EM8BTEM – UPS Toulouse

Introduction à la Modélisation Moléculaire

(dynamique moléculaire, *docking* et criblage virtuel)

Georges Czaplicki, UPS / IPBS-CNRS
cgeorge@ipbs.fr, tél. : 05 61 17 54 04

Dynamique moléculaire

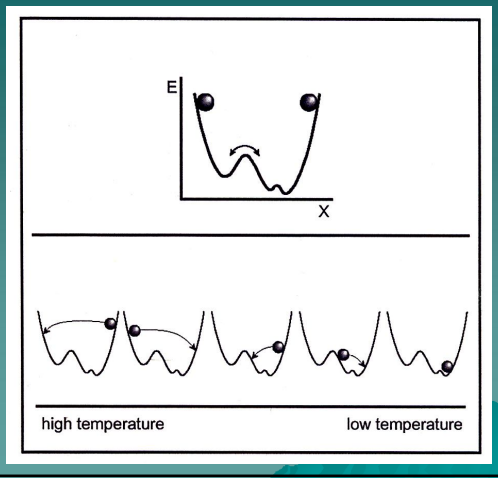


Table 1.2. The evolution of molecular mechanics and dynamics.

Period	System and Size[a]	Trajectory Length[b] [ns]	CPU Time/Computer[c]
1973	Dinucleoside (GpC) in vacuum (8 flexible dihedral angles)	—	—
1977	BPTI, vacuum (58 residues, 885 atoms)	0.01	
1983	DNA, vacuum, 12 & 24 bp (754/1530 atoms)	0.09	several weeks each Vax 780
1984	GnRH, vacuum (decapeptide, 161 atoms)	0.15	
1985	Myoglobin, vacuum (1423 atoms)	0.30	50 days VAX 11/780
1985	DNA, 5 bp (2800 atoms)	0.50	20 hrs Cray X-MP
1989	Phospholipid Micelle (≈ 7,000 atoms)	0.10	
1992	HIV protease (25,000 atoms)	0.10	100 hrs. Cray Y-MP
1997	Estrogen/DNA (36,000 atoms, multipoles)	0.10	22 days HP-735 (8)
1998	DNA, 24 bp (21,000 atoms, PME)	0.50	1 year, SGI Challenge
1998	β-heptapeptide in methanol (≈ 5000/9000 atoms)	200	8 months, SGI-Challenge (3)
1998	Villin headpiece (36 residues, 12,000 atoms, cutoffs)	1000	4 months, 256-proc. Cray T3D/E
1999	bc_1 complex in phospholipid bilayer (91,061 atoms, cutoffs)	1	75 days, 64 450-MHz-proc. Cray T3E
2001	C-terminal β-hairpin of protein-G (177 atoms, implicit solvent)	38000	~ 8 days, 5000 proc. Folding@home megacluster
2002	channel protein in lipid membrane (106,189 atoms, PME)	5	30 hrs, 500 proc. LeMieux terascale system; 50 days, 32 proc. Linux (Athlon)

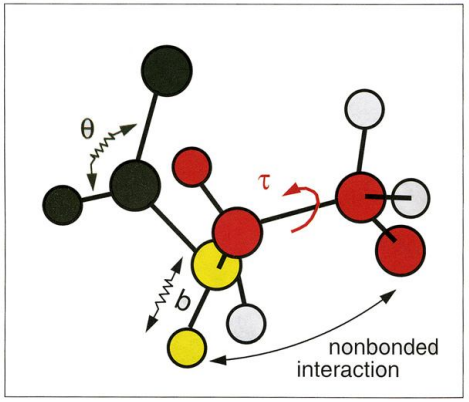


Figure 7.3. A molecule is considered a mechanical system in which particles are connected by springs, and where simple physical forces dictate its structure and dynamics.

Dynamique moléculaire

Loi de Newton :

$$ma = F$$

$$a = \frac{dv}{dt}$$

$$v = \frac{dr}{dt}$$

$$m\frac{dv}{dt} = F$$

$$m\frac{d^2r}{dt^2} = -\frac{dV}{dr}$$

$$dv = a \cdot dt$$

$$v = \int a \cdot dt \xrightarrow{a \approx const} at + v_0$$

$$dr = v \cdot dt$$

$$r = \int v \cdot dt = \int (at + v_0)dt \xrightarrow{a \approx const} \frac{1}{2}at^2 + v_0 t + r_0$$

Dynamique moléculaire

$$m_i \mathbf{a}_i(t) = \mathbf{F}_i(t)$$

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = -\frac{\partial V}{\partial \mathbf{r}_i}$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t \cdot \mathbf{v}(t) + \frac{\Delta t^2 \cdot \mathbf{a}(t)}{2}$$

$$\mathbf{a}(t + \Delta t) = \frac{\mathbf{F}(t + \Delta t)}{m}$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \Delta t \cdot \frac{\mathbf{a}(t) + \mathbf{a}(t + \Delta t)}{2}$$

$$f(v)dv = \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \cdot e^{-\frac{mv^2}{2kT}} \cdot 4\pi v^2 dv$$

$$\left\langle \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} \right\rangle = \frac{N_f kT}{2}$$

$$T = \frac{1}{N_f k} \cdot \sum_{i=1}^{N} m_i \mathbf{v}_i^2$$

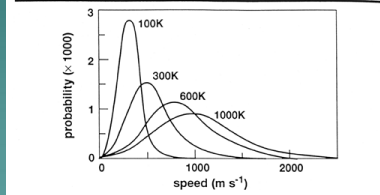


Figure 5–6. Maxwell–Boltzmann Distribution of Velocity of Water at Various Temperatures

Distribution of molecular velocities at equilibrium as predicted by the Maxwell–Boltzmann equation. The *Discover* program assigns random initial velocities to a system of atoms such that the overall distribution of velocities matches a Maxwell–Boltzmann distribution for the desired temperature.

Choice of parameters

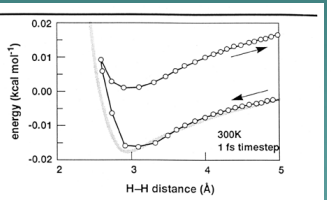


Figure 5–1. Numerical Integration of Energy from Molecular Dynamics H–H Trajectory, 1 fs Timestep

Integrated energy calculated numerically from a dynamics trajectory of two colliding hydrogen atoms (circles) is compared with the analytical energy curve (thick line).

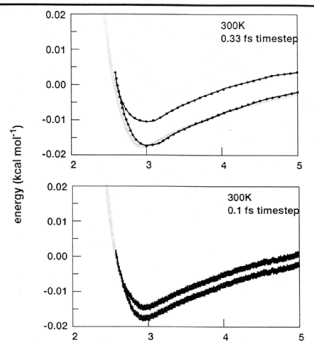


Figure 5–2. Numerical Integration of Energy from Molecular Dynamics H–H Trajectory, 0.33 and 0.1 fs Timesteps

Energy integration errors decrease with smaller time steps. Compared to 0.016 kcal error with 1 fs time steps, the 0.33 fs time step has a 0.006 error, and the 0.1 fs time step a 0.001 error. The cost for increased accuracy is the computational burden to compute more steps. For most simulations, a 1-fs time step is a good compromise between numerical accuracy and computational efficiency.

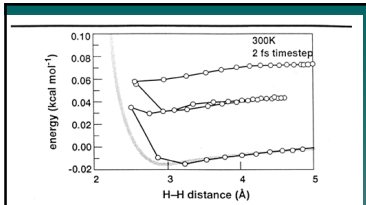


Figure 5–3. Numerical Integration of Energy from Molecular Dynamics H–H Trajectory, 2 fs Timestep

A 2-fs time step causes the rebounding force to be underestimated, robbing the colliding hydrogens of sufficient escape velocity and resulting in the two hydrogens being "bound" by their van der Waals forces.

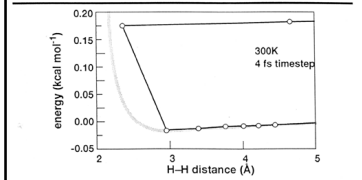


Figure 5–4. Numerical Integration of Energy from Molecular Dynamics H–H Trajectory, 4 fs Timestep

With a time step of 4 fs, the atoms travel too far in a single step, interpenetrating each other's van der Waals radii before the forces are recalculated. By this time, the forces are so large that the atoms are flung apart at a temperature equivalent to hundreds of thousands of degrees.

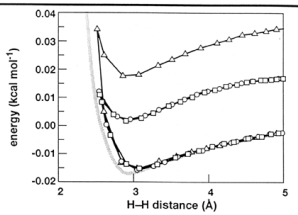


Figure 5–5. Integration Errors for H–H Collisions at Several Temperatures

Integration errors observed for an H–H collision for an initial velocity equal to the mean velocity appropriate for 300 K (squares), 1200 K (circles) and 30,000 K (triangles).

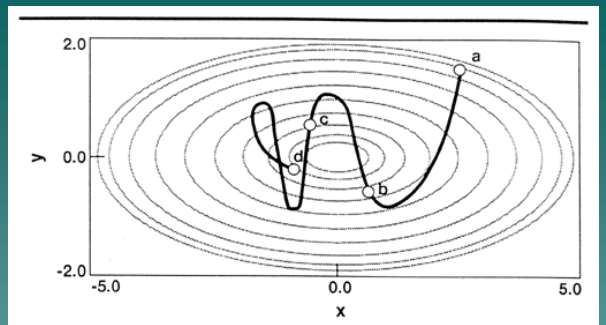
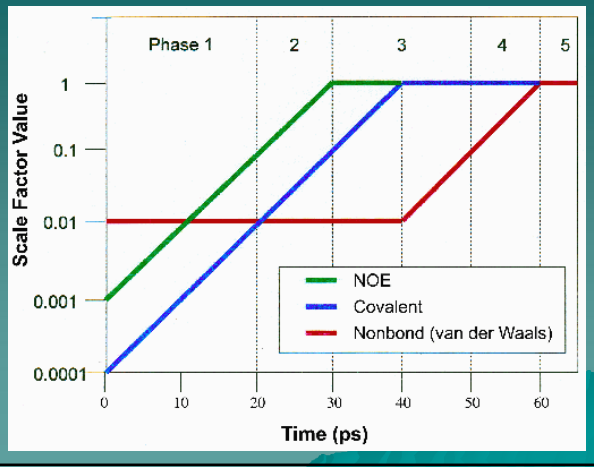


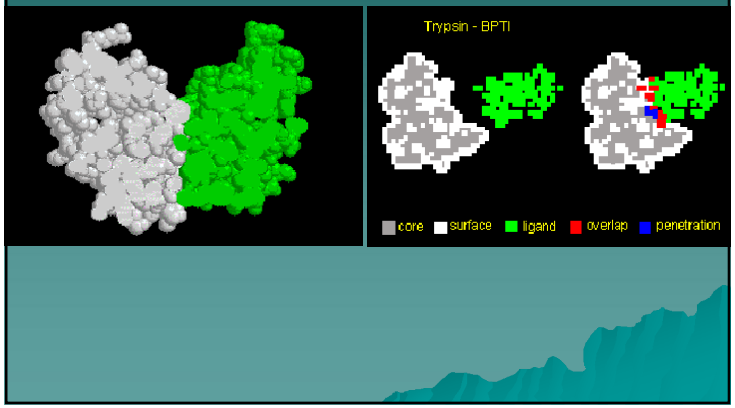
Figure 5–7. Dynamics Trajectory on a Simple Energy Surface

Dynamics trajectory of a particle beginning at rest at point **a** and continuing with constant total energy (kinetic + potential) until point **d**. On this elliptical energy surface, minimization from any of the points **a**, **b**, **c**, or **d** converges to the same point. Thus, minima can be used to characterize many nearby points sampled during dynamics. In molecules having many more degrees of freedom, periodic minimization during a dynamic calculation results in relatively fewer structures (compared to the number of discrete dynamic structures) for use in detailed structural and energetic studies.

Finding global minima with SA + MD



Docking techniques



Geometric docking

$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} \bar{a}_{l,m,n} \cdot \bar{b}_{l+\alpha,m+\beta,n+\gamma}$$

$$\bar{a}_{l,m,n} = \begin{cases} 1 & \text{surface of the molecule} \\ \rho & \text{inside the molecule} \\ 0 & \text{outside the molecule} \end{cases}$$

$$\bar{b}_{l,m,n} = \begin{cases} 1 & \text{surface of the molecule} \\ \delta & \text{inside the molecule} \\ 0 & \text{outside the molecule} \end{cases}$$

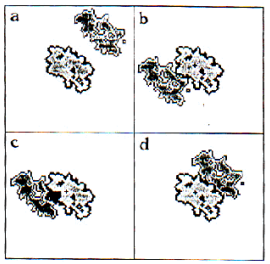


FIG. 2. Different relative positions of molecules **a** and **b**, illustrated by the cross sections $\bar{a}_{40,m,n}$ and $\bar{b}_{40,m,n}$ from Fig. 1. The relative orientation of the molecules is as in the known α–β dimer. (*a*) No contact. (*b*) Limited contact. (*c*) Penetration. The penetrated area is represented in black. (*d*) Good geometric match, as indicated by the extensive overlap of complementary surface layers.

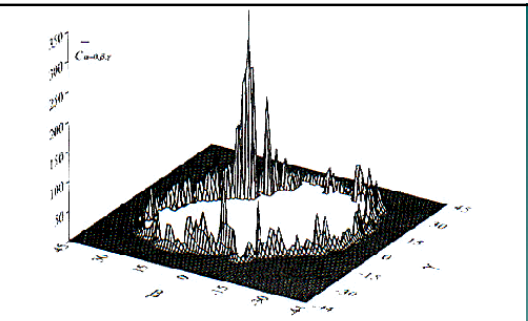


FIG. 3. Cross section (at $\alpha = 0$) through a 3D correlation function $C_{\alpha,\beta,\gamma}$. The correlation function shown was calculated for the α and β subunits of hemoglobin, oriented as in the dimer (from 2HHB, see Figs. 1 *c* and *d*). The correlation value at each shift vector $\{\theta,\beta,\gamma\}$ is represented by the height of the graph. The prominent peak at $\{\alpha = 0, \beta = 14, \gamma = 17\}$ corresponds to the correct match between the molecules (see Fig. 2*d*). Other intermolecular surface contacts (such as in Fig. 2*b*) give rise to the low positive correlation values around the center of the graph. The negative correlation values caused by penetration (see Fig. 2*c*) are omitted, leaving the empty area at the center.

$$X_{o,p,q} = \sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} x_{l,m,n} \cdot e^{-2\pi i \frac{ol+pm+qn}{N}}$$

$$C_{o,p,q} = A^{*}_{o,p,q} \cdot B_{o,p,q}$$

$$C = DFT\{\bar{c}\}$$
$$A^{*} = cnjg\left[DFT\{\bar{a}\}\right]$$
$$B = DFT\{\bar{b}\}$$

$$\bar{c}_{\alpha,\beta,\gamma} = \frac{1}{N^3}\sum_{o=1}^{N}\sum_{p=1}^{N}\sum_{q=1}^{N} C_{o,p,q} \cdot e^{2\pi i \frac{o\alpha+p\beta+q\gamma}{N}}$$

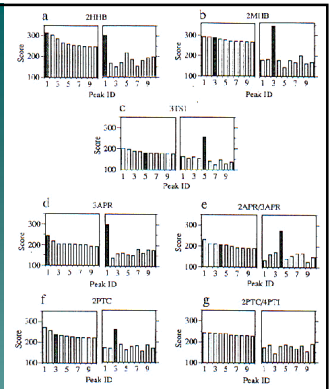
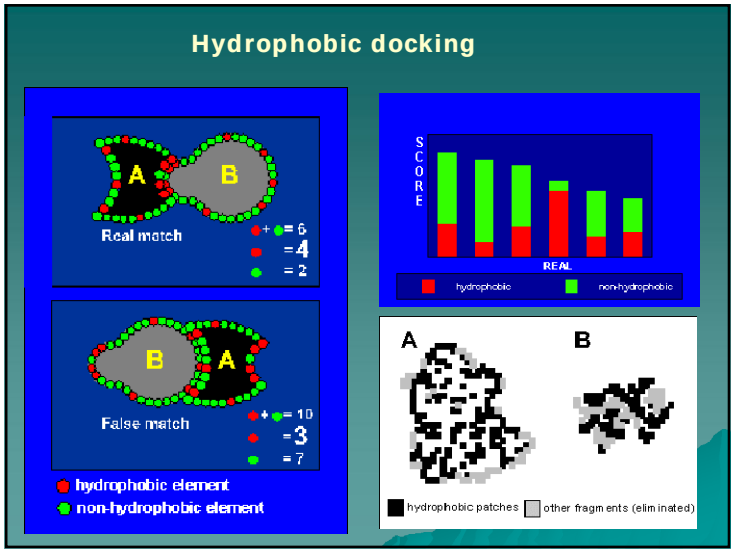


FIG. 4. Correlation results for different pairs of molecules. The pairs are identified by their respective codes (see text). In each panel, the histogram on the left shows the 10 highest correlation peaks obtained in the scan stage ($\eta = 1.0$–1.2 Å), sorted by their score. Each of these peaks was obtained at a different relative orientation of the molecules and corresponds to a potential geometric match. The shaded peak in each histogram corresponds to the known complex between the molecules considered. The histogram on the right side of each panel shows the scores obtained at the discrimination stage ($\eta = 0.7$–0.8 Å), for the 10 orientations singled out in the scan stage. Note that in the discrimination stage the spurious peaks (plain) are suppressed, whereas the correct peak (shaded) becomes prominent.



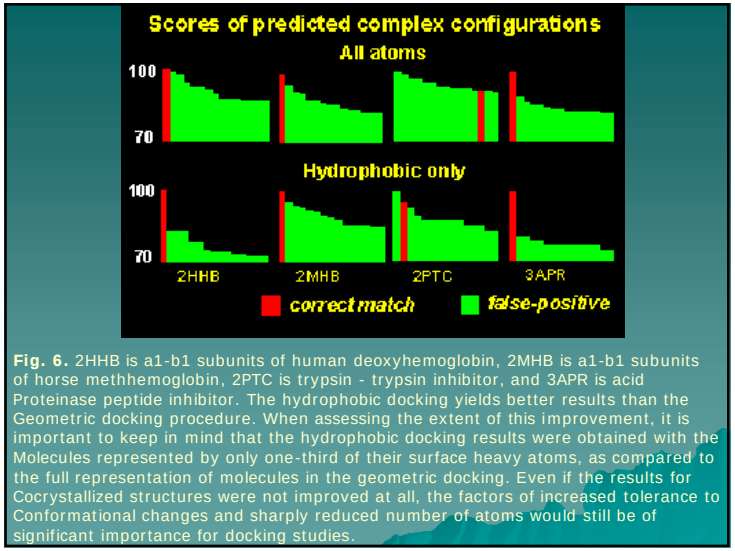
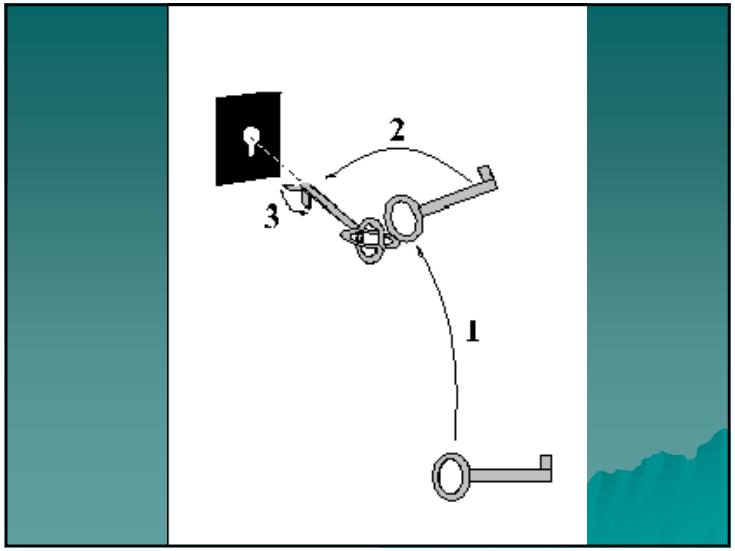
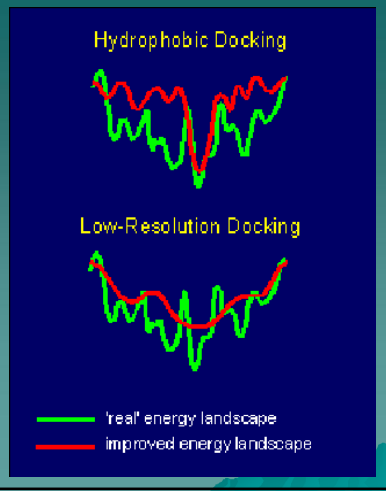
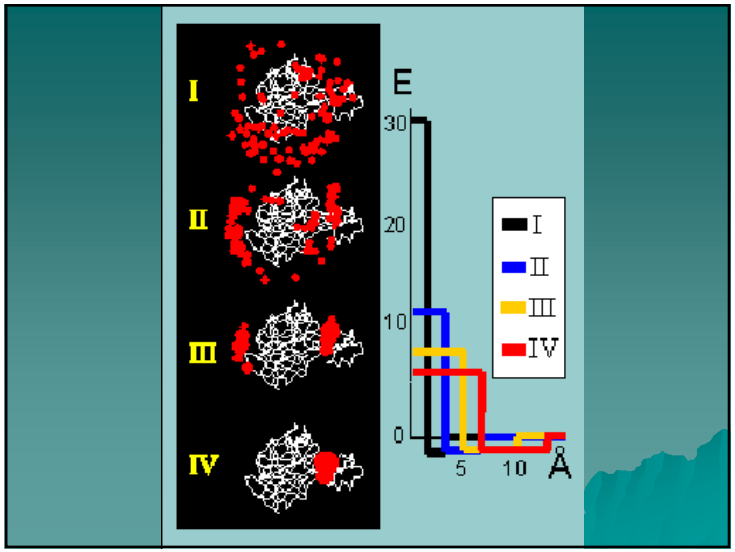


Fig. 6. 2HHB is a1-b1 subunits of human deoxyhemoglobin, 2MHB is a1-b1 subunits of horse methhemoglobin, 2PTC is trypsin - trypsin inhibitor, and 3APR is acid Proteinase peptide inhibitor. The hydrophobic docking yields better results than the Geometric docking procedure. When assessing the extent of this improvement, it is important to keep in mind that the hydrophobic docking results were obtained with the Molecules represented by only one-third of their surface heavy atoms, as compared to the full representation of molecules in the geometric docking. Even if the results for Cocrystallized structures were not improved at all, the factors of increased tolerance to Conformational changes and sharply reduced number of atoms would still be of significant importance for docking studies.



Multiple minima problem

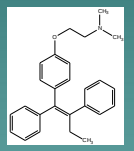


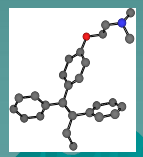


Aspects bioinformatiques du criblage virtuel

Phase initiale du criblage virtuel : Préparation

- Choix de la base de données
  (Zinc, Shapes…)
- Filtrage de la BdD
  élimination des molécules indésirables : toxiques, trop réactives…
- Choix des meilleurs candidats
  règles de Lipinski, élimination de *frequent hitters*…
- Filtrage spécifique
  ADME : absorption / distribution / métabolisme / excrétion
- Conversion de structures
  2D → 3D

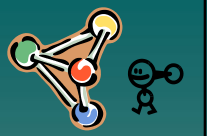


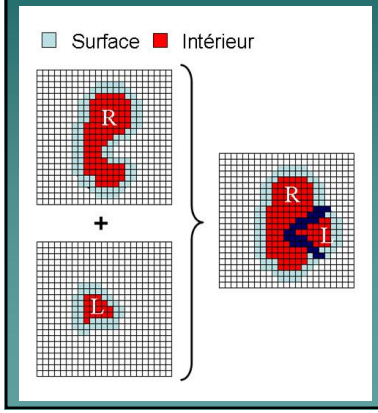


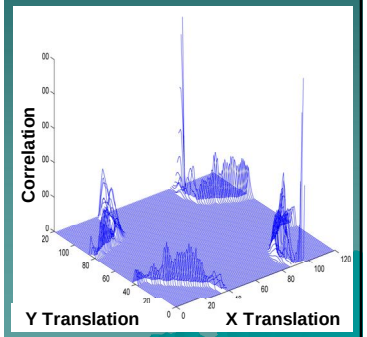
Tamoxifène

2e phase : Docking

1. Système : protéine rigide – ligand rigide
complémentarité de surfaces (GRAMM, FTDock, HEX…)

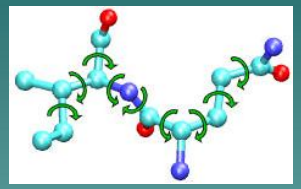




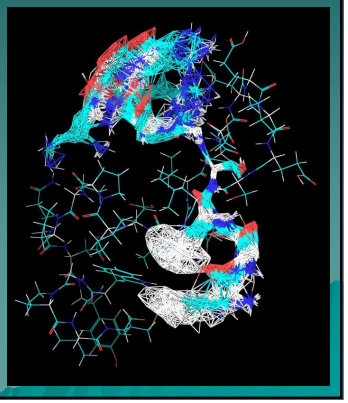


2e phase : Docking

2. Système : protéine rigide – ligand flexible
(FlexX, Gold, Dock, AutoDock…)



Liaisons rotatables d'un ligand flexible permettent l'exploration de l'espace conformationnelle de la molécule, et la création d'une famille des solutions.



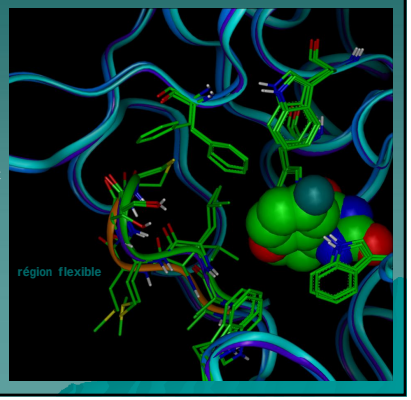
2e phase : Docking

3. Système : protéine flexible – ligand flexible
(FlexE, mais…)

Il faut prendre en compte les mouvements à petite échelle (changements conformationnels de chaînes latérales), à grande échelle (mouvements autour des charnières), ainsi que les transitions vers la forme plus ordonnée lors de la liaison avec le ligand.

La meilleure solution est de travailler avec un ensemble de structures d'un récepteur, utilisant les bibliothèques de rotamères pour la modélisation des chaînes latérales.

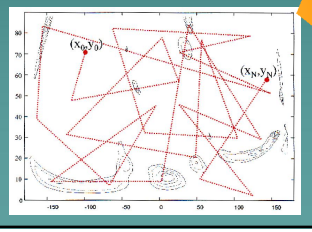
On peut simuler la dynamique moléculaire du récepteur et de faire le *docking* du ligand avec un nombre important de structures du récepteur, extraites de la trajectoire MD. Cette approche nécessite un temps très long de l'exécution.

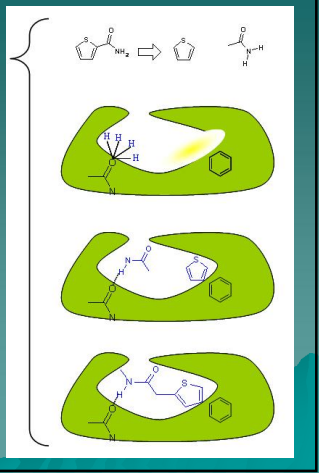


2e phase : Docking

Algorithmes :

- **construction incrémentale**
  (Dock, FlexX)
- **algorithme génétique**
  (Gold, AutoDock)
- **recuit simulé et Monte Carlo**
  (AutoDock, MCDock, ICM)
- **autres**





3e phase : Scoring

Prédiction d'énergie libre d'une liaison protéine-ligand :

- **interactions polaires** (liaisons hydrogènes, interactions ioniques)
- **interactions apolaires** (interactions lipophiliques et aromatiques)
- **changements d'entropie** (perte de la mobilité du ligand)
- **effets de la désolvation** (hydrophobicité)

Calcul de scores :

fonctions empiriques, basées sur les champs de force, ou potentiels de la force moyenne :

$$E = \sum_{i=1}^{N_{lig}} \sum_{j=1}^{N_{rec}} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right]$$

4e phase : Filtrage

- Élimination de faux positifs
  (propriétés topologiques)

- Recherche de meilleurs solutions :
  - multiple *scoring*
  - multiple *ranking*
  - *clustering* en familles

- Obtention de la liste finale (*hitlist*)
  (un jeu limité de meilleurs candidats)

Comparaison de méthodes de Docking

CAPRI: Critical Assessment of PRediction of Interactions

Résultats typiques :

Predictor	Software	Algorithm	T1	T2	T3	T4	T5	T6	T7
Abagyan	ICM	FF			+			+	+
Camacho	CHARMM	FF						+	+
Gardiner	GAPDOCK	GA		+					
Sternberg	FTDOCK	FFT						+	
Bates	Guided Docking	FF							+
Ten Eyck	DOT	FFT						+	
Vakser	GRAMM	FFT							
Olson	Harmony	?							
Weng	ZDOCK	FFT		+	+				+
Eisenstein	MolFit	FFT			+			+	+
Wolfson	BUDDA/PPD	GH							+
Iwadate	TSCF	FF							
Ritchie	Hex	SPF			+			+	
Palma	BIGGER	GF						+	
Gray	?	MC					+	+	

Comparaison de méthodes de Docking

Conclusions :

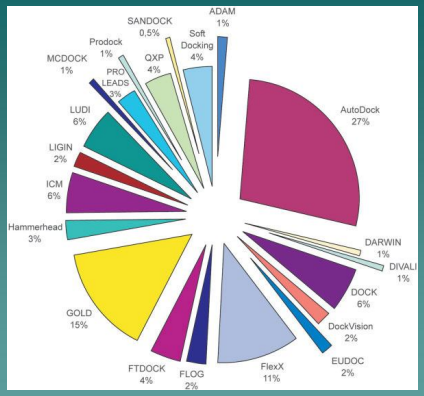
- Il n'existe pas une seule technique supérieure aux autres

  (les approches sont plus ou moins performantes, en fonction du système étudié et de la fonction de *scoring* choisie)

- Les meilleures programmes : Gold, FlexX, Dock, AutoDock

  (pas nécessairement dans cet ordre)

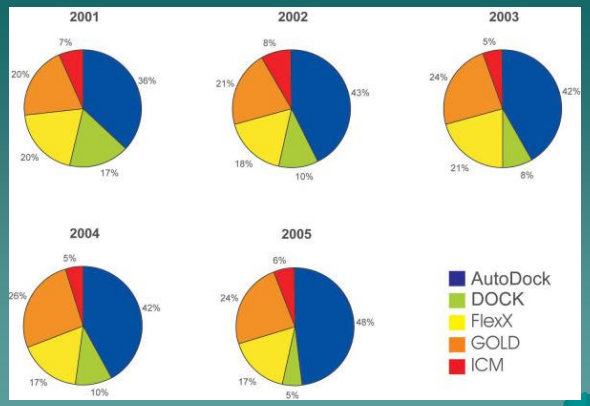
Quelques données statistiques sur les logiciels de Docking



Nombre de citations de logiciels de Docking (ISI Web of Science, 2005)

Sergio Filipe Sousa, Pedro Alexandrino Fernandes, and Maria Joao Ramos, PROTEINS: SFB **65**:15–26 (2006)

Quelques données statistiques sur les logiciels de Docking



Les logiciels de Docking le plus utilisés (ISI Web of Science, 2005)

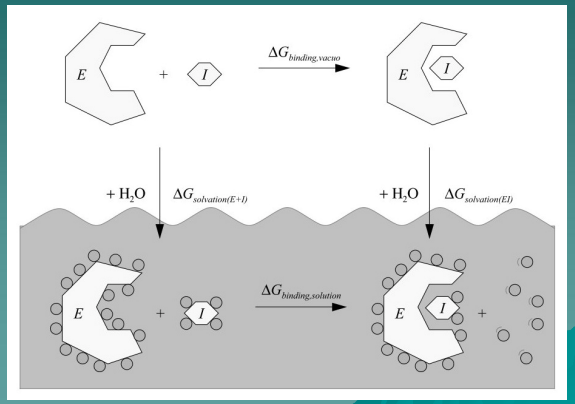
Sergio Filipe Sousa, Pedro Alexandrino Fernandes, and Maria Joao Ramos, PROTEINS: SFB **65**:15–26 (2006)

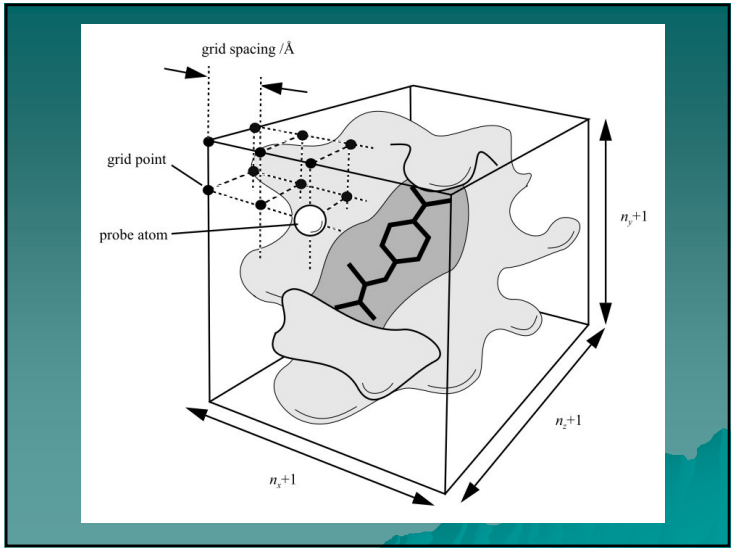
Ressources bioinformatiques nécessaires :

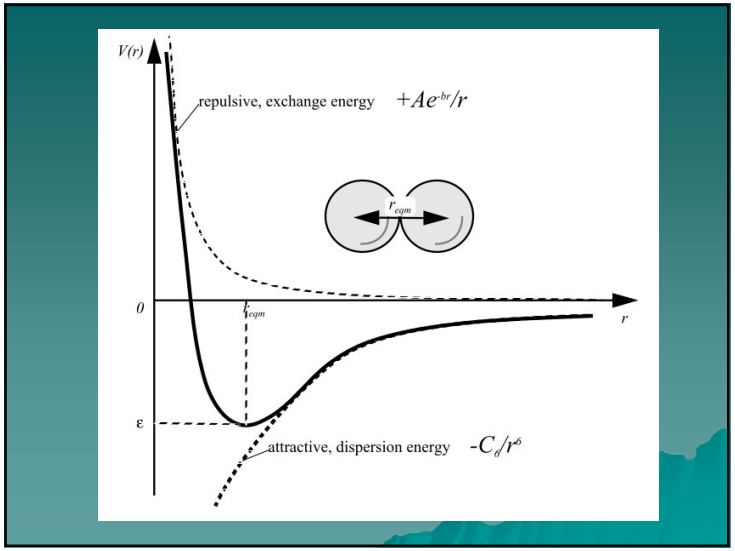
- Cluster de calcul sous Linux
- Serveur de fichier
- Logiciels libres ou versions académiques
- Bases de données en local (Zinc, Shapes…)
- Scripts de la maison (filtrage de bases de données, docking avec AutoDock)

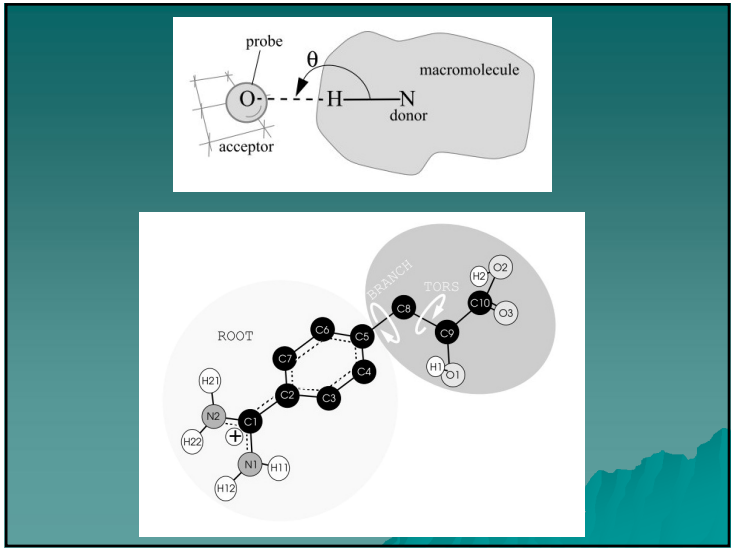
AutoDock

Automated Docking of Flexible Ligands to Receptors









Exemple : recherche des inhibiteurs de Pol beta

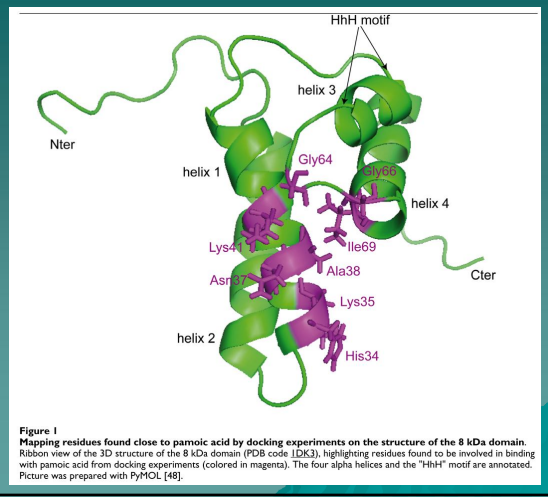
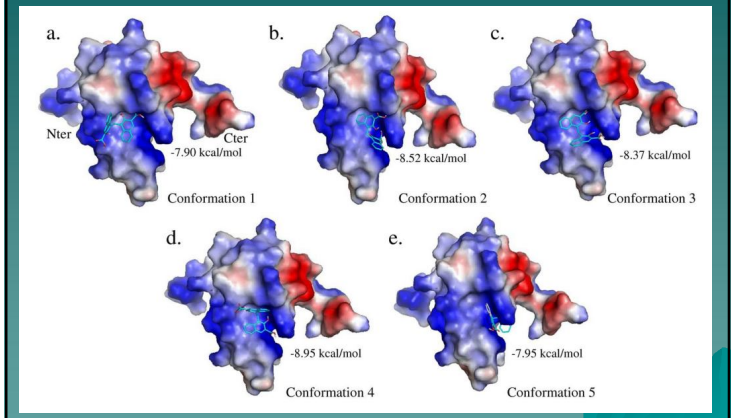


Figure 1
Mapping residues found close to pamoic acid by docking experiments on the structure of the 8 kDa domain. Ribbon view of the 3D structure of the 8 kDa domain (PDB code 1DK3), highlighting residues found to be involved in binding with pamoic acid from docking experiments (colored in magenta). The four alpha helices and the "HhH" motif are annotated. Picture was prepared with PyMOL [48].

Resultats de docking



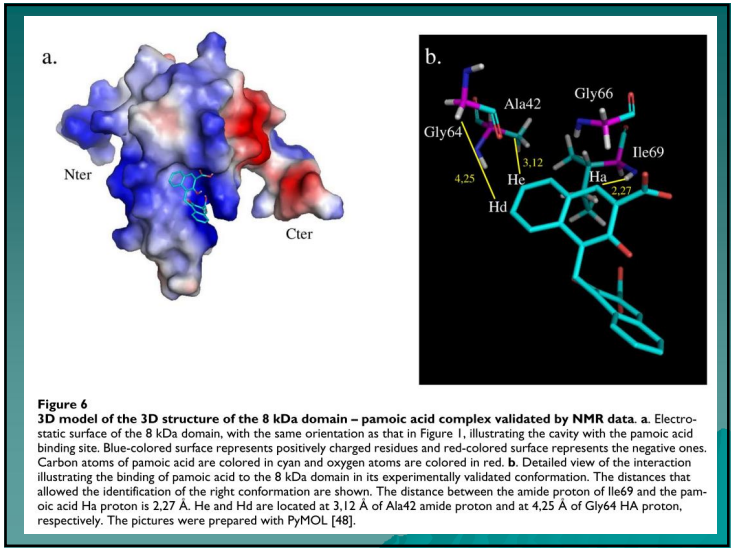
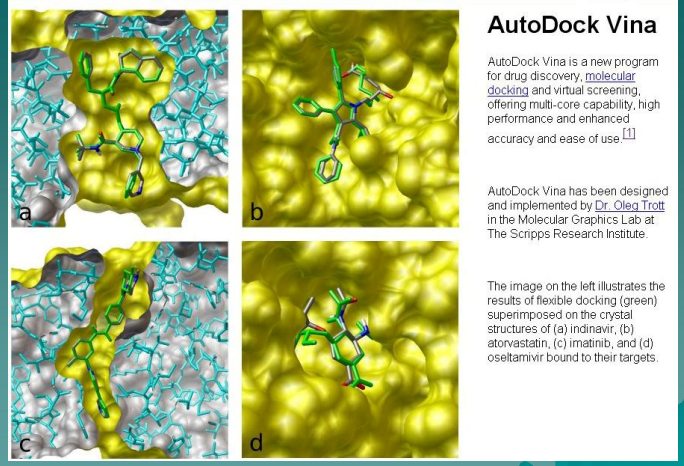


Figure 6
3D model of the 3D structure of the 8 kDa domain – pamoic acid complex validated by NMR data. **a**. Electrostatic surface of the 8 kDa domain, with the same orientation as that in Figure 1, illustrating the cavity with the pamoic acid binding site. Blue-colored surface represents positively charged residues and red-colored surface represents the negative ones. Carbon atoms of pamoic acid are colored in cyan and oxygen atoms are colored in red. **b**. Detailed view of the interaction illustrating the binding of pamoic acid to the 8 kDa domain in its experimentally validated conformation. The distances that allowed the identification of the right conformation are shown. The distance between the amide proton of Ile69 and the pamoic acid Ha proton is 2,27 Å. He and Hd are located at 3,12 Å of Ala42 amide proton and at 4,25 Å of Gly64 HA proton, respectively. The pictures were prepared with PyMOL [48].



AutoDock Vina

AutoDock Vina is a new program for drug discovery, molecular docking and virtual screening, offering multi-core capability, high performance and enhanced accuracy and ease of use.[1]

AutoDock Vina has been designed and implemented by Dr. Oleg Trott in the Molecular Graphics Lab at The Scripps Research Institute.

The image on the left illustrates the results of flexible docking (green) superimposed on the crystal structures of (a) indinavir, (b) atorvastatin, (c) imatinib, and (d) oseltamivir bound to their targets.