

# Exploration des mécanismes de repliement des protéines par dynamique moléculaire

Bernard GILQUIN

DSV/DIEP/LSP

CEN Saclay

91191 GIF sur YVETTE

## Abstract

Comment se replient les protéines ? Cette question est ancienne. En introduction nous rappellerons ce qu'est le paradoxe de Levinthal et comment on est passé de la notion de chemin de repliement à la notion de paysage énergétique. Les simulations de dynamique moléculaire ont permis d'aborder la compréhension du processus de repliement au niveau atomique. Cependant l'échelle de temps des processus de repliement ( de l'ordre de la milliseconde) n'est pas accessible aux simulations numériques (de l'ordre de la nanoseconde). Plusieurs auteurs ont donc proposé de simuler le dépliement des protéines par dynamique moléculaire. En admettant le principe de micro-réversibilité l'étude du processus de dépliement renseigne sur celui de repliement. Cependant, il est nécessaire d'accélérer le dépliement en introduisant un biais afin que les états dépliés soient accessibles aux échelles de temps des simulations. Nous présenterons un exemple de ce qui a été réalisé dans le cas de l'étude de protéines de petite taille suivant un repliement simple, globalement à deux états. Nous présenterons ensuite ce que nous avons réalisé dans le cas d'une protéine de taille plus importante et pour laquelle le processus de repliement est plus complexe car il existe un intermédiaire transitoire de repliement. C'est le cas du lysozyme pour lequel les simulations de dépliement permettent d'accéder au mécanisme atomique de repliement et de comprendre pourquoi des mutants de cette protéine se replient plus lentement et forment des fibres amyloïdiques. Ainsi les intermédiaires de repliement seraient à l'origine de formes pathogènes des protéines observées dans les maladies neuro-dégénératives. Enfin nous montrerons comment à partir de plusieurs simulations longues de dynamique moléculaire, le paysage énergétique pour de petites protéines peut être calculé.

## Introduction

Comment les protéines se replient-elles ? Cette question est ancienne puisqu'elle se pose depuis qu'Anfinsen, dans les années 60 a montré que le repliement des protéines est réversible et n'est fonction que de la nature de la séquence primaire. De ce point de vue le repliement apparaît comme un processus spontané et rapide. Levinthal a fait remarquer que le nombre de conformations accessibles par une protéine est très grand (de l'ordre de  $10^{30}$ ) et qu'une recherche aléatoire ne permet pas de trouver la forme native dans des temps de l'ordre de grandeur des temps de repliements (sec-mn). Pour résoudre ce paradoxe dit de Levinthal on pouvait penser que durant son repliement la protéine adopte une suite de conformations ou d'états intermédiaires constituant le chemin de repliement (Figure 1b). Ces intermédiaires pouvant être aussi bien des intermédiaires productifs, le long du chemin, que non productifs, des impasses. A partir des années 95, cette vision du chemin de repliement a été préférée à une autre vue dite « new view ». Sont alors introduites les notions de paysage énergétique ou d'entonnoir de repliement ( Figure 1). L'axe vertical représente l'énergie interne de la chaîne polypeptidique. Les axes horizontaux représentent les conformations de la chaîne. L'extension le long de ses axes indique le nombre de conformations accessibles. Pour l'état natif ce nombre est réduit ( le bas de l'entonnoir) et plus les états sont dépliés plus le nombre d'états conformationnels accessibles est grand et l'entropie conformationnelle élevée. Alors le paradoxe de Levinthal apparaît alors comme une question mal posée. Le paradoxe de Levinthal conçoit le repliement comme une recherche conformationnelle dans un paysage énergétique plat (Figure 1a). Le paysage énergétique peut être plus ou moins complexe. Un

paysage énergétique complexe conduira à la présence de multiples chemins de repliement et à une cinétique de repliement complexe (Figure 1d).

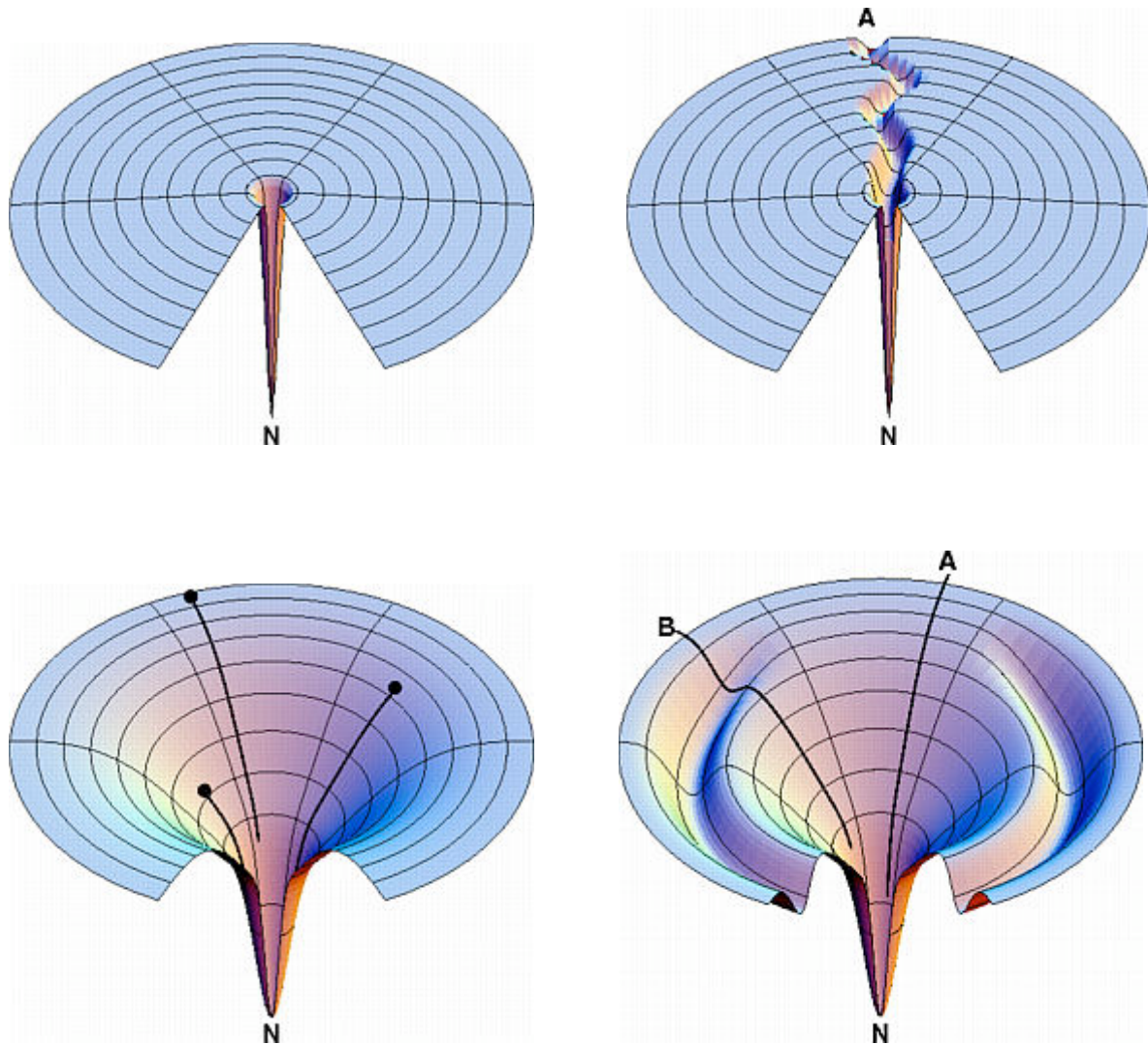


Figure 1 a) Paysage énergétique plat de type trou d'un terrain de golf correspondant au paradoxe de Levinthal b) paysage énergétique présentant un chemin de repliement de A vers l'état natif N. c) paysage énergétique idéal qui guide les conformations vers la conformation native N d) paysage énergétique pour lequel il existe un repliement rapide de A vers N et un processus lent de B vers N avec un état intermédiaire transitoire ( d'après (1) )

### Etude du repliement des protéines par des simulation de dépliement

#### **Mécanismes de repliement simple.**

Il importe donc d'étudier aussi bien des systèmes simples que des protéines dont le repliement est plus complexe afin comprendre les différents mécanismes de repliement au niveau atomique. Plusieurs modèles ont été proposés. Selon le premier modèle, le « framework model » ou le model « diffusion-collision » les structure secondaires se forment en premier avant de s'assembler pour former la structure native (Figure 2). A l'opposé, le modèle du « collapse hydrophobe », privilégie l'assemblage des résidus hydrophobes pour former le cœur de la protéine suivi de la structuration des éléments de structures secondaires.

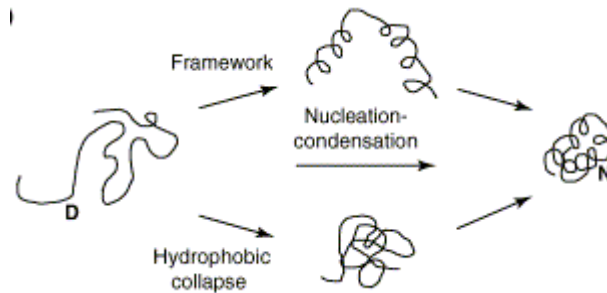
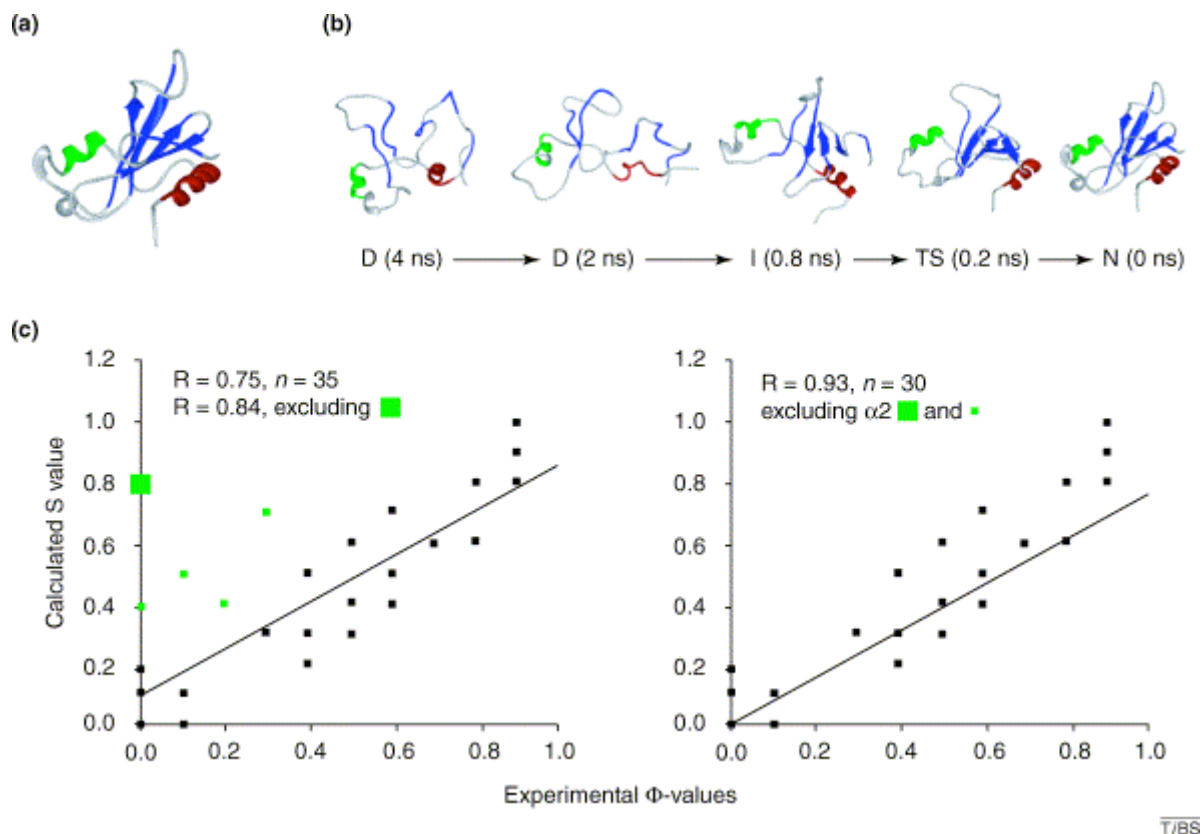


Figure 2 : Représentation schématique des deux modèles de repliements (d'après 2)

Alan Fersht et Valerie Daggett ont étudié expérimentalement et par les méthodes de la dynamique moléculaire le repliement de plusieurs protéines de petite taille (barnase, CI2, engrailed homeodomain) (2). L'approche expérimentale suivie par A. Fersht consiste à mesurer les grandeurs thermodynamiques et cinétiques caractérisant le repliement d'une protéine pour la protéine native et pour plusieurs mutants ponctuels. Ces mesures permettent de déterminer l'énergie libre de l'état natif,  $\Delta G_{N-D}$  et celle de l'état de transition,  $\Delta G_{TS-D}$  et de calculer la quantité  $\phi = (\Delta G_{TS-D} - \Delta G'_{TS-D}) / (\Delta G_{N-D} - \Delta G'_{N-D})$ . Si  $\phi$  est proche de 0 la structure de l'état de transition pour cette position mutée est proche de celle de l'état déplié. Si  $\phi$  est proche de 1 la structure de l'état de transition à cette position est proche de celle de la protéine repliée. Ainsi la mesure expérimentale de  $\phi$  permet de caractériser structurellement l'état de transition. En parallèle V. Daggett a réalisé des simulations du dépliement de ces protéines par dynamique moléculaire à température élevée (225°C). Les trajectoires sont analysées pour identifier le passage par l'état de transition. Celui-ci correspond à une portion de la trajectoire après l'échappement à l'état natif. A partir de cette ensemble de conformations correspondant à l'état de transition, un paramètre S équivalent au paramètre expérimental  $\phi$  est calculé. Une bonne corrélation entre les paramètres S et  $\phi$  est trouvée (Figure 3).



T/BS

Figure 3 : Repliement de la barnase a) structure RMN de la barnase, b) instantanés de la trajectoire de dépliage à haute température (225°C) simulée dans l'eau et présentée dans l'ordre inverse, c) comparaison des valeurs expérimentales de  $\Phi$  et des valeurs calculées de S pour l'état de transition présenté en b.

L'ensemble de ces travaux a permis à ces auteurs de proposer un nouveau modèle du repliement : le modèle « nucléation-condensation ». Ce modèle unifie les modèles « framework » et « collapse hydrophobe » : le repliement démarre autour d'un noyau de repliement (folding nucleus) et à partir de ce germe de repliement présent dans l'état de transition, la structure secondaire et la structure tertiaire se construisent en parallèle. Si le couplage entre la formation de la structure secondaire et la structure tertiaire est faible on se rapproche du « framework » modèle.

### Mécanismes de repliement complexe

Cependant l'étude du repliement de protéines de plus grande taille a montré que leur mécanisme de repliement peut être plus complexe. Lorsque la taille des protéines augmente les barrières dans le paysage énergétique seront vraisemblablement plus grandes et le processus de repliement plus complexe. Alors plusieurs processus différents de repliement seront possibles. Une des protéines qui présente un processus de repliement complexe et qui a été abondamment étudiée est le lysozyme de poule. C. Dobson (3) a montré par plusieurs types d'expériences, en particulier des expériences RMN d'échange, que le repliement de cette molécule est complexe : une partie des molécules (environ 20%) se replie sans apparition

d'intermédiaire de repliement selon une échelle de temps de l'ordre de 100ms. Approximativement 10% des molécules se replient très lentement selon un processus qui est limité par l'isomérisation des prolines. Pour la majorité des molécules (70%) apparaît un intermédiaire transitoire de repliement. L'échelle de temps est alors de l'ordre de 400ms. Cet intermédiaire a pu être caractérisé. La protéine lysozyme comporte deux domaines : un domaine  $\beta$  constitué de 4 hélices alpha et d'une hélice  $3_{10}$ , et un domaine  $\alpha$  constitué d'un feuillet  $\beta$  à trois brins et d'une hélice  $3_{10}$  (Figure 4). Il a été montré qu'au sein de l'intermédiaire transitoire de repliement, seul le domaine  $\beta$  est partiellement structuré.

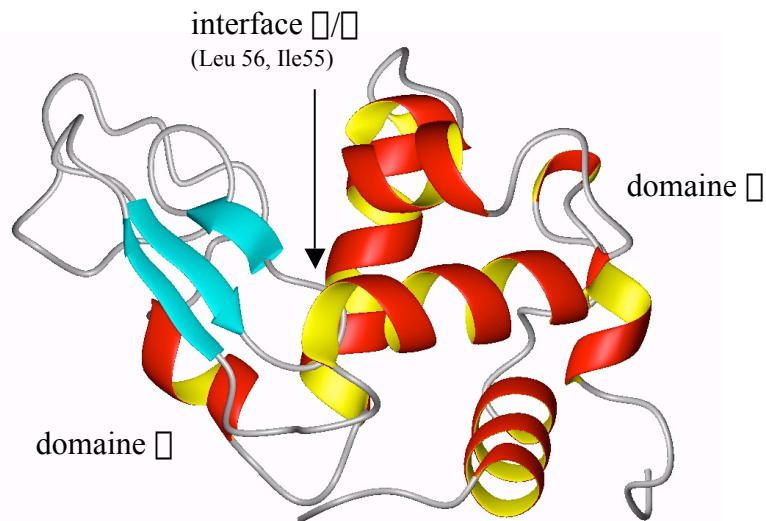


Figure 4 : Représentation schématique de la structure cristalline du lysozyme. Sur la droite, le domaine  $\alpha$  est constitué de cinq hélices (en rouge et jaune). Sur la gauche, le domaine  $\beta$  est constitué de trois brins  $\beta$  (en bleu) et d'une hélice

Afin d'appréhender au niveau atomique le processus de repliement du lysozyme nous avons réalisé des simulations de dynamique moléculaire de dépliement. Un des problèmes majeurs dans la réalisation de simulations de dépliement, est l'échelle de temps des processus de repliement ou dépliement qui est de l'ordre de la milli-seconde à la seconde, c'est à dire d'environ 11 ordres de grandeur plus grand que les temps de simulation réalisable actuellement par dynamique moléculaire. C'est pourquoi les simulations doivent être réalisées à des températures supérieures à la température ambiante ou utiliser une force additionnelle qui favorise le dépliement. La première solution favorise les contributions entropiques et peut donc introduire des artefacts et altérer le mécanisme de dépliement (et donc de repliement). La deuxième solution doit être appliquée avec prudence en introduisant une force faible. Nous avons choisi d'appliquer un potentiel qui dépend de l'écart de tous les atomes de la protéine par rapport à la structure native. Ce type de contrainte ne suppose a priori aucun changement de volume ou de rayon de giration. Les mouvements induits sont ceux qui nécessitent le plus faible changement énergétique.

Pour ce faire nous avons réalisé des simulations de dépliement par dynamique moléculaire en utilisant la méthode PEDC pour écarter la protéine de son état natif (4). Ceci est réalisé par l'addition d'une contrainte sur l'écart par rapport à la structure native. Cette écart a été augmenté progressivement jusqu'à atteindre la valeur de 12\_ pour une durée totale de simulation de 12ns. L'application de cette procédure au lysozyme permet de déterminer les régions de la protéine qui sont les plus rapidement déformées et celles qui résistent le mieux et qui donc constituent des motifs structuraux stables. Deux trajectoires indépendantes ont été réalisées de façon à dégager les caractéristiques communes des deux processus de dépliements.

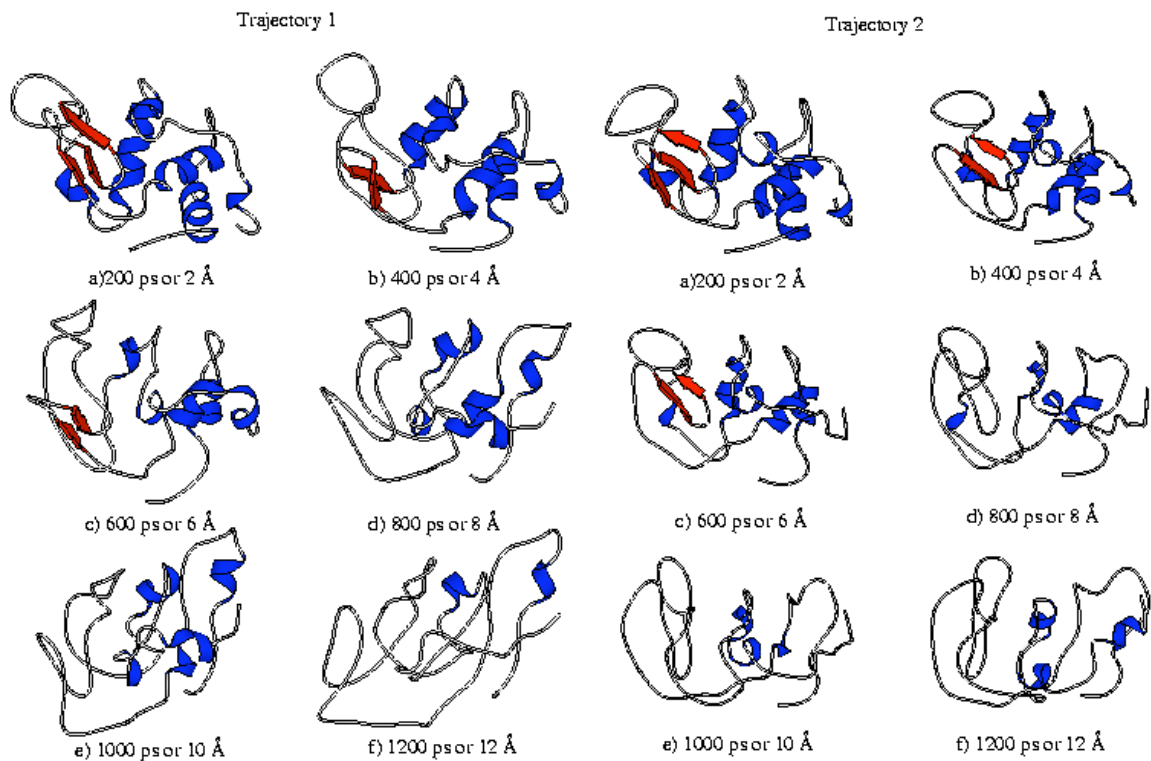


Figure 5 : Représentation schématique du dépliement (a -> f) ou du repliement (f -> a) du lysozyme à six étapes différentes d'une trajectoire de dépliement obtenus par simulation de dynamique moléculaire. Les hélices  $\alpha$  sont représentées par des serpents en bleu et les feuilletts  $\beta$  par des flèches rouges.

La même chronologie des événements est observée dans les deux simulations (figure 5) : le site actif perd sa configuration, l'hélice C présente une cassure en son milieu et un début de dépliement dans sa partie N-terminale, le feuillet  $\beta$  se déplie en se réduisant à deux brins tandis que les hélices A, B, C et D restent en grande partie repliées, enfin la structure du domaine  $\beta$  est complètement perdue alors que les hélices présentent encore 40% de leur structure secondaire en hélice. On observe que le dépliement des hélices est extrêmement progressif et qui est dû à une hydratation progressive qui démarre des extrémités. Les liaisons hydrogène qui maintiennent la structure en hélice sont détruites et remplacées par des liaisons hydrogène avec des molécules d'eau. Ces simulations montrent que le dépliement du domaine  $\beta$  est corrélé à la perte des contacts hydrophobes entre ce domaine et le domaine  $\alpha$ . Le brin

central du feuillet  $\beta$  est le plus stable. Ceci n'est pas du seulement à sa position centrale dans le feuillet mais à la présence du coude  $\beta$  Ile55 et Leu56 qui est inséré dans le cœur hydrophobe du domaine  $\beta$ . Les résidus Ile 58 et Ile 98 contribuent également à former une interface hydrophobe entre les domaines  $\beta$  et  $\beta'$ . Le dépliement du domaine  $\beta$  est consécutif à la désagrégation de cette interface hydrophobe. Cette interface hydrophobe est nécessaire à la structuration du domaine  $\beta$  en feuillet.

L'étude cinétique de l'échange des protons amides durant le repliement montre que 23 protons du domaine  $\beta$  sont rapidement protégés alors que dans le domaine  $\beta'$  seuls 4, localisés dans les boucles le sont. D'autres expériences confirment l'existence d'un intermédiaire de repliement présentant de nombreux protons protégés dans le domaine  $\beta$ . Cet intermédiaire est analogue au états partiellement dépliés qui ont pu être caractérisés dans divers solvants. Des facteurs de protection théoriques ont été calculés à partir de l'ensemble des conformations générées dans les deux trajectoires de dépliements. Il apparaît que 14 protons amides sont protégés dans le domaine  $\beta$  et 9 dans le domaine  $\beta'$ . Ces simulation apparaissent donc globalement en accord avec les données expérimentales. Cependant le domaine  $\beta$  apparaît plus structuré expérimentalement que dans nos simulations.

Néanmoins ces simulations permettent de décomposer au niveau atomique le processus de repliement du lysozyme et d'éclairer ses propriétés particulières. Ainsi pour se replier le domaine  $\beta$ , commence par insérer deux de ses résidus ( Leu56 et Ile55 ) dans le domaine  $\beta'$  en cours de repliement. Si ces résidus sont mutés d'hydrophobes en hydrophiles ( en particulier l'Ile55 muté en Thr) alors leur capacité à s'insérer dans le domaine  $\beta'$  en cours de repliement est moindre et la protéine se replie moins vite. Elle va rester plus longtemps dans une forme partiellement dépliée dans laquelle le domaine  $\beta$  n'est pas replié. Dans cet état intermédiaire de repliement, les domaines  $\beta$  de deux protéines différentes, peuvent s'associer et conduire à la formation d'une fibre. C'est effectivement ce qu'il se produit pour ce mutant du lysozyme qui forme des fibres amyloïdiques. D'une manière plus générale les intermédiaires de repliement pourraient conduire à des formes mal repliées des protéines. Ces formes mal repliées ou partiellement dépliées ont tendance généralement à s'agréger et à former des fibres amyloïdiques. Ainsi les intermédiaires de repliement seraient à l'origine de formes pathogènes des protéines observées dans les maladies neuro-dégénératives.

D'autres méthodes ont été utilisées pour simuler le dépliement du lysozyme (action sur le rayon de giration, action sur les cavités internes, réalisation de dynamique à haute température) (5). Généralement il n'a pas été observé de différence dans le dépliement des deux domaines. Ce résultat peut être rapproché des observations expérimentales qui indiquent qu'à 50°C il n'est plus observé de différence entre le repliement des deux domaines. On peut supposer que les méthodes utilisées accélèrent trop le dépliement pour que des différences de comportement entre les deux domaines soient observées. Par des simulations à température moyennement élevée, des auteurs ont par contre observé une différence dans le comportement des deux domaines (6). Il apparaît donc que, selon la méthode utilisée pour accélérer le dépliement afin de le rendre accessible à des simulations d'une durée de l'ordre de la nanoseconde et donc réalisable sur les ordinateurs dont on dispose actuellement, les résultats peuvent être concordants ou différents. Il est donc essentiel de confronter les résultats des simulations à des données expérimentales. C'est dans cette esprit qu'ont été réalisées des simulations de dépliement ou celui-ci est favorisé par une force de traction qui tend à éloigner la partie N-terminale de la partie C-terminale (7). Les résultats ont été comparés à ceux obtenus par des simulations à hautes température. Réalisés sur quatre protéines différentes

(deux composées de feuillets  $\beta$  uniquement et deux autres d'hélices  $\alpha$ ), les résultats montrent que même si les processus de dépliements ne sont pas identiques, des propriétés communes sont observées i) l'existence d'un noyau de repliement, (les même feuillets  $\beta$  sont les plus lent à se déplier quelque soit la méthode), ii) les deux protéines structurées en hélice  $\alpha$  se déplient plus progressivement que les deux protéines en feuillet  $\beta$ .

### **Etude de repliement des protéines par dynamique moléculaire**

Avec la croissance de la puissance des ordinateurs, il est maintenant possible de réaliser de simulations de dynamique moléculaires de repliement pour des protéines de petite taille. Ces simulations sont réalisées en en tenant compte explicitement du solvant et en considérant explicitement tous les atomes de la protéine. Ce type d'études est difficile à entreprendre car il nécessite un grand nombre d'heures de calcul. Cependant il est essentiel car il permet d'accéder au paysage énergétique. Une telle étude a été réalisée dès 1995 par Brooks (8) sur un paquet de trois hélices (46 résidus). Plusieurs trajectoires de 9ns ont été réalisées. Les structures ont été regroupées selon une fonction qui tient compte des différences en terme de liaison hydrogène du squelette, de surface accessible et de l'arrangement du cœur. La variable utilisée pour suivre le repliement/dépliement est le rayon de giration. A partir de ces statistiques a pu être tracée l'évolution de l'énergie libre en fonction du rayon de giration. De cette étude les auteurs peuvent conclure : le profil énergétique permet de distinguer trois régions : la région autour de la conformation native, une région plus réduite proche de l'état natif correspondant à un état transitoire, et la région correspondant à des conformations dépliées. L'état natif est environ 1,1 kcal/mol plus bas que l'état intermédiaire et 2,6 kcal/mol plus bas que l'état déplié. L'état natif est favorisé par les interaction protéine-protéine et défavorisé par l'interaction avec le solvant. L'entropie de la protéine augmente avec le dépliement alors que l'entropie protéine-solvant diminue. L'accroissement de l'exposition au solvant durant le dépliement entraîne une diminution de la mobilité des molécules d'eau autour de la protéine ( c'est ce qu'on appel l'effet hydrophobe). Ainsi ces calculs, même si ils ne sont réalisés sur des protéines de petite taille, et donc suivant un repliement simple, permettent d'analyser en détail le paysage énergétique du repliement.

Depuis d'autres simulations du repliement ont été réalisées, en utilisant de nouveaux outils qui permettent d'échantillonner plus efficacement le paysage énergétique (replica exchange molecular dynamics, REMD) (9). La variable utilisé pour suivre le dépliement/repliement est généralement la fraction de contact natif qui est préférée au rayon de giration. Globalement ils arrivent aux même conclusions : « la formation de la structure secondaire est couplé avec l'exclusion du solvant et la formation du cœur ».

### **Conclusion**

Les études par dynamique moléculaire du repliement ou du dépliement des protéines ont permis de mettre au point des méthodes d'exploration du paysage énergétique. Ces méthodes utilisées massivement sur un grand nombre de protéines permettront vraisemblablement d'améliorer la prédiction ab-initio de repliement. En effet actuellement les méthodes de prédiction ne fournissent que des modèles à basse résolution (déviation de l'ordre de 4 Å par rapport à la structure tridimensionnelle déterminée expérimentalement). Une des voies pour améliorer ces méthodes est l'exploration systématique du paysage énergétique. C'est alors que la prédiction de la structure pourra servir de base à la conception de nouveaux médicaments. Enfin, on voit à travers l'exemple du lysozyme comment l'étude de son repliement et de son dépliement par dynamique moléculaire est nécessaire à la compréhension

du mécanisme moléculaire qui est à l'origine de la formation d'états conformationnels mal repliés.

## Références

1. Dill K.A., Chan H.S., Nat. Stru. Biol. 1997, 4, 10-19 From Levinthal to pathways to funnels
2. Daggett V., Fersht A.R., TIBS 2003, 28, 18-25 Is there a unifying mechanism for protein folding ?
3. Radford S.E., Dobson C.M., Evans P.A., Nature 1992 358,302-358., The folding of hen lysozyme involves partially structured intermediates and multiple pathways.
4. Gilquin B., Guilbert C. Perahia D., Proteins 2000, 41, 58-74 Unfolding of hen egg lysozyme by molecular dynamics simulation at 300K: Insight into the role of the interdomain interface.
5. Hünenberger P.H., Mark A.E., van Gunsteren W.F. Proteins 1995, 21, 196-213 Computational approaches to study protein unfolding: hen egg white lysozyme as a case study.
6. Kazmirski S.L., Daggett V. J Mol Biol. 1998, 284,793-80. Non-native interactions in protein folding intermediates: molecular dynamics simulations of hen lysozyme.
7. Paci E., Karplus M., PNAS 2000, 97,6521-6526, Unfolding proteins by external forces and temperature : the importance of topology and energetics.
8. Boczek E.M., Brooks III C.L. Science 1995, 269, 393-396 First-principles calculation of the folding free energy of a three-helix bundle protein.
9. Garcia A.E., Onuchic J.N., PNAS 2003, 100, 13898-13903 Folding a protein in a computer : an atomic description of the folding/unfolding of protein A.