

NANOPHYSIQUE

INTRODUCTION PHYSIQUE AUX NANOSCIENCES

Ch. 9. Les Protéines

James Lutsko

Lecture 10 part 2, 2020-2021

Les Protéines

- Introduction: structure, properties
- Protéine interactions
- Pliage

De polymers a protéines

Polymer: molecule composee de multiple repetant units chains ou 3-D networks

Homopolymer: $A+A+A \rightarrow AAAA$

Copolymer: $A+B+A+B \rightarrow ABAB$

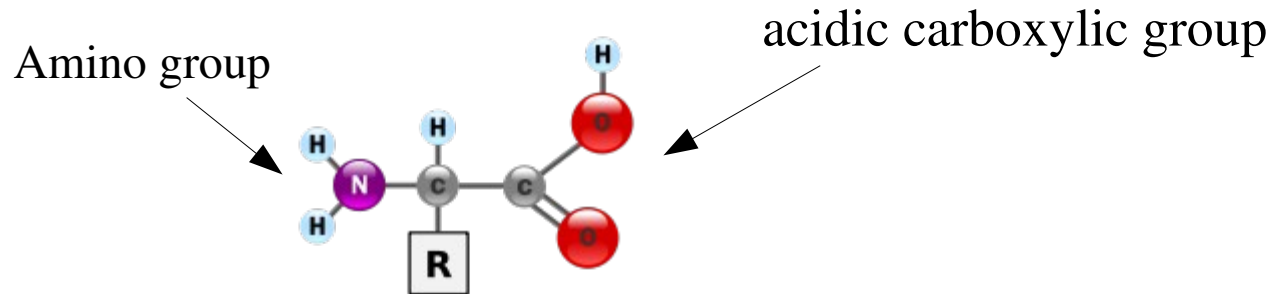
Covalent liens

Catalysis, DNA replication, Transport, ...

Protéine: A molecule composed of polymers of amino acids joined together by peptide bonds. It can be distinguished from fats and carbohydrates by containing nitrogen.

Acide Aminé

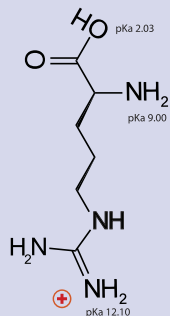
Acide Aminé: Une molécule composée de (i) le groupe aminé (NH₂); (ii) le groupe acidic carboxylic (COOH); (iii) une atom d'hydrogen (H); (iv) et un groupe organic(R) attaché au l'atom charbon. Donc, ayant la formule de NH₂CH(R)COOH.



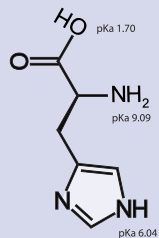
A. Amino Acids with Electrically Charged Side Chains

Positive

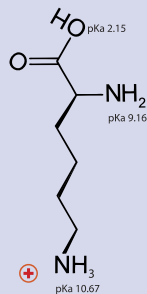
Arginine
(Arg) **R**



Histidine
(His) **H**

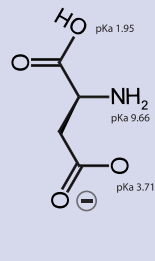


Lysine
(Lys) **K**

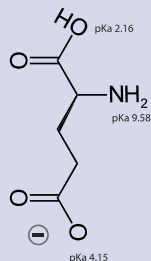


Negative

Aspartic Acid
(Asp) **D**

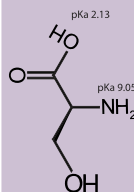


Glutamic Acid
(Glu) **E**

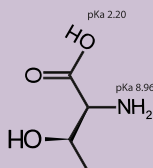


B. Amino Acids with Polar Uncharged Side Chains

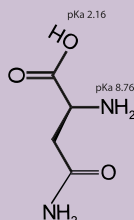
Serine
(Ser) **S**



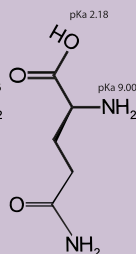
Threonine
(Thr) **T**



Asparagine
(Asn) **N**

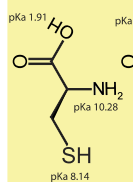


Glutamine
(Gln) **Q**

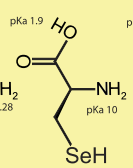


C. Special Cases

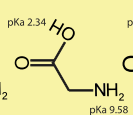
Cysteine
(Cys) **C**



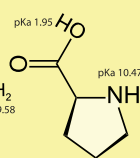
Selenocysteine
(Sec) **U**



Glycine
(Gly) **G**

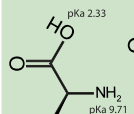


Proline
(Pro) **P**

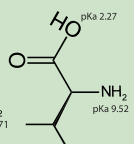


D. Amino Acids with Hydrophobic Side Chain

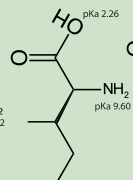
Alanine
(Ala) **A**



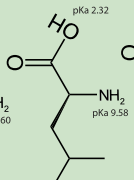
Valine
(Val) **V**



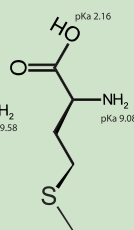
Isoleucine
(Ile) **I**



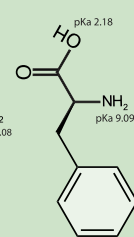
Leucine
(Leu) **L**



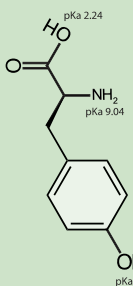
Methionine
(Met) **M**



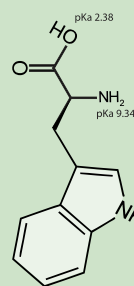
Phenylalanine
(Phe) **F**



Tyrosine
(Tyr) **Y**



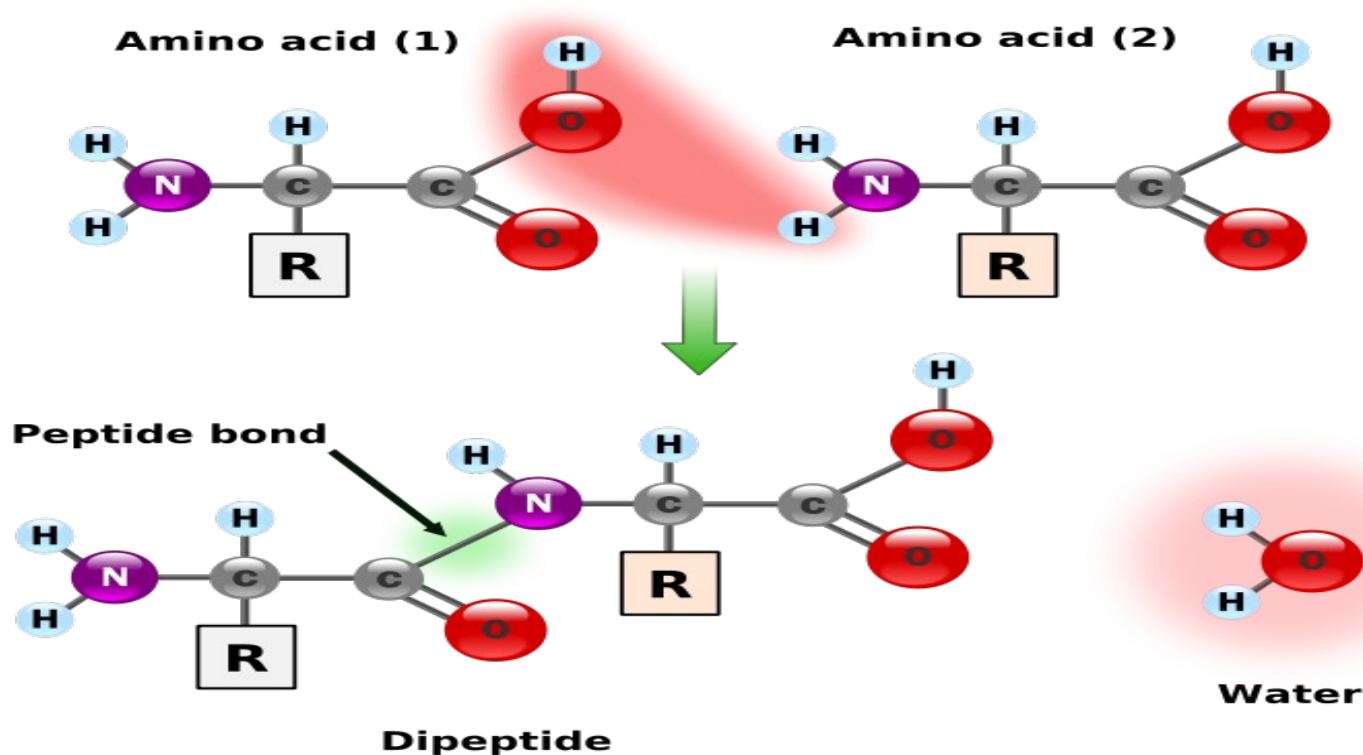
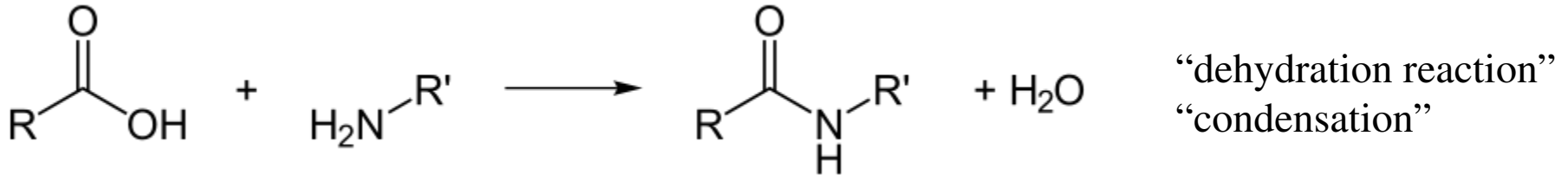
Tryptophan
(Trp) **W**



Peptides

Peptide: court polymer chain (linear, unbranched, chaque amino acid bonded to deux voisin).

Peptide Bond: lien (covalent) entre un amino group et un carboxylic group.



Protéine: Structure

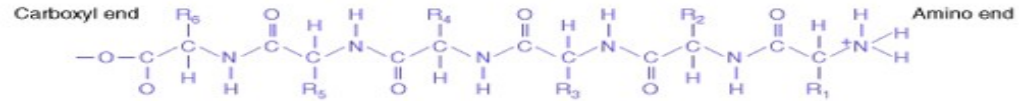
Protéine: une longue chaîne polymérique (linéaire, non ramifiée, avec chaque acide aminé collé à deux voisins).

Residue: acide aminé dans une protéine

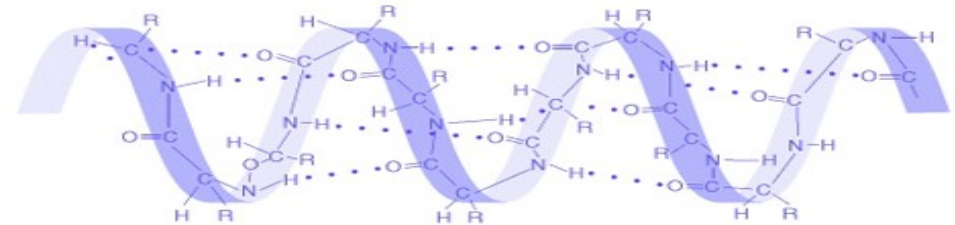
Backbone: la série liée de N, C, O.

Peptide <~ 20-30 residues <~ protein

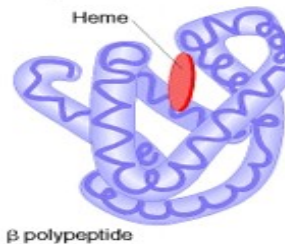
(a) Primary structure



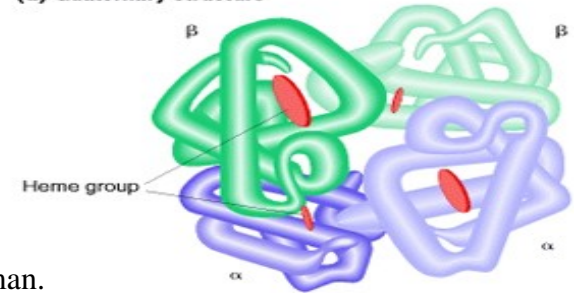
(b) Secondary structure



(c) Tertiary structure



(d) Quaternary structure



Modern Genetic Analysis, Griffiths AJF, Gelbart WM, Miller JH, et al., 1999 W. H. Freeman.

Primary structure: acide aminé sequence

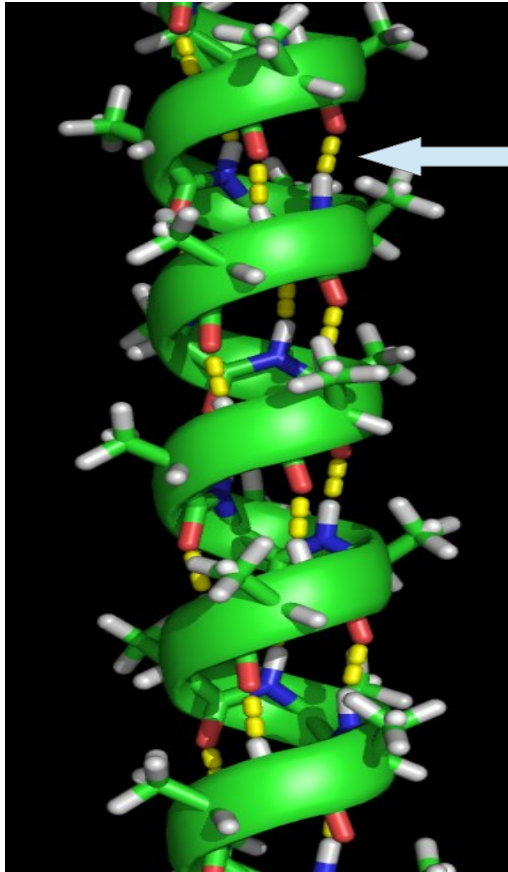
Secondary structure: 3-D structures composé de segments locaux définis par les liens d'hydrogène entre “backbone” amino and carboxyle groupes (α -helix, β -sheet, turns (hairpin turn),...)

Tertiary structure: la forme d'une seule protéine molécule; la relation spatiale des structures secondaires. Lié aux coordonnées atomiques.

Quaternary structure: arrangement de protéines multiples dans un seul complexe.

Protéine Secondary Structure: α -helix

α -helix



H-bond

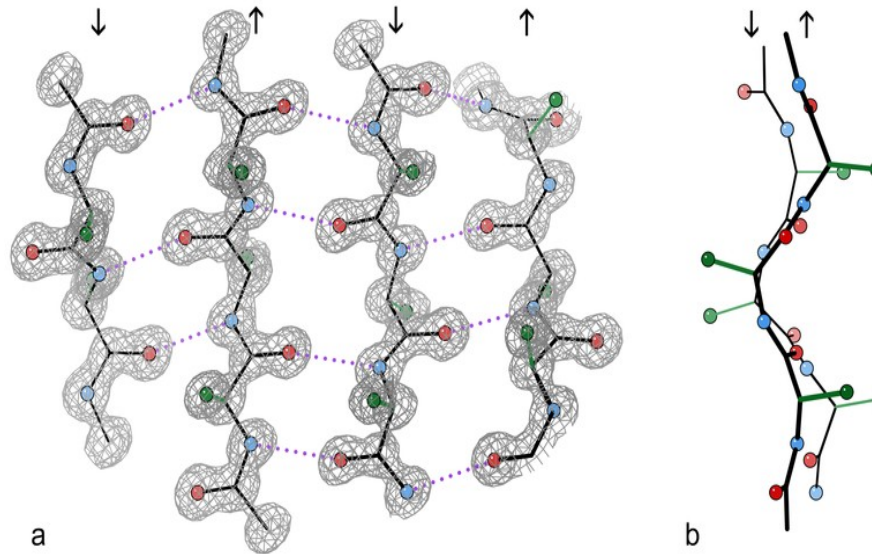
Liens entre residue n et $n+4$ (“ $n+4 \rightarrow n$ bonding”)

4-40 residues possible

Typique: 10 residues qui correspondent au 3 turns.

Protéine Secondary Structure: β -sheet

β -strand: chaîne de peptide entièrement prolongée

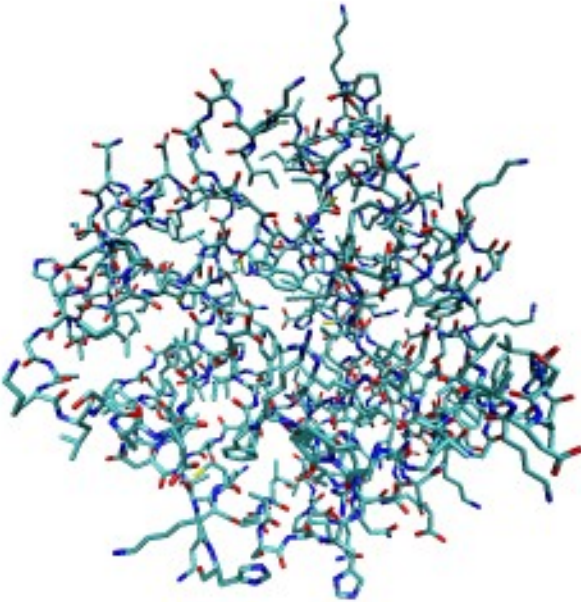


β -sheet:
réticulés β -
strands

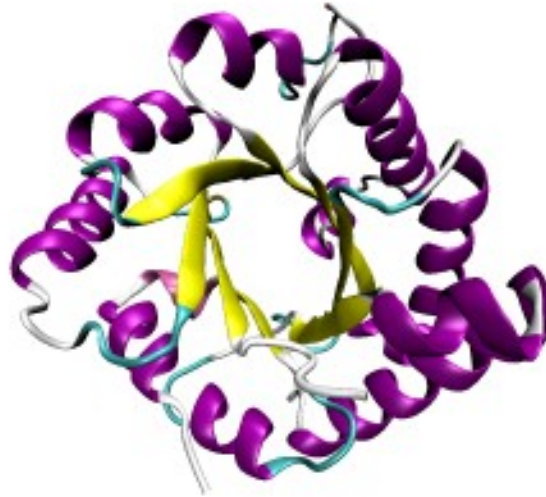
Exemple d'un fragment de feuillet β à quatre chaînes antiparallèles extrait de la structure cristalline de l'enzyme catalase (résolution 0,88 Å). a) Vue de face, montrant les liaisons hydrogènes (en pointillés) entre les groupes NH et CO des acides aminés adjacents. Les flèches indiquent l'orientation des chaînes, et les contours de densité d'électron entourent les atomes autres que l'hydrogène. Les atomes d'oxygène sont donnés en rouge, ceux d'azote en bleu. Les atomes d'hydrogène sont omis pour plus de simplicité. Dans le même but, seul le premier carbone des radicaux est montré (en vert). b) vue par côté des deux chaînes centrales montrant la torsion à droite des chaînes l'une par rapport à l'autre, ainsi que les plis de chacune d'elle qui orientent les carbones portant les radicaux des acides aminés alternativement de part et d'autre de celles-ci.

Protéine Structure: examples

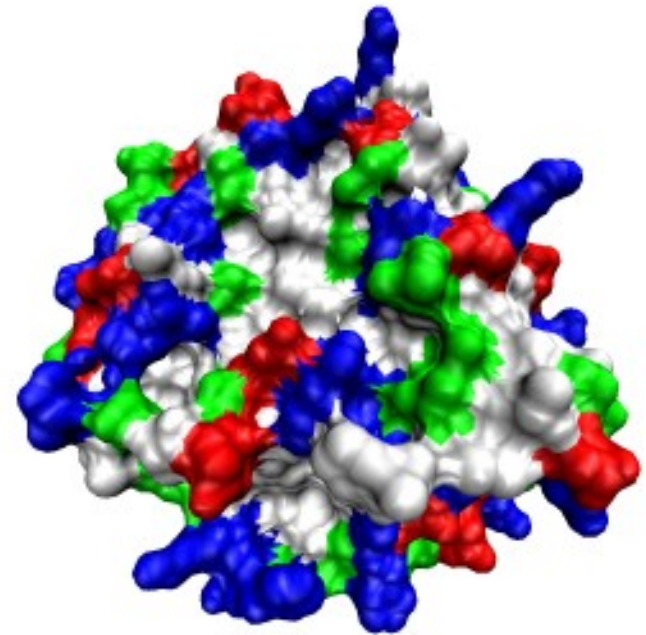
triose phosphate isomerase



All atom

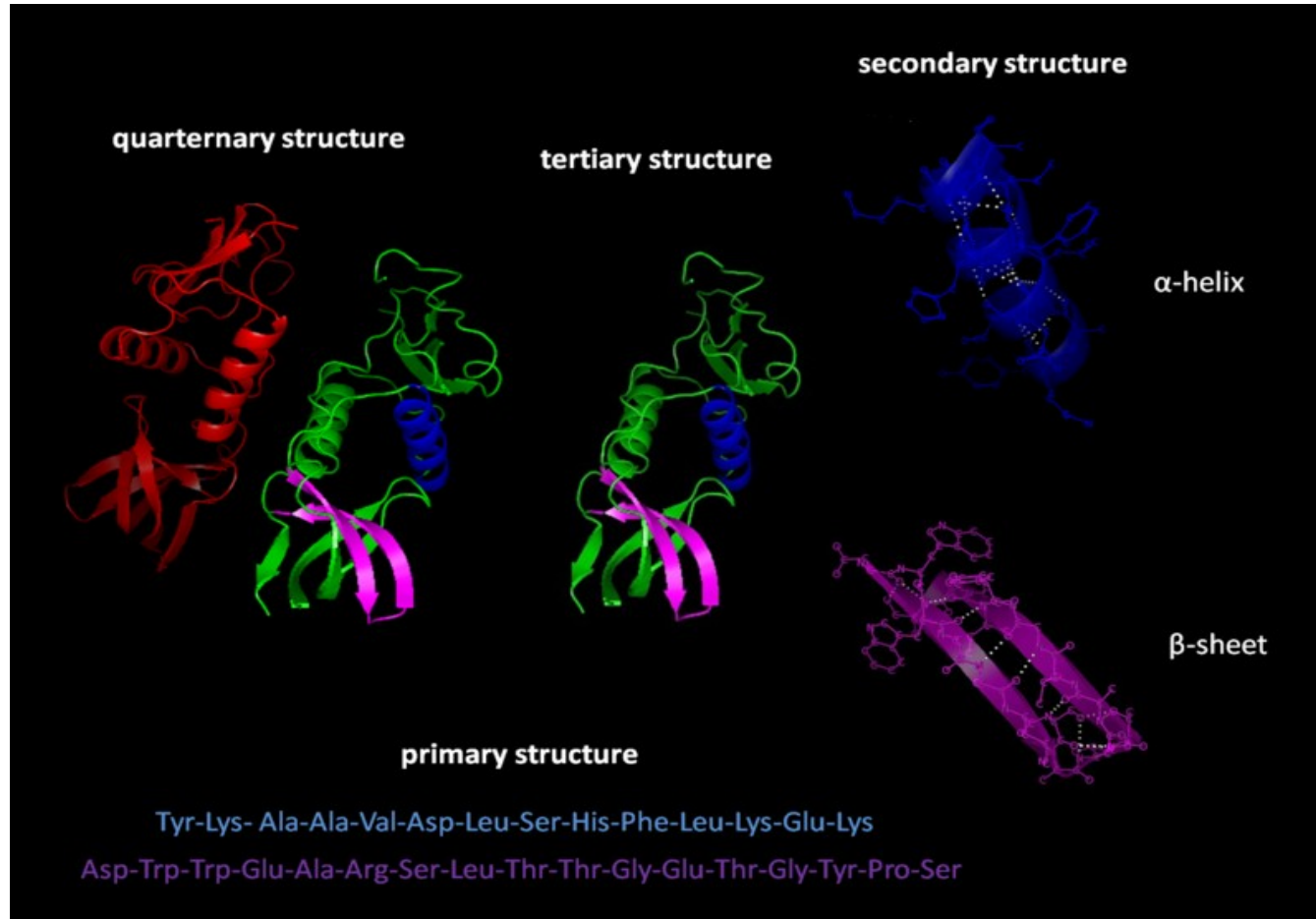


Secondary et
tertiary (α -
helices and β -
sheets)



Solvent-accessible surface
representation colored by residue
type (acidic residues red, basic
residues blue, polar residues green,
nonpolar residues white)

Protéine Structure: examples



Protéine: domains

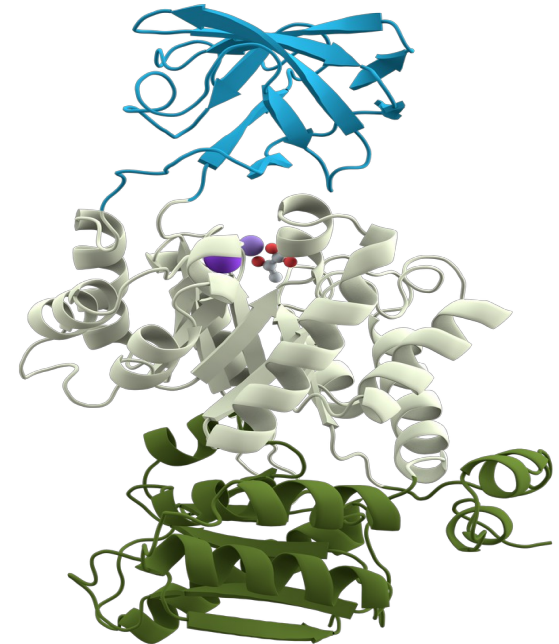
Structural Domain: is an element of the protein's overall structure that is self-stabilizing and often folds independently of the rest of the protein chain.

36-392 residues

Many domains are not unique to the protein products of one gene or one gene family but instead appear in a variety of proteins.

Domains often are named and singled out because they figure prominently in the biological function of the protein they belong to; for example, the "calcium-binding domain of calmodulin".

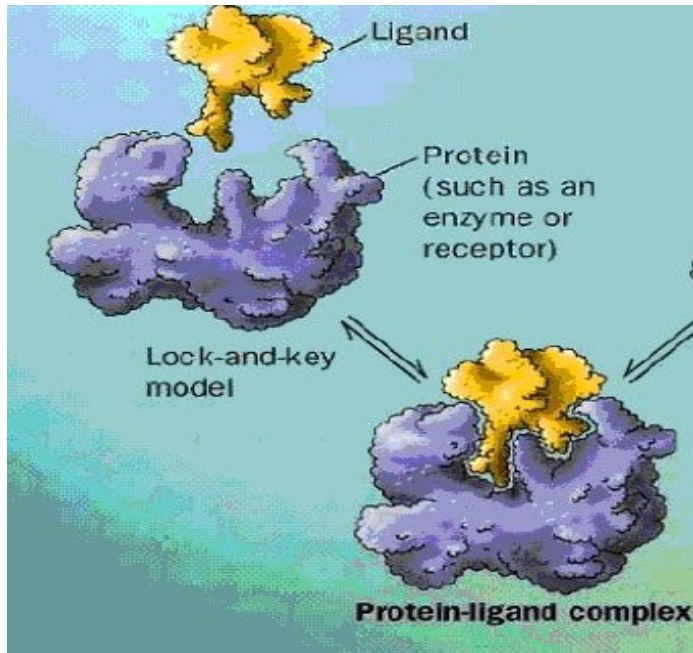
Because they are independently stable, domains can be "swapped" by genetic engineering between one protein and another to make chimeras.



Pyruvate kinase, a protein with three domains.

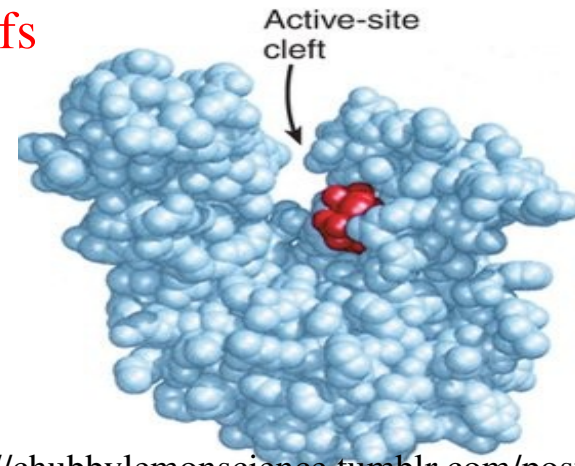
Protéine: la relation entre structure et fonction

“Lock and key”



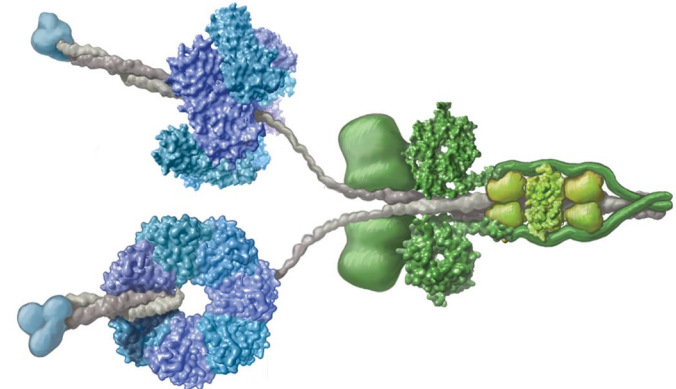
<http://student.biology.arizona.edu/>

Accessibilité des sites actifs



<http://chubbylemonscience.tumblr.com/post/3162713115/enzymes-shape-structure-and-what-it-does>

Molecular moteurs: propriétés mécaniques



<http://physics.berkeley.edu/research/yildiz/research.html>

Protéine: détermination de la structure

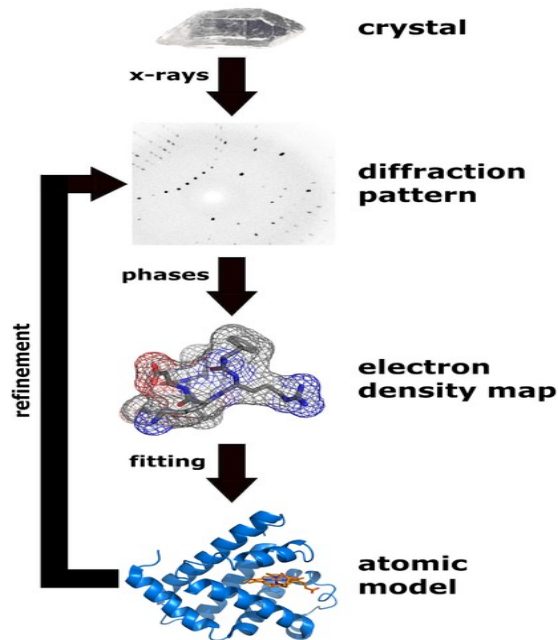
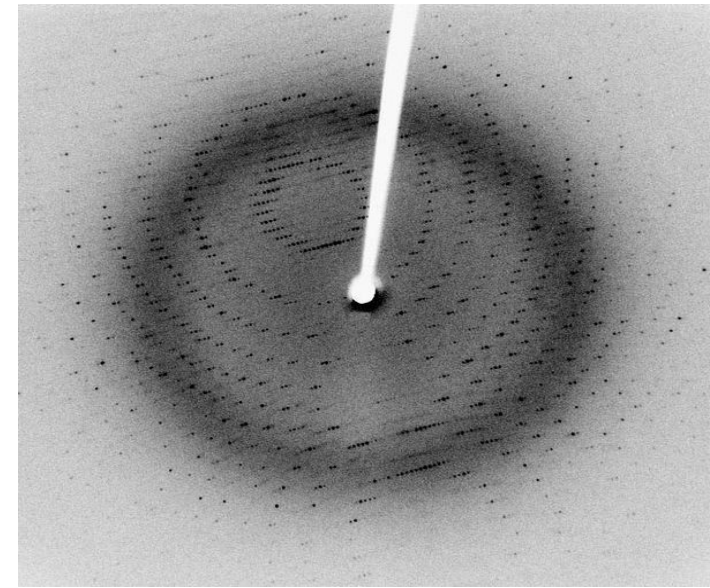


Image: Thomas Splettstoesser
http://en.wikipedia.org/wiki/X-ray_crystallography



X-ray diffraction pattern of crystallized 3Clpro, a SARS protease. (2.1 Angstrom resolution). Jeff Dahl, 2006

Diffusion statique

$$I(\mathbf{q}) \sim S(\mathbf{q}) = 1 + \rho \int [g(r) - 1] e^{i2\pi \mathbf{q} \cdot \mathbf{r}} d\mathbf{r}$$

Besoin des cristaux de haute qualité.

Diffusion dynamique

$$\frac{\langle I(\mathbf{q}, t) I(\mathbf{q}, t + t') \rangle}{\langle I \rangle^2} = I_0^2 [1 + \gamma e^{-2D_c q^2 t}]$$

$$D_c \sim \frac{k_B T}{6 \pi \eta R_h}$$

Hydrodynamic radius

Protéine: la but en bref

- ADN \rightarrow ARN \rightarrow les Protéines
- Les Protéines \rightarrow Fonction (catalyseurs, moteurs, etc.)
- But: comprendre les conséquences (fonction) en termes de la cause (ADN, ...)
- Car fonction et structure sont liés, il faut comprendre la structure des protéines.
- Expérience: a besoin des cristaux de haute qualité \rightarrow on doit comprendre protéine cristallisation et croissance.
- Théorie: pour prédire leur structure, on doit comprendre le pliage des protéines.

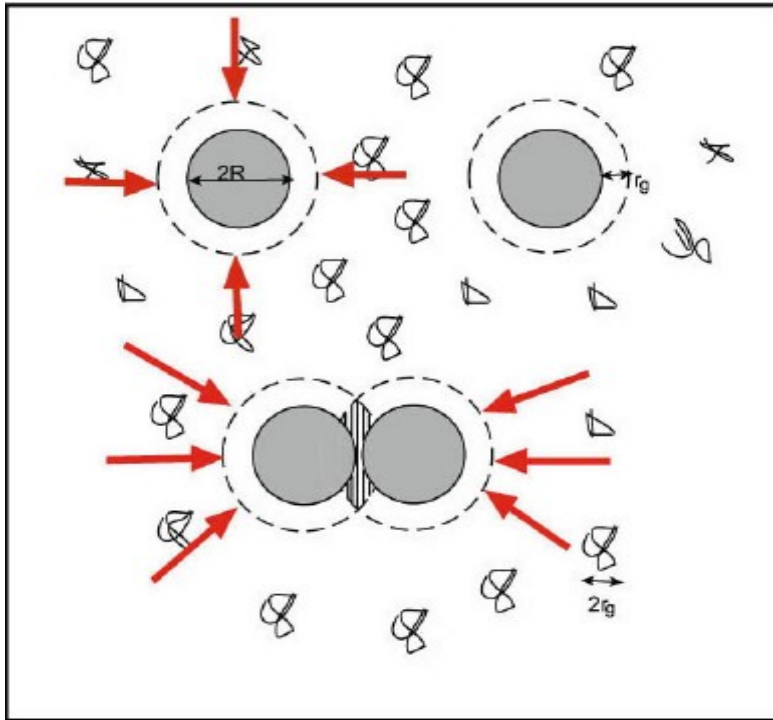
Les Protéines

- Introduction
- Protéine interactions
- Pliage

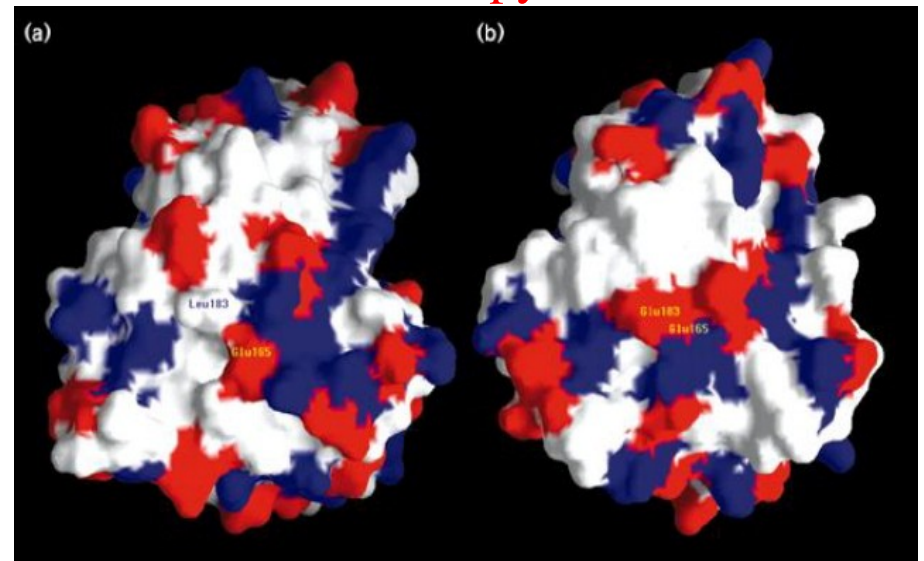
Protéines: beaucoup des interactions

Excluded volume, electrostatic, VdW forces, hydrophobic/hydrophilic forces ...

Depletion forces



Anisotropy

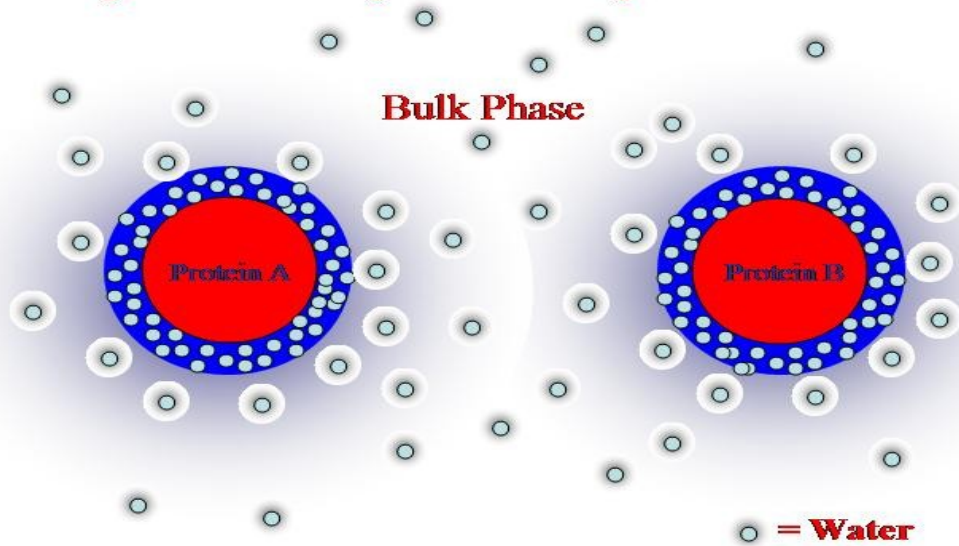


triosephosphate isomerase: des régions chargées
rouge (-) et bleu(+)

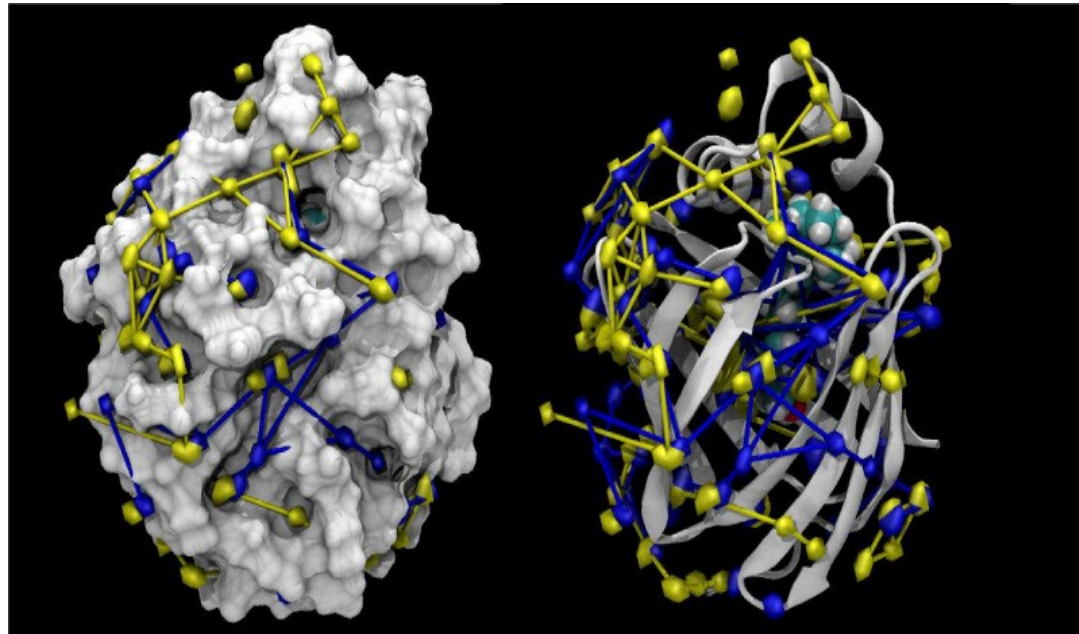
S. Velanker, et al, Structure **5** 751 (1997)

Protéines: beaucoup des interactions

Hydration Layer and Repulsive Forces



Hydration sites (from MD)



Protéine: des modeles d'interaction simple

DLVO interaction $V(r) = V_{HS}(r) + V_{el}(r) + V_{vdW}(r)$

Petite patricule: $V_{vdW}(r) \sim -C/r^6$

Grande “particule” spherique, composite: $V_{vdW}(r) = -\frac{A_H}{12} \left[\frac{\sigma^2}{r^2} + \frac{\sigma^2}{r^2 - \sigma^2} - 2 \log \frac{r^2 - \sigma^2}{r^2} \right]$

Hamaker constante: $A_H = \frac{3}{4} k_B T \left(\frac{\epsilon_1 - \epsilon_2}{\epsilon_1 + \epsilon_2} \right)^2 + \frac{(\epsilon_1^2 - \epsilon_2^2)^2}{(\epsilon_1^2 + \epsilon_2^2)^{3/2}}$ $\epsilon_1 = \text{particule}$
 $\epsilon_2 = \text{medium}$

Keesom-Debye interaction

London dispersion interaction

Protéine: des modeles d'interaction simple

DLVO interaction

$$V(r) = V_{HS}(r) + V_{el}(r) + V_{vdW}(r)$$

$$V_{vdW}(r) = -\frac{A_H}{12} \left[\frac{\sigma^2}{r^2} + \frac{\sigma^2}{r^2 - \sigma^2} - 2 \log \frac{r^2 - \sigma^2}{r^2} \right]$$

Solution: des ions avec densité n_j et charge z_j

Protéine: surface charge Ze , ($Ze \sim 12$ pour lysozyme)

$$\mu_j(\mathbf{r}) = \mu_j^0 + e z_j V(\mathbf{r}) \quad d\mu_j^0(\mathbf{r}) = -k_B T d \ln n_j(\mathbf{r}) \quad d\mu_j(\mathbf{r}) = 0 \Rightarrow n_j(\mathbf{r}) = n_j^{(0)} e^{-z_j e V(\mathbf{r}) / k_B T}$$

Dehors les protéines: $\nabla^2 V = 0$

Dehors les protéines: $\epsilon_0 \epsilon_{sol} \nabla^2 V = -\rho_f$, free charge density $\rho_f(\mathbf{r}) = \sum_{j=1}^N e z_j n_j(\mathbf{r})$

Sur la interface: $\epsilon_p \nabla V_{int} \cdot \mathbf{n} - \epsilon_{sol} \nabla V_{ext} \cdot \mathbf{n} = \sigma / \epsilon_0$

Resultat:
$$V(r) = \frac{Ze}{\epsilon_{sol}(1 + \kappa a)} \frac{e^{\kappa(a-r)}}{r}$$

Gunton, Shiryayev and Pagan, "Protein condensation", Cambridge, 2007.

Protéine: des modeles d'interaction simple

DLVO interaction

$$V(r) = V_{HS}(r) + V_{el}(r) + V_{vdW}(r)$$

$$V_{el}(r) = \frac{Ze}{\epsilon_{sol}(1 + \kappa a)} \frac{e^{\kappa(a-r)}}{r}$$

↖
rayon

$$V_{vdW}(r) = -\frac{A_H}{12} \left[\frac{\sigma^2}{r^2} + \frac{\sigma^2}{r^2 - \sigma^2} - 2 \log \frac{r^2 - \sigma^2}{r^2} \right]$$

Depletion forces: pertinente quand “nonabsorbing” polymère est ajouté aux solutions (e.g. polyethylene glycol “PEG”).

Asakura-Oosawa:

$$V(r) = \begin{cases} \infty, & r \leq 2a \\ -\frac{4\pi}{3} d^3 n_{PEG} k_B T \left[1 - \frac{3r}{4d} + \frac{r^3}{16d^3} \right], & 2a \leq r \leq 2d \\ 0, & 2d < r \end{cases}$$

$d = (a + R_{PEG})$

Polymer reference interaction site model “PRISM”:

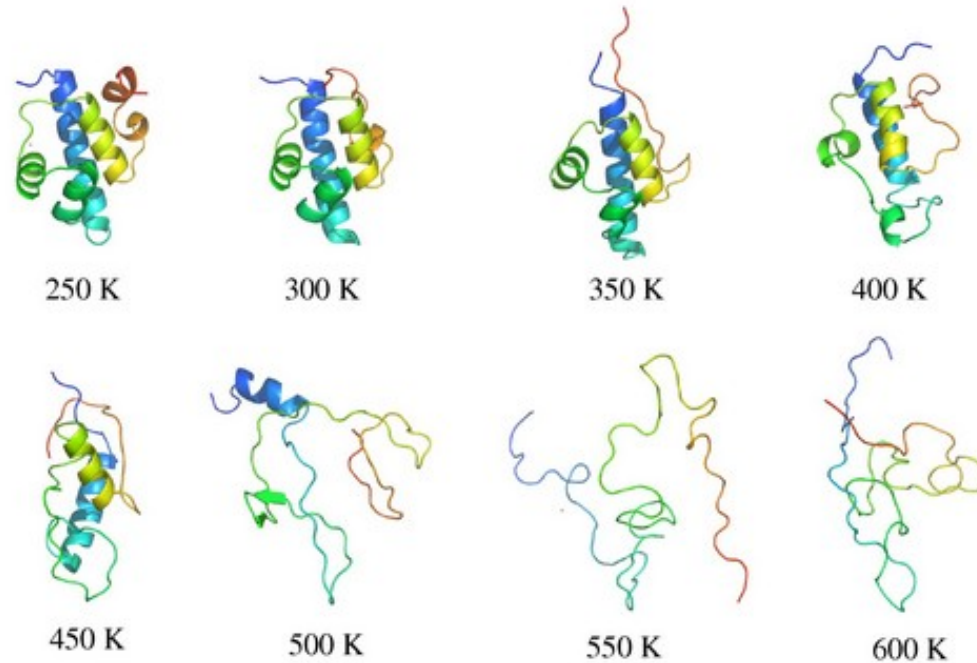
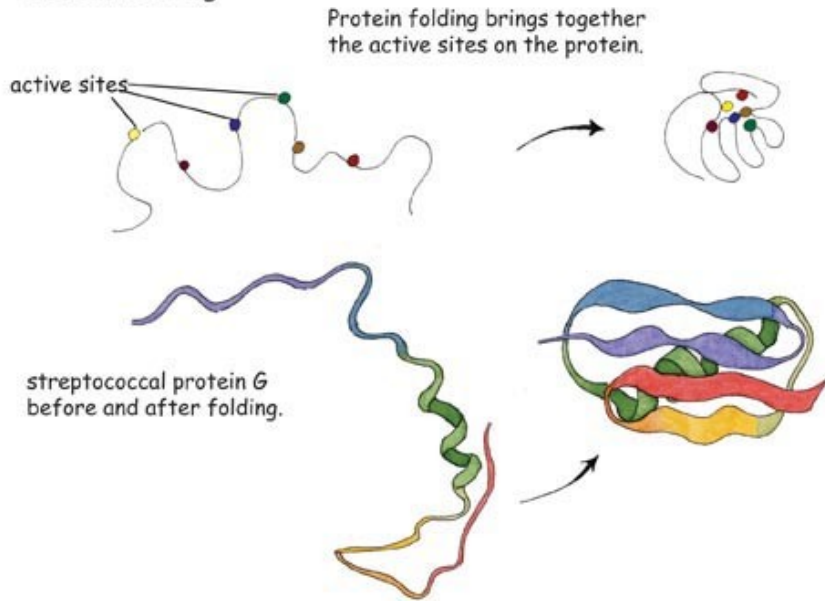
$$V(r) = \begin{cases} \infty, & r \leq 2a \\ -k_B T \ln \left[1 + \frac{\pi z a^2}{3 \sigma r} e^{-(r-2a)/\psi} \right], & 2a \leq r \end{cases}$$

Les Protéines

- Introduction
- Protéine interactions
- **Pliage**

Protéines Pliage

Protein Folding



©Sheri Amsel
www.exploringnature.org

http://www.almaden.ibm.com/st/past_projects/PFolding/

Protéines Pliage

REVIEW

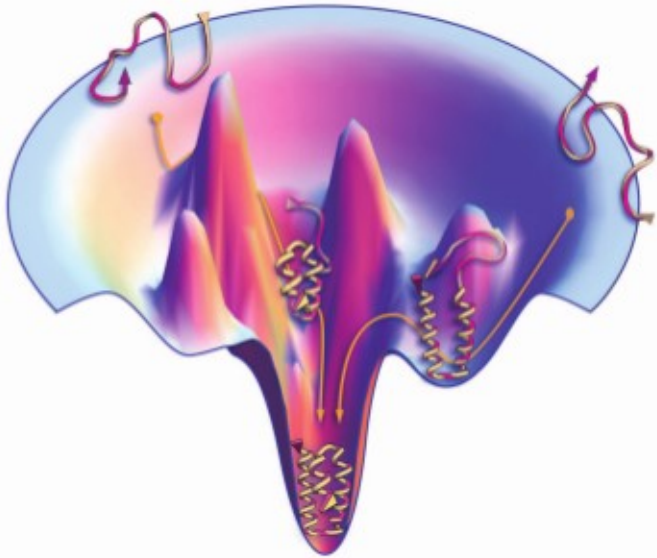


Fig. 3. Proteins have a funnel-shaped energy landscape with many high-energy, unfolded structures and only a few low-energy, folded structures. Folding occurs via alternative microscopic trajectories.

“Perhaps the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates, and it is more complicated than has been predicated by any theory of protein structure. Though the detailed principles of construction do not yet emerge, we may hope that they will do so at a later stage of the analysis.”

John Kendrew prix Nobel (avec Max Perutz) 1962
pour la détermination de la structure de la myoglobine.

Free-energy landscape for protein folding:
Shallow local minima, one deep minimum

The Protein-Folding Problem, 50 Years On
Ken A. Dill and Justin L. MacCallum
Science 338, 1042 (2012);

Protéines Pliage

Problems (d'après Dill):

1. Comment dériver la 3D structure du 1D sequence acid amino?
2. Il y a beaucoup de possibilités mais le processus est très vite: comment?
3. Peut on prédire, par ordinateur, la structure d'un protein?

REVIEW

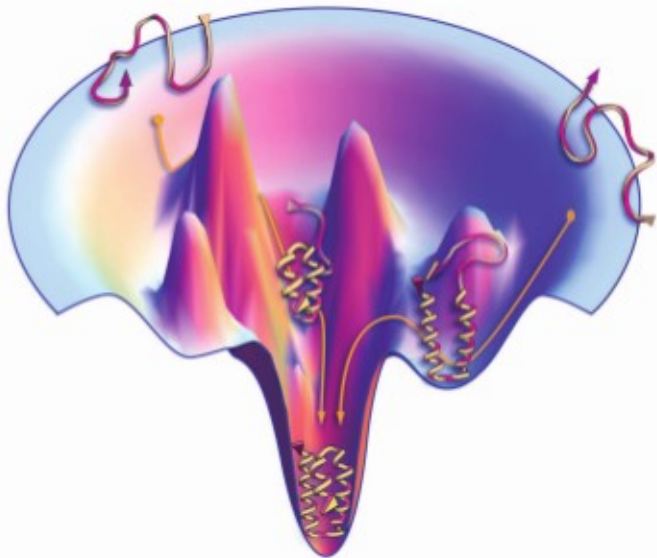
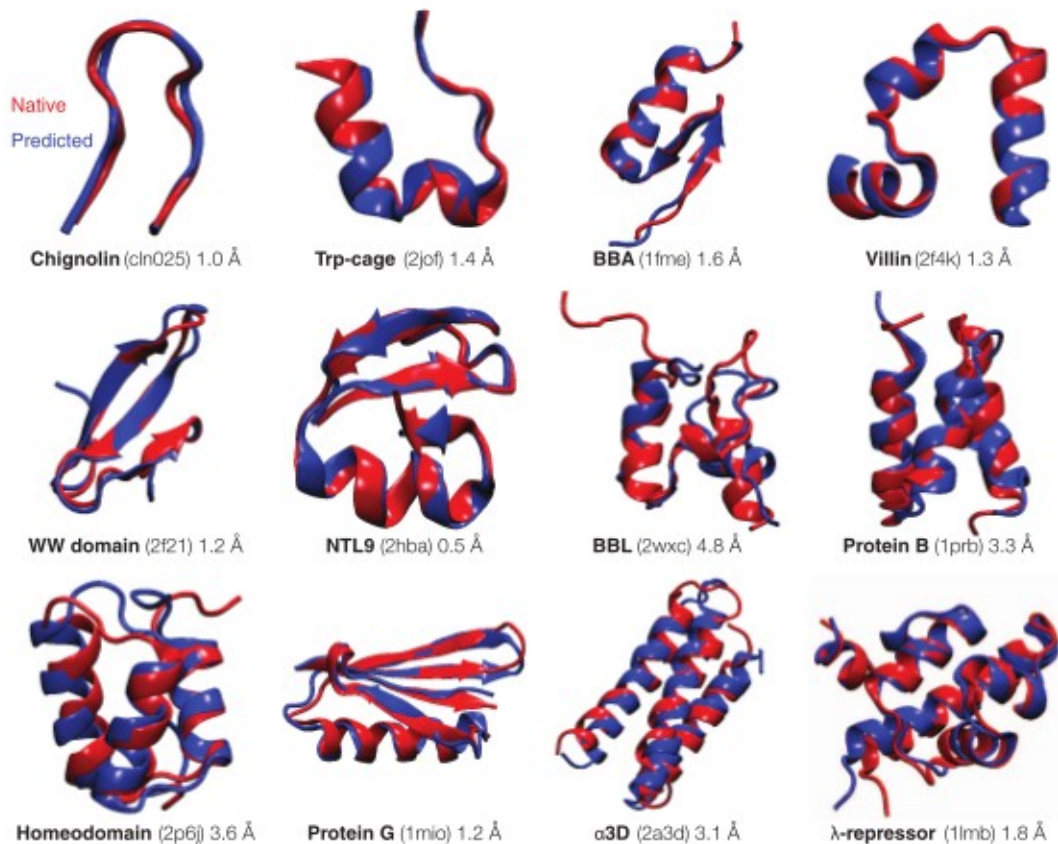


Fig. 3. Proteins have a funnel-shaped energy landscape with many high-energy, unfolded structures and only a few low-energy, folded structures. Folding occurs via alternative microscopic trajectories.

The Protein-Folding Problem, 50 Years On
Ken A. Dill and Justin L. MacCallum
Science 338, 1042 (2012);

Protéines Pliage

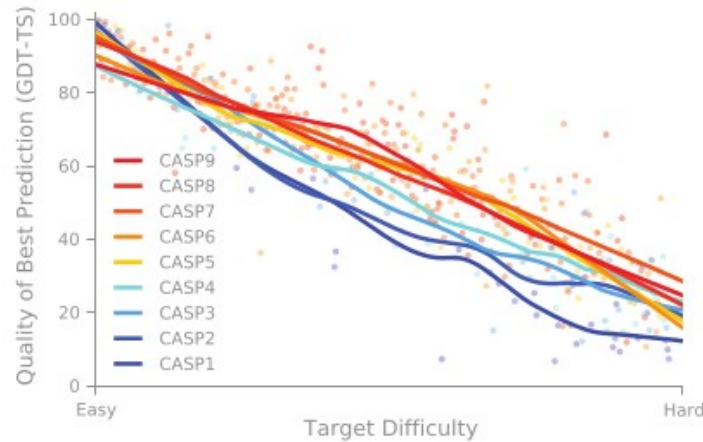


Pour quelques petites protéines: oui...

Fig. 2. Modern physical models can compute the folded structures of some small proteins. Using a high-performance custom computer called Anton (48), Shaw and co-workers observed reversible folding and unfolding in more than 400 events across 12 small proteins to structures within 4.5 Å of the experimental structure (15). The experimental structures are shown in red, and the computed structures are blue. Shown are the name, PDB identifier, and RMSD (root-mean-square deviation between alpha carbon atoms) between the predicted and experimental structures. [Adapted with permission (15)]

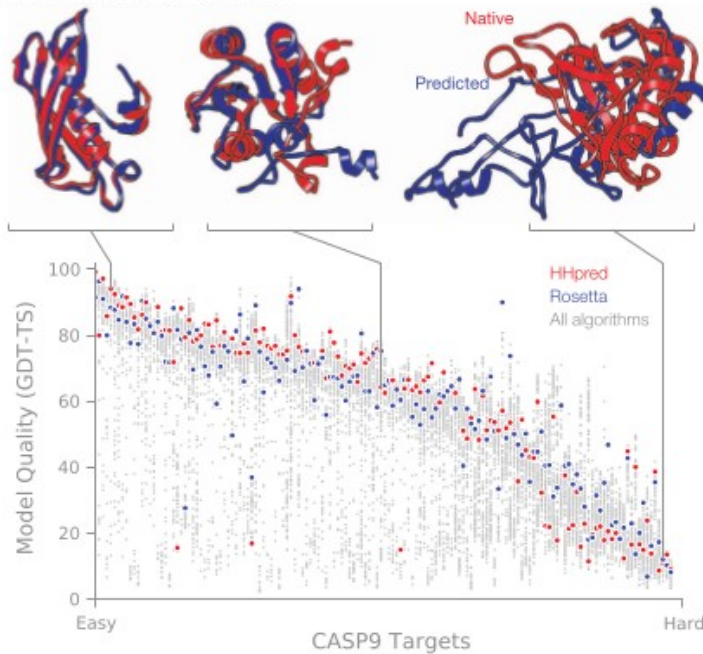
Protéines Pliage

A Historical CASP Performance



... mais pas pour toutes.

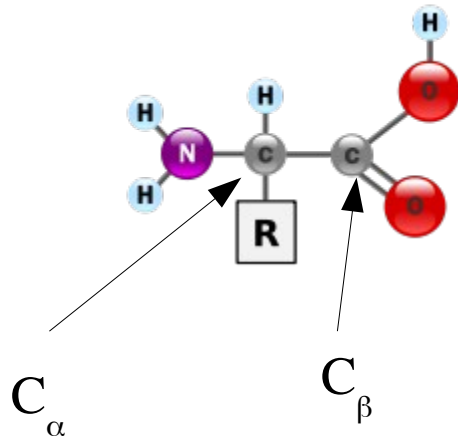
B Performance in CASP9



The Protein-Folding Problem, 50 Years On
Ken A. Dill and Justin L. MacCallum
Science 338, 1042 (2012);

Protéines Pliage: les modeles “atomique”

Molecular Dynamics (MD): résous les équations de Newton numériquement.



$$\frac{d^2}{dt^2} \mathbf{q}_i = - \frac{\partial U(\mathbf{q}^N)}{\partial \mathbf{q}_i}$$

Typiquement, on utilise le logiciel standard (open source) tel que “GROMACS” ou “LAMPS”

C_α seulement; C_α et C_β; “all atom” (sauf H).

$$U_{C_\alpha} = \sum_{\text{bonds}} \epsilon_r (r - r_0)^2 + \sum_{\text{angles}} \epsilon_\theta (\theta - \theta_0)^2 + \sum_{\text{backbone}} \epsilon_D F_D(\phi) \\ + \sum_{\text{contacts}} \epsilon_C \left[5 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{10} \right] + \sum_{\text{non-contacts}} \epsilon_{NC} \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6$$

Dihedral potential

$$F_D(\phi) = [1 - \cos(\phi - \phi_0)] + \frac{1}{2} [1 - \cos(3(\phi - \phi_0))]$$

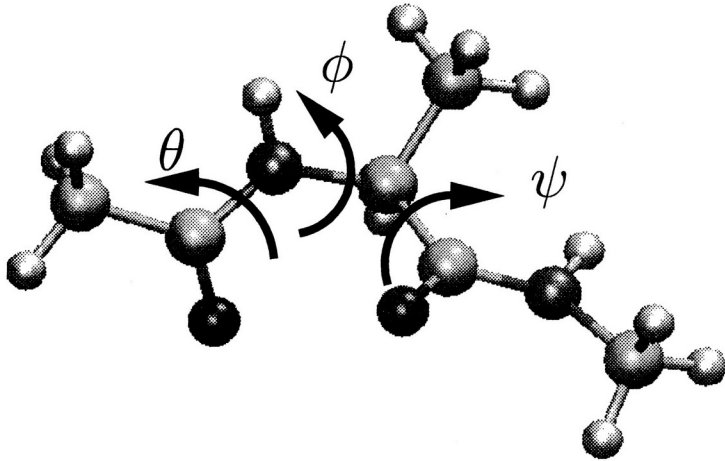
On utilise termes qui dépendent sur autant que 4 voisins.

Protéines Pliage: les modeles stochastique

$$\dot{\mathbf{q}}_i = \mathbf{b}_i(\mathbf{q}) + \mathbf{Q}_{ij} \eta_j(t)$$

$$\frac{d}{dt} \frac{\delta L}{\delta \dot{\mathbf{q}}_i(t)} - \frac{\delta L}{\delta \mathbf{q}_i(t)} = 0$$

$$S_{eff} = -\frac{1}{2} \int_0^T dt \underbrace{\left(\frac{1}{2} (\dot{\mathbf{q}}_i(t) - \mathbf{b}_i(\mathbf{q})) \mathbf{Q}_{ij}^{-2} (\dot{\mathbf{q}}_j(t) - \mathbf{b}_j(\mathbf{q})) + \frac{\partial b_i}{\partial q_i} \right)}_{\text{Lagrangian } L}$$



Alanine dipeptide molecule
(CH₃-CONH-CHCH₃-CONH-CH₃).

“Reaction coordinates of bimolecular isomerization”, Peter G. Bolhuis, Christoph Dellago and David Chandler, PNAS **97** 5877 (2000).

The molecule in vacuum has two stable conformers: the C_{7eq} state with $\phi \approx -86^\circ$ and $\psi \approx 68^\circ$ and the C_{ax} state with $\phi \approx 50^\circ$ and $\psi \approx -50^\circ$. In solution, these positions shift slightly, for instance, the C_{7eq} state is located around $\phi \approx -80^\circ$ and $\psi \approx 160^\circ$.

Protéines Pliage: les modeles stochastique

$$\dot{\mathbf{q}}_i = b_i(\mathbf{q}) + Q_{ij} \eta_j(t)$$

$$\frac{d}{dt} \frac{\delta L}{\delta \dot{\mathbf{q}}_i(t)} - \frac{\delta L}{\delta \mathbf{q}_i(t)} = 0$$

$$S_{\text{eff}} = -\frac{1}{2} \int_0^T dt \underbrace{\left(\frac{1}{2} (\dot{\mathbf{q}}_i(t) - b_i(\mathbf{q})) Q_{ij}^{-2} (\dot{\mathbf{q}}_j(t) - b_j(\mathbf{q})) + \frac{\partial b_i}{\partial q_i} \right)}_{\text{Lagrangian } L}$$

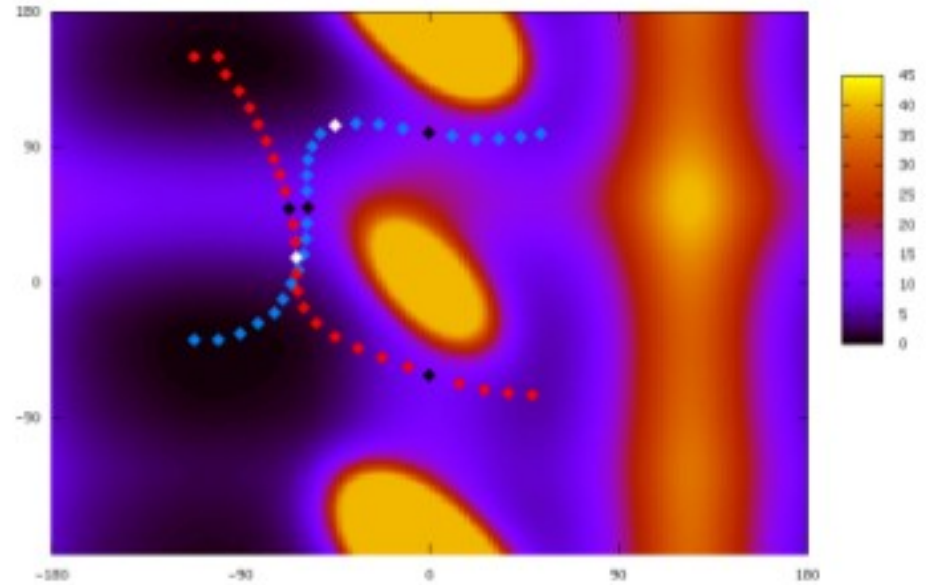
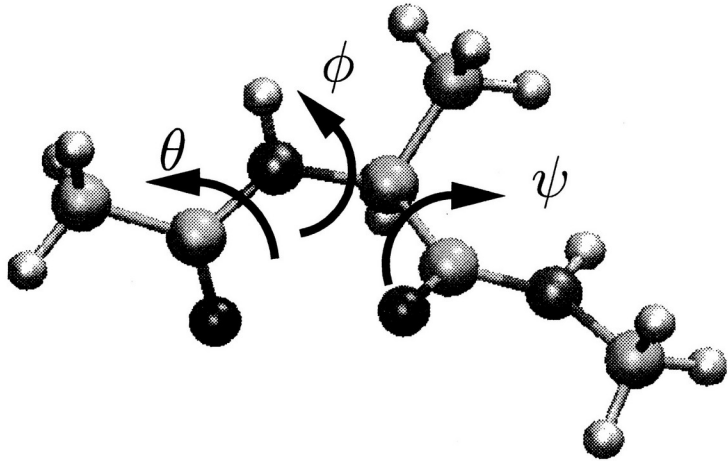


FIG. 1 (color online). Dominant Folding Paths for the $C7_{ax} \rightarrow C7_{eq}$ (red squares, from bottom right to top left) and $\alpha_L \rightarrow \alpha_R$ (blue squares, from bottom left to top right) transitions. In the background, the free-energy profile for the ψ and ϕ dihedrals is shown (in units of kJ/mol). Black and white squares identify the minimum residence time conformations and the commitment analysis transition states, respectively.

“Quantitative Protein Dynamics from Dominant Folding Pathways”, M. Sega, P. Faccioli, F. Pederiva, G. Garberoglio, and H. Orland, PRL **99**, 118102 (2007)