

Richard Lavery : « La modélisation des molécules de la vie »

Le besoin de modèles

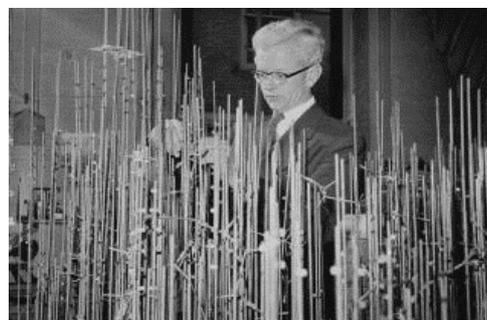
Depuis toujours les scientifiques, comme les ingénieurs, les architectes et même les enfants, ont eu besoin de construire des modèles pour les aider à comprendre le monde complexe qui nous entoure. Néanmoins, les modèles dont je vais parler ici ont dû attendre la fin du dix-neuvième siècle pour voir le jour. La raison principale pour cela est que nous allons parler du monde des atomes et leurs assemblages pour former des molécules et des macromolécules. Même si l'existence des atomes a été postulée par le philosophe grec Démocrite [1], 400 ans avant notre ère, il a fallu attendre les années 1900 pour accumuler suffisamment d'évidence en faveur de l'existence des atomes pour convaincre le monde scientifique. A ce sujet, il est remarquable de noter que les chimistes, qui avaient utilisé les formules pour décrire la constitution des molécules depuis le début du dix-neuvième siècle (par exemple, H₂O, deux "parts" d'hydrogène pour une "part" d'oxygène), ont été parmi les plus difficiles à convaincre que "part" pouvait se traduire par atome et ceci malgré les travaux de leurs illustres prédécesseurs, notamment Lavoisier et Dalton, en faveur de la théorie atomique [1].

C'est donc aux alentours de 1900 que les premiers modèles représentant le nombre et l'organisation spatiale des atomes au sein des molécules ont vu le jour. Ces modèles comme vous pouvez le voir dans l'illustration, ressemblent beaucoup aux modèles que l'on trouve dans les salles de cours et les laboratoires aujourd'hui. Il y a naturellement différents types de représentation pour satisfaire les besoins des utilisateurs. Certaines emploient des sphères tronquées pour illustrer l'espace occupé par chaque atome (modèles Corey-Pauling-Kolton ou CPK des années '50), tandis que d'autres se concentrent sur la conformation des liaisons qui sont représentées par des fils métalliques (modèles Dreiding des années '60).

De tels modèles fonctionnent bien pour des molécules composées de quelques dizaines d'atomes, mais posent des problèmes pour construire des macromolécules formées de milliers, voir de dizaines de milliers d'atomes. Cette difficulté se fait ressentir au début des années soixante quand les premières structures des protéines ont été obtenues par les cristallographes à Cambridge. A



Un jeu de modèles datant de 1905

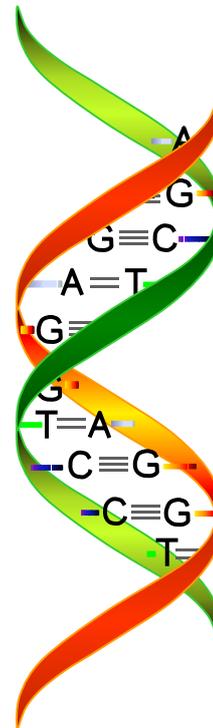


John Kendrew construisant un modèle de myoglobine

partir de ce moment, il a fallu chercher d'autres moyens de modélisation plus rapides à mettre en place, moins chers, et plus maniables. C'est l'ordinateur et le passage aux modèles virtuels qui a fourni la réponse. Mais avant de parler de ces développements il y a un exemple remarquable de modélisation "classique" qui mérite discussion.

L'ADN - un exemple phare du vingtième siècle

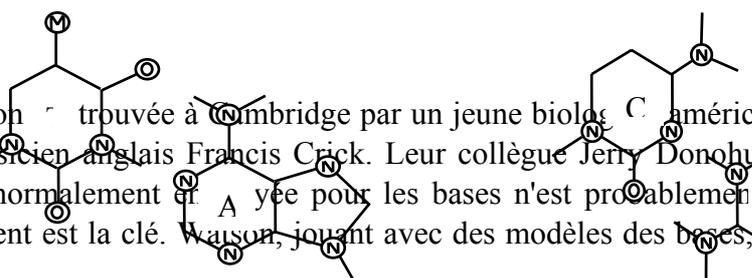
L'histoire de l'ADN (acide désoxyribonucléique) commence en 1869 quand Friedrich Miescher isole une substance de noyaux des cellules humaines qu'il dénomme "nucléine". Il s'agit en fait d'un mélange complexe de protéines et d'ADN. Il faut attendre le travail des chimistes du vingtième siècle et notamment les efforts de Phoebus Levene à l'Institut Rockefeller à New York pour connaître la structure chimique de la molécule qui se révèle être de longues chaînes composées d'une alternance de phosphates et de sucres. Sur chaque sucre est accrochée une base. Quatre bases sont identifiées : adénine (A), cytosine (C), guanine (G) et thymine (T). En formulant l'hypothèse que ces quatre bases se répètent de façon régulière le long de la chaîne d'ADN (par exemple, ACGT-ACGT-ACGT-.....) Levene relègue l'ADN à la famille de polymères jouant probablement un rôle structural au sein de la cellule. Mais, Levene se trompe et comme Oswald Avery, un autre scientifique de l'Institut Rockefeller, montre en 1944, l'ADN a le pouvoir de transformer des bactéries. L'ADN porte donc le message génétique et une course est lancée pour trouver sa structure et comprendre son fonctionnement.



Plusieurs informations sont connues. Chargaff démontre que les bases sont présentes dans des rapports fixes de telle façon que le rapport de concentrations $[A]/[G]$ est égale au rapport $[T]/[C]$. Astbury et ensuite Rosalind Franklin obtiennent des clichés de diffraction des rayons X à partir des fibres d'ADN et démontrent que la molécule possède une structure hélicoïdale.

Linus Pauling, un des plus grands chimistes du vingtième siècle propose une structure qui ne peut pas être correcte puisqu'il met les bases à l'extérieur et les phosphates en contact au centre de la structure sans tenir compte du fait qu'ils sont chargés négativement et ne peuvent pas se rapprocher ainsi [2].

La solution est trouvée à Cambridge par un jeune biologiste américain James Watson et le physicien anglais Francis Crick. Leur collègue Jerry Donohue explique que la formule normalement utilisée pour les bases n'est probablement pas correcte. Ce changement est la clé. Watson, jouant avec des modèles des bases, voit qu'il peut les



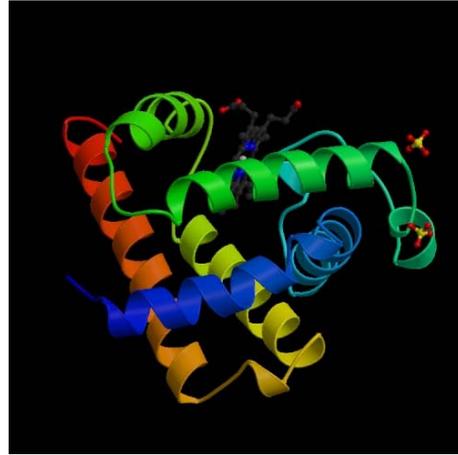
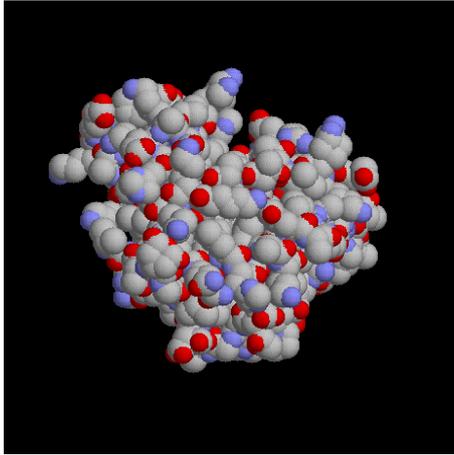
assembler par paire : A avec T, G avec C. Les deux paires ont exactement la même forme et elles peuvent être placées, non pas à l'extérieur de la structure hélicoïdale, mais au centre.

Il crée ainsi la fameuse double hélice qui est devenue une des icônes de notre époque [3]. La construction du modèle, guidée par les informations expérimentales, donne un résultat si simple et si beau qu'il est accepté immédiatement. La double hélice est non seulement compatible avec les données expérimentales, mais suggère également comment l'information génétique passe d'une cellule à une autre. En effet, il suffit de séparer les deux chaînes et de fabriquer de nouvelles doubles hélices en copiant l'information: A dans le premier chaîne donne son complément T, T dans le deuxième chaîne donne son complément A. Il est probable que nous ne verrons jamais plus un modèle moléculaire qui aura un tel impact.

L'arrivée des ordinateurs

Pour aller plus loin avec la modélisation des molécules de la vie, il a fallu une autre étape clé - l'arrivée des ordinateurs. L'envie de calculer plus vite et avec plus de précision a inspiré les ingénieurs depuis longtemps. Qu'il s'agit d'obtenir les tables logarithmiques sans erreur, d'effectuer des calculs de balistique, ou de comprendre une réaction de fission nucléaire, les capacités de calcul humaines sont rapidement dépassées. Quelques pionniers du dix-neuvième siècle comme Charles Babbage ont tenté de résoudre les problèmes à l'aide d'une machine [4]. Plus précisément, une machine "universelle" capable d'effectuer différents types de calcul en suivant une suite d'instructions. Des ingénieurs comme Jacquard travaillant pour l'industrie de soie à Lyon, ont fourni le moyen d'écrire de tels programmes sur des ensembles de cartes perforées. Les plans de Babbage ont été bon (la Musée des Sciences de Londres vient de construire et de faire marcher des éléments de l'ordinateur de Babbage, et notamment son imprimante) mais en 1850 il n'avait ni les bons matériaux ni des outils de fabrication suffisamment précis. Les ordinateurs ont réellement vu le jour pendant la deuxième guerre mondiale quand les calculs rapides sont devenus indispensables pour casser des codes et pour faire avancer le projet Manhattan vers la production de la bombe atomique.

La disponibilité des ordinateurs pour des travaux non militaires date des années soixante. Les chimistes et biologistes n'ont pas attendu pour profiter de leurs possibilités, non seulement pour effectuer des calculs, mais aussi pour créer une nouvelle façon de visualiser des molécules, d'abord par des images fixes imprimées sur papier et ensuite par des images animées grâce au couplage entre l'ordinateur et l'écran cathodique. Dès 1966, Cyrus Levinthal à MIT a mis au point un système capable de représenter la structure d'une protéine et de la manipuler dans l'espace [5]. Depuis, les moyens de visualisation ont progressé de façon remarquable et même un ordinateur familial permet de se plonger dans le monde fascinant des macromolécules biologiques à travers des représentations toujours plus belles. Je vous encourage d'ailleurs de se procurer un des logiciels de visualisation gratuits tels que VMD [6] et d'entreprendre votre propre voyage au sein des protéines et des acides nucléiques (dont les structures sont librement accessibles dans la banque RSCB). Les coordonnées de VMD et du RCSB sont indiquées dans la liste des sites internet ci-dessous.

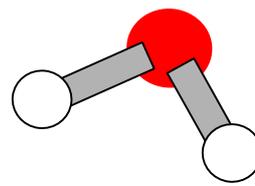
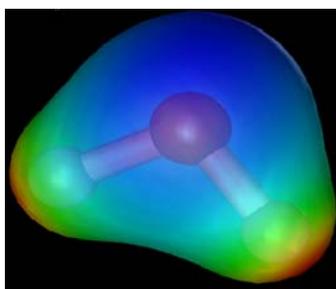


En parallèle, avec le développement du graphisme, les logiciels permettant de modéliser mathématiquement le comportement des molécules ont vu le jour, initialement pour satisfaire les besoins de la spectroscopie en interprétant les spectres en termes de vibrations moléculaires et ensuite pour modéliser la structure, la dynamique et les interactions des molécules dans leur environnement biologique, c'est à dire, entourées de l'eau et d'ions et soumises aux effets de l'excitation thermique. Le développement de tels logiciels continue aujourd'hui en ciblant des représentations moléculaires toujours plus près de la réalité et la capacité de modéliser des systèmes toujours plus grands et plus complexes.

Les molécules

Qu'est qu'il faut pour créer une molécule virtuelle au sein de l'ordinateur ? Peut-on modéliser la dynamique d'une molécule, le processus d'assemblage d'un complexe multi-moléculaire ou le fonctionnement d'un enzyme ? Pour commencer à répondre à ces questions, il faut se rappeler que les molécules, même les macromolécules de la vie, sont très petites. Leur taille se mesure en dizaines de nanomètres (un nanomètre est un mille milliardième d'un mètre, 10^{-9} m) et il faut, par exemple, empiler 30,000 protéines pour atteindre l'épaisseur d'une feuille de papier ! A ces dimensions, c'est la mécanique quantique qui règne; les électrons forment un nuage de densité électronique autour des noyaux atomiques, et obéissent à la fameuse équation de Schrödinger. Dans ce monde quantique toute la chimie est possible, les électrons et les noyaux peuvent être perturbés par des interactions avec la lumière, d'autres rayonnements ou d'autres molécules et les électrons peuvent s'échanger entre différentes molécules en formant et en brisant des liaisons chimiques. Néanmoins, les calculs associés sont complexes et, malgré le progrès remarquable de la chimie quantique, ils sont encore prohibitifs pour la plupart des systèmes macromoléculaires.

Dans ce cas, si on accepte de se limiter aux études de la structure, la dynamique conformationnelle et les interactions physiques des macromolécules nous pouvons retourner vers la mécanique classique de Newton. Dans ce monde les atomes deviennent des billes (avec des tailles et des charges électrostatiques qui dépendent de leurs identités chimiques et de la molécule à la quelle ils appartiennent) et les liaisons chimiques deviennent des ressorts.



D'autres termes simples représentent la déformation des angles de valence, la rotation des angles dièdres et l'équilibre entre l'attractivité des atomes à longue portée et leur répulsion à courte portée. On crée ainsi un "champ de force" qui permet de calculer l'énergie d'un système moléculaire, d'optimiser sa structure en minimisant cette énergie ou encore de suivre sa dynamique à une température donnée (en intégrant l'équation de Newton $\text{Force} = \text{Masse} \times \text{Accélération}$ dans le temps). Quelques décennies de recherche ont permis de raffiner de champs de force suffisamment pour obtenir des résultats en bon accord avec l'expérience. Combinés avec la puissance croissant des ordinateurs, ils sont devenus un moyen efficace pour étudier le comportement des macromolécules biologiques. Nous retournerons vers la double hélice de l'ADN pour montrer un exemple.

Le physicien sonde l'ADN

Dans notre domaine, un des développements les plus excitants de ces dernières années a été la possibilité de manipuler directement une seule molécule [7]. L'ADN est un bon candidat pour de telles expériences puisque, malgré un diamètre de seulement deux nanomètres, sa longueur peut atteindre des centimètres. L'envie de manipuler une seule molécule résulte de l'observation que l'évaporation d'une gouttelette d'eau pouvait étirer l'ADN bien au delà de sa longueur naturelle [8]. Par la modification chimique des extrémités de la molécule (un sort de "scotch" moléculaire), il était ensuite possible d'attraper une molécule d'ADN et de la fixer sur une extrémité à une surface et sur l'autre à une microbille en polystyrène. En tirant sur la microbille il est devenu possible de suivre l'extension de la molécule et de mesurer les forces exercées. Les résultats ont été surprenants puisqu'il s'avère que l'ADN ne se comporte pas comme un ressort simple. Au delà d'une certaine force (environ 70 picoNewtons), la molécule est capable de presque doubler sa longueur sans que la force exercée augmente [9]. L'explication structurale de ce phénomène est venue de la modélisation. En étirant la double hélice dans le monde virtuel de l'ordinateur, nous avons constaté qu'il y a en fait deux chemins d'étirement, soit en déroulant la double hélice, soit en diminuant son diamètre par l'inclinaison des paires de bases. En réalité il est probable que ces deux chemins participent à former la structure étirée qui porte désormais le nom d'ADN-S (*stretched DNA*) [9, 10].

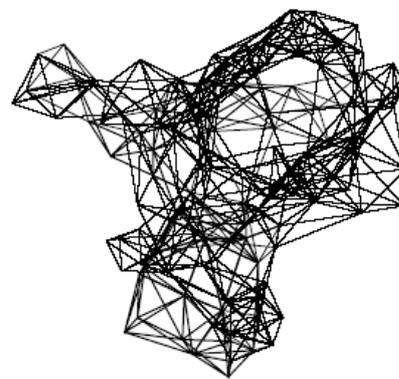
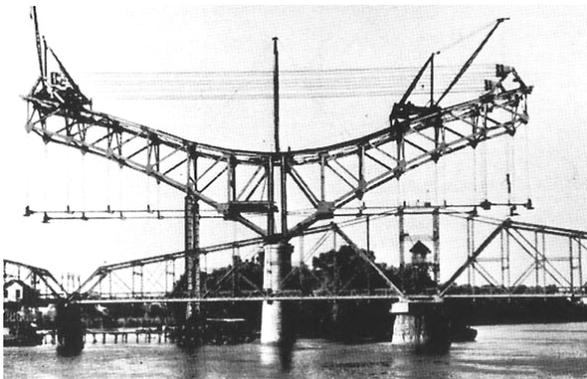


L'entortillement de l'ADN et la structure de la forme dit ADN-P

Par la suite, l'ajout d'une microbille magnétique a permis de contrôler à la fois l'étirement et l'entortillement de la molécule [11]. A nouveau les résultats ont été surprenants. En diminuant le nombre de tours de la double hélice on arrive à séparer les deux brins, mais en augmentant le nombre de tours on a constaté que la molécule s'allonge et qu'on pouvait atteindre une rotation de presque 160° entre les paires de bases successives (contre seulement 34° dans la conformation usuelle de l'ADN). La modélisation de ce phénomène a permis de postuler une nouvelle forme de la double hélice qui se ressemble étrangement à la structure incorrecte proposée par Linus Pauling avant le succès de Watson et Crick. Cette structure, qui est maintenant appelée ADN-P (P pour Pauling) se distingue par la position des bases, qui sont à l'extérieur de la structure, tandis que les brins phosphodiester sont entrelacés au centre [12].

Ces expériences et la modélisation qui a suivi ont montré la complexité de la mécanique de l'ADN. Elles ont aussi servi de base pour une nouvelle domaine scientifique, la physique des molécules uniques, qui continue de fournir des informations sur une gamme de systèmes biologiques (complexes protéine-ADN, la chromatine, les moteurs moléculaires, le fonctionnement des virus, ...) qui sont difficilement accessibles par d'autres types d'expérience.

Partir des ponts pour arriver aux protéines



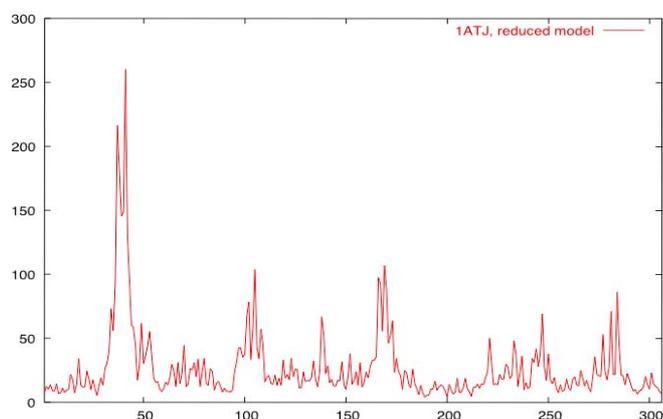
Pour continuer sur le thème de la mécanique des macromolécules, j'aimerais parler un peu des protéines. Les protéines ont des structures plus complexes que l'ADN. D'abord, elles sont formées de polymères (polypeptides) composés de 20 types de sous unités différentes (les acides aminés), plutôt que seulement quatre types pour l'ADN (les nucléotides portant les bases A, C, G et T). Ensuite, dans la plupart des cas, le repliement de la chaîne des acides aminés conduit à des structures compactes et globulaires.

Cette complexité illustre l'importance de la structure des protéines, mais la structure seule n'est pas suffisante pour tout comprendre. On peut raisonnablement assumer que leurs propriétés mécaniques sont également importantes compte tenu des

travaux accomplis par des protéines. Ainsi, plusieurs protéines appartiennent à la catégorie des enzymes et sont capables de catalyser des réactions chimiques avec une spécificité remarquable, d'autres jouent un rôle structural au sein de nos cellules et d'autres encore fonctionnent en tant que moteurs miniatures. Toutes ces tâches nécessitent non seulement des structures particulières, mais aussi des propriétés mécaniques appropriées.

Nous avons tenté de mettre au point des techniques de modélisation pour étudier ces propriétés [13]. Plus précisément, à partir des fluctuations spatiales des acides aminés lors d'une simulation dynamique, nous avons pu calculer des constantes de force correspondant à la difficulté de déplacer un résidu donné par rapport au reste de la structure [14]. Pour accélérer les calculs nous avons utilisé un modèle protéique plus simple comportant seulement quelques points pour chaque acide aminé (plutôt qu'une dizaine d'atomes) et nous avons également remplacé le champ de force classique avec de simples ressorts entre tous les résidus proches. Ainsi modélisée, la protéine ressemble à un objet élastique où la densité de ressorts reflète le repliement de la chaîne polypeptidique.

Compte tenu de la simplicité de notre modèle, nous étions surpris de voir que les propriétés mécaniques des différents résidus pouvaient varier de façon importante au sein d'une seule protéine. La figure montre le résultat à travers le "spectre" de constantes de force pour les acides aminés d'une



péroxydase. Cette protéine contient un groupement hème qui joue un rôle central en catalysant la cassure d'une liaison peroxyde, $R-O-O-R'$, pour former deux alcools, $R-OH$ et $R'-OH$, par l'addition de deux atomes d'hydrogène et de deux électrons. Un petit nombre de résidus sont particulièrement difficiles à déplacer et ont des constantes de force très élevées. Il s'avère que ces résidus sont exactement ceux qui maintiennent le groupement hème en place et sont donc des résidus clés pour le fonctionnement de la protéine.

Après l'étude d'environ 100 protéines, nous avons pu démontrer que les résidus ayant un rôle fonctionnel ont presque toujours des propriétés mécaniques exceptionnelles. Ils sont, dans l'ensemble, tenus de façon beaucoup plus rigide au sein de leurs structures protéiques que les autres résidus. Nous pouvons conclure que cette propriété est importante pour l'activité protéique et que l'évolution a choisi le repliement complexe de chaque protéine non seulement pour placer les résidus clés au bon endroit, mais aussi pour assurer qu'ils y restent.

L'avenir de la modélisation en biologie

Il est toujours dangereux de parler de l'avenir. Néanmoins, dans le domaine de la modélisation on peut faire deux prédictions concernant les développements possibles et souhaitables sans prendre trop de risques.

Premièrement, les ordinateurs vont continuer à progresser en puissance comme en capacité de stockage. Depuis les années quarante jusqu'à nos jours la puissance des processeurs a doublé environ tous les 18 mois. En même temps, les mémoires ont changé de quelques octets à des kilooctets, puis des mégaoctets, des gigaoctets et maintenant des téraoctets. Aujourd'hui certaines voitures ont plus de puissance de calcul que les capsules Apollo des années soixante-dix ! Au delà de la puissance des processeurs individuels, il est aussi devenu courant d'assembler de dizaines, des centaines, voir des milliers de processeurs pour multiplier la puissance disponible. De telles machines sont traditionnellement construites dans les bâtiments des centres de calculs, mais il est aussi possible de créer une machine virtuelle composé d'ordinateurs indépendants. Les efforts de projets tels que "*Screensaver Lifesaver*". Ce projet cible la conception de nouveaux médicaments contre le cancer grâce aux calculs effectués par un logiciel de sauvegarde d'écran installé volontairement par des particuliers sur leurs propres PC (voir la liste des sites internet ci-dessous). Les résultats de *Lifesaver* montrent la puissance de cette approche puisque les calculs effectués ont largement dépassé la puissance des gros centres de calcul conventionnels avec plus de 450,000 heures de calcul sur un total de 3.5 millions de PC à travers le monde.

Deuxièmement, malgré la puissance de calcul qui sera disponible, elle sera toujours insuffisante pour modéliser toute la complexité des systèmes vivants. Aujourd'hui nos efforts portent sur une meilleure compréhension de la structure et de la dynamique de macromolécules individuelles, sur les interactions macromolécule-ligand ou sur les interactions entre deux macromolécules. En revanche, au sein de la cellule, chaque macromolécule se trouve en contact avec des dizaines d'autres dans un milieu hétérogène et dense qui, de surcroît, évolue dans le temps. La plupart de complexes qui se forment dans ce milieu impliquent de multiples macromolécules. Un nombre très important de petites molécules entre et sort des cellules et voyage entre les différents compartiments cellulaires pour passer des messages chimiques, tandis qu'un système de fibres et de moteurs se charge de déplacer des objets moléculaires plus encombrants et participe dans les mouvements et les interactions de la cellule. Finalement, les cellules sont protégées et partitionnées par des membranes lipidiques comportant une gamme impressionnante de canaux et de récepteurs qui se chargent de la communication avec le monde extracellulaire. Par comparaison, nos efforts de modélisation semblent un peu timides. Comprendre la complexité des systèmes vivants au niveau moléculaire nécessitera non seulement toute la puissance informatique disponible, mais aussi toute la créativité des chercheurs pour mettre au point de nouveaux modèles et de nouveaux algorithmes de modélisation.

Remerciements

La recherche aujourd'hui implique plus des équipes que des individus. Je souhaite remercier mes collègues qui ont contribué aux travaux présentés ici. Notamment, pour la modélisation des acides nucléiques et leur manipulation, Anne Lebrun et Krystyna

Zakrzewska, et nos collègues de l'Ecole Normale Supérieure de Paris, Jean-François Allemand, David Bensimon, Didier Chatenay et Vincent Croquette, et pour l'étude de la mécanique des protéines, Fabien Cailliez, Isabelle Navizet, et Sophie Sacquin-Mora. Je remercie également les autres membres du Laboratoire de Biochimie Théorique à Paris avec qui j'ai eu le plaisir de travailler et le CNRS qui a fourni les moyens d'accomplir ce travail.

Quelques sites Internet

Charles Babbage et ses merveilleuses machines à calculer :

www.charlesbabbage.net

www.fourmilab.ch/babbage/sketch.html

L'histoire des ordinateurs :

www.hitmill.com/computers/computerhxl.html#BB

L'histoire de la visualisation des molécules sur ordinateur :

www.umass.edu/microbio/rasmol/history.htm

Le logiciel ORTEP :

<http://www.ornl.gov/sci/ortep/ortep.html>

Les impressionnantes images de David Goodsell :

www.scripps.edu/mb/goodsell/

Les calculs à très grand échelle pour lutter contre le cancer "Screensaver Lifesaver" :

www.chem.ox.ac.uk/curecancer.html

"Imprimer" en trois dimensions :

3dmoleculardesigns.com/

La RCSB : les structures des protéines accessibles à tous:

www.rcsb.org

Visualiser les macromolécules chez vous avec VMD :

www.ks.uiuc.edu/Research/vmd/

Références

1. Pullman, B. (1998). *The atom in the history of human thought* (Oxford: Oxford University Press).
2. Pauling, L., and Corey, R.B. (1953). Structure of the nucleic acids. *Nature* *171*, 346.
3. Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* *171*, 737-738.
4. Swade, D. (2001). *The cogwheel brain* (London: Abacus).
5. Levinthal, C. (1966). Molecular model building by computer. *Scientific American* *214*, 42-52.
6. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph* *14*, 33-38, 27-38.
7. Lavery, R., Lebrun, A., Allemand, J.-F., Bensimon, D., and Croquette, V. (2002). Structure and mechanics of single biomolecules: experiment and simulation. *J Phys (Cond. Mat.)* *14*, R383-R414.
8. Bensimon, D., Simon, A.J., Croquette, V.V., and Bensimon, A. (1995). Stretching DNA with a receding meniscus: Experiments and models. *Physical Review Letters* *74*, 4754-4757.
9. Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J.L., Chatenay, D., and Caron, F. (1996). DNA: an extensible molecule. *Science* *271*, 792-794.
10. Lebrun, A., and Lavery, R. (1996). Modelling extreme stretching of DNA. *Nucleic Acids Res* *24*, 2260-2267.
11. Strick, T.R., Allemand, J.F., Bensimon, D., and Croquette, V. (1998). Behavior of supercoiled DNA. *Biophys J* *74*, 2016-2028.
12. Allemand, J.F., Bensimon, D., Lavery, R., and Croquette, V. (1998). Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proc Natl Acad Sci U S A* *95*, 14152-14157.
13. Navizet, I., Cailliez, F., and Lavery, R. (2004). Probing protein mechanics: residue-level properties and their use in defining domains. *Biophys J* *87*, 1426-1435.
14. Sacquin-Mora, S., and Lavery, R. (2006). Investigating the local flexibility of functional residues in hemoproteins. *Biophys J* *90*, 2706-2717.