



Les modes normaux de basse fréquence des protéines

Yves-Henri Sanejouand

► **To cite this version:**

Yves-Henri Sanejouand. Les modes normaux de basse fréquence des protéines. Biophysique [physics.bio-ph]. Université Claude Bernard - Lyon I, 2007. <tel-00258781>

HAL Id: tel-00258781

<https://tel.archives-ouvertes.fr/tel-00258781>

Submitted on 25 Feb 2008

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

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

MÉMOIRE

présenté à

l'Université Claude Bernard - Lyon 1

en vue de l'obtention de

L'HABILITATION À DIRIGER DES RECHERCHES

par

Yves-Henri Sanejouand

LES MODES NORMAUX DE VIBRATION DE BASSE FRÉQUENCE DES PROTÉINES

Soutenue le 5 novembre 2007, devant le jury composé de :

Président : Gilbert DELEAGE
Rapporteur : Daniel BORGIS
Rapporteur : Marc DELARUE
Rapporteur : Richard LAVERY
Examineur : Florent CALVO
Examineur : Philippe DERREUMAUX
Examineur : Catherine ETCHEBEST

Les modes normaux de vibration de basse fréquence des protéines

Yves-Henri Sanejouand¹

Laboratoire Joliot Curie² et laboratoire de Physique³
Ecole Normale Supérieure
46 allée d'Italie, 69364 Lyon Cedex 07

Résumé

Les mouvements de basse fréquence des protéines, tels qu'on peut les calculer via l'approximation des petits déplacements (approximation harmonique), ressemblent souvent beaucoup à des mouvements fonctionnels, tels qu'on peut les observer par cristallographie des rayons X. Ce résultat semble très robuste puisqu'il a été obtenu, tout d'abord, en partant d'une description des protéines à l'échelle atomique puis, plus récemment, de descriptions à "gros grains" (des modèles de type réseau élastique). S'appuyant sur ce constat, des applications variées ont été proposées, notamment pour faciliter la résolution de structures ou pour améliorer leur qualité (ajustement dans des enveloppes obtenues par cryomicroscopie électronique, remplacement moléculaire...). Ce mémoire est une revue des résultats obtenus, dans laquelle mes contributions sont plus particulièrement mises en avant. Parmi les développements les plus récents, sont évoquées des tentatives de faire le lien entre modèles à l'échelle atomique et modèles à gros grains, ou bien de complexifier ces derniers, en y incluant des termes anharmoniques.

Mots-clés : approximation harmonique, changement de conformation, modèles de réseau élastique, robustesse, anharmonicité, localisation non-linéaire d'énergie.

¹yves-henri.sanejouand@ens-lyon.fr

²USR 3010 du CNRS.

³UMR 5672 du CNRS.

Table des matières

1	Introduction	4
2	L'approximation harmonique	6
3	L'approximation RTB	9
4	Le lien avec les mouvements fonctionnels	13
5	Les modèles de réseau élastique	15
5.1	L'idée de Monique Tirion	15
5.2	L'apport d'Ivet Bahar	16
5.3	La synthèse de Konrad Hinsen	16
5.4	Elnémo	17
5.5	Comparaison avec les données expérimentales	18
5.5.1	Les facteurs B	18
5.5.2	Les changements de conformation	19
5.6	Comparaison avec les modes normaux standards	20
5.7	Vers des modèles mixtes	22
6	Applications	28
6.1	Intégration des équations du mouvement	28
6.2	Validation de changements de conformation	30
6.3	Ajustement dans des enveloppes à basse résolution	32
6.4	Remplacement moléculaire	32
6.5	Prédiction du changement de conformation du CD4	33
6.5.1	A partir des modes de vibration	33
6.5.2	Rebondissement	34
6.5.3	Entrelacement de domaines	35
6.5.4	Perspectives	38
6.6	Autres prédictions	38
7	Un modèle de réseau élastique anharmonique	39
7.1	Introduction	39
7.2	Le modèle choisi	39
8	Perspectives	41
8.1	Localisation d'énergie et activité catalytique	41
8.2	Vers la prédiction de changements de conformation	41
9	Remerciements	43
10	Curriculum vitae	44

1 Introduction

Les protéines sont des hétéropolymères qui peuvent adopter une structure tridimensionnelle très précisément définie. Cette structure est cependant flexible, et ceci s'avère parfois indispensable au bon fonctionnement de la protéine. Par exemple, la flexibilité de l'isopropylmalate déshydrogénase (IPMDH), une enzyme impliquée dans la biosynthèse de la leucine (voir Fig.1), a été étudiée en déterminant ses constantes de vitesse d'échange hydrogène-deutérium. Lorsque l'expérience est faite à température ordinaire, on constate que l'échange hydrogène-deutérium est beaucoup plus rapide lorsque l'enzyme provient d'*E. Coli* que lorsqu'elle provient de *Th. Thermophilus*. Par contre, les vitesses d'échange s'avèrent très similaires lorsqu'elles sont déterminées à la température de fonctionnement optimum de l'enzyme, à savoir 48 degrés lorsque l'enzyme provient d'*E. coli* et 70 degrés lorsqu'elle provient de *Th. Thermophilus*. Pourtant, les structures tridimensionnelles de ces enzymes sont très semblables : les positions de leurs C_α diffèrent d'environ 1Å, en moyenne.⁴ Le taux d'identité de leurs séquences est aussi élevé : plus de 50%. Malgré cela, l'activité de l'IPMDH de *Th. Thermophilus* est inhabituellement faible à température ordinaire. Tout ceci suggère que des mouvements d'origine thermique sont nécessaires au bon fonctionnement de l'IPMDH de *Th. Thermophilus*[1].

Il semble probable que le changement de conformation qui doit se produire lors de l'expression de la fonction de cette enzyme soit la conséquence d'un mouvement de "fermeture" de ses domaines structuraux. On observe déjà un mouvement de ce type, quoique d'amplitude modeste, lorsque l'isopropylmalate est lié à l'enzyme (code PDB 1A05⁵) mais la forme adoptée par le complexe ternaire, en présence de l'isopropylmalate et du cofacteur, le NAD, qui pourrait bien être complètement "fermée", n'est pas encore connue.⁶

Les simulations de dynamique moléculaire permettent d'étudier les mouvements d'origine thermique des protéines mais, actuellement, il n'est guère envisageable d'aller nettement au-delà de la microseconde lorsqu'on s'intéresse à des protéines de plusieurs centaines d'acides aminés et ce, même en utilisant une représentation simplifiée des effets du solvant et en ayant accès à des moyens de calcul hors norme. On peut cependant s'affranchir de cette limitation en étudiant ces mouvements dans le cadre d'une approximation, l'approximation harmonique (ou approximation des petits déplacements), qui permet de décrire le mouvement de chacun des atomes d'un système comme une combinaison de modes de vibration indépendants les uns des autres.

Dès 1976, Martin Karplus et ses collaborateurs ont tenté de prédire le mouvement fonctionnel du lysozyme humain en supposant qu'il ressemble à son mode de vibration de plus basse fréquence et qu'il s'agit d'un mouvement "de charnière" (hinge-bending) au

⁴Le "root-mean-square deviation" obtenu en superposant les structures 1XAA et 1CM7. Ce qui sera noté par la suite : C_α -r.m.s.

⁵Il s'agit cependant d'une enzyme provenant d'un autre organisme : *Thiobacillus Ferrooxidans*.

⁶Les mouvements de ce type ont cependant des constantes de vitesse bien plus élevées que ceux qui sont responsables des échanges hydrogène-deutérium mesurés par G.A. Petsko et ses collaborateurs[1]. Une hypothèse supplémentaire est donc nécessaire pour expliquer leurs résultats : celle d'un lien entre les mouvements fonctionnels et les mouvements les plus lents de l'enzyme, à savoir ceux qui sont impliqués dans son dépliement (et/ou son repliement), puisque les échanges hydrogène-deutérium les plus lents ne doivent se produire que lorsque la protéine est partiellement dépliée. Un tel lien a déjà été suggéré par ailleurs[2].



FIG. 1 – Structure de l’isopropylmalate déshydrogénase de *Th. Thermophilus* (code PDB 1XAA). Dessin réalisé avec Molscript[3].

cours duquel deux domaines quasi-rigides se déplacent l’un par rapport à l’autre[4]. Puis, au début des années quatre-vingt, dans les groupes de Martin Karplus, Nobuhiro Go et Michael Levitt, les modes de vibration de peptides[5], puis de petites protéines, furent calculés : le BPTI[6, 7, 8, 9], la crambine, la ribonucléase[10] et le lysozyme humain[10, 11]. Signalons aussi le travail méconnu de Robert W. Harrison qui calcula dès cette époque les modes de vibration d’une protéine de taille plus standard, l’hexokinase (environ 450 acides aminés) en utilisant la méthode de Ritz-Raleigh[12] et une base de fonctions spline. Comme il le précise ”using a spline is equivalent to using a thin deformable strip or rod as a virtual structure”[13]. Il constata de plus que ”for the native open form there were (at least) two modes that had strong components along the conformation change”[13], mais sans donner plus de détails. Comme nous le verrons dans la suite de ce texte, son approche, qui revient à réduire la dimension du système étudié, a depuis été reprise avec succès. Quand à son principal résultat, il a été amplement confirmé. Dans le cas de nombreuses protéines, on trouve en effet qu’un ou deux de leurs modes de vibration suffisent pour décrire avec une très bonne précision leur changement de conformation fonctionnel.

Cependant, dès que les simulations de dynamique moléculaire de protéines en solvant explicite devinrent possibles, il semble que l’intérêt pour les calculs de modes de vibration *per se* ait fortement diminué, peut être parce que les premiers calculs étaient effectués *in vacuo*⁷ et qu’il ne paraissait pas évident qu’on puisse y inclure un jour une description suffisamment réaliste de l’environnement.

C’est maintenant chose possible, depuis qu’on a été proposés des modèles de solvant ”implicites”, tels EEF1[14], ACE[15], *etc*, qui permettent de tenir compte de l’énergie libre de solvation, à peu près telle qu’on peut l’obtenir avec une approche de type Poisson-Boltzman[16]. De plus, comme nous le verrons dans la suite de ce texte, on peut main-

⁷Avec, au mieux, une constante diélectrique fonction de la distance.

tenant calculer, au moins de manière approximative, les modes de vibration de basse fréquence de protéines ou de complexes protéiques de très grande taille (par exemple ceux du ribosome[17]) et, depuis les premières études de modèles (très) simplifiés des protéines[18], on comprend désormais mieux pourquoi l'approximation harmonique permet d'extraire des informations dynamiques pertinentes, malgré ses limitations les plus évidentes, qui sont détaillées ci-après.

2 L'approximation harmonique

Pour de petits déplacements des coordonnées de ses N atomes, V , l'énergie potentielle d'une protéine, peut être approximée par un développement en série limitée :

$$V = V_0 + \sum_{i=1}^{3N} \left(\frac{\partial V}{\partial r_i} \right)_0 (r_i - r_i^0) + \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \left(\frac{\partial^2 V}{\partial r_i \partial r_j} \right)_0 (r_i - r_i^0)(r_j - r_j^0) + \dots \quad (1)$$

où r_i est la i -ème coordonnée et où V_0 est l'énergie potentielle de la structure dont on veut calculer les modes de vibration ; l'indice "0" est associé aux coordonnées de la conformation considérée. Comme en physique classique V est définie à une constante additive près, on peut poser : $V_0 = 0$. De plus, si V_0 est un minimum de la fonction V , alors on a pour tout i :

$$\left(\frac{\partial V}{\partial r_i} \right)_0 = 0 \quad (2)$$

Enfin, si les déplacements atomiques sont suffisamment petits, on peut négliger les termes d'ordre supérieur du développement en série. On obtient alors :

$$V = \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \left(\frac{\partial^2 V}{\partial r_i \partial r_j} \right)_0 (r_i - r_i^0)(r_j - r_j^0) \quad (3)$$

En d'autres termes, si l'on ne s'intéresse qu'aux petits mouvements des atomes d'une protéine au voisinage d'une configuration d'équilibre, on peut considérer que son énergie potentielle a une forme quadratique. Or il se trouve que, sauf en présence de contraintes dépendant explicitement du temps[19], l'énergie cinétique de tout système peut aussi être mise sous une forme quadratique et que, dans ce cas, les équations du mouvement ont une solution analytique, de la forme suivante :

$$r_i(t) = r_i^0 + \frac{1}{\sqrt{m_i}} \sum_{k=1}^{3N} C_k a_{ik} \cos(2\pi\nu_k t + \Phi_k)$$

où m_i est la masse associée à la coordonnée i et où C_k et Φ_k , respectivement l'amplitude et la phase du mode de vibration k , dépendent des conditions initiales, c'est-à-dire des positions et des vitesses des atomes à $t = 0$. En particulier, C_k est fonction de E_k , l'énergie totale du mode k . Notamment, lorsque l'équipartition de l'énergie est respectée, on a :

$$C_k = \frac{\sqrt{2k_B T}}{2\pi\nu_k} \quad (4)$$

où T est la température et k_B la constante de Boltzmann. Ce qui signifie que l'amplitude du mode k est inversement proportionnelle à ν_k , sa fréquence. De fait, dans le cadre de

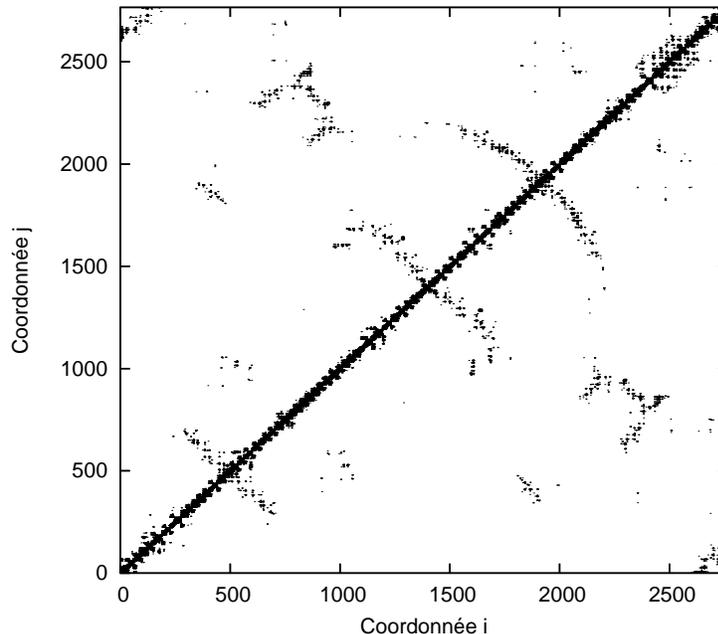


FIG. 2 – Éléments non-nuls du hessien de l'énergie potentielle de la protéase du virus de l'immunodéficience humaine (code PDB 1HHP). Cette protéine a 99 acides aminés. Les motifs de part et d'autre de la diagonale sont dus aux structures secondaires. Chaque point correspond à l'un des 40000 éléments les plus grands de la matrice.

l'approximation harmonique, on trouve que les modes de fréquence inférieure à $30\text{-}100\text{ cm}^{-1}$ sont responsables de l'essentiel des déplacements des atomes d'une protéine[5].⁸

On obtient les $3N$ valeurs des fréquences en diagonalisant \mathbf{H} , le Hessien de l'énergie potentielle pondéré par les masses atomiques, c'est-à-dire la matrice dont l'élément ij s'écrit :⁹

$$\left(\frac{\partial^2 V}{\sqrt{m_i m_j} \partial r_i \partial r_j} \right)_0$$

Diagonaliser \mathbf{H} , cela signifie l'exprimer dans un nouveau repère tel que, pour tout i différent de j , on ait :

$$\left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0 = 0$$

Les valeurs propres de la matrice, c'est-à-dire les éléments diagonaux de la matrice transformée, permettent d'obtenir les fréquences de vibration du système étudié, puisque l'on a alors[19, 20] :

$$\left(\frac{\partial^2 V}{\partial^2 q_i} \right)_0 = (2\pi\nu_i)^2$$

tandis que les nouvelles coordonnées, dites coordonnées normales, sont des combinaisons linéaires des coordonnées initiales, telles que, pour la coordonnée k :

$$q_k = \sum_i^{3N} a_{ik} \sqrt{m_i} (r_i - r_i^0) \quad (5)$$

⁸Un mouvement de fréquence 1 cm^{-1} a une période d'environ 33 picosecondes.

⁹La pondération par les masses permet de ne pas tenir compte du Hessien de l'énergie cinétique[19, 20].

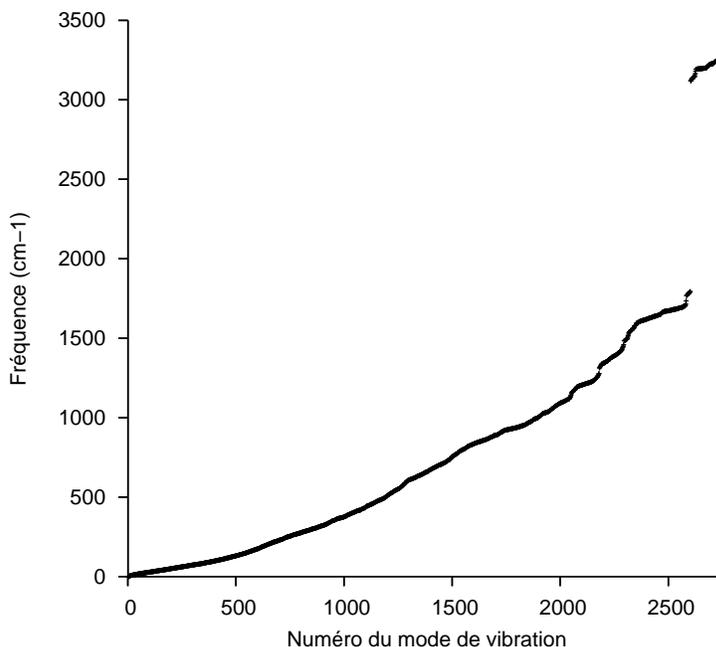


FIG. 3 – Spectre de la protéase du virus de l’immunodéficience humaine (code PDB 1HHP). Les fréquences sont données par ordre croissant.

où a_{ik} est l’élément i du vecteur propre k de \mathbf{H} .

En pratique, \mathbf{H} étant une matrice réelle symétrique, il est toujours possible de la diagonaliser. Lorsque sa taille n’est pas trop grande, c’est-à-dire lorsqu’on peut la stocker dans la mémoire vive d’un ordinateur, les algorithmes les plus courants permettent de le faire avec un coût en calcul (CPU) proportionnel à nN^2 , où n est le nombre de valeurs et de vecteurs propres que l’on souhaite déterminer.

Dans le cas des protéines, \mathbf{H} est d’ordinaire une matrice creuse, puisque les interactions entre atomes lointains ne sont en général pas calculées. La Figure 2 montre la répartition des éléments non-nuls de \mathbf{H} dans le cas de la protéase du virus de l’immunodéficience humaine (VIH-1), une petite protéine de 99 acides aminés, 922 atomes (et donc : $3N=2766$ coordonnées¹⁰), lorsque \mathbf{H} est calculé avec un programme de modélisation moléculaire standard, tel CHARMM[21], tandis que la Figure 3 montre l’ensemble des fréquences de vibration de cette protéine, telles qu’elles sont obtenues à partir des valeurs propres de \mathbf{H} . Dans la gamme des hautes fréquences, chaque bande correspond à un motif chimique. Par exemple, vers 3300 cm^{-1} , il s’agit de vibrations des liaisons chimiques azote-hydrogène des liaisons peptidiques.¹¹ C’est l’existence de ces modes ”localisés”, dont les fréquences sont caractéristiques de la présence de certains groupements chimiques dans une molécule, qui a permis de faire de la spectroscopie infra-rouge une méthode d’analyse structurale fort utile.

On ne remarque par contre aucune structuration particulière du spectre dans la gamme

¹⁰Dans la représentation ”atomes étendus”, c’est-à-dire en ne prenant en compte que les hydrogènes impliqués dans des liaisons chimiques ”polaires”.

¹¹Dans CHARMM, les valeurs des hautes fréquences sont nettement sous-estimées. Un champ de force incluant des termes supplémentaires est nécessaire afin de mieux rendre compte de cette partie du spectre[22].

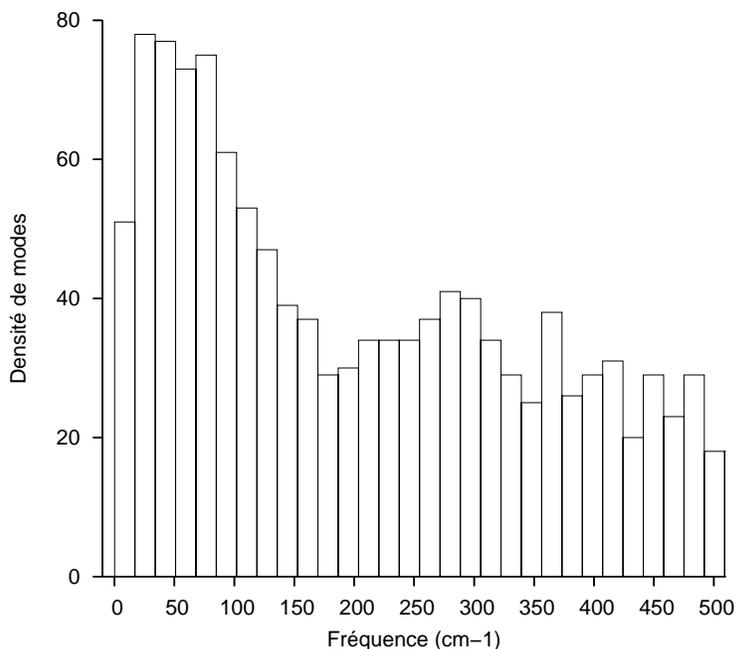


FIG. 4 – Densité de modes dans la partie basse fréquence du spectre de la protéase du virus de l’immunodéficience humaine.

des basses fréquences, qu’on décrit d’ordinaire en termes de densité de modes par unité de fréquence. On constate alors qu’il y a un pic de densité vers 50 cm^{-1} (voir Fig. 4). En deçà de cette valeur, les vibrations sont en général très délocalisées, c’est-à-dire qu’un grand nombre d’atomes y sont impliqués. Dans le cas des plus basses fréquences, les mouvements sont ceux de domaines structuraux entiers. Nous verrons par la suite qu’ils ressemblent parfois beaucoup aux mouvements fonctionnels de la protéine étudiée.

3 L’approximation RTB

Pour pouvoir obtenir les modes de vibration d’une protéine de grande taille, il faut soit renoncer à une description à l’échelle atomique, soit renoncer à calculer l’ensemble de son spectre. Partant de cette seconde option, plusieurs méthodes ont été proposées. Celle de Ritz-Raleigh, appliquée tout d’abord, comme rappelé dans l’Introduction, au cas de l’hexokinase par R.W. Harrison, a ensuite été reprise par Minc-Honc Hao et Stephen C. Harvey[23]. La méthode de Lanczos, tout d’abord testée par Bernard R. Brooks et Martin Karplus dans le cas du lysozyme[11], a par la suite été appliquée sous une forme plus sophistiquée¹² au cas d’une grosse protéine, la citrate synthase sous forme dimérique (858 acides aminés), par O. Marques¹³ et moi-même[24]. Ce fut pour un bref moment la plus grosse protéine dont l’analyse en modes normaux (de basse fréquence) ait été faite. Ce ”record” fut rapidement battu par Lilianne Mouawad et David Perahia, qui utilisaient DIMB, une méthode originale développée et implémentée par eux dans CHARMM[25, 26].

Simultanément, nous avons proposé, Georges Trinquier, Philippe Durand et moi-même[27],

¹²Un algorithme de type Bloc-Lanczos implémenté par Osni Marques dans le logiciel BLZPACK. Celui-ci est disponible sur sa page Web : <http://crd.lbl.gov/~osni/>.

¹³Alors au CERFACS, à Toulouse. O. Marques travaille actuellement au Lawrence Berkeley National Laboratory.

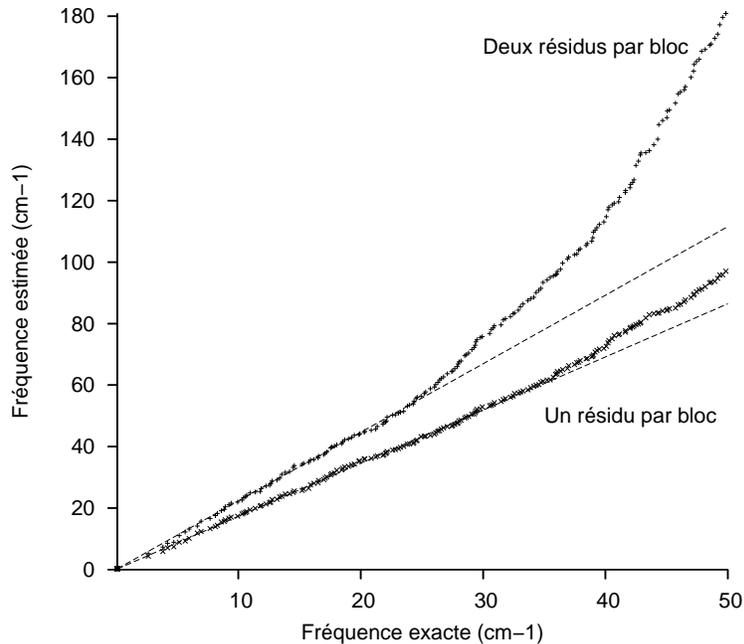


FIG. 5 – Fréquences de la protéase du virus de l’immunodéficience humaine, calculées avec l’approximation RTB, en fonction de la valeur des fréquences des modes de vibration ayant le même numéro d’ordre. L’approximation RTB est utilisée en mettant un (les croix) ou deux (les signes ”plus”) résidus d’acide aminé dans chaque bloc. Les lignes tiretées sont des fits linéaires.

une approximation particulièrement efficace dont le principe est le suivant.¹⁴ Au lieu de diagonaliser \mathbf{H} , une matrice $3N \times 3N$, on diagonalise \mathbf{H}_b , une matrice qui est l’expression de \mathbf{H} dans un espace de dimension $6n_b$:

$$\mathbf{H}_b = \mathbf{P}^t \mathbf{H} \mathbf{P}$$

où \mathbf{P} est une matrice $3N \times 6n_b$ orthogonale construite avec les vecteurs décrivant les trois mouvements de translation et les trois mouvements de rotation de chacun des n_b ”blocs” du système étudié, et où \mathbf{P}^t est sa transposée. Dans le cas des protéines, on fait donc l’hypothèse que chacun des modes que l’on souhaite calculer peut être correctement décrit comme une combinaison linéaire de mouvements de blocs parfaitement rigides.¹⁵ \mathbf{A}_p , la matrice des $6n_b$ pseudo-vecteurs propres de \mathbf{H} , est ensuite obtenue à partir de \mathbf{A}_b , la matrice des vecteurs propres de \mathbf{H}_b :¹⁶

$$\mathbf{A}_p = \mathbf{P} \mathbf{A}_b$$

Comme on le voit sur la Figure 5, les fréquences calculées avec l’approximation RTB, obtenues à partir des valeurs propres de \mathbf{H}_b , sont toujours plus grandes que les fréquences exactes. Cependant, dans le cas des fréquences inférieures à $25\text{-}35\text{ cm}^{-1}$, elles le sont en proportion, c’est-à-dire que l’on a [28] :

$$\nu_b \simeq d_p \nu_s$$

¹⁴L’idée sur laquelle repose cette méthode est due à Philippe Durand, fondateur du laboratoire de Physique Quantique et actuellement professeur émérite à l’Université Paul Sabatier de Toulouse.

¹⁵D’où le nom de l’approximation. RTB signifie en effet : Rotations-Translations de Blocs.

¹⁶ \mathbf{H}_b est une matrice d’ordre $6n_b$. Elle est donc beaucoup plus rapide à diagonaliser que \mathbf{H} , qui est d’ordre $3N$.

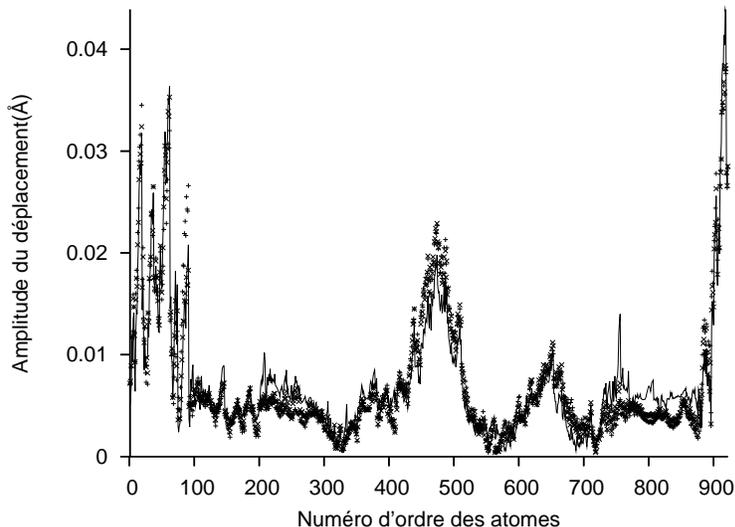


FIG. 6 – Mode de vibration de plus basse fréquence de la protéase du virus de l’immunodéficience humaine, calculé de manière standard (ligne pleine) ou avec l’approximation RTB, en mettant un (les croix) ou deux (les signes ”plus”) acides aminés dans chaque bloc.

où ν_s et ν_b sont, respectivement, les fréquences exactes et approximées, avec, dans le cas de la protéase du virus de l’immunodéficience humaine, $d_P = 1.73 \pm 0.03$, lorsqu’il y a un résidu d’acide aminé dans chaque bloc, et $d_P = 2.23 \pm 0.01$, lorsqu’il y en a deux. Dans le cadre de la thèse de Florence Tama,¹⁷ nous avons étudié ainsi cinq autres cas[28] : la crambine, l’adénylate kinase, la triglycérine lipase, l’alcool déshydrogénase et l’énolase et nous avons obtenu $d_P = 1.7 \pm 0.1$, lorsqu’il y a un résidu d’acide aminé dans chaque bloc, et $d_P = 2.1 \pm 0.2$, lorsqu’il y en a deux. La valeur de d_P ne semble donc pas dépendre de la taille de la protéine étudiée (la crambine n’a que 46 acides aminés tandis que l’énolase en a 436). Elle n’est pas non plus significativement affectée lorsqu’on modifie la façon dont on calcule les interactions non-liées. En effet, dans notre étude initiale[28], la distance de coupure (cutoff) pour les interactions électrostatiques était de 8.5Å alors que dans le cas présent elle est de 9.0Å . Plus significativement, les effets de solvant n’étaient pas pris en compte ($\epsilon = 1$) tandis qu’ici le modèle de solvant implicite EEF1[14] a été utilisé (dans sa version 1.1[29]). Ce modèle inclus notamment une neutralisation des extrémités des chaînes latérales des acides aminés chargés (arginine, aspartate, glutamate et lysine). Des valeurs similaires (1.8 et 2.0 , respectivement) ont été obtenues par Guohui Li et Qiang Cui[30], lors de l’étude qui a illustré leur implémentation de l’approximation RTB dans CHARMM[21].¹⁸ Cette dernière fut effectuée avec une distance de coupure de 8.0Å et une constante diélectrique fonction de la distance. Guohui Li et Qiang Cui ont aussi regardé le cas d’un fragment d’ADN double brin. Avec un nucléotide dans chaque bloc, ils ont obtenu $d_P = 2.7$, ce qui confirme que d_P dépend du nombre d’atomes mis dans chaque bloc mais ce d’une manière non triviale puisqu’un nucléotide n’a, en moyenne, que deux fois plus d’atomes qu’un résidu d’acide aminé.¹⁹

¹⁷Thèse qui fut co-dirigée par François-Xavier Gadéa et moi-même, à Toulouse, ainsi que par Nobuhiro Go, à Kyoto.

¹⁸Sous l’acronyme BNM -pour Block Normal Modes.

¹⁹Dans la représentation ”atomes étendus”.

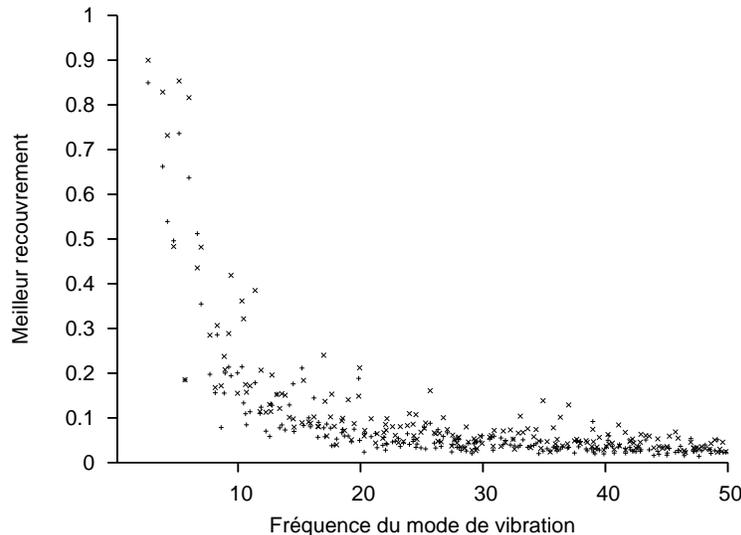


FIG. 7 – Calcul des modes de vibration de plus basse fréquence de la protéase du virus de l’immunodéficience humaine, exact et approché. Pour chaque vecteur propre calculé de manière exacte, on donne le recouvrement du vecteur approché le plus similaire obtenu avec l’approximation RTB, en mettant un (les croix) ou deux (les signes ”plus”) acides aminés dans chaque bloc.

Ceci dit, en dépit de cette nette surestimation des valeurs propres, les vecteurs propres de \mathbf{H} , tels qu’ils sont calculés avec l’approximation RTB, s’avèrent être très proches des vecteurs exacts. Dans le cas du vecteur associé au mode de vibration de plus basse fréquence de la protéase du VIH-1, les différences sont minimales, que le calcul soit effectué en mettant un ou deux résidus d’acide aminé dans chaque bloc (voir Fig. 6).

Pour quantifier le degré de ressemblance entre deux vecteurs i et j , on peut calculer leur recouvrement, I_{ij} , défini ci-après comme étant le carré de leur produit scalaire :

$$I_{ij} = \left(\sum_k^{3N} a_{ki} a_{kj} \right)^2 \quad (6)$$

où l’on a supposé que les vecteurs i et j sont de norme unitaire. Notez que l’on a, puisque les vecteurs propres de \mathbf{H} sont par définition orthogonaux :

$$\sum_i^{3N} I_{ij} = \sum_j^{3N} I_{ij} = 1$$

Dans le cas de deux ensembles de vecteurs propres, comme l’ordre des vecteurs peut être différent, pour chaque vecteur d’un des ensembles, on détermine quel est le vecteur de l’autre ensemble pour lequel le recouvrement est maximum. Dans la Fig. 7, on compare ainsi les vecteurs propres de \mathbf{H} exacts et approchés, tels qu’ils sont calculés avec l’approximation RTB. On constate que cette approximation ne permet de déterminer précisément que les modes de vibration de la protéase du VIH-1 de fréquence inférieure à 10 cm^{-1} . Nous verrons par la suite qu’il s’agit de modes particulièrement ”robustes” [31].

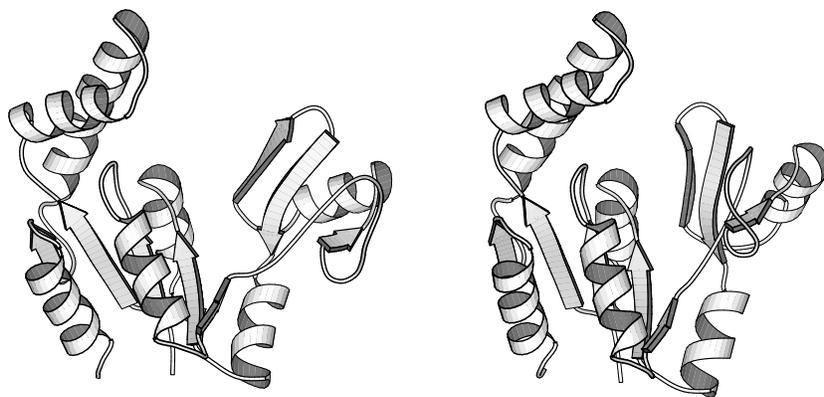


FIG. 8 – Changement de conformation de la guanylate kinase de *S. Cerevisiae*, de la forme ouverte (à gauche) à la forme fermée (à droite).

4 Le lien avec les mouvements fonctionnels

Pour mettre en évidence ce lien, on calcule cette fois le recouvrement (voir eqn. 6) entre chacun des modes de vibration et le vecteur $\vec{\Delta r}$, qui est relié à la façon dont se déplacent les atomes lorsque la protéine change de conformation :

$$\vec{\Delta r} = \vec{r}_b - \vec{r}_a \quad (7)$$

où \vec{r}_a et \vec{r}_b sont, respectivement, les vecteurs position des atomes dans les conformations "a" et "b", après superposition de ces deux conformations par la méthode des moindres carrés.²⁰ Notez qu'en procédant ainsi on suppose que, pour aller de la conformation "a" à la conformation "b", les atomes se déplacent en ligne droite, ce qui n'est manifestement pas exact.²¹

Malgré cela, on a observé dans de nombreux cas qu'il existe au moins un mode de vibration qui ressemble beaucoup au changement de conformation tel qu'il est connu expérimentalement, par cristallographie des rayons X le plus souvent. Après Robert W. Harrison qui constata, comme déjà rappelé précédemment, que dans le cas de l'hexokinase "for the native open form there were (at least) two modes that had strong components along the conformation change" [13], des comparaisons plus quantitatives furent effectuées pour les cas de la citrate synthase dimérique[24] et de l'hémoglobine[26]. Dans le premier cas, avec Osni Marques, nous avons observé que le mode de plus basse fréquence de la forme fermée (obtenue en présence d'un des ligands) a un recouvrement de 0.25 avec le mouvement d'ouverture de la citrate synthase tandis que, dans le second cas, David Perahia et Lilianne Mouawad ont observé que le deuxième mode de plus basse fréquence de la forme T de l'hémoglobine (en absence d'oxygène) a un recouvrement de près de 0.7 avec la mouvement qui se produit lorsque l'hémoglobine passe en forme R (en présence d'oxygène).

Pour illustrer ce que signifie ce type de résultat, prenons le cas de la guanylate kinase de *S. Cerevisiae*. En présence d'un de ses ligands, la guanosine-5'-monophosphate, elle change de conformation, passant d'une forme "ouverte" (code PDB 1EX6) à une forme "fermée"

²⁰Avec par exemple Profit, qui est disponible sur la Toile : <http://www.bioinf.org.uk/software/profit/>

²¹Pour éviter d'avoir à faire cette supposition, il faudrait connaître le chemin de transition qui permet de passer d'une conformation à l'autre.

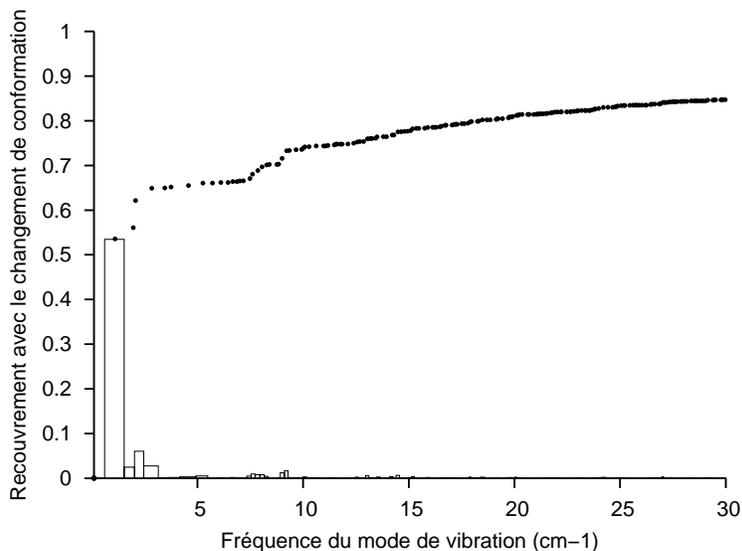


FIG. 9 – Comparaison du changement de conformation de la guanylate kinase de *S. Cerevisiae* et de chacun des modes de vibration de la forme ouverte, calculés avec une approche standard. Histogramme : recouvrement. Cercles noirs : recouvrement cumulé.

(code PDB 1EX7; voir Fig. 8). Cela correspond à un déplacement de 3.6\AA de chacun de ses C_α , en moyenne,²² et de plus de 10\AA pour les acides-aminés 40-43. Si on calcule les modes de vibration de la forme ouverte et qu'on détermine le recouvrement de chacun d'eux avec le changement de conformation, on constate que celui du mode de plus basse fréquence est supérieur à 0.5 (voir Fig. 9) et que seulement six modes, parmi les 5340 que compte cette protéine, ont un recouvrement avec le changement de conformation de plus de 0.01. De fait, les quatre modes de plus basse fréquence ont un recouvrement cumulé d'environ 0.65, ce qui signifie qu'une combinaison linéaire de déformations de la protéine le long de ces quatre modes là suffit pour décrire très correctement son changement de conformation. Le mode de plus basse fréquence donne d'ailleurs déjà une très bonne idée du type de mouvement qui se produit (voir Fig. 10).

Ce type de résultat peut paraître paradoxal. En effet, utiliser la théorie des modes normaux dans le cas d'une protéine, cela signifie qu'on étudie son comportement dynamique après avoir fait l'hypothèse qu'il s'agit d'un solide qui oscille autour de sa position d'équilibre. Ce qui ne devrait pas être un bon point de départ pour essayer de deviner ses "changements d'état". De plus, avant de calculer les modes de vibration d'une protéine, il faut la déformer significativement, de façon à ce qu'elle soit dans une configuration d'équilibre (voir eqn. 2). En d'autres termes, il faut minimiser son énergie. Dans le cas de la guanylate kinase, les atomes C_α se déplacent d'environ 1.3\AA durant ce processus, ce qui n'est pas négligeable, comparé à l'amplitude de son changement de conformation (3.6\AA). Par ailleurs, on sait qu'à température ordinaire une protéine peut explorer un nombre astronomique de minima d'énergie[32]. De fait, d'infimes modifications du protocole suivi pour effectuer la minimisation d'énergie permettront d'atteindre des minima différents. On doit donc en conclure, au moins en ce qui concerne leurs modes de basse fréquence, que tous ces minima doivent se ressembler énormément[33].

²² C_α -r.m.s.

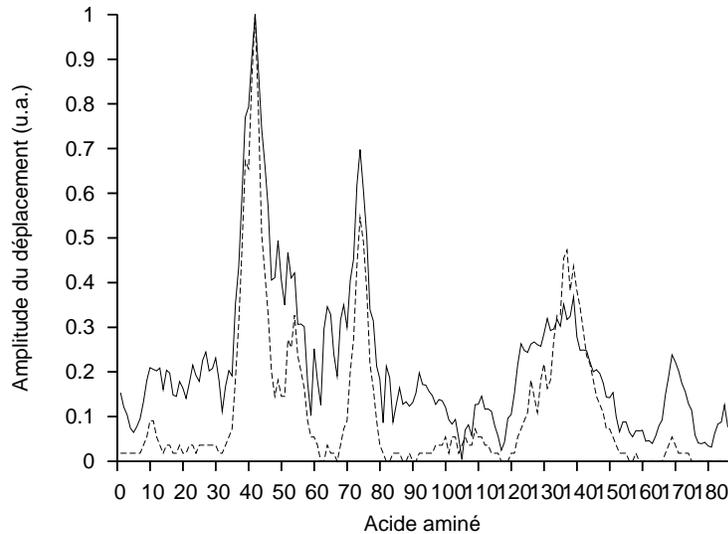


FIG. 10 – Comparaison du changement de conformation de la guanylate kinase de *S. Cerevisiae* (ligne continue) et du mode de plus basse fréquence de la forme ouverte (ligne tiretée). Les deux ensembles de déplacements ont été normalisés.

5 Les modèles de réseau élastique

5.1 L'idée de Monique Tirion

Puisqu'on peut obtenir des résultats qui semblent avoir un sens biologique en considérant que le comportement dynamique d'une protéine est proche de celui d'un solide, plus précisément, en admettant que son énergie a une forme quadratique (voir eqn. 3), on peut se demander ce qui se passe si on simplifie encore plus la description du paysage énergétique qui lui est accessible à température ordinaire.

Le premier pas dans cette direction a été fait par Monique Tirion[18], qui a proposé de remplacer les interactions "non-liées", c'est-à-dire les interactions électrostatiques et de Lennard-Jones, ainsi que celles associées aux angles dièdres, par des termes de la forme suivante :²³

$$E_{nb} = \frac{1}{2} \sum_{d_{ij}^0 < R_c} k_{enm} (d_{ij} - d_{ij}^0)^2 \quad (8)$$

où d_{ij} est la distance entre les atomes i et j , d_{ij}^0 étant leur distance dans la structure étudiée (la structure cristallographique le plus souvent) tandis que R_c est une distance de coupure.²⁴ k_{enm} est par ailleurs supposé ne pas dépendre du type des atomes en interaction, ce qui signifie que cette constante se sert qu'à spécifier les unités énergétiques employées. R_c est donc le seul vrai paramètre du modèle. Néanmoins, comme l'a montré M. Tirion, ce modèle permet de reproduire la partie basse fréquence du spectre de l'actine, ainsi que les fluctuations des positions de ses atomes, tels qu'ils sont calculés avec l'approximation harmonique et une fonction énergie empirique "standard"[18].

²³Les termes énergétiques correspondant aux liaisons covalentes et aux angles de valence étaient conservés de manière implicite puisque M. Tirion appliquait l'approximation harmonique en passant par un jeu de coordonnées internes ne comprenant que les angles dièdres, à l'instar de ce qui avait été fait dans les groupes de N. Go[6, 8] et M. Levitt[9, 10].

²⁴Un "cutoff".

Le modèle proposé par M. Tirion a un avantage majeur par rapport aux modèles utilisés jusque là : par définition, l'énergie de la structure étudiée est minimale ($E_{nb} = 0$). On peut donc calculer les modes de vibration de n'importe quelle structure, sans avoir au préalable à la déformer de manière significative. Ce modèle a un autre avantage, de nature pratique cette fois : plus besoin d'avoir été "initié" aux subtilités des langages de commandes des (gros) programmes de modélisation moléculaire pour pouvoir calculer les modes normaux de basse fréquence d'une protéine, puisqu'il n'y a plus besoin d'avoir accès aux (nombreux) paramètres des fonctions énergétiques dont ils font usage. De plus, les calculs effectués sont plus aisément reproductibles, puisqu'ils sont tous faits en partant de la même structure (la structure cristallographique, par exemple).

5.2 L'apport d'Ivet Bahar

Pour tenter d'aller encore plus loin, Ivet Bahar, Burak Erman et leurs collaborateurs ont proposé deux simplifications supplémentaires[34] du modèle de M. Tirion. Tout d'abord, ils ont remplacé les interactions interatomiques par des interactions entre résidus d'acides aminés. En effet, si l'on peut faire l'hypothèse que les interactions interatomiques ne dépendent pas du type des atomes impliqués, c'est qu'elles sont sans doute moyennées et que les interactions "utiles" se font, de fait, à une échelle supérieure. Ensuite, puisque le modèle de M. Tirion revient à tendre des ressorts entre atomes situés à moins de R_c Ångströms les uns des autres, ils ont décidé d'étudier la matrice d'adjacence du réseau ainsi défini, c'est-à-dire la matrice dont l'élément ij est tel que $a_{ij} = -1$, si les acides aminés i et j sont en contact, a_{ij} étant nul s'ils ne le sont pas, l'élément diagonal a_{ii} étant défini ainsi :²⁵

$$a_{ii} = - \sum_{j \neq i} a_{ij} \quad (9)$$

Les vecteurs propres et les valeurs propres de cette matrice $N \times N$ permettent de calculer l'amplitude des déplacements des N résidus de la protéine étudiée. La corrélation avec les valeurs expérimentales, obtenues à partir de celles des facteurs B isotropes (les facteurs de Debye-Waller), s'avère être souvent très bonne[34]. De fait, les meilleures corrélations obtenues à ce jour entre valeurs calculées et valeurs expérimentales l'ont été en utilisant des variantes de ce modèle[35], dit du "réseau gaussien" (Gaussian Network Model), et non pas en utilisant des modèles bien plus détaillés, tous-atomes par exemple. Cette supériorité du modèle le plus simple proposé à ce jour pour décrire une protéine à l'échelle de l'acide aminé reste à expliquer. On la suppose bien sûr provisoire (voir partie 5.7)...

5.3 La synthèse de Konrad Hinsen

Konrad Hinsen a ensuite proposé[36] de conserver la représentation "à gros grains" proposée par I. Bahar et B. Erman, mais en se contentant d'étendre la relation (8) à toutes les paires d'atomes retenus (les carbones α le plus souvent), y compris ceux qui sont liés covalamment.

Deux versions de ce modèle, dit du "réseau anisotrope" (Anisotropic Network Model), selon la terminologie proposée par I. Bahar[37], sont couramment utilisées. Dans la première,

²⁵Ce qui différencie la matrice d'adjacence de la matrice de contact.

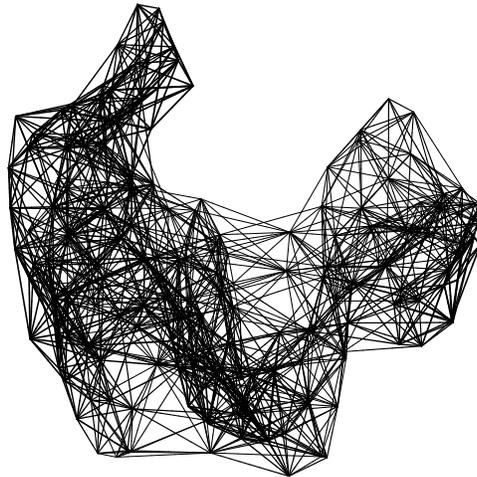


FIG. 11 – Le réseau élastique construit pour la forme ouverte de la guanylate kinase de *S. Cerevisiae*. Chaque ligne représente un ressort établi entre deux C_α situés à moins de 10\AA l’un de l’autre. L’orientation est la même que dans la Fig. 8.

k_{enm} est une fonction exponentielle de la distance entre paires d’atomes[36]. Dans la seconde, comme proposé initialement par M. Tirion, k_{enm} est constant et une distance de coupure permet de définir la liste des atomes qui sont en interaction[37, 38]. En pratique, pour éviter d’avoir des groupes d’atomes isolés, ou bien insuffisamment liés au reste du réseau, ce qui entraîne l’apparition de modes de basse fréquence sans grande signification physique, voire de modes de fréquence nulle supplémentaires, la distance de coupure choisie doit être assez grande. D’une étude à l’autre, elle varie de 8 [38] à 16\AA [39] (alors que l’on considère que les distances typiques entre acides aminés en contact, entre leurs centres de gravité par exemple, sont plutôt de l’ordre de $6-7\text{\AA}$ [40]).

Notez que dans le cas de cette seconde version, les éléments de la matrice qu’il faut diagonaliser pour obtenir les modes de basse fréquence d’un système ont une expression qui n’est pas beaucoup plus compliquée que ceux de la matrice d’adjacence. Par exemple, pour l’élément a_{kl} correspondant aux coordonnées x et y des atomes i et j :

$$a_{kl} = -k_{enm} \frac{(x_i - x_j)(y_i - y_j)}{\sqrt{m_i m_j} d_{ij}^2}$$

les éléments diagonaux étant obtenus de la même façon que ceux de la matrice d’adjacence (eqn. 9). La Figure 11 montre à quoi ressemble le réseau élastique correspondant, dans le cas de la forme ouverte de la guanylate kinase, lorsque $R_c = 10\text{\AA}$. C’est ce type de modèles, finalement assez proches de ceux dits en ”éléments finis”[12], courants dans le domaine de l’analyse des propriétés dynamiques de structures rigides (voitures, ponts, etc) que nous allons considérer plus en détails dans la suite de ce texte.

5.4 Elnémo

Ces modèles de réseau élastique sont faciles à implémenter dans des programmes eux-mêmes simples à utiliser,²⁶ puisque le nombre de paramètres requis est minimal. Cependant, le temps qu’il faut pour installer un programme, pour se familiariser avec ses ”sorties”, pour commencer à avoir confiance dans les résultats obtenus, etc., peuvent

²⁶Les miens sont disponibles sur la Toile : <http://ecole.modelisation.free.fr/modes.html>.

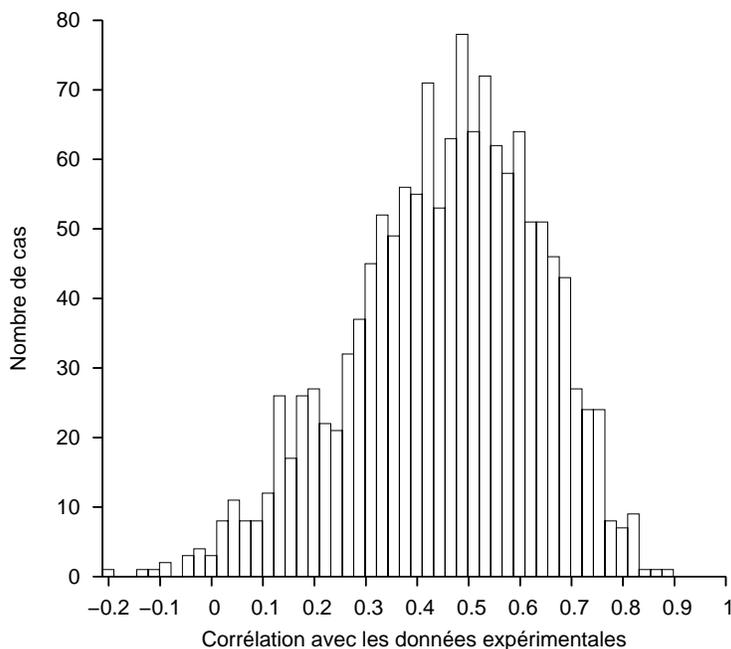


FIG. 12 – Statistique des corrélations observées entre fluctuations atomiques calculées et observées (facteurs B), pour un ensemble de 1405 protéines.

être des facteurs inhibants. C’est pourquoi nous avons décidé, avec Karsten Suhre, de développer un site Web qui puisse permettre à l’utilisateur de calculer et de visualiser les modes de vibration de la protéine de son choix. Ce site, appelé Elnémo (pour Elastic Network Model)²⁷ fut installé à l’Institut de Génomique Structurale, à Marseille²⁸ (<http://www.igs.cnrs-mrs.fr/elnetmo/>). Il effectue plusieurs milliers de calculs chaque année, et ce depuis son ”ouverture”, fin 2003.²⁹ Ces calculs peuvent être faits avec le modèle élastique décrit ci-dessus (un atome par résidu, une distance de coupure) mais aussi avec un modèle élastique tous-atomes, très proche de celui proposé par M. Tirion. Dans ce dernier cas, les modes de vibration sont calculés avec l’approximation RTB[41].

Par la suite, plusieurs groupes ont mis en ligne leur propre site de calcul de modes normaux. Il y a désormais un grand choix : Elnémo[41], donc, mais aussi WEBnm[42], NOMAD-ref[43], ANM-Webserver[44], etc.³⁰

5.5 Comparaison avec les données expérimentales

5.5.1 Les facteurs B

Dans le cadre de l’approximation harmonique, on peut calculer de manière analytique les fluctuations des coordonnées de chaque atome autour de sa position d’équilibre. On a en

²⁷L’accent grave, ce doit être pour la ”french touch”.

²⁸Une version ”miroir” a été installée récemment à l’institut Joliot-Curie : <http://elnemo.ens-lyon.fr>.

²⁹Karsten Suhre fut le ”développeur” du site initial, mon rôle ayant consisté à adapter mes programmes de façon à pouvoir, notamment, diagnostiquer la majorité des erreurs commises par les utilisateurs (fichier PDB vide, etc.). Je suis en train de mettre en place le site miroir, que je compte faire évoluer indépendamment du site initial.

³⁰Voir ma page de liens, où j’espère les avoir tous recensés : <http://yh.sanejouand.com>.

effet, par exemple pour la coordonnée x de l'atome i :

$$\langle \Delta x_i^2 \rangle = k_B T \sum_k \frac{a_{ik}^2}{4\pi^2 m_i \nu_k^2} \quad (10)$$

Or ces fluctuations sont responsables du "flou" observé lorsqu'on cherche à déterminer la position des atomes dans une maille cristalline par cristallographie des rayons X.³¹ Dans le cas de la plupart des structures de protéines, on quantifie cela via un facteur associé à chaque atome, le facteur B, pour lequel on a :³²

$$B_i = \frac{8\pi^2}{3} \langle \Delta x_i^2 + \Delta y_i^2 + \Delta z_i^2 \rangle \quad (11)$$

La Figure 12 montre la statistique des valeurs des corrélations observées entre fluctuations atomiques calculées (avec l'eqn. 10) et observées (à partir de l'eqn. 11). Les calculs furent effectués pour un ensemble de 1405 protéines bien résolues ($R < 1.8\text{\AA}$), non redondantes (moins de 50% d'identité entre paires de séquences), en ne prenant en compte que les C_α ($R_c=10\text{\AA}$). En moyenne, la corrélation est de 0.45 ± 0.18 , mais des valeurs proches de 0.9 sont observées dans certains cas. En ne prenant en compte que des structures à très haute résolution ($R < 1.0\text{\AA}$) la corrélation peut atteindre une valeur de 0.68 ± 0.11 (pour 83 cas, avec $R_c=16\text{\AA}$ [39]). Ceci suggère que les facteurs B pourraient être déterminés directement à partir des coordonnées, lors de la résolution d'une structure cristallographique. L'efficacité de cette approche a été démontrée il y a une quinzaine d'années[45, 46, 47, 48], sans pour autant avoir été utilisée depuis, à ma connaissance, par les cristallographes...

Des fluctuations de la position des atomes sont aussi obtenues lorsqu'une structure est déterminée par résonance magnétique nucléaire (RMN). En effet, cette technique permet d'obtenir un ensemble de conformations compatibles avec les distances inter-atomiques déterminées expérimentalement. Cristian Micheletti et ses collaborateurs ont montré que les fluctuations des positions des atomes dans cet ensemble de conformations étaient aussi souvent fortement corrélés avec celles calculées à partir de l'eqn. 10[49].

5.5.2 Les changements de conformation

Dans l'article dans lequel elle présentait le premier des modèles de réseau élastique pour les protéines,³³ M. Tirion concluait ainsi : "Tests performed on a periplasmic maltodextrin binding protein (MBP) indicate that the slowest modes do indeed closely map the open form into the closed form", annonçant qu'il s'agissait d'un travail "in preparation"[18]. Quelques années plus tard, l'article suivant de M. Tirion se faisant attendre,³⁴ avec Florence Tama, nous avons confirmé le résultat brièvement décrit par M. Tirion, en utilisant

³¹Le désordre cristallin, ainsi que les phonons, y contribuent aussi.

³²Lorsque le nombre de taches de diffraction obtenues le permet, c'est-à-dire lorsque la qualité du cristal est suffisamment bonne, on décrit la fluctuation de la position d'un atome via une matrice 3×3 symétrique. On obtient ainsi des facteurs B dits anisotropes.

³³Nobuhiro Go, à qui l'on doit à la fois les "modèles de Go"[50] et parmi les tout premiers calculs de modes normaux de protéines[6] aurait sans doute aussi du être le premier à étudier ce type de modèles. Dans les "modèles de Go", on ne tient en effet compte que des contacts observés dans la structure native. N. Go n'a cependant publié que des études de versions "sur réseau" de son modèle. Il soulignait pourtant dès 1983 que "in modes with frequencies below 120 cm^{-1} the protein molecule behaves like a continuous elastic body"[8].

³⁴M. Tirion n'a publié aucun autre article depuis, à ma connaissance.

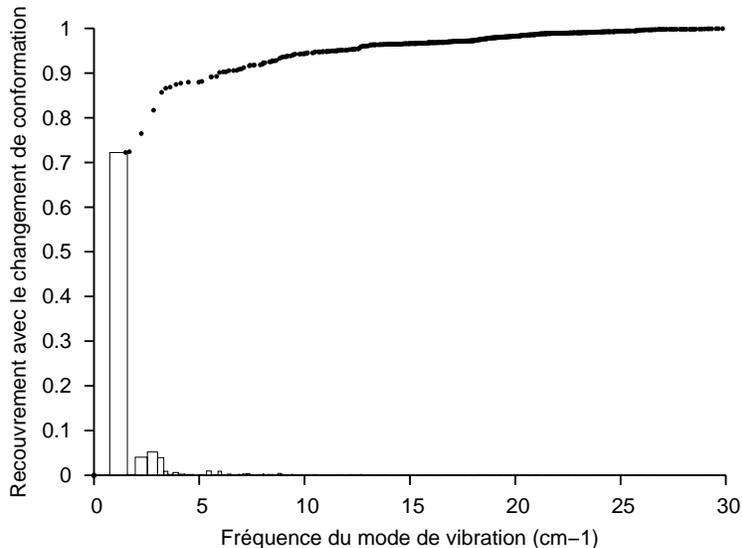


FIG. 13 – Comparaison du changement de conformation de la guanylate kinase de *S. Cerevisiae* et de chacun des modes de vibration de la forme ouverte, calculés avec un modèle de réseau élastique. Histogramme : recouvrement. Cercles noirs : recouvrement cumulé.

toutefois un modèle plus simple que le sien (que les C_α , $R_c=8\text{\AA}$). Dans le cas de la MBP, le recouvrement entre le changement de conformation (codes PDB 1OMP et 1ANF) et son mode de plus basse fréquence est en effet élevé : 0.74[38]. Par ailleurs, en considérant au total une vingtaine de cas, nous avons constaté que le lien mode de vibration - changement de conformation devenait évident (recouvrement de plus de 0.25) dès lors que ce dernier est suffisamment collectif, c'est-à-dire qu'il implique un nombre suffisamment grand d'atomes[38].

Avec Marc Delarue, nous avons ensuite vérifié, dans le cas d'un ensemble de protéines de fonction semblable, de forme similaire mais de repliements différents, des polymérases ADN-dépendantes, que leurs changements de conformation peuvent être très correctement décrits comme étant une combinaison linéaire d'un tout petit nombre de leurs modes de basse fréquence[51], un résultat confirmé aussi par l'équipe de Mark Gerstein, lors d'une étude de près de 4000 cas[52].

Rétrospectivement, on s'aperçoit que ce résultat n'est qu'une conséquence logique de ceux qu'avaient obtenus K. Hinsen. Ce dernier avait en effet montré qu'on peut utiliser les modes normaux calculés avec un modèle de réseau élastique pour identifier les domaines structuraux des protéines[36, 53]. Or on sait depuis longtemps que, lorsqu'on étudie les structures associées au mouvement fonctionnel de grande amplitude d'une protéine, dans la plupart des cas, on trouve qu'il s'agit d'un mouvement de domaines[54].

5.6 Comparaison avec les modes normaux standards

Si l'on reprend le cas de la guanylate kinase de *S. Cerevisiae* et que l'on calcule cette fois les modes de vibration de sa forme ouverte avec un modèle de réseau élastique ($R_c=10\text{\AA}$), on constate que ceux-ci permettent de décrire son changement de conformation fonctionnel (au moins) aussi bien que ceux calculés précédemment, avec un modèle bien plus détaillé (comparez les Fig. 9 et 13). On peut en conclure que les modes concernés doivent être très

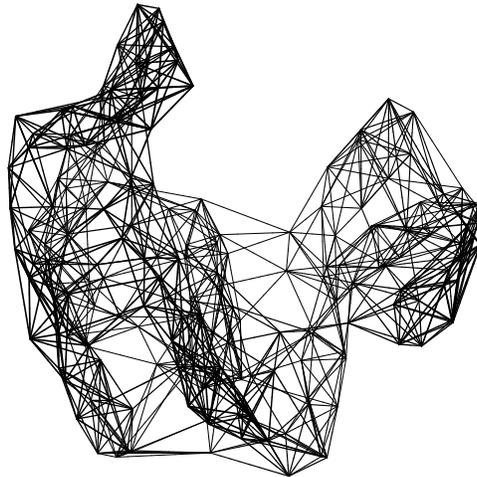


FIG. 14 – Un autre type de réseau, construit pour la forme ouverte de la guanylate kinase de *S. Cerevisiae*. Chaque ligne représente un ressort. La plupart des C_α sont liés à dix de leurs voisins.

similaires, ce qui signifie que déterminer les modes de vibration de basse fréquence d'une protéine permet d'extraire une information qui reste présente lorsque la description devient grossière[55].³⁵ Etant donnée la façon dont sont d'ordinaire construits les modèles de réseau élastique, cette information peut être de deux types : soit il s'agit de la distribution des masses de la protéine dans l'espace, soit il s'agit de la répartition des portions rigides à l'intérieur de la structure.

Pour trancher entre ces deux hypothèses extrêmes, avec Samuel Nicolay,³⁶ nous avons conçu un modèle de réseau élastique dans lequel, pour chaque atome, le nombre d'interactions est (à peu près) constant. En conséquence de quoi la rigidité à l'intérieur de la structure devient (presque) uniforme[31]. Ce modèle est construit ainsi. Dans un premier temps, toutes les paires d'atomes sont "liées" entre elles. Puis le lien entre les deux atomes les plus lointains est retiré. Puis celui entre les deux atomes les plus lointains encore liés et ainsi de suite, si ce n'est que lorsque, pour la paire d'atomes considérée, l'un ou l'autre ont un nombre de liens restant égal à un nombre donné, N_v , ce lien est conservé. Bien sûr, ce faisant, à la fin du processus, certains atomes ont plus d'interactions que ce qui est souhaité. Par exemple, dans le cas de la guanylate kinase, si l'on souhaite que N_v soit égal à dix, dans le réseau ainsi construit (voir Fig. 14) on observe que chaque C_α est en interaction avec 10.1 ± 0.5 de ses voisins, en moyenne. Notez qu'avec un critère de distance, le nombre d'interactions de chaque C_α est en général bien plus élevé que cela. En effet, avec $R_c = 10\text{\AA}$, il y en a en moyenne 15.9 ± 5.0 .

Si l'on compare les modes de vibration de basse fréquence de deux réseaux construits, l'un avec un critère de distance (par exemple, $R_c = 10\text{\AA}$) l'autre en fixant le nombre (minimal) d'interactions de chaque C_α (par exemple, $N_v = 10$), on constate que certains d'entre eux sont peu "perturbés" par ce changement de critère[31]. La Figure 15 montre les meilleurs recouvrements obtenus, dans le cas de la forme ouverte de la guanylate kinase. On voit que les quatre modes de plus basse fréquence sont presque parfaitement conservés (recouvrement de 0.85 et plus) et, si l'on se reporte à la Figure 13, on constate qu'il s'agit des modes qui sont impliqués dans le changement de conformation. Dans ce

³⁵On parle désormais de "modèles à gros grains" (pour "coarse grained models").

³⁶Alors en thèse sous la direction d'Alain Arneodo.

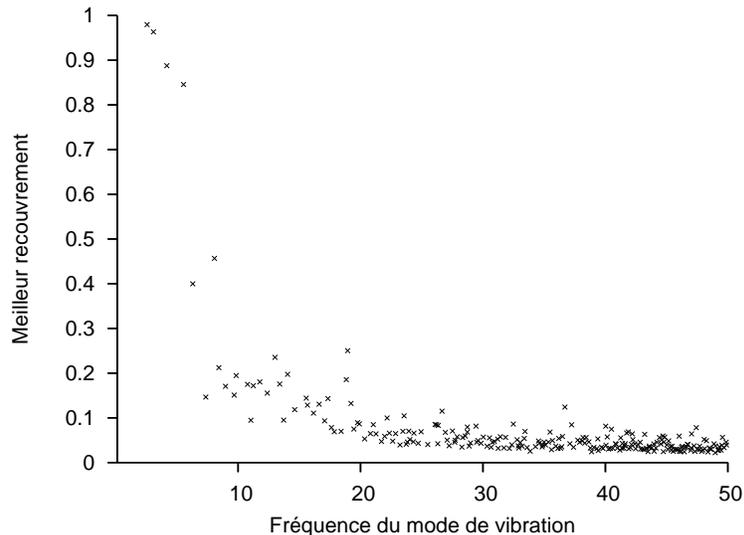


FIG. 15 – Calcul des modes de vibration de plus basse fréquence de la guanylate kinase avec deux réseaux élastiques différents. Pour chaque vecteur propre calculé avec un critère de distance ($R_c = 10\text{\AA}$), on donne le recouvrement du vecteur approché le plus similaire obtenu en fixant le nombre (minimal) d’interactions ($N_v = 10$).

cas-là, l’ordre de ces modes n’est d’ailleurs lui aussi pas affecté par le changement de type de représentation.

Avec S. Nicolay, nous avons étudié ainsi une vingtaine de cas et nous avons montré que, dès lors que le changement de conformation est de suffisamment grande amplitude (C_α -r.m.s. de plus de 2\AA), alors il existe de tels modes ”robustes” (jusqu’à onze, dans le cas de l’ADN polymérase du bactériophage RB69). De plus, parmi ceux-ci se trouvent la (grande) majorité des modes impliqués dans le changement de conformation[31]. Nous avons par ailleurs obtenu des résultats similaires en comparant dans une demi-douzaine de cas les modes obtenus avec deux représentations à l’échelle atomique : l’une, associée à une fonction énergie standard (celle de CHARMM[21]), l’autre correspondant au modèle initialement proposé par M. Tirion[18].

L’approche que nous avons proposée permet donc, d’une part, d’identifier un petit ensemble de modes potentiellement importants pour la fonction d’une protéine, d’autre part, de conclure que la distribution des masses dans l’espace est bien l’information pertinente, c’est-à-dire celle que le calcul des modes de vibration permet d’extraire, au moins en ce qui concerne le lien avec les mouvements fonctionnels.

5.7 Vers des modèles mixtes

En fait, quand on compare les Figures 9 et 13, il est tentant de conclure que c’est le modèle de réseau élastique, pourtant de loin le plus simple, qui permet d’obtenir les résultats les plus convaincants : le mode de plus basse fréquence a en effet un recouvrement de plus de 0.7 avec le changement de conformation (contre 0.5 dans le cas du modèle standard). De plus, quatre modes seulement suffisent pour décrire près de 90% du changement de

conformation alors qu’avec le modèle standard les quatre modes les plus ”impliqués” dans le changement de conformation ne peuvent en décrire que 65% environ.

Ce résultat n’est pas atypique : avec Florence Tama, nous avons effectué ce genre de comparaison dans le cas de cinq autres protéines et, quatre fois sur cinq, nous avons constaté que le mode qui ressemble le plus au changement de conformation a un recouvrement plus important lorsqu’il a été calculé avec un modèle de réseau élastique[38]. De même, avec Samuel Nicolay, dans cinq autres cas, nous avons constaté que les modes robustes obtenus à partir des modes de vibration d’un modèle standard ne permettent pas de décrire le changement de conformation aussi bien que ceux obtenus à partir des modes de vibration d’un modèle de réseau élastique[31]. Cet effet est d’ordinaire faible, mais il semble être (presque) systématique, ce qui a intrigué Martin Karplus qui m’a suggéré d’étudier cela plus en détail,³⁷ dans l’esprit de l’un de ses premiers articles consacrés aux modes de vibration des protéines[7].

Pour ce faire, une étude d’une dizaine de cas est menée. Ceux-ci sont spécifiés dans la Table 1, tout comme l’amplitude de la déformation due à la minimisation d’énergie préalable, nécessaire lorsqu’il s’agit de calculs de modes normaux standards. Dans ce cas, les calculs furent effectués avec un modèle de solvation implicite (EEF1.1[29]) et la minimisation d’énergie fut faite en deux temps. Tout d’abord, avec les positions des C_α fixes. Puis sans aucune contrainte, en effectuant 10000 pas de l’algorithme ”Steepest descent” suivis d’autant de pas que nécessaire (jusqu’à parfois près de 100000) de l’algorithme Newton-Raphson inclus dans CHARMM[21] (ABNR), ce, jusqu’à ce que les forces qui s’exercent sur chacun des atomes soient d’environ 10^{-4} kcal/mole.Å.³⁸ Comme on le voit dans la Table 1, ce protocole entraîne des déformations qui vont de 0.7Å, dans le cas de la protéase du VIH-1 sous forme dimérique, jusqu’à 3.2Å, dans le cas de l’adénylate kinase. Parfois, l’amplitude de la déformation est de l’ordre de grandeur de celle du changement de conformation, comme dans les cas de l’ADN polymérase β ou de la protéase du VIH-1 sous forme monomérique, ce qui suggère qu’elle pourrait être la cause principale de la ”dégradation” observée, puisque lorsqu’on se sert d’un modèle de réseau élastique on part, a priori, de la structure expérimentale elle-même.

Pour vérifier cette hypothèse, pour chacun des dix cas considérés, trois calculs de modes normaux furent effectués : un standard, à partir de la structure relaxée (celle obtenue à l’issue de la minimisation d’énergie) et deux avec un modèle de réseau élastique ne prenant en compte que les C_α ($R_c=10\text{Å}$), le premier étant effectué en partant de la structure cristallographique, le second en partant de la structure relaxée. Ensuite, la capacité de chacun de ces ensembles de modes à décrire un changement de conformation fut évaluée en déterminant n_{eff} , le nombre de modes ”nécessaires” (utiles), de la façon suivante[31] :

$$n_{eff} = \exp\left(-\sum_{i=1}^{N_m} \alpha I_i \ln(\alpha I_i)\right) \quad (12)$$

où N_m est le nombre de modes de basse fréquence pris en compte ($N_m = 100$ ci-après), et où α est un terme de normalisation tel que :

$$\sum_{i=1}^{N_m} \alpha I_i = 1$$

³⁷Travail en cours.

³⁸Ce qui permet d’obtenir six fréquences vraiment (numériquement) nulles.

Protéine	Taille	Codes PDB	Amplitude (Å)
Protéase du VIH-1 (mono)	99	1HHP, 1A30	1.1 (1.9)
Pyrophosphokinase	158	1HKA, 1Q0N	1.9 (1.3)
Lysozyme T4	162	178L, 158L	4.1 (0.8)
Erythropoïétine	166	1BUY, 1EER	4.1 (1.5)
Guanylate kinase	186	1EX6, 1EX7	3.6 (1.2)
Protéase du VIH-1 (dim)	198	1HHP, 1A30	2.0 (0.7)
Adénylate kinase	218	4AKE, 1AKE	7.1 (3.2)
Glutamine bp	224	1GGG, 1WDN	5.3 (1.4)
LAO bp	238	2LAO, 1LST	4.7 (1.4)
ADN polymérase β	335	1BPX, 1BPY	2.8 (2.5)

TAB. 1 – Amplitude des changements de conformation étudiés. Le premier code PDB est celui de la structure étudiée, tandis que le second correspond à celui de la structure ayant subi le changement de conformation. Lorsque les modes sont calculés de manière standard, l'énergie de la structure est au préalable minimisée. L'amplitude de la déformation correspondante est donnée entre parenthèses (C_α -r.m.s.). La protéase du VIH-1 est étudiée soit sous forme monomérique, soit sous forme dimérique.

tandis que I_i est le recouvrement (voir eqn. 6) entre le mode i et $\vec{\Delta r}$, le vecteur décrivant le changement de conformation (voir eqn. 7).

Cette mesure est moins sensible aux petits changements dans le détail de la structure étudiée que la valeur du plus grand recouvrement observé entre un mode et le changement de conformation, mesure que nous avons eu tendance à privilégier précédemment[38, 51]. En effet, si, par exemple, un changement de conformation peut être bien décrit par une combinaison linéaire de deux modes de fréquence voisine, il suffira de modifier de manière minimale la structure étudiée pour que la valeur du plus grand recouvrement double, ce qui ne sera pas le cas en ce qui concerne n_{eff} , c'est-à-dire l'estimation du nombre de modes nécessaires à la description. Notez cependant que ces deux mesures donnent le plus souvent des indications similaires, comme l'illustre la Figure 16 qui regroupe les résultats obtenus pour les dix changements de conformation pris en compte et les trois ensembles de modes calculés dans un premier temps pour chacune des protéines.

Dans la Figure 17, pour chaque changement de conformation de la Table 1, la valeur de n_{eff} obtenue avec un modèle de réseau élastique est donnée en fonction de celle obtenue avec une approche standard, une flèche indiquant son évolution lorsqu'on passe de la structure cristallographique à la structure relaxée. Tout d'abord, on constate que dans chacun des dix cas considérés, il faut moins de modes pour décrire le changement de conformation lorsque ceux-ci ont été calculés avec un modèle de réseau élastique et à partir de la structure cristallographique (les cercles noirs sont tous situés sous la diagonale). De plus, à une exception près, d'ailleurs peu significative, il en faut plus lorsqu'on part non pas de la structure cristallographique mais de la structure relaxée (les cercles vides sont au dessus des cercles noirs et les flèches pointent donc toutes vers le haut). Notez que cet effet peut être très important, comme dans les cas de l'ADN polymérase β ou de la protéase du VIH-1 sous forme monomérique. Ceci va dans le sens de notre hypothèse initiale, puisqu'il s'agit des deux cas pour lesquels l'amplitude de la déformation due à la minimisation

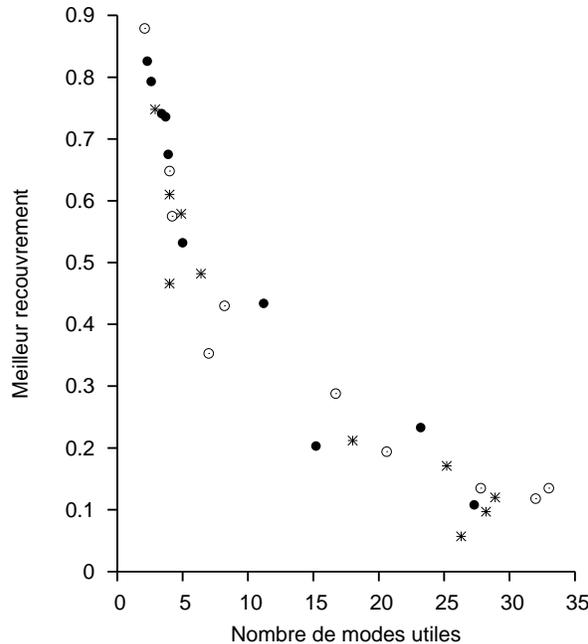


FIG. 16 – Meilleur recouvrement observé entre un mode et un changement de conformation, en fonction du nombre de modes nécessaires (”utiles”) à la description de ce dernier. Dix changements de conformation différents ont été considérés. Les modes normaux ont été obtenus soit avec un modèle de réseau élastique, à partir de la structure cristallographique (cercles noirs) ou d’une structure dont l’énergie a été minimisée (cercles vides), soit avec une approche standard (étoiles).

d’énergie est du même ordre de grandeur que celle du changement de conformation (voir Table 1). Notez qu’en ce qui concerne l’ADN polymérase β , on passe d’une description à ”quelques modes” ($n_{eff} = 5.0$),³⁹ proche de ce qui a été montré précédemment dans le cas de la guanylate kinase ($n_{eff} = 3.7$; voir Fig. 13) à une description beaucoup plus diffuse ($n_{eff} = 20.6$), qui ressemble plus à ce que l’on attend lorsqu’on projette un mouvement collectif sur une base de coordonnées dont certaines sont collectives (voir Fig. 18).

La déformation due aux minimisations d’énergie semble donc bien être responsable de la (souvent très relative) dégradation de la relation modes de basse fréquence - changement de conformation, lorsqu’on passe d’un modèle de réseau élastique à une représentation beaucoup plus détaillée. On peut donc penser que, dans le cadre d’une approche standard, il faudrait pouvoir se passer de la minimisation d’énergie. Si l’on calcule directement les valeurs et vecteurs propres de \mathbf{H} à partir d’une structure cristallographique, on constate, d’une part, que \mathbf{H} a un grand nombre de valeurs propres négatives, d’autre part, que n_{eff} a une valeur plus élevée que lorsqu’on étudie une structure dont l’énergie a été minimisée. Même si les valeurs propres négatives sont associées à des mouvements le plus souvent localisés (ceux de quelques atomes dont l’énergie d’interaction est particulièrement élevée), il semble que cela ”perturbe” significativement les autres vecteurs propres de \mathbf{H} (sans doute parce qu’ils doivent, par définition, être orthogonaux les uns aux autres). Partir de structures dont l’énergie a été partiellement minimisée améliore nettement la qualité des résultats obtenus, sans permettre toutefois de faire mieux (en ce qui concerne les valeurs de n_{eff}) que ce qu’on obtient avec un modèle de réseau élastique, en partant de la structure cristallographique (résultats non montrés).

³⁹Pour plus de détails, voir [51].

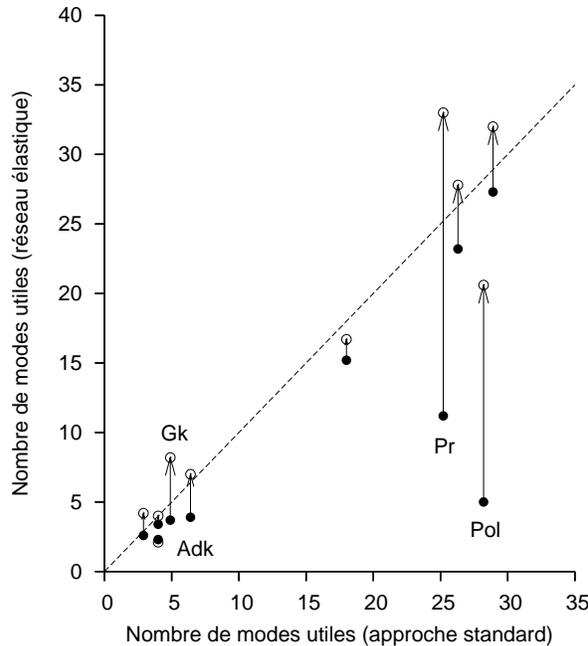


FIG. 17 – Evolution du nombre de modes nécessaires à la description d’un changement de conformation, lorsque ceux-ci ont été obtenus avec un modèle de réseau élastique et lorsqu’on passe de la structure cristallographique (cercles noirs) à une structure dont l’énergie a été minimisée (cercles vides), en fonction du nombre de modes nécessaires à cette description, lorsque ceux-ci ont été obtenus avec une approche standard. Les cas de la guanylate kinase (Gk), de l’adénylate kinase (Adk), de la protéase du VIH-1 sous forme monomérique (Pr) et de l’ADN polymérase β (Pol) sont étiquetés. La ligne pointillée permet de mieux visualiser la diagonale.

Pour pouvoir étudier, dans le cadre d’une approche standard, des structures voisines de la structure cristallographique, mais d’énergie parfaitement minimisée, nous avons choisi d’y inclure un réseau élastique, c’est-à-dire de décrire nos systèmes avec une fonction énergie de la forme suivante :

$$E = E_{Charmm} + \frac{1}{2}k_{enm} \sum_{d_{ij}^0 < R_c} (d_{ij} - d_{ij}^0)^2$$

où E_{Charmm} est la fonction énergie de CHARMM[21], telle que nous l’avons utilisée jusqu’ici, et où k_{enm} est la constante de force des ressorts établis entre les paires de C_α situés à moins de R_c Ångströms les uns des autres dans la structure cristallographique ($R_c=10$ Å ci-après, comme lors des calculs précédents). Des modèles de ce type ont déjà été proposés, avec l’idée de ne décrire à l’échelle atomique que les portions les plus intéressantes d’une structure, ses sites de fixation par exemple[56].

Dans notre cas, lorsque k_{enm} a une valeur élevée, la minimisation d’énergie permet aux chaînes latérales et aux liaisons peptidiques de se réarranger, sans que la position relative des C_α ne change de manière significative. En pratique, pour chacune des quatre protéines ”étiquetées” dans la Figure 17, une série de minimisations d’énergie a été effectuée, avec des valeurs de k_{enm} décroissantes ($k_{enm}=100, 90, \dots, 10, 9, \dots, 1, 0.9, \dots, 0.1, 0.09, \dots, 0.01, 0$ kcal/mole.Å²), en partant à chaque fois de la structure obtenue à l’issue de la minimisation précédente. Pour chaque structure obtenue, les modes normaux ”standards” ont été calculés, puis comparés au changement de conformation correspondant.⁴⁰

⁴⁰Notez qu’avec de tels modèles mixtes, il s’agit de ”vrais” calculs standards : la protéine est décrite

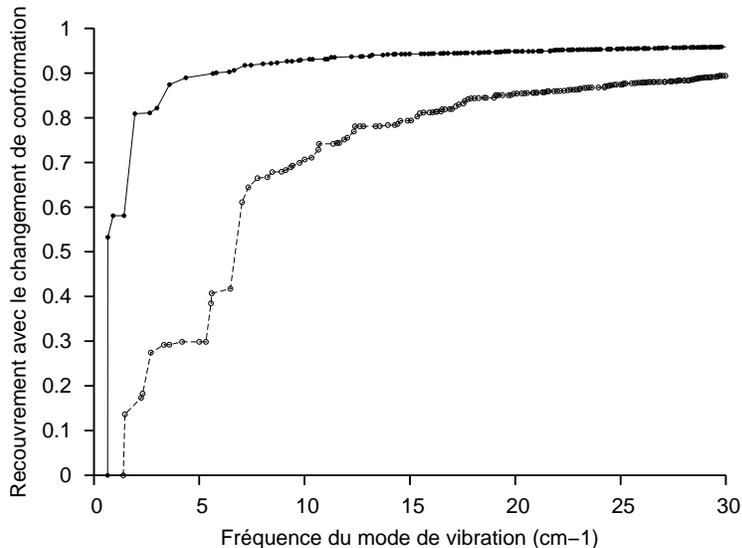


FIG. 18 – Description du changement de conformation de l’ADN polymérase β , avec deux ensembles de modes calculés avec un modèle de réseau élastique. Cercles noirs et ligne continue : en partant de la structure cristallographique. Cercles vides et ligne brisée : en partant de la structure relaxée.

La Figure 19 montre la variation du nombre de modes utiles (voir eqn. 12) par rapport à ce que l’on a obtenu, au mieux, avec un modèle de réseau élastique, c’est-à-dire en partant de la structure cristallographique, en fonction de l’amplitude de la déformation due à la minimisation. Sauf dans le cas de la protéase du VIH-1 sous forme monomérique (non montré), les modes de vibration de la plupart des structures à plus de ~ 0.1 et à moins de $\sim 0.5\text{\AA}$ de la structure cristallographique permettent de mieux décrire le changement de conformation correspondant. Par exemple, dans le cas de la guanylate kinase, pour les structures à environ 0.5\AA , on passe de $n_{eff}=3.7$ (voir Fig. 13) à $n_{eff}=2.6$ (variation de -1.1). A ce stade du processus k_{enn} est de l’ordre de $0.5 \text{ kcal/mole.\AA}^2$.

Pour confirmer le fait qu’une faible déformation permet de réduire le nombre de modes nécessaire à une description correcte du changement de conformation fonctionnel d’une protéine, on peut suivre le recouvrement cumulé. Dans la Fig. 20, le cumul est effectué pour les dix modes qui contribuent le plus à la description. On constate que notre protocole permet d’améliorer progressivement cette quantité, l’amplitude de déformation optimale étant plus ou moins marquée. Dans le cas où elle l’est le plus nettement, c’est-à-dire dans le cas de l’ADN polymérase β , on peut penser que cela est du à la nature de la structure étudiée. En effet, il s’agit de celle d’un complexe binaire, protéine-ADN (code PDB 1BPX), le changement de conformation considéré se produisant suite à la fixation d’un troisième partenaire (un nucléotide). Or nos calculs furent effectués en partant de la protéine ”isolée”. Il est donc fort probable que cette dernière se déforme en s’éloignant plus nettement de la structure cristallographique que ne le ferait le complexe protéine-ADN.

Dans le cas de la protéase du VIH-1 sous forme monomérique (cas non montré), les modes de vibration du modèle mixte permettent de mieux décrire son changement de conformation que ceux obtenus avec une approche standard (lorsque $k_{enn}=2-5 \text{ kcal/mole.\AA}^2$), mais

à l’échelle atomique, son énergie, qui inclut un terme de solvatation implicite, est minimisée, on obtient six valeurs propres numériquement nulles, etc.

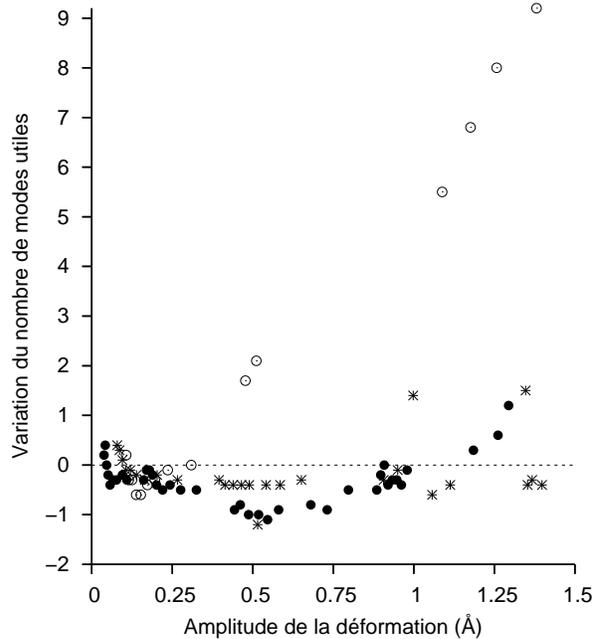


FIG. 19 – Variation du nombre de modes utiles, lorsqu’on passe d’un modèle de réseau élastique à un modèle mixte, en fonction de l’amplitude de la déformation (C_α -r.m.s.) due aux minimisations d’énergie. Etoiles : adénylate kinase. Cercles pleins : guanylate kinase. Cercles vides : ADN polymérase β .

la qualité de la description ($n_{eff}=14-16$) reste en deçà de ce qu’on peut obtenir avec un modèle de réseau élastique ($n_{eff}=11.2$; notez toutefois que même dans ce dernier cas, il s’agit d’un nombre relativement élevé).

6 Applications

6.1 Intégration des équations du mouvement

Puisqu’on peut décrire très correctement les changements de conformation de nombreuses protéines avec quelques modes seulement, on pourrait croire que prédire ces derniers est une tâche relativement aisée, puisqu’il devrait suffire d’explorer le voisinage d’une structure connue en se déplaçant le long des quelques coordonnées correspondantes. Plus précisément, il devrait suffire d’inverser la relation (5), ce qui donne :

$$r_i = r_i^0 + \frac{1}{\sqrt{m_i}} \sum_k^{N_m} a_{ik} q_k \quad (13)$$

où N_m est le nombre de modes utilisés ($N_m = 3N$, au maximum) et où q_k est l’amplitude choisie pour le mode k . Le principe serait d’explorer l’espace des conformations correspondant à une valeur de N_m faible, moins de dix par exemple.

Il y a là plusieurs difficultés. La première : le paysage énergétique au voisinage d’une structure est complexe et il n’est pas du tout évident d’y trouver les barrières d’énergie les plus basses. La seconde : dans ce type de recherche, les imperfections de la fonction énergie utilisée vont introduire un biais favorisant une famille de minima et de points de transition qui peut ne pas inclure ceux qui sont sur le(s) chemin(s) suivi(s) par la protéine pour passer d’une conformation à une autre. De fait, toute une série de méthodes ont été proposées pour essayer de résoudre une version plus simple du problème, celle de la

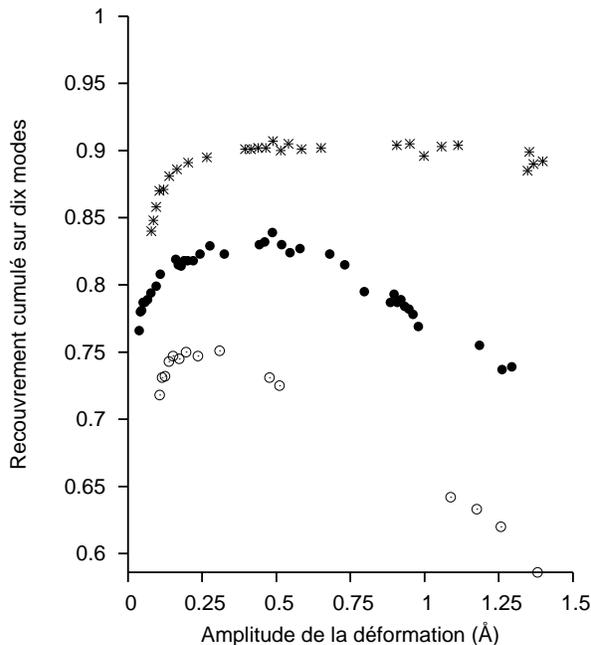


FIG. 20 – Qualité de la description d’un changement de conformation avec dix modes de vibration d’un modèle mixte, en fonction de l’amplitude de la déformation (C_{α} -r.m.s.) due aux minimisations d’énergie. Etoiles : adénylate kinase. Cercles pleins : guanylate kinase. Cercles vides : ADN polymérase β .

recherche du (ou des) chemin(s) suivi(s), lorsqu’on connaît déjà les structures des points de départ et d’arrivée, notamment par Ron Elber et Martin Karplus[57], Anass Ech-Cherif El-Kettani et Jean Durup[58], Stefan Fischer et Martin Karplus[59], Christophe Guilbert, David Perahia et Lilianne Mouawad[60], *etc.*

Une autre difficulté, sans doute un peu moins évidente : dans l’approximation harmonique, à une interaction locale forte, une liaison covalente par exemple, ne correspond qu’un seul mode de vibration, c’est-à-dire une seule coordonnée. Ne pas faire varier cette coordonnée, cela ne signifie donc pas maintenir la longueur de cette liaison constante, mais maintenir constante la *projection* de cette liaison le long de la direction qu’elle avait lorsque les modes de vibration ont été calculés. De ce fait, tout déplacement le long d’un mode qui entraîne un changement de l’orientation d’une liaison covalente sera pénalisé.

Avec David Perahia, lors de mon travail de thèse[20], nous avons rencontré cette difficulté en essayant d’utiliser l’eqn. 13 pour résoudre les équations du mouvement. En effet, comme l’indique l’eqn. 4, à énergie donnée, plus la fréquence d’un mode est basse plus son amplitude est grande. Être en mesure de négliger les modes de haute fréquence aurait plusieurs avantages : d’une part, permettre de résoudre les équations du mouvement avec un pas d’intégration plus grand, puisque ce dernier est relié à la valeur de la plus haute fréquence qui doit être prise en compte, d’autre part, rapprocher les trajectoires de ce qu’elles sont en réalité, puisqu’on sait que les modes de haute fréquence ne sont pas ”activés” (excités) à température ordinaire ($h\nu > k_bT$). De fait, des méthodes permettant de résoudre les équations du mouvement en présence de contraintes appliquées aux longueurs des liaisons covalentes ont été proposées très précocément[61]. Elles permettent d’utiliser un pas d’intégration deux à trois fois plus grand. Malheureusement, du fait de l’artefact mentionné ci-dessus, lorsqu’on résout les équations du mouvement dans le sous-espace associé aux seuls modes de basse fréquence, si on peut en effet le faire avec un pas d’intégration bien plus grand que d’ordinaire (jusqu’à vingt-trente fois), l’am-

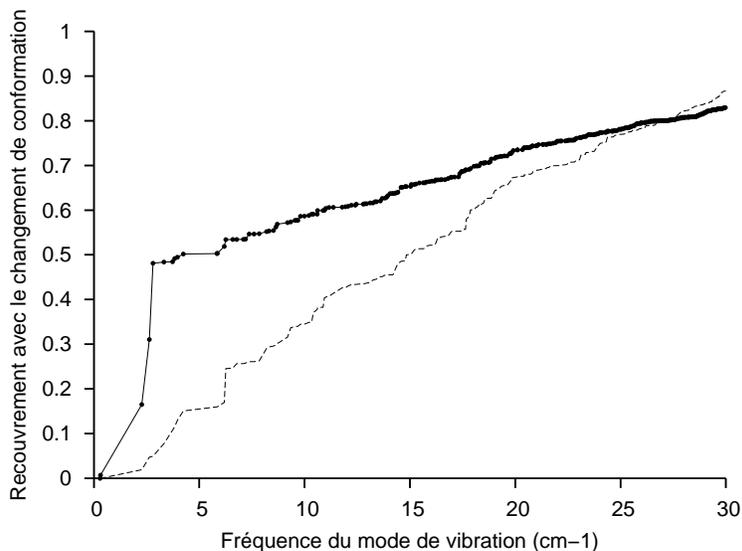


FIG. 21 – Comparaison du changement de conformation proposé pour la rhodopsine lors de son activation par la lumière et de chacun des modes de vibration de sa forme "inactive", calculés avec un modèle de réseau élastique. Cercles noirs et ligne continue : recouvrement cumulé, dans le cas du mouvement proposé par G. Nikiforovich et G. Marshall. Ligne brisée : recouvrement cumulé, dans le cas du mouvement observé par K. Palczewski *et al.*

plitude des mouvements atomiques obtenus est réduite (d'un facteur deux environ[20]). Bien sûr, si le système étudié est moins hétérogène qu'une protéine, cet artefact peut ne pas se manifester, comme cela a été montré par la suite dans le cas de cristaux ou de verres[62, 63].

Dans le cas des protéines, avec Juan Elezgaray, nous avons proposé deux façons d'utiliser le sous-espace associé aux modes de basse fréquence pour résoudre les équations du mouvement. Dans la première, un poids faible est donné aux interactions topologiques (liaisons covalentes, angles de valence, angles dièdres) ce qui permet de se débarrasser des modes de vibration localisés et, ainsi, de l'artefact décrit ci-dessus. Cela nous a permis de produire des trajectoires d'une petite protéine, le BPTI, en utilisant un pas d'intégration de 10-20 fsec (au lieu de 1-2 fsec, avec les approches traditionnelles) tout en obtenant des amplitudes correctes pour les mouvements atomiques[64]. Dans la seconde, on considère qu'une protéine est un ensemble de corps semi-rigides, les coordonnées associées à chacun de ces corps correspondant, d'une part, à leurs mouvements de rotation et de translation, d'autre part, à quelques-uns de leurs modes de basse fréquence. Cette méthode, très proche de celle proposée précédemment sous le nom de MBO(N)D[65], permet d'effectuer des trajectoires à haute température (800K), sans que la protéine ne se dénature[66]. Notez qu'une température élevée permet d'accélérer l'exploration de la surface d'énergie.

6.2 Validation de changements de conformation

Puisqu'on observe un lien peu évident entre les mouvements fonctionnels des protéines et leurs modes de vibration, on peut faire l'hypothèse que lorsqu'on n'observe pas ce lien, c'est parce qu'il ne s'agit pas d'un mouvement (vraiment) fonctionnel, au moins lorsque le mouvement concerné est (plus ou moins) hypothétique.

Par exemple, la structure de la forme inactive (en absence de lumière) de la rhodopsine, une protéine membranaire intégrale, a été obtenue en 2000[67]. Comme la rhodopsine est le prototype d'une famille de protéines très importantes du point de vue pharmacologique, celle des récepteurs couplés aux protéines G, un effort important a été fait pour essayer de "deviner" la nature de son changement de conformation, dans les millisecondes qui suivent l'absorption d'un photon.⁴¹ Notamment, en introduisant une cystéine à l'extrémité cytoplasmique de l'hélice C et une autre à diverses positions dans l'hélice F, puis en modifiant ces cystéines de façon à ce que leur interaction soit détectable, d'une façon qui dépende de leur distance,⁴² Gobind Khorana et ses collaborateurs ont montré que des mouvements de grande amplitude se produisent lors de l'activation[68]. Par la suite, Gregory Nikiforovich et Garland Marshall ont proposé un modèle compatible avec les variations de distances mesurées par G. Khorana *et al* (code PDB 1OV0 [69]). On peut donc comparer les modes de la meilleure structure connue de la rhodopsine (code PDB 1U19) avec le mouvement proposé par G. Nikiforovich et G. Marshall.⁴³ Comme on le voit sur la Figure 21, ce dernier peut être assez bien décrit en ne prenant en compte que les trois premiers modes de la rhodopsine. En comparant avec les Figures 9 et 13, on constate cependant que dans le cas de ce mouvement de la rhodopsine beaucoup trop de modes sont nécessaires pour arriver à en décrire plus de 80% (recouvrement cumulé de plus de 0.8). On peut donc penser que le mouvement proposé par G. Nikiforovich et G. Marshall est une bonne première approximation, mais qu'il demande à être affiné, afin de mieux ressembler à un changement de conformation "naturel".

À l'aune de notre méthode de validation, ce mouvement semble en tout cas plus convaincant que celui proposé récemment par Krzysztof Palczewski et ses collaborateurs, sur une base pourtant expérimentale. Ceux-ci ont en effet réussi à faire croître des critaux de rhodopsine qui résistent à une illumination et à déterminer la structure de la conformation ainsi obtenue (code PDB 2I37[70]). Pourtant, quand on compare le mouvement ainsi défini aux modes de la rhodopsine, on constate qu'aucun de ces derniers ne permet de le décrire correctement, fut-ce de manière partielle. On doit donc en conclure qu'il ne s'agit probablement pas du changement de conformation fonctionnel de la rhodopsine. De fait, le mouvement observé par K. Palczewski *et al* est de relativement faible amplitude (C_{α} -r.m.s. de 1-2Å), par rapport à ce que les données obtenues par G. Khorana *et al* suggèrent (C_{α} -r.m.s. de 4-5Å[68, 69]).

Dans le cas du canal mécanosensible d'*E. Coli*, comme nous l'avions constaté avec Hélène Valadié, Jean-Jacques Lacapère et Catherine Etchebest, si l'on compare le mouvement d'ouverture proposé par Sergei Sukharev et ses collaborateurs[71] aux modes de vibration de la forme fermée (code PDB 1MSL), on constate qu'une dizaine de ces derniers suffisent pour le décrire très correctement (recouvrement cumulé d'environ 0.8[72]). On peut donc en conclure que dans ce cas le calcul des modes de vibration "valide" le mouvement proposé (plus précisément : qu'il s'agit d'un argument en sa faveur).

⁴¹Passage de l'état "Meta I" à l'état "Meta II".

⁴²Via la technique du "Spin labelling".

⁴³Travail en cours.

6.3 Ajustement dans des enveloppes à basse résolution

Les progrès réalisés ces dernières années dans le domaine de la cryomicroscopie électronique ont permis d'obtenir des structures à "basse" résolution⁴⁴ pour de nombreuses protéines. L'un des avantages de cette technique est qu'elle ne requiert aucune préparation préalable de l'échantillon étudié : il suffit de le congeler. Cela a permis, notamment, d'obtenir de nombreuses conformations différentes de systèmes complexes, tels le ribosome.

Dans le cas où la structure d'une conformation est déjà connue à haute résolution, des déformations le long de ses modes normaux de basse fréquence (eqn. 13) peuvent être faites de façon à ajuster la structure dans l'enveloppe obtenue par cryomicroscopie. Ce type d'applications a été testé par Florence Tama lors de son séjour postdoctoral chez Charles L. Brooks, en prenant pour exemple les données de Mikel Valle et Joachim Frank[17]. Cette approche, reprise depuis par M. Delarue *et al*[73], K. Hinsen *et al*[74], a notamment permis de proposer une structure à haute résolution pour un complexe entre un canal membranaire et un ribosome en train de fonctionner[75].

Récemment, avec Karsten Suhre et Jorge Navazza, nous avons mis au point NORMA, un ensemble de scripts désormais disponible sur la Toile (<http://elnemo.ens-lyon.fr/NORMA>) qui permet d'effectuer ce type d'ajustement flexible de manière "presse-bouton". Des structures de protéines du rotavirus ont depuis été déterminées ainsi[76].

6.4 Remplacement moléculaire

De nombreuses structures cristallographiques sont obtenues par la technique dite du "remplacement moléculaire". Celle-ci s'applique lorsqu'on connaît déjà une structure "proche" de celle que l'on souhaite déterminer. On se sert alors de cette dernière pour résoudre le problème des phases, en lieu et place de méthodes plus coûteuses en temps humain, et aussi plus aléatoires, comme la technique du remplacement isomorphe.

Cependant, si la structure que l'on souhaite déterminer a une conformation significativement différente de celle qui est déjà connue, le remplacement moléculaire ne permet pas d'obtenir une solution. C'est pourquoi, avec K. Suhre, nous avons proposé d'utiliser cette technique en partant d'ensembles de conformations générées en déformant une structure le long de ses modes de basse fréquence (eqn. 13). En pratique, les déformations le long d'un ou deux modes sont systématiquement échantillonnées, l'amplitude maximum de la déformation étant fixée.⁴⁵ Puis chaque conformation obtenue est testée, celles qui permettent de résoudre le problème des phases devant correspondre à des déformations d'amplitude voisine[77]. Au même moment que nous, Marc Delarue et Philippe Dumas proposaient une méthode voisine dans laquelle l'échantillonnage est remplacé par une optimisation de l'amplitude des déformations[73]. D'autres variantes ont été proposées depuis[78, 79], et quelques structures ont été résolues en utilisant l'une ou l'autre de ces approches (voir, par exemple, les refs. [80, 81]). La nôtre vient d'être implémentée dans PHASER[82], un logiciel largement utilisé dans la communauté des cristallographes, ce qui devrait permettre de vulgariser son utilisation.

⁴⁴Facteur R de 8-20Å.

⁴⁵Au-delà de 2Å de C_α-r.m.s., il faut s'affranchir d'effets artefactuels, par exemple en minimisant l'énergie des conformations ainsi obtenues.

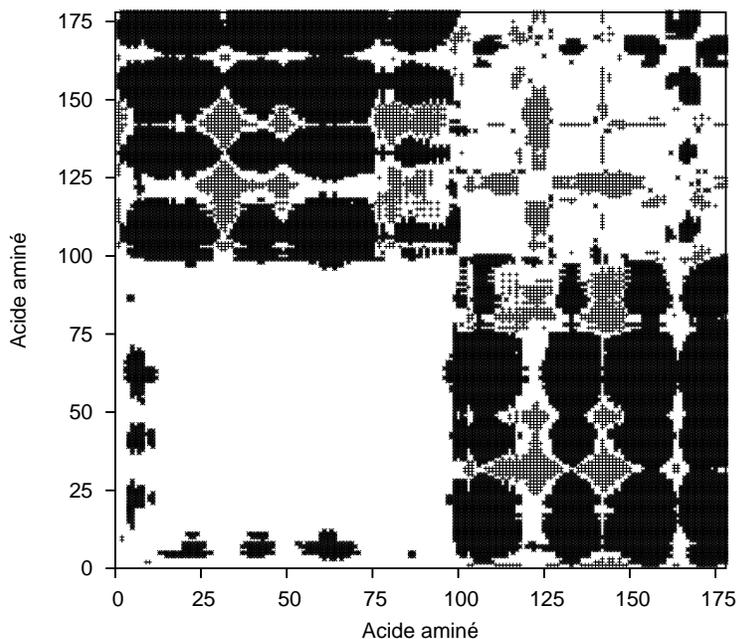


FIG. 22 – Variation des distances entre paires de résidus, lors du mouvement de plus basse fréquence des deux premiers domaines du CD4. Ne sont indiquées que les paires de résidus dont la distance varie d’au moins 50% de plus que ne le fait la moyenne des paires. Un même signe (Plus ou Etoile) indique que les distances considérées varient dans le même sens.

6.5 Prédiction du changement de conformation du CD4

6.5.1 A partir des modes de vibration

Les résultats que nous avons obtenus avec Osni Marques, dans le cas de la citrate synthase[24] (voir p13), m’avaient convaincu qu’il devrait être possible de ”deviner” la nature du changement de conformation fonctionnel d’une protéine, puisque cela revient à être capable de choisir celui qui, parmi ses modes de basse fréquence, doit ressembler le plus à ce mouvement.

Songeant, de manière manifestement trop directe,⁴⁶ à des applications thérapeutiques, j’avais tenté de trouver des protéines impliquées dans des maladies bien étudiées (pour avoir suffisamment de données expérimentales), mais elles-mêmes quelque peu délaissées.⁴⁷ Mon choix s’était porté sur le CD4, le récepteur du VIH-1,⁴⁸ en partie parce que l’idée de chercher des molécules capables d’empêcher l’entrée du virus dans une cellule me paraissait plus raisonnable que la voie privilégiée alors, à savoir l’empêcher de s’y multiplier.⁴⁹

Le CD4 est une protéine transmembranaire dont la partie extracellulaire comporte quatre domaines de type immunoglobuline. Le VIH-1 se fixe sur le premier, le plus loin des quatre par rapport à la membrane. A l’époque,⁵⁰ la structure de cette partie extracellulaire n’était connue que par fragments et c’est pourquoi les calculs de modes de vibration furent effectués en partant de structures ne comprenant que les deux premiers domaines. Dans les trois cas considérés,⁵¹ les résultats furent très similaires : les trois modes de plus basse

⁴⁶L’impatience de la jeunesse, sans doute.

⁴⁷Pour n’avoir qu’une pile de taille raisonnable d’articles à lire.

⁴⁸Au moment où j’ai commencé à travailler sur ce sujet, les co-récepteurs n’étaient pas connus.

⁴⁹Des ”inhibiteurs d’entrée” ont pris place ces dernières années dans l’arsenal thérapeutique[83].

⁵⁰Il y a un peu plus de dix ans, donc.

⁵¹Trois structures avaient été déterminées par les groupes de R.W. Hendrickson[84] et S.C. Harisson[85] ; codes PDB : 1CDH, 1CDI et 3CD4. Ces structures sont à moins de un Ångströms de C_{α} -r.m.s. les unes

fréquence correspondent à des mouvements quasi-rigides d'un domaine immunoglobuline par rapport à l'autre[86, 87]. Ceci n'était pas vraiment inattendu, même si l'interface entre les deux premiers domaines du CD4 est particulièrement grande[84]. Ce qui fut par contre une surprise, c'est la façon dont les boucles BC et FG du second domaine (résidus 120-127 et 162-167, respectivement) "accompagnent" le mouvement du premier domaine, comme on peut le deviner sur la Figure 22, qui montre les variations de distances entre paires de résidus lors du mouvement de plus basse fréquence.⁵²

Or, c'est dans ces boucles-là qu'ont été trouvées les deux seules mutations du second domaine du CD4 qui abaissent de manière significative son affinité pour Gp120, la protéine de surface du VIH-1.⁵³ Dans le premier cas, il s'agit d'une triple mutation : P121S + P122K + G123V[88] ; dans le second, de l'insertion d'un dipeptide, Ser-Arg, après Asn 164[89]. Une autre résultat expérimental semble aussi indiquer que l'affinité du CD4 pour Gp120 dépend (en partie) de portions de la structure qui ne sont pas en interaction directe avec Gp120. En effet, si l'on insère les résidus 36-62 du CD4 humain en lieu et place de leurs homologues du CD4 du rat,⁵⁴ on trouve que la chimère ainsi obtenue a une affinité deux fois moindre pour Gp120[90].

De plus, les boucles BC et FG sont proches ou font partie d'épitopes reconnus par des anticorps monoclonaux capables d'altérer l'infectivité du VIH-1, sans toutefois perturber de manière significative l'interaction du CD4 et de Gp120.⁵⁵ On verra ci-après que ce fait, qui m'avait paru être un argument fort en faveur de la nécessité d'un mouvement relatif des deux premiers domaines du CD4, préalablement à la fusion cellule-VIH-1[87], peut avoir (au moins) un autre type d'explication.

6.5.2 Rebondissement

En 2002, Philip Hogg et son équipe ont montré qu'il se passe effectivement "quelque chose" au niveau du second domaine du CD4, et que ce quelque chose est nécessaire au processus qui permet au VIH-1 d'entrer dans une cellule[93]. Il s'agit cependant d'une série d'évènements manifestement plus complexe qu'un simple mouvement relatif de deux domaines structuraux. Tout d'abord, à la surface de certains types cellulaires, le CD4 peut dimériser. De plus, ce dimère est covalent[94] et ce sont les cystéines 130 et 159⁵⁶ qui lient les deux monomères, via une paire de ponts disulfures[93]. De plus, si l'on fait réagir le CD4 avec un composé capable de se greffer sur une paire de fonctions thiols libres, on constate que l'entrée du VIH-1 est fortement inhibée, un effet que l'on observe pas dans le cas du mutant C130A + C159A[93]. Or, dans le CD4 sous forme monomérique, toutes les cystéines sont impliquées dans des ponts disulfures[84, 85] (y compris les cystéines 130 et 159).

Tout se passe donc comme si le pont disulfure du second domaine devait se rompre, pour que l'entrée du VIH-1 dans la cellule devienne possible. De fait, ce pont disulfure

des autres.

⁵²Obtenu en partant comme précédemment[86] de la structure 3CD4, avec un calcul standard, mais cette fois en incluant le terme de solvant implicite EEF1.1, développé entretemps[14, 29].

⁵³À l'époque où l'on cherchait à déterminer quel est l'épitope reconnu par Gp120.

⁵⁴Les deux séquences ont un taux d'identité de l'ordre de 50%.

⁵⁵L'anticorps, tel 5A8[91, 92], le CD4 et Gp120 pouvant former des complexes ternaires.

⁵⁶Dans le second domaine, donc.

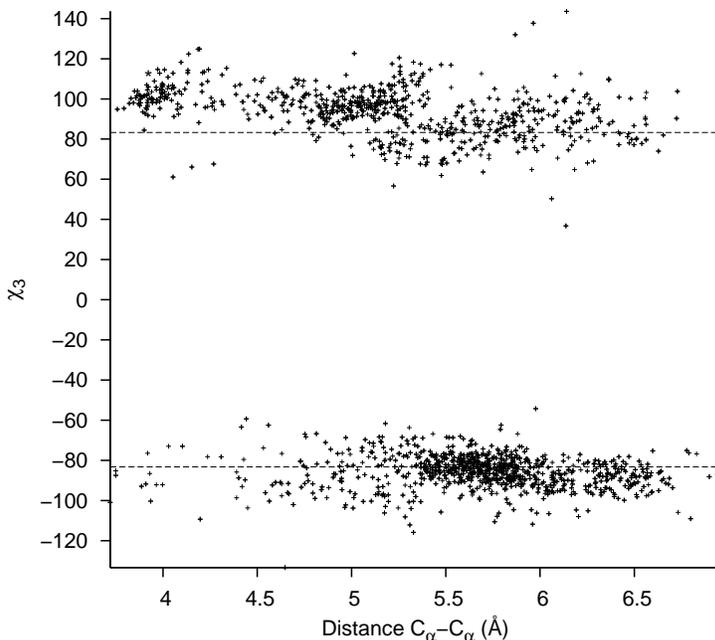


FIG. 23 – Angle dièdre d’un pont disulfure, en fonction de la distance entre les C_α des deux cystéines, pour un ensemble de 1908 structures non redondantes et de bonne qualité ($R < 1.8\text{\AA}$). Les lignes pointillées indiquent les valeurs d’équilibre, telles qu’observées dans des systèmes modèles[95, 96].

est atypique : alors que dans un domaine immunoglobuline standard le pont disulfure lie ensemble les deux feuillets β qui se font face, dans le second domaine du CD4 il lie deux brins d’un même feuillet. De plus, ces deux brins étant côte à côte, ce pont ne peut se former qu’avec une conformation inhabituelle, moins favorable énergétiquement parlant. De fait, la distance entre les C_α des cystéines 130 et 159 est de 3.9\AA , parmi les plus courtes observées (voir Fig. 23), et l’angle dièdre du pont disulfure qu’elles forment (χ_3) a une valeur (positive) nettement plus grande que ce l’on observe d’ordinaire : 109° (au lieu de ± 83 ; voir Fig. 23). Ce pont disulfure fait d’ailleurs partie d’une catégorie de ponts qui s’avèrent souvent réactifs, que Philip Hogg et ses collaborateurs ont proposé d’appeler ”cross-strand disulfides” [97].

6.5.3 Entrelacement de domaines

Quoiqu’il en soit, ce pont est enfoui au coeur du second domaine et il n’est pas évident d’imaginer comment, une fois rompu, les cystéines 130 et 159 peuvent ensuite participer à des ponts disulfures inter-monomères. La solution (probable) est donnée par un cas aux caractéristiques similaires, celui du prion. Lorsque la structure de cette protéine fut déterminée par RMN[98], ce fut sous forme monomérique. Mais lorsqu’elle le fut ensuite par cristallographie des rayons X, ce fut cette fois sous une forme dimérique[99]. De plus si, dans les deux cas, un pont disulfure est établi entre les cystéines 179 et 214, dans le second, il lie covalamment les deux monomères. Cela s’avère possible grâce à un phénomène qu’on peut appeler ”entrelacement de domaines”.⁵⁷ Le principe est le suivant : la protéine est constituée de deux domaines A et B, reliés par une boucle, L (comme ”linker”). Quand elle dimérise, on peut observer soit un dimère normal, dans lequel les monomères sont accolés, $AB : B'A'$, soit une forme ”entrelacée”, $AB' : BA'$. Si l’on compare

⁵⁷Une traduction de ”3D domain-swapping” due à Joseph Parello.

ces deux formes (comparaison possible par exemple dans le cas du CD2[100, 101]; voir ci-dessous), on constate que seule la conformation de la boucle L change, passant d'une forme "coudée", dans le dimère accolé, à une forme "étendue", dans le dimère entrelacé, ce qui lui permet de "sauter" d'un monomère à l'autre. De plus, si, comme dans le cas du prion, un pont disulfure est établi entre une cystéine du domaine A et une autre du domaine B alors, dans la forme entrelacée, ces deux ponts disulfures lient les monomères entre eux, contribuant notamment à la stabilité de cette forme dimérique.⁵⁸ Il est d'ailleurs possible que la capacité qu'a le prion à adopter une forme entrelacée ait un rapport avec sa pathogénicité, puisque de nombreuses mutations impliquées dans des maladies génétiques sont localisées au voisinage ou entre les cystéines 179 et 214 (D178N, V180I, T183A..., E211Q, Q212P⁵⁹).

Quoiqu'il en soit, j'ai voulu vérifier que c'est bien ce qui se passe dans le cas du CD4. D'un point de vue pratique, si l'on connaît la forme dimérique normale (accolée), il doit (ou devrait) être relativement facile de dire si ce type de changement de conformation peut ou non se produire. En effet, la plupart des interactions qui stabilisent les deux types de dimères sont identiques, puisque seules changent les interactions que L fait avec A et B, ainsi que la conformation de L. Dans le cas du CD2, la forme dimérique accolée est connue (code PDB 1HNG), en tant que paire de "voisins" dans la maille cristalline[100].⁶⁰ Partant de cette forme, on peut essayer de retrouver la forme entrelacée et voir si des protocoles de modélisation standard permettent de dire laquelle des boucles du CD2 est le linker. J'avais choisi le protocole suivant, au cours duquel chaque boucle est testée. Pour ce faire, elle est coupée en position i et reliée à l'acide aminé $i+1$ de l'autre monomère. A ce stade, une énergie topologique importante est introduite (la longueur de la "nouvelle" liaison entre les acides aminés i et $i+1$ peut être de plus de dix Ångströms). La boucle est donc relaxée, en présence de contraintes harmoniques sur le reste de la structure. Lors d'une première étude de "faisabilité", cette étape fut effectuée via une combinaison de minimisations d'énergie et de simulations de dynamique moléculaire à température ordinaire. De plus, pour augmenter les chances d'atteindre les meilleures conformations du linker (celles de plus basse énergie), de nombreux points de départ différents furent testés (notamment trois points de coupure : soit au début, soit au milieu, soit à la fin de la boucle). Au total, pour chaque boucle, vingt relaxations indépendantes furent effectuées, et parmi les vingt conformations de plus basse énergie ainsi obtenues, les dix meilleures furent retenues. Les résultats sont les suivants[104] : le linker observé expérimentalement est bien la boucle qui se trouve avoir, à la fin du protocole d'optimisation, à la fois la tension la plus basse et l'énergie d'interaction avec le reste de la structure la plus favorable. De plus, les quatre conformations de plus basse énergie du linker ainsi obtenues sont à 1.8-2Å de la conformation observée dans la forme cristalline entrelacée (alors que les autres sont à plus de 4Å).

Dans le cas du CD4, en partant d'une forme accolée trouvée dans la maille du cristal de la structure 1G9M[105],⁶¹ on constate qu'une forme entrelacée est tout aussi possible (voir

⁵⁸ Dans le cas du prion, la forme monomérique est la plus stable et c'est le processus de cristallisation qui "sélectionne" la forme entrelacée[99].

⁵⁹Voir <http://www.expasy.ch>, séquence PRIO_HUMAN.

⁶⁰Les deux monomères d'une forme entrelacée ont d'ordinaire une interface beaucoup moins importante que lorsqu'il s'agit d'un dimère "standard"[102, 103].

⁶¹Deux autres formes furent aussi considérées, sans succès[104]. La structure 1G9M est celle d'un com-

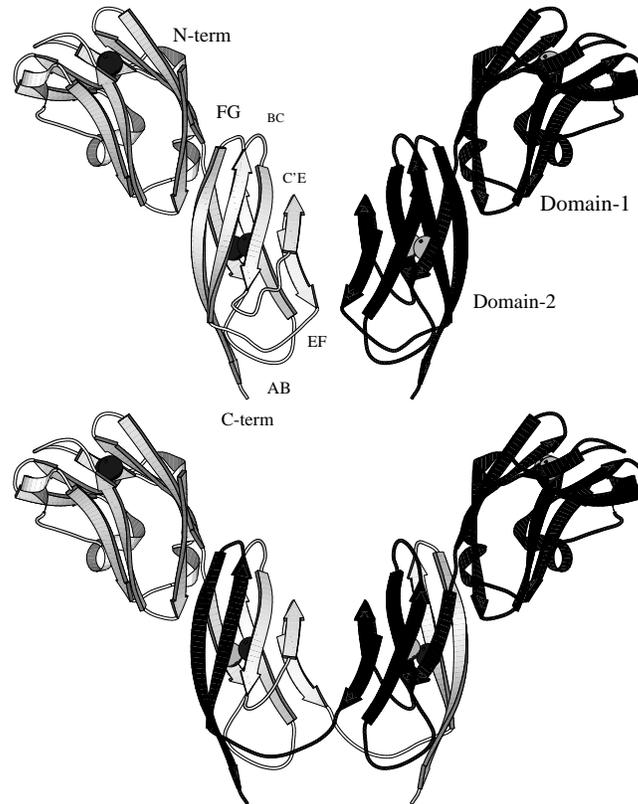


FIG. 24 – Dimérisation du CD4. De la forme accolée, observée dans la maille du cristal de la structure 1G9M (en haut), à la forme entrelacée (en bas). C'est la boucle EF qui joue le rôle de linker. Les sulfures des cystéines sont représentés par des sphères. Dans le second domaine, ils sont liés via un pont disulfure qui est, soit intra-monomérique (forme accolée), soit inter-monomérique (forme entrelacée).

Fig. 24), et que le linker le plus probable se situe, dans la séquence, entre les deux cystéines qui forment le pont disulfure impliqué dans sa dimérisation covalente[104]. Dans un tel cas, l'entrelacement de domaines se traduit donc bien par un échange de ponts disulfures, le pont intra-monomère devenant dans le dimère un pont inter-monomères. Philip Hogg et ses collaborateurs ont depuis obtenu des données expérimentales qui suggèrent fortement que la forme entrelacée du CD4 pourrait être la forme fonctionnelle *in vivo*, lorsque le CD4 joue son rôle de co-activateur du complexe majeur d'histocompatibilité (CMH). Ceci nous a conduit, Bryan Schmidt et moi-même, à proposer un modèle pour le complexe CD4-CMH(II)-TCR[106].

Par ailleurs, la géométrie du CD4 sous forme dimérique permet de proposer une explication en ce qui concerne la façon dont agissent certains anticorps monoclonaux. Sur la Figure 25, on voit par exemple que l'épitope de MT151, qui inhibe fortement la formation de syncytia due à l'infection par le VIH-1 sans toutefois perturber l'interaction du CD4 et de Gp120[107, 108, 89, 109, 110], est situé entre les deux monomères. On peut donc penser que lorsque MT151 interagit avec le CD4, la dimérisation de ce dernier ne doit plus

plexe Gp120-CD4-anticorps monoclonal. Le fait qu'une forme accolée compatible avec un entrelacement de domaines soit trouvée dans un tel cas est un argument en faveur de l'hypothèse que cette forme joue un rôle "fonctionnel".

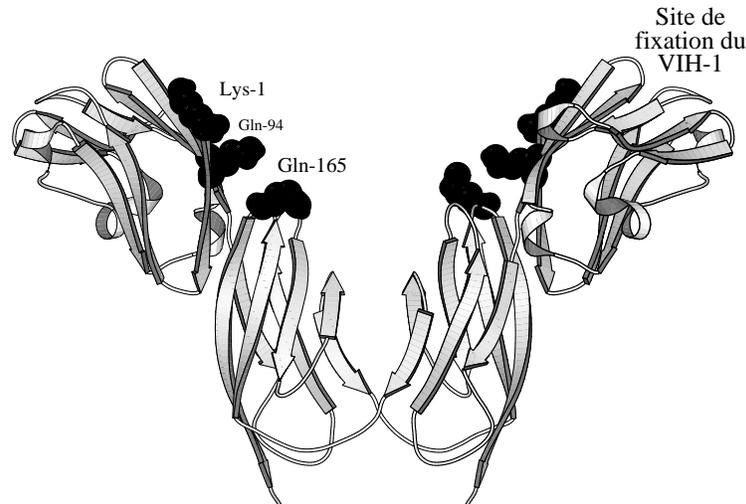


FIG. 25 – Les trois acides-aminés connus pour faire partie de l'épitope de l'anticorps monoclonal MT151.

être possible. Réciproquement, cela suggère que la dimérisation du CD4 est nécessaire, au moins lors de certaines étapes du processus infectieux. On voit alors aussi qu'un mouvement relatif des deux premiers domaines du CD4 devrait permettre de pouvoir positionner plus précisément les deux sites de fixation du CD4, augmentant l'affinité globale de ce dernier pour Gp120, ce qui permet de mieux comprendre comment des mutations à l'interface entre les deux domaines pourraient affecter cette dernière.

6.5.4 Perspectives

Le CD4 n'est pas le seul récepteur viral dont la partie extracellulaire soit constitué d'une série de domaines de type immunoglobuline. Ce n'est pas non plus le seul dont le site de fixation du virus est situé au sommet du premier domaine, loin de la surface cellulaire. D'autres cas sont connus, tels ICAM-1 (cinq domaines), le récepteur du rhinovirus, PVR (trois domaines), le récepteur du poliovirus, etc. De plus, le site de fixation d'un autre virus (l'herpès de type 7) se situe lui aussi sur le premier domaine du CD4. Il est donc tentant de faire l'hypothèse que ces virus utilisent une stratégie similaire pour pénétrer dans une cellule. En d'autres termes, ils exploiteraient le même type de "point faible", qu'il pourrait s'avérer utile de connaître plus en détails...

6.6 Autres prédictions

Par la suite, bien d'autres tentatives ont été faites pour essayer de relier structure et fonction en étudiant les propriétés dynamiques d'une protéine via le calcul de ses modes de vibration, notamment par Ivet Bahar, Robert Jernigan et leurs collaborateurs. Avec le modèle du réseau gaussien (voir p16), ils ont par exemple étudié une t-ARN synthétase[111], la reverse transcriptase du VIH-1[112], la tryptophane synthase[113], etc. Dix ans plus tard, il est devenu (relativement) difficile de trouver une protéine d'intérêt biologique dont la structure est connue et dont les modes de vibration n'ont pas déjà été analysés en détails...

7 Un modèle de réseau élastique anharmonique

7.1 Introduction

L'approximation harmonique (voir partie 2) permet, à partir d'une structure, d'obtenir des informations dynamiques beaucoup plus riches que ce que l'on pourrait en attendre, *a priori*. Il est cependant bien évident que les contributions des termes anharmoniques sont importantes[5, 114], ne serait-ce que lorsque des échanges d'énergie se produisent à l'intérieur d'une structure. De manière générale : chaque fois que le système est hors d'équilibre. En effet, dans l'approximation harmonique, lorsque de l'énergie est donnée à un mode de vibration à $t=0$, celle-ci y reste *ad eternam*.⁶² Cette propriété, à la base de la généralité de cette approximation, puisque c'est elle qui permet de résoudre les équations du mouvement et ce, même lorsque le nombre de degrés de liberté est grand, est tout à fait incongrue : lorsqu'on donne de l'énergie à un mode, celle-ci doit diffuser vers les autres modes, jusqu'à ce qu'une situation d'équilibre soit atteinte, dans laquelle chaque mode a $k_B T$ d'énergie, en moyenne.

7.2 Le modèle choisi

Si l'on souhaite introduire des termes anharmoniques de manière contrôlée, la façon la plus directe de procéder consiste à étudier des systèmes à basse température[115], ou bien à poursuivre le développement en série de la fonction énergie potentielle standard (voir eqn. 1), en incluant les termes cubiques[116, 117], quartiques, etc. Avec Francesco Piazza et Paolo de los Rios, ce, dans le cadre de la thèse de Brice Juanico,⁶³ nous avons choisi d'explorer une autre voie, qui consiste à complexifier peu à peu les modèles de réseau élastique, en essayant d'y mettre en évidence des phénomènes non-linéaires déjà connus par ailleurs. Nous avons commencé notre série d'études en adoptant la forme suivante pour la fonction énergie, directement inspirée du travail pionnier d'Enrico Fermi, John Pasta et Stanislaw Ulam[118] :

$$E_p = k_{enm} \sum_{d_{ij}^0 < R_c} \left[\frac{1}{2} (d_{ij} - d_{ij}^0)^2 + \frac{\lambda_{anh}}{4} (d_{ij} - d_{ij}^0)^4 \right] \quad (14)$$

Lorsque $\lambda_{anh}=0$, on retrouve la forme proposée par M. Tirion (voir eqn. 8). Dans un premier temps, nous avons choisi d'étudier le cas correspondant à $\lambda_{anh}=1$. On sait en effet qu'alors, au moins dans le cas de chaînes, ou de pavages réguliers, d'oscillateurs, on peut mettre en évidence un phénomène singulier : la localisation non-linéaire d'énergie[119]. De plus, ce phénomène peut se manifester dans des conditions qui permettent de ne faire aucune hypothèse particulière en ce qui concerne ses caractéristiques (fréquence, localisation, etc) : lors du refroidissement d'un système par ses extrémités (dans le cas d'une chaîne) ou par ses bords (dans le cas d'un pavage)[120, 121, 122]. D'un point de vue physique, ce phénomène correspond à l'existence de modes de vibration non-linéaires, dont les fréquences sont (le plus souvent) supérieures aux plus hautes fréquences harmoniques du système.

Dans un premier temps, nous avons vérifié que dans le cas de représentations simplifiées de protéines dont le comportement est parfaitement harmonique (linéaire), si on ne met que

⁶²On dit que les modes ne sont pas "couplés".

⁶³Soutenue en décembre 2007.

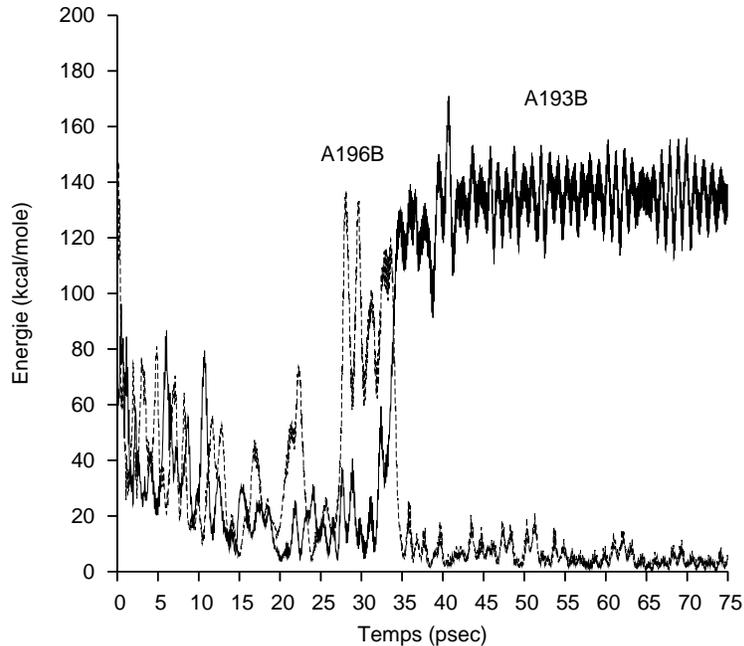


FIG. 26 – Energie de deux acides-aminés du second monomère de la citrate synthase, en fonction du temps. A $t=0$, la température est de 10000K. Vers $t=20$ psec, du fait du refroidissement par la surface, elle est de 300K. A $t=75$ psec, 75% de l'énergie du système est concentrée sur un seul acide-aminé.

les atomes de surface en contact avec le solvant, la dissipation d'énergie hors de la protéine peut devenir extrêmement lente[123], ainsi qu'on l'observe pour des systèmes dans lesquels le phénomène de localisation non-linéaire d'énergie peut se manifester[121, 124].

Dans un second temps, nous avons cherché à mettre en évidence ce phénomène dans des protéines, en décrivant celles-ci avec notre modèle non-linéaire (eqn. 14). La dissipation d'énergie s'effectue, via un terme de friction,⁶⁴ par la surface, définie comme étant composée par les acides aminés qui ont plus de 25\AA^2 en contact avec le solvant. Ainsi, pour les systèmes que nous avons étudiés, environ la moitié des acides aminés font partie de la surface. Par ailleurs, pour avoir des échelles de temps comparables à celles des "vraies" protéines, on pose : $k_{enm} = 5 \text{ kcal/mole.\AA}^2$. Si de plus, comme précédemment, $R_c=10\text{\AA}$, alors les valeurs des plus basses fréquences sont voisines de celles qu'on obtient avec une fonction énergie de type CHARMM[21].⁶⁵

La Figure 26 montre l'évolution de l'énergie de deux acides aminés de la citrate synthase (code PDB 1IXE) en fonction du temps.[125] Alors que, du fait de la friction en surface, l'énergie totale du système décroît rapidement (elle est inférieure à 200 kcal/mole à $t=75$ psec), l'énergie de chacun de ces deux acides aminés fluctue fortement, jusqu'à atteindre des valeurs supérieures aux valeurs initiales (60 kcal/mole par acide aminé, en moyenne, à $t=0$). Notez que vers $t=32$ psec, l'énergie passe de l'alanine 196B à l'alanine 193B, où elle reste jusqu'à la fin de la simulation. Dans d'autres simulations, effectuées avec des conditions initiales différentes (mais toujours en partant d'une température élevée), l'énergie se localise sur d'autres sites, la threonine 192B et la threonine 208A, notamment.

⁶⁴ $\beta=2 \text{ psec}^{-1}$, une valeur standard pour un atome d'une protéine dans un environnement aqueux[116].

⁶⁵On associe par ailleurs à chaque C_α une masse correspondant à la masse moyenne d'un acide aminé : 110 a.m.u.

Il est tentant de se demander si cette capacité qu'ont certains sites à concentrer, puis à retenir, de l'énergie, ne pourrait pas être utilisée par les macromolécules biologiques, notamment lorsque des barrières d'énergie doivent être franchies, par exemple lors des processus catalytiques. Cette suggestion avait été faite dès la fin des années soixante-dix par A.S. Davydov, qui pensait que la localisation d'énergie pourrait avoir lieu via dans des réseaux de liaisons hydrogènes, notamment ceux impliqués dans les structures secondaires des protéines[126].

Dans notre cas, la localisation se produit sur les sites les plus "rigides" de la protéine. Or il se trouve que Richard Lavery et ses collaborateurs ont récemment montré que les acides aminés impliqués dans le processus catalytique ont tendance à être situés dans les portions les plus rigides des enzymes[127], un fait qui n'est pas évident à comprendre, *a priori*. En effet, dans un modèle à gros grains, les portions les plus rigides sont aussi les plus denses, ce qui semble incompatible au premier abord avec la nécessité qu'ont les résidus du site actif d'entrer en contact avec les substrats de l'enzyme. Notre résultat permet en tout cas de suggérer pourquoi ce pourrait être nécessaire : pour y amener un surcroît d'énergie, par rapport à ce qui est normalement "accessible" à température ordinaire.

8 Perspectives

8.1 Localisation d'énergie et activité catalytique

Les résultats brièvement commentés dans la partie précédente ont été obtenus avec une représentation très peu réaliste des protéines. Il s'agira donc (notamment) d'essayer de montrer que ce phénomène de localisation non-linéaire d'énergie peut aussi être mis en évidence quand on utilise une représentation plus détaillée.

8.2 Vers la prédiction de changements de conformation

Le constat suivant est désormais bien établi : il existe un lien entre le(s) changement(s) de conformation fonctionnel(s) d'une protéine et ses modes de vibration de basse fréquence. Il s'agira, d'une part, de mieux comprendre l'existence même de ce lien. En particulier de répondre à la question : comment se fait-il qu'on puisse l'observer lorsqu'on étudie des modèles qui ne prennent en compte aucun des détails de la structure ? (notamment : la séquence en acides aminés de la protéine étudiée). Il s'agira, d'autre part, d'essayer d'exploiter ce lien à des fins prédictives.

Pour ce faire, une première voie consiste à déformer la protéine étudiée le long de ses modes de basse fréquence et à voir si l'on peut générer ainsi des conformations compatibles avec des données expérimentales d'origine diverse (cristallographie, RMN, FRET, etc). Plusieurs applications de ce type ont été récemment proposées (voir la partie 6). Il s'agit maintenant d'essayer de faire mieux connaître cette approche, en continuant à multiplier les exemples, mais aussi en rendant son utilisation (encore) plus simple.

Une seconde voie, complémentaire, consiste à utiliser les méthodes développées récemment afin d'explorer une surface d'énergie libre le long de quelques coordonnées, en utilisant comme coordonnées celles qui sont associées aux modes de basse fréquence d'une protéine.

La plus connue de ces méthodes est sans doute la "méta-dynamique", proposée par Michele Parrinello et son équipe[128].

De manière plus générale, le problème de la prédiction d'un changement de conformation est analogue à celui du repliement des protéines. Comme ce dernier, il a deux facettes indépendantes. Tout d'abord, il faut trouver une façon d'explorer les possibilités conformationnelles d'une protéine qui permette de trouver la conformation recherchée. Avoir la possibilité de limiter le nombre de degrés de liberté (en privilégiant les coordonnées associées aux modes de basse fréquence) devrait permettre de simplifier grandement cet aspect du problème. L'autre point concerne la capacité à "reconnaitre" la conformation recherchée, c'est-à-dire à évaluer correctement son énergie (libre). On peut séparer ces problèmes de deux façons. Tout d'abord, en ajoutant des termes énergétiques qui favorisent la conformation recherchée, dans l'esprit des "modèles de Go" [50, 129] qui se sont avérés si féconds pour l'étude du repliement des protéines. Ceci permet de tester (et/ou de développer) les méthodes de recherche conformationnelle mais aussi d'avoir une idée des termes qui dominent l'état de transition. Enfin, on peut choisir d'étudier des protéines dont la séquence est telle que la conformation recherchée soit particulièrement stable. De telles séquences pourront être obtenues, par exemple, via les techniques de type "conception rationnelle de protéines".⁶⁶

⁶⁶"Protein design". Je participe au projet PROTEUS (voir <http://migale.jouy.inra.fr/proteus>), dont l'objectif est d'appliquer cette technique à grande échelle.

9 Remerciements

Ils vont bien sûr tout d'abord à mes collaborateurs, sans qui la plus grande partie, pour ne pas dire l'essentiel, des résultats que j'ai présentés dans ce mémoire auraient vu le jour ailleurs, autrement, et, pour certains, peut-être même pas encore. Dans l'ordre alphabétique : François Colonna-Césari[20], Marc Delarue[51], Philippe Durand[27], Juan Elezgaray[64, 66, 130, 131], Catherine Etchebest[72], Florent-Xavier Gadéa[28], Philip Hogg[106], Brice Juanico,⁶⁷ Martin Karplus,⁶⁸ Gilles Marcou[66, 131], Osni Marques[24, 28], Samuel Nicolay[31], David Perahia[20], Francesco Piazza[123], Paolo de los Rios[123], Bryan Schmidt[106], Karsten Suhre[41, 77, 132], Florence Tama[28, 38], Orlando Tapia[20, 133], Georges Trinquier[27, 134, 135, 136, 137], Hélène Valadier[72] et Jeannine Yon-Kahn[20].

Une pensée aussi pour tous ceux dont je n'ai pas fait mention ici, nos travaux communs ayant été consacrés à d'autres sujets : Fabienne Alary[138, 139], David Allouche[140, 141, 142, 143], Pierre Coutos-Thévenot[144], Jean Durup[138, 139], Jean-Pierre Daudey[141, 142], Simon Mathieu,⁶⁹ Ivan Ortega-Blake[142], Nicole Paillous[140], Joseph Parello[143], Jacques Perié[145, 146], Xavier Periole[141, 142], Alejandro Ramirez-Solis[142] et Michèle Willson[145, 146].

Mes remerciements vont enfin aux membres du jury, qui m'ont fait l'honneur d'accepter de juger ce mémoire et qui, pour la plupart, ont aussi accepté d'en être d'éventuels rapporteurs.⁷⁰

⁶⁷Travail soumis pour publication.

⁶⁸Travail en cours de rédaction.

⁶⁹Travail en cours.

⁷⁰Pour en retenir finalement trois, l'Université Claude Bernard demande cinq noms...

10 Curriculum vitae

Depuis Janvier 2003 : Affecté au Laboratoire de Physique (UMR 5672) de l'E.N.S.-Lyon. Travail dans le laboratoire transdisciplinaire Joliot-Curie (USR 3010), dans l'équipe "Chromatine et Génôme" dirigée par Françoise Argoul.

1999-2002 : Affecté au Centre de Recherche Paul Pascal (UPR 8641), à Bordeaux. Travail dans le groupe "Structure et Dynamique des Macromolécules Biologiques", dirigé par Alain Arnéodo.

1991-1999 : Affecté au laboratoire de Physique Quantique (UMR 5626), à Toulouse. Travail dans le groupe "Dynamique des protéines", dirigé par Jean Durup, puis dans le groupe "Biodynamique", co-dirigé par Florent-Xavier Gadéa et moi-même.

Printemps 1991 : Recruté par le C.N.R.S.

1990-1991 : Séjour post-doctoral au Scripps Clinic Research Institute, à San Diego (USA). Travail avec Donald Bashford sur la nature du "globule fondu".

Juin 1990 : Thèse intitulée "Étude théorique des mouvements internes de grande amplitude de la décaalanine et du fragment C-terminal de la protéine ribosomale L7/L12", effectuée sous les directions de Jeannine Yon-Kahn et de David Perahia, au laboratoire d'Enzymologie Physico-Chimique et Moléculaire, en collaboration avec François Colonna-Cesari et Orlando Tapia (Université d'Uppsala).

1986 : Service militaire, au Centre de Recherche du Service de Santé des Armées, à Nantes puis Paris.

1985 : DEA d'Enzymologie, Université de Paris XI. Stage avec David Perahia.

1984 : DEA de Neurosciences, Université de Paris VI. Stage chez Jacques Goldberg et Remy Chauvin (Sorbonne). Etude de l'effet des drogues sur le comportement constructeur de l'araignée *Zygiella x-notata*. Travail en partie effectué au Laboratoire National de Contrôle Anti-Dopage.

1983 : Maîtrise de Biochimie, Université de Paris VI. Options Psychophysiologie et Embryologie Moléculaire.

1979 : Baccalauréat, série C (Paris).

1962 : Montée à Paris.

1961 : Naissance (Lyon).

Références

- [1] Zavodszky P, Kardos J, Svingor A, Petsko GA. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc Natl Acad Sci USA* 1998;95(13) :7406–7411.
- [2] Miyashita O, Onuchic J, Wolynes P. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci USA* 2003;100(22) :12570–12575.
- [3] Kraulis P. Molscript : a program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 1991 ;24 :946–950.
- [4] McCammon JA, Gelin BR, Karplus M, Wolynes P. The hinge-bending mode in lysozyme. *Nature* 1976 ;262 :325–326.
- [5] Levy R, Perahia D, Karplus M. Molecular dynamics of an alpha-helical polypeptide : temperature dependance and deviation from harmonic behavior. *Proc Natl Acad Sci USA* 1982;79 :1346–1350.
- [6] Noguti T, Go N. Collective variable description of small-amplitude conformational fluctuations in a globular protein. *Nature* 1982;296 :776–778.
- [7] Brooks B, Karplus M. Harmonic dynamics of proteins : normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA* 1983; 80(21) :6571–6575.
- [8] Go N, Noguti T, Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc Natl Acad Sci USA* 1983;80 :3696–3700.
- [9] Levitt M, Sander C, Stern P. Normal-mode dynamics of a protein : Bovine pancreatic trypsin inhibitor. *Int J Quant Chem* 1983;10 :181–199.
- [10] Levitt M, Sander C, Stern P. Protein normal-mode dynamics : trypsin inhibitor, crambin, ribonuclease and lysozyme. *J Mol Biol* 1985;181 :423–447.
- [11] Brooks BR, Karplus M. Normal modes for specific motions of macromolecules : application to the hinge-bending mode of lysozyme. *Proc Natl Acad Sci USA* 1985; 82 :4995–4999.
- [12] Geradin M, Rixen D. *Theorie des vibrations : application a la dynamique des structures*. Paris : Masson, 1996.
- [13] Harrison R. Variational calculation of the normal modes of a large macromolecules : Methods and some initial results. *Biopolymers* 1984;23 :2943–2949.
- [14] Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins* 1999 ;35(2) :133–152.
- [15] Schaefer M, Karplus M. A comprehensive analytical treatment of continuum electrostatics. *J Phys Chem* 1996;100(5) :1578–1599.
- [16] Sharp K, Honig B. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *The Journal of Physical Chemistry* 1990; 94(19) :7684–7692.
- [17] Tama F, Valle M, Frank J, Brooks III CL. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc Natl Acad Sci USA* 2003;100(16) :9319–9323.

- [18] Tirion M. Low-amplitude elastic motions in proteins from a single-parameter atomic analysis. *Phys Rev Lett* 1996 ;77 :1905–1908.
- [19] Goldstein H. *Classical Mechanics*. Reading, MA : Addison-Wesley, 1950.
- [20] Sanejouand YH. Ph. D. Thesis. Orsay, France : Université de Paris XI, 1990.
- [21] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm : A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* 1983 ;4 :187–217.
- [22] Dauchez M, Derreumaux P, Vergoten G. Vibrational Molecular Force Field of Model Compounds with Biologic Interest. 11. Harmonic Dynamics of Both homers of Glucose in the Crystalline State. *Journal of Computational Chemistry* 1992 ; 14(3,263-277).
- [23] Hao MH, Harvey S. Analyzing the normal mode dynamics of macromolecules by the component synthesis method. *Biopolymers* 1992 ;32 :1393–1405.
- [24] Marques O, Sanejouand YH. Hinge-bending motion in citrate synthase arising from normal mode calculations. *Proteins* 1995 ;23 :557–560.
- [25] Mouawad L, Perahia D. Dimb : Diagonalization in a mixed basis. A method to compute low-frequency normal modes for large macromolecules. *Biopolymers* 1993 ; 33 :569–611.
- [26] Perahia D, Mouawad L. Computation of low-frequency normal modes in macromolecules : improvements to the method of diagonalization in a mixed basis and application to hemoglobin. *Comput Chem* 1995 ;19(3) :241–246.
- [27] Durand P, Trinquier G, Sanejouand YH. A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers* 1994 ;34 :759–771.
- [28] Tama F, Gadea FX, Marques O, Sanejouand YH. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins* 2000 ;41(1) :1–7.
- [29] Lazaridis T. Effective energy function for proteins in lipid membranes. *Proteins* 2003 ;52(2) :176–192.
- [30] Li G, Cui Q. A coarse-grained normal mode approach for macromolecules : an efficient implementation and application to $ca(2+)$ -atpase. *Biophys J* 2002 ;83(5) :2457–2474.
- [31] Nicolay S, Sanejouand YH. Functional modes of proteins are among the most robust. *Phys Rev Lett* 2006 ;96(7) :078104.
- [32] Elber R, Karplus M. Multiple conformational states of proteins : A molecular dynamics analysis of myoglobin. *Science* 1987 ;235 :318–321.
- [33] Lamy A, Souaille M, Smith J. Simulation evidence for experimentally detectable low-temperature vibrational inhomogeneity in a globular protein. *Biopolymers* 1996 ;39(3) :471–478.
- [34] Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 1997 ;2 :173–181.
- [35] Kondrashov D, Cui Q, Phillips G. Optimization and evaluation of a coarse-grained model of protein motion using x-ray crystal data. *Biophys J* 2006 ;91(8) :2760–2767.
- [36] Hinsen K. Analysis of domain motions by approximate normal mode calculations. *Proteins* 1998 ;33 :417–429.

- [37] Atilgan A, Durell S, Jernigan R, Demirel M, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 2001 ; 80(1) :505–515.
- [38] Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. *Prot Eng* 2001 ;14(1) :1–6.
- [39] Kondrashov DA, Van Wynsberghe AW, Bannen RM, Cui Q, Phillips GN Jr. Protein structural variation in computational models and crystallographic data. *Structure* 2007 ;15(2) :169–177.
- [40] Miyazawa S, Jernigan R. Estimation of effective inter-residue contact energies from protein crystal structures : quasi-chemical approximation. *Macromolecules* 1985 ; 18 :534–552.
- [41] Suhre K, Sanejouand YH. Elnémo : a normal mode server for protein movement analysis and the generation of templates for molecular replacement. *Nucl Ac Res* 2004 ;32 :W610–W614.
- [42] Hollup S, Salensminde G, Reuter N. Webnm@ : a web application for normal mode analyses of proteins. *BMC Bioinformatics* 2005 ;6(1) :52. ISSN 1471-2105.
- [43] Lindahl E, Azuara C, Koehl P, Delarue M. Nomad-ref : visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucl Ac Res* 2006 ;34 :W52–W56.
- [44] Eyal E, Yang LW, Bahar I. Anisotropic network model : systematic evaluation and a new web interface. *Bioinformatics* 2006 ;22(21) :2619–2627.
- [45] Diamond R. On the use of normal modes in thermal parameter refinement : theory and application to the bovine pancreatic trypsin inhibitor. *Acta Cryst A* 1990 ; 46(6) :425–435.
- [46] Kidera A, Go N. Refinement of Protein Dynamic Structure : Normal Mode Refinement. *Proc Natl Acad Sci USA* 1990 ;87(10) :3718–3722.
- [47] Kidera A, Go N. Normal mode refinement : crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data. *J Mol Biol* 1992 ;225(2) :457–75.
- [48] Kidera A, Inaka K, Matsushima M, Go N. Normal mode refinement : crystallographic refinement of protein dynamic structure. II. Application to human lysozyme. *J Mol Biol* 1992 ;225(2) :477–86.
- [49] Micheletti C, Lattanzi G, Maritan A. Elastic properties of proteins : insight on the folding process and evolutionary selection of native structures. *J Mol Biol* 2002 ; 321 :909–921.
- [50] Taketomi H, Ueda Y, Go N. Studies on protein folding, unfolding and fluctuations by computer simulation. i. the effect of specific amino acid sequence represented by specific inter-unit interactions. *Int J Pept Prot Res* 1975 ;7(6) :445–459.
- [51] Delarue M, Sanejouand YH. Simplified normal modes analysis of conformational transitions in dna-dependant polymerases : the elastic network model. *J Mol Biol* 2002 ;320 :1011–1024.
- [52] Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M. Normal mode analysis of macromolecular motions in a database framework : developing mode concentration as a useful classifying statistic. *Proteins* 2002 ;48(4) :682–695.

- [53] Hinsen K, Thomas A, Field MJ. Analysis of domain motions in large proteins. *Proteins* 1999;34 :369–382.
- [54] Gerstein M, Krebs W. A database of macromolecular motions. *Nucl Acid Res* 1998; 26 :4280–4290.
- [55] Tama F, Brooks C. Symmetry, form, and shape : Guiding principles for robustness in macromolecular machines. *Annu Rev Biophys Biomol Struct* 2006;35 :115–33.
- [56] Neri M, Anselmi C, Cascella M, Maritan A, Carloni P. Coarse-grained model of proteins incorporating atomistic detail of the active site. *Phys Rev Lett* 2005; 95(21) :218102.
- [57] Elber R, Karplus M. A method for determining reaction paths in large molecules : Application to myoglobin. *Chemical Physics Letters* 1987;139 :375–380.
- [58] Ech-Cherif El-Kettani M, Durup J. Theoretical determination of conformational paths in citrate synthase. *Biopolymers* 1992;32 :561–574.
- [59] Fischer S, Karplus M. Conjugate peak refinement : an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chemical Physics Letters* 1992;194 :252–261.
- [60] Guilbert C, Perahia D, Mouawad L. A method to explore transition paths in macromolecules. Applications to hemoglobin and phosphoglycerate kinase. *Computer Physics Communications* 1995;91 :263–273.
- [61] Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints : molecular dynamics of n-alkanes. *J Comput Phys* 1977;23 :327–341.
- [62] Space B, Rabitz H, Askar A. Long time scale molecular dynamics subspace integration method applied to anharmonic crystals and glasses. *The Journal of Chemical Physics* 1993;99(11) :9070–9079.
- [63] Askar A, Space B, Rabitz H. Subspace method for long time scale molecular dynamics. *Journal of Physical Chemistry* 1995;99(19) :7330–7338.
- [64] Elezgaray J, Sanejouand YH. Modeling large-scale dynamics of proteins. *Biopolymers* 1998;46 :493–501.
- [65] Chun H, Padilla C, Chin D, Watanabe M, Karlov V, Alper H, Soosaar K, Blair K, Becker O, Caves L, Nagle R, Haney D, Farmer B. Mbo(n)d : A multibody method for long-time molecular dynamics simulations. *J Comput Chem* 2000;21(3) :159–184.
- [66] Elezgaray J, Marcou G, Sanejouand YH. Exploring the natural conformational changes of the C-terminal domain of calmodulin. *Phys Rev E* 2002;66 :31908–31915.
- [67] Palczewski K, Kumasaka T, Hori T, Behnke C, Motoshima H, Fox B, Le Trong I, Teller D, Okada T, Stenkamp R, Yamamoto M, Miyano M. Crystal structure of rhodopsin : A G protein-coupled receptor. *Science* 2000;289 :739–745.
- [68] Farrens DL, Altenbach C, Yang K, Hubbell WL, Khorana HG. Requirement of Rigid-Body Motion of Transmembrane Helices for Light Activation of Rhodopsin. *Science* 1996;274(5288) :768–770.
- [69] Nikiforovich G, Marshall G. Three-dimensional model for meta-II rhodopsin, an activated G-protein-coupled receptor. *Biochemistry* 2003;42(30) :9110–9120.

- [70] Salom D, Lodowski DT, Stenkamp RE, Trong IL, Golczak M, Jastrzebska B, Harris T, Ballesteros JA, Palczewski K. Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *PNAS* 2006 ;103(44) :16123–16128.
- [71] Sukharev SI, Blount P, Martinac B, Kung C. Mechanosensitive channels of *Escherichia Coli* : the *mscL* gene, protein, and activities. *Annual Review of Physiology* 1997 ;59(1) :633–657.
- [72] Valadie H, Lacapere JJ, Sanejouand YH, Etchebest C. Dynamical properties of the *mscL* of *Escherichia Coli* : a normal mode analysis. *J Mol Biol* 2003 ;332 :657–674.
- [73] Delarue M, Dumas P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc Natl Acad Sci USA* 2004 ;101(18) :6957–6962.
- [74] Hinsén K, Reuter N, Navaza J, Stokes DL, Lacapere JJ. Normal mode-base fitting of atomic structure into electron density maps : Application to sarcoplasmic reticulum Ca-atpase. *Biophys J* 2005 ;88(2) :818–827.
- [75] Mitra K, Schaffitzel C, Shaikh T, Tama F, Jenni S, Brooks 3rd C, Ban N, Frank J. Structure of the *E. Coli* protein-conducting channel bound to a translating ribosome. *Nature* 2005 ;438 :318–324.
- [76] Jiang X, Jayaram H, Kumar M, Ludtke SJ, Estes MK, Prasad BVV. Cryoelectron Microscopy Structures of Rotavirus NSP2-NSP5 and NSP2-RNA Complexes : Implications for Genome Replication. *J Virol* 2006 ;80(21) :10829–10835.
- [77] Suhre K, Sanejouand YH. On the potential of normal mode analysis for solving difficult molecular replacement problems. *Acta Cryst D* 2004 ;60 :796–799.
- [78] Jeong JI, Lattman EE, Chirikjian GS. A method for finding candidate conformations for molecular replacement using relative rotation between domains of a known structure. *Acta Crystallographica Section D* 2006 ;62(4) :398–409.
- [79] Trapani S, Abergel C, Gutsche I, Horcajada C, Fita I, Navaza J. Combining experimental data for structure determination of flexible multimeric macromolecules by molecular replacement. *Acta Crystallographica Section D* 2006 ;62(5) :467–475.
- [80] Phan J, Shi ZD, Burke Jr TR, Waugh DS. Crystal structures of a high-affinity macrocyclic peptide mimetic in complex with the *grb2* sh2 domain. *J Mol Biol* 2005 ;353(1) :104–115.
- [81] Abergel C, Chenivesse S, Byrne D, Suhre K, Arondel V, Claverie JM. Mimivirus tyrrs : preliminary structural and functional characterization of the first amino-acyl trna synthetase found in a virus. *Acta Cryst F* 2005 ;61(2) :212–215.
- [82] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *Phaser* crystallographic software. *Journal of Applied Crystallography* 2007 ;40(4) :658–674.
- [83] Moore JP, Doms RW. The entry of entry inhibitors : A fusion of science and medicine. *PNAS* 2003 ;100(19) :10598–10602.
- [84] Ryu S, Kwong P, Truneh A, Porter T, Arthos J, Rosenberg M, Dai X, Xuong N, Axel R, Sweet R, Hendrickson W. Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature* 1990 ;348 :419–426.
- [85] Wang J, Yan Y, Garret T, Liu J, Rodgers D, Garlick R, Tarr G, Husain Y, Reinherz E, Harrison S. Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature* 1990 ;348 :411–418.

- [86] Sanejouand YH. Normal-mode analysis suggests important flexibility between the two n-terminal domains of CD4 and supports the hypothesis of a conformational change in CD4 upon HIV binding. *Prot Eng* 1996 ;9(8) :671–677.
- [87] Sanejouand YH. Rôle du changement de conformation du CD4 lors de la fusion VIH/cellule. *C R Acad Sci Paris, Sciences de la vie* 1997 ;320 :163–170.
- [88] Clayton L, Hussey R, Steinbrich R, Ramachandran H, Husain Y, Reinherz E. Substitution of murine for human CD4 residues identifies amino acids critical for HIV-gp120 binding. *Nature* 1988 ;335 :363–366.
- [89] Mizukami T, Fuerst T, Berger E, Moss B. Binding region for human immunodeficiency virus (HIV) and epitopes for HIV-blocking monoclonal antibodies of the CD4 molecule defined by site-directed mutagenesis. *Proc Natl Acad Sci USA* 1988 ; 85 :9273–9277.
- [90] Schockmel G, Somoza C, Davis S, Williams A, Healey D. Construction of a binding site for human immunodeficiency virus type 1 gp120 in rat CD4. *J Exp Med* 1992 ; 175(1) :301–304.
- [91] Burkly L, Olson D, Shapiro R, Winkler G, Rosa J, Thomas D, Williams C, Chisholm P. Inhibition of HIV infection by a novel CD4 domain 2-specific monoclonal antibody. dissecting the basis for its inhibitory effect on HIV-induced cell fusion. *J Immunol* 1992 ;149(5) :1779–1787.
- [92] Moore J, Sattentau Q, Klasse P, Burkly L. A monoclonal antibody to CD4 domain 2 blocks soluble CD4-induced conformational changes in the envelope glycoproteins of human immunodeficiency virus type 1 (HIV-1) and HIV-1 infection of CD4+ cells. *J Virol* 1992 ;66(8) :4784–4793.
- [93] Matthias L, Yam P, Jiang X, Vandegraaff N, Li P, Pountourios P, Donoghue N, Hogg P. Disulfide exchange in domain 2 of CD4 is required for entry of HIV-1. *Nat Immunol* 2002 ;3(8) :727–732.
- [94] Lynch G, Sloane A, Raso V, Lai A, Cunningham A. Direct evidence for native CD4 oligomers in lymphoid and monocytoid cells. *Eur J Immunol* 1999 ;29 :2590–2602.
- [95] Rauk A. Chiroptical properties of disulfides - abinitio studies of dihydrogen disulfide and dimethyl disulfide. *J Am Chem Soc* 1984 ;106(22) :6517–6524.
- [96] Katz B, Kossiakoff A. The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. *J Biol Chem* 1986 ;261(33) :15480–15485.
- [97] Hogg P. Disulfide bonds as switches for protein function. *Trends in Biochemical Sciences* 2003 ;28 :210–214.
- [98] Zahn R, Liu A, Luhrs T, Riek R, von Schroetter C, Lopez Garcia F, Billeter M, Calzolari L, Wider G, Wuthrich K. NMR solution structure of the human prion protein. *PNAS* 2000 ;97(1) :145–150.
- [99] Knaus K, Morillas M, Swietnicki W, Malone M, Surewicz W, Yee V. Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat Struct Biol* 2001 ;8(9) :770–774.
- [100] Bodian D, Jones E, Harlos K, Stuart D, Davis S. Crystal structure of the extracellular region of the human cell adhesion molecule CD2 at 2.5a resolution. *Structure* 1994 ;2(8) :755–766.

- [101] Murray A, Lewis S, Barclay A, Brady R. One sequence, two folds : a metastable structure of CD2. *Proc Natl Acad Sci USA* 1995;92 :7337–7341.
- [102] Bennett M, Choe S, Eisenberg D. Domain swapping : Entangling alliances between proteins. *Proc Natl Acad Sci USA* 1994;91(8) :3127–3131.
- [103] Liu Y, Eisenberg D. 3D domain swapping : as domains continue to swap. *Protein Sci* 2002;11 :1285–1299.
- [104] Sanejouand YH. Domain-swapping of CD4 upon dimerization. *Proteins : Structure, Function and Bioinformatics* 2004;57(1) :205–212.
- [105] Kwong P, Wyatt R, Robinson J, Sweet R, Sodroski J, Hendrickson W. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 1998;393 :648–659.
- [106] Maekawa A, Schmidt B, de St Groth BF, Sanejouand YH, Hogg PJ. Evidence for a domain-swapped CD4 dimer as the coreceptor for binding to class II MHC. *J Immunol* 2006;176 :6873–6878.
- [107] Sattentau Q, Dalgleish A, Weiss R, Beverley P. Epitopes of the CD4 antigen and HIV infection. *Science* 1986;234 :1120–1123.
- [108] Jameson B, Rao P, Kong L, Hahn B, Shaw G, Hood L, Kent S. Location and chemical synthesis of a binding site for HIV-1 on the CD4 protein. *Science* 1988;240 :1335–1339.
- [109] Peterson A, Seed B. Genetic analysis of monoclonal antibody and HIV binding sites on the human lymphocyte antigen CD4. *Cell* 1988;54 :65–72.
- [110] Ashkenazi A, Presta L, Marsters S, Camerato T, Rosenthal K, Fendly B, Capon D. Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. *Proc Natl Acad Sci USA* 1990;87 :7150–7154.
- [111] Bahar I, Jernigan RL. Vibrational dynamics of transfer RNAs : comparison of the free and synthetase-bound forms. *J Mol Biol* 1998;281(5) :871–884.
- [112] Bahar I, Erman B, Jernigan RL, Atilgan AR, Covell DG. Collective motions in HIV-1 reverse transcriptase : Examination of flexibility and enzyme function. *J Mol Biol* 1999;285(3) :1023–1037.
- [113] Bahar I, Jernigan R. Cooperative fluctuations and subunit communication in tryptophan synthase. *Biochemistry* 1999;38(12) :3478–3490. ISSN 0006-2960.
- [114] Hayward S, Kitao A, Go N. Harmonicity and anharmonicity in protein dynamics : a normal mode analysis and principal component analysis. *Proteins* 1995;23(2) :177–186.
- [115] Moritsugu K, Miyashita O, Kidera A. Vibrational energy transfer in a protein molecule. *Phys Rev letters* 2000;85(18) :3970–3973.
- [116] Sagnella D, Straub J, Thirumalai D. Time scales and pathways for kinetic energy relaxation in solvated proteins : Application to carbonmonoxy myoglobin. *J Chem Phys* 2000;113(17) :7702–7711.
- [117] Leitner DM. Vibrational energy transfer in helices. *Phys Rev Lett* 2001;87(18) :188102.
- [118] Segré E, editor. *Collected papers of Enrico Fermi*. Chicago : University of Chicago Press, 1965.

- [119] Dauxois T, Khomeriki R, Piazza F, Ruffo S. The anti-FPU problem. *Chaos* 2005 ; 15(1) :015110.
- [120] Flach S, Kladko K, Willis CR. Localized excitations in two-dimensional hamiltonian lattices. *Phys Rev E* 1994 ;50(3) :2293–2303.
- [121] Tsironis GP, Aubry S. Slow relaxation phenomena induced by breathers in nonlinear lattices. *Phys Rev Lett* 1996 ;77(26) :5225–5228.
- [122] Piazza F, Lepri S, Livi R. Cooling nonlinear lattices toward energy localization. *Chaos* 2003 ;13(2) :637.
- [123] Piazza F, De Los Rios P, Sanejouand YH. Slow energy relaxation of macromolecules and nanoclusters in solution. *Phys Rev Lett* 2005 ;94 :145502.
- [124] Piazza F, Lepri S, Livi R. Slow energy relaxation and localization in 1D lattices. *Journal of Physics A : Mathematical and General* 2001 ;34(46) :9803–9814.
- [125] Juanico B, Sanejouand YH, Piazza F, De Los Rios P. Discrete breathers in nonlinear network models of proteins. *Phys Rev Lett* 2007 ;99 :238104.
- [126] Davydov A. Solitons and energy transfer along protein molecules. *J Theor Biol* 1977 ;66 :379–387.
- [127] Sacquin-Mora S, Laforet E, Lavery R. Locating the active sites of enzymes using mechanical properties. *Proteins* 2007 ;67(2) :350–359.
- [128] Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci USA* 2002 ; 99(20) :12562–12566.
- [129] Go N, Taketomi H. Respective roles of short- and long-range interactions in protein folding. *Proc Natl Acad Sci USA* 1978 ;75(2) :559–563.
- [130] Elezgaray J, Sanejouand YH. Modal dynamics of proteins in water. *J Comput Chem* 2000 ;21(14) :1274–1282.
- [131] Elezgaray J, Marcou G, Sanejouand YH. Coupling overall rotations with modal dynamics. *Theor Chem Acc* 2001 ;106 :62–68.
- [132] Suhre K, Navaza J, Sanejouand YH. Norma : a tool for flexible fitting of high resolution protein structures into low resolution electron microscopy derived density maps. *Act Cryst D* 2006 ;62(9) :1098–1100.
- [133] Sanejouand YH, Tapia O. Low-frequency domain motions in the carboxy terminal fragment of L7/L12 ribosomal protein studied with molecular dynamics techniques : are these movements model independant ? *J Phys Chem* 1995 ;99(15) :5698–5704.
- [134] Trinquier G, Sanejouand YH. Which effective property of amino acids is best preserved by the genetic code ? *Prot Eng* 1998 ;11(3) :153–169.
- [135] Trinquier G, Sanejouand YH. New protein-like properties of cubic lattice models. *Phys Rev E* 1999 ;59(1) :942–946.
- [136] Sanejouand YH, Trinquier G. L'apport des modèles sur réseau cubique à l'étude des propriétés des protéines. *Bull Soc Fr Phys* 2000 ;125 :25–27.
- [137] Sanejouand YH, Trinquier G. Protein-like properties of simples models. Berlin : Springer, 2003 ; pp. 147–153, pp. 147–153.
- [138] Alary F, Durup J, Sanejouand YH. Molecular dynamics study of the hydration structure of an antigen-antibody complex. *J Phys Chem* 1993 ;97(51) :13864–13876.

- [139] Durup J, Alary F, Sanejouand YH. Molecular Dynamics simulation of an antigen-antibody complex : hydration structure and dissociation dynamics. New-York : Plenum Publ. Co, 1994 ; pp. 127–137, pp. 127–137.
- [140] Fabiano AS, Allouche D, Sanejouand YH, Paillous N. Synthesis of a cationic pyro-phosphoribide derivative and studies of its aggregation process in aqueous solution. *Photochem Photobiol* 1997 ;66(3) :336–345.
- [141] Periole X, Allouche D, Daudey JP, Sanejouand YH. Simple two-body cation-water interaction potentials derived from ab initio calculations. comparison to results obtained with an empirical approach. *J Phys Chem B* 1997 ;101(25) :5018–5025.
- [142] Periole X, Allouche D, Ramirez-Solis A, Ortega-Blake I, Daudey JP, Sanejouand YH. New effective potentials extraction method for the interaction between cations and water. *J Phys Chem B* 1998 ;102(43) :8579–8587.
- [143] Allouche D, Parello J, Sanejouand YH. Ca/Mg exchange in parvalbumin and other EF-hand proteins. A theoretical study. *J Mol Biol* 1999 ;285(2) :857–873.
- [144] Gomes E, Sagot E, Gaillard C, Laquitaine L, Poinot B, Sanejouand YH, Delrot S, Coutos-Thévenot P. Nonspecific lipid-transfer protein genes expression in grape (*Vitis Sp.*) cells in response to fungal elicitor treatments. *Mol Path Micr Int* 2003 ; 16(5) :456–464.
- [145] Willson M, Alric I, Perie J, Sanejouand YH. Yeast hexokinase inhibitors designed from the 3D enzyme structure rebuilding. *J Enzyme Inhibition* 1997 ;12 :101–121.
- [146] Willson M, Sanejouand YH, Perie J, Hannaert V, Opperdoes F. Sequencing, modeling, and selective inhibition of *Trypanosoma Brucei* hexokinase. *Chem & Biol* 2002 ;9 :839–847.

Index

- ACE, voir Solvant
Adénylate kinase, 10, 23
ADN polymérase β , 23, 24, 27
Approximation harmonique, 4, 6, 15, 19, 29, 39
- Barrière d'énergie, 28, 40
Base, 4, 25
BNM, voir RTB
- Catalyse, 40
CD2, 35, 36
CD4, 33, 34, 36–38
Changement de conformation, 4, 5, 13, 14, 19–21, 23, 26, 31, 33, 36, 41
CHARMM, 8–10, 22, 23, 25, 40
Chemin de transition, 13, 28
Citrate synthase, 9, 13, 33, 40
Coordonnées, 6, 8, 15, 17, 19, 25, 28–30, 41
Cristallographie, 13, 15, 16, 19, 32, 35, 36, 41
Cryomicroscopie électronique, 32
- Densité de modes, 9
Diagonalisation, 8, 9, 17
DIMB, 9
Discrete Breather, voir Localisation non-linéaire d'énergie
Domaines, 4, 9, 20, 33, 36, 38
Dynamique moléculaire, 4, 5, 36
- Echange hydrogène-deutérium, 4
EEF1, voir Solvant
ENM, 15–27, 39
Entrelacement de domaines, 35–37
Equations du mouvement, 6, 29
Equipartition de l'énergie, 6
- Facteurs B, 16, 18, 19
Friction, 40
- GNM, 16, 38
Guanylate kinase, 14, 17, 20, 21, 25, 27
- Hessien, 7
Hexokinase, 4, 9, 13
- Isopropylmalate dehydrogenase, 4
- Lanczos, 9, 13
Liaison hydrogène, 40
Localisation non-linéaire d'énergie, 39–41
- Métadynamique, 41
Minimisation d'énergie, 14, 23, 25–27, 32, 36
- Modèle de réseau
 Anharmonique, 39–41
 Anisotrope, voir ENM
 Elastique, voir ENM
 Gaussien, voir GNM
- Modes normaux
 Calcul standard, 15, 20, 22–27, 34
 Robustesse, 22, 23
- Mouvement fonctionnel, voir Changement de conformation
- Orthogonalité, 9, 12, 25
- PHASER, 32
Pont disulfure, 34–37
Prion, 35, 36
Protéase du VIH-1, 8, 12, 23, 25, 27
- Résonance magnétique nucléaire, voir RMN
Recouvrement, 12–14, 19, 21, 22, 24, 27, 31
- Rhodopsine, 30, 31
Rigidité, 4, 10, 17, 21, 30, 34, 41
Ritz-Raleigh, 4, 9
RMN, 19, 35, 41
Robustesse, 12
RTB, 9, 10, 12, 18
- Solvant, 4, 5, 10, 34, 40
Spin labelling, 30
- Température, 4, 6, 14, 15, 29, 30, 36, 39–41
- Valeurs propres, 7, 8, 10, 12, 16, 25