une manière générale, plus spécifique que l'enzyme de foie de cheval (Bille, 1987).

Cependant, outre son substrat naturel, l'YADH est capable de réduire et d'oxyder toute une série d'autres substrats. Un aperçu des réactivités envers des alcools à 1,2,3,4,5 et 8 carbones est donné à la figure (26) (Schimpfessel, 1967). Elle réagit mieux avec des alcools à plus longue chaîne qu'avec le méthanol (chaîne plus courte). Elle est également capable de réagir avec des alcools aromatiques tels que le l'alcool benzylique (Klinman, 1976).

I.4.3 L'ALCOOL DESHYDROGENASE DE FOIE DE CHEVAL:

I.4.3.1 L'apoenzyme

L'alcool déshydrogénase de foie de cheval (HADH) à été séquencée en 1970 par Jörnvall. Elle est dimérique, sa MM est de 80 Kd et elle contient 374 résidus par monomère (Fasman 1989: p163). Le dimère peut être circonscrit dans un ellipsoïde de 45Å *60Å*110Å. Le détail de sa structure secondaire est donnée à la figure (27) (Jörnvall, 1976:p40). La structure cristallographique du complexe apoenzyme/analogue du coenzyme, à été déterminée à une résolution de 2,4 Å en 1976 (Eklund, 76). L'analogue du coenzyme est de l'ADP ribose qui est en fait du NADH dont le cycle nicotinamide est absent (fig 28) . Cette structure cristallographique révèle que chaque monomère est constitué de deux domaines distincts, l'un fixant le coenzyme et l'autre responsable de la catalyse (fig 29).

a)Le domaine catalytique:

Le domaine catalytique est constitué des résidus 1-175 et 319-374 . Il contient les deux ions zinc. L'un se situe au niveau du site catalytique proprement dit fig . Il est coordonné à une histidine (His 67) et à deux cystéines (Cys 46 et Cys 174). Le quatrième ligand est une molécule d'eau dans l'apoenzyme et est remplacée par le substrat dans l'holoenzyme . Le second Zinc est également tétracoordonné mais par quatre cystéines (Cys 97,100,103,111). Il se situe dans un loop à la surface de la protéine. Son rôle est mal défini.

Le site actif est majoritairement hydrophobe avec, en particulier, la Phe 93 dont la chaîne latérale fait protrusion dans la poche de fixation du substrat. On y trouve cependant deux résidus hydrophiles, la Ser 48 et l'His 51 qui sont importants pour la catalyse comme nous le verrons par la suite.

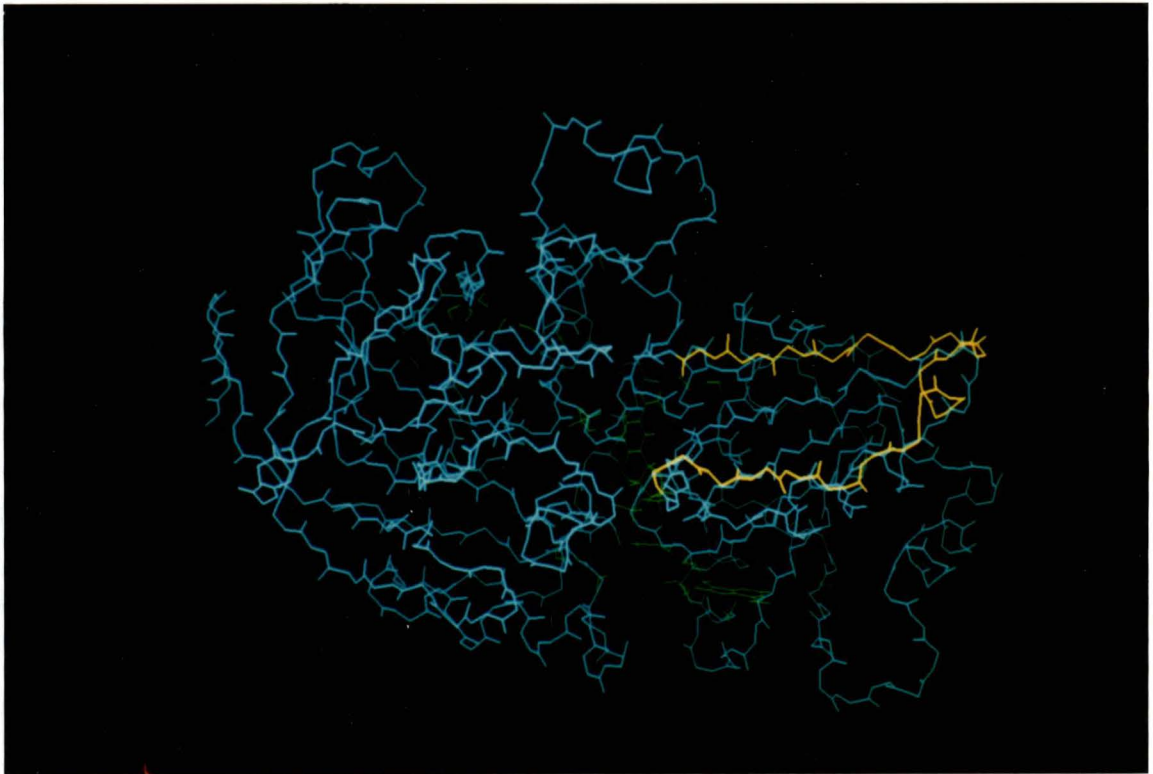


Figure 30

Zone interfaciale entre les monomères de l'alcool déshydrogénase de foie de cheval. Dans un but de clarté, seule la chaîne principale de la protéine est représentée. Le peptide établissant le contact est coloré en jaune. Cette structure secondaire forme un plan β quand la protéine est sous forme dimérique (ce plan β apparaît au centre du dimère à la figure 31). Le substrat, le coenzyme et les molécules d'eau constitutives apparaissent en vert.

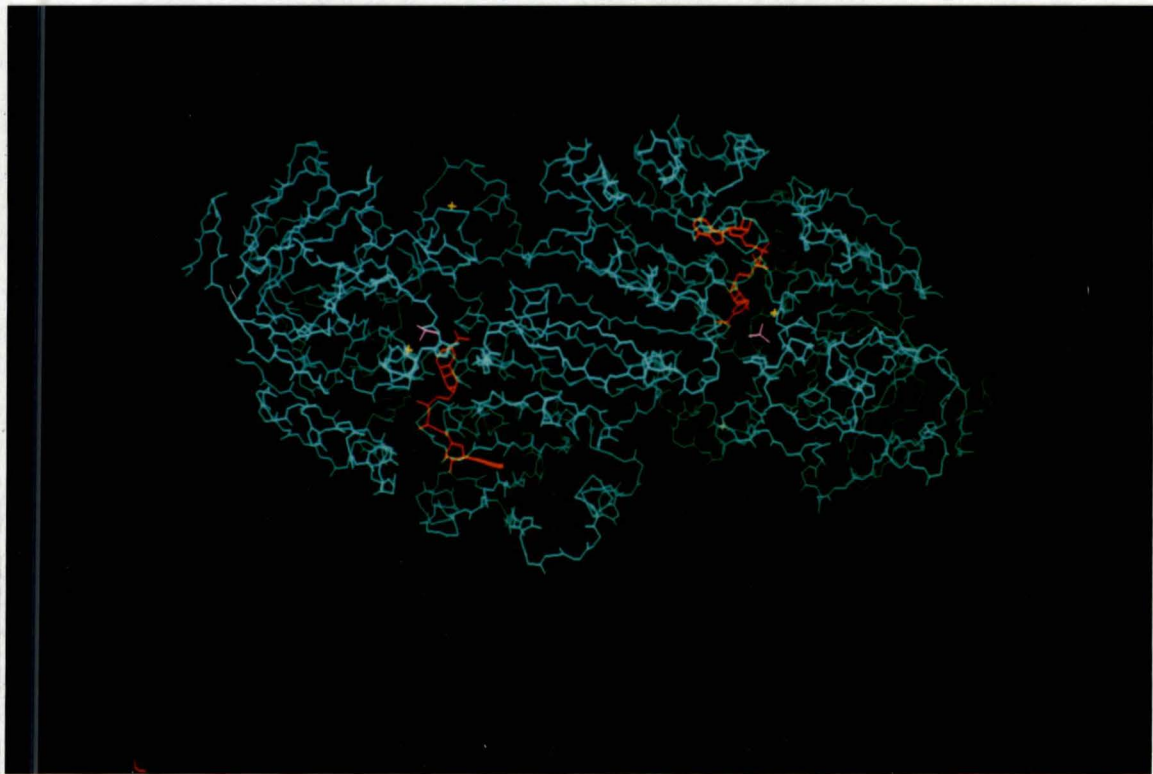


Figure 31

Structure cristallographique de la chaîne principale de l'alcool déshydrogénase de foie de cheval. Le coenzyme est coloré en rouge, le diméthyl sulfoxyde (DMSO, analogue du substrat) est coloré en violet et les atomes de zinc en jaune. Les molécules d'eau ne sont pas représentées dans cette figure.

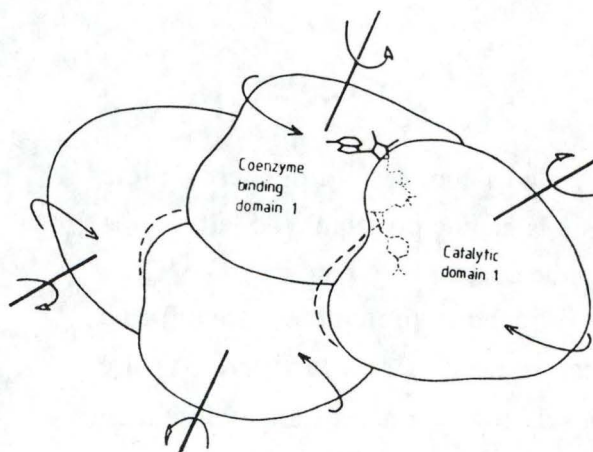


Figure 32

Représentation schématique des changements conformationnels induits par la fixation du coenzyme dans l'alcool déshydrogénase de foie de cheval dimérique.

b) Le domaine de fixation du coenzyme:

L'ADP-ribose se fixe dans un sillon parcourant la surface du domaine et qui plonge dans la poche hydrophobe séparant les deux domaines. Le ribose terminal se situe à environ 6Å de l'ion Zn^{++} .

Ce domaine est également responsable de l'interaction entre monomères. L'interface est constituée des résidus 299 à 317. La structure secondaire de l'interface est montrée à la figure (30).

I.4.3.2 L'holoenzyme:

L'holoenzyme a été cristallisée sous forme de complexe ternaire avec du NAD^+ et du diméthyl sulfoxyde (DMSO), qui est un inhibiteur compétitif de l'enzyme (Eklund, 1981). La structure a été déterminée à 2,9 Å de résolution (fig 31).

a) Changements induits par la fixation du coenzyme:

La fixation du coenzyme entraîne une rotation de $1,5^\circ$ des domaines autour des axes perpendiculaires aux plans b (fig 32). Ceci a pour effet de "refermer" le sillon existant entre ces domaines (Eklund, 1984). De même, le loop comprenant les résidus 294 à 297 subit une légère modification pour permettre aux domaines de se rapprocher.

Au niveau du site actif, la fixation du coenzyme entraîne des modifications de la distribution électronique du zinc. Il serait, dès lors, plus apte à fixer et à activer le substrat (Eklund, 1981).

Des études cristallographiques ont été réalisées avec des analogues du NADH ainsi que sur des enzymes débarrassées de leur zinc catalytique. Il semble que seul le cycle nicotinamide soit important pour induire ces changements conformationnels.

b) Etude de la fixation du coenzyme :

La distance entre l'atome C6 du cycle adénine et l'atome C2 du cycle nicotinamide dans le site actif de l'enzyme est de 15 Å.

La figure (33) donne un aperçu des ponts hydrogène formés entre les résidus du site actif contenant le coenzyme ou un analogue (ADP-ribose). Les molécules d'eau constitutives sont également représentées (Eklund, 1984). Seul le groupe carboxamide du coenzyme participe à des ponts hydrogène avec les résidus du site actif. La présence d'une charge positive sur le C4 du NAD^+ doit être compensée dans le site actif. Le complexe Zn^{++} /ligand protéique (deux Cys et une His) ne possède pas de charge négative globale. Il est donc possible que le NAD^+ influence le pK du substrat ce qui induit la polarisation du groupe carbonyle.

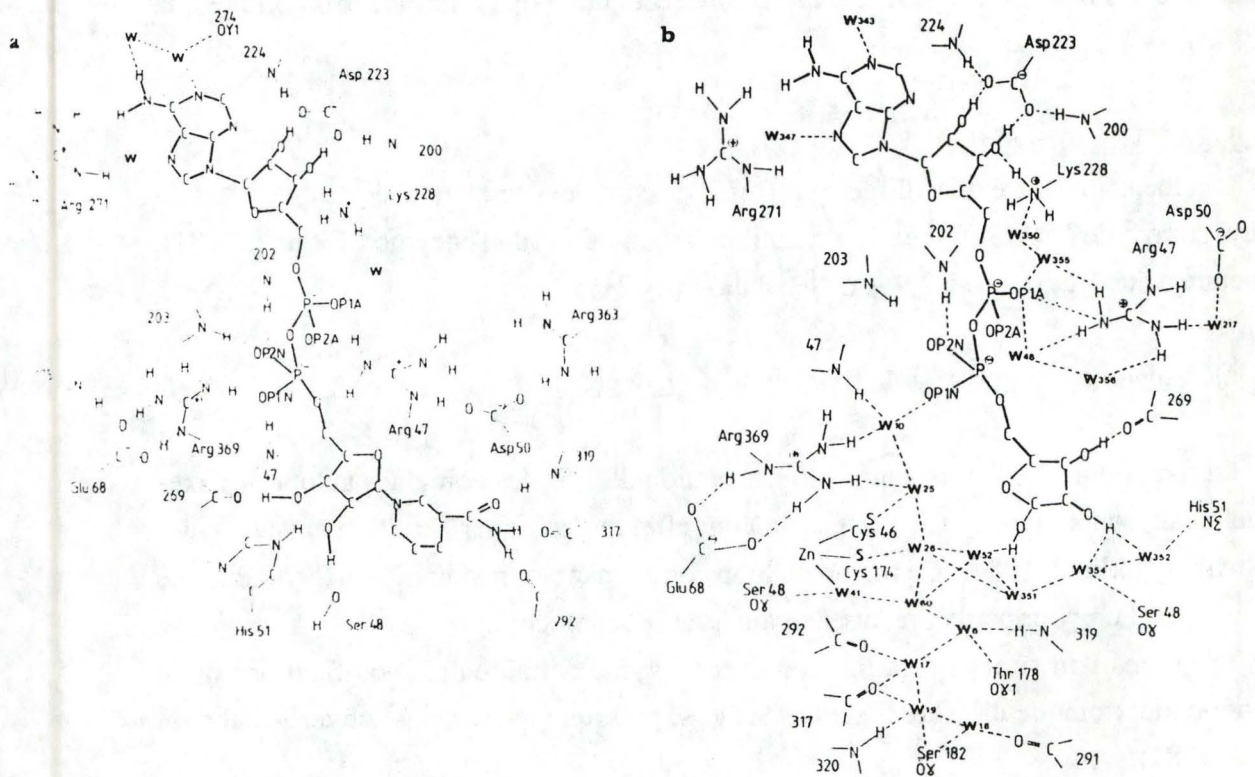


Figure 33

Illustration du site actif de l'alcool déshydrogénase de foie de cheval contenant le NAD⁺ (A), l'ADP-ribose (B) . Les ponts hydrogènes sont représentés en pointillés, et les molécules d'eau par des "w" (Eklund, 1984).

Scheme I

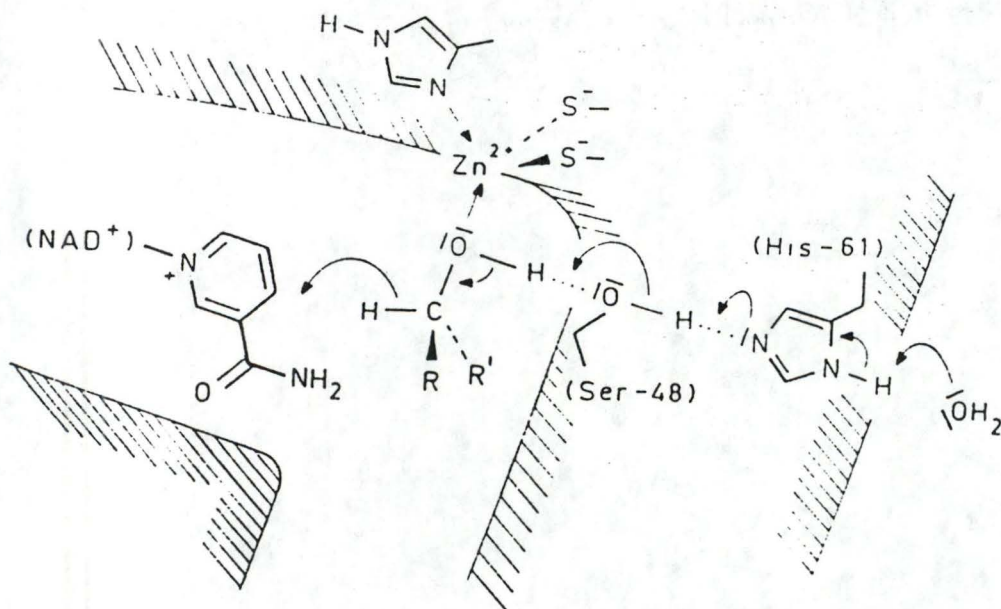
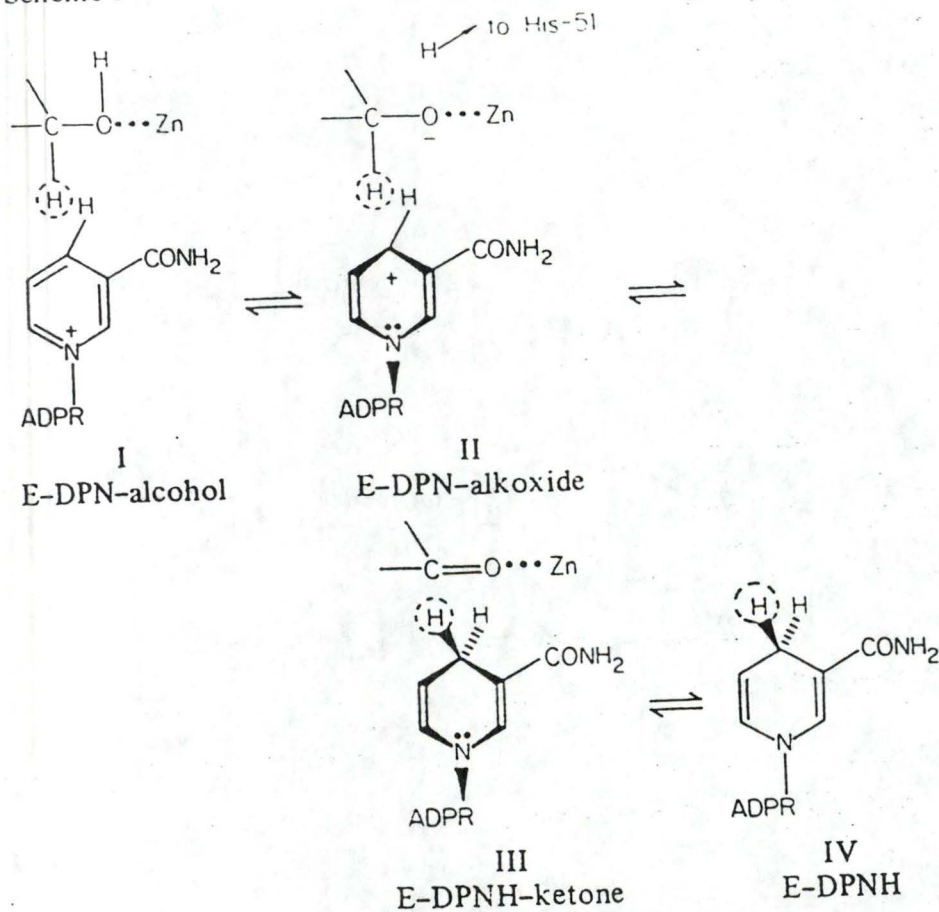


Figure 34

Au dessus: Illustration des interactions entre les différents résidus participant au mécanisme catalytique dans l'alcool déshydrogénase de foie de cheval (Kaptein, 1989)

En dessous: Mécanisme réactionnel de l'alcool déshydrogénase de foie de cheval proposé par Cook (1981).

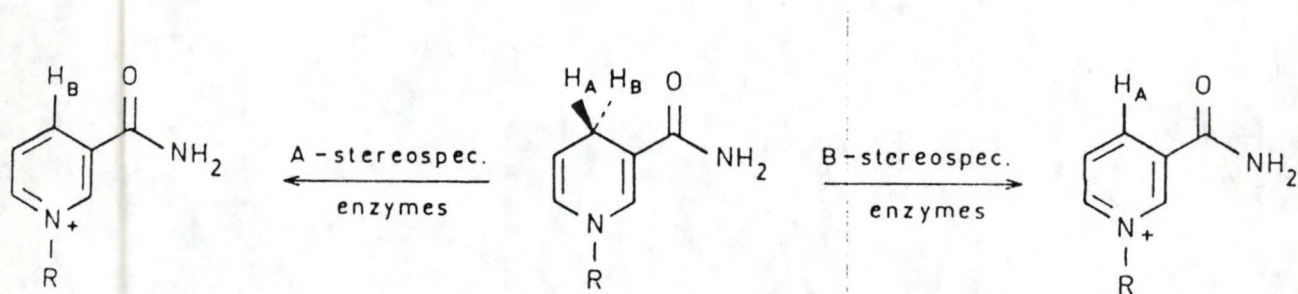


Figure 35

Description schématique de la stéréospécificité des enzymes vis à vis du NAD⁺ (Kaptein, 1989).

c) Fixation du substrat:

Des études en RMN avaient suggéré que l'éthanol était relié au Zn^{++} via une molécule d'eau (Branden *et al*, 1980: p76). L'étude cristallographique de l'enzyme réalisée par Schneider *et al* (1983) a clairement montré la coordination directe du substrat au cofacteur.

d) Mécanisme réactionnel

Branden *et al.* (1980: p71-76) ont proposé une fixation ordonnée du coenzyme puis du substrat (ordered BiBi). Les changements conformationnels qui résultent de la fixation du coenzyme constitueraient l'étape limitante de la réduction de l'acétaldéhyde. Le mécanisme est illustré à la figure (34).

- Fixation du NADH puis de l'acétaldéhyde ce qui entraîne les modifications structurelles et chasse l'eau du site actif.
- La présence de l'ion Zn^{++} polarise le groupe carbonyle et donne un alkoxyde.
- Transfert direct de l'hydrure depuis le NADH vers le $C^{\delta+}$ de l'alkoxyde. Les molécules d'eau du milieu pourraient interférer avec le transfert de l'ion hydrure (Schellenberg, 1977); l'hydrophobicité du site favorise donc la réaction.
- Transfert d'un proton sur l'oxygène de l'alkoxyde via le réseau de ponts hydrogène existant entre la Ser 48, l'His 51 et le ribose du NADH.
- Libération du NAD^+ ; l'enzyme reprend sa configuration initiale.
- Libération de l'éthanol.

- Le système capte un proton dans le milieu.

e) Stéréospécificité de la réaction:

La poche dans laquelle viennent se loger le coenzyme et le substrat est relativement étroite, ce qui peut expliquer la stéréospécificité de leur fixation.

Le coenzyme est dit A-spécifique en rapport avec la face du cycle nicotinamide qui est présentée au substrat (fig 35). La face B est, elle, présentée au Zn^{++} . Si l'on essaye de faire pivoter le cycle de 180° autour de l'axe C1-N1, le groupe carboxamide entre en collision avec les résidus 46 et 174.

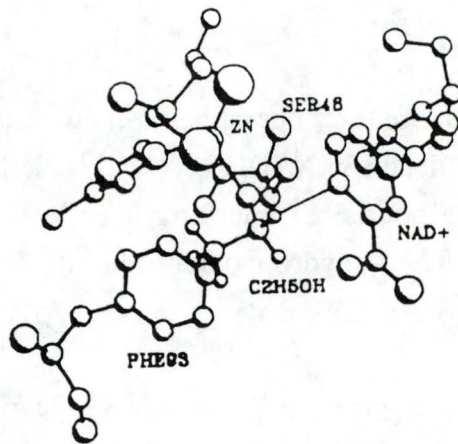
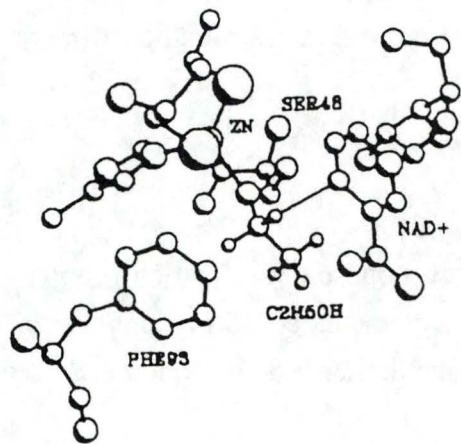


Figure 36

Orientation de l'éthanol dans le site actif de l'alcool déshydrogénase de foie de cheval (Eklund, 82).

Une seule orientation du substrat par rapport au NADH permet le transfert correct de l'hydrure (fig 36). Une rotation de l'éthanol par rapport à l'axe C=O entrainerait des contacts entre son groupe méthyl et la chaîne latérale de la Phe 93.

I.5 IDEE GENERALE DU TRAVAIL:

La structure tridimensionnelle de l'YADH I est inconnue. Elle présente cependant un certain nombre de caractéristiques communes avec l'alcool déshydrogénase de foie de cheval:

I.5.1 MECANISME CATALYTIQUE:

Toutes les deux utilisent le même coenzyme ainsi que l'ion Zn^{++} comme cofacteur. De même, on peut remarquer une similitude dans l'affinité vis à vis de certains substrats (propanol, butanol,...) même si la spécificité de l'YADH est plus restreinte que celle de l'HADH (Bille, 1987). Les mécanismes eux-mêmes sont très semblables. Il semble que l'enzyme de levure subit également des changements conformationnels lors de la fixation du coenzyme (Ganzhorn *et al*, 1987).

I.5.2 PROPRIETES ENZYMATIQUES:

Des alignements de séquences ont révélé une similarité d'environ 45%. L'identité est globalement de 25%, bien que plus élevée dans les zones correspondant au site actif de l'enzyme de cheval.

Sur base de ces alignements, des mutations ont été réalisées au niveau des résidus supposés appartenir au site actif chez la levure (Ganzhorn *et al*, 1987, 1988 ; Plapp *et al*, 1987). Il en ressort une série d'observations confirmant l'hypothèse selon laquelle la similarité de structure entre ces deux enzymes est importante.

Malgré les connaissances accumulées sur l'enzyme de levure ainsi que le perfectionnement des techniques de modélisations, aucun modèle plus rigoureux qu'une simple superposition de la séquence de YADH I sur la structure tridimensionnelle de HADH n'a été envisagé. Un tel modèle ne permet pourtant que l'évaluation des perturbations locales de la structure engendrées par la mutation d'un ou plusieurs résidus. Les modifications de charges partielles des atomes, d'encombrements stériques, ainsi que les répercussions à "longue distance" ne sont pas prises en compte mais jouent pourtant un rôle fondamental (il arrive souvent que la modification d'un résidu de surface entraîne l'inactivation totale de l'enzyme).

L'étude et par la même occasion, la modification des propriétés catalytiques de l'enzyme nécessite donc un modèle précis et dynamique de sa structure tridimensionnelle. C'est ce que nous nous proposons de réaliser dans le cadre de ce travail.

II. MATERIEL ET METHODES:

II.1 SUPPORT INFORMATIQUE:

Toutes les manipulations ont été réalisées sur un "IRIS PERSONAL COMPUTER" (SILICON GRAPHICS) disposant d'un processeur de 12,5 Méga Herz ce qui lui permet de traiter 10 millions d'instructions par seconde (MIPS). Il possède deux disques durs de 380 Mégaoctets (380 millions de caractères mémorisables) chacun et d'une mémoire centrale de 12 Mégaoctets . Son système d'exploitation est le "UNIX 2" développé par les laboratoires AT&T Bell.

II.2 BANQUES DE DONNEES:

Nous avons utilisé la "Protein Data Bank"(PDB) du Brookhaven National Laboratory- Cambridge, USA). Elle contient les coordonnées cartésiennes des atomes lourds de structures protéiques déterminées par cristallographie. Chaque structure y est répertoriée sous un nom de code.L'édition dont nous disposons date de juillet 1989 et contient les coordonnées de 434 structures macromoléculaires(protéines, virus,...).

II.3 LOGICIELS DE MODELISATION:

Nous disposons des logiciels INSIGHT II, MOLEDT et DISCOVER de chez BIOSYM (San Diego). Ces trois logiciels peuvent être utilisés séparément ou simultanément.

II.3.1 INSIGHT (VERSION II):

INSIGHT a deux fonctions principales: Il réalise des modèles graphiques tridimensionnels à partir des coordonnées cristallographiques des atomes. Ces modèles peuvent être manipulés, transformés et conservés sous forme d'images (fichier binaire incluant les couleurs, le mode de représentation,...) ou sous forme de fichiers utilisables par DISCOVER. Il joue donc un rôle d'interface graphique entre les banques de données (PDB,...), l'utilisateur et DISCOVER. Il permet également la construction *de novo* de molécules à partir d'une librairie contenant les fichiers complets des différents aa, de molécules organiques, de radicaux organiques diverses et même de certains atomes.

BIOSYM Molecular Modeling Tools: System Architecture

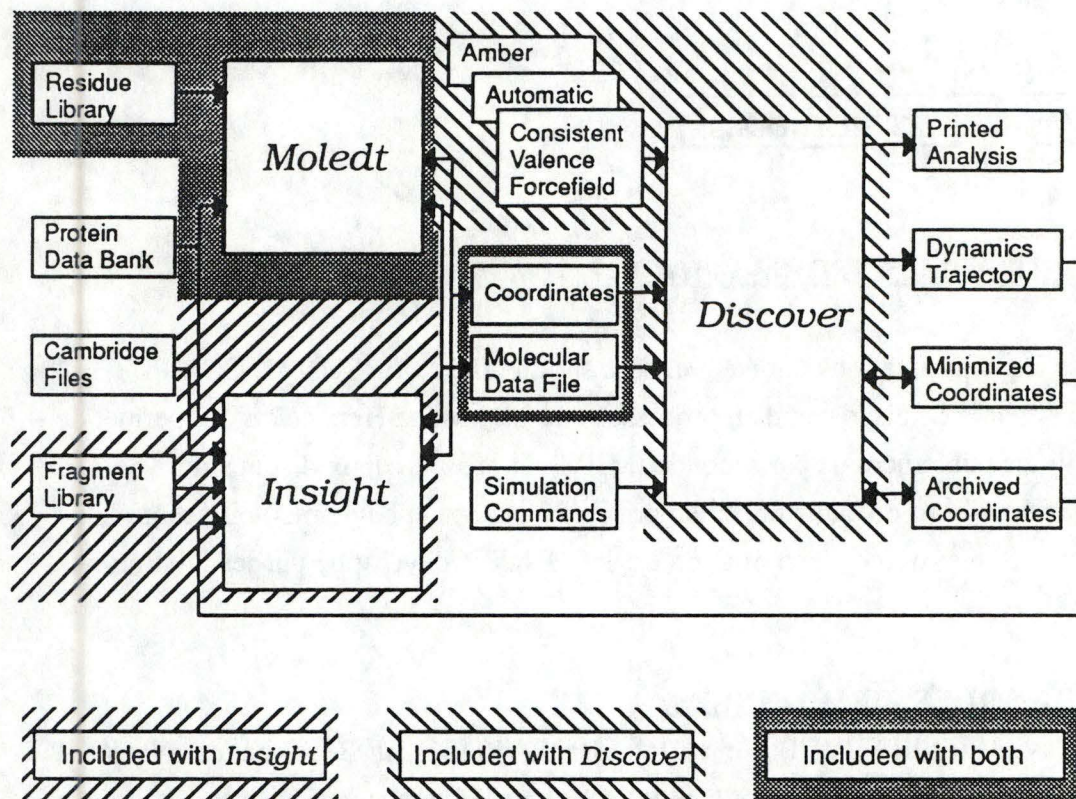


Figure 37

Description simplifiée des interactions possibles entre les différents modules des logiciels de modélisation moléculaire INSIGHT et DISCOVER (Discover, 89).

$$E_{pot} = \sum_b D_b [1 - e^{-a(b-b_0)^2}] + \frac{1}{2} \sum_{\theta} H_{\theta} (\theta - \theta_0)^2 + \frac{1}{2} \sum_{\phi} H_{\phi} [1 + s \cos(n\phi)] + \frac{1}{2} \sum_x H_x \chi^2 +$$

(1)
(2)
(3)
(4)

$$\sum_b \sum_{b'} F_{bb'} (b - b_0) (b' - b_0') + \sum_{\theta} \sum_{\theta'} F_{\theta\theta'} (\theta - \theta_0) (\theta' - \theta_0') +$$

(5)
(6)

$$\sum_b \sum_{\theta} F_{b\theta} (b - b_0) (\theta - \theta_0) + \sum_{\phi} F_{\phi\theta\theta'} \cos \phi (\theta - \theta_0) (\theta' - \theta_0') + \sum_x \sum_{x'} F_{xx'} \chi \chi' +$$

(7)
(8)
(9)

$$\sum \epsilon [(r^*/r)^{12} - 2(r^*/r)^6] + \sum q_i q_j / \epsilon r_{ij} + \sum \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right]$$

(10)
(11)
(12)

Figure 38

Equation d'énergie potentielle utilisées dans le CVFF (Discover, 1989). Les différents termes de l'équation sont illustrés schématiquement à la figure 39.

II.3.2 MOLEDT:

MOLEDT est un éditeur des fichiers créés par INSIGHT ou DISCOVER c-à-d qu'il permet la manipulation et la modification de ces fichiers. Les fonctions de MOLEDT tendent à être de plus en plus reprises dans le logiciel INSIGHT.

II.3.3 DISCOVER (1989)

DISCOVER est un logiciel de calcul. Il réalise les minimisations d'énergie, la dynamique moléculaire ainsi que des calculs d'énergie libre. Rappelons que ces calculs sont réalisés sur base d'un Force Field(F.F.). Plusieurs F.F. sont supportés par DISCOVER; le "Consistent Valence F.F" (CVFF; Hagler et al, 1979) et le "Assisted Model Building with Energy Refinement" (AMBER;Weiner et al, 1984).

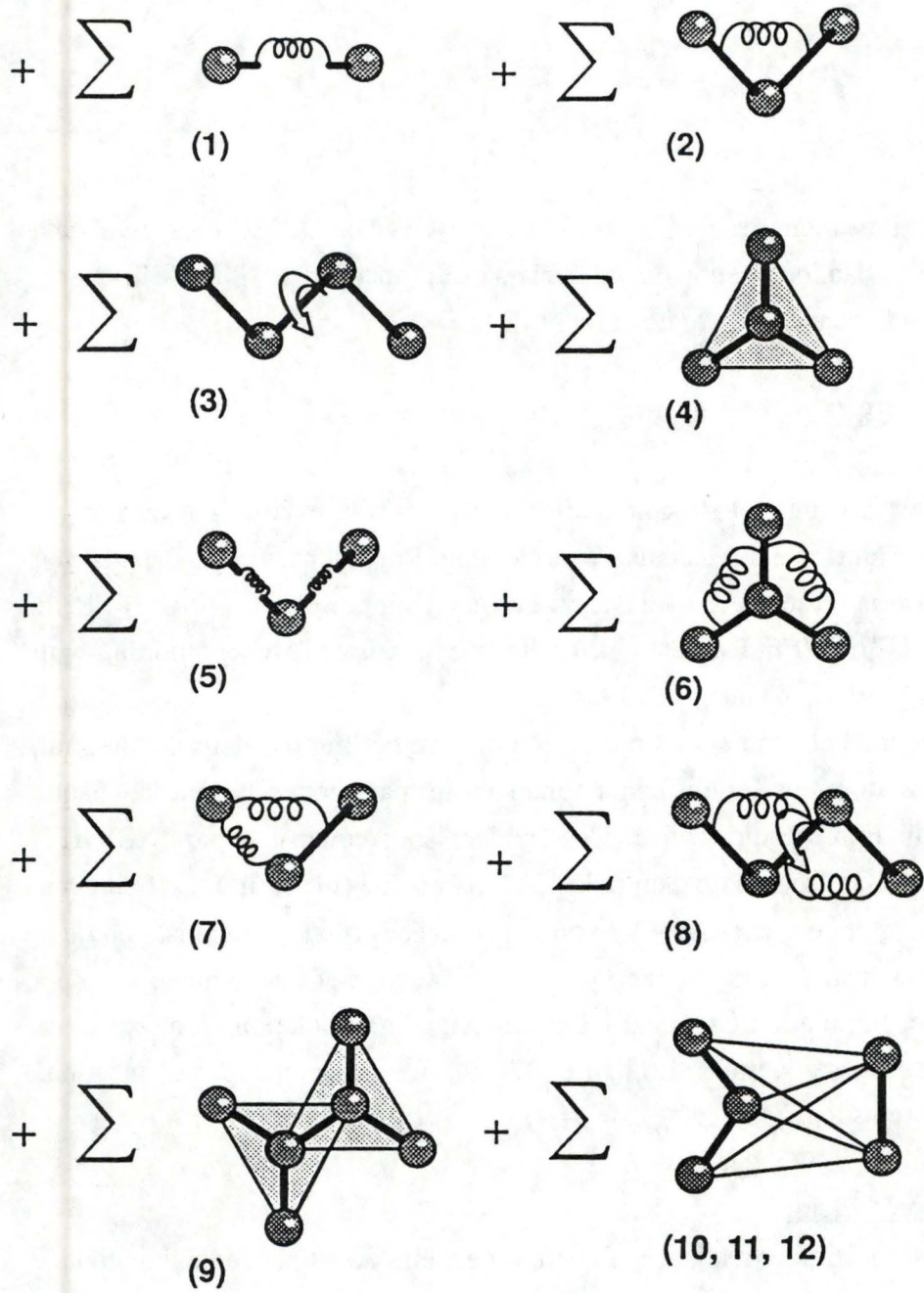
Pour être reconnue et traitée, une molécule doit être définie par deux fichiers: un fichier de coordonnées atomiques et un fichier contenant les paramètres assignés à chaque atome en fonction du Force Field utilisé. Ces fichiers sont construits par INSIGHT. DISCOVER réalise les calculs sur base d'un fichier de commande (Input file). Ce fichier est écrit par le manipulateur et contient toutes les instructions nécessaires à la réalisation de la simulation désirée. Ces instructions sont écrites dans un langage et une syntaxe propre à DISCOVER (Discover Simulation Language). Les résultats des simulations sont écrits par DISCOVER dans des fichiers "output". La figure (37) résume l'ensemble des connections existant entre ces différents logiciels.

II.3.3.1 Le Force Field:

Le CVFF a été choisi pour réaliser ce travail. Il peut être décrit par l'équation donnée à la figure (38)

La figure (39) illustre les interactions décrites par chacun des termes. Les termes 1 à 4 représentent les fonctions énergie potentielle des interactions covalentes.

Il faut souligner que ce F.F utilise une fonction dite "morse" pour les interactions covalentes (fig 39). Bien que plus représentative de la réalité, cette fonction peut tolérer ou générer des longueurs de liaisons covalentes irréalistes si la structure de la molécule est mal définie en début de minimisation. On utilise donc cette fonction en fin de minimisation alors qu'une fonction de type "harmonique" (fig 39) est préférable en début de minimisation car elle permet d'ajuster les liaisons covalentes à leur longueur optimale. Les termes 5 à 9 sont des "cross terms" (voir introduction: § I.3.4.1). Enfin, les termes 10 et 11 représentent les interactions non covalentes (voir § I.3.4.1). Ces dernières regroupent les interactions de Van der Waals (10) et les interactions électrostatiques (11). Les interactions par ponts hydrogène sont reprises par le dixième terme de la fonction énergie potentielle. Néanmoins, une fonction décrivant explicitement l'énergie de ces interactions est disponible (terme 12).



Panel A

Panel B

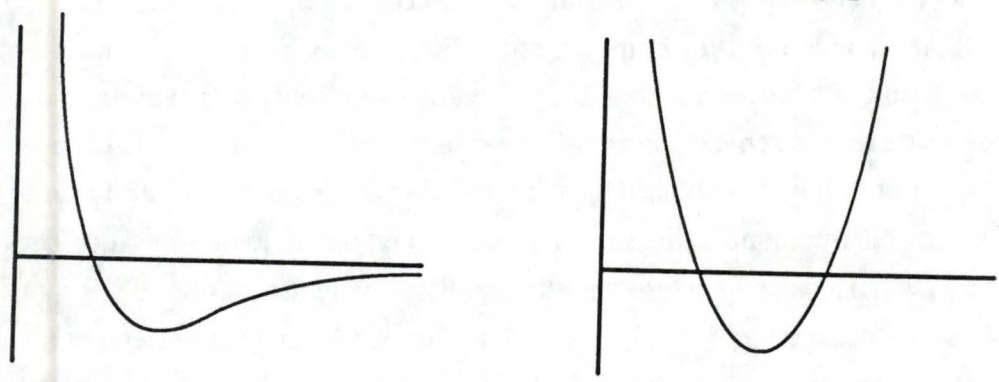


Figure 39

En haut: Représentation schématique des termes de l'équation d'énergie potentielle du CVFF (Discover, 1989)

En bas: Différence entre l'allure de la fonction d'énergie potentielle morse (A) et quadratique (B) pour la liaison C-H (Discover, 1989)

Hagler *et al* (1979) ont montré que l'utilisation d'une fonction particulière pour ce type d'interaction n'améliore pas l'évaluation de l'énergie d'une molécule. Une description détaillée du CVFF est donné dans le manuel d'utilisation de DISCOVER 5.1 (Discover, 1989, p 5-5 à 5-14).

II.4 ALIGNEMENTS DE SEQUENCES:

L'alignement de séquence utilisé est celui proposé par Jörnvall en 1978. Des alignements de séquences peuvent être réalisés grâce au logiciel développé par le "Genetics Computer Group" (GCG; J. Devereux , 1987)

Le module "GAP" réalise des alignement de deux séquences par l'algorithme de Needleman et Wunsch (1970). La similarité entre deux résidus est évaluée sur base de la matrice de Dayhoff (1972) (voir § I.3.5.1). Les insertions/délétions ainsi que leurs longueurs peuvent être pondérées. Un autre module:"BESTFIT" utilise la méthode d'alignement proposée par Smith et Waterman (1981) qui est une adaptation de la méthode précédente utilisant une valeur de 1 pour les identités et une valeur de -0,9 pour les aa différents. Les délétions/insertions ainsi que leurs longueurs sont également pondérables.

II.5 MODELISATION DE L'APOENZYME DE LEVURE (YADH)

Toutes les opérations sont réalisées en INSIGHT. Elles consistent à:

- Transférer le fichier de la banque PDB contenant les coordonnées de l'holoenzyme (Code= 6ADH) à l'écran .Une image de l'holoenzyme dimérique cristallisée ne contenant pas d'atomes hydrogène est ainsi obtenue.
- Sélectionner un des monomère puisque les deux sont identiques.
- Placer les hydrogènes
- Remplacer les résidus de l'enzyme de cheval par ceux de levure sur base de l'alignement choisi. La conformation de la chaîne principale de l'ancien résidu est conservée pour le nouveau résidu (Summers & Karplus, 1989). La conformation de la chaîne latérale du résidu précédent n'est appliquée à la nouvelle chaîne latérale que si il n'y a pas de contacts avec d'autres résidus environnants. Sinon, certains angles de torsion sont modifiés de manière à supprimer ces contacts.
- Les délétions consistent simplement à enlever les résidus concernés puis à joindre les extrémités libres par un long pont covalent.
- il n'existe pas de démarche rigoureuse pour les insertions. D'une manière générale, l'aa à insérer provient de la librairie de résidus. Le lien peptidique entre les résidus bordant la délétion est supprimé et le nouveau résidu est intégré à la chaîne en modifiant la configuration de celle-ci le moins possible fig(40).

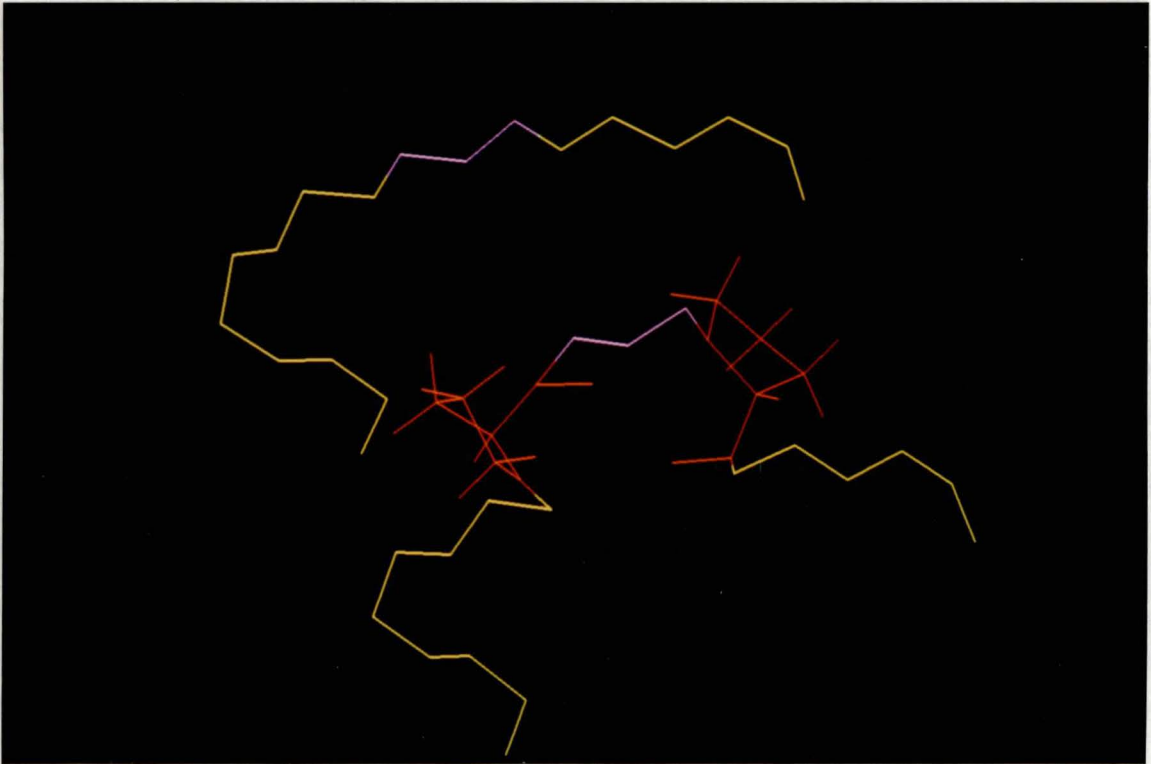


Figure 40

Comparaison entre la chaîne peptidique avant insertion (au dessus) et après insertion (en dessous). Les résidus insérés sont colorés en rouge.

- Vérifier la séquence mutée ainsi que des résidus N et C terminaux. Ces derniers doivent être reconnus comme tels par INSIGHT et DISCOVER.
- Assigner un potentiel et une charge partielle à chaque atome de l'enzyme .

II.6 MODELISATION DES MOLECULES ANNEXES:

Pour que la minimisation donne des résultats cohérents, il est nécessaire d'intégrer le substrat, le coenzyme, le cofacteur et les molécules d'eau constitutive dans l'apoenzyme. De cette manière, on évite les déformations artificielles du site actif lors de la minimisation.

II.6.1 LE COENZYME:

A partir de l'image du NAD⁺ disponible dans la structure PDB, il faut construire un coenzyme dont les atomes possèdent les caractéristiques conformationnelles et physicochimiques les plus proches possible de la réalité. Le modèle du coenzyme utilisé dans ce travail à été utilisé réalisé comme suit:

- Placer les hydrogènes
- Placer les liaisons doubles partielles, simples et doubles entre les atomes du coenzyme. Les caractéristiques des liaisons du cycle nicotinamide nous ont été cordialement fournies par Mr L. Hevesi.
- Attribuer un potentiel à chaque atome ainsi qu'une charge partielle.
- Vérifier la charge globale du NAD⁺ qui doit être -1. Si ce n'est pas le cas, une fonction d'INSIGHT permet d'attribuer une charge partielle donnée aux atomes d'une molécule.

II.6.2 LE SUBSTRAT:

L'éthanol n'existant pas en temps que tel dans la librairie de résidus, il est élaboré de la manière suivante à partir du glycol :

- Amener la molécule glycol (HOCH₂-CH₂OH) à l'écran (à partir de la librairie de résidus).
- Remplacer un groupe hydroxyl par un H.
- Assigner les potentiels atomiques et les charges partielles aux atomes de l'éthanol.
- Vérifier la charge globale qui doit être nulle.

II.6.3 L'EAU CONSTITUTIVE:

- Le fichier PDB de l'HADH cristallisée avec de L'ADP-ribose (Code= 5ADH) comme coenzyme contient également 300 molécules d'eau dont 40 font partie du site actif en présence de l'ADP-ribose. Les molécules d'eau constitutives de l'enzyme sont sélectionnées de la manière suivante:
- Vérifier par superposition des chaînes principales que 5ADH n'est pas trop différente de 6ADH. D'une manière générale, la superposition de deux molécules est réalisée sur base d'un nombre identique d'atomes dans les deux structures. Ces structures sont superposées de façon à minimiser la somme des carrés des écarts des distances entre les atomes désignés comme correspondant.
- Les molécules d'eau extérieures sont supprimées.
- Sur base de la publication de Eklund (1984), on ne conserve que les molécules d'eau du site contenant le NAD^+ (21 dans ce modèle).

II.7 CONSTRUCTION DE L'HOLOENZYME:

Les positions des atomes du coenzyme, du substrat, et des molécules d'eau étant connues dans la structure PDB, on peut facilement remplacer les images existantes par les molécules construites à partir de INSIGHT selon les étapes décrites ci-dessous:

- Superposer les atomes équivalents de la structure PDB et de la molécule corrigée.
- Supprimer l'image PDB.
- Associer la nouvelle structure à l'apoenzyme.
- Contrôler les potentiels et les charges partielles pour toutes les molécules.
- Construire les fichiers nécessaires à DISCOVER pour la réalisation de la minimisation.

Rem: L'ION Zn^{++}

Le modèle proposé n'inclut que le Zn^{++} catalytique selon l'hypothèse de Kaptein (1989).

Beaucoup d'atomes ne sont pas reconnus par DISCOVER, les paramètres n'étant pas intégrés dans le Force Field. Tel est le cas pour l'ion Zn^{++} . Il exerce pourtant une force d'attraction sur les atomes qui lui sont coordonnés. Sa suppression entraînerait des modifications de la structure du site actif lors de la minimisation.

La solution choisie a été de fixer les atomes coordonnés au Zn^{++} dans leur position de départ durant toute la minimisation de manière à conserver la configuration observée dans la structure connue.

II.8 LA MINIMISATION D'ENERGIE:

Plusieurs méthodes de minimisation sont disponibles sur DISCOVER. Le choix de la méthode ainsi que ses options d'application sont modulables en fonction des caractéristiques du modèle. Rappelons que l'ensemble des instructions nécessaires à la réalisation et au contrôle du processus de minimisation sont regroupées dans un fichier de commandes (fichier.inp).

Plusieurs paramètres sont à contrôler lors d'une minimisation:

1) Le nombre d'itérations:

Une itération de la minimisation d'énergie consiste à calculer le gradient négatif de l'énergie potentielle de chaque atome de la protéine par rapport à sa position relative dans la protéine et à déplacer ces atomes le long de ce gradient négatif, jusqu'à ce que l'énergie globale de la protéine passe par un minimum. Il est admis qu'un modèle demande 6 fois plus d'itérations qu'il y a d'atomes dans la protéine pour atteindre un niveau énergétique correspondant à une conformation stable.

Il est important de contrôler régulièrement l'état d'avancement d'une minimisation pour permettre d'éventuelles corrections du modèle. La minimisation se décompose donc en étapes au bout desquelles on considère les résultats obtenus.

Enfin, il faut savoir que le temps calcul nécessaire à une minimisation est en général très élevé. Pour éviter de perdre des heures de travail suite à un arrêt intempestif de la machine, il convient de stocker périodiquement le modèle au stade où il en est. Cette fonction permet également de suivre l'évolution des différents paramètres au cours du processus.

2) la valeur absolue de la dérivée maximum de l'énergie

Cette valeur est le baromètre de la minimisation. Elle renseigne sur les tensions existant au sein de la molécule. Plus la dérivée de l'énergie potentielle d'un atome est élevée, plus la force à laquelle il est soumis est grande et plus il déstabilise la structure de la molécule.

On considère en général que la minimisation est achevée quand la dérivée maximale est inférieure ou égale à 0.01 en valeur absolue. Une liste des atomes associés aux n plus grandes dérivées peut révéler des zones de tension au sein du modèle.

3) L' énergie potentielle du modèle:

Les énergies potentielles associées à chaque type d'interactions (covalente, torsion, VDW,...) permettent de mieux cerner les interactions responsables d'une dérivée élevée.

4) Vitesse d'exécution de la minimisation:

Le temps calcul nécessaire à l'ordinateur pour mener à bien une série de minimisations se révèle très rapidement comme le principal facteur limitant de la conception du modèle. A titre d'exemple, une minimisation de L'YADH en conjugate gradient dure environ 24 heures par 1000 itérations. Aussi, il est important de limiter la taille des modèles tout en leur conservant un maximum de précision. Pour cela, une série d'options sont proposées sur DISCOVER; nous nous contenterons d'examiner celles utilisées dans la suite de ce travail.

Les minimisations d'énergie sont réalisées en général sur des molécules sous vide ce qui permet de diminuer la taille du système de quelques milliers d'atomes.

Or, le solvant exerce une influence importante sur la conformation de la molécule. Des molécules d'eau placées entre des atomes chargés atténuent leurs interactions électrostatiques. Dans le vide, les interactions électrostatiques n'étant pas atténuées, des atomes même très éloignés s'influencent selon leur charge propre. Privés de l'effet d'écran du solvant, des atomes de la surface de la protéine risquent d'influencer d'autres atomes qui leur seraient normalement indifférents. Pour remédier à ce problème, les énergies d'interactions électrostatique sont posées comme étant inversement proportionnelle au carré de la distance séparant deux atomes. Dès lors, les interactions à faible distance sont toujours présentes alors qu'à plus longue distance, elles sont atténuées.

Le nombre d'interactions à évaluer est largement dominé par les interactions non covalentes. Nous avons vu qu'au delà d'une certaine distance, l'énergie de ces interactions devient rapidement négligeable. On peut donc considérer uniquement les interactions non covalentes dont la distance interatomique n'excède pas une valeur donnée (*Cutoff distance*). Au-delà de cette valeur, l'énergie est considérée comme nulle. Cela a comme effet de diminuer considérablement le nombre d'équations à traiter.

L'utilisation des "Cross Terms" augmente également considérablement le nombre d'équations à traiter. D'autre part, ils ne sont utiles que lors d'une évaluation précise de l'énergie potentielle. On utilise donc les "Cross Terms" qu'en fin de minimisation, quand on désire un modèle plus précis.

```

! INPUT FILE FOR DISCOVER GENERATED BY INSIGHT
!
!
begin simulation
* add-automatic bond torsion valence out-of-plane
!
!
cutoff = 12.000000
cutdis = 10.000000
swtdis = 1.5
nrneib = 19
!
!
fixed atom list generation clear
!
fix residues
* molecule 1
* molecule 2
* molecule 3 to 23
!
fixed atom list generation
* add main atoms
* molecule 24
!
!
set dielectric = 1.0*r
loop=0
100 minimize using steepest descent
* no cross terms
* no morse
* for 200 iterations
* until the maximum derivative is less than 0.010 kcal/A
archive as file number 1
loop=loop+1
if loop.lt.3 then 100
!
!
print energy/residue
print largest 10 derivatives
!
end

```

Figure 41

Détail d'un fichier de commande de Discover.

Une liste contenant l'ensemble des atomes dont la distance de séparation est inférieure à cette limite sert de base au calcul de l'énergie d'interaction non covalente. Chaque déplacement d'atome est susceptible de modifier cette liste. Sa réévaluation requiert un temps non négligeable. Cette opération n'est donc réitérée que toutes les n itérations ($n = nrneib$ est déterminé par le manipulateur).

Il est également possible de ne minimiser qu'une partie de la molécule (Par exemple uniquement le site actif). Pour cela, une fonction (*Fix residue* ou *Fixed atom list generation*) permet de fixer les résidus ou les atomes se situant en dehors de la zone concernée. Bien qu'exerçant toujours une influence sur les atomes restés libres, ni leur énergie potentielle, ni la dérivée de leur énergie potentielle n'est calculée.

5) Détail d'un fichier de commande(Discover Input file):

Un fichier de commande est illustré à la figure(41). La syntaxe et les instructions sont celles que demande DISCOVER pour réaliser la simulation suivante.

Le système contient 24 molécules:

L'apoenzyme (mol. 24), le NAD⁺ (mol.1), l'éthanol (mol.2) et les molécules d'eau constitutive (mol.3-23).

La minimisation sera réalisée avec les paramètres suivants:

- Méthode de "steepest descent" pour 600 itérations.
- "cutoff distance" de 12 Å et mise à jour de la liste des interactions non covalentes toutes les 19 itérations.- Arrêt de la minimisation lorsque la dérivée maximale est de 0.01 en valeur absolue.
- Les résidus des molécules 1 à 23 sont fixés ainsi que la chaîne principale de la molécule 24.
- Pas d'utilisation des "cross terms".
- La fonction harmonique (quadratique) est utilisée pour l'évaluation de l'énergie des liaisons covalentes.
- Stockage du modèle en cours de minimisation toutes les 200 itérations.
- Impression des énergies pour chaque résidu et impression des 10 plus grandes dérivées.

II.9 ETAPES DE LA MINIMISATION:

Toute la minimisation est réalisée avec une distance de "cutoff" de 12Å et la liste des atomes impliqués dans les interactions non covalentes est remise à jour toutes les 19 itérations. L'élaboration du modèle d'YADH à été effectuée selon les étapes décrites ci-dessous:

- Relaxation de l'apoenzyme de levure par la méthode de Steepest descent. Les résidus coordonnés au zinc sont fixés dans leur position initiale.
- Placer la molécule à pH =7.
 - Supprimer les hydrogènes de l'apoenzyme en INSIGHT
 - Sauver cette version de la molécule sous une forme lisible par DISCOVER
 - Reprendre cette molécule en INSIGHT.
 - Placer les hydrogènes en considérant la molécule comme étant à pH=7 (une fonction de INSIGHT permet de réaliser cette opération automatiquement).
 - Assigner les charges et les potentiels atomiques corrects.
 - Sauver l'holoenzyme sous une forme reconnaissable par DISCOVER.
- Minimisation des liaisons covalentes d'énergie trop élevée (dues aux délétions et aux insertions). Une fonction (*scale*) permet multiplier l'énergie d'un type d'interaction particulier par un facteur donné. La dérivée de l'énergie de ce type d'interaction est également multipliée par ce facteur. Lors de la minimisation, la contribution de cette dérivée au gradient de l'énergie potentielle globale de chaque atome sera donc accrue et le déplacement atomique sera orienté de manière à favoriser la diminution de l'énergie de ce type d'interaction.

7000 itérations ont été nécessaires pour la résorption des ponts covalents introduits par les insertions et les délétions.

- Vérification des angles de torsion

Une liste d'angles présentant une mauvaise conformation est établie sur base d'une comparaison entre les angles de torsions de HADH et ceux du modèle avant minimisation. Cette liste est établie de la manière suivante:

- Les résidus identiques doivent conserver les mêmes angles de torsion.
- Les valeurs d'angles particulières dans l'HADH doivent être conservées chez YADH
- Les aa de HADH ayant été remplacés par des Pro doivent avoir un angle ϕ de -80° .
- Les angles de torsion des résidus de YADH correspondant à des résidus de l'HADH admettant des angles ϕ et ψ particuliers doivent être vérifiés (*ex:* Gly et Ala qui sont les seuls résidus à admettre des angles ϕ et ψ positifs).
- Les angles ω doivent être proches de 180° en valeur absolue.
- Vérifier les angles de torsions autour des délétions et des insertions.

Une fois cette liste établie, elle peut servir de base pour la vérification des angles de torsion du modèle en cours de minimisation. 7000 itérations supplémentaires ont été nécessaires pour obtenir des angles de torsions adéquats. Une fonction (*Force*) permet d'appliquer une force sur certains angles de torsion ce qui leur permet de passer certaines barrières énergétiques et d'adopter une conformation donnée.

- Minimisation du site actif seul:

Le site actif de l'holoenzyme est isolé du reste de la protéine et son énergie est minimisée et son énergie est minimisée en considérant toutes les interactions possible entre les différents atomes. Cette opération est réalisée de la manière suivante:

- Sélectionner les atomes de l'holoenzyme se situant dans un volume de 15\AA de rayon autour de l'atome OG1 de la Thr 48 (en INSIGHT).
- Sauver cette zone sous une forme utilisable par DISCOVER.
- Fixer les résidus des extrémités libres dans leurs positions initiales.
- Minimiser le site contenant coenzyme et substrat avec la méthode de Conjugate Gradient et avec les Cross Terms jusqu'à obtenir une dérivée maximale de 0,01.

- Réintégrer le site actif minimisé dans l'holoenzyme complet. La méthode la plus aisée est de remplacer les coordonnées du site actif non minimisé par celles du site minimisé dans le fichier de coordonnées de l'holoenzyme.

II.10 MINIMISATION DE L'HADH:

La structure de l'holoenzyme est celle de PDB. Les hydrogènes sont ajoutés à pH 7 et les cosubstrats (coenzyme, substrat) sont placés dans les sites adéquats selon la même procédure que celle suivie pour l'YADH. Les Zn^{++} (deux chez HADH) sont absents du modèle mais les atomes qui leurs sont coordonnés sont fixés dans leur position initiale.

La minimisation est réalisée par la méthode de steepest descent et avec la fonction quadratique évaluant l'énergie des interactions covalentes (harmonique) selon les étapes suivantes:

- Ajustement des hydrogènes de l'apoenzyme par une minimisation du modèle dont tous les atomes lourds ont été fixés, durant 300 itérations .
- Ajustement des chaînes latérales en minimisant le système dont tous les atomes ont été fixés, sauf ceux des chaînes latérales de l'apoenzyme, durant 600 itérations.
- Minimisation globale du système dont seul les atomes liés aux Zn^{++} ont été fixés, durant 800 itérations.

II.11 VALIDATION DU MODELE:

II.11.1 CALCUL DU COMPACTAGE DES RESIDUS

Le compactage des résidus de l'HADH ainsi que celui du modèle de YADH à été réalisé selon la méthode de Richards (1974) grâce à un programme réalisé par Feytmans (non publié). Le compactage du ième résidu est calculé de la manière suivante:

$$P(i) = \frac{\sum_{j=1}^N C_{i,j}}{A_i}$$

- où
- P(i) est la valeur du compactage du résidu i
 - N est le nombre de résidus dans la protéine
 - C_{i,j} est le nombre d'atome des résidus i et j (et i-j > 2 en val. absolue) se situant à moins d'une distance définie par l'utilisateur (3,6 Å dans notre cas).
 - A_i est la surface maximale accessible au solvant du résidu i

La valeur obtenue permet de se rendre compte de la proportion de la surface accessible au solvant du résidu i qui est en contact avec d'autres atomes de la protéine. Cette valeur donne donc une mesure du compactage des atomes de la protéine autour de chaque résidu.

II.11.2 CALCUL DES MOMENTS HYDROPHOBIOQUES

La méthode utilisée est quelque peu différente de celle présentée dans le §I.2.3.5. Il s'agit de calculer les vecteurs radiaux d'hydrophobicité moyens d'une fenêtre de 7 résidus se déplaçant sur la séquence polypeptidique (VRH7; Feytmans, 1988). Ce vecteur se calcule de la manière suivante.

- Calcul de la longueur, de l'orientation et de la taille du vecteur associé à chaque résidu. La taille du vecteur est donnée par la valeur absolue de l'hydrophobicité du résidu selon l'échelle consensus établie par Eisenberg (1984). La direction est donnée par la liaison entre le C α et le C β de l'aa . Son sens est donnée par le signe de l'hydrophobicité du résidu.

- On calcule les coordonnées du centre de la protéine correspondant aux coordonnées moyenne x,y et z de tous les atomes.

- Chaque vecteur est projeté orthogonalement sur l'axe liant le centre de la protéine et le Ca du résidu. Une valeur positive indique un vecteur pointant vers le centre de la protéine et une valeur négative indique un vecteur pointant vers l'extérieur.

- Le VRH7 est calculé par la somme des projections des vecteurs associés aux aa de la fenêtre considérée.

- On peut alors réaliser un profil d'hydrophobicité sur un graphe de VRH7 en fonction de la séquence. Des structures supposées semblables doivent présenter des profils d'hydrophobicité similaires. L'évaluation de leur similarité peut être mesurée par le coefficient de corrélation entre les VRH7 des résidus analogues dans les séquences.

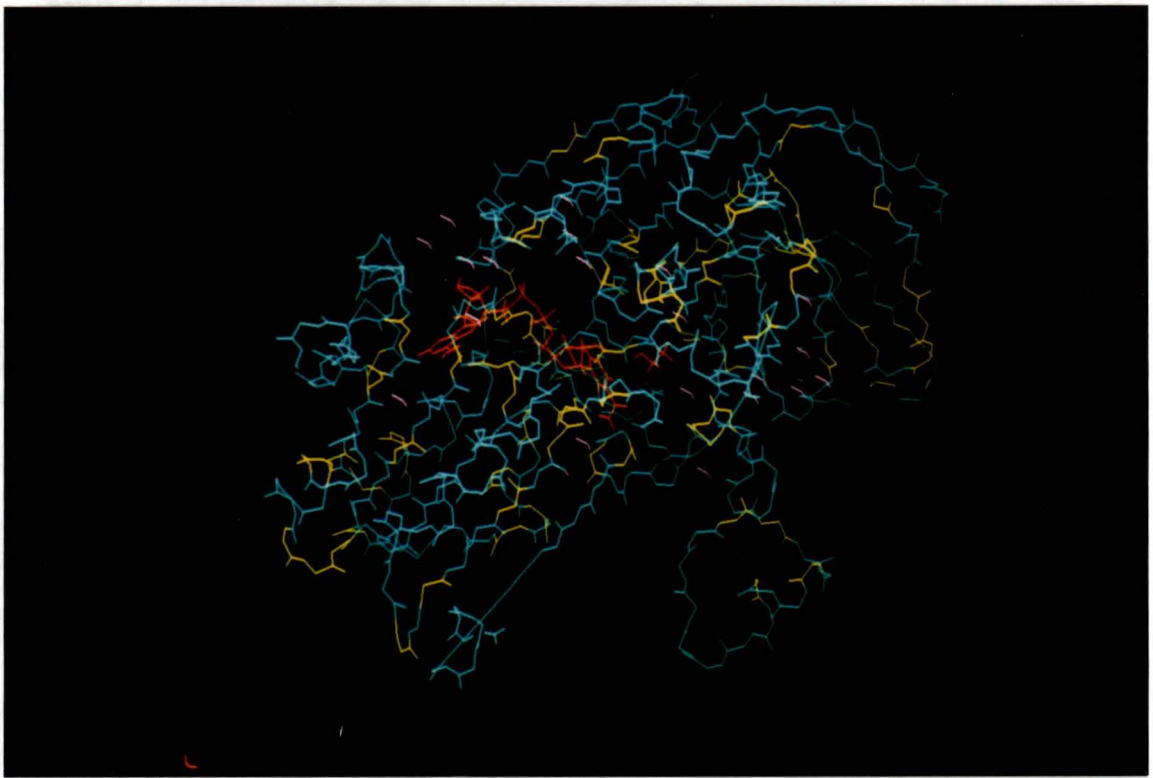


Figure 43

Les résidus conservés dans les deux enzymes sont colorés en jaune dans le modèle initial de l'alcool déshydrogénase de levure (avant minimisation d'énergie). Le substrat et le coenzyme apparaissent en rouge, les molécules d'eau apparaissent en violet.

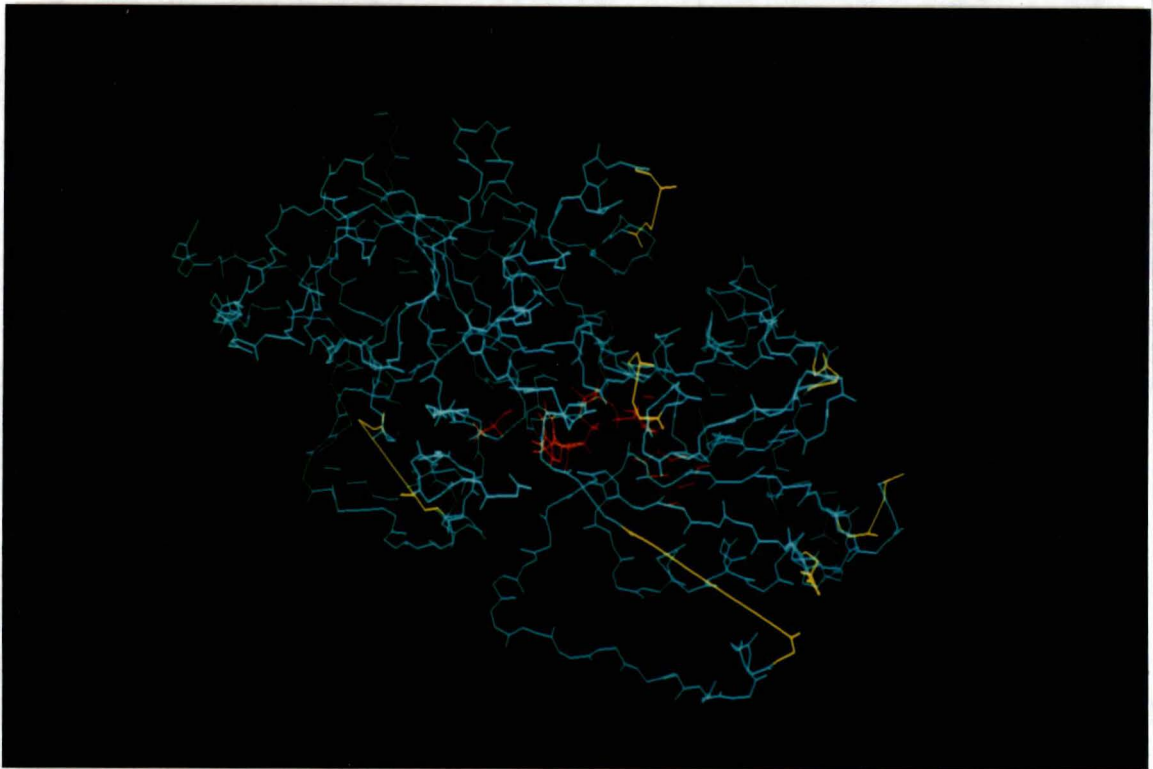


Figure 44

Les longues liaisons covalentes introduites par les délétions sont situées en jaunes dans le modèle initial (avant minimisation d'énergie) de l'alcool déshydrogénase de levure. Le coenzyme et le substrat apparaissent en rouge.

III. RESULTATS ET DISCUSSIONS:

III.1 LES ALIGNEMENTS:

III.1.1 L'ALIGNEMENT DE JÖRNVALL:

L'alignement proposé par Jörnvall (1977b & Jörnvall *et al*, 1978) à été retenu pour servir de base à la construction du modèle d'YADH fig(42). Il présente une identité globale de 25% bien que les résidus correspondant au site actif présentent une identité de l'ordre de 50%. Les résidus identiques dans les deux séquences sont situés dans le monomère d'YADH (avant minimisation) à la figure (43).

Deux grandes délétions apparaissent dans la structure initiale de YADH (fig 44). La première concerne les résidus 118 à 140 (numérotation de l'enzyme de cheval). Elle se situe dans une boucle externe et ne devrait pas induire, à priori, de grandes modifications dans la conformation de l'enzyme. L'autre délétion (311-314) se situe dans un plan β et est proche des aa fixant le groupe carboxamide du coenzyme (317 et 319) . Elle devrait entraîner un glissement de la chaîne principale de l'enzyme et donc modifier les résidus fixant le coenzyme. Cependant, les interactions entre les aa de l'enzyme et le coenzyme s'effectuent via les groupements amine et carboxylique de la chaîne principale; elles sont donc non spécifiques d'aa particuliers.

Il est important qu'un alignement n'introduise pas de résidus chargés dans la partie hydrophobe de l'enzyme ce qui provoquerait des modifications fondamentales de sa structure spatiale. Selon l'auteur, tous les résidus hydrophobes situés à l'intérieur de l'HADH sont conservés dans l'YADH. De même, toutes les glycines importantes pour la configuration de l'enzyme de foie de cheval sont conservées chez l'enzyme de levure.

Numbers refer to the sequence of the horse protein.

Function	Horse liver	Yeast			
Adenine binding pocket					
			Interior	198 Phe	Ser
				222 Val	Ile
				224 Ile	Gly
				243 Pro	Phe
				250 Ile	Val
				274 Thr	Ala
				277 Thr	Ala
				271 Arg	Ser
				273 Asp	Ala
Adenosine ribose binding					
			223 Asp	Asp	
			199 Gly	Gly	
			269 Ile	Ser	
			225 Asn	Gly	
228 Lys	Lys				
Pyrophosphate binding					
			47 Arg	His	
	269 Ile	Ser			
Nicotinamide ribose	293 Gly	Gly			
Nicotinamide	178 Thr	Thr			
Substrate binding pocket					
			57 Leu	Trp	
			93 Phe	Trp	
			110 Phe	Asn	
			116 Leu	Leu	
			140 Phe	Tyr	
			141 Leu	Thr	
			296 Pro	Ala	
318 Ile	Ile				
Acid-base system					
			48 Ser	Thr	
	51 His	His			
Ligands to active site zinc atom					
			46 Cys	Cys	
			67 His	His	
	174 Cys	Cys			

Table 6

Comparaison des aa du site actif chez les deux enzymes selon l'alignement de Jörnvall.(Jörnvall, 1977b)

Table 7

Comparaison des résidus non identiques dans le site actif des deux enzymes du point de vue de leur hydrophobicité.

HADH		YADH	
résidu	hydrophobicité	résidu	hydrophobicité
Ser 48	-0.6	Thr 46	-1.2
Arg 47	12.3	His 47	3
Phe 93	-3.7	Trp 93	-1.9
Gly 175	-1	Ala 175	-1.6
Leu 200	-2.8	Ala 200	-1.6
Ile 224	-3.1	Gly 224	-1
Ile 269	-3.1	Ala 269	-1.6
Val 294	-2.6	Met 294	-3.4
Phe 319	-3.7	Val 319	-3.6
Leu 362	-2.8	Met 362	-3.4
Ile 368	-3.1	Gly 368	-1
TOTAL:	-14.2		-15.3

Les résidus participant au mécanisme catalytique chez l'HADH se retrouvent dans l'YADH. Les aa du site de fixation du coenzyme (table 6) présentent une similarité de plus de 60%. Notamment, les aa 178, 199, 223, 228 et 293 participant activement à la fixation du coenzyme (Eklund, 1984), sont identiques dans les deux enzymes. Les aa non identiques sont très semblables quant à leur hydrophobicité comme le montre la table (7). Les valeurs sont attribuées sur base de l'échelle d'hydrophobicité d'Eisenberg (Eisenberg, 1984). Il faut cependant se rendre compte que les valeurs de l'hydrophobicité de deux aa sont insuffisantes pour évaluer leur similarité car l'orientation de leur chaîne latérale est également très importante. Deux résidus similaires dont les chaînes latérales pointent dans des directions différentes risquent d'influencer différemment les résidus qui leur sont proches. C'est pourquoi a été développée la notion de "moment hydrophobique" (Eisenberg, 1986) qui intègre ces deux paramètres. Les résidus formant la poche de fixation du substrat sont en général plus volumineux chez l'YADH ce qui pourrait expliquer sa spécificité plus stricte (Jörnvall *et al*, 1978). En particulier, la Ser 48 et la Phe 93 de l'HADH sont remplacées respectivement par une Thr et un Trp chez l'enzyme de levure.

Ces hypothèses n'ont pas été vérifiées puisqu'aucun modèle, autre qu'une simple superposition de la séquence de YADH sur la structure de HDAH, n'a été construit.

III.1.2 AUTRES ALIGNEMENTS:

Les séquences d'HADH et de YADH ont été alignées grâce aux modules GAP et BESTFIT du logiciel du GCG Laboratory (Univ. of Wisconsin). Deux alignements ont été réalisés par chaque module, l'un avec un facteur 5 de pondération des délétions et l'autre avec un facteur de 3. L'allure générale des alignements est très semblable (fig 45). Les deux grandes délétions (résidus 119-139 et 310-315) de l'alignement de Jörnvall se retrouvent légèrement décalées dans tous les alignements réalisés avec GCG (résidus 117-144 et 304-307). Les pourcentages de similarité sont de 47,5% (GAP) et 48% (BESTFIT). Les pourcentages d'identité sont respectivement de 22% et 23%.

Ces alignements ne sont pas fondamentalement différents de celui proposé par Jörnvall. D'autres alignements avaient été testés par Jörnvall *et al* (1978) mais présentaient moins d'identités. D'autre part, un alignement multiple de 16 alcool déshydrogénases différentes (Jörnvall et Persson, 87) donne lieu au même alignement que précédemment (sauf l'insertion 58B du premier alignement qui n'est pas conservée dans l'alignement multiple et est remplacée par une seconde insertion en position 57C. Nous avons donc choisi de baser notre modèle sur l'alignement proposé par Jörnvall en 1977.

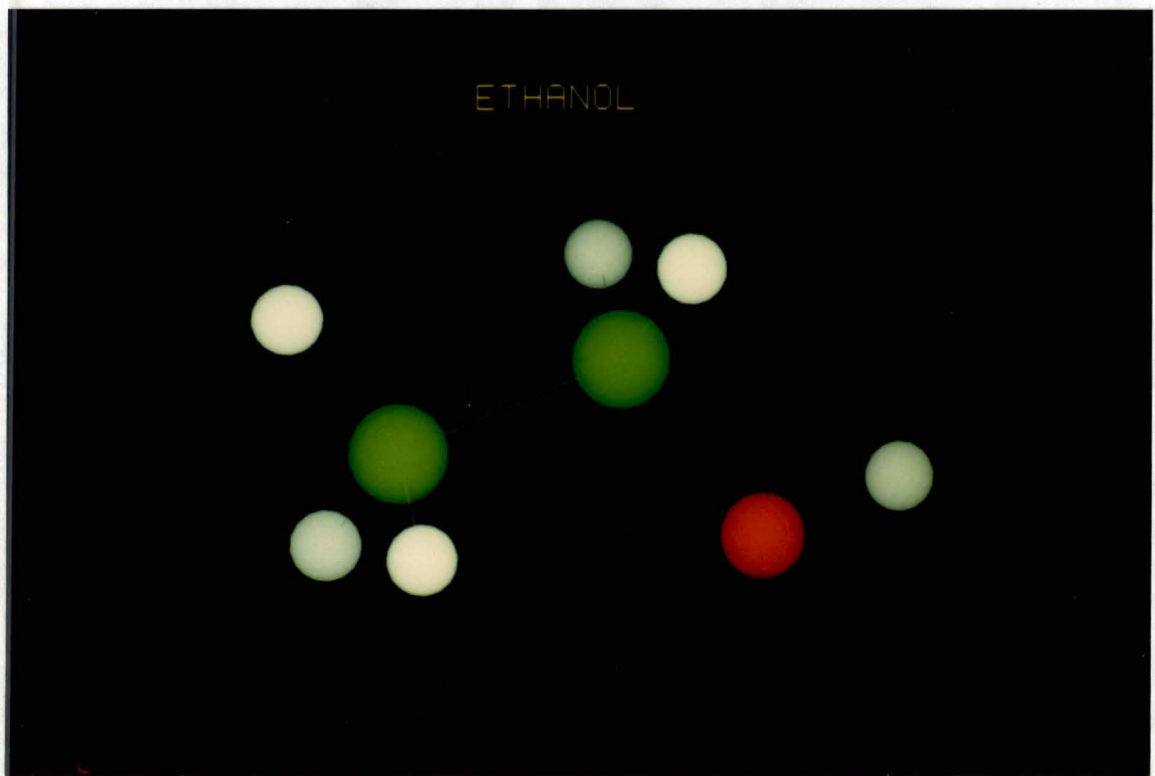


Figure 46

Représentation schématique de l'éthanol. Les hydrogènes sont en blancs, les carbones en vert et l'oxygène en rouge.

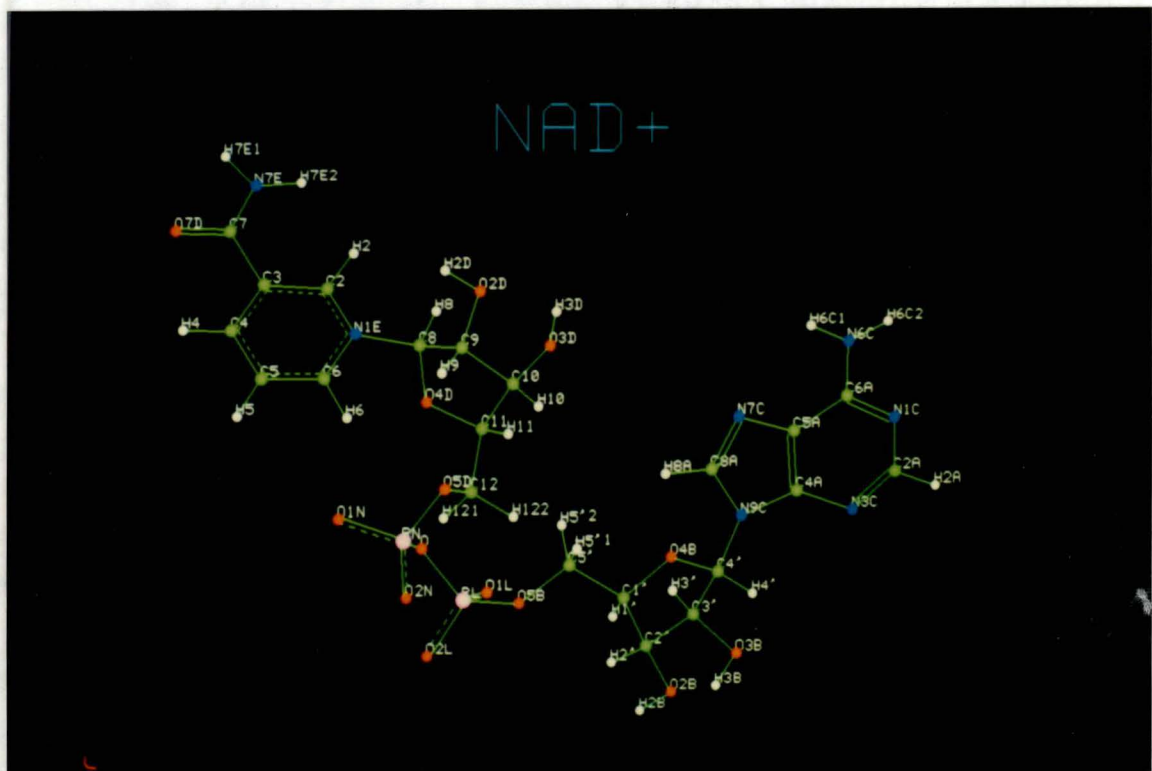


Figure 47

Représentation schématique du NAD⁺. Les noms des atomes sont indiqués. Les carbones sont colorés en vert, les azotes en bleu, les oxygènes en rouge, les phosphore en violet et les hydrogènes en blanc. Les types de liaisons sont également représentés; les pointillés symbolisent une résonance.

III.2 MODELE AVANT MINIMISATION:

La structure de la chaîne principale de l'YADH avant minimisation est donnée à la figure (44) . Les longueurs des chaînes peptidiques délétées sont reprises dans la table (8) ci-dessous (numérotation de l'HADH):

Délétion entre les aa n°	longueur (Å)
<u>118-140</u>	<u>13.32</u>
180-182	3.09
216-218	1.30
255-257	3.26
285-287	2.87
<u>310-315</u>	<u>9.92</u>

Son énergie potentielle totale est de 1388760 Kcal/mole et sa dérivée maximum de 79630 (Un modèle est considéré comme relaxé lorsque sa dérivée maximale est inférieure à 0,01)! L'énergie des interactions covalentes est à elle seule de 1018277 Kcal/mole ce qui est compréhensible au vu de la longueur de certaines délétions. La charge globale de l'apoenzyme est nulle .

Les modèles du substrat (l'éthanol) et du coenzyme (NAD⁺) sont donnés aux figures (46 et 47). Tous les atomes du cycle du nicotinamide sont coplanaires. La charge positive est délocalisée uniquement à l'intérieur du cycle. La charge globale du coenzyme est de -1 (deux groupement phosphates négatifs et une charge positive délocalisée sur le cycle nicotinamide).

III.3 RELAXATION DE LA MOLECULE:

Les mutations introduites dans la structure cristallographique de l'HADH ont créé des tensions entre certains atomes ce qui explique sa dérivée maximale élevée. Si certaines précautions ne sont pas prises lors de la minimisation d'une telle structure, ces tensions risquent de provoquer une redistribution incontrôlable des atomes et la conformation du modèle peut être totalement modifiée. Il est donc important de relaxer progressivement la molécule pour éviter que des zones importantes (le site actif et le site de fixation du coenzyme) ne perdent leur conformation native.

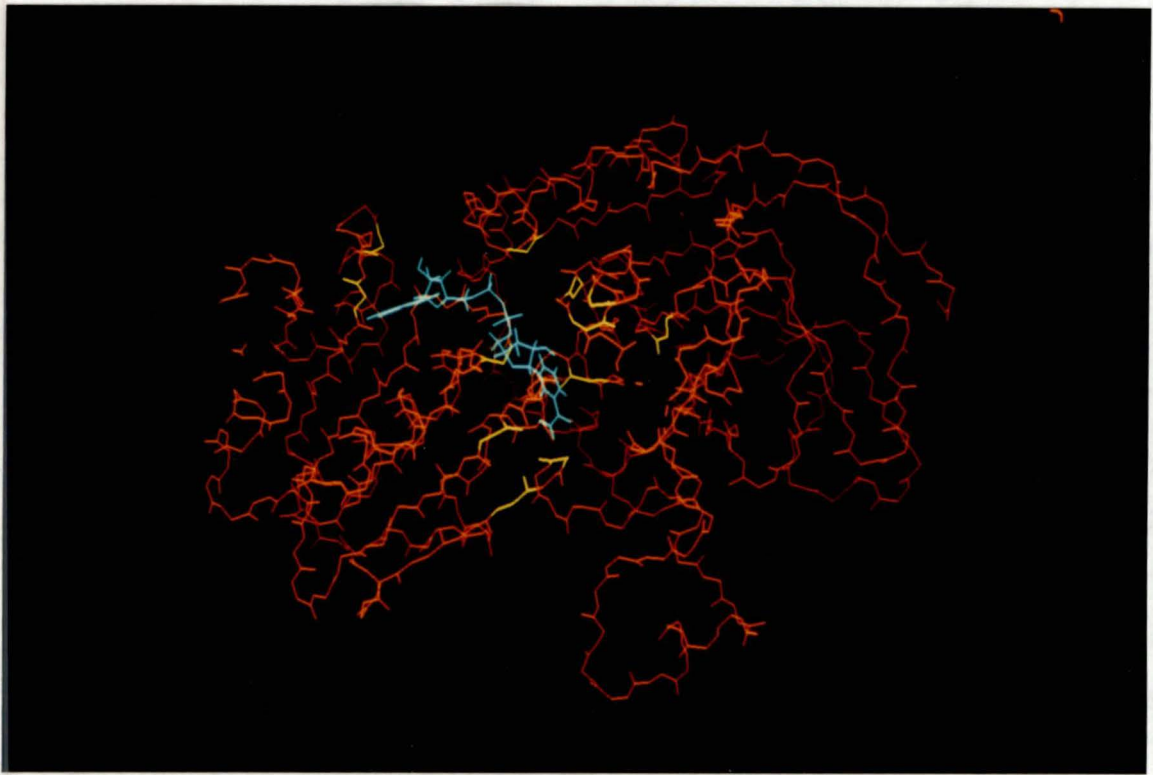


Figure 48

Les résidus ayant été fixés durant la première partie de la minimisation d'énergie apparaissent en jaune dans le modèle tridimensionnel de l'alcool déshydrogénase de levure. Le coenzyme est coloré en bleu et est également fixé.

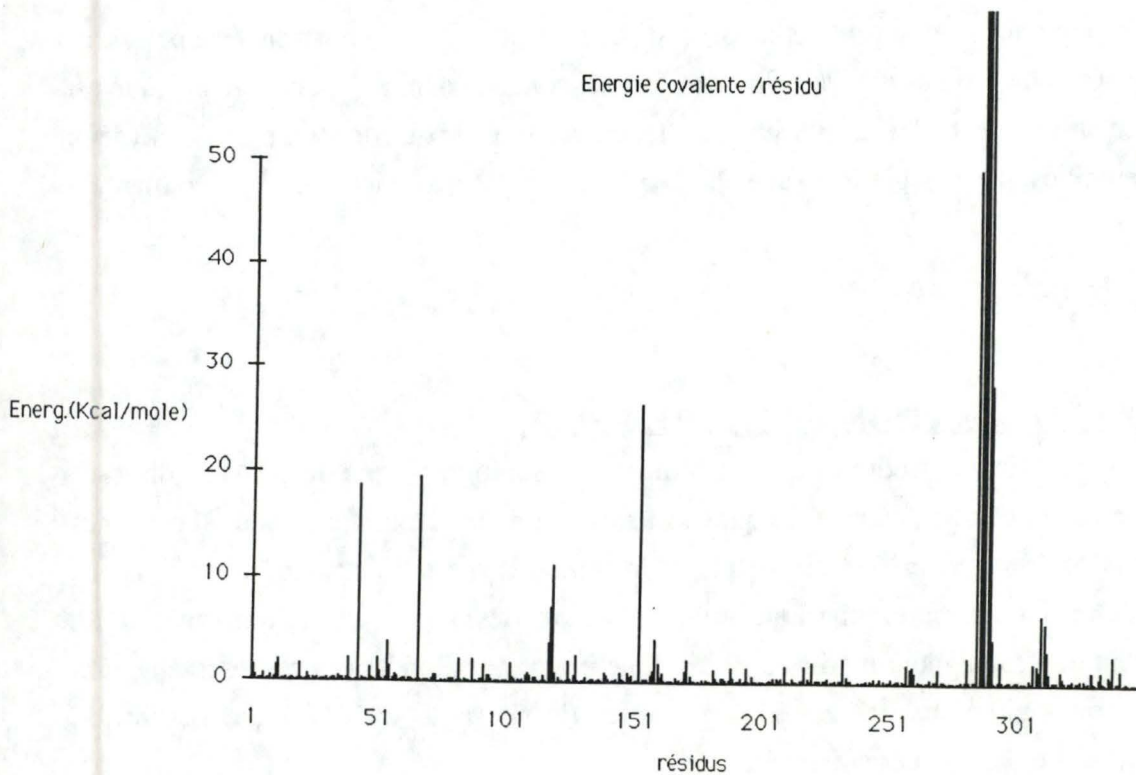


Figure 49

Histogramme de distribution de l'énergie des liaisons covalentes du modèle de l'alcool déshydrogénase de levure ayant été relaxé pendant 600 itérations.

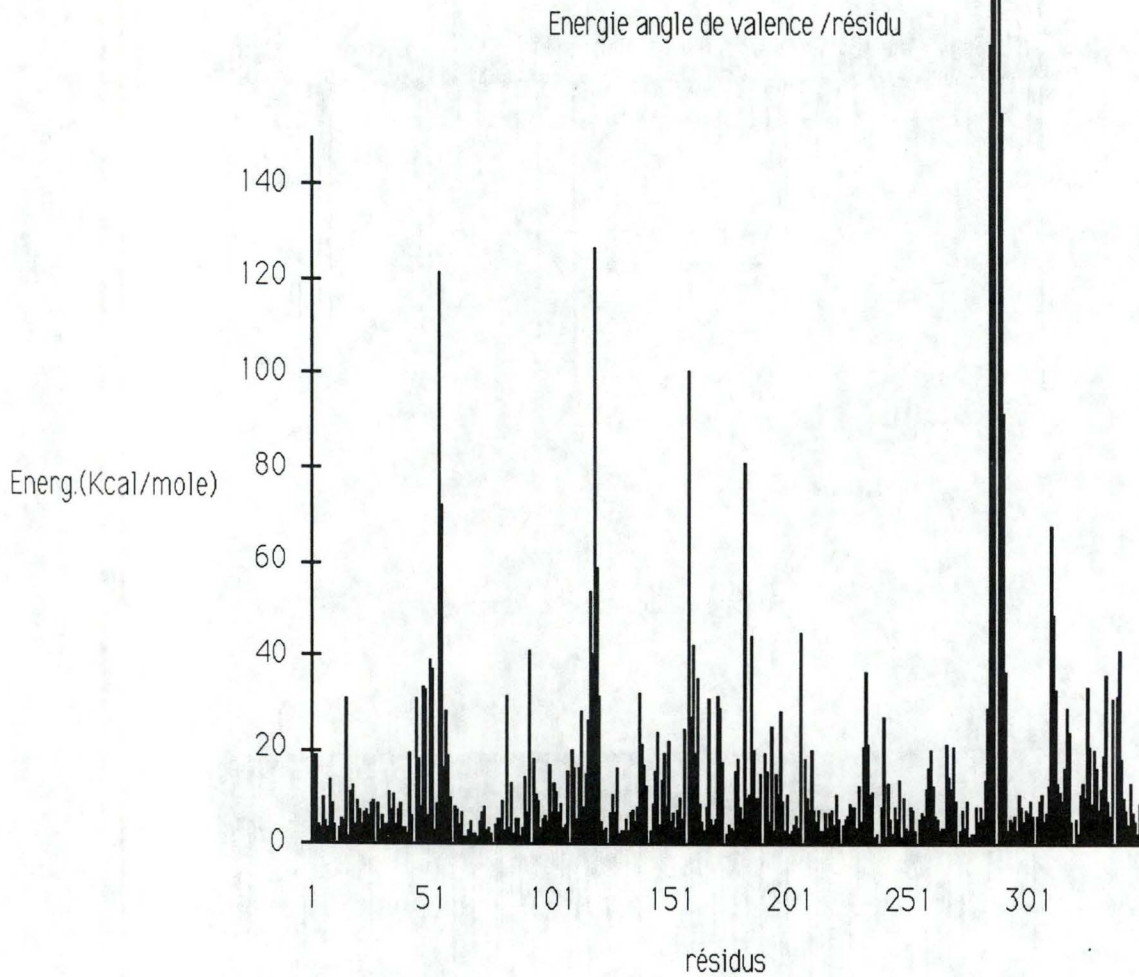
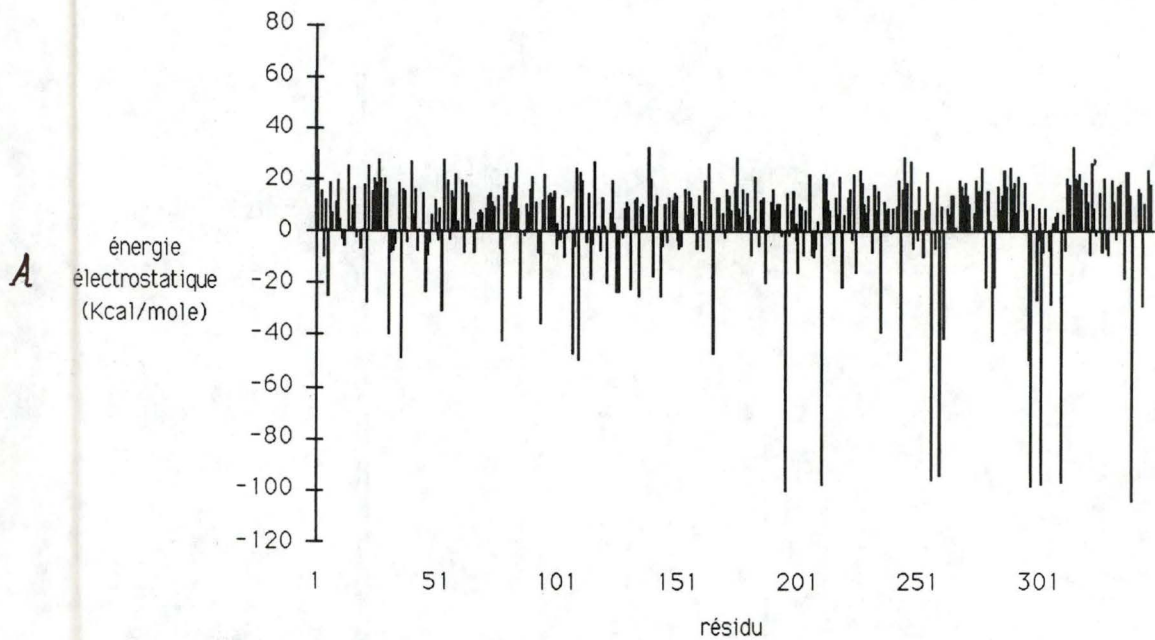


Figure 50

Histogramme de distribution de l'énergie des angles de valence du modèle de l'alcool déshydrogénase de levure ayant été relaxé pendant 600 itérations.

Interactions de Coulomb



Energie électrostatique /résidu

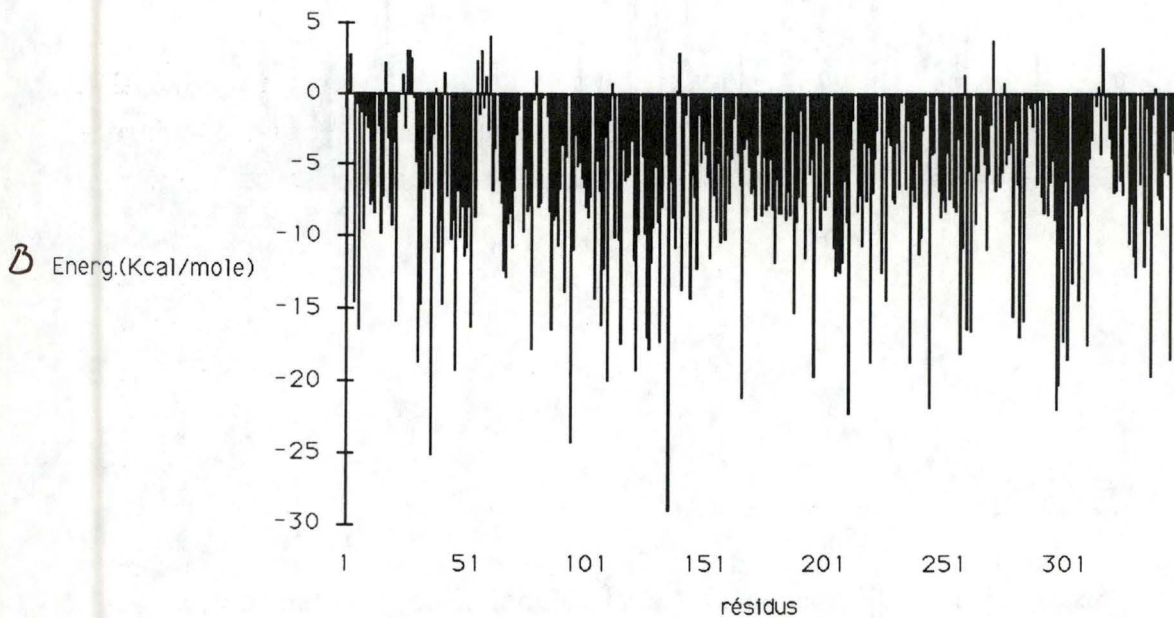


Figure 51

Histogrammes de distribution de l'énergie d'interactions électrostatiques du modèle d'alcool déshydrogénase de levure dans le vide (A) (toutes les valences sont saturées) et à pH 7 (B).

Dans un premier temps, les atomes des résidus du site actif ainsi que ceux des ligands du Zn^{++} , du coenzyme et du substrat (fig 48) sont donc fixés dans leur position initiale. Ils seront progressivement libérés dans la suite de la minimisation lorsque le reste de la molécule sera stabilisé.

Suite à une première minimisation de 600 itérations en "steepest descent", l'énergie potentielle totale du système est passée de 1400000 Kcal/mole à 16000 Kcal/mole et la dérivée maximum de 79630 à 142. La figure (49) nous montre que les énergies des liaisons covalentes entre les aa bordant les délétions sont très élevées. Les angles de valences de ces aa présentent également des énergies trop élevées (fig 50).

La seconde partie de la minimisation d'énergie va donc être consacrée à réduire l'énergie de ces interactions.

III.4 MINIMISATION DES LIAISONS COVALENTES:

Pour tenter de modéliser la protéine dans des conditions proches de celles rencontrées en milieu physiologique, l'apoenzyme est placé à pH 7. Cette opération introduit des charges supplémentaires sur les résidus dont le pK est supérieur à 7 (Glu, Asp, His). La figure (51) montre la différence dans la distribution des charges sur les résidus de l'YADH sous vide et à pH 7. Dans cette dernière, la majorité des résidus possèdent une énergie d'interaction électrostatique négative, correspondant à une force d'attraction entre les atomes. La charge globale de l'apoenzyme est maintenant de -1.

Les 7000 itérations suivantes sont réalisées par la méthode du "conjugate gradient" et sans les "cross terms". Nous avons mentionné précédemment que le modèle est considéré en l'absence de solvant, ce qui permet de limiter le nombre d'atome du système. Les énergies d'interactions électrostatiques sont donc calculées à partir d'une fonction inversement proportionnelle au carré de la distance séparant les atomes ce qui atténue les interactions entre des atomes éloignés (cfr § II.8). Une autre fonction (*scale 1-4*) permet d'atténuer les interactions de Van der Waals entre deux atomes chargés séparés par trois liaisons covalentes successives. L'atténuation de ces interactions par un facteur 0.5 permettra aux deux angles de valence formés par les trois liaisons covalentes de se relaxer sans subir l'influence des interactions électrostatiques entre deux atomes en question. Les mêmes résidus que précédemment restent fixés dans leur position initiale. La figure (52) montre l'évolution de la dérivée maximum et de l'énergie potentielle totale durant ces 7000 itérations. Elle permet de suivre les différentes étapes de la minimisation des liaisons covalentes:

Evolution de la dérivée et de l'énergie

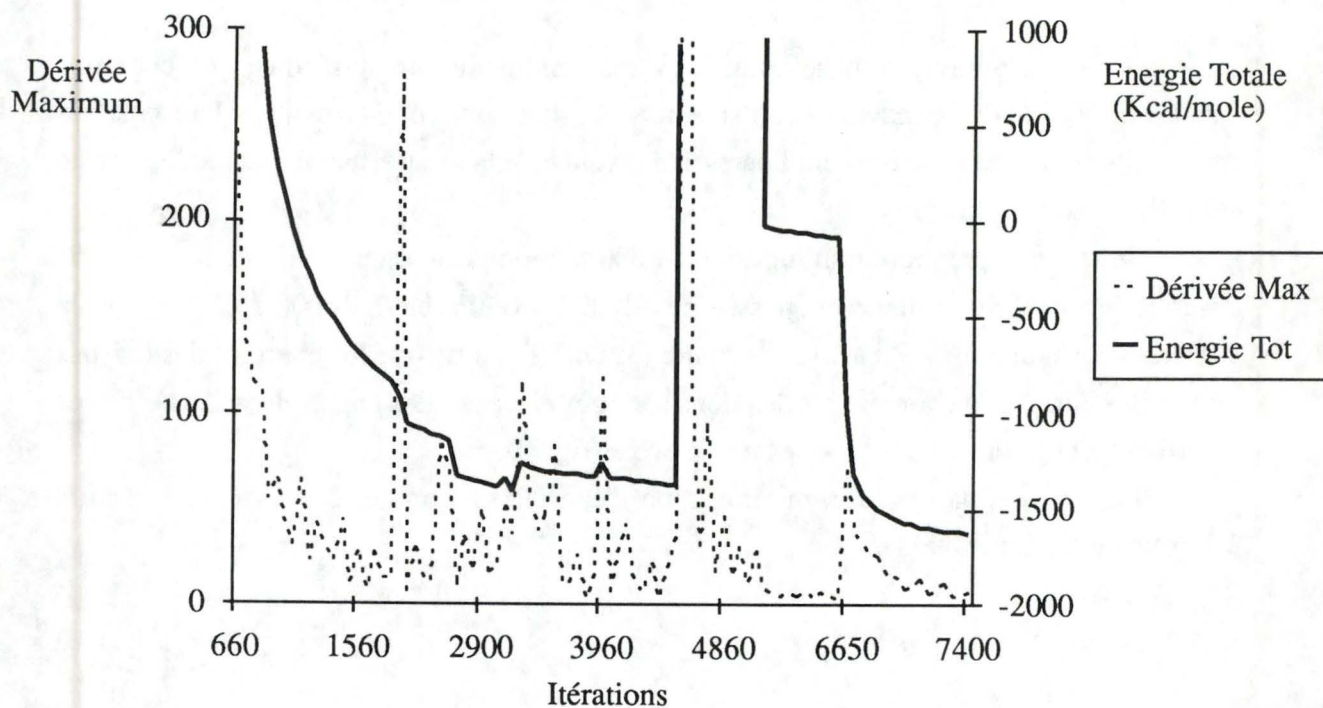


Figure 52

Evolution de l'énergie potentielle totale et de la dérivée maximale d'énergie potentielle depuis le stade des 660 itérations jusqu'au stade des 7550 itérations.

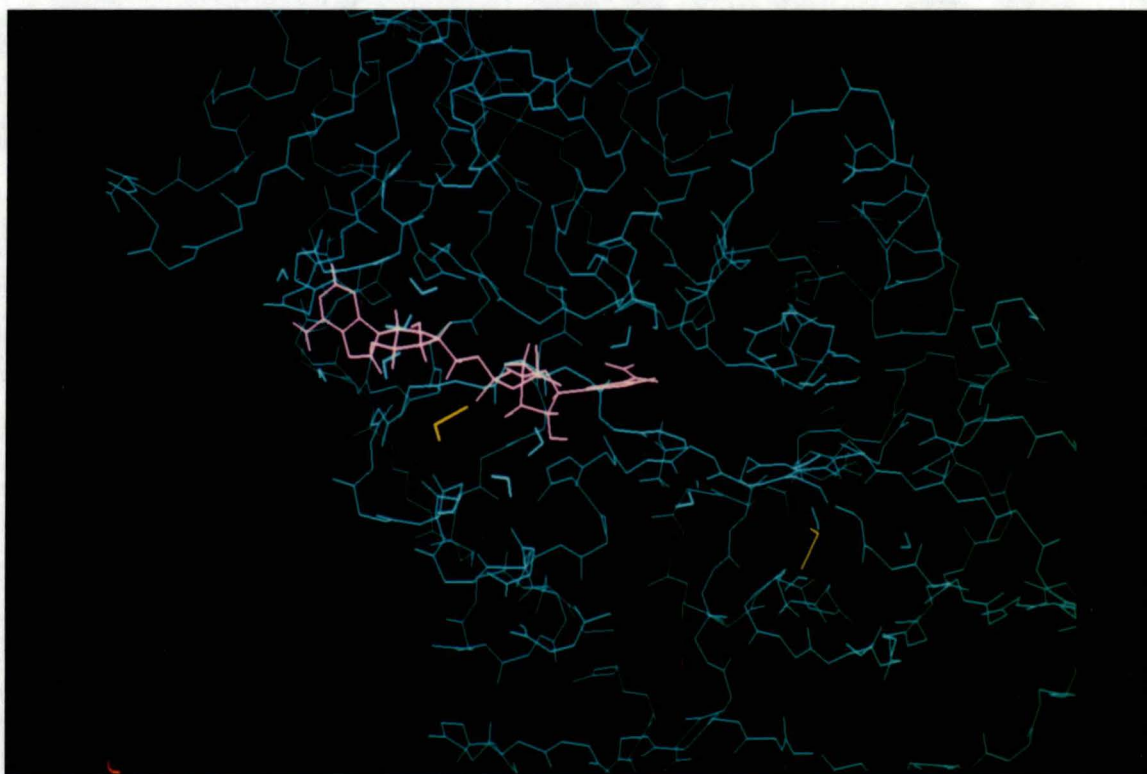


Figure 53

Vue d'une partie du modèle du monomère d'alcool déshydrogénase de levure.

Les molécules d'eau dont un des hydrogènes tend à former une liaison avec un atome voisin sont colorées en jaune. Le coenzyme est coloré en violet.

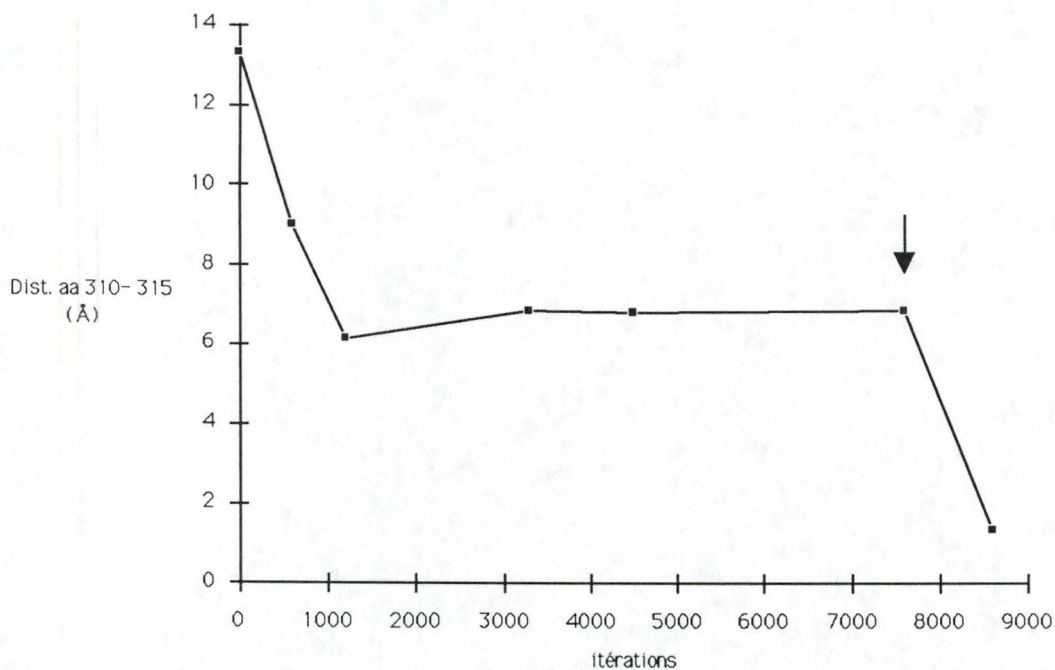


Figure 54

Evolution de la longueur de la liaison covalente entre les aa 310 et 315 au cours de la minimisation de l'énergie du modèle enzymatique.

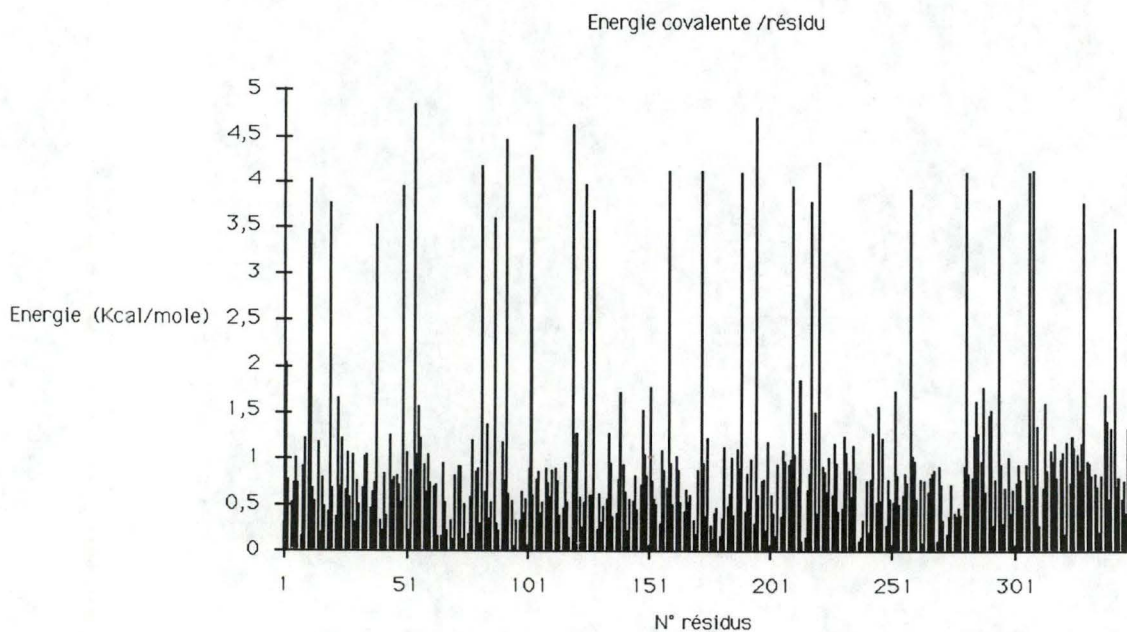


Figure 55

Histogramme de distribution de l'énergie des liaisons covalentes du modèle de l'alcool déshydrogénase de levure ayant été relaxé pendant 7600 itérations.

III.4.1 Etapes:

Après 3700 itérations, deux molécules d'eau montrent des énergies covalentes très élevées table 9:

Mol. n°	Energ. cov.
9	89 Kcal/m
24	83 Kcal/m

La (fig 53) montre qu'un de leurs hydrogènes tend à former un pont hydrogène avec une molécule voisine (une avec une autre molécule d'eau et l'autre avec un groupement "OH" du coenzyme) ce qui étire la liaison covalente entre cet hydrogène et l'oxygène. Ces anomalies provoquent une instabilité plus grande du niveau de la dérivée maximale visible sur la figure (52) entre 1900 et 4000 itérations La minimisation est poursuivie après avoir réajusté la longueur de la liaison covalente des molécules d'eau.

Après 4500 itérations, toutes les délétions sont résorbées sauf celles entre les aa 310-315 ce qui provoque la stagnation de l'énergie maximale. Pour tenter de recréer un lien peptidique normal entre ces deux aa, 600 et 1500 itérations sont effectuées en multipliant les interactions covalentes respectivement par 100 et par 10. Ceci induit une augmentation brutale de l'énergie totale et de la dérivée maximale visibles sur le graphe de la figure 52. Malgré cette manipulation, la taille de la délétion stagne à une valeur de 6 Å. Deux aa proches de la délétion, les aa 317 et 319, appartiennent au site actif et sont fixés dans leur position initiale. Ils doivent donc empêcher la distance séparant les atomes situés de part et d'autre de la délétion. De plus, la fonction employée pour le calcul des interactions covalentes est non quadratique c-à-d que l'énergie des liaisons covalentes n'augmente que lentement en fonction de la longueur de la liaison. La minimisation a donc été poursuivie par 1000 itérations en utilisant la fonction quadratique (harmonique) pour décrire l'énergie des liaisons covalentes et en fixant uniquement les atomes bordant la poche du site actif et du site de fixation du NAD⁺. Suite à ces modifications, la longue liaison covalente introduite par la délétion se résorbe en à peine 500 itérations (fig 54).

L'énergie covalente de chaque résidu du modèle au stade actuel est donnée à la figure (55). D'une manière générale, les aa dont la chaîne latérale comprend un cycle à 6 carbones possèdent des énergies de liaisons covalentes plus élevées. Ceci est dû aux tensions existant dans les liaisons du cycle. Néanmoins, aucune liaison covalente ne présente une énergie supérieure à la moyenne.

Une liste des angles de torsion ϕ , ψ et ω est imprimée à la fin de cette première étape. Elle va permettre la vérification et l'ajustement si besoin est, de la conformation des

Energie angle de torsion /résidu

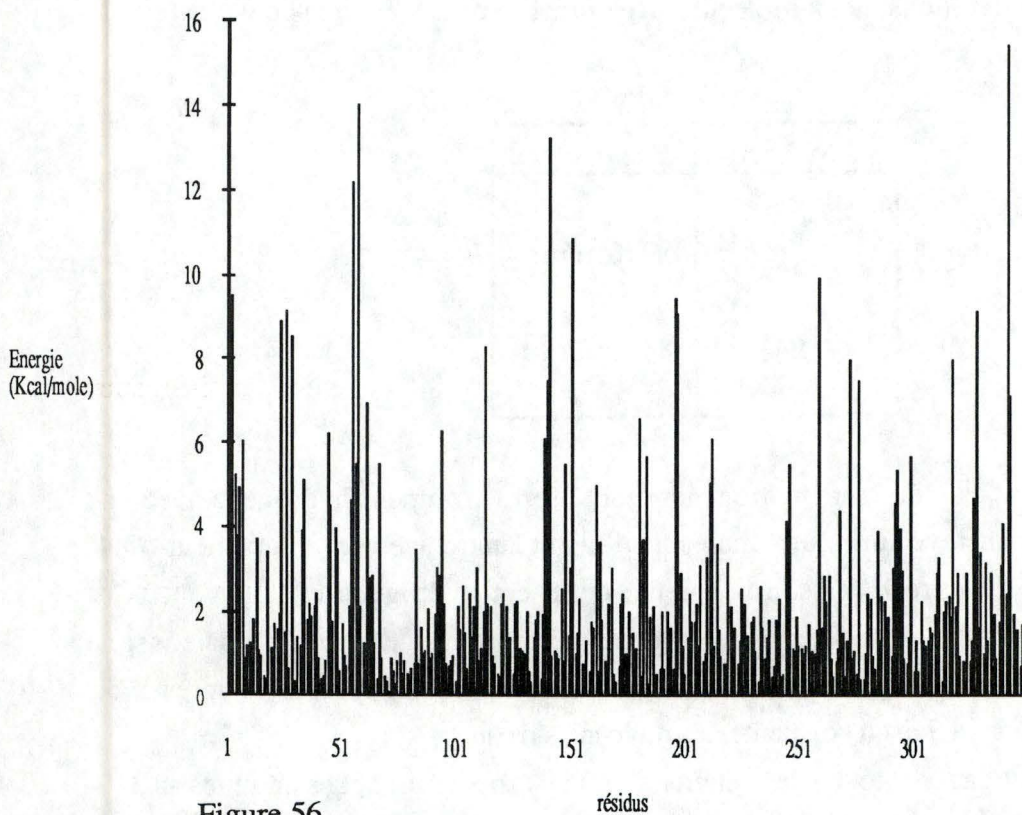


Figure 56

Histogramme de distribution de l'énergie des angles de torsion du modèle de l'alcool déshydrogénase de levure ayant été relaxé pendant 7600 itérations.

Evolution de la dérivée et de l'énergie

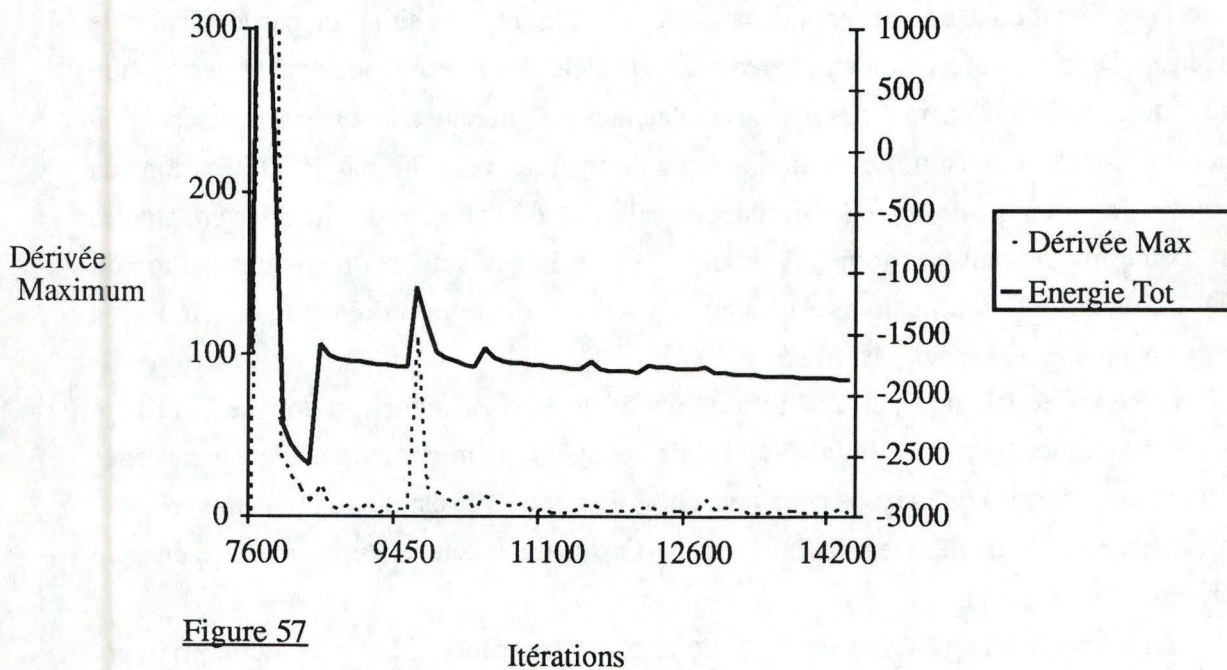


Figure 57

Evolution de l'énergie potentielle totale et de la dérivée maximale d'énergie potentielle du modèle d'alcool déshydrogénase de levure depuis le stade des 7600 itérations jusqu'au stade des 14400 itérations.

angles de torsion des aa de l'enzyme. Ceci constitue la troisième étape de l'élaboration du modèle.

III.5 VERIFICATION DES ANGLES DE TORSION:

L'énergie totale des angles de torsion pour chaque aa de la protéine est donné à la figure (56). Certains aa présentent une énergie beaucoup trop élevée. Une liste d'angles de torsion présentant une conformation incorrecte est établie sur base d'une comparaison entre l'HADH et l'YADH (voir § II.9). 7000 itérations supplémentaires vont être nécessaires pour amener les énergies de ces angles à des valeurs acceptables (c-à-d du même ordre de grandeur que la moyenne des énergies de tous les angles de torsion de la protéine). L'évolution de l'énergie totale ainsi que de la dérivée maximale au fur et à mesure des itérations est donnée à la figure (57). Une première série de 2600 itérations dont 200 sont réalisées en multipliant les énergies des angles de torsion par 100 permettent à quelques angles d'adopter la conformation requise (premier pic sur le graphe). Les angles dont la conformation n'a pas été corrigée sont forcés (voir § II.9) durant un nombre d'itérations donné avec une énergie de 100 Kcal/ rad² jusqu'à obtenir la valeur désirée. Ces modifications des paramètres de la minimisation engendrent des fluctuations du niveau de la dérivée maximale. L'énergie potentielle globale, quant à elle, reste stable durant ces manipulations car la contribution de l'énergie de ces angles de torsion à l'énergie totale est relativement faible.

La table (10) ci-après donne pour chaque angle de torsion, la valeur observée, la valeur attendue, le nombre total d'itérations et le nombre d'itérations au cours desquelles il a été forcé (force) à adopter une conformation adéquate. Les aa numérotés "B" sont ceux correspondant à des insertions dans la séquence de YADH.

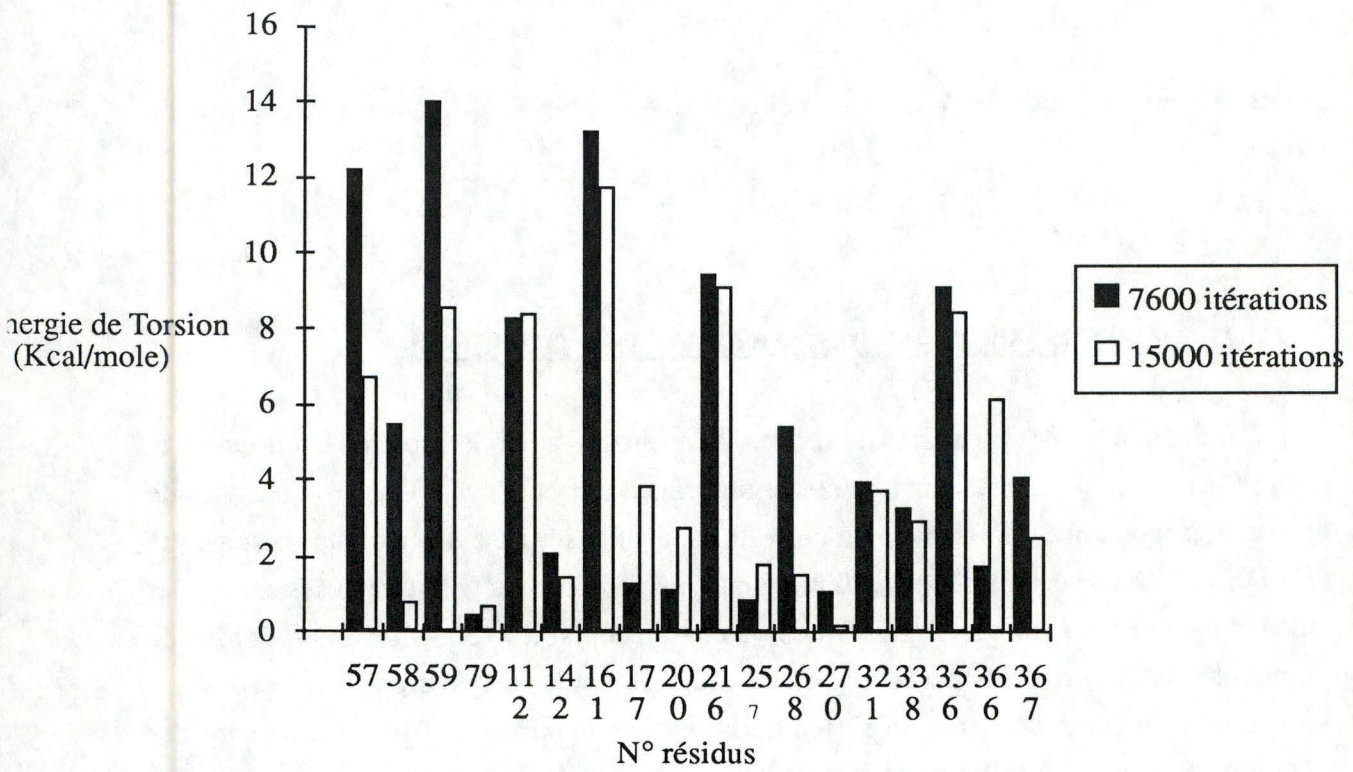


Figure 58

Histogramme montrant la différence entre l'énergie des angles de torsion des aa de la table 11 avant et après correction.

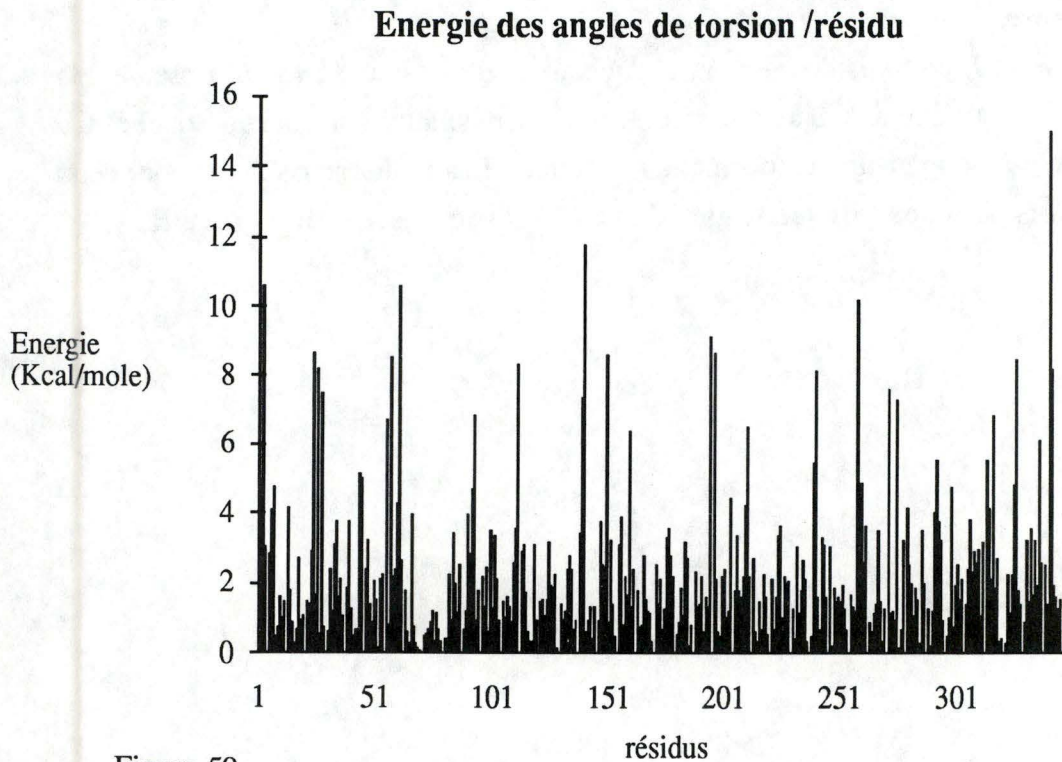


Figure 59

Histogramme de distribution de l'énergie des liaisons covalentes du modèle de l'alcool déshydrogénase de levure ayant été relaxé pendant 14500 itérations.

N° aa	ϕ		ψ		ω		nombre d'itérations pour régularisation	
	Obs.	Att.	Obs.	Att.	Obs.	Att.	force	total
Pro 57B	63.6	-80	-	-	-	-	400	4000
Pro 58 B	-	-	-	-	-5.3	-180	400	4000
Leu 58	-	-	-	-	8.5	180	200	1500
Asn 79	67.24	-100	28.1	120	-	-	300	2000
Pro 112	-	-	3.65	50	-	-	100	500
His 142	97.1	-120	-26.5	70	-	-	400	4000
Ile 177	-	-	-34.6	-70	-	-	100	500
Ala 200	-	-	-	-	13	180	400	4000
Tyr 216	74.5	-60	-	-	129	180	300	2000
Ala 257	-	-	-1.7	-60	-	-	100	500
Val 268	-	-	27.2	140	-	-	300	2000
Val 270	62.2	-80	-	-	62.7	180	200	1500
Gly 320	-	-	-	-	3.9	180	100	500
Ser 321	50.7	-80	-	-	-	-	300	2000
Arg 338	-	-	-	-	5	180	300	2000
Pro 356	-	-	17	50	-	-	100	500
Gln 366	-	-	-	-	0.8	180	100	500
Val 367	45	-80	-	-	-	-	200	1500

Une fois les angles corrigés, 3000 itérations sans "force" ont permis d'ajuster les angles dans la conformation la plus stable possible. La figure (58) met en évidence la différence d'énergie de torsion des résidus concernés avant et après correction des angles. La figure (59) donne les énergies des angles de torsion de chaque résidu à ce stade de la minimisation (l'énergie du modèle a été minimisée durant 14600 itérations). les prolines conservent des énergies plus élevées que la moyenne générale ce qui peut être expliqué par la conformation très stricte de leur angle ϕ (-80°).

Energie angles de valence /résidu

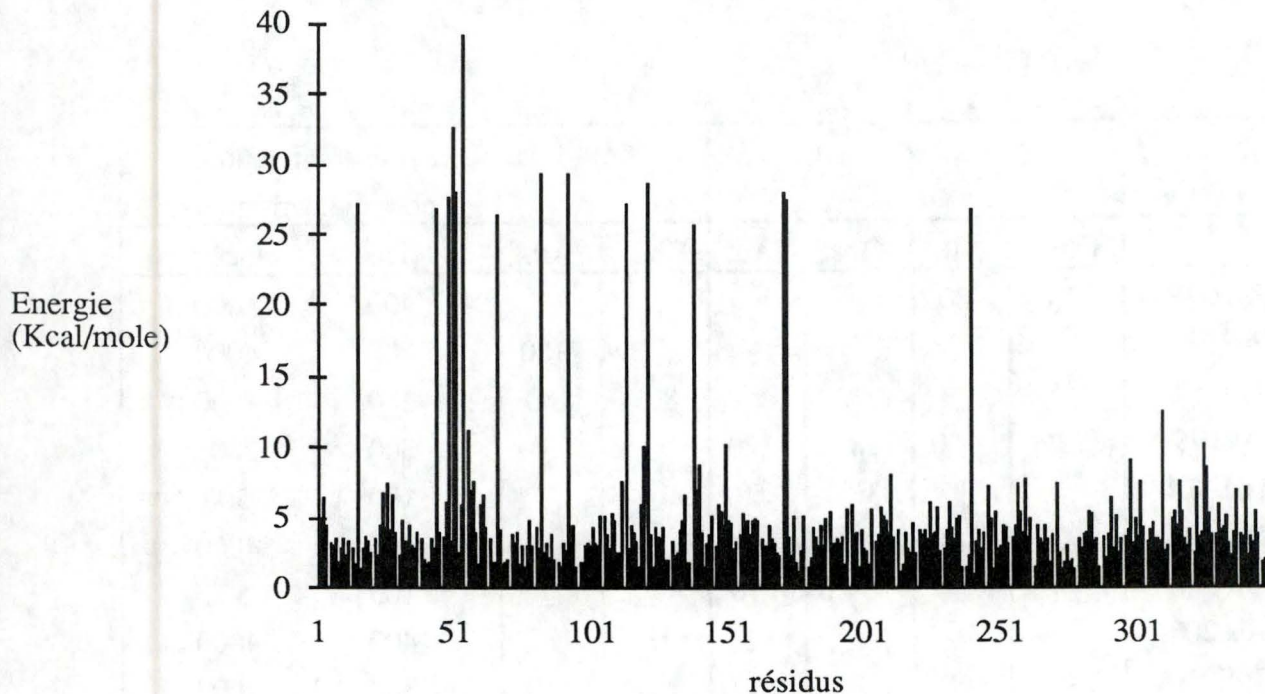


Figure 60

Histogramme de distribution de l'énergie des angles de valence du modèle de l'alcool déshydrogénase de levure ayant été relaxé pendant 14500 itérations.

Energie de Van der Waals/résidu

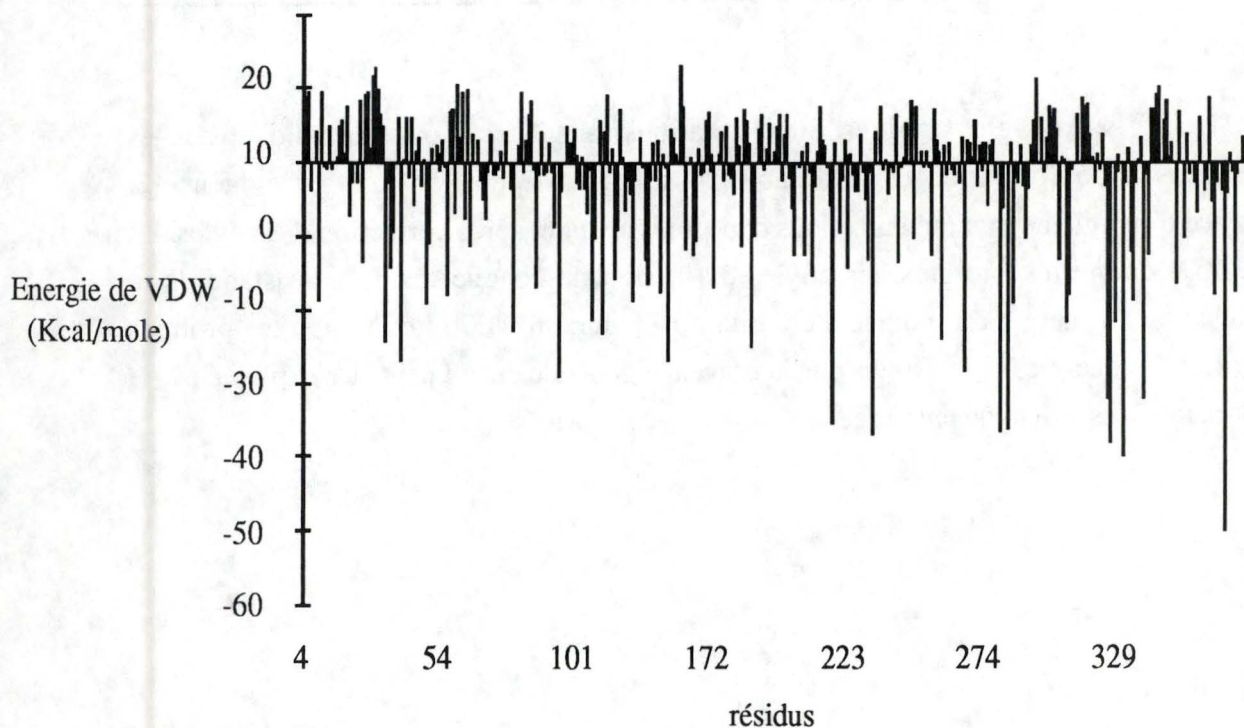


Figure 61

Histogramme de distribution de l'énergie de Van der Waals du modèle de l'alcool déshydrogénase de levure ayant été relaxé pendant 15000 itérations.

III.6 LIBERATION PROGRESSIVE DES ATOMES FIXES:

A ce stade de la minimisation, l'énergie et la dérivée totale du modèle sont suffisamment faibles pour que l'on puisse commencer à libérer progressivement les atomes fixés sans risquer de modifier le site actif suite à des interactions trop élevées avec des atomes voisins.

La figure (60) nous montre également que les angles de valence sont tous à un niveau énergétique d'un même ordre de grandeur. Le facteur d'atténuation des interactions 1-4 peut donc être également retiré.

3000 itérations sont effectuées sans aucun facteur d'atténuation ou de multiplication. Les atomes encore fixés sont ceux liés à l'ion zinc ainsi que les atomes participant au mécanisme catalytique:

His 67 atome NE2

Cys 46 atome SG

Cys174 atome SG

Thr 48 atome OG1

His 51 atome NE2

NAD⁺ atome C4

Ethanol atome O3

L'énergie globale a augmenté suite à la suppression du facteur d'atténuation 1-4. La figure (61) **graphe()** des énergies de Van der Waals montre qu'ici encore, certains résidus présentent des énergies plus élevées que la moyenne. Ces résidus sont tous des Tyr et des Trp. Leur chaîne latérale est volumineuse et donc encombrante, ce qui peut expliquer ces énergies élevées. Néanmoins, 200 itérations sont effectuées en multipliant les énergies d'interactions non covalentes par 100 de manière à relaxer les éventuelles interactions d'énergie trop élevées qui subsisteraient dans le modèle. La minimisation est ensuite poursuivie jusqu'au stade des 19000 itérations sans facteur de multiplication. Au terme de cette série d'itérations, il est utile de considérer les atomes présentant les 10 dérivées les plus importantes. Ils permettront mettre en évidence les zones les plus instables de la protéine. La table (11) ci-dessous donne la liste des atomes possédant les dix plus grandes dérivées d'énergie potentielle.

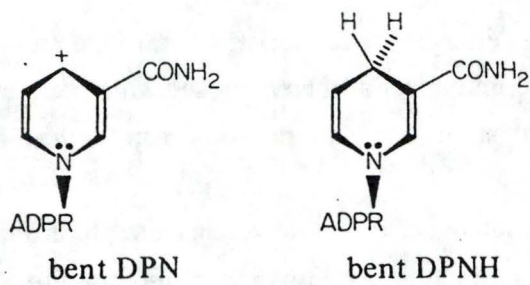


Figure 62

Schéma de la structure "bateau" adoptée par le NAD^+ dans le site actif de l'enzyme (Cook, 1981).

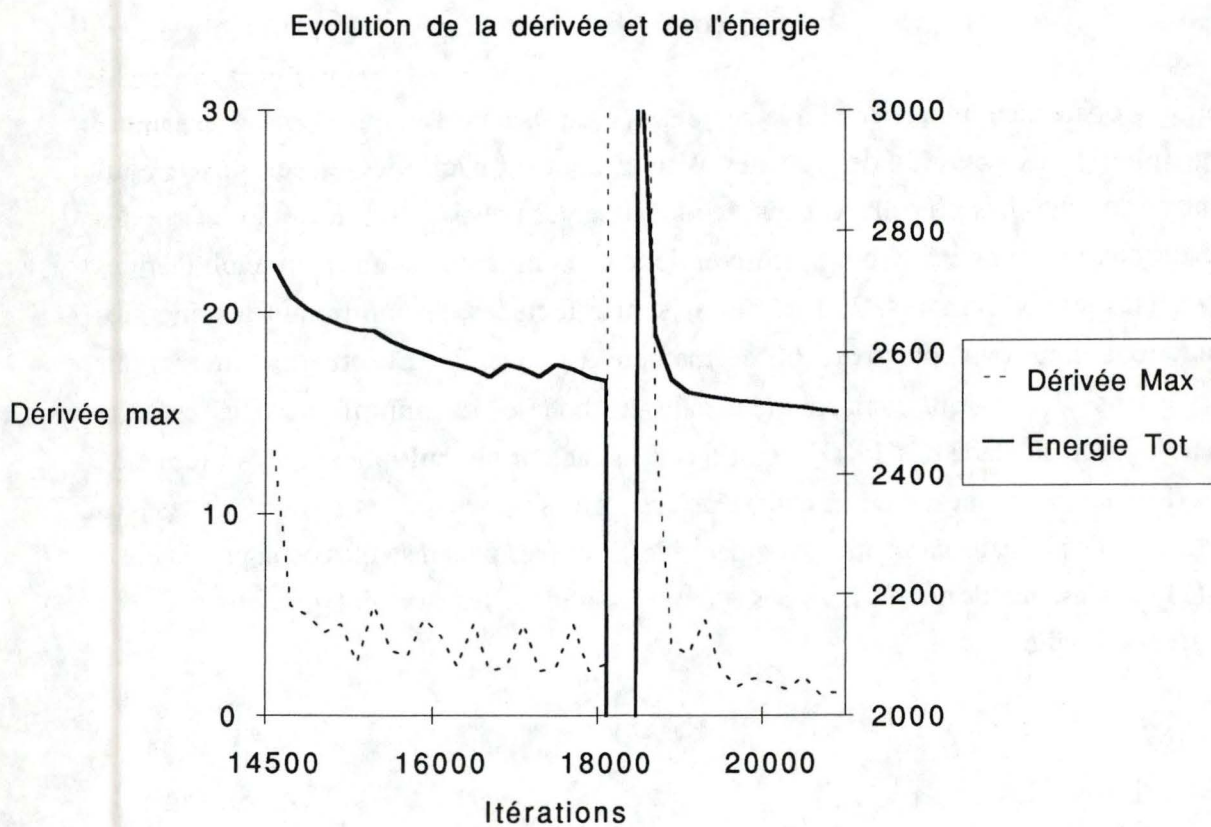


Figure 63

Evolution de l'énergie potentielle totale et de la dérivée maximale d'énergie potentielle du modèle de l'alcool déshydrogénase depuis le stade des 14550 itérations jusqu'au stade des 20800 itérations.

CARTESIAN DERIVATIVES OF ENERGY FUNCTION

Atom Name	Residue Name No.	Molecule Number	Derivative (kcal/mole-Angstrom)		
			dv/dx	dv/dy	dv/dz
N1E	NAD+ B3H	2	-1.597217	-0.411870	0.671642
C2	NAD+ B3H	2	1.299417	-0.250565	-0.411817
PL	NAD+ B3H	2	0.670695	0.914650	-0.371682
C6	NAD+ B3H	2	0.937686	0.519767	-0.524548
O2L	NAD+ B3H	2	-0.697290	-0.372769	0.378155
C8	NAD+ B3H	2	0.671332	0.377456	-0.095613
HN	ILE 316	1	0.370453	-0.629606	0.254621
N6C	NAD+ B3H	2	0.054761	0.467801	-0.574055
O1L	NAD+ B3H	2	0.099081	-0.571800	-0.185250
O	SER 315	1	-0.443129	0.397459	-0.111897

Presque toutes ces dérivées appartiennent au NAD⁺. Rappelons que chaque fois que le modèle est stocké dans la mémoire de l'ordinateur, les atomes possédant la dérivée minimale et maximale sont renseignés dans le fichier de résultats (output file). Nous sommes donc remontés à travers les étapes précédentes pour savoir quels avaient été les atomes associés aux dérivées maximales. L'atome N1E du NAD⁺ apparaissait systématiquement comme celui possédant la dérivée d'énergie maximale. Le NAD⁺ et en particulier l'atome N1E, est donc l'élément limitant la décroissance de la dérivée du modèle. Il est donc probable que le modèle du NAD⁺ employé dans ce travail n'est pas tout à fait conforme à la réalité. Selon Cook (1981), le cycle nicotinamide du coenzyme adopterait une conformation dite "bateau" (fig 62). Cette conformation permettrait peut-être un meilleur ajustement du coenzyme dans sa poche de fixation. Néanmoins, la dérivée n'a cessé de diminuer régulièrement jusqu'à présent, ce qui indique que bien que le coenzyme soit limitant dans la minimisation d'énergie, ses atomes continuent de s'ajuster.

Enfin, après 19600 itérations, tous les atomes ne participant pas directement à la fixation du Zn⁺⁺ sont libérés. Rappelons que les atomes ligands de l'ion zinc sont fixés dans leur position initiale de manière à mimer la présence de l'ion. la figure (63) résume l'évolution de l'énergie potentielle globale et de la dérivée maximale à partir du stade des 14600 itérations et durant les 6200 itérations suivantes. La suppression de l'atténuation 1-4 ainsi que la multiplication des énergies des interactions non covalentes y apparaissent clairement (variations importantes du niveau d'énergie globale).

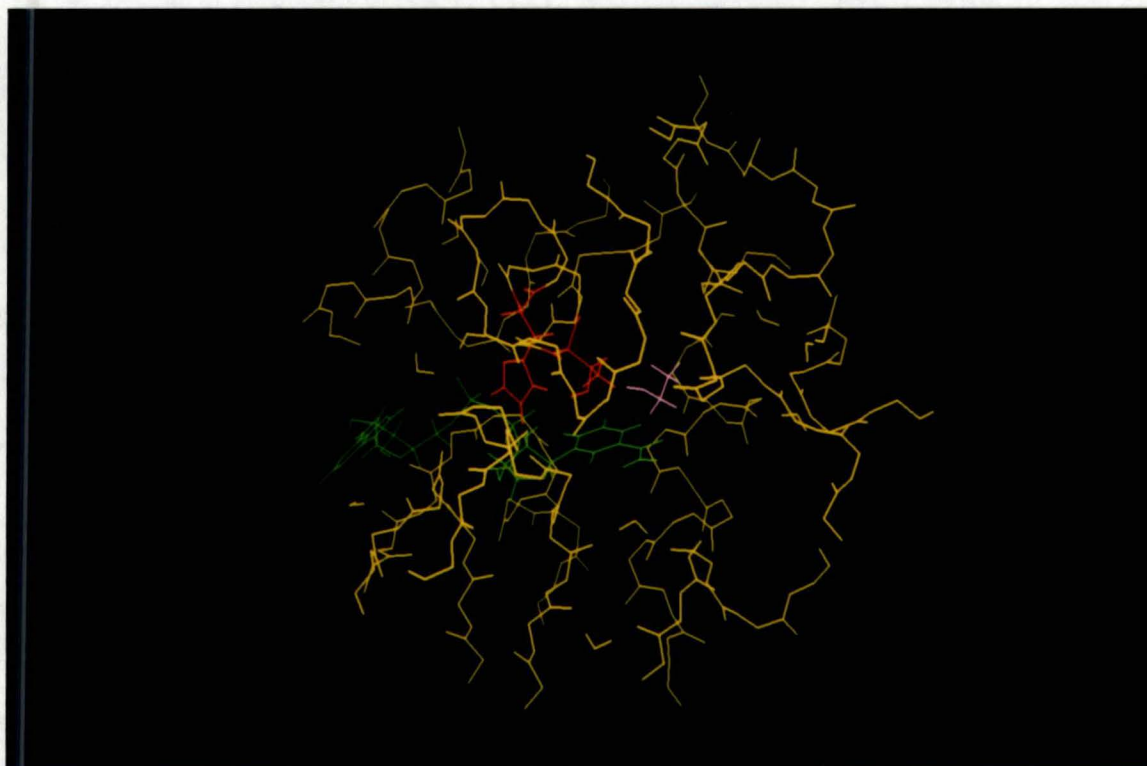


Figure 64

Représentation du site actif de l'enzyme ayant été isolé et minimisé seul. Le coenzyme est coloré en vert et le substrat en violet. Les deux aa participant probablement au mécanisme catalytique (Thr 48 et His 51) sont représentés avec leur chaîne latérale en rouge.

III.7 MINIMISATION DU SITE ACTIF:

L'énergie du modèle a été minimisée durant 20800 itérations. La décroissance de l'énergie potentielle totale ralentit nettement et la dérivée maximale se stabilise autour de 1 ce qui indique que la structure ne doit plus être très différente de la conformation recherchée. Pour tenter d'ajuster la position des atomes du site actif le plus précisément possible, celui-ci est isolé du reste de la molécule (fig 64) et son énergie est minimisée par la méthode du "conjugate gradient" et avec les "cross terms" durant 3000 itérations. Seuls les atomes ligands du Zn^{++} sont fixés. La dérivée maximale atteinte après les 3000 itérations est de 0.005. La superposition des atomes du site actif avant et après minimisation révèle un écart-type des distances séparant les atomes des deux modèles de 0.4 Å. Ces résultats indiquent que la configuration du site actif a atteint un niveau d'énergie potentielle suffisamment faible et est donc représentatif d'une structure énergétiquement plausible. Le site actif minimisé est réintégré dans le modèle et le système complet est minimisé avec les mêmes paramètres que précédemment (mais sans les "cross terms") pour 600 itérations de manière à réajuster les résidus du site actif dans l'enzyme complète.

Le modèle ainsi obtenu peut être considéré comme suffisamment représentatif d'une conformation stable pour servir de base aux études ultérieures de l'YADH.

III.8 MINIMISATION DE L'HADH:

Pour pouvoir comparer notre modèle d'YADH au modèle cristallographique de HADH, il faut que ce dernier soit préalablement traité selon la même méthode de minimisation d'énergie. L'analyse cristallographique d'une protéine ne permet pas de situer tous les atomes avec exactitude (notamment les hydrogènes). Il est donc nécessaire d'affiner le modèle de l'HADH.

L'holoenzyme est construite suivant la même méthode que l'YADH. Les hydrogènes sont placés en considérant l'apoenzyme à pH 7. La relaxation de la structure de HADH est réalisée par la méthode de "steepest descent", sans "cross terms" et avec une fonction quadratique décrivant l'énergie des liaisons covalentes. La fonction de calcul de l'énergie électrostatique utilisée est inversement proportionnelle au carré de la distance séparant les atomes interagissant. L'énergie de ce modèle est tout d'abord minimisée durant 300 itérations en fixant tous les atomes lourds dans leur position initiale. Cette manipulation permet aux atomes hydrogènes de s'ajuster par rapport aux atomes lourds de la protéine dont la position est plus précise. Suite à cette opération, l'énergie potentielle totale et la dérivée d'énergie maximale de la structure cristallographique de l'HADH sont respectivement de 10583 Kcal/mole et -5968. Elle n'est donc pas exempte de tensions internes.

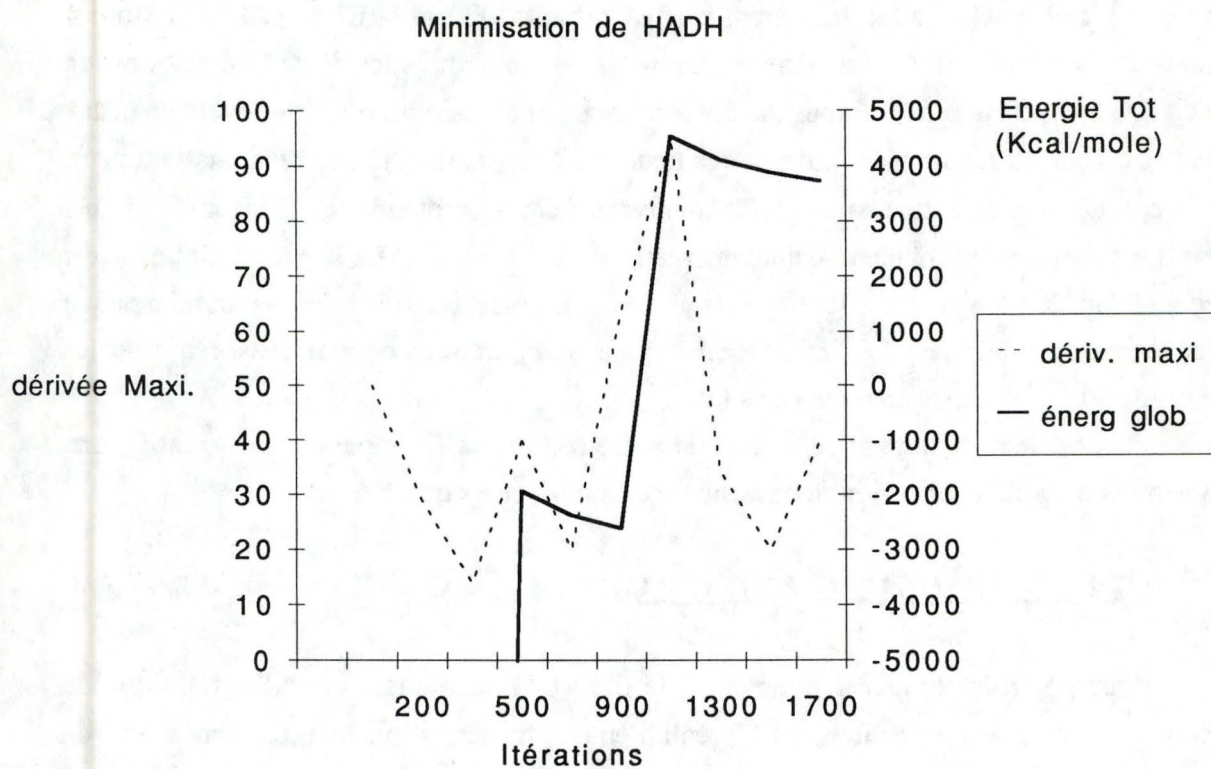


Figure 65

Evolution de l'énergie potentielle totale et de la dérivée d'énergie potentielle maximale du modèle cristallographique de l'alcool déshydrogénase de foie de cheval durant sa minimisation d'énergie.

La position des atomes de la chaîne principale est en général plus précise que celle des atomes des chaînes latérales. On réalise une série de 600 itérations en fixant les atomes de la chaîne principale dans leur position initiale de manière à ajuster uniquement les chaînes latérales. Les atomes de Zn^{++} ne pouvant être intégrés au modèle, les atomes auxquels ils sont coordonnés sont fixés durant toute la minimisation d'énergie. Enfin, les atomes de la chaîne principale sont libérés et l'énergie de l'holoenzyme est minimisée durant 800 itérations.

La variation de l'énergie potentielle globale et de la dérivée maximale d'énergie de l'HADH durant la minimisation sont données à la figure (65). Les étapes de la relaxation de la molécule font intervenir de plus en plus d'atomes, ce qui explique que l'énergie globale de la protéine augmente au cours des étapes successives. Quant à la dérivée maximale, elle semble osciller aux alentours de 20. La table (12) ci-dessous donne la liste des atomes possédant les 10 plus grandes dérivées d'énergie potentielle après 1800 itérations:

CARTESIAN DERIVATIVES OF ENERGY FUNCTION

```
=====
```

Atom Name	Residue Name	Molecule Derivative (kcal/mole.Å)			Name
		dv/dx	dv/dy	dv/dz	
-----	-----	-----	-----	-----	
N1E	NAD+	-33.088366	-7.052820	40.807262	
C2	NAD+	29.302978	4.422078	-36.318056	
C6	NAD+	25.397589	9.791085	-32.424504	
C3	NAD+	-18.842511	-4.269492	24.177657	
C5	NAD+	-14.617256	-9.174088	23.171866	
C4	NAD+	12.974658	4.881343	-18.155397	
C7	NAD+	4.828715	2.562041	-8.499873	
C8	NAD+	5.945627	0.668852	-7.371072	
H2	NAD+	-3.587073	-4.020728	7.523113	
H6	NAD+	-5.714652	1.533156	4.239820	

Tous ces atomes appartiennent une fois de plus au NAD^+ . Le dernier atome de cette liste, l' H6 du coenzyme, possède une dérivée de 5,7 en valeur absolue; les atomes de la protéine ont donc une dérivée inférieure à 5,7 ce qui est proche du niveau atteint dans le modèle d'YADH.

La superposition des deux structures de HADH, avant et après minimisation, donne un écart type des distances séparant les atomes superposés de 0,45 Å.

III.9. VALIDATION DU MODELE:

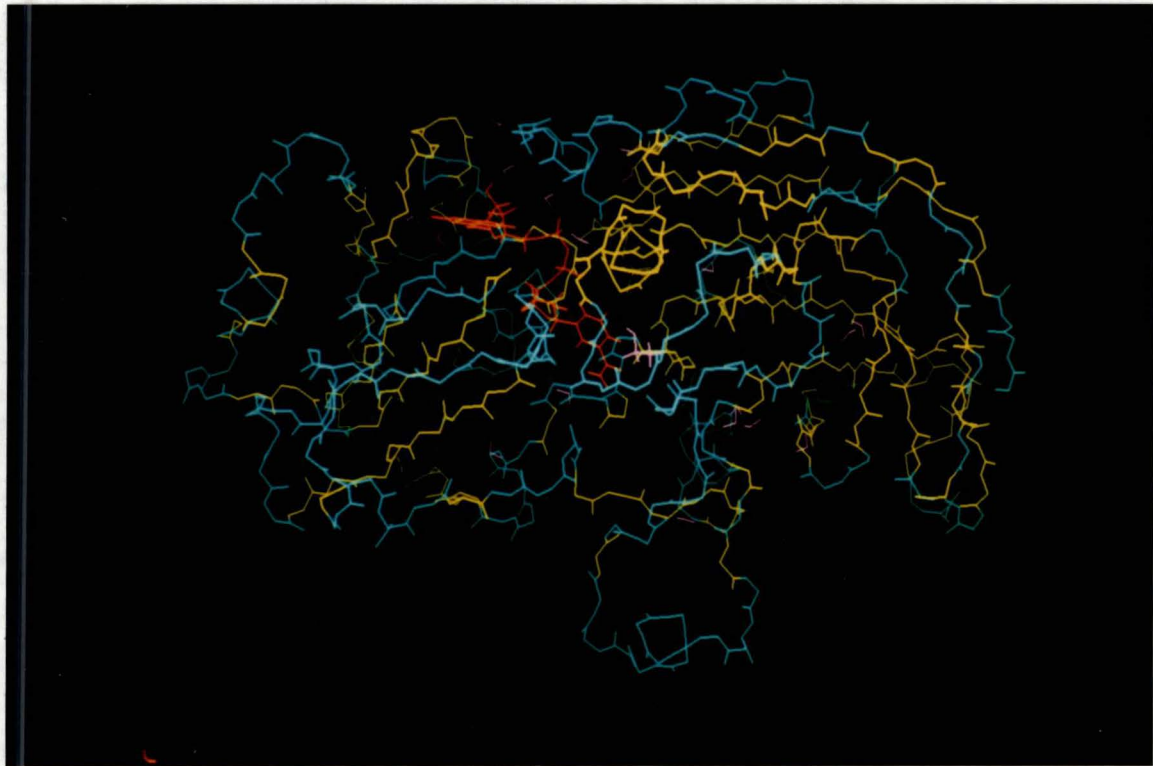


Figure 66

Les structures tridimensionnelles des deux modèles enzymatiques ont été superposées de manière à minimiser la somme des carrés des écarts entre les $C\alpha$ des deux enzymes. Les résidus de l'alcool déshydrogénase de levure dont les $C\alpha$ se situent à moins de $1,5\text{\AA}$ de ceux de l'alcool déshydrogénase de foie de cheval sont colorés en jaunes.

Vecteurs Radiaux d'Hydrophobicité

A

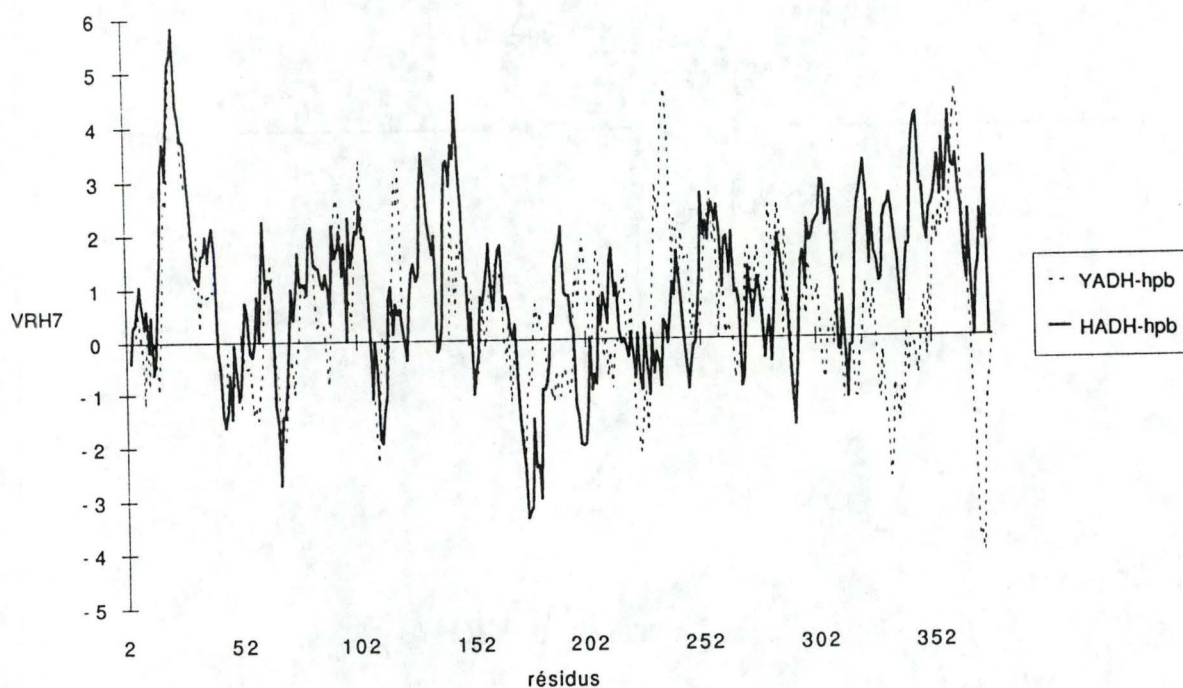


Figure 67

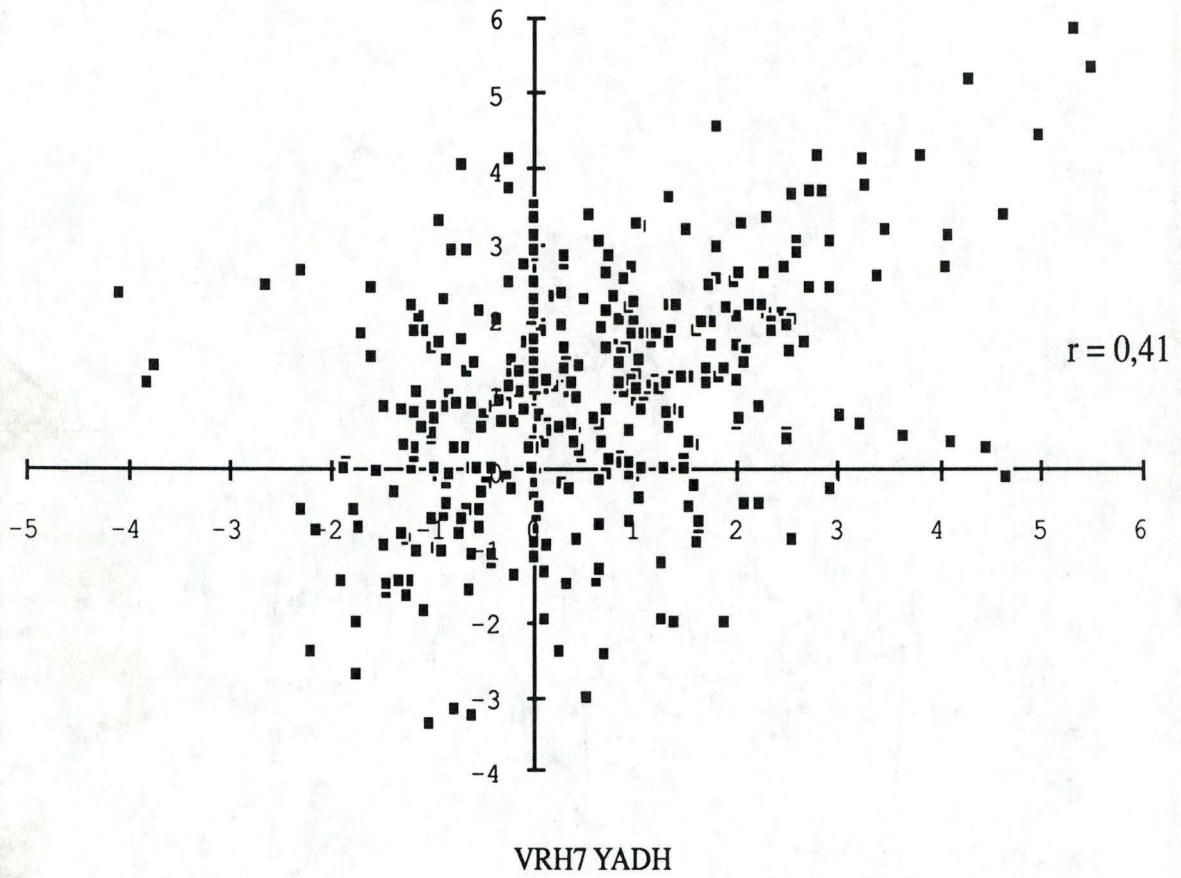
A) Superposition des profils des VRH7 des modèles d'alcool déshydrogénase de levure et de foie de cheval.

B) Corrélation entre les VRH7 des résidus des deux modèles enzymatiques selon l'alignement de séquence proposé par Jörnvall (1978).

C) Corrélation entre les VRH7 des résidus des deux modèles enzymatiques selon l'alignement tridimensionnel des deux structures.

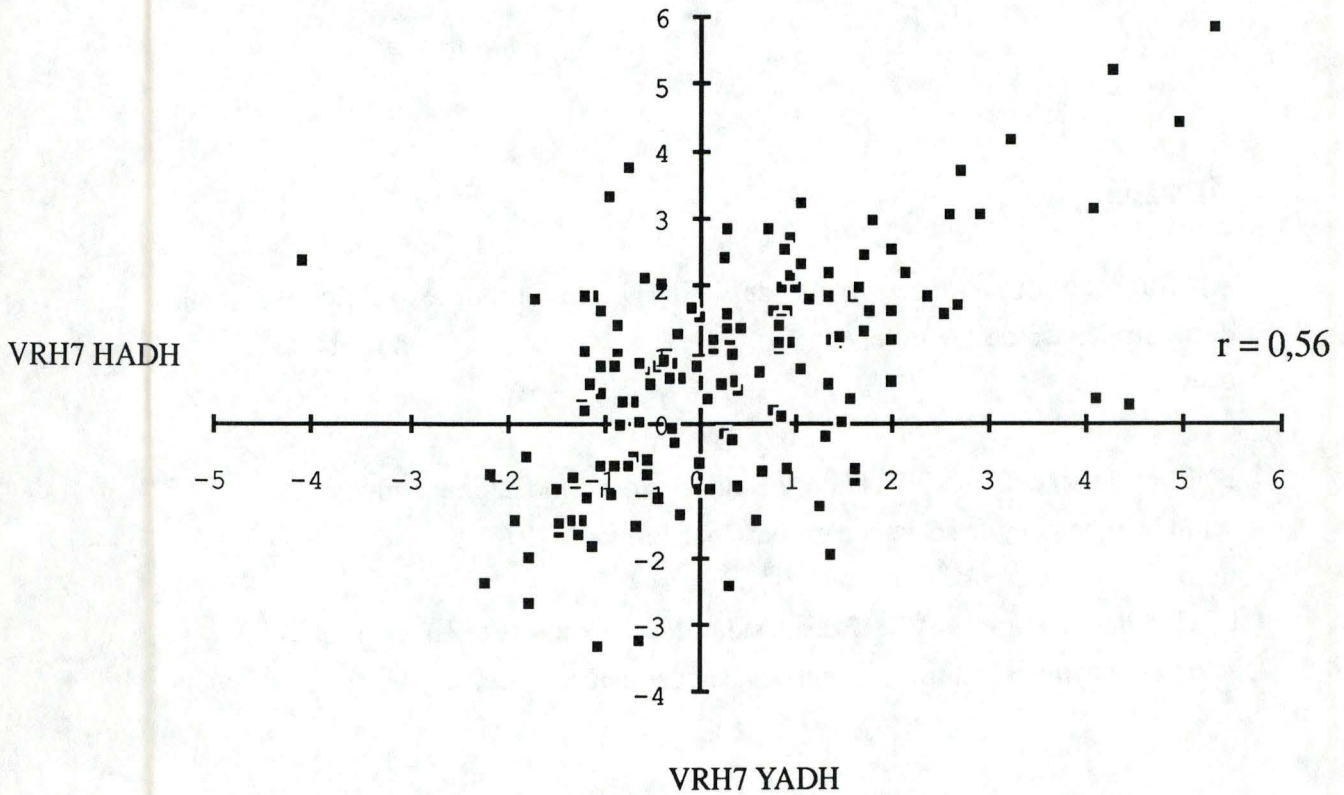
Corrélation VRH7

B



Corrélation VRH7

C



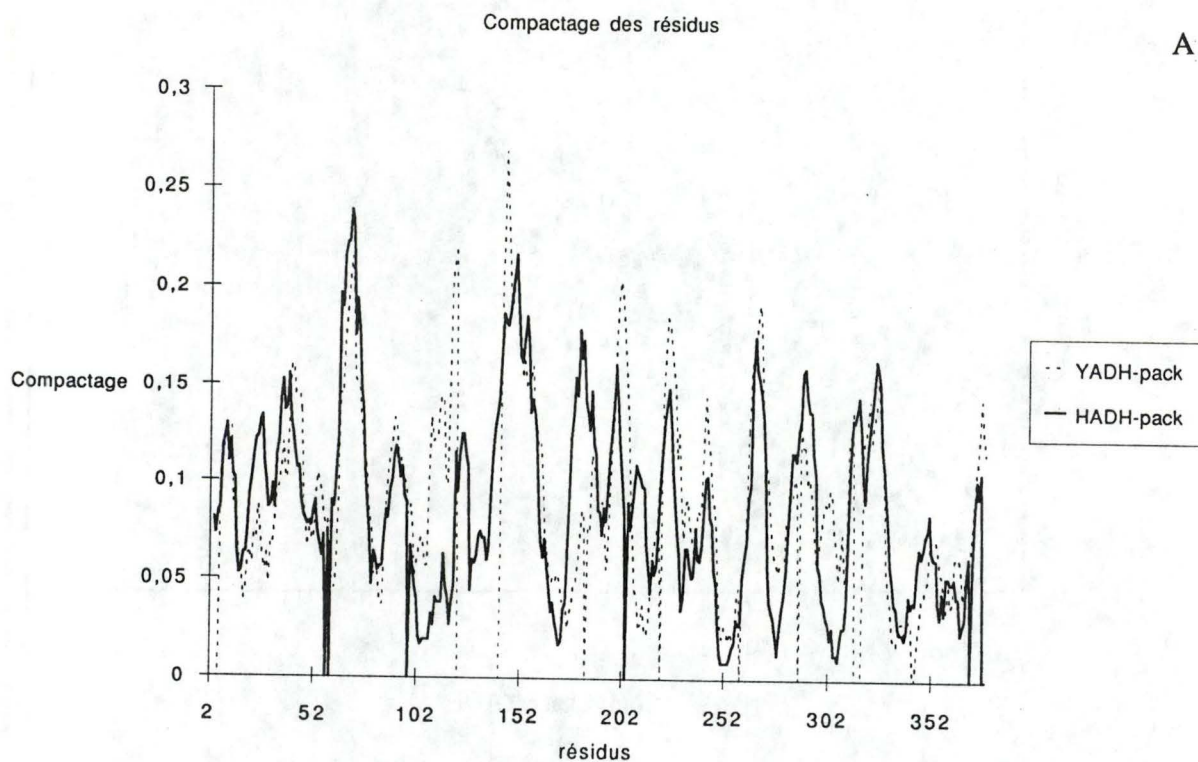


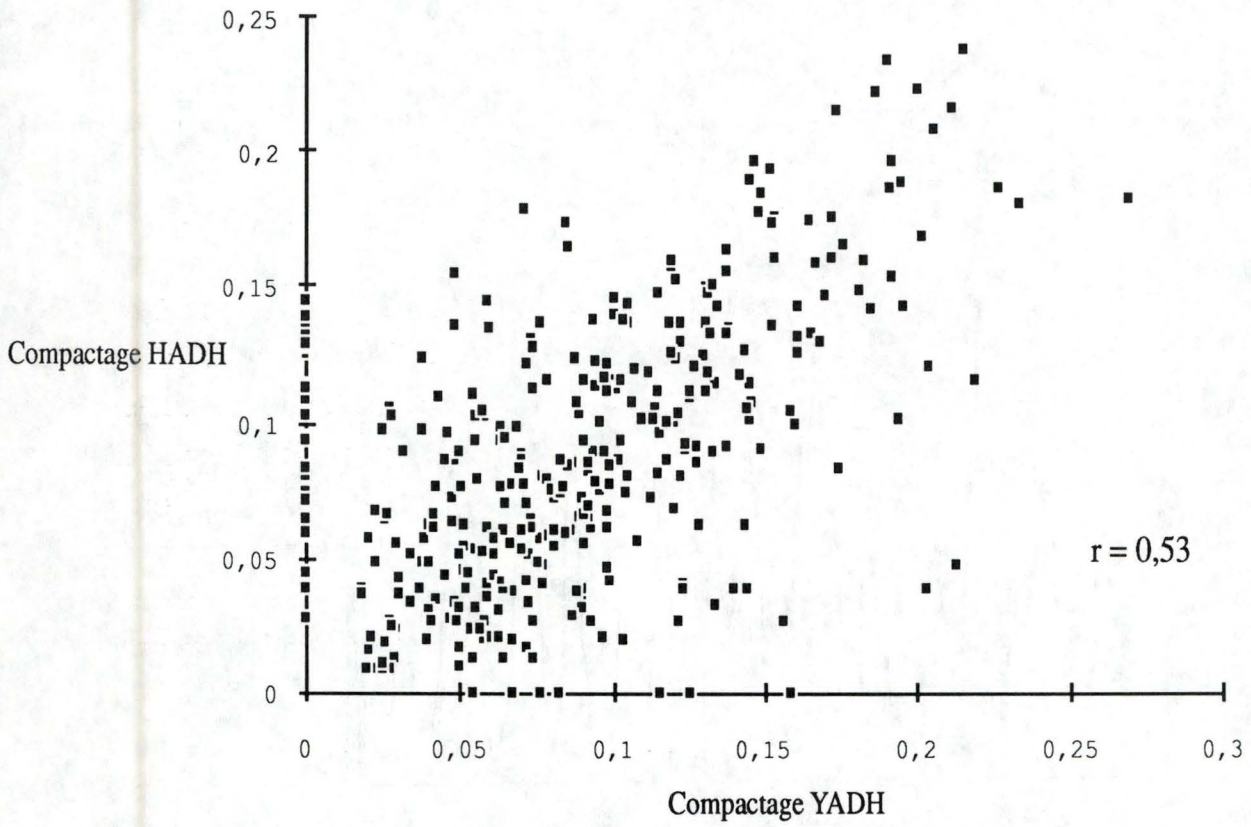
Figure 68

A) Superposition des profils de compactage des résidus des modèles d'alcool déshydrogénase de levure et de foie de cheval.

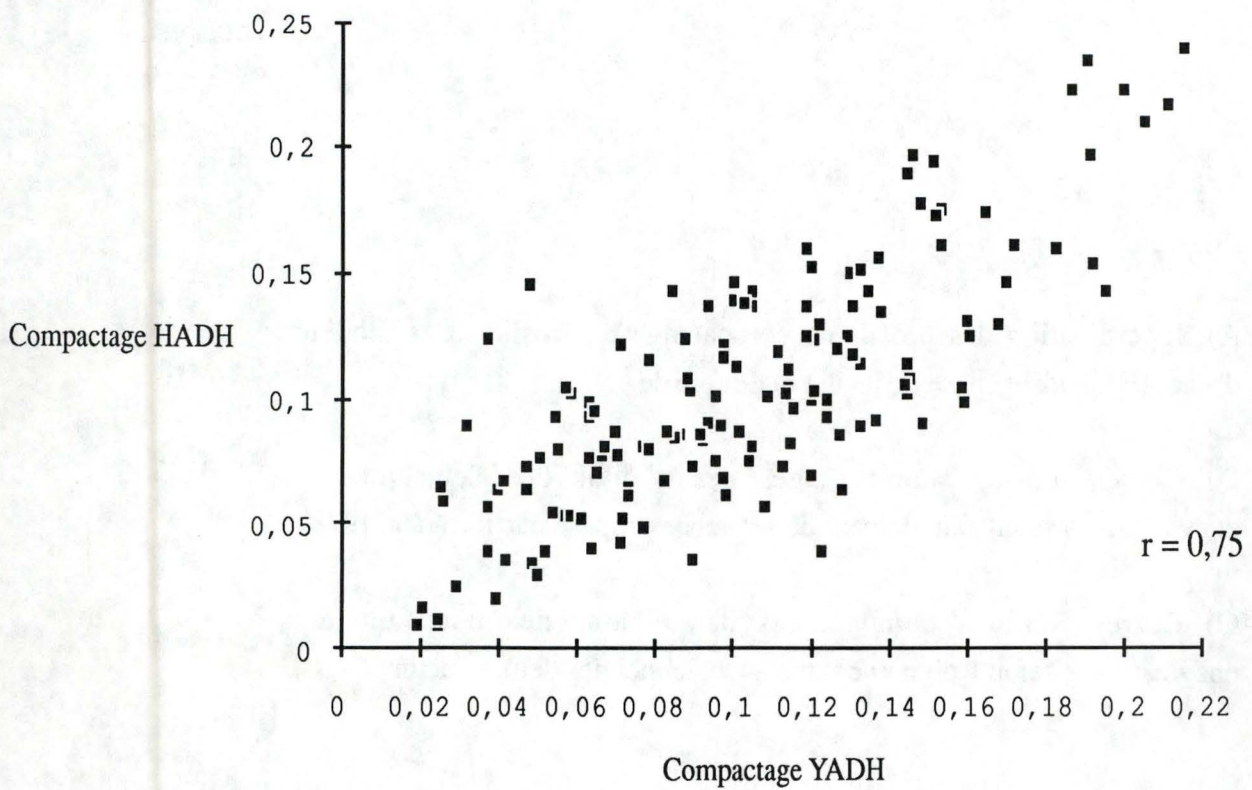
B) Corrélation des compactages des résidus des deux modèles enzymatiques selon l'alignement de séquence proposé par Jörnvall (1978)

C) Corrélation des compactages des résidus des deux modèles enzymatiques selon l'alignement tridimensionnel des deux structures.

Corrélation Compactage



Corrélation Compactage



III.9.1.CORRELATION ENTRE LES VECTEURS RADIAUX

D'HYDROPHOBICITE:

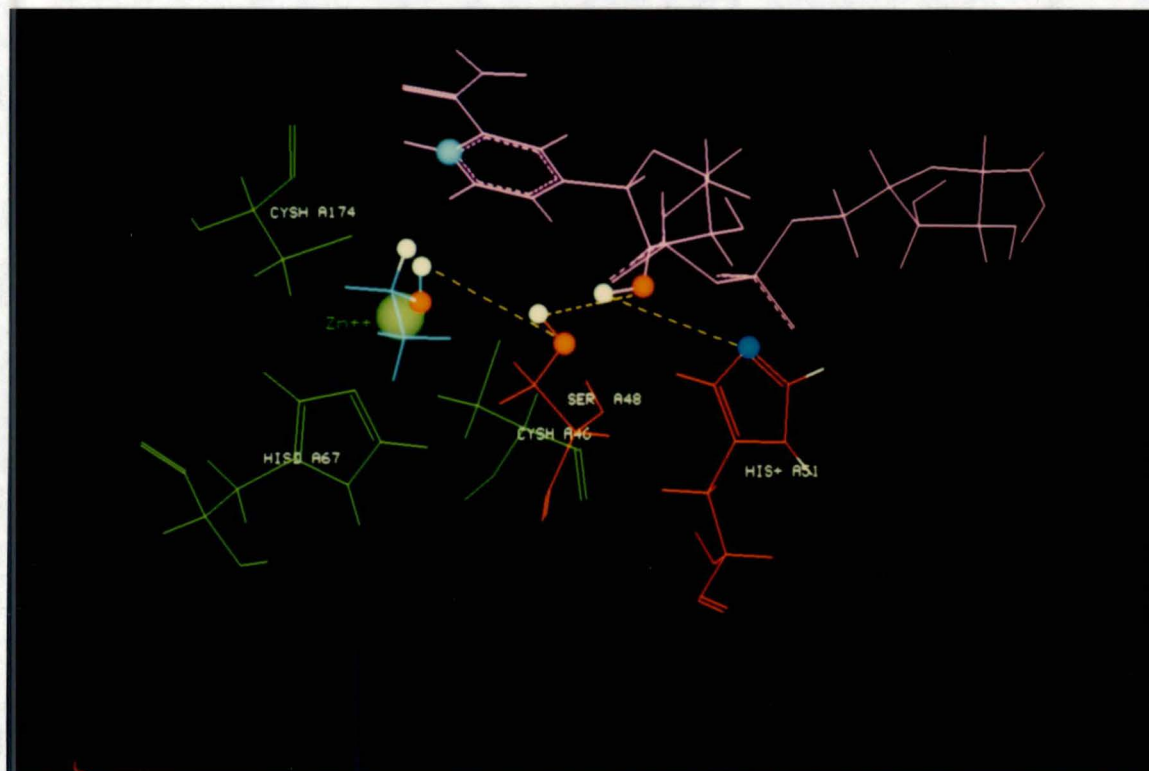
La corrélation (r) entre les vecteurs radiaux d'hydrophobicité (VRH7 ; Feytmans, 88) des deux modèles enzymatiques permet une appréciation quantitative de la validité du modèle de l'YADH. Un coefficient de corrélation significatif (r est significativement différent de zéro quand la probabilité "P" d'obtenir cette valeur, alors que la distribution des "r" est centrée sur zéro, est inférieure à 0,05) ne constitue pas une preuve de l'exactitude du modèle mais permet d'affirmer que la séquence d'une protéine n'est pas incompatible avec la structure proposée. Par contre un coefficient de corrélation non significatif (non significativement différent de zéro) permet d'affirmer que le modèle est incorrect.

Le coefficient de corrélation entre les VRH7 des modèles de YADH et HADH a été calculé sur base de l'alignement des séquences et sur base d'un alignement tridimensionnel des structures minimisées.

Dans l'alignement tridimensionnel, les deux modèles sont superposés de manière à minimiser la somme des carrés des écarts entre tous les atomes de leurs résidus identiques. Les VRH7 sont ensuite calculés et corrélés uniquement pour les résidus dont les C α sont distants de moins de 1,5 Å. 152 résidus ont été retenus selon cette méthode ce qui correspond à une similarité de 41% . La distribution de ces résidus dans les deux séquences est donnée à la figure (66).

La figure (67) montre les profils de distribution des VRH7 des deux modèles ainsi que la corrélation entre ces VRH7 selon les deux alignements. Quel que soit l'alignement, les coefficients de corrélation obtenus sont hautement significatifs ($P < 0.001$). Il est cependant plus élevé pour l'alignement tridimensionnel, ce qui indique que ces résidus en particulier ont une grande probabilité de se trouver dans une conformation adéquate. L'Arg 319, apparaît comme possédant des VRH7 opposés dans les deux séquences ce qui traduirait une mauvaise conformation. Cet aa présente effectivement une énergie d'angles de torsion élevée ($E_p \text{ torsion} = 11.7 \text{ Kcal/mole}$). Si l'on compare les angles de torsion de ce résidu dans les deux modèles table 13 (ci-après), seul l' angle ϕ est un peu plus élevé dans l'YADH.

A



B

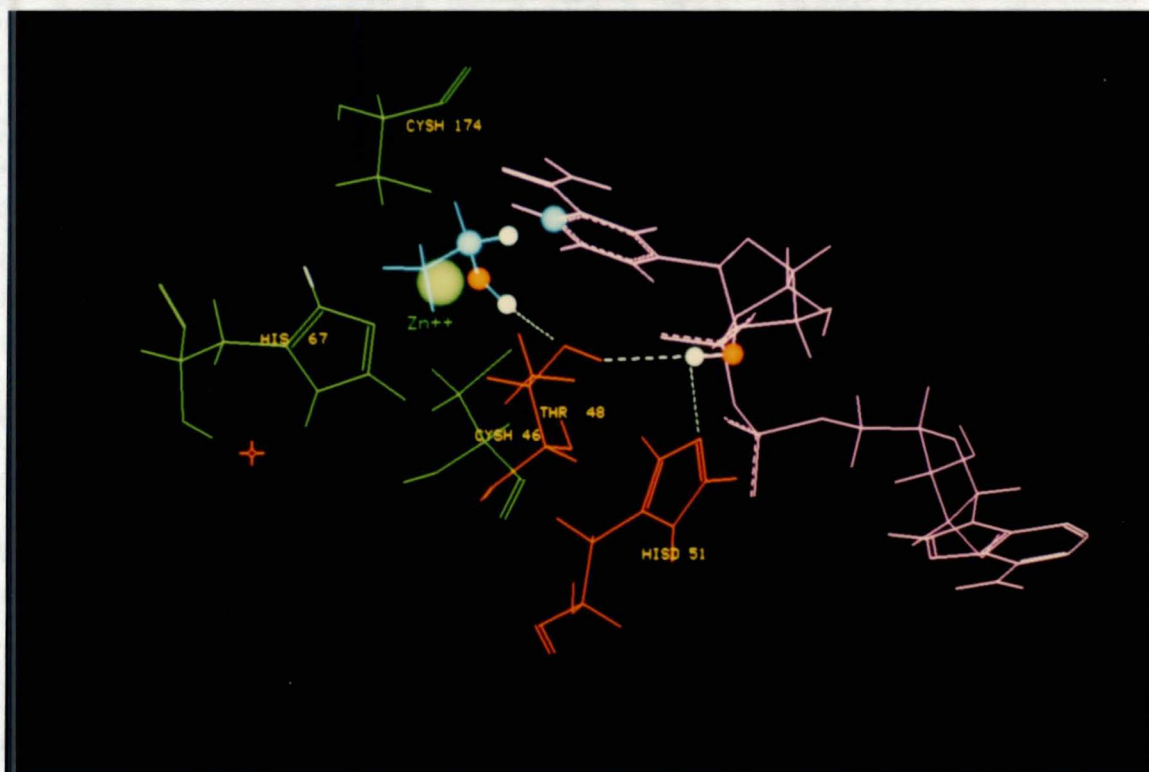


Figure 69

Vue des sites actifs des deux modèles enzymatiques (HADH en A et YADH en B) ainsi que des ponts hydrogènes supposés exister entre les résidus participant au mécanisme catalytique. Les trois résidus fixant le zinc et le zinc lui-même apparaissent en vert. Les liaisons coordinatives entre le zinc et les atomes auquel il est coordonné ne sont pas représentées. Le substrat (quatrième élément auquel le zinc est coordonné) est représenté en bleu clair. Le carbone 4 du coenzyme (violet) échangeant l'ion hydrure avec le substrat est représenté sous forme d'une sphère bleue clair. Les rayons des sphères sont proportionnels aux rayons atomiques réels.

Enzyme	ϕ	ψ	ω
HADH	-138.9	101.9	174.32
YADH	-168.27	107.78	175.74

Il est donc probable que les angles de torsions de cet aa dans l'HADH ne sont pas adaptés à la structure du modèle d'YADH. Une dynamique moléculaire leur permettrait probablement de passer les barrières énergétiques qui les maintiennent dans des conformations instables et d'adopter d'autres conformations énergétiquement plus favorables.

III.9.2 VERIFICATION DU COMPACTAGE:

Les profils de compactage de même que les coefficients de corrélation (fig 68) ont été évalués pour des valeurs calculées à partir d'une fenêtre de 7aa se déplaçant le long des séquences alignées comme précédemment. Une fois encore, les coefficients obtenus sont hautement significatifs ce qui traduit une occupation du volume interne du modèle très proche de la valeur maximale.

III.10 EXPLOITATION DU MODELE:

III.10.1. COMPARAISON DES SITES ACTIFS:

Les positions relatives des résidus participant au mécanisme catalytique sont très semblables dans les deux modèles (fig 69). Les distances entre les atomes formant le réseau de ponts hydrogène nécessaire au transfert de protons sont reprises dans la table (14) suivante:

atomes	distance YADH	distance HADH
éthanol: OH - Thr48: O	1.77 Å	3.28 Å
Thr /Ser 48: OH - NAD+ribose: O2D	3.38 Å	4.13 Å
NAD+ ribose: H2D - His51: N	3.10 Å	3.32 Å
<u>Thr48: OH - His51: N</u>	<u>4.38 Å</u>	<u>4.84 Å</u>

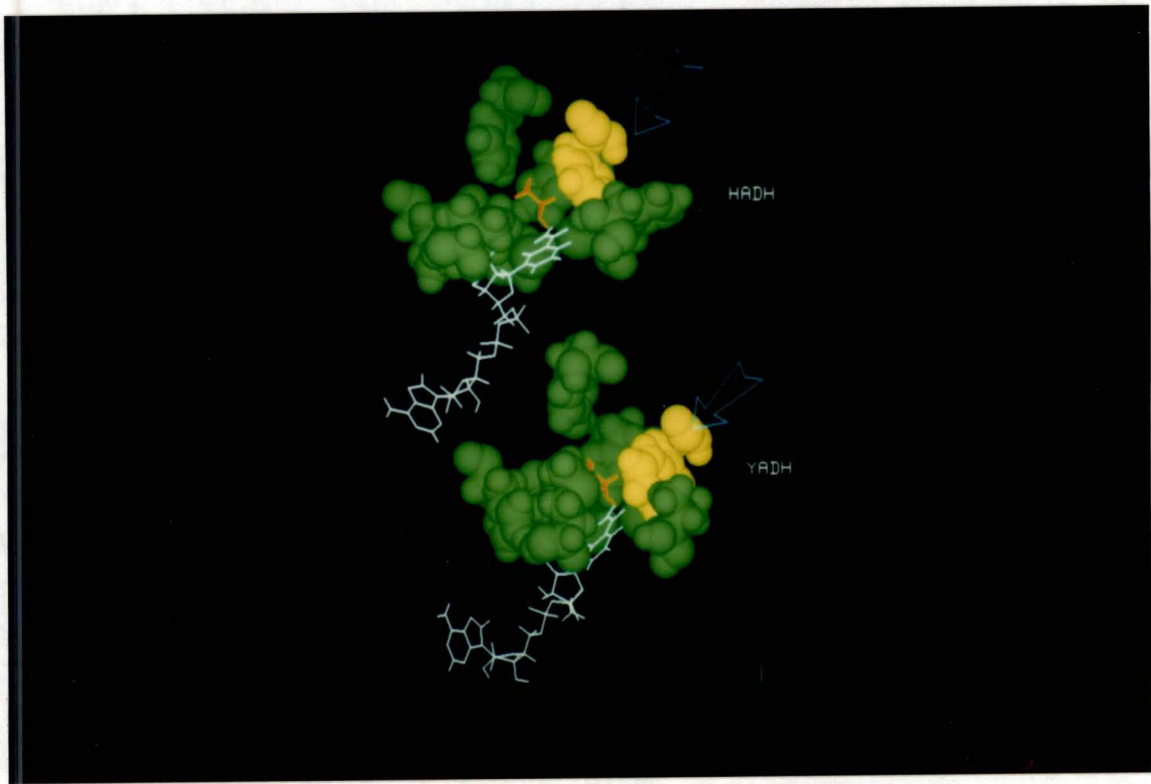


Figure 70

Illustration de la différence entre les positions de la Phe 93 (en jaune) dans les modèles des sites actifs de l'alcool déshydrogénase de levure (YADH) et de foie de cheval (HADH). le substrat apparaît en rouge au fond de sa poche de fixation et le coenzyme en blanc.

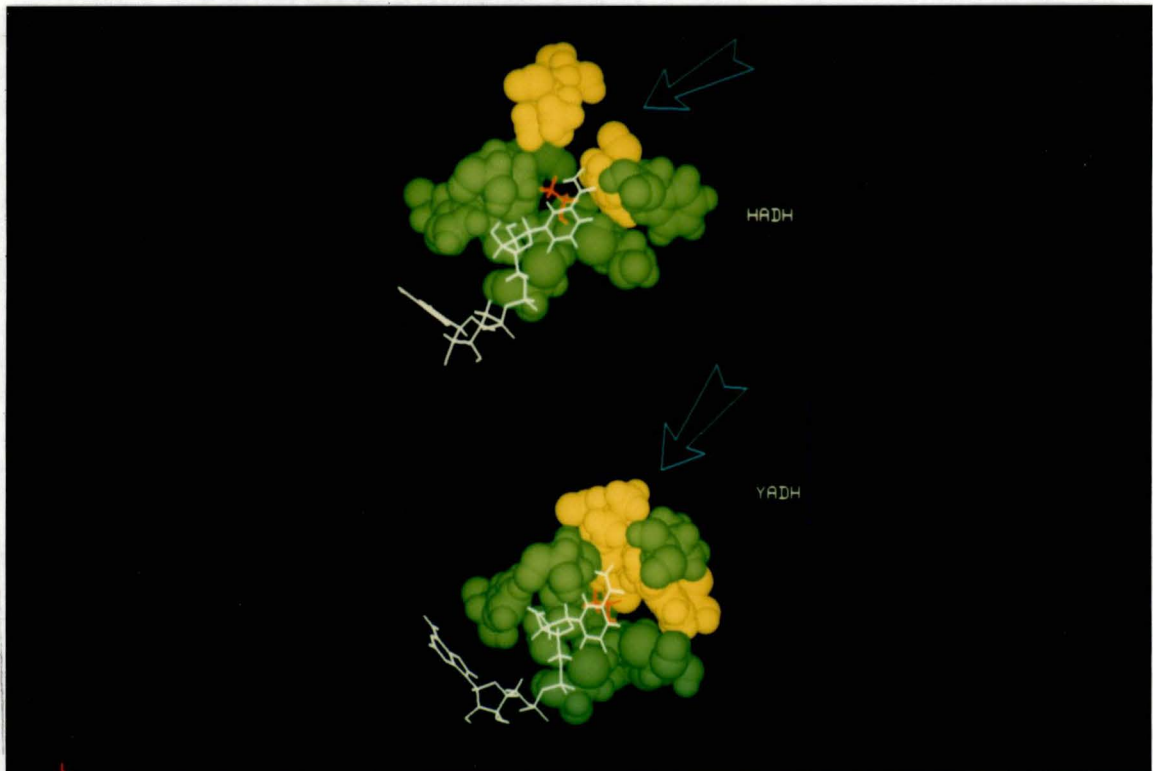


Figure 71

Illustration de la différence de position de l'Asp 115 dans les modèles des sites actifs de l'alcool déshydrogénase de levure (YADH) et de foie de cheval (HADH). Les modes de représentation sont identiques à ceux de la figure 70.

Selon Summers et Karplus (1989) la distance limite entre deux atomes s'échangeant un hydrogène est de 3,6Å. Les distances relevées dans l'YADH ne sont donc pas incompatibles avec les hypothèses expliquant le mécanisme catalytique. Le modèle d'HADH présente une trop grande distance entre la Ser 48 et le ribose du coenzyme pour qu'un ponts hydrogène ne se forme directement. Il convient malgré tout d'être prudent dans les hypothèses à en tirer car nous avons vu précédemment que la conformation du NAD⁺ est sujette à discussions (voir §II.6).

Il semble que l'oxygène du groupement hydroxyl de la Thr 48 soit trop éloigné de l'azote de l'His 51 pour qu'il puisse former un pont hydrogène. Il est probable que l'oxygène 2D du ribose associé au cycle nicotinamide (voir figure 47) joue le rôle d'intermédiaire dans la formation de liens hydrogène entre ces deux résidus. Une molécule d'eau pourrait également jouer ce rôle (rappelons que la structure cristallographique de l'holoenzyme n'inclut pas l'eau constitutive et que celle de notre modèle proviennent de la structure de l'apoenzyme cristallisée avec un analogue du NAD⁺).

La spécificité de l'YADH est plus stricte que celle de HADH (Jörnvall, 78). Selon cet auteur, ce phénomène est en partie dû au remplacement de la Phe 93 de l'HADH par un Trp chez l'YADH. Ce dernier étant plus volumineux, il interfère avec la fixation du substrat dans le site actif. Selon notre modèle, le résidu Trp 93 occupe effectivement un volume un peu plus important dans le site actif de l'YADH que dans celui de l'HADH (fig 70). Cependant, la modification la plus importante se situe au niveau des aa 115 et 116 dont la chaîne principale "referme" le site actif de l'YADH (fig 71). Dans l'HADH, le Ca de l'Asp 115 se situe à une distance de 10.40 Å du C1 de l'éthanol alors que dans l'YADH, cette distance n'est plus que de 5.60 Å! Il est peu probable qu'une minimisation plus poussée de l'enzyme de foie de cheval fasse apparaître le même phénomène surtout si l'on considère le niveau très bas déjà atteint par la dérivée maximale de cette dernière. Le rétrécissement du site actif qui découle de ce mouvement pourrait expliquer la spécificité plus stricte de l'YADH, d'autant que les résidus 115 et 116 bordent la voie d'accès du substrat au site actif (fig 72).

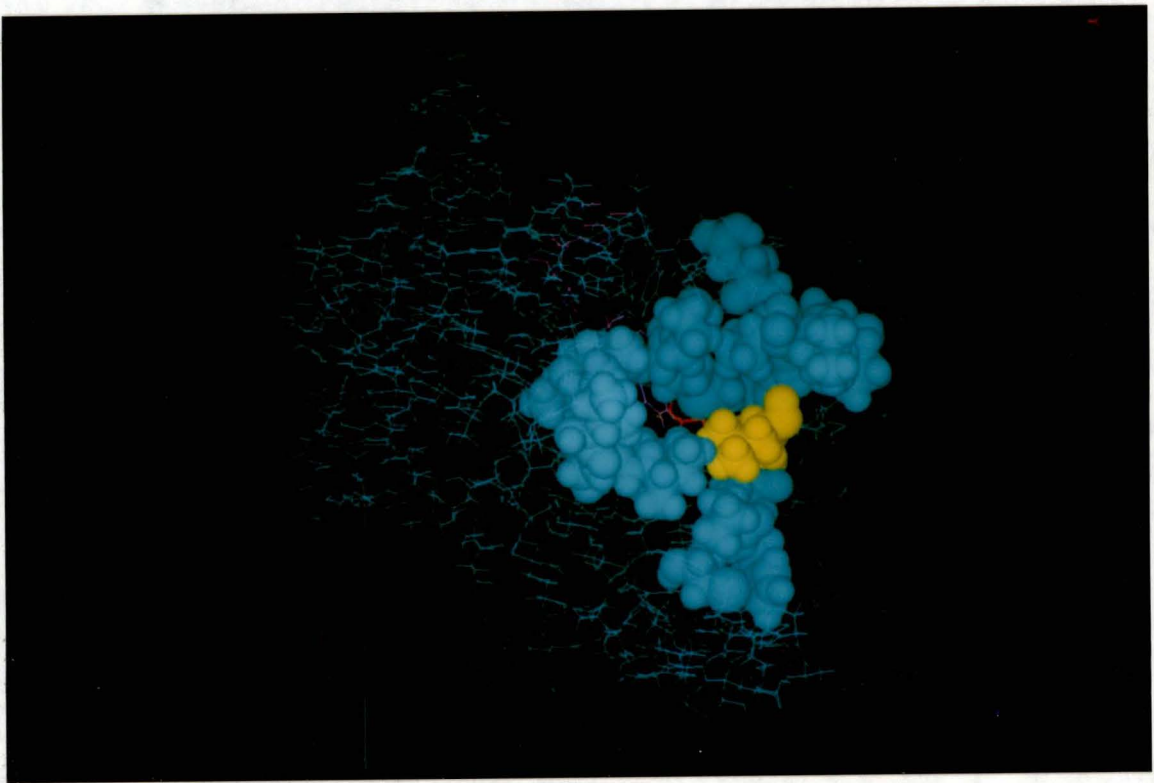


Figure 72

Représentation d'une partie de la surface du monomère d'alcool déshydrogénase de levure comprenant l'Asp 115 (en jaune).

L'ouverture de la poche de fixation du substrat apparaît clairement (le substrat est visible en rouge au fond de cette dernière). L'Asp 115 borde cette ouverture. Le coenzyme apparaît en violet sur cette photographie.



Figure 73

Modélisation des peptides obtenus après lyse trypsinique de l'alcool déshydrogénase de levure et possédant une activité. La chaîne principale de ces peptides est représentée sous forme de ruban. Le coenzyme et le substrat sont représentés respectivement en violet et en rouge. Les résidus catalytiques apparaissent en jaune.

III.10. 2. MODELISATION DU PEPTIDE OBTENU PAR LYSE TRYPSINIQUE:

Le peptide obtenu après lyse trypsinique de l'YADH et possédant une activité alcool déshydrogénase (P.Hayen, 89) à été modélisé en simulant un digestion de l'enzyme par la trypsine. Cet enzyme protéolytique rompt les chaînes peptidiques de manière spécifique au niveau des arginines et des lysines . D'autre part, les peptides isolés possèdent au moins un atome de zinc. Le fragment d'YADH retenu sur base de ces critères est illustré à la figure (73). Il est composé des peptides formés par les aa: Lys 41 à Thr 59, Lys 60 à Lys 81 et Lys 92 à Lys182. Il contient les 3 résidus fixant le zinc ainsi que la Thr 48 et l'His 51. Selon la configuration de la structure de la protéine native, ces trois chaînes peptidiques forment une poche pouvant accueillir le substrat et le coenzyme.

On ne peut évidemment pas affirmer que ce modèle correspond à la réalité puisque la lyse s'effectue en condition dénaturante pour l'enzyme. De plus, la perte d'une grande partie de la chaîne peptidique modifie totalement l'environnement des résidus du site actif (proximité du solvant, absence de certains aa,...) et donc sa configuration spatiale. Néanmoins, le peptide possède une activité alcool déshydrogénase, qui ne pourrait pas exister sans une disposition précise du coenzyme, du substrat , de l'ion Zinc et des aa responsables de la catalyse.

L'élaboration d'un modèle plus précis demanderait plus d'informations sur les propriétés catalytiques et structurales (orientation des chaînes peptidiques, activité vis à vis d'autres substrats, études cinétiques).

III.10. 3. STRUCTURE QUATERNAIRE:

L'YADH est une enzyme tétramérique. Très peu d'informations sont disponibles sur les caractéristiques de l'enzyme active ce qui rend l'élaboration d'un modèle très difficile. Comme nous l'avons vu précédemment, les tétramères sont en général, constitués de deux dimères associés de manière symétrique. Les zones d'association sont souvent constituées de plages hydrophobes circonscrites par des résidus chargés.

Un modèle dimérique de l'YADH a d'abord été construit sur base du dimère de l'HADH. Les résidus des monomères ont ensuite été colorés en fonction de leur hydrophobicité selon l'échelle consensus hydrophobicité d'Eisenberg (Eisenberg, 86). La figure (74) illustre le résultat obtenu. Une zone hydrophile de résidus chargés négativement entourant une surface plus hydrophobe apparaît moins prononcée dans l'HADH que dans l'YADH . Peut être s'agit-il là d'une zone d'association entre les dimères. Si cette hypothèse se vérifiait, il serait alors possible que la boucle formée par les aa 117 à 140 (en vert sur la figure) empêche les dimères de s'associer chez HADH. Une manière de tester cette hypothèse serait de remplacer un des aa hydrophobe de la zone en question par un aa hydrophile et de contrôler si cette mutation influence la formation des tétramères. En supposant que les dimères s'associent par ces zones, il resterait à expliquer pourquoi la forme dimérique est inactive et la forme tétramérique active.

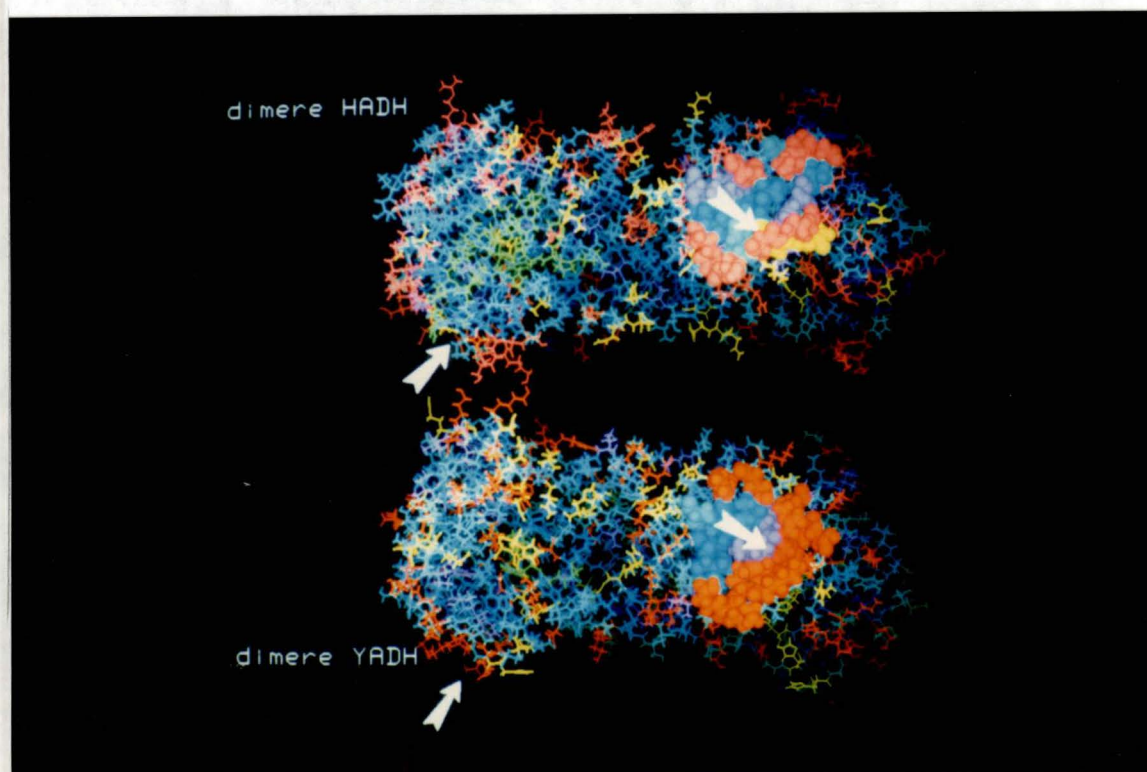


Figure 74

Essais de modélisation d'un dimère d'alcool déshydrogénase de levure. Le dimère d'alcool déshydrogénase de foie de cheval est également représenté par comparaison. Les résidus sont colorés selon leur hydrophobicité (les bleus sont hydrophobes et les rouges sont chargés). Les aa 117 à 141 sont colorés en vert. Les zones surfacées mettent en évidence des interfaces potentielles entre dimères. Les flèches blanches situent ces surfaces sur chaque monomère.

IV PERSPECTIVES ET CONCLUSIONS:

La prédiction d'une structure tertiaire à partir d'une séquence en aa n'étant pas envisageable actuellement, le modèle du monomère d'alcool déshydrogénase de *Saccharomices cerevisae* a été élaboré sur base de la structure d'une enzyme similaire: l'alcool déshydrogénase de foie de cheval. Nous ne remettons pas en question la validité de cette approche (CFR: § I.3.4), mais l'utilisateur d'un tel modèle ne doit jamais oublier que si il est très probable que deux séquences similaires adoptent des conformations globalement similaires, les positions relatives des atomes ne sont pas respectées . Un alignement des séquences de YADH et d'HADH à été réalisé de manière à maximiser leur similarité. L'alignement utilisé accorde une grande importance à la conservation des résidus participant au mécanisme catalytique chez HADH et peu d'importance à la taille des délétions et des insertions. La similarité entre les séquences de ces deux enzymes étant assez faible, il est peu rigoureux d'assimiler la structure native de l'enzyme de levure à la structure native de l'enzyme de foie de cheval. Pour tenter d'approcher la structure réelle de l'enzyme ou tout au moins de disposer d'une structure plus plausible, l'énergie du modèle a été minimisée jusqu'à obtenir une conformation stable. Cette conformation est influencée par les paramètres de la minimisation. Ces paramètres sont principalement l'absence de solvant et d'ions dans le milieu environant, le ph 7 et la minimisation de la protéine sous forme monomérique. Les interactions prises en compte dans le calcul de l'énergie potentielle de l'enzyme sont donc très particulières. La conformation obtenue après minimisation présente des différences assez importantes par rapport à la structure initiale. La superposition des modèles de HADH et YADH de manière à minimiser la somme des carrés des écarts entre les atomes des aa identiques montre un écart type de 2.3 Å entre les atomes des deux structures. Les variations au niveau du site actif consistent essentiellement en un rétrécissement de la poche de fixation du substrat chez YADH.

Quoi qu'il en soit, la structure présentée dans ce travail doit être considérée comme une approche d'une structure énergétiquement plausible de l'YADH mais certainement pas comme une image de la structure native. Dans l'état actuel de nos connaissances de la structure spatiale de l'YADH, la configuration des aa du site actif, l'absence de remplacement de résidus hydrophobes par des résidus chargés, la corrélation entre les profils d'hydrophobicité et le compactage des résidus de ce modèle semble indiquer que le modèle obtenu représente une structure tridimensionnelle vraisemblable. Comme tout modèle, il devra être corrélé à d'autres résultats expérimentaux pour pouvoir juger de sa validité et des améliorations à apporter.

Il ouvre pourtant des perspectives d'exploitations intéressantes. Des mutations de certains résidus peuvent être testées sur le modèle. Puisqu'il présente une structure vraisemblable, les conséquences d'une mutation donnée sur sa structure devraient être le reflet des conséquences de la même mutation sur la protéine native. Deux applications concrètes sont envisagées actuellement. Le clonage du gène de YADH a été réalisé par D. Prozzi (1988). La surproduction de l'enzyme mène à sa précipitation sous forme de corps d'inclusions et la solubilisation de ces corps d'inclusion résulte en une perte totale de l'activité enzymatique. Une application peut être de tenter de solubiliser l'enzyme en remplaçant un ou plusieurs aa hydrophobes de la surface de la protéine par un aa plus hydrophile. Cette technique a déjà été appliquée avec succès auparavant (Andrew *et al*, 1990). Ces mutations permettraient par ailleurs d'étudier les aa responsables de la tétramérisation de l'enzyme. Une autre exploitation intéressante serait de muter certains résidus du site actif de manière à augmenter l'activité de l'YADH vis à vis du méthanol. Ceci permettrait de régénérer du NADH à partir d'un substrat beaucoup moins coûteux que l'éthanol. L'approche consisterait à comparer le site actif du modèle de l'YADH avec celui d'une autre enzyme possédant une activité méthanol déhydrogénase.

IV. I CONCLUSIONS GENERALES:

Les modèles tridimensionnels présentent de nombreux avantages. Ils permettent tout d'abord une meilleure compréhension des relations entre la structure et la fonction d'une protéine au sein d'un organisme vivant. Jusqu'à présent, les aspects structuraux des protéines étant relativement mal connus, ils intervenaient peu dans l'élaboration des hypothèses proposées sur le rôle particulier de certains de leurs résidus. Les techniques du génie génétique offrent actuellement aux chercheurs la possibilité de tester les effets de mutations ponctuelles dans une protéine. Il est cependant très difficile de prévoir, *a priori*, les conséquences de la mutation d'un aa particulier car la conformation d'une protéine résulte d'un réseau complexe d'interactions entre ses différents atomes. Les méthodes de modélisation de structures tridimensionnelles permettent d'envisager des mutations rationnelles de la molécule, leur simulation et la prédiction des modifications conformationnelles qu'elles sont susceptibles de provoquer. Une autre application de la modélisation moléculaire combinée à la technique de mutagenèse dirigée, est la modification volontaire de la spécificité, voire de la fonction, d'une protéine. Les premières réalisations dans ce domaine sont très prometteuses même si elles sont encore d'application relativement limitée. Enfin, la simulation des interactions entre plusieurs molécules peut se révéler d'une grande utilité dans la conception de substances thérapeutiques spécifiques. Les perspectives dans ce domaine de la biologie sont donc multiples mais beaucoup de progrès sont encore nécessaires avant de pouvoir maîtriser pleinement "l'outil" moléculaire.

BIBLIOGRAPHIE

- Anfinsen C. B.; Haber E. (1961)
"The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain." *Proc natl. Acad. Sci. USA*, 47: 1309-1314
- Bille V. (1987)
"Développement d'une méthode d'immobilisation enzymatique avec modification du microenvironnement. Répercussions sur les propriétés d'un système enzymatique immobilisé."
Dissertation présentée pour l'obtention du grade de Docteur en Sciences, FUNDP, Namur.
- Birktoft J.; Blow D. M. (1972)
J. Mol. Biol., 68: 187-240
- Blundell T.L.; Sibanda B. L. ; Sternberg M. J. E.; Thornton J.M. (1987)
"Knowledge-based Prediction of Protein Structures and the Design of Novel Molecules." *Nature*, 326: 347-352.
- Bränden C. I.; Jörnvall H. ; Eklund H.; Furugen B. (1975)
"Alcohol Dehydrogenase." in "The Enzymes" Boyer *et al* eds. New York Academic Press. Vol 11 Chap 3: 103-190.
- Bränden C. I. ; Eklund H. (1980)
"Structure and Mechanism of Liver Alcohol Dehydrogenase." Jeffery J. ed., in "Dehydrogenase Requiring Nicotinamide Coenzymes." *Experientia Suppl* N°36, Basel, Birkhäuser-Verlag .
- Bühner M.; Sund H. (1969)
"Yeast Alcohol Dehydrogenase:-SH Groups, Disulphide Groups, Quaternary Structure, and Reactivation by Reductive Cleavage of Disulphide Groups." *Eur. J. Biochem.*, 11: 73-79
- Chotia C. (1974)
"Hydrophobic Bonding and Accessible Surface Area in Proteins." *Nature*, 248: 338-339
- Chou P. Fasman G. D. (1974)
"Prediction of Protein Conformation." *Biochem.*, 13: 222-245
- Connolly M. (1985)
"Atomic Size Packing Defects in Proteins" *Int. j. Peptide Protein Res.*, 28: 360-363
- Cook P. F.; Oppenheimer N. J. ; Cleland W. W. (1981)
"Secondary Deuterium and Nitrogen-15 Isotopes Effects in "Enzyme-Catalyzed Reaction. Chemical Mechanism of Liver Alcohol Dehydrogenase." *Biochem.*; 20: 1817-1825
- Creighton T. E. (1984)
"Proteins; Structure and Molecular Properties."(International Student Edition), New York, W. H. Freeman and Co

- Creighton T. E. (1985)
"The Problem of How Proteins Adopt Folded Conformations." *J. Phys. Chem.*, 89: 2452-2459
- Creighton T. H. (1989)
"Protein structure, a Practical Approach." , Oxford, Rickwood & Hames Eds. ; Oxford University Press, .
- Dayhoff M. O. (1972)
"Atlas of Protein Sequence and Structure." Natl. Biom. Res. Washington D.C.
- Devereux J. et al (1987)
Genetic Computer Group (GCG) Univ. of Wisconsin, Public.:*Nucl. Ac. Res.*, 12: 387-395
- Dickenson C.; Dickinson F. M. (1977)
"A study of the Ionic Properties of the Essential Histidine Residue of Yeast Alcohol Dehydrogenase in Complexes of the Enzyme with its Coenzymes and Substrates." *Biochem. J.*, 161: 73-82
- Eisenberg D. ;Weiss R. M. ; Terwilliger T. (1982)
"Hydrophobic Moments and Protein Structure." *Faraday Symp. Chem. Soc.*, 17: 109-120
- Eisenberg D.; Weiss R. M. ; Terwilliger T.; Wilcox W. (1984)
"The Hydrophobic Moment Detects Periodicity in Protein Hydrophobicity." *Proc. Natl. Acad. Sci. USA*, 81: 140-144
- Eklund H.; Nordström B.; Zeppezauer E.; Söderlund G. ; Ohlsson I.; Boiwe T.; Söderberg B. O.; Tapia O. ; Bränden C. I.n (1976)
"Three-dimensional Structure of Horse Liver Alcohol Dehydrogenase at 2.4 Å Resolution." *J. Mol. Biol.*, 102: 27-59
- Eklund H. ; Samama J.P.; Wallen L. Bränden C. I.; Åkeson Å.; Jones T. A. (1981)
"Structure of a Triclinic Ternary Complex of Horse Liver Alcohol Dehydrogenase at 2.9 Å Resolution." *J. Mol. Biol.*, 146: 561-587
- Eklund H.; Plapp B. V.; Bränden C. I. (1982)
"Binding of Substrate in a Ternary Complex of Horse Liver Alcohol Dehydrogenase." *J. Biol. Chem.*, 257: 14349-14358
- Eklund H.; Samama J. P. ; Jones A. (1984)
"Cristallographic Investigation of Nicotinamide Adenine Dinucleotide Binding to Horse Liver Dehydrogenase." *Biochem.*, 23: 5982-5996
- Fasman G. D. (1989 b)
" Practical Handbook of Biochemistry and Molecular Biology." Cleveland, CRC Press © 1989
- Feytmans E.; Bazan J. F. (1988)
"A three-dimensional Model of the CD4 N-terminal Domain Evaluated by a Simple Method using Vectors of Radial Hydrophobicity."
- Feytmans E.; Depiereux E. (1990)
"Use of Multivariate Profile of Similarities between Several Protein Sequences." 20 th Meeting of the FEBS, Budapest, August/1990.

- Fitch W. M.; Smith T. F. (1983)
"Optimal Sequence Alignment." Proc. Natl. Acad. Sci. USA, 80: 1382-1386
- Frühbeis H. Klein R. ; Wallmeier H. (1987)
"Computer-Assisted Molecular Design (CAMD)- An Overview." Angew. Chem. Int. Ed. Engl. 26: 403-418
- Ganzhorn A. J.; Plapp B. V. (1988)
"Carboxyl Groups near the Active Site Zinc Contributor to Catalysis in Yeast Alcohol Dehydrogenase." J. Biol. Chem., 263: 5446-5454
- Ganzhorn A. J.; Green D. W. ; Hershey A. D.; Gould R. M.; Plapp B. V. (1987)
"Kinetic Characterisation of Yeast Alcohol Dehydrogenase." J. Biol. Chem., 262: 3754-3761
- Greer J. (1980)
"Model for Haptoglobin Heavy Chain Based upon Structural Homology." Proc. Natl. Acad. Sci. USA, 77: 3393-3397
- Greer J. (1981)
"Comparative Model-building of the Mammalian Serine Proteases." J. Mol. Biol., 153: 1027-1042
- Gribskov M.; Mc Lachlan A. D.; Eisenberg D. (1987)
"Profile analysis: Detection of Distantly Related Proteins." Proc. Natl. Acad. Sci. USA, 84: 4355-4358
- Hagler A. T.; Lifson S. *et al* (1979)
"Consistent Valence Force Field Studies of Intermolecular Forces in Hydrogen Bonded Crystals. A Benchmark for the Objective Comparison of Alternative Force Fields;" J. Am. Chem. Soc., 101: 813-819
- Hayen P. (1989)
"Recherche d'une Activité Alcool Déshydrogenase à partir de Fragments Peptidiques d'une Enzyme Native." Mémoire présentée en vue de l'obtention du grade de Licencié en Sciences. FUNDP, Namur.
- Hendrickson W. A. (1987)
"X- Ray diffraction." in "Protein Engineering" © 1987 Oxender D.L.; Fox C.F. ed; New York, Alan R. Liss.
- Horjales E. ; Branden C. I. (1985)
"Docking of Cyclohexanol- derivatives into the Active Site of Liver Alcohol dehydrogenase." J. Biol. Chem., 260: 15445-15451
- Jaenicke R. (1987)
"Folding and Association of Proteins." Prog. Biophys. Molec., 49: 117-155
- Johnson M. S.; Doolittle R. F. (1986)
"A Method for the Simultaneous of Three or More Amino Acid Sequences." J. Mol. Evol., 23: 267-278
- Jörnvall H. (1977a)
"The Primary Structure of Yeast Alcohol Dehydrogenase." Eur. J. Biochem., 72: 425-442
- Jörnvall H. (1977 b)
"Differences between Alcohol Dehydrogenases." Eur. J. Biochem., 72: 443-452

- Jörnvall H. Eklund H. ; Bränden C.I. (1978)
"Subunit Conformation of Yeast Alcohol Dehydrogenase." *J. Biol. Chem.*, 253: 8414-8419
- Jörnvall H.; Persson B. ; Jeffery J. (1987)
"Characteristics of Alcohol/ Polyol Dehydrogenases." *Eur. J. Biochem.*, 167: 195-201
- Kaptein B. (1989)
"Synthetic Models for the Active Site of Alcohol Dehydrogenase." *Diss. Doct. Wisk. en Natuurwetenschap, Univ of Groningen, Van Denderen ed.*,
- Klinman J. P. (1972)
"The Mechanism of Enzyme-catalyzed Reduced Nicotiamide Adenine Dinucleotide-dependent Reductions." *J. Biol. Chem.*, 247: 7977-7987
- Klinman J. P. (1975)
" Acid Base Catalysis in the Yeast Alcohol Dehydrogenase Reaction." *J. Biol. Chem*, 250: 2569-2573
- Klinman J. P. (1976)
" Isotope Effects and Structure-Reactivity Correlations in the Yeast Alcohol Dehydrogenase Reaction." *Biochem.*, 15: 2018-2026
- Leskovac V. ; Parkov-Pericin D. (1975)
"Evidence for a Histidine and a Cysteine Residue in the Substrate-Binding Site of Yeast Alcohol Dehydrogenase." *Biochem. J.*, 145: 581-590
- Levitt M. (1978)
"Conformational Preferences of Amino-acids in Globular Proteins." *Biochem*, 17: 4277-4285
- Linderström K. U.; Schellman J.P. (1959)
" Protein Structure and Enzyme Activity." in "THE ENZYMES", Boyer, 2nd Ed, vol1 :443-510, New York, Academic Press.
- Lim V. I. (1974)
"Algorithm for Prediction of α -helical and β -Structural Regions in Globular Proteins." *J. Mol. Biol.*, 88: 873-894
- Maizel J. V.; Lenk R. P. (1981)
"Enhanced Graphic Matrix Analysis of Nucleic Acid and Protein Sequences." *Proc. Natl. Acad. Sci. USA*, 78: 7665-7669
- Mc Cammon J. A. (1984)
" Protein Dynamics." *Rep. Prog. Phys.*, 47: 1-46
- Mc Cammon A. J.; Harvey S. (1988)
"Dynamics of Proteins and Nucleic Acids." © 1987 Cambridge University Press; Cambridge
- Miller S. (1989)
"The Structure of Interface between Subunits of Dimeric and Tetrameric Proteins." *Prot. Eng.*, 3: 77-83

- Montelione G. T.; Scheraga H. A. (1989)
"Formation of Local Structures in Protein Folding." *Acc. Chem. Res.*, 22: 70-76
- Murali C.; Creaser E.H. (1986)
"Protein Engineering of Alcohol Dehydrogenase- 1. Effects of two amino-acid changes in the active site of yeast ADH-1." *Prot. Eng.*, 1: 55-57
- Murata M. Richardson J. S. (1985)
"Simultaneous Comparison of Three Protein sequences." *Proc. Natl. Acad. Sci. USA*, 82: 3073-3077
- Needleman S. B.; Wunsch C. D. (1970)
"A general Method Applicable to the Search for Similarities in the Amino Acid sequence of Two Proteins." *J. Mol. Biol.*, 48: 443-453
- Nozaki Y. ; Tanford C. (1971)
"The Solubility of Amino-acids and Two Glycine Peptides in Aqueous Ethanol and Dioxane Solutions." *J. Mol. Biol.*, 246: 2211-2217
- Pakula A. A.; Sauer R. T. (1990)
"Reverse Hydrophobic Effects Relieved by Amino-acid Substitution at protein Surface." *Nature*, 344: 363-364
- Plapp B.V.; Ganzhorn A.J. ; Gould R. M.; Green D. W. ; Hershey A.D. (1987)
"Structures and Fonction in Yeast Alcohol Dehydrogenase." *Enzymology and Molecular Biology of Carbonyl Metabolism.*: p 227-236; New York, Alan R. Liss © 1987.
- Picworth J. ; True blood K. (1985)
" Crystal Structure Analysis." 2nd Ed. Oxford Univ Press.
- Prozzi D. (1989)
" Bio/ 19 Rapport 1989, Sous-projet 1." (non publié)
- Ptitsyn O. B.; Finkelstein A. V. (1983)
"Theory of Protein Secondary Structure and Algorithm of its Prediction." *Biopol.*, 22: 15-25
- Ramachandran *et al* (1963)
" Stereochemistry of Polypeptide Chain Configuration." *J. Mol. Biol.*, 7: 55-59
- Richards F.M. (1974)
"The Interpretation of Protein Structures: Total Volume, group Volume distributions and Packing Density." *J. Mol. Biol.*, 82: 1-14
- Rose G. D. ; Dworkin J. (1989)
"The Hydrophobicity Profile." Fasman G. D. New York, Plenum Press; © 1989
- Sanger F.; Tuppy H. (1951)
"The Amino Acid Sequence in the Phenylalanyl Chain of Insulin." *Biochem. J.*, 53: 463-490
- Schellenberg K. A. (1977)
" Mechanism of Hybride Transfert." in "Pyridine Nucleotide Dependent Dehydrogenase." Sund H., ed; New York, Springer- Verlag.

- Schimpfessel L. (1967)
"Présence et Régulation de la Synthèse de Deux Alcools Déshydrogénases chez la Levure *Saccharomyces cerevisiae*." *Biochim. Biophys. Acta.*, 151: 317-329
- Schneider G.; Eklund H.; Zeppezauer E. C.; Zeppezauer M. (1983)
"Crystal Structures of the Active Site in Specifically Metal-depleted and Cobalt-substituted Horse Liver Alcohol Dehydrogenase Derivatives." *Proc. Natl. Acad. Sci. USA*, 80: 5289-5293
- Schulz G. E. (1977)
"Structural Rules for Globular Proteins." *Angew. Chem. Int. Ed. Engl.*, 16: 23-32
- Schulz G. E.; Schirmer R. H. (1985)
"Principles of Protein Structure." 7^{ème} édition, Charles Cantor ed. ; New York, Springer Advanced Texts in Chemistry
- Smith T. F. ; Waterman M. S. (1981)
"Comparison of Biosequences." *Adv. Appl. Math.*, 2: 482-489
- Stryer L. (1988)
"Biochemistry" 3d Ed. W. H. New York, Freeman and Co
- Subbiah S. et Harrison S. C. (1989)
"A Method for Multiple Sequence Alignment with Gaps." *J. Mol. Biol.*, 209: 539-548
- Temler; Kägi (1973) Abs. of Int. Cong. of Biochem. p 109
- Vasques M. Scheraga H.A. (1985)
"Use of Buildup and Energy Minimization Procedures to Compute Low-energy Structures
- Von Heijne G. (1989)
"Sequence Analysis in Molecular Biology, Treasure Trove or Trivial Pursuit?"
- Weiner S. J. ; Kollman P.A. ; Case D. ; Singh CH.; Ghio C.; Alagona G. ; Profeta S.; Weiner P. (1984)
"A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins." *J. Am. Chem. Soc.*, 106: 765-784
- Wilbur W. J.; Lipman D. (1983)
"Rapid Similarity Searches of Nucleic-acid and Protein Data Bank." *Proc. Natl. Acad. Sci. USA*, 80: 726-730
- Williamson V.M. ; Paquin CH. E. (1987)
"Homology of *Saccharomyces cerevisiae* ADH4 to iron-activated alcohol dehydrogenase from *Zymomonas mobilis* " *Mol. Gen. Genet.*, 209: 374-381
- Wills C. ; Jörnvall H. (1979)
"The Two Major Isozymes of Yeast Alcohol Dehydrogenase." *Eur. J. Biochem.*, 99: 323-331
- Wolfenden R. V.; Cullis P. M. (1979)
"Water, Protein Folding and Genetic Code." *Science*, 206: 575-577