NANOPHYSIQUE
INTRODUCTION PHYSIQUE AUX NANOSCIENCES

Ch. 9. Les Protéines

James Lutsko

Lecture 10, 2019-2020

Les Protéines

- Introduction
- Protéine interactions
- Condensation
- Pliage

Les Protéines

- Introduction
- Protéine interactions
- Condensation
- Pliage

De polymers a protéines

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

Homopolymer: A+A+A → AAAA
Copolymer: A+B+A+B → ABAB

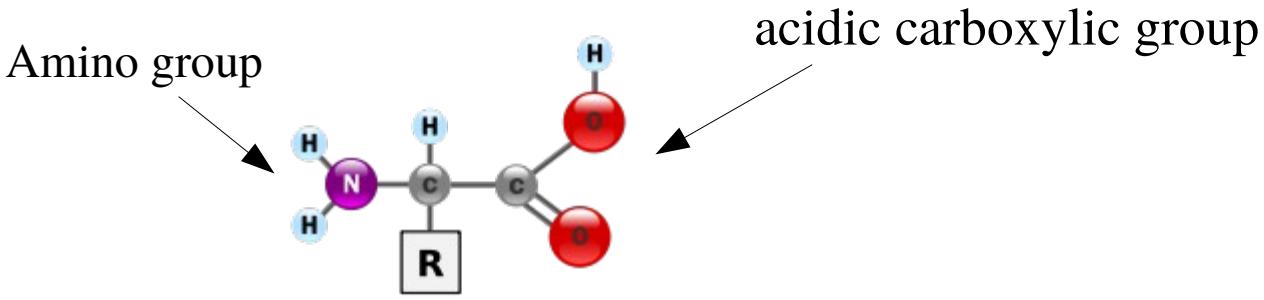
Covalent liens

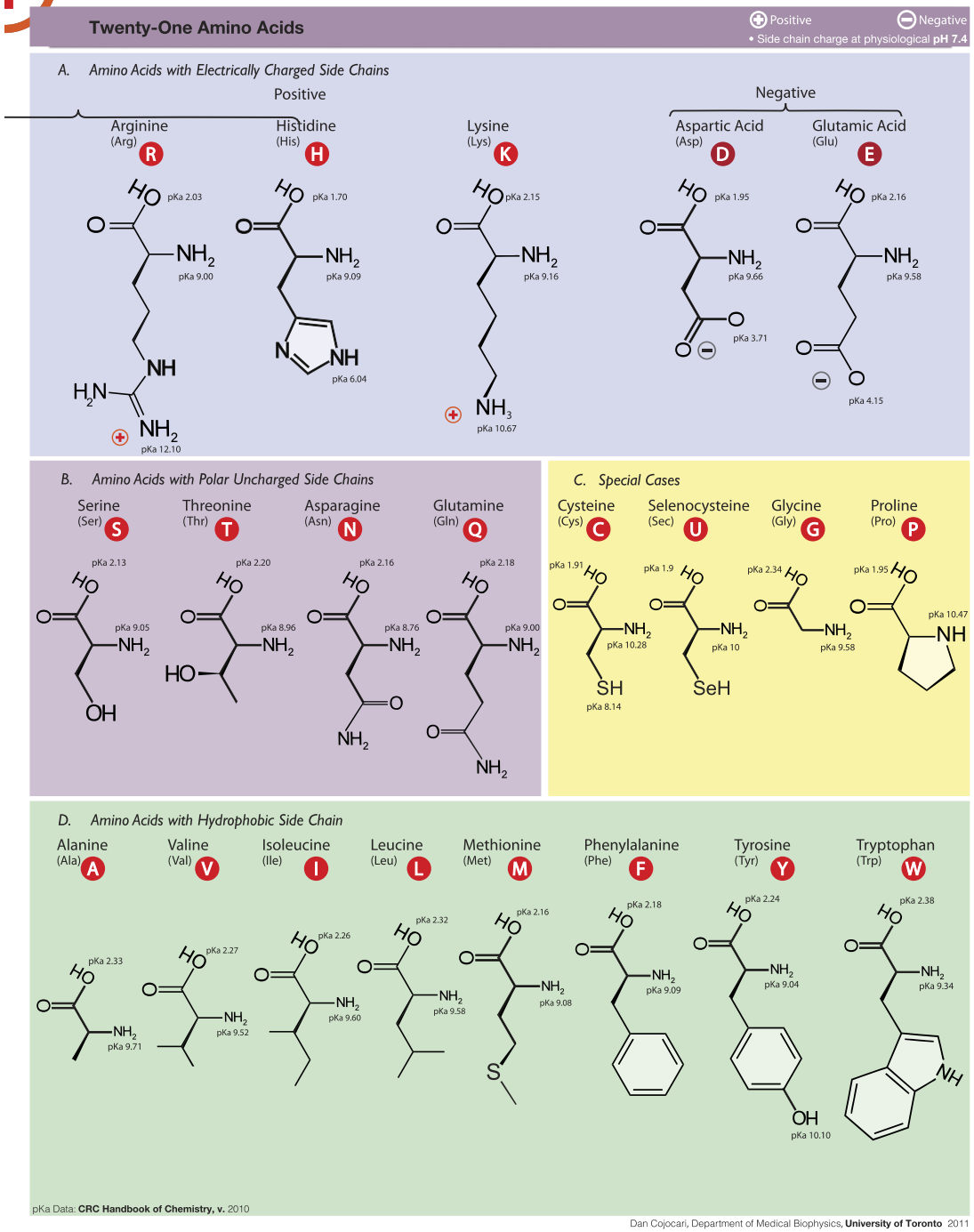
Catalysis, DNA replication, Transport, ...

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 molecule 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) attacheé au l'atom charbon. Donc, ayant la formule de NH2CHRCOOH.



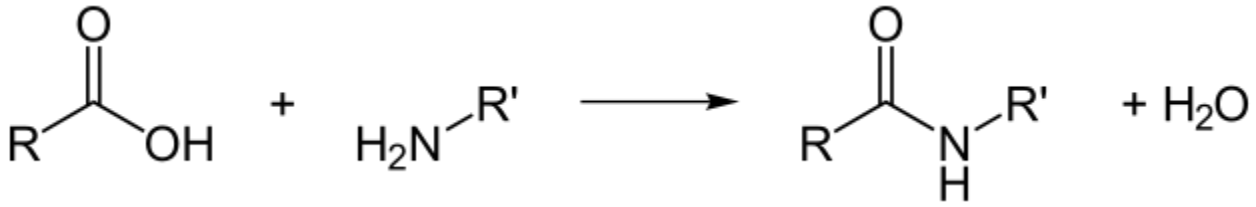


From http://en.wikipedia.org/wiki/File:Amino_Acids.svg

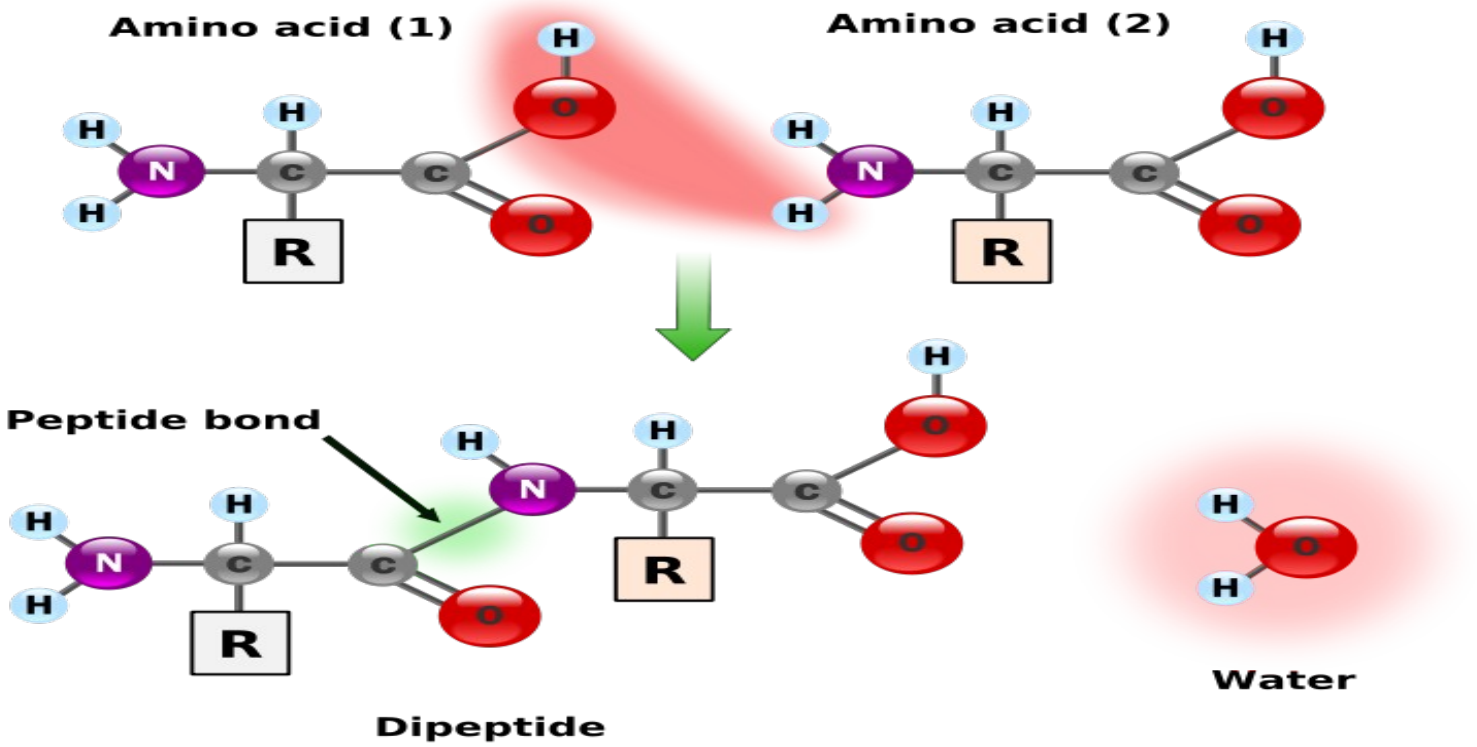
Peptides

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.



"dehydration reaction"
"condensation"



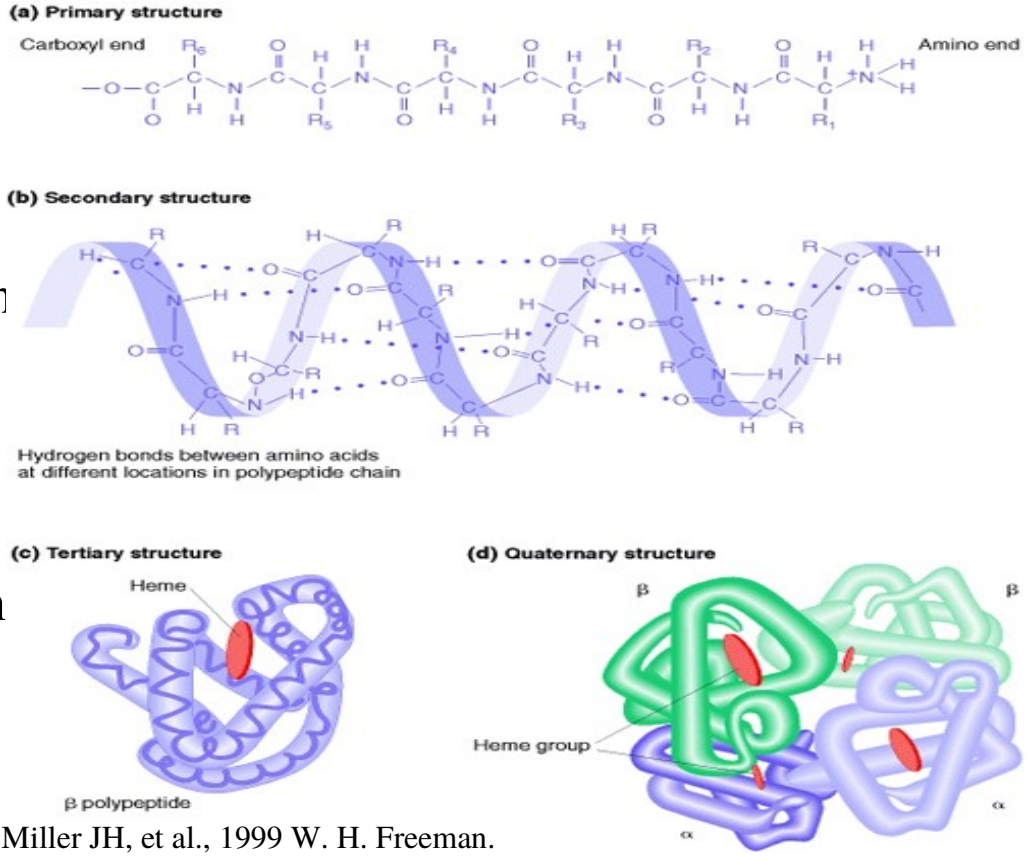
Protéine: Structure

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

Residue: acide aminé dans une protéin

Backbone: le série liée de N, C, O.

Peptide <~ 20-30 residues <~ protein



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

Primary structure: acide aminé sequence

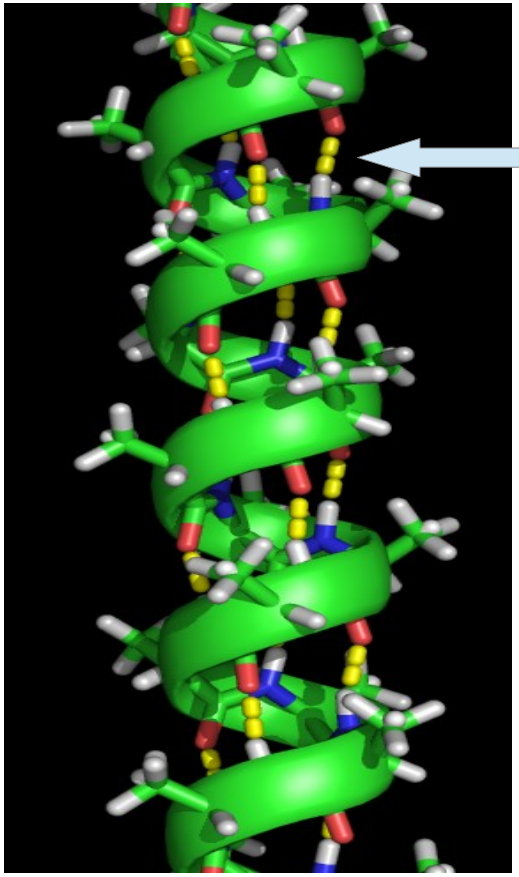
Secondary structure: 3-D structures composé de segments locaux définis par les liens d'hydrogène entre "backbone" amino and carboxyle groupes (α-helix, β-sheet, turns (hairpin turn),...)

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

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

Protéine Secondary Structure: α-helix

α-helix



H-bond

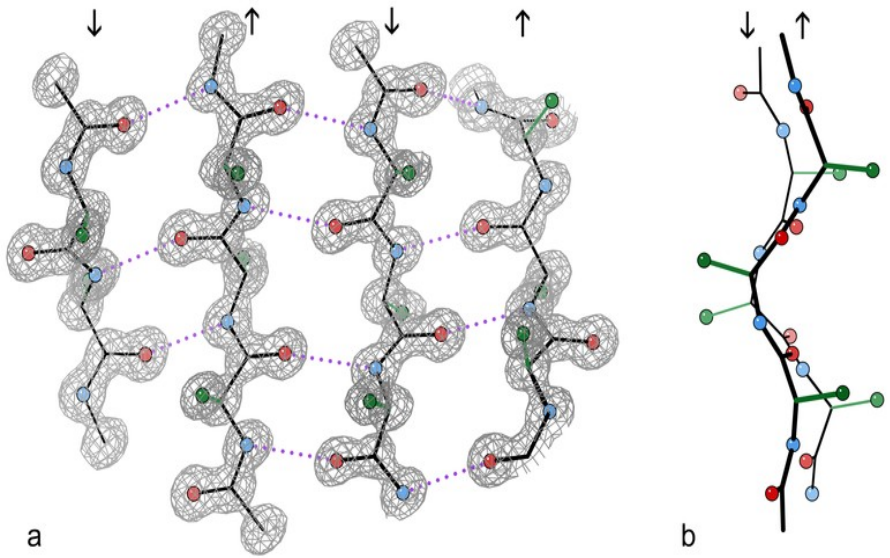
Liens entre residue n et n+4 (“n+4 → n bonding”)

4-40 residues possible
Typique: 10 residues qui correspondent au 3 turnes.

http://en.wikipedia.org/wiki/File:Alpha_helix.png

Protéine Secondary Structure: β-sheet

β-strand: chaîne de peptide entièrement prolongée

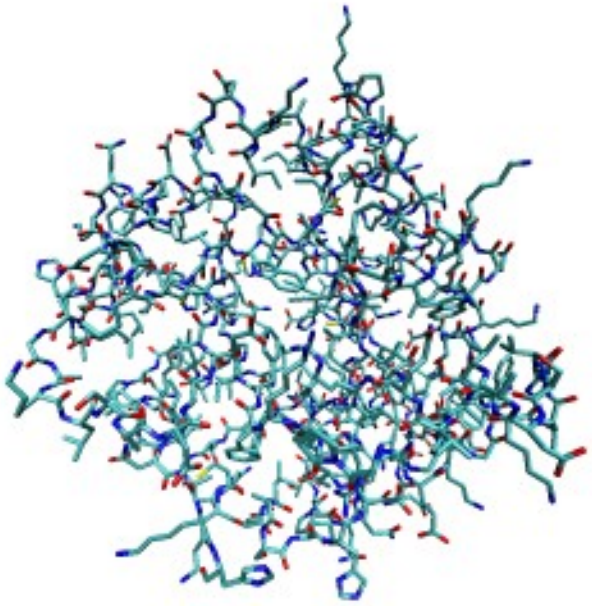


β-sheet: réticulés β-strands

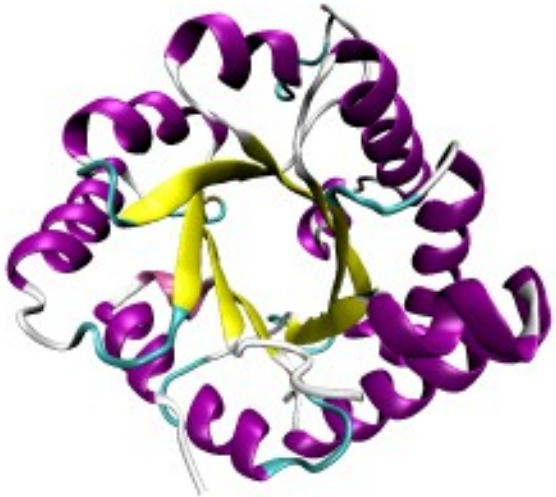
Exemple d'un fragment de feuillet β à quatre chaines antiparallèles extrait de la structure cristalline de l'enzyme catalase (résolution 0,88 Å). a) Vue de face, montrant les liaisons hydrogènes (en pointillés) entre les groupes NH et CO des acides aminés adjacents. Les flèches indiquent l'orientation des chaines, 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 chaines centrales montrant la torsion à droite des chaines 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

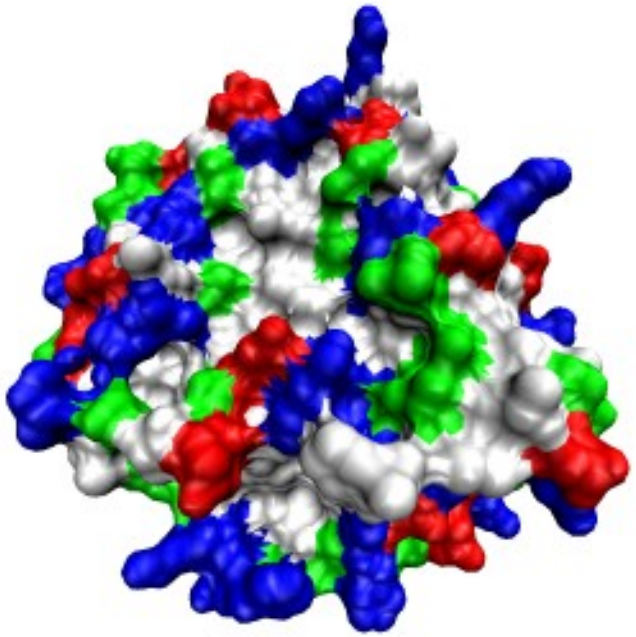
triose phosphate isomerase



All atom



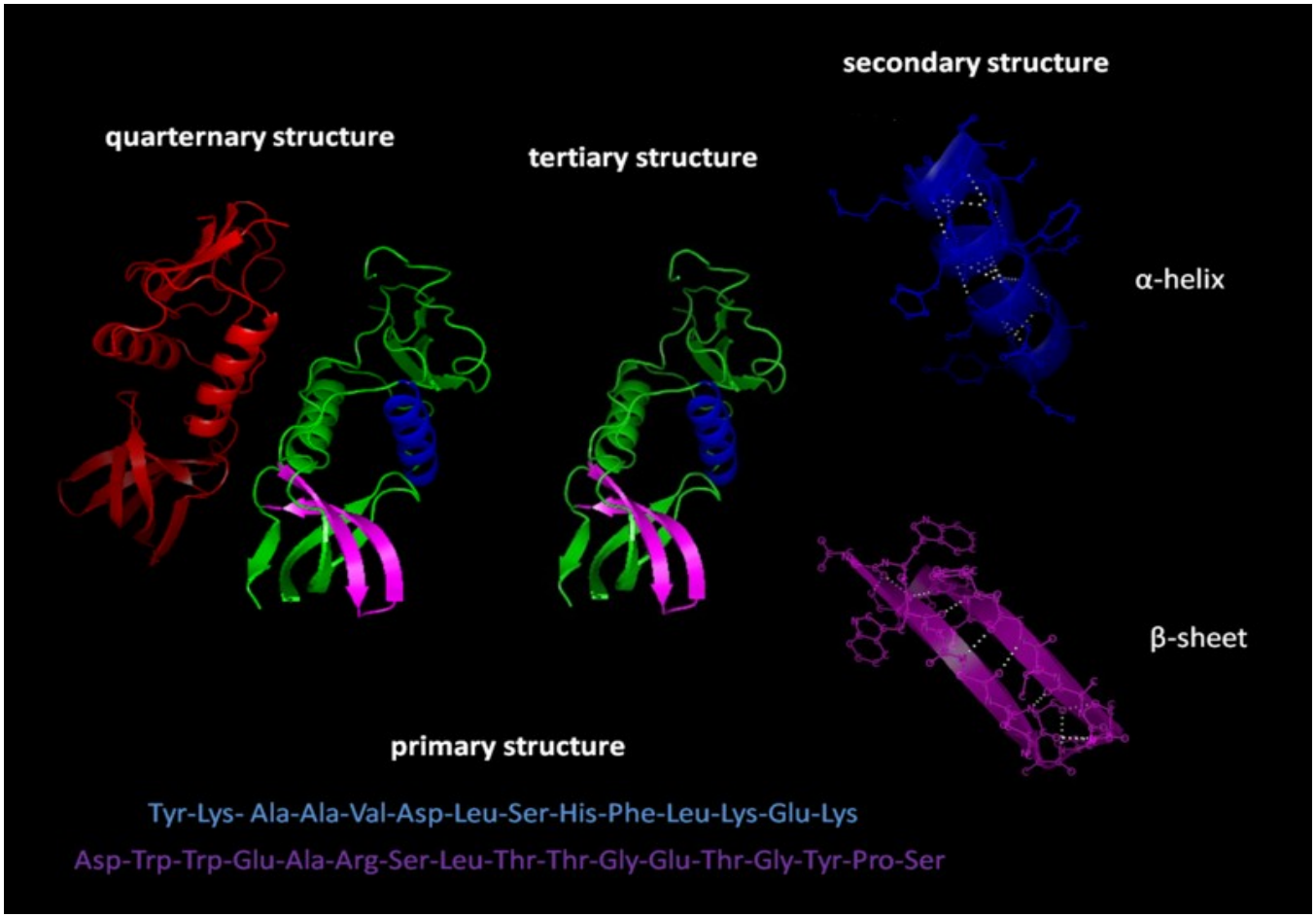
Secondary et tertiary (α-helices and β-sheets)



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

http://en.wikipedia.org/wiki/Protein

Protéine Structure: examples



http://en.wikipedia.org/wiki/Protein_tertiary_structure

Protéine: domains

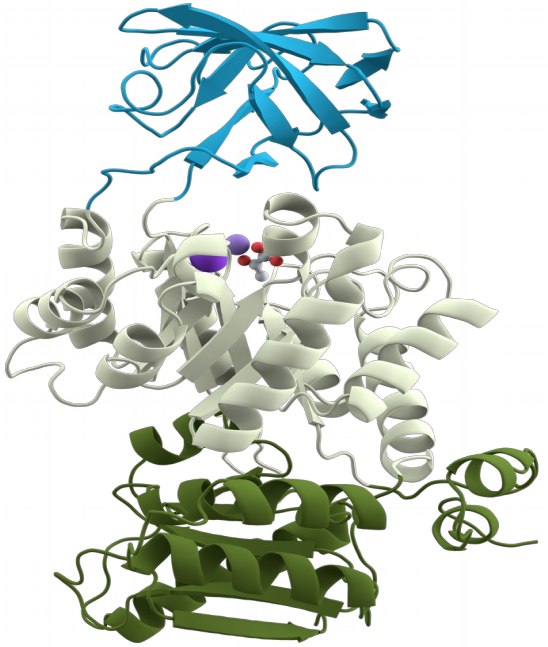
Structural Domain: is an element of the protein's overall structure that is self-stabilizing and often folds independently of the rest of the protein chain.

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".

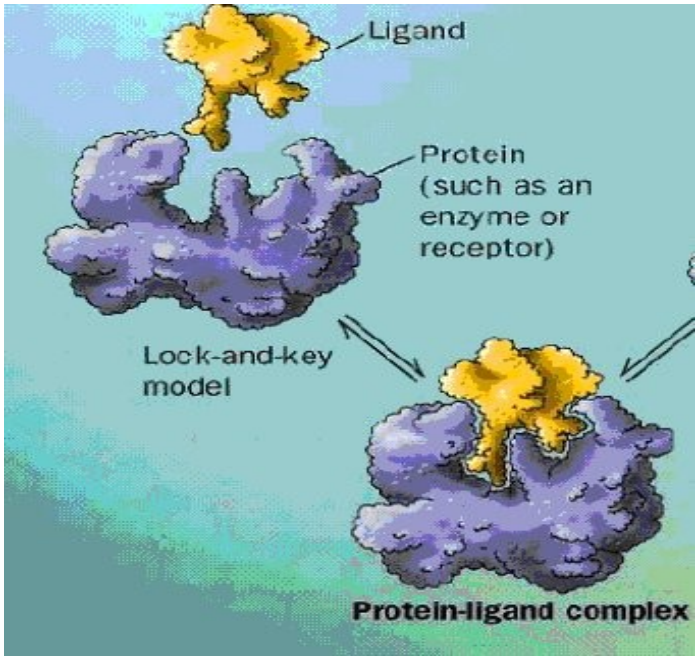
Because they are independently stable, domains can be "swapped" by genetic engineering between one protein and another to make chimeras.



Pyruvate kinase, a protein with three domains.

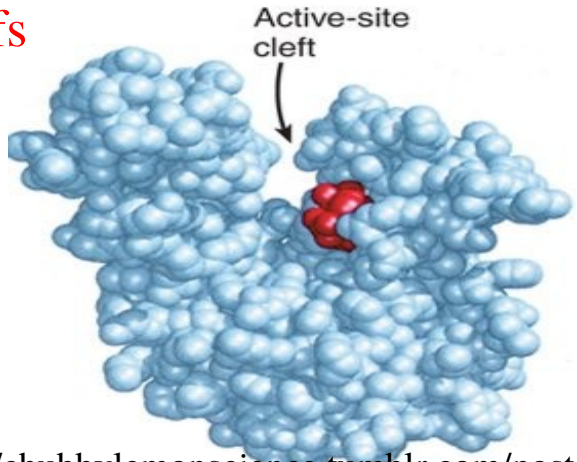
Protéine: la relation entre structure et fonction

"Lock and key"



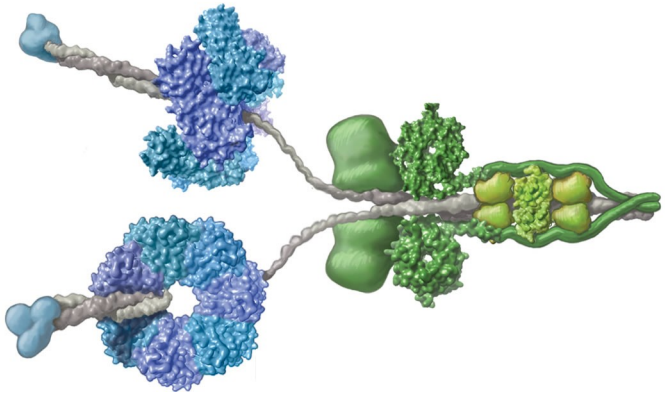
http://student.biology.arizona.edu/

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

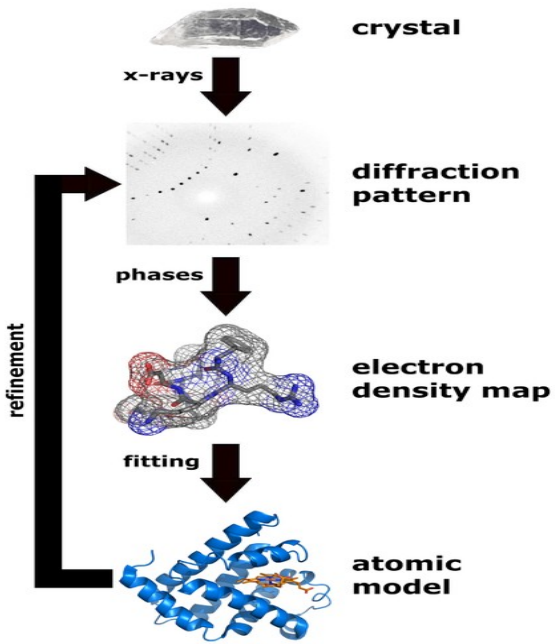
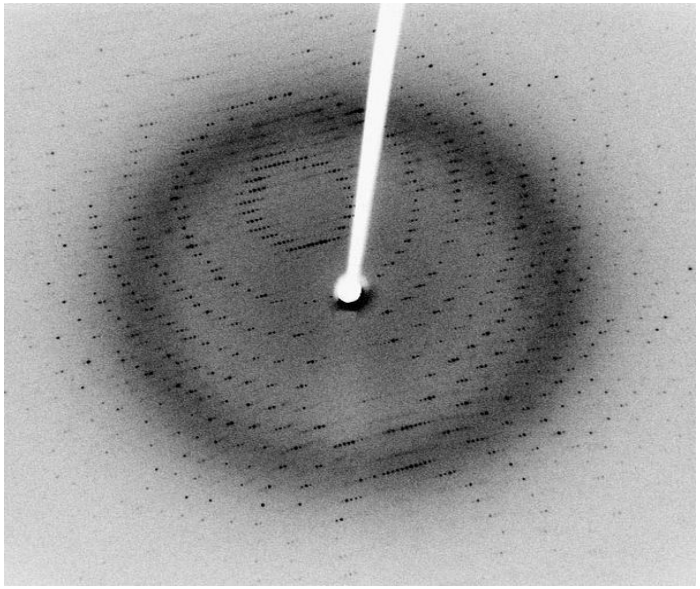


Image: Thomas Splettstoesser
http://en.wikipedia.org/wiki/X-ray_crystallography



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

Diffusion statique

$$I(\boldsymbol{q}) \sim S(\boldsymbol{q}) = 1 + \rho \int [g(r) - 1] e^{i 2\pi \boldsymbol{q}\cdot\boldsymbol{r}} d\boldsymbol{r}$$

Besoin des cristaux de haute qualité.

Diffusion dynamique

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

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

Hydrodynamic radius

Protéine: la but en bref

- ADN → ARN →les Protéines
- Les Protéines → 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→ on doit comprendre protéine cristallisation et croissance.
- Théorie: pour predire leur structure, on doit comprendre le pliage des protéines.

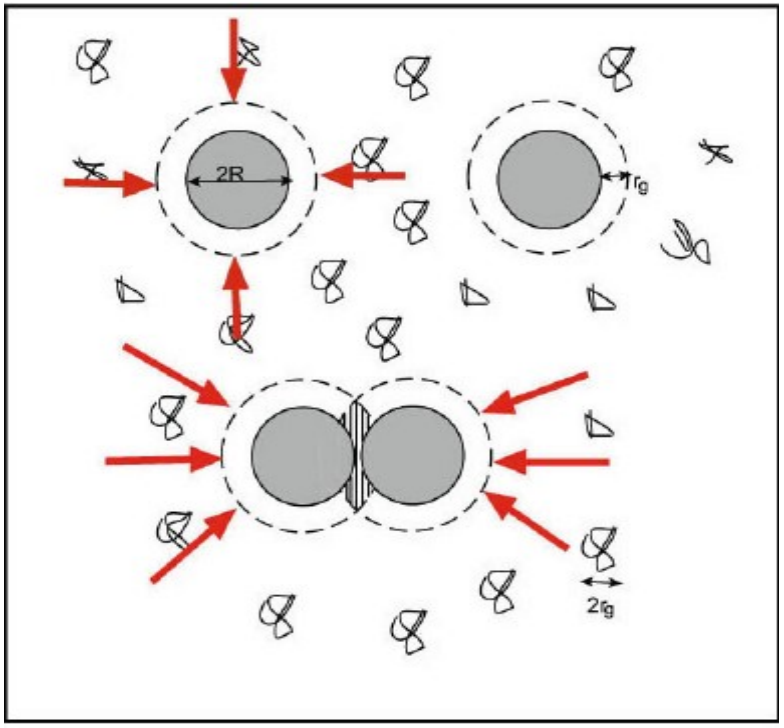
Les Protéines

- Introduction
- Protéine interactions
- Condensation
- Pliage

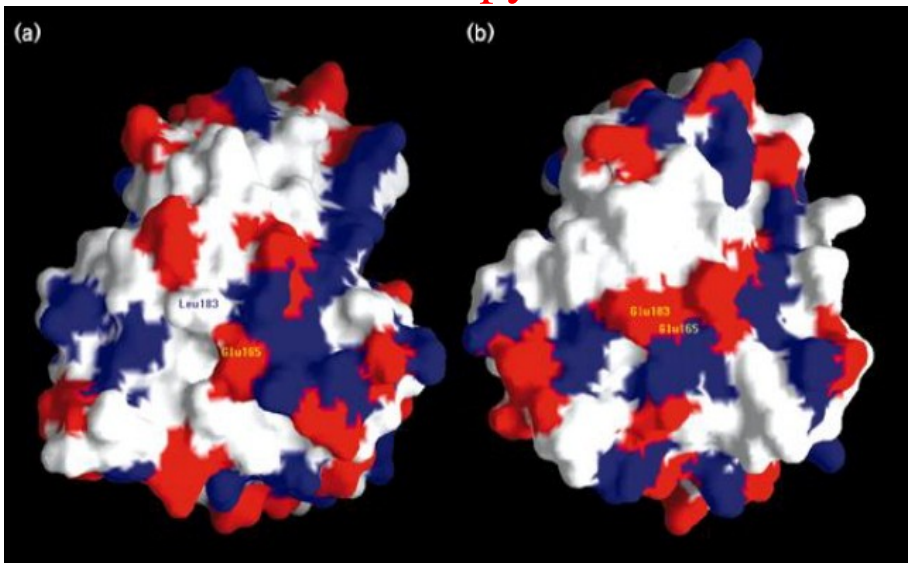
Protéines: beaucoup des interactions

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

Depletion forces



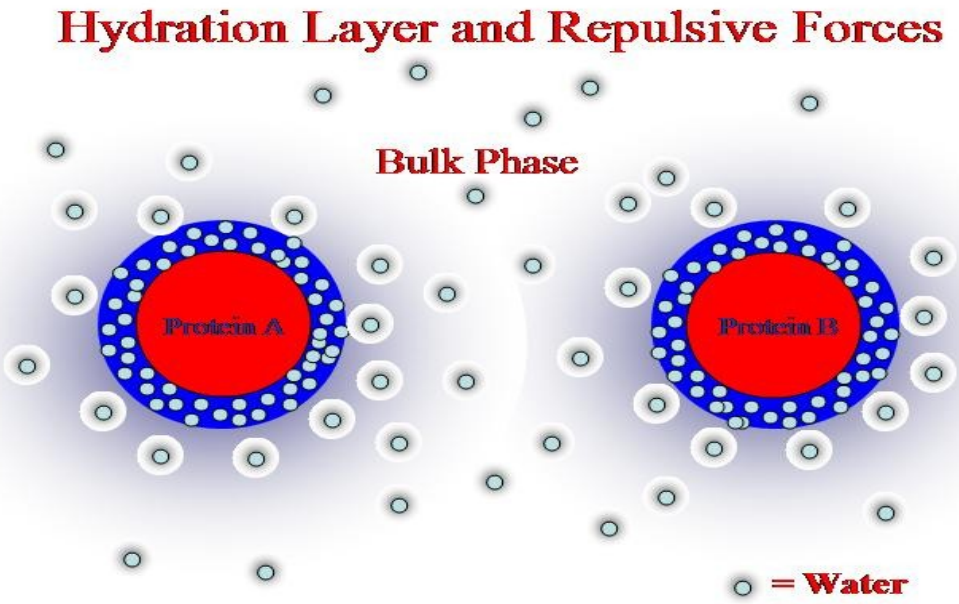
Anisotropy



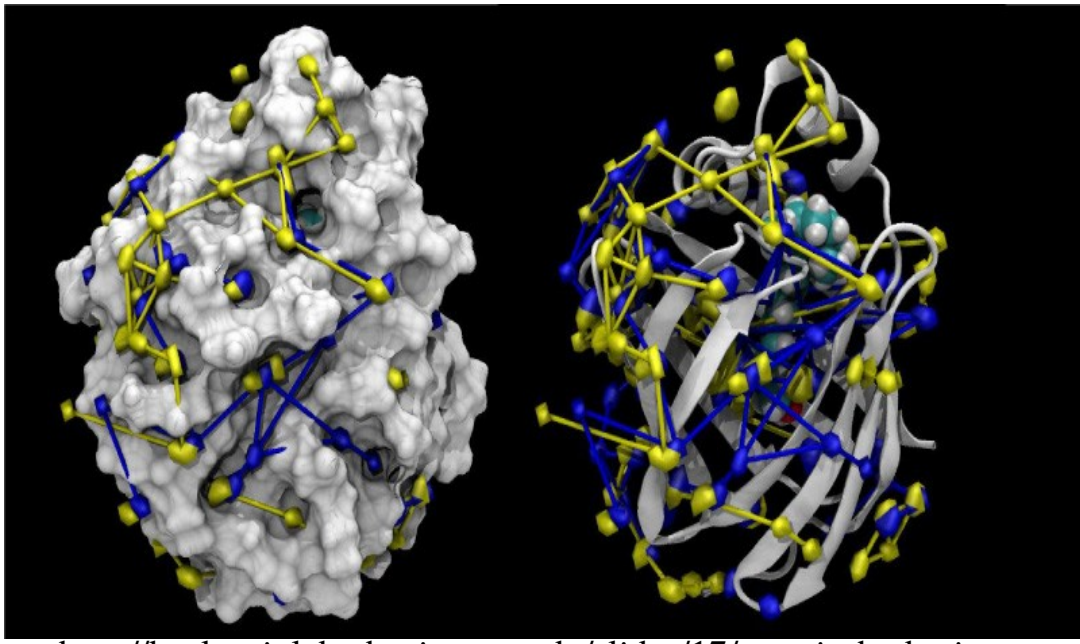
triosephosphate isomerase: des régions chargées rouge (-) et bleu(+)

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

Protéines: beaucoup des interactions



Hydration sites (from MD)



http://becksteinlab.physics.asu.edu/slides/17/protein-hydration

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: $A_H=\frac{3}{4}k_BT\left(\frac{\epsilon_1-\epsilon_2}{\epsilon_1+\epsilon_2}\right)^2+\frac{\left(\epsilon_1^2-\epsilon_2^2\right)^2}{\left(\epsilon_1^2+\epsilon_2^2\right)^{3/2}}$

$\epsilon_1=$ particule

$\epsilon_2=$ medium

Keesom-Debye interaction

London dispersion interaction

Gunton, Shiryayev and Pagan, “Protein condensation”, Cambridge, 2007.

Protéine: des modeles d'interaction simple

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

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

Solution: des ions avec densite n_j et charge z_j

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

$\mu_j(\boldsymbol{r})=\mu_j^0+ez_jV(\boldsymbol{r})$ $d\mu_j^0(\boldsymbol{r})=-k_BTd\ln n_j(\boldsymbol{r})$ $d\mu_j(\boldsymbol{r})=0\Rightarrow n_j(\boldsymbol{r})=n_j^{(0)}e^{-z_jeV(\boldsymbol{r})/k_BT}$

Dehors les protéines: $\nabla^2V=0$

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

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

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:

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

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

Polymer reference interaction site model "PRISM":

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

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

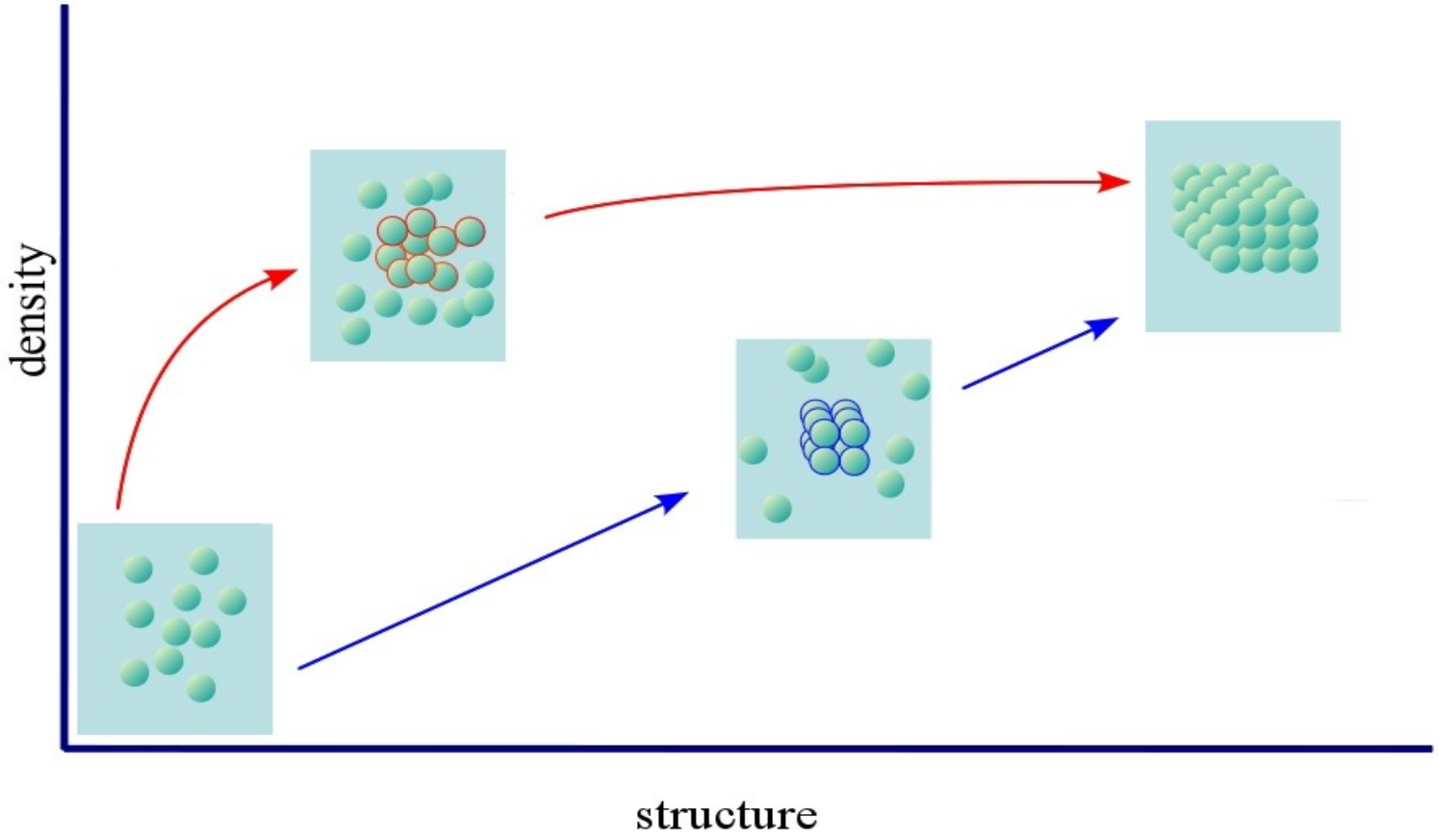
Les Protéines

- Introduction
- Protéine interactions
- Condensation
- Pliage

Nucléation Protéines: un bref historique

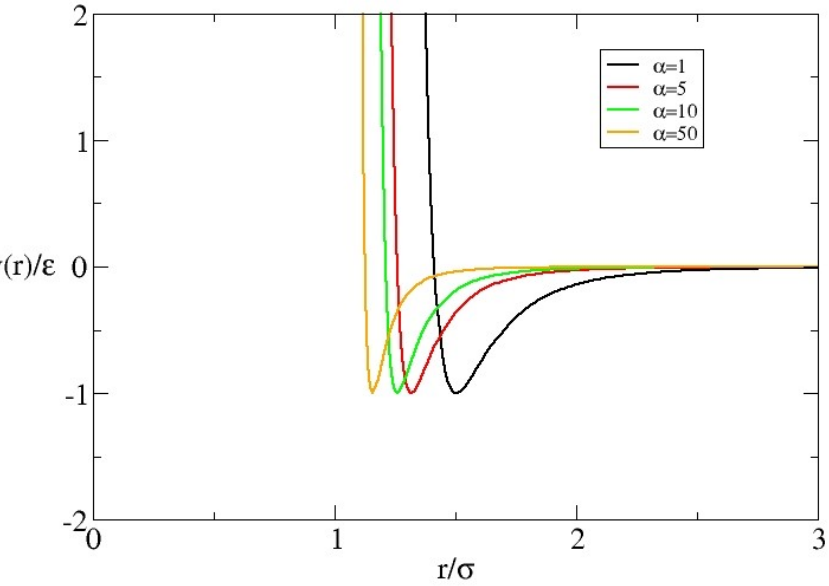
- Crystal quality correlated with osmotic virial coefficient. (George and Wilson, Acta Crystallogr., Sect. D 1994)
- Phase diagram of globular proteins mapped onto simple liquids. (Rosenbaum, Zamora, Zukoski, PRL 1996)
- Simulations studies (ten Wolde and Frenkel, Science, 1997) and theoretical models (Talanquar and Oxtoby, JCP 1998; Sear JCP 2001; Shiryayev and Gunton JCP 2004) show lower barriers and enhanced nucleation rates near metastable critical point.
- Experimental evidence for two-step mechanism throughout phase diagram (Vekilov, P. G. Cryst. Growth Des. 2004); theoretical evidence for proteins and simple fluids (Lutsko and Nicolis, PRL 2006; Adv. Chem. Phys. 151 2012).
- Simulation/theory demonstrating importance of both free energy landscape and dynamics (Hedges and Whitelam JCP 2012).

Protéine: nucléation à deux étapes

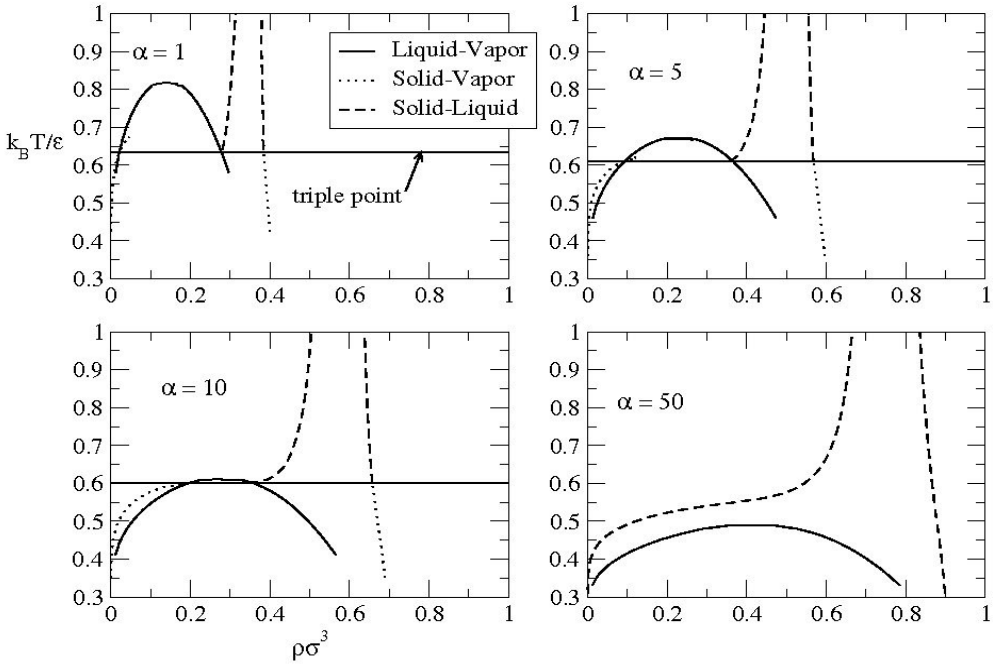


Ten Wolde-Frenkel potential

ten Wolde and Frenkel,SCIENCE 1/7 VOL. 277 1/7 26 SEPTEMBER 1997



$$v(r)=\begin{cases}\infty, & r<\sigma \\ \dfrac{4\epsilon}{\alpha^2}\left(\left(\dfrac{1}{\left(\frac{r}{\sigma}\right)^2-1}\right)^6-\alpha\left(\dfrac{1}{\left(\frac{r}{\sigma}\right)^2-1}\right)^3\right), & r>\sigma\end{cases}$$



Lutsko and Nicolis, J. Chem. Phys. 122, 244907 (2005)

Two-step nucleation

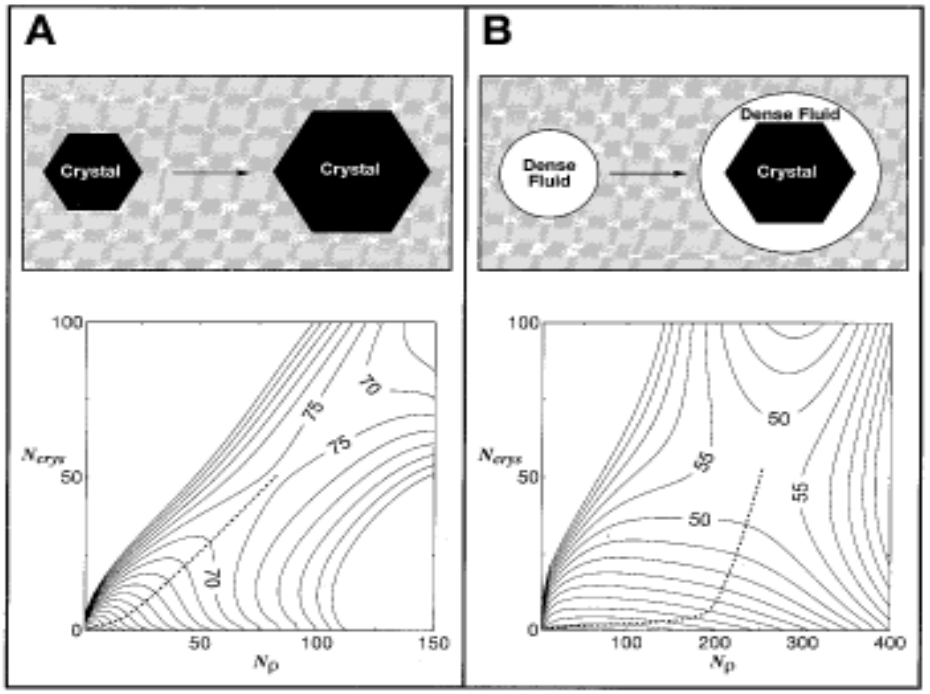
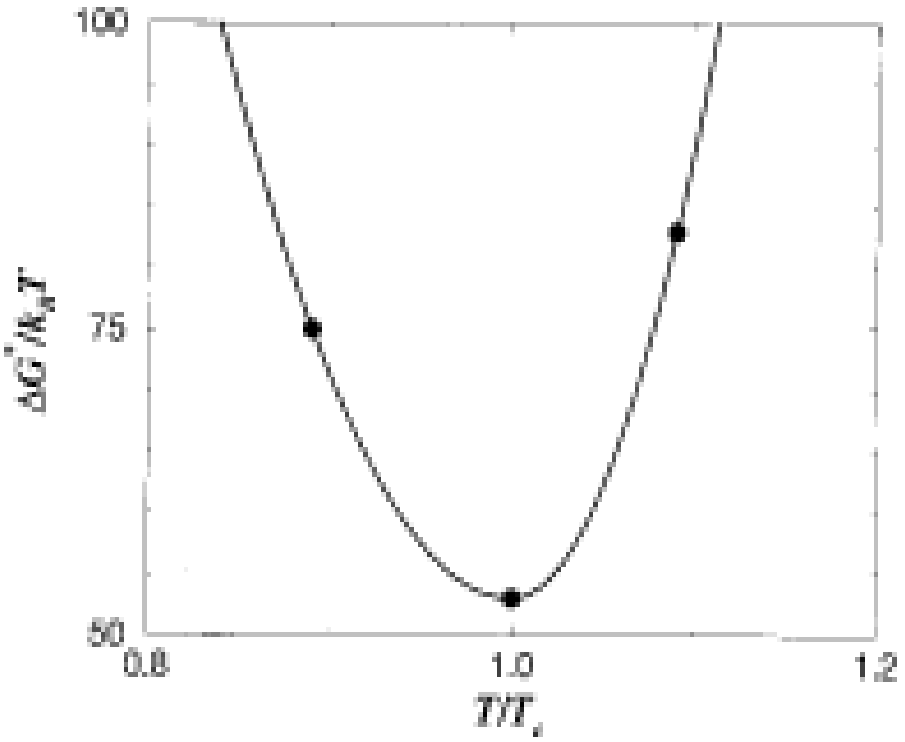


Fig. 2. Contour plots of the free-energy landscape along the path from the metastable fluid to the critical crystal nucleus for our system of spherical particles with short-range attraction. The curves of constant free energy are drawn as a function of N_p and N_{crys} and are separated by $5k_BT$. (**A**) The free-energy landscape well below the critical temperature ($T/T_c = 0.89$). The lowest free-energy path to the critical nucleus is indicated by a dashed curve. This curve corresponds to the formation and growth of a highly crystalline cluster. (**B**) As (A), but for $T = T_c$. In this case, the free-energy valley (dashed curve) first runs parallel to the N_p axis (formation of a liquidlike droplet), and then moves toward a structure with a higher crystallinity (crystallite embedded in a liquidlike droplet). The free-energy barrier for this route is much lower than the one in (A).

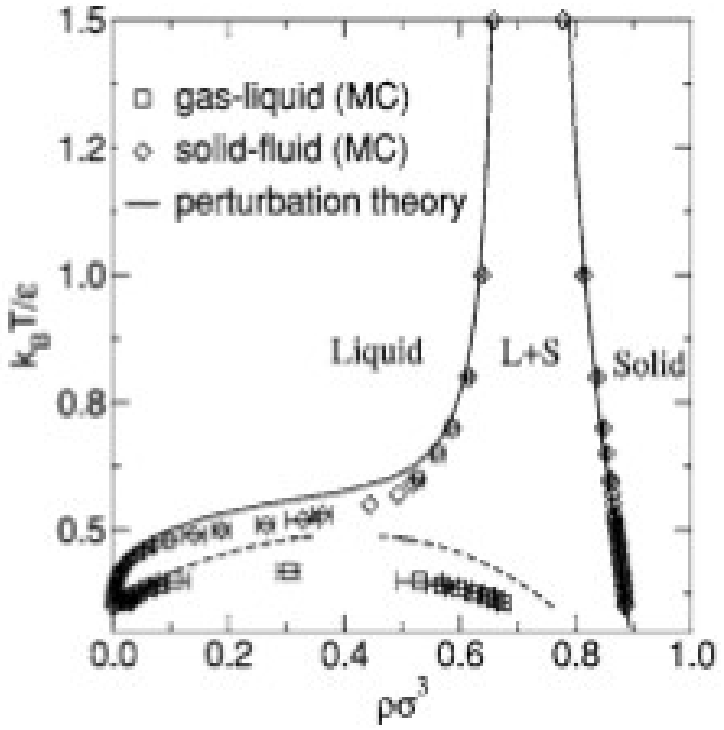
ten Wolde and Frenkel,SCIENCE 1⁄7 VOL. 277 1⁄7 26 SEPTEMBER 1997

Protéine: nucléation à deux étapes



ten Wolde and Frenkel,Science **277**, 1975 (1997)

Protéine: nucléation à deux étapes

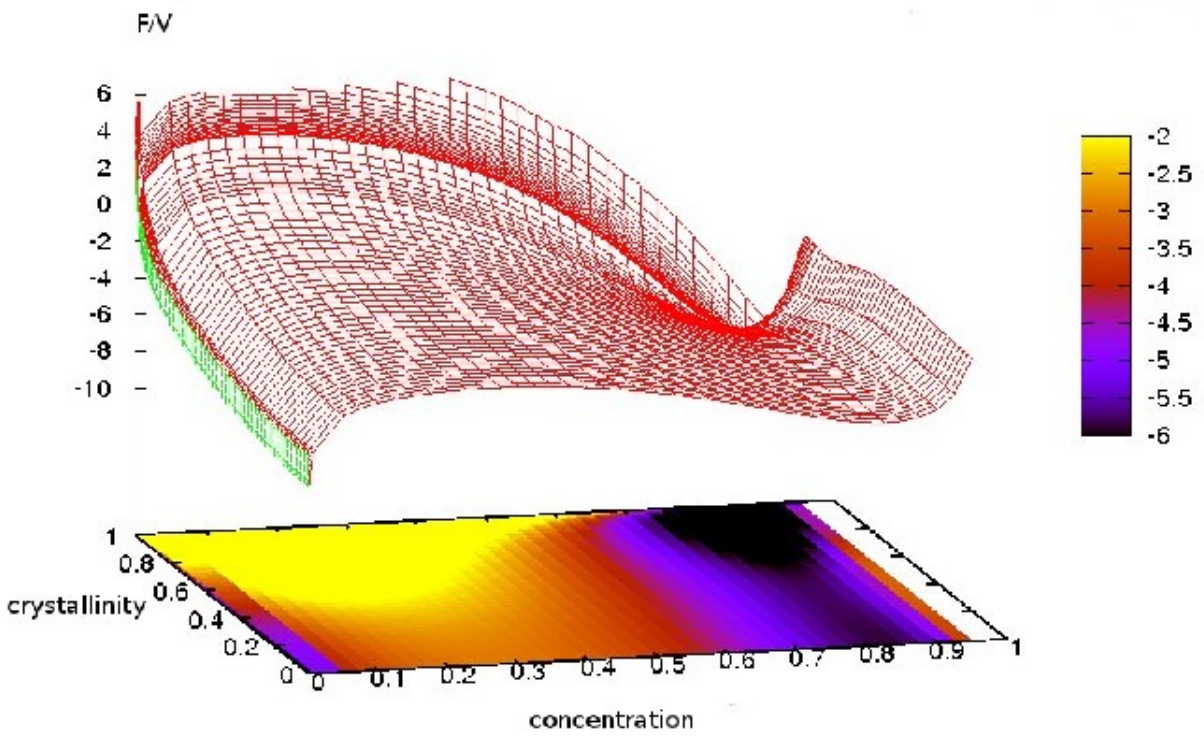


Thermodynamics via DFT

$$F[\rho] = F_{FMT}[\rho] + V\left(f_{liq}(\bar{\rho}) - \frac{1}{V}F_{FMT}(\bar{\rho})\right)$$

Lutsko and Nicolis, PRL 96, 046102 (2006)

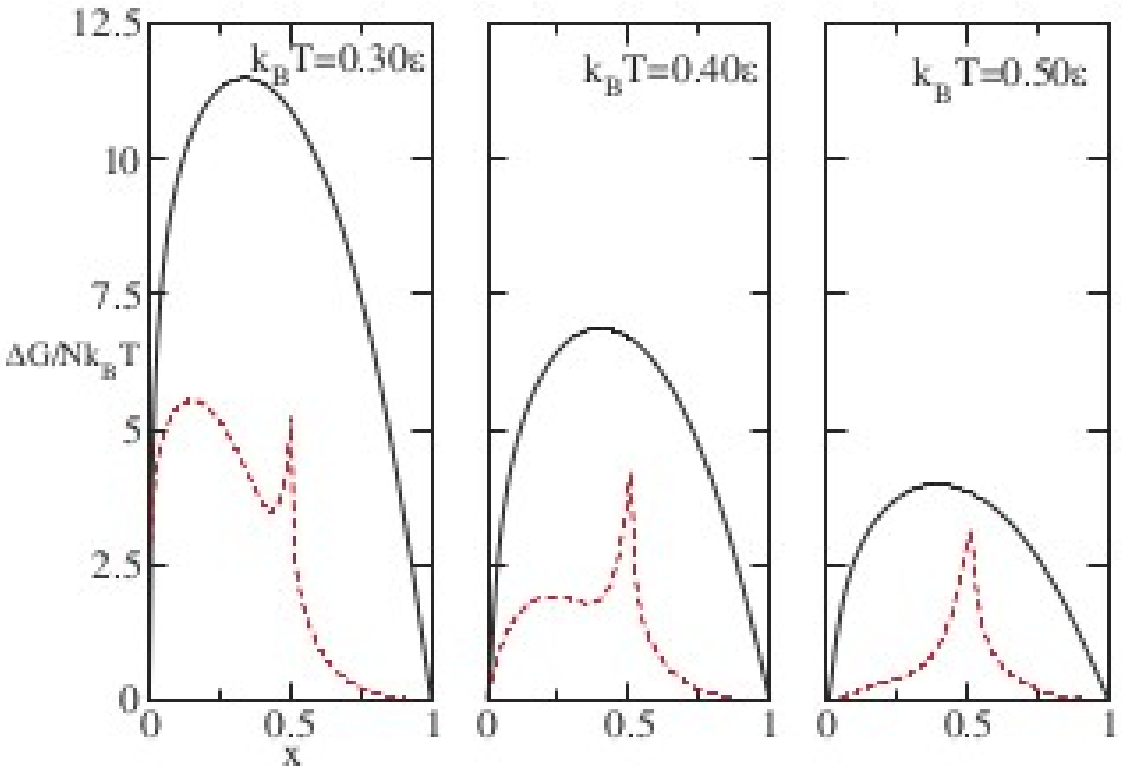
Protéine: nucléation à deux étapes



Surface d'energie libre calcule avec DFT

Lutsko and Nicolis, PRL 96, 046102 (2006)

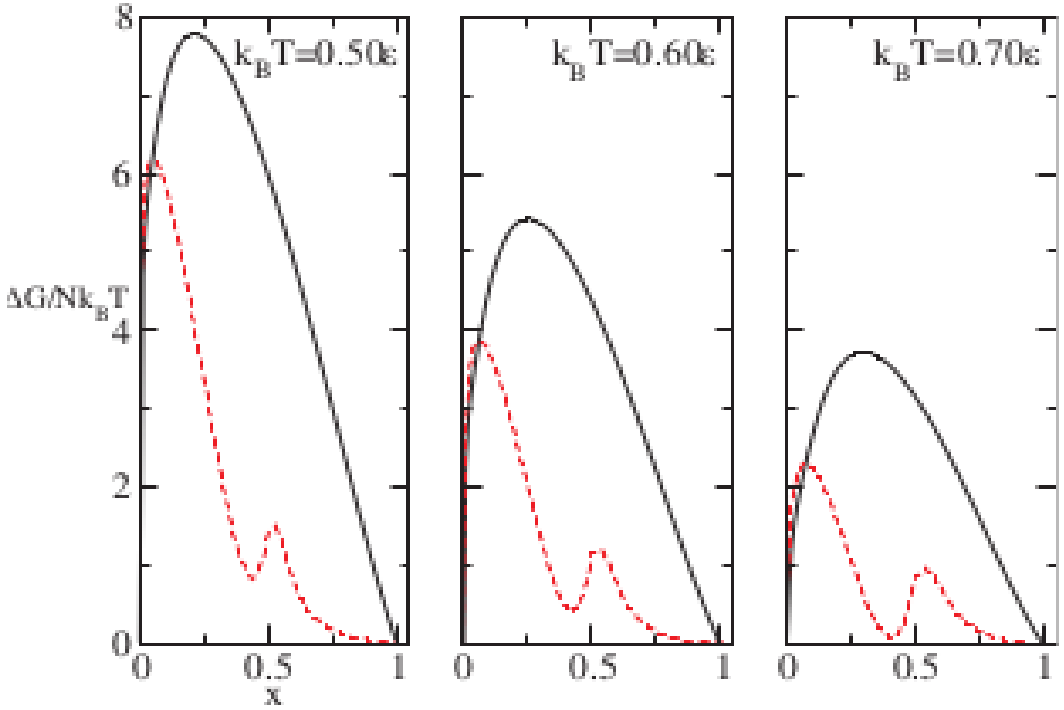
Protéine: nucléation à deux étapes



Nucleation pathways in “globular proteins”.

Lutsko and Nicolis, PRL 96, 046102 (2006); Lutsko Adv. Chem. Phys. 151, 137 (2012).

Fluide simple: nucléation à deux étapes

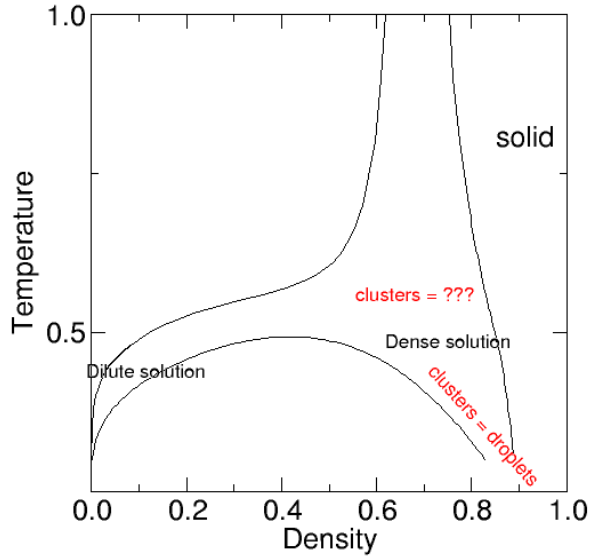


Barriers for simple fluid (LJ)

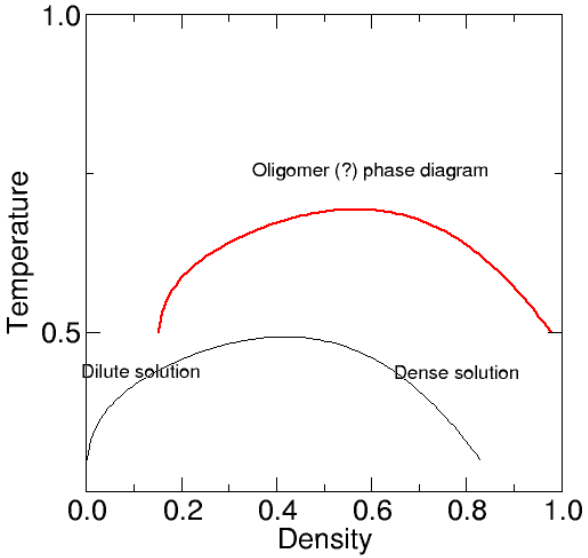
Lutsko and Nicolis, PRL 96, 046102 (2006); Lutsko Adv. Chem. Phys. 151, 137 (2012).

The mysterious protein clusters

- Protein clusters became of interest because of “two step nucleation”



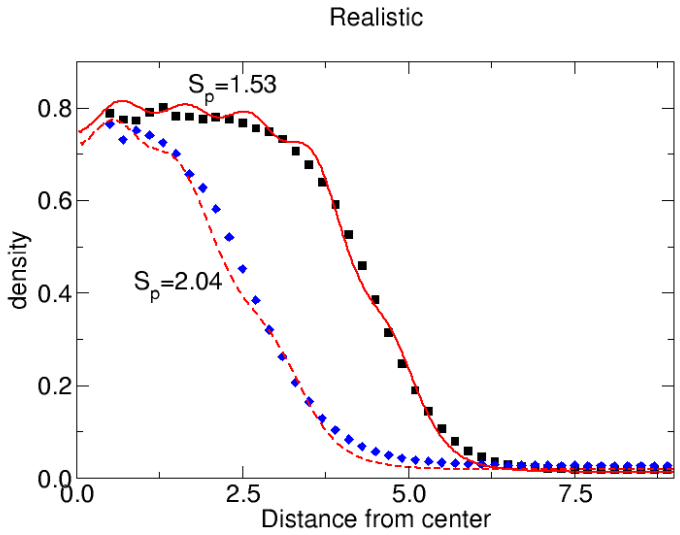
Hypothesis: second species



- Clusters above the critical point cannot be metastable droplets

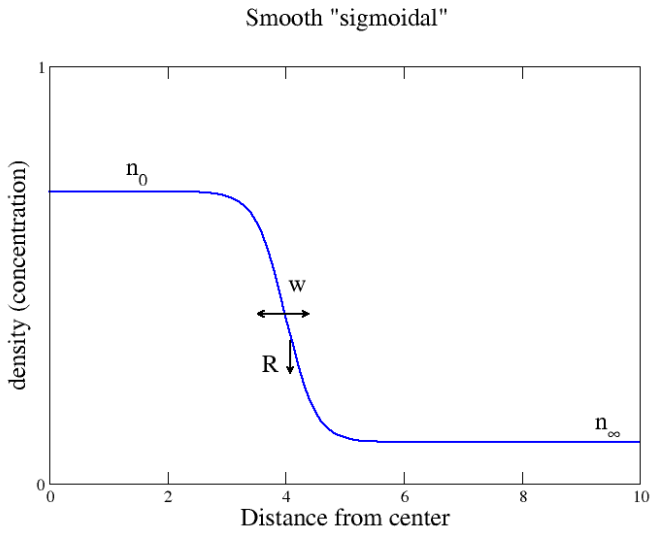
- One hypothesis: they are a seperate species: an oligomer, a misfolded protein, … (Pan, Vekilov and Lubchenko, J. Phys. Chem. B 114, 7620 (2010).)

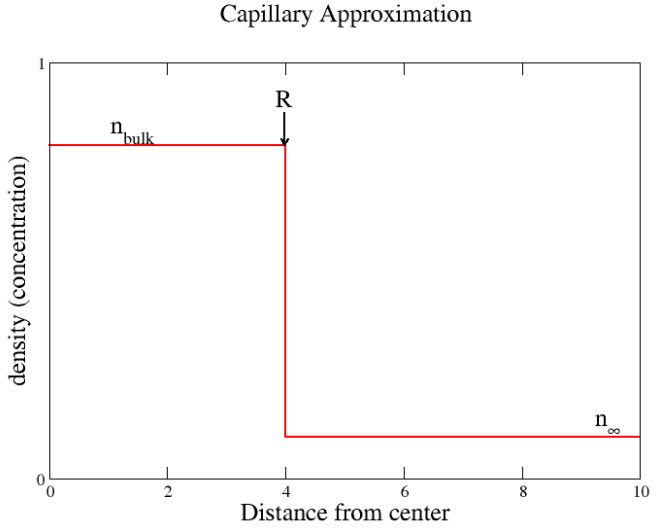
Different levels of description of droplets (with spherical symmetry)



Simulation data: P. ten Wolde and Daan Frenkel, J. Chem. Phys. 109, 9901 (1998).

Calculations: JFL, J. Chem. Phys. 129, 244501 (2008).





Part I: Capillary Model

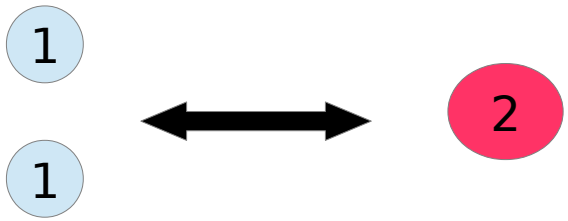
Two processes: super-critical growth and chemical reaction

1. Growth of a super-critical droplet (CNT <==> capillary approximation)

$$\frac{dR}{dt} = D n^{(\infty)} \frac{\beta P\left(n^{(0)}\right) - \beta P\left(n^{(\infty)}\right)}{\left(n^{(0)} - n^{(\infty)}\right)^2} R^{-1} \Rightarrow R \propto \sqrt{t}$$

(JFL J. Chem. Phys. 136:034509, 2012)

2. Chemical reaction between species



$$m_2 = 2 m_1$$

$$R_2 = 1.26 \times R_1$$

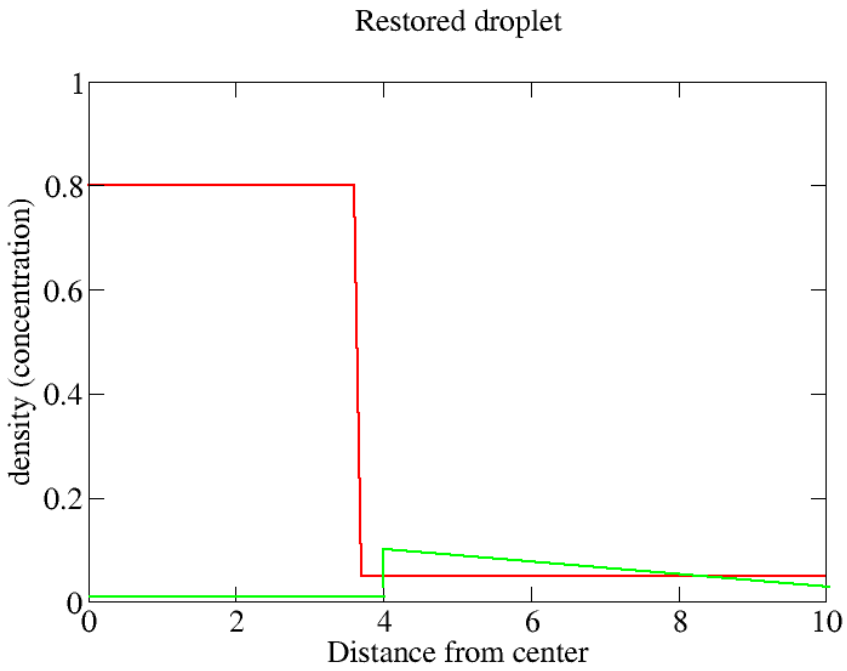
$$\frac{d n_1}{dt} = 2 k_{21} n_2 - 2 k_{12} n_1^2$$

$$\left(n_1 + 2 n_2 = \text{constant}\right)$$

$$\frac{d n_2}{dt} = k_{12} n_1^2 - k_{21} n_2$$

Reaction+excluded volume+fast diffusion = extinction

Reaction converts dimers to monomers



Excluded volume effects (e.g. entropy) forces expulsion of monomers: radius decreases so density remains at optimal value.

Within droplet, net effect is extinction reaction.

$$\frac{dN}{dt} = -k_{21} N$$

$$\frac{dR}{dt} = -k_{21} R/3$$

Stabilization of post-critical droplets

Combining growth and extinction reaction happen simultaneously, we get

$$\frac{dR}{dt} = D n_2^{(\infty)} \frac{\beta P\left(n_2^{(0)}\right) - \beta P\left(n_2^{(\infty)}\right)}{\left(n_2^{(0)} - n_2^{(\infty)}\right)^2} R^{-1} - k_{21} R/3$$

There is a stationary radius

$$R_{stable} = \sqrt{3 D n_2^{(\infty)} \frac{\beta P\left(n_2^{(0)}\right) - \beta P\left(n_2^{(\infty)}\right)}{k_{21}\left(n_2^{(0)} - n_2^{(\infty)}\right)^2}}$$

- For small clusters*, the 1/R term dominates giving growth
- For large clusters, the second term dominates causing dissolution

Stability

* “small” must still be larger than the critical cluster for nucleation

Part 2: Microscopic Model

(Dynamic Density Functional Theory)

Microscopic (DDFT) description of supercritical droplet

$$\frac{\partial n(r,t)}{\partial t} = D\nabla n(r,t)\nabla\frac{\delta F[n]}{\delta n(r,t)}$$

(Dynamic Density Functional Theory: From Mesoscopic nucleation theory: JFL J. Chem. Phys. 135:161101, 2011; 136:034509, 2012)

$$F[n] = \int \{\underbrace{f(n(r,t))}_{\text{volumetric}} + \underbrace{\frac{1}{2}K(\nabla n)^2}_{\text{surface tension}}\} d\boldsymbol{r}$$

(Squared-gradient approximation)

$$f(n) = \underbrace{n\ln(n) - n}_{\text{ideal gas}} + f^{(ex)}(n)$$

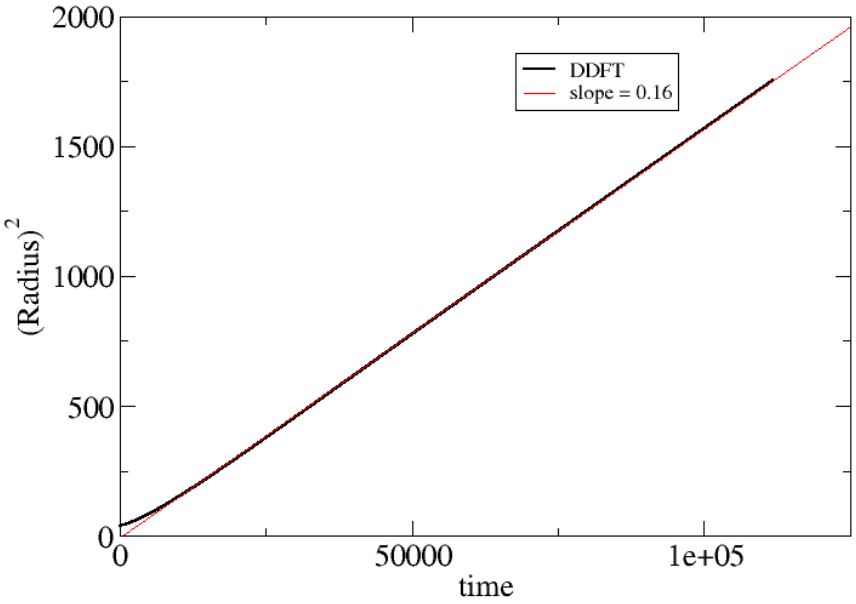
(Bulk Helmholtz free energy)

Note: $$n \to 0 \Rightarrow f^{(ex)} \to 0 \Rightarrow \frac{\partial n(r,t)}{\partial t} = D\nabla^2 n(r,t)$$

This can therefore be viewed as a generalization of the diffusion equation with a density-dependent diffusion constant.

DDFT description of supercritical droplet: Comparison to capillary model

- Bulk Helmholtz free energy: for Lennard-Jones
- Squared-gradient coefficient: calculated from LJ potential



$$\frac{dR^2}{dt} = 2Dn_2^{(\infty)} \frac{\beta P\left(n_2^{(0)}\right) - \beta P\left(n_2^{(\infty)}\right)}{\left(n_2^{(0)} - n_2^{(\infty)}\right)^2}$$

Capillary slope = 0.017

Observed slope = 0.016

DDFT description of reacting mixture

$$\frac{\partial n_1(r,t)}{\partial t} = D_1 \nabla n_1(r,t) \nabla \frac{\delta F[n_1,n_2]}{\delta n_1(r,t)} + 2k_{21} n_2(r,t) - 2k_{12} n_1^2(r,t)$$

$$\frac{\partial n_2(r,t)}{\partial t} = D_2 \nabla n_2(r,t) \nabla \frac{\delta F[n_1,n_2]}{\delta n_2(r,t)} - k_{21} n_2(r,t) + k_{12} n_1^2(r,t)$$

Model

$$F[n_1,n_2] = \int \{ f_1(n_1) + \frac{1}{2} K_1 (\nabla n_1)^2 \} d\mathbf{r}$$

$$+ \int \{ f_2(n_2) + \frac{1}{2} K_2 (\nabla n_2)^2 \} d\mathbf{r}$$

$$+ \int \{ f_3^{(ex)}(n_1,n_2) + \frac{1}{2} K_3 (\nabla n_1) \cdot (\nabla n_2) \} d\mathbf{r}$$

Assumptions

$$D_1 = D_2$$

$$f_1(n_1) = f_{hs}(n_1;d)$$

$$f_2(n_2) = f_{LJ}(n_2)$$

$$K_1 = K_2 = K_{LJ}$$

$$K_3 = 0$$

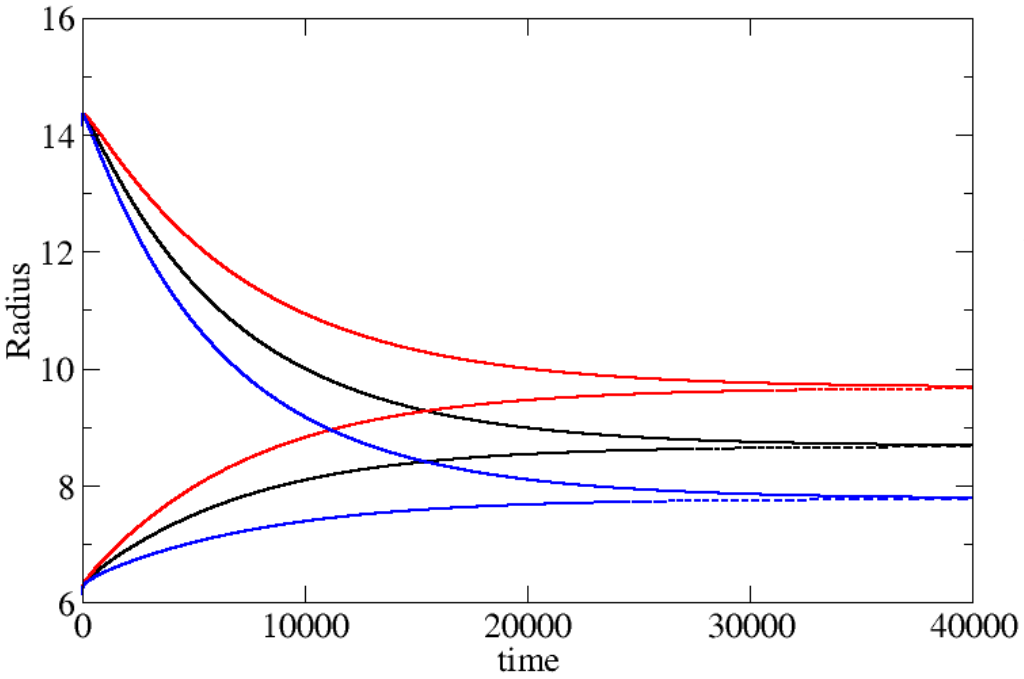
$$f_3^{(ex)}(n_1,n_2) = f_{hs}^{(ex)}(n_1+n_2;d) - f_{hs}^{(ex)}(n_1;d) - f_{hs}^{(ex)}(n_2;d)$$

(Excluded volume effect)

Post-critical droplets are stable

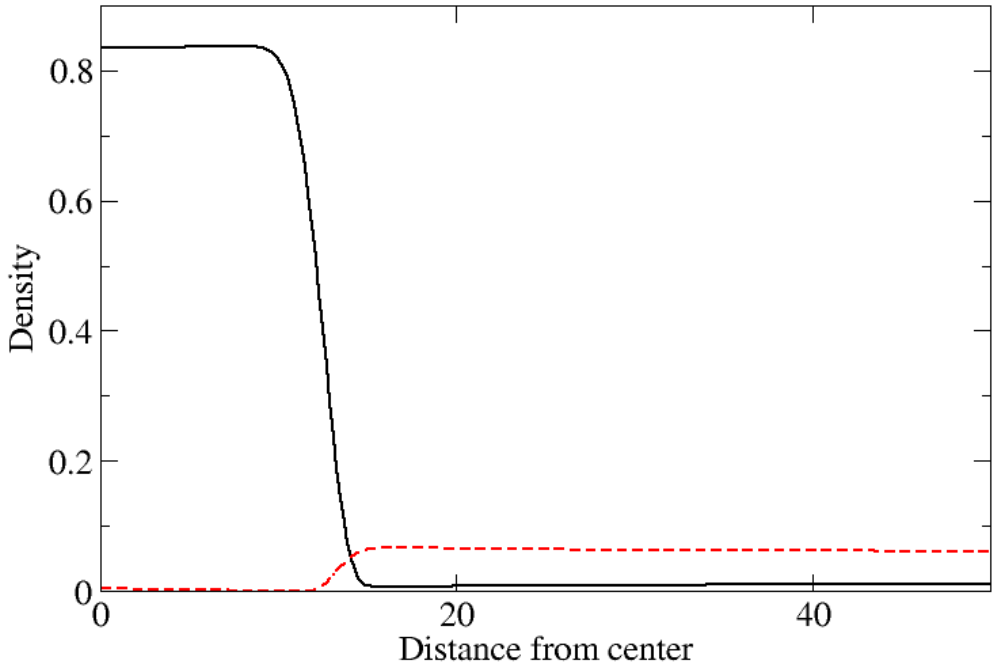
Protocol:

- Find critical cluster for pure dimer system
- Increase critical cluster radius by ΔR
- Add monomer
- Solve DDFT equations numerically
- Repeat for three different values of k_{21}
  (same equil. Monomer and dimer densities)
- For each k, do for large and small initial displacements



Convergence to same result independently of initial condition ==> empirical proof of stability.

Stable Droplet structure



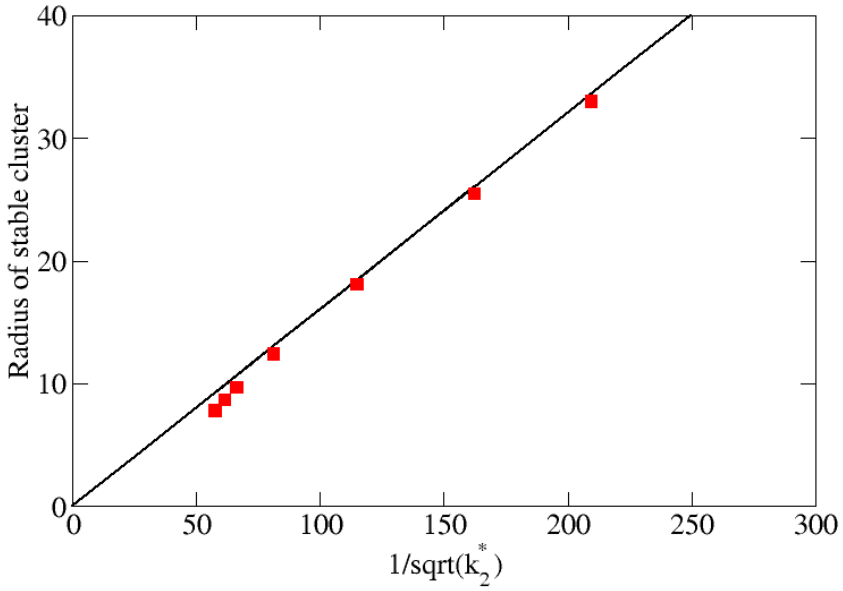
Confirmation that monomer is expelled

==> effective extinction reaction within cluster.

Scaling with reaction rate

$$R_{stable} = \sqrt{3 D n_2^{(\infty)} \frac{\beta P\left(n_2^{(0)}\right) - \beta P\left(n_2^{(\infty)}\right)}{k_{21}\left(n_2^{(0)} - n_2^{(\infty)}\right)^2}}$$

(capillary model)



Summary

- Capillary model and numerical cDFT show that droplets can be stable
- This supports the hypothesis that the nucleation precursors are a “complex” of protein molecules – problem solved!

JFL and Grégoire Nicolis, "Mechanism for the stabilization of protein clusters above the solubility curve", Soft Matter, 12, 93 (2016)

However, if the reaction rates are allowed to depend on the local concentrations, then everything changes:

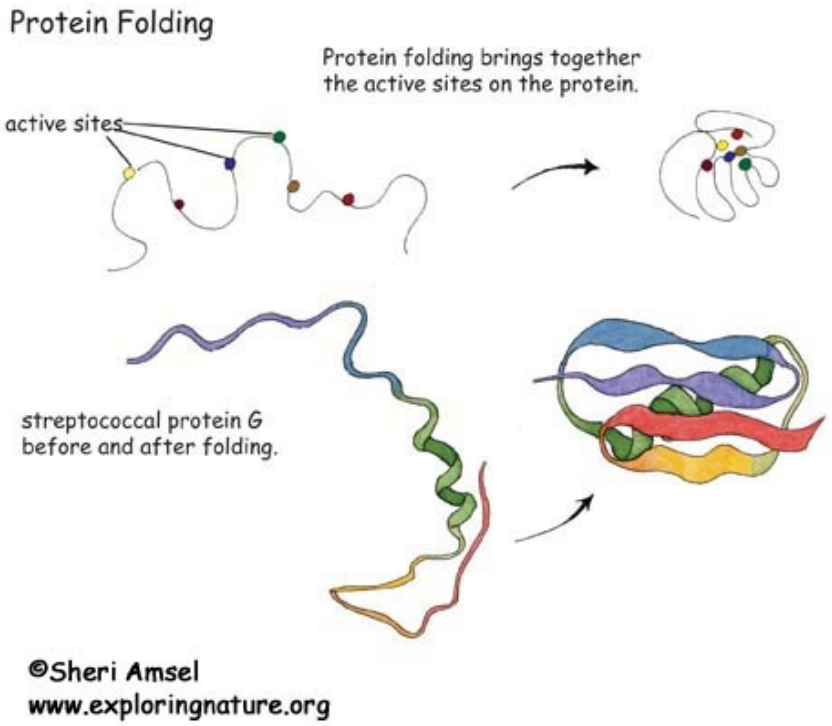
- Dimers in a condensed droplet do not “want” to become monomers: reaction is suppressed.
- This disrupts the stabilization mechanism – all stability is lost.
- The problem remains open ...

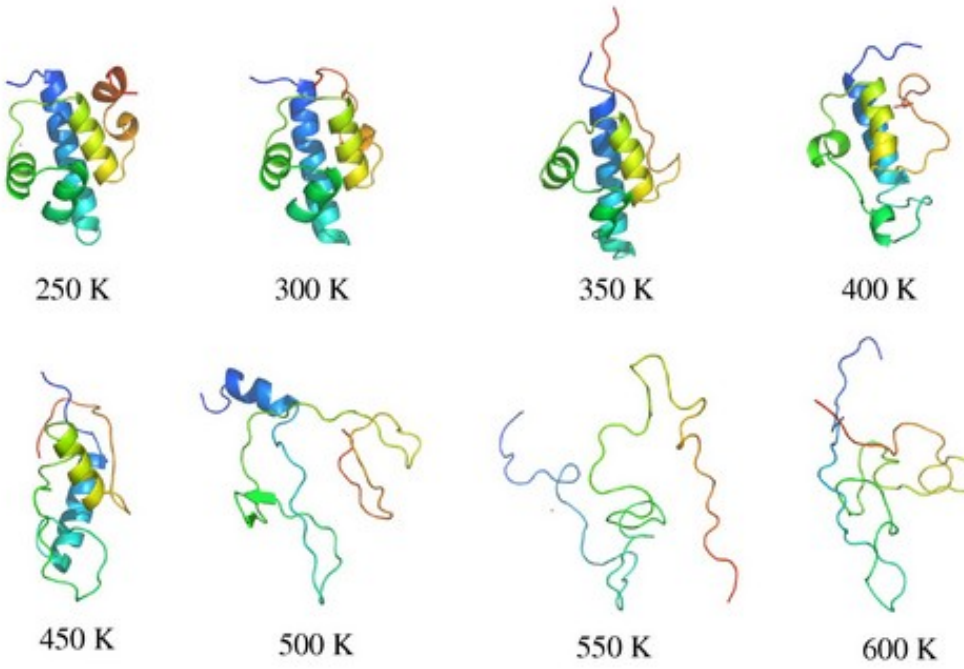
JFL , "Mechanism for the stabilization of protein clusters above the solubility curve: The role of non-ideal chemical reactions", J. Phys. Cond. Matt., 28, 244020 (2016)

Les Protéines

- Introduction
- Protéine interactions
- Condensation
- Pliage

Protéines Pliage





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

Protéines Pliage

REVIEW

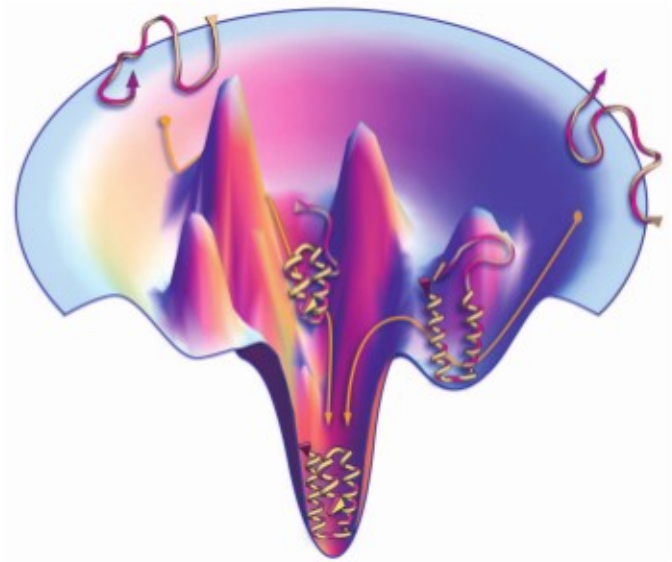


Fig. 3. Proteins have a funnel-shaped energy landscape with many high-energy, unfolded structures and only a few low-energy, folded structures. Folding occurs via alternative microscopic trajectories.

"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 myoglobin.

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'aprés 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?

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

Protéines Pliage

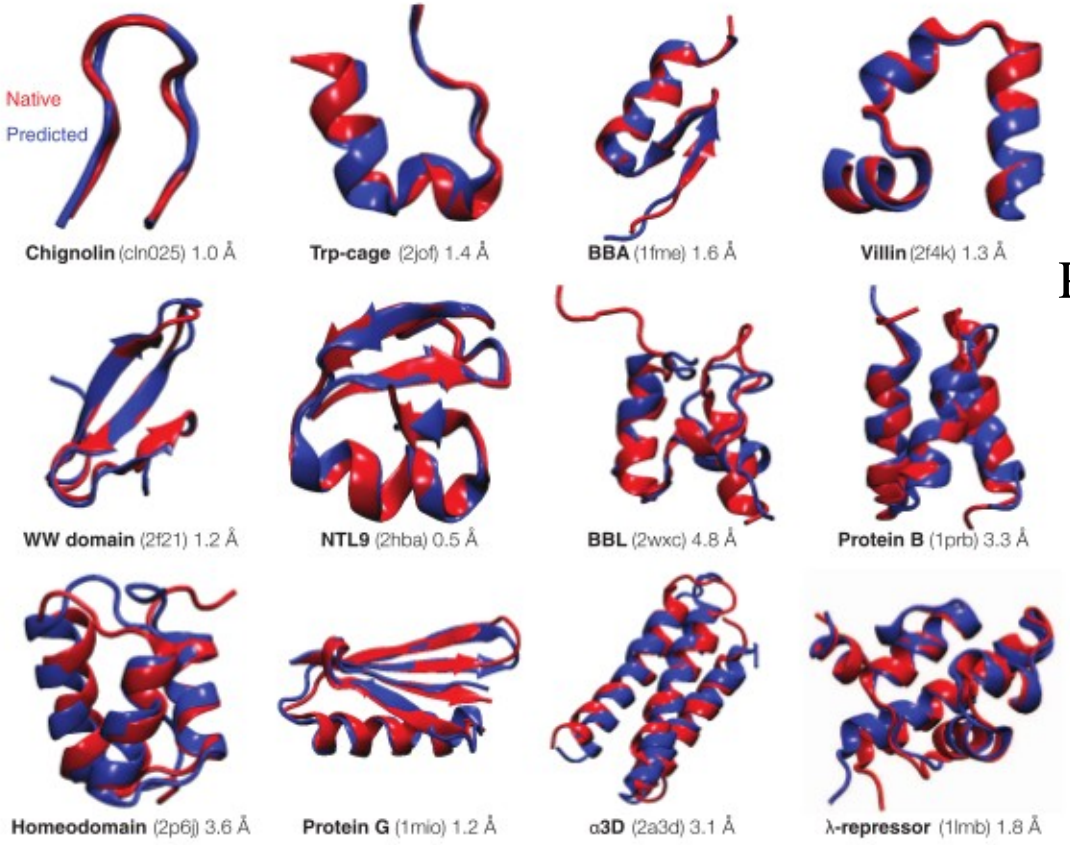
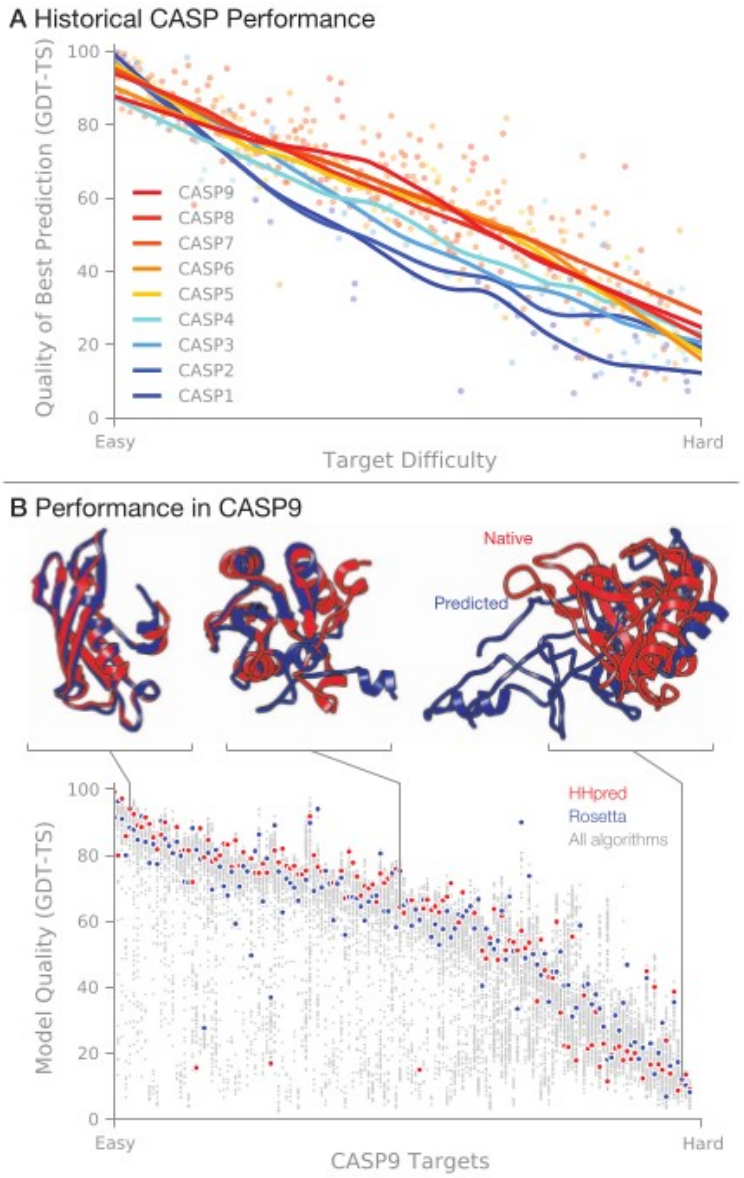


Fig. 2. Modern physical models can compute the folded structures of some small proteins. Using a high-performance custom computer called Anton (*48*), Shaw and co-workers observed reversible folding and unfolding in more than 400 events across 12 small proteins to structures within 4.5 Å of the experimental 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*)]

Pour quelques petites protéines: oui...

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

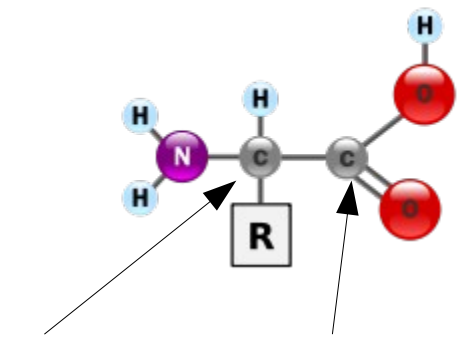
Protéines Pliage



… mais pas pour toutes.

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.

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

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

Over-damped stochastic dynamics:

$$\frac{d}{dt}\boldsymbol{q}_i = -D_{ij}\frac{\partial U(\boldsymbol{q}^N)}{\partial \boldsymbol{q}_j} + Q_{ij}\eta_j(t), \quad \langle \eta_i(t)\eta_j(t')\rangle = \delta_{ij}\delta(t-t')$$

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

Noise probabilities:

$$P(\eta_i(t)=\bar{\eta}_i) \equiv P(\bar{\eta}_t) = \pi^{-1/2} e^{-\frac{\bar{\eta}^2}{2}}$$

$$P(\eta_i(t_1)=\bar{\eta}_i(t_1), \eta_i(t_2)=\bar{\eta}_i(t_2)) = \pi^{-2/2} e^{-\frac{\bar{\eta}^2(t_1)+\bar{\eta}^2(t_2)}{2}}$$

Path Probabilities:

$$P(\boldsymbol{q}(t);\dot{\boldsymbol{q}}(t)) \sim \pi^{-1/2}\exp\left[-\frac{1}{2}\left(\dot{\boldsymbol{q}}_i(t)-b_i(\boldsymbol{q})\right)Q_{ij}^{-2}\left(\dot{\boldsymbol{q}}_j(t)-b_j(\boldsymbol{q})\right)\right]$$

$$P(\boldsymbol{q}_0, t=0; \boldsymbol{q}_f, t=T) = \int D\boldsymbol{q}(\tau)\exp\left[-\frac{1}{2}\int_0^T dt\left(\frac{1}{2}\left(\dot{\boldsymbol{q}}_i(t)-b_i(\boldsymbol{q})\right)Q_{ij}^{-2}\left(\dot{\boldsymbol{q}}_j(t)-b_j(\boldsymbol{q})\right)+\frac{\partial b_i}{\partial q_i}\right)\right]$$

Protéines Pliage: les modeles stochastique

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

$$P(\boldsymbol{q}_0, t=0; \boldsymbol{q}_f, t=T) = \int D\boldsymbol{q}(\tau)\exp\left[-S_{eff}\right]$$

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

Most likely path:

$$\frac{\delta S_{eff}}{q_i(t)} = 0$$

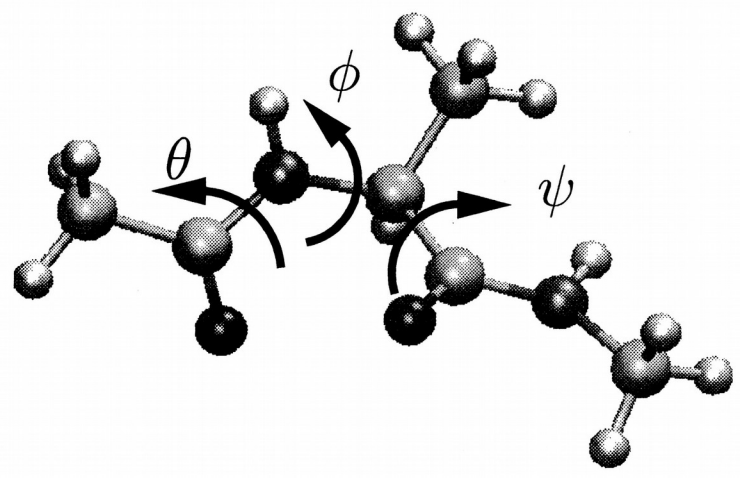
$$\frac{d}{dt}\frac{\delta L}{\delta \dot{q}_i(t)} - \frac{\delta L}{\delta \dot{q}_i(t)} = 0$$

Protéines Pliage: les modeles stochastique

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

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

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



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).

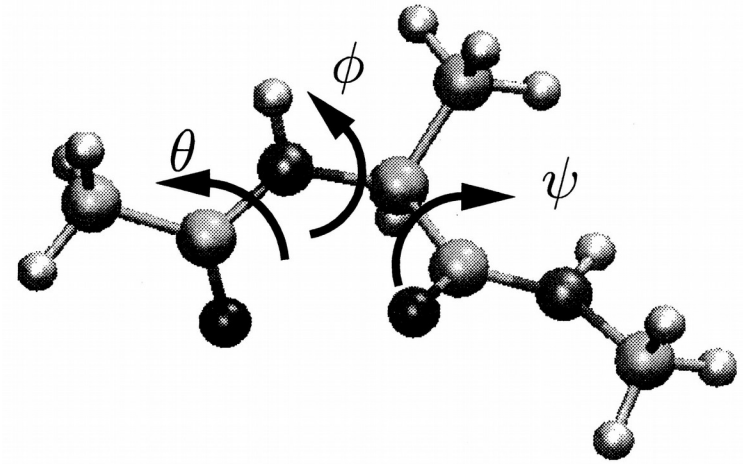
The molecule in vacuum has two stable conformers: the C_{7eq} state with $\varphi \approx -86°$ and $\psi \approx 68°$ and the C_{ax} state with $\varphi \approx 50°$ and $\psi \approx -50°$. In solution, these positions shift slightly, for instance, the C_{7eq} state is located around $\varphi \approx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

Protéines Pliage: les modeles stochastique

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

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

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



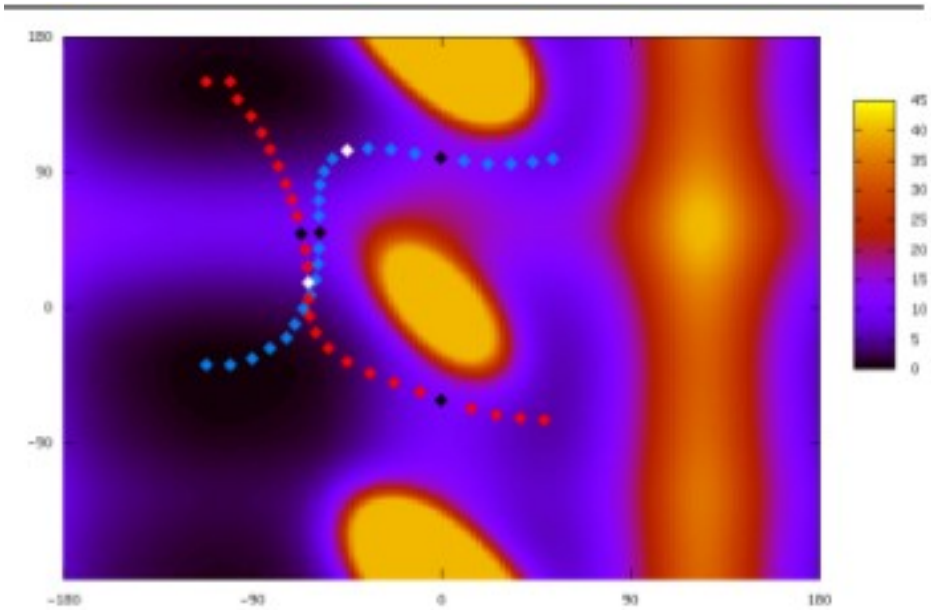


FIG. 1 (color online). Dominant Folding Paths for the $C7_{ax} \rightarrow C7_{eq}$ (red squares, from bottom right to top left) and $\alpha_L \rightarrow \alpha_R$ (blue squares, from bottom left to top right) transitions. In the background, the free-energy profile for the ψ and ϕ dihedrals is shown (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)