



Protein Folding & Thermodynamics

Structural Bioinformatics

Protein Folding & Dynamics

9 feb 2017

Protein Folding & Dynamics

- Introduction to folding
- Dynamics
- Reading:
 - Chapter 12: Introduction to Protein Folding and Simulation
- Background reading:
 - Folding and flexibility (Branden & Tooze Chapter 6)

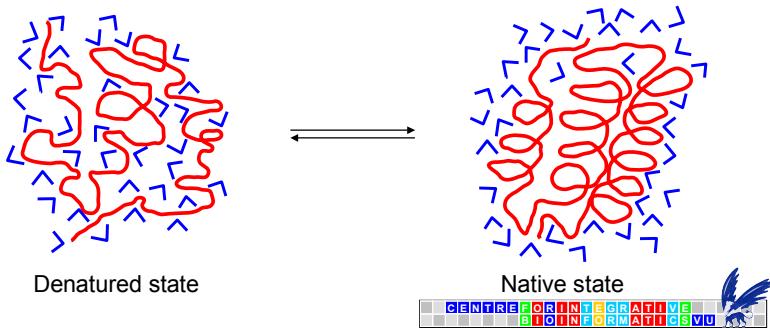


[2] 9 feb 2017 Protein Folding & Dynamics

This lecture will give a short introduction on the physical process by which chains of amino acids adopt a three-dimensional structure that we call a ‘protein’: the protein folding process. Also, we will touch on the related topic of protein (native) dynamics, i.e. the flexibility and movability that a protein structure has in its native state. Finally, we will start the introduction of thermodynamics as it relates to dynamics at the molecular scale.

Active protein conformation

- Active conformation of protein is the native state
- unfolded, denatured state
 - high temperature
 - high pressure
 - high concentrations 'denaturant' (e.g., urea 8 M)
- Equilibrium between two forms



[3] 9 feb 2017 Protein Folding & Dynamics



We know that the functional, active form (or conformation) of a protein is called the 'native state'. There are different ways to denature (or unfold) proteins, e.g. by high temperature, high pressure, or by chemicals that (somehow) distort the stabilizing effects that keep the protein in its native state.

It is important to realize that under all conditions, a **dynamic equilibrium** exists between the protein in its native conformation and the 'denatured state'. Denaturation simply has the effect of **shifting** the balance, from mostly native under stabilizing conditions. To mostly denatured under unfolding conditions.

Anfinsen's Theorem (1950's)

- Primary structure determines tertiary structure:

"In the mid 1950's Anfinsen began to concentrate on the problem of the **relationship between structure and function** in enzymes. [...] He proposed that the **information determining the tertiary structure** of a protein resides in the **chemistry of its amino acid sequence**. [...] It was demonstrated that, after cleavage of disulfide bonds and disruption of tertiary structure, many proteins could **spontaneously refold** to their native forms. This work resulted in general acceptance of the '**thermodynamic hypothesis**'."

– Nobel Prize Chemistry 1972 –

www.nobel.se/chemistry/laurates/1972/anfinsen-bio.html



[4] 9 feb 2017 Protein Folding & Dynamics

Anfinsen's work helped found of what we now call the central dogma of molecular biology: DNA makes RNA makes protein. Other than that, there is (assumed to be) nothing else that is needed to obtain a functional, folded, protein. His work showed that, at least in principle, the 'translation' from protein chain to functional protein structure required no external input.

Anfinsen: Reversible Folding

- Anfinsen performed un-folding/re-folding experiments!
- Molecular basis for the central dogma of molecular biology:

DNA makes RNA makes Protein

– and –

Sequence → Structure → Function

- Without reversible folding, ‘something’ could interfere in the step from sequence to structure (from DNA/RNA to Protein)



[5] 9 feb 2017 *Protein Folding & Dynamics*

Levinthal's paradox (1969)

- Denatured protein re-folds in ~ 0.1 - 1000 seconds
- Protein with e.g. **100** amino acids each with **2** torsions (ϕ and ψ)
Each can assume **3** conformations (1 trans, 2 gauche)
 $3^{100 \times 2} \approx 10^{95}$ possible conformations!
- Or:
100 amino acids with **3** possibilities in Ramachandran plot
(α , β , L): $3^{100} \approx 5 \cdot 10^{47}$ conformations
- If the protein can visit one conformation in one ps (10^{-12} s)
exhaustive search costs $5 \cdot 10^{47} \times 10^{-12}$ s = $5 \cdot 10^{35}$ s $\approx \mathbf{1.6 \cdot 10^{28}}$ years!
(the lifetime of the **universe** $\approx \mathbf{10^{10}}$ years...)

[6] 9 feb 2017 *Protein Folding & Dynamics*

C. Levinthal, J. Chim. Phys. Phys.-Chim. Biol., 1968, 65, 44–45



But a paradox remained regarding the folding of proteins. We know experimentally folding takes place within seconds, and occasionally minutes or hours. This hold for *in vivo* folding ‘at the ribosome’ as well as for *in vitro* folding experiments in the lab.

Given the vast number of possible ways in which a polypeptide chain could potentially fold into a protein structure, a folding protein could not possibly search through all that conformational space, even during the lifetime of the universe!

Levinthal's paradox



Protein folding problem:

- Predict the 3D structure from sequence (bioinf.)
- Understand the folding process (biophysics)

[7] 9 feb 2017 Protein Folding & Dynamics

C. Levinthal, J. Chim. Phys. Phys.-Chim. Biol., 1968, 65, 44–45

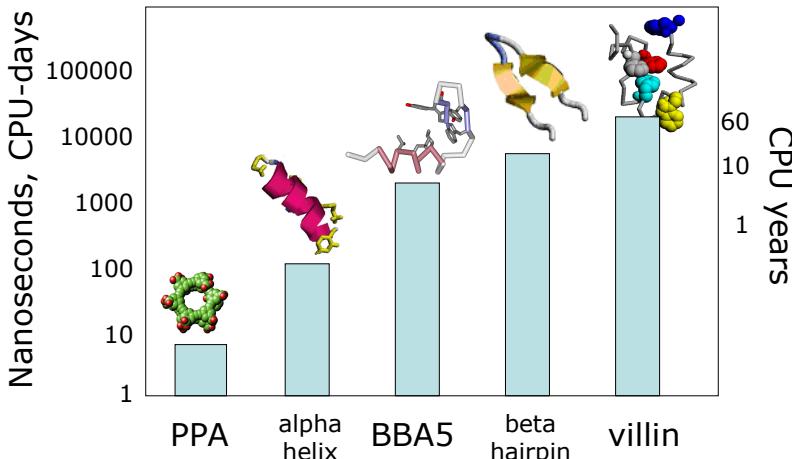


This is known as Levinthal's paradox.

Nowadays, it is widely accepted that the answer to this paradox is that, clearly, the protein does not have to search through this vast conformational space. Apparently at some fundamental level the problem is simpler than it appears to us now.

One way of getting some insight into that is through computer simulations of the dynamics of proteins (more about that in later lectures).

What to fold? ... fastest folders



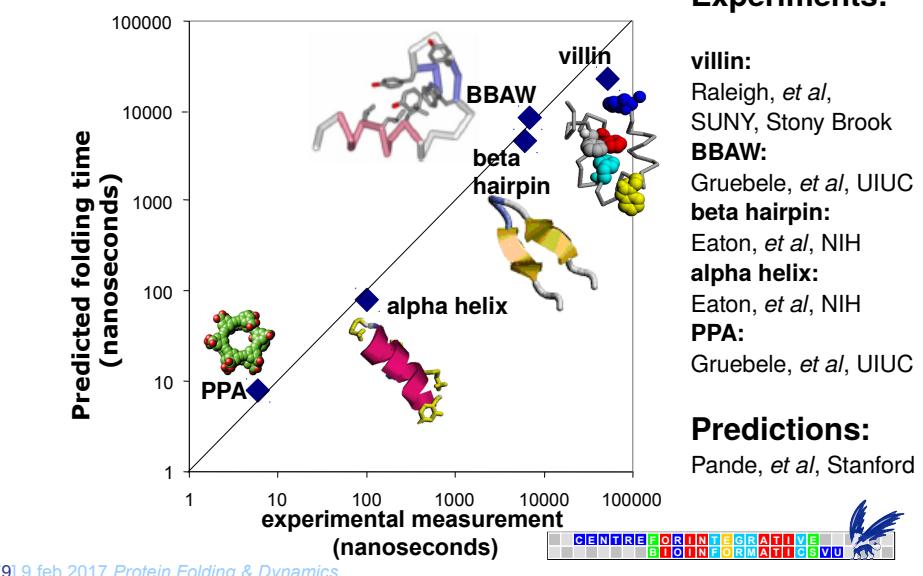
Pande et al. "Atomistic Protein Folding Simulations on the Submillisecond Time Scale
Using Worldwide Distributed Computing" *Biopolymers* (2003) **68** 91–109

[8] 9 feb 2017 Protein Folding & Dynamics

Given the limitations of these simulations (on the order of nanoseconds of protein motion per day of computation), people started with simple polypeptides that were (experimentally) known to fold very very fast (less than milliseconds).

More recently, as microseconds and even miliseconds are coming within reach for selected large-scale applications, this has been extended to larger proteins and/or slower folding proteins [references needed]. One example is the Shaw et al, 2010, paper discussed later in this lecture, and again in later lectures.

Rates: predicted vs experiment

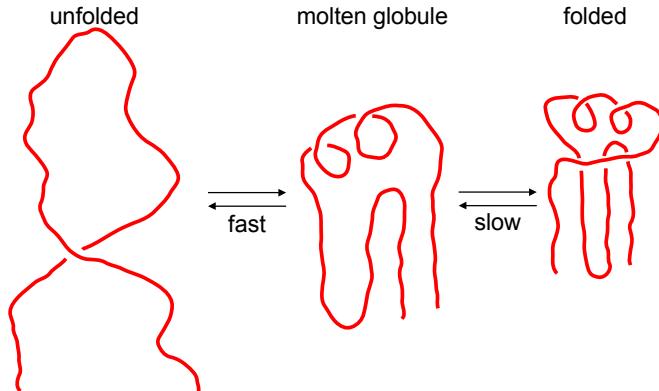


Surprisingly, the folding times observed from simulation matches very good to the experimentally known folding times for these molecules. There are concrete numbers in a later slide, but these folding times are far shorter than would naively be expected based on the size of the conformational space (Levinthal's paradox).

We will now explore some features of protein molecules that may be responsible for this observed fast folding.

Molten globule

- First step: hydrophobic collapse
- Molten globule: globular structure, not yet correct folded
- Local minimum on the free energy surface



[10] 9 feb 2017 Protein Folding & Dynamics



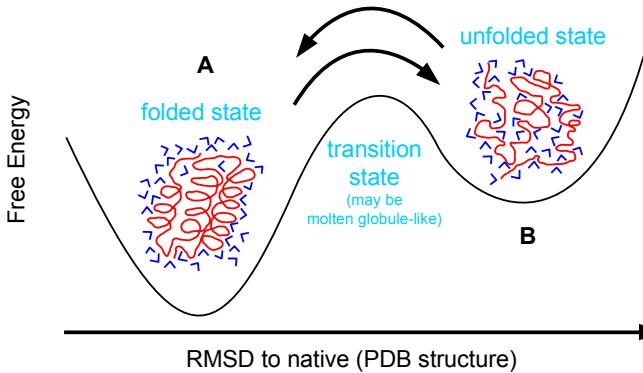
A common theme in protein folding is ‘hydrophobics inside’. Given that this hydrophobic effect is a major contributor to protein stability, it may well be an important aspect of the process of folding.

The idea is as follows. In an unfolded protein chain, many hydrophobic sidechains will be in contact with water. This is highly unfavourable, thus a first, and fast, step will be the aggregation of many of these sidechains; this is called the hydrophobic collapse and basically is the step that forms the hydrophobic protein interior.

The structure that is now created is characterized by a much smaller (average) radius than the unfolded state, however, not all native contacts have yet been formed. In particular, β -strands of high contact order will not have formed, while α -helices and low contact order β -sheets (like β -hairpins) may already be present. This state is called the ‘molten globule’ to indicate that the folding chain has achieved compactness, but the interior has not settled into a stable structure and may therefore be thought of as being liquid-like.

Protein folding as a “two-state process”

- From the folded state, some protein molecules can (and will) unfold
- From the unfolded state, some protein molecules (re-)fold (continually)
- ⇒ Dynamic equilibrium: A (folded) \leftrightarrow B (unfolded)



[11] 9 feb 2017 Protein Folding & Dynamics



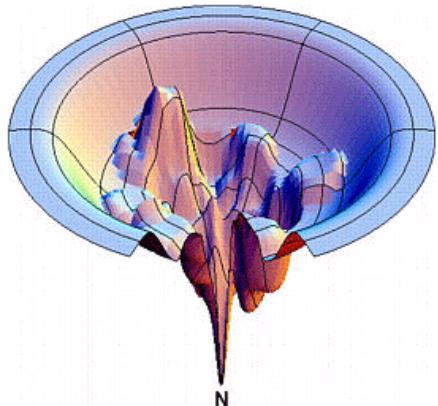
The simplest way to think about protein folding is as a transition between two more or less well defined states: the folded (native, functional) state, and the unfolded (denatured, non-functional). Sketching this in a diagram of free energy (measuring stability – lower is more stable) versus distance to native (here, RMSD to the crystal structure), yields two ‘vallies’ corresponding to those states.

Protein molecules are not fixed in one of the states; instead, a folded protein molecule will, after some time, spontaneously un-fold, and usually after a short amount of time, again re-fold. This is called a dynamic equilibrium. Moreover, here, the folded state has the lowest free energy, indicating that it is the most stable (note, that in different conditions, like high temperature, this may be different and the native state may not be the most stable with the lowest free energy). When free energy levels change, so will the stability of the states, and the equilibrium will shift along with that; e.g. as the folded state becomes less stable (with respect to the unfolded) when temperature is increased, we will find fewer molecules folded, and increasingly more unfolded.

In reality, things are often more complicated; both states are not so well defined, and may actually consist out of multiple (sub)states. We will go a bit further into this in the rest of this lecture, and in later lectures.

Folded state vs. Unfolded state

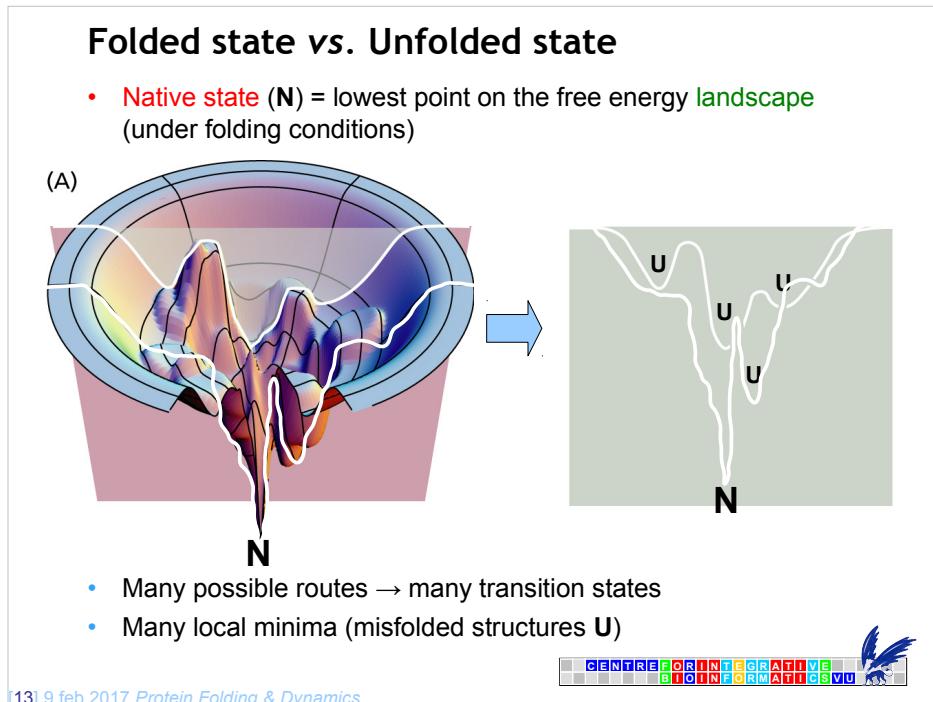
- Native state (**N**) = lowest point on the free energy **landscape** (under folding conditions)



12 | 9 feb 2017 Protein Folding & Dynamics



A related notion that visualizes the folding process is called the ‘folding funnel’, where the whole free energy landscape (more about that later) is depicted in two dimensions (horizontal) with the free energy being the third (vertical). This still is a strong simplification of the actual process! In the middle of this landscape is the lowest point, with lowest free energy and therefore (by definition) the most stable, native state. At the upper rim (and stretching outwards in all directions, which is not drawn here), at high free energies, is the unfolded state. Inbetween is a rugged landscape with hills and valleys, representing barriers that a folding protein molecule must cross, and trapped ‘local minimum’ states that can delay the folding process.

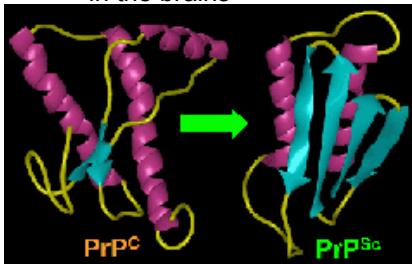


The folding funnel representation conveys the notion that, instead of a simple two-state model going from a single defined (unfolded) state to another defined (native) state, rather the folding process proceeds from a largely heterogeneous state (the whole of the rim ‘highlands’) through a multitude of possible pathways and intermediate states into the native state.

By the way, particularly low local minima, sometimes close to the native state, represent misfolded proteins. These are highly stable, almost like the native state, but not functional. In the worst case, they may be dysfunctional in that they impair the function of other proteins in the cell, or even the function of a cell as a whole.

Alternative folding: prions

- Prion proteins are found in the brains
- Function unknown
- Two forms
 - normal alpha-structure
 - harmful beta-structure
- beta-structure can aggregate and form 'plaques'
 - Blocks certain tissues and functions in the brains



[14] 9 feb 2017 Protein Folding & Dynamics



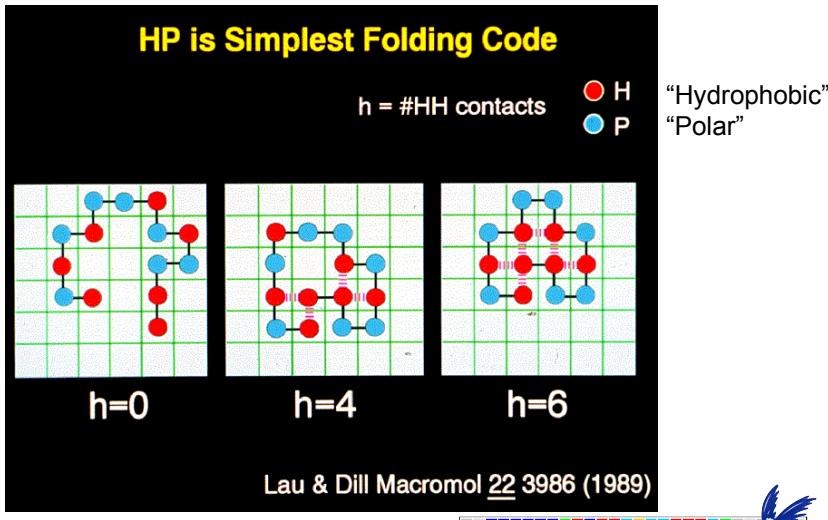
http://www.cnsi.ucla.edu/ar/paper?paper_id=191528



One particular such misfolded state is the β -fibril. Basically a structured aggregate of proteins in which the proteins themselves have changed conformation into a mostly β -sheet state, and these β -sheets have joined between protein molecules.

This process can be autocatalytic due to the fact that the unsatisfied ends of the β -fibril can induce a conformational change in the (still) native protein molecules. These will subsequently elongate the β -sheet in the β -fibril, producing another 'sticky' end.

Simple model for protein folding: ‘HP’



[15] 9 feb 2017 Protein Folding & Dynamics

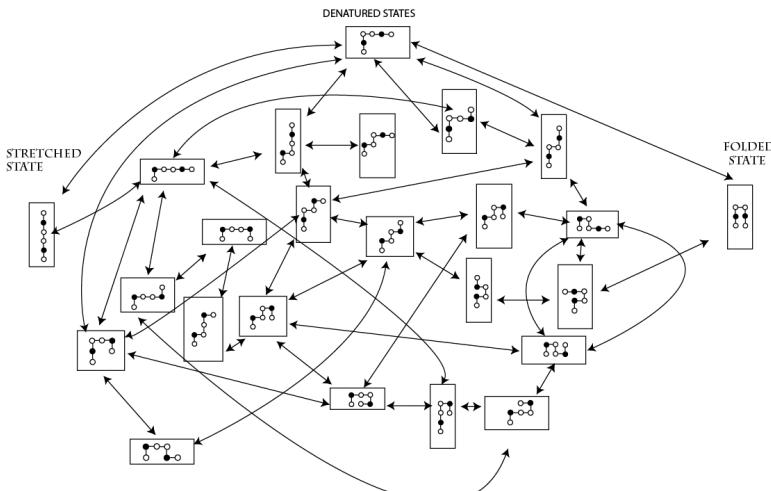


We now switch to an extremely simplified model that has been used since (at least) the early 1970’s to study some aspects of protein folding: the 2D lattice HP model.

Here, a ‘protein’ is represented as a string of H (hydrophobic) and P (polar) beads, which live on a rectangular grid. Being simple, we are now able to enumerate (all) possible conformations for a given sequence. Lau & Dill state that a 10-mer can obtain 2034 conformations. This, of course, rises exponentially with chain length.

For background reading, if you’re interested to see what such models could be used for, see for example “Phase diagram of a model protein derived of the conformations by exhaustive Enumeration” by Dinner, Sali, Karplus, and Shakhnovich, J. Chem. Phys. 101 (2), 15 July 1994 – although they used a slightly more complicated interaction potential. A 3D version of this type of model is used in the Van Dijk et al., 2015 paper.

Introducing Entropy



(Martin Gruebele – LatticeProtein.ppt)
 [16] 9 feb 2017 Protein Folding & Dynamics



Here you get an impression of several conformational states for a HPHHPH six-mer, and the conformational changes (transitions) that can occur between them. This gives you an impression of how complex a conformational space, and hence the folding landscape of a protein, can be.

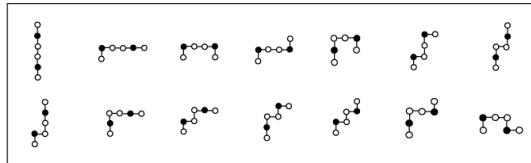
Each state has different number of ‘options’

“Energy” here is a function
of the number of contacts

ENERGY= - 0.0 kJ

14 Conformations

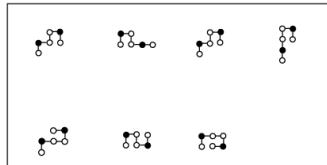
No contacts



ENERGY= - 0.25, =0.5 kJ

7 Conformations

1 P-P contact



Here, energy decreases and at the same time the number of ‘possible’ states decreases as well.
For real proteins,

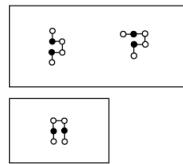
this is a lot more complicated.
Energy differences are small. There will be multiple low-energy states, but not all of them will have the same number of possible states. Moreover, water needs to be counted in as well!



ENERGY= - 1 kJ

2 Conformations

1 H-H contact



ENERGY= - 1.5 kJ

1 Conformation

1 P-P + 1 H-H

(Martin Gruebele – LatticeProtein.ppt)

[17] 9 feb 2017 Protein Folding & Dynamics

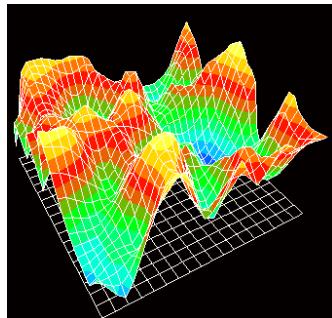
Here you can see that different states of folding (going from no interactions at the top, to the maximum of 2 at the bottom) result in different numbers of possible conformations. This is the basis for entropy: high entropy means large(r) conformational freedom.

Hence, the folded state can be said to be entropically challenged.

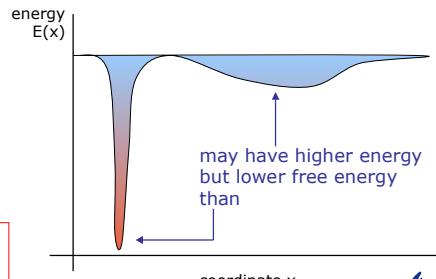
But we are now only looking at the protein. In a real system, this exists in a water(y) environment, and the water entropy turns out to be sensitive to the (folding) state of the protein. More on this in later lectures.

Folding energy

- Each protein conformation has a certain energy and a certain flexibility (entropy)
- Corresponds to a point on a **multidimensional** free energy surface



$$\Delta G = \Delta H - T\Delta S$$



Three coordinates per atom
3N-6 dimensions possible



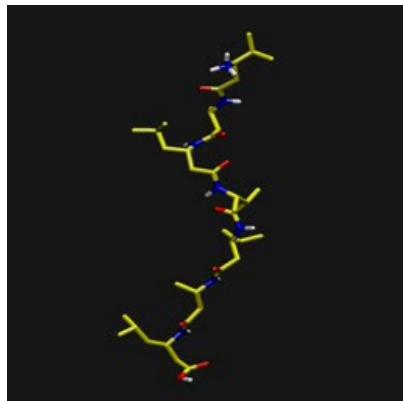
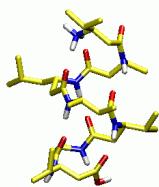
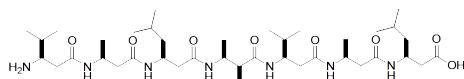
[18] 9 feb 2017 Protein Folding & Dynamics

One way to relate conformations with free energy is through the image of a ‘free energy landscape’ (the protein folding funnel shown before is another example of such a free energy landscape). Here, it is simplified to two (left) or one (right) dimension only. But realize that for a protein of N atoms, there are in principle 3N-6 dimensions (3 for x, y, z, and -6 because overall rotation and translation are irrelevant).

A small, 100 residue, protein has about 1000 atoms, so we are talking about a 3000-dimensional space (or energy landscape). In that space, each point is a particular conformation of all the atoms of the protein – shift even one atom, and you shift to the next point in that space, the so-called ‘conformational space’.

Peptide folding from simulation

- A small (β -)peptide forms helical structure according to NMR



- Computer simulations
of the atomic motions:
molecular dynamics

[19] 9 feb 2017 Protein Folding & Dynamics

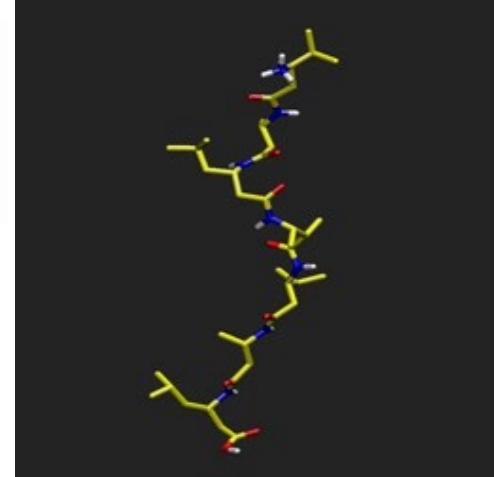
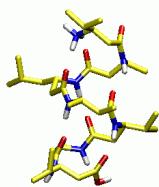
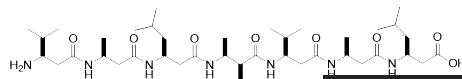


We will now dive back into protein folding. A particularly simple example is this peptide consisting of 7 amino acids. To be precise, they are β -amino acids, the main chain contains an extra carbon atom compared to naturally occurring peptides (natural amino acids are α -amino acids). This extra carbon atom is now called the alpha carbon, and the one the sidechain connects to is the beta carbon (this makes sense to chemists).

The main chain in this peptide therefore has three rotational angles per residue, or 21 in total for this 7-residue peptide. Assuming roughly three options per dihedral, the total conformational space would be $3^{21} \sim 10^{10}$ or ten billion.

Peptide folding from simulation

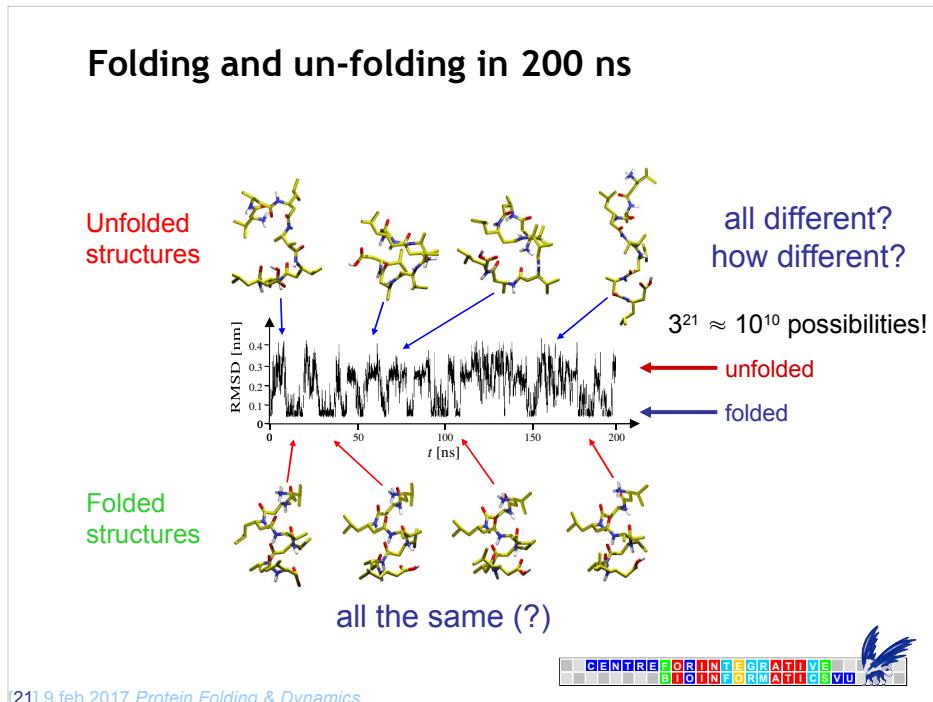
- A small (beta-)peptide forms helical structure according to NMR



- Computer simulations
of the atomic motions:
molecular dynamics

[20] 9 feb 2017 Protein Folding & Dynamics

From NMR we know this peptide is predominantly α -helical. The movie shows the dynamics of this peptide over about 100 ns. Among the conformations are curved, possibly helical-like, but clearly also many others.



A more concise overview of conformations encountered is given by calculating the RMSD with respect to the helical conformation (as determined by NMR). A low RMSD therefore means ‘helical’, a high RMSD means ‘something else’.

We can see that the peptide leaves its native conformation, and, very importantly, also finds it again! This happens several times over the course of 200 ns.

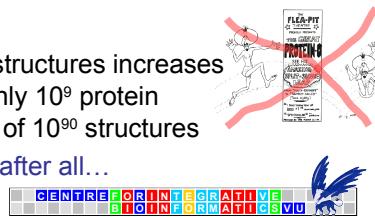
Importantly, this simulation of about 200 ns, never created 10 billion conformations. As a rough estimate, one ‘new’ conformation is visited every 1 ps, amounting to ‘only’ 200 million conformations.

Surprising result

- Number of **relevant** non-folded structures is very much smaller than the number of **possible** non-folded structures

	Number of aminoacids in protein chain	Folding time (exp/sim) (seconds)	Number possible structures	Number relevant (observed) structures
peptide	10	10^{-8}	$3^{20} \approx 10^9$	10^3
protein	100	10^{-2}	$3^{200} \approx 10^{90}$	10^9

- If the number of **relevant** non-folded structures increases proportionally with the **folding time**, only 10^9 protein structures need to be visited in stead of 10^{90} structures
- Folding-mechanism perhaps simpler after all...



22 | 9 feb 2017 Protein Folding & Dynamics

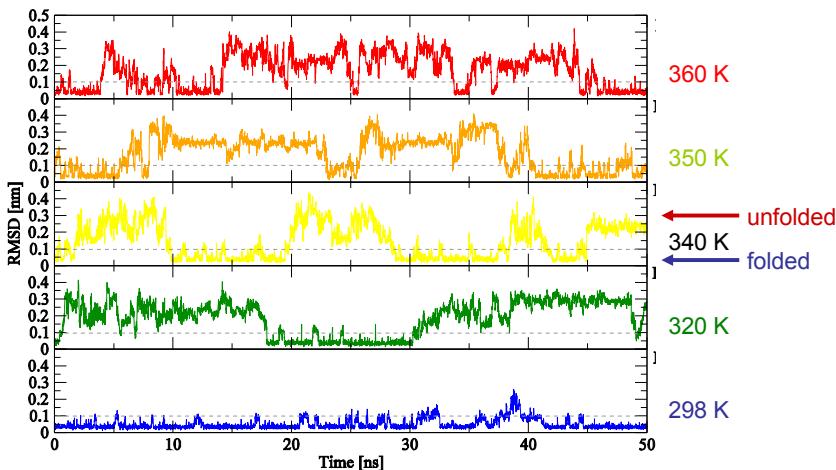


Even stronger, if we cluster the unfolded conformations with a distance criterion of 0.1 Ångstrom, we find about 1000 clusters, i.e. only 1000 clearly identifiably different conformations.

This peptide is estimated from the simulations to fold in about 10 ns. Extrapolating to a protein of 100 residues with a known (experimental) folding time of 10 ms (a million times slower), we may expect the protein to visit a million times more different conformations, i.e. about a billion.

This gives a concrete way out of Levinthal's paradox. A folding protein of 100 residues could potentially visit 10^{90} conformations but in reality only needs to visit a billion (10^9) to find the native state. If each takes about 1 ps (10^{-12} s), the protein could fold in 10^{-3} s = 1 ms. This is a realistic folding time for a small, 100-residue protein.

Temperature dependence



Daura, et al. 1999 Angew. Chem. Int. Ed., 38: 236–240
van Gunsteren, et al. 2001 Angew. Chem. Int. Ed., 40: 351–355.
23] Feb 29, 2016 Simulation and (Thermo)Dynamics

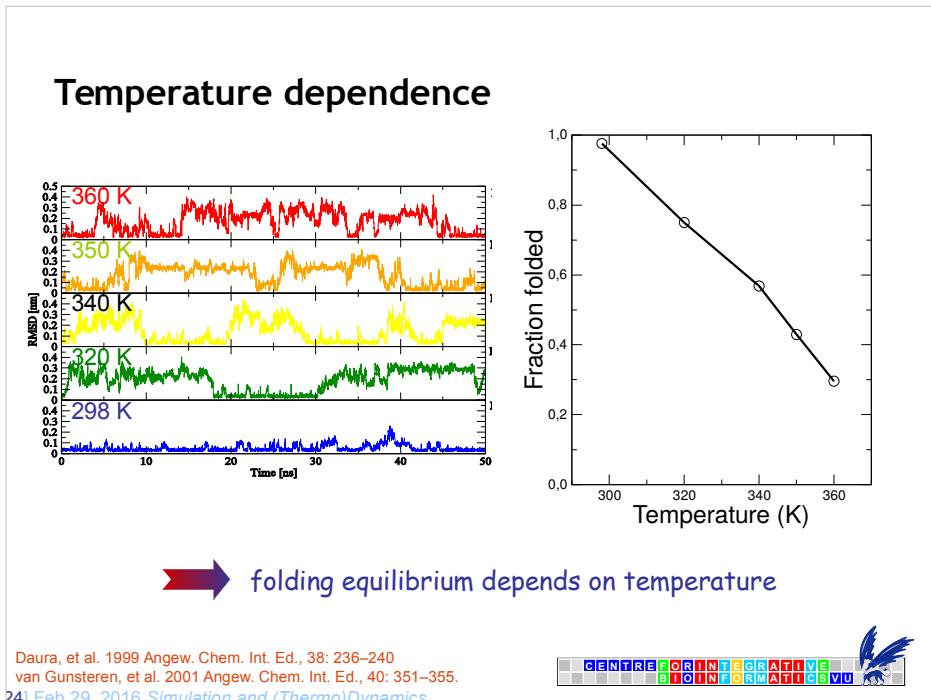


The equilibrium that we are sampling in these simulations is the folding-unfolding equilibrium for this peptide. Here, we show that the equilibrium is dependant on the temperature, as one would expect. The middle plot (at 340 K) is the one shown before, where on average the peptide is folded 50% of the time (and unfolded the other 50%).

At lower temperatures, we see a tendency of a larger fraction of time spent in the folded state.

At higher temperatures we see the opposite; a larger fraction of time spent in the *unfolded* state(s).

Note that the temperature changes are fairly small, only 10-20K up, and 20-40K down.



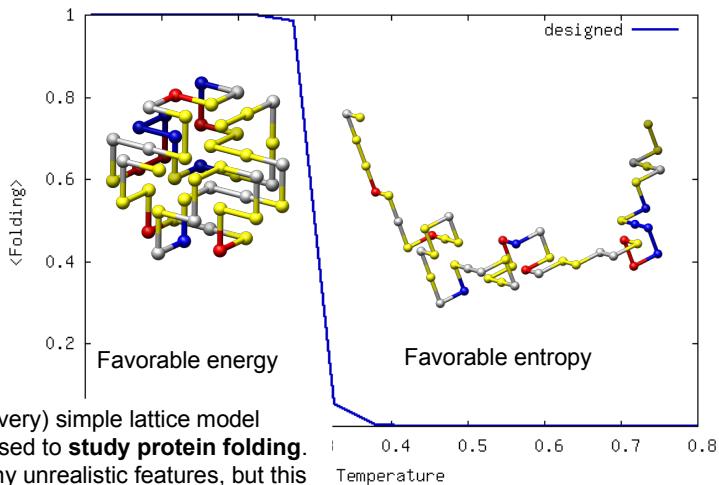
The equilibrium that we are sampling in these simulations is the folding-unfolding equilibrium for this peptide. Here, we show that the equilibrium is dependant on the temperature, as one would expect. The middle plot (at 340 K) is the one shown before, where on average the peptide is folded 50% of the time (and unfolded the other 50%).

At lower temperatures, we see a tendency of a larger fraction of time spent in the folded state.

At higher temperatures we see the opposite; a larger fraction of time spent in the *unfolded* state(s).

Note that the temperature changes are fairly small, only 10-20K up, and 20-40K down.

Temperature dependence for a ‘Protein’



This is a (very) simple lattice model which is used to **study protein folding**. It has many unrealistic features, but this temperature dependence is **realistic**, and is also observed for real proteins.

[25] 1 feb 2016 Introduction to Protein Structure



The graph shows an overview of the folding process as a function of temperature. All proteins denature (unfold) at higher temperature. (These results are from a simplified model of protein folding with reduced units for temperature, hence the scale from 0 to 0.8.) The vertical axis shows the extent of folding, 1 meaning folded and 0 unfolded (typically measured by the fraction of native contacts.)

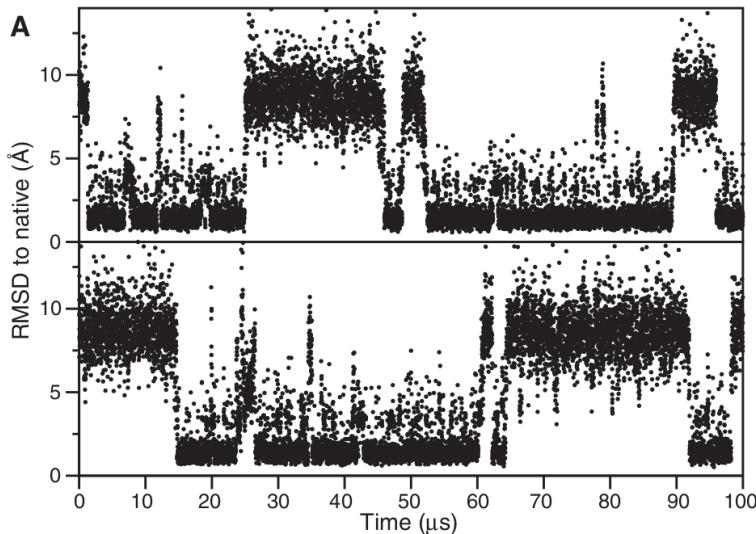
At low temperature, the native state is stable and therefore the protein is folded. This folding is driven by the energetically favourable conformation where hydrophobic residues (yellow) are ‘shielded’ in the interior of the protein structure.

At high temperature, entropic effects win out over the energetic effects. This makes the unfolded state more stable. The unfolded state, naturally, has a higher entropy than the folded state, but has to pay the energetic cost of exposing hydrophobic residues to the water.

Compared to the curve shown for a (small) peptide on the previous slide, we see that the protein here has a more ‘switching’ behaviour: nearly completely stable below some critical temperature, and nearly completely unstable above it.

Note that, even though this model is extremely simple, this temperature-dependent folding stability represents that of real proteins quite realistically. There is more about this in the Van Dijk et al, 2014 paper.

Reversible microsecond (small) protein folding

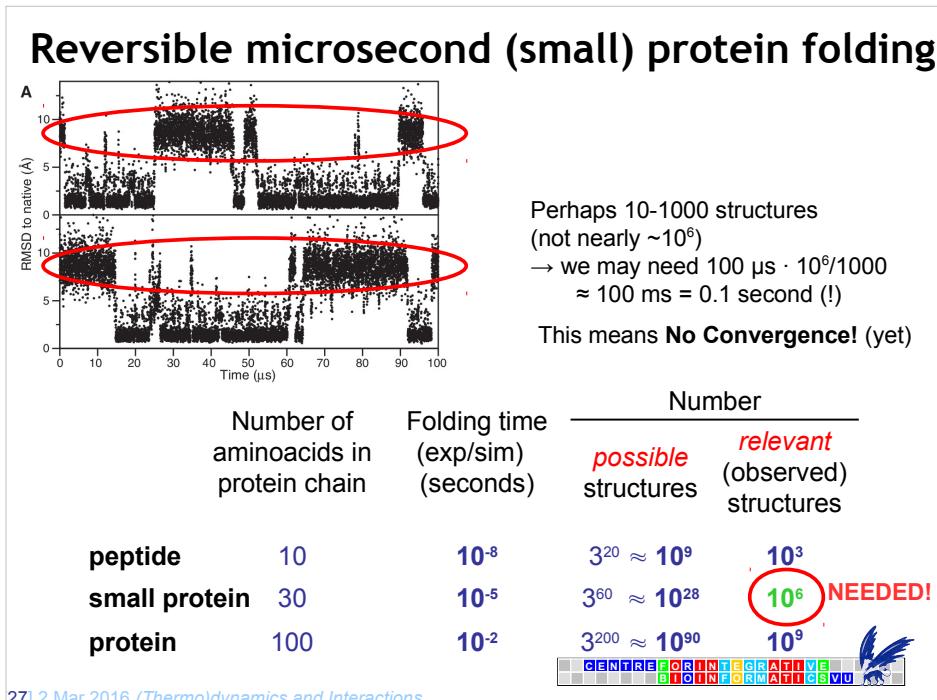


Shaw et al. Science. 2010 330:341-6. doi: 10.1126/science.1187409.

26 | 2 Mar 2016 (Thermo)dynamics and Interactions

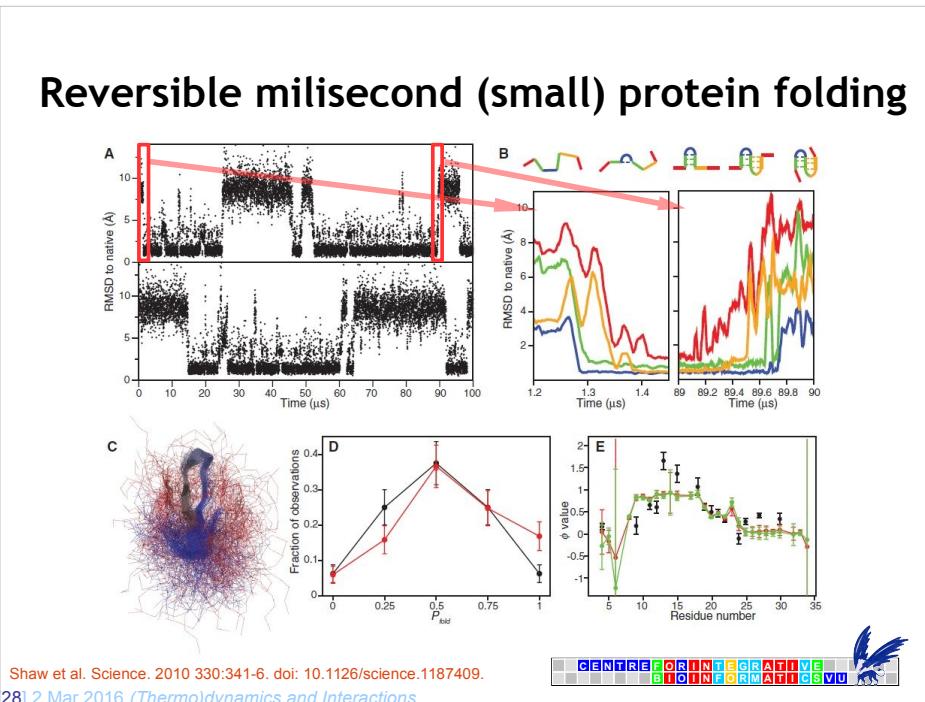


Although the simulations clearly show reversibility, only a relatively small number of unfolding and re-folding events are observed (about 10 or so; take a cut-off of 5 Å to discriminate folded from unfolded). From this, one may get a first rough estimate of the free energy of unfolding (at the temperature of these two simulations). But it is far too little to accurately estimate the probabilities of states, or to start drawing conclusions about properties of the ensemble of unfolded states.



If we now take again a look at the table comparing chain length, folding time and number of possible vs. relevant structures, we can add the protein that was used in the Shaw *et al.* paper. It has 30 residues, so something of the order of 10^{28} conformations. But the known folding time of $10 \mu\text{s}$ (10^{-5} s) gives an estimate of a million relevant conformations. We have no direct data (the paper does not report number of clusters), but from the figure it seems unlikely the simulations have sampled more than 1000 conformations; more likely something in the range of 10–100; let's be generous and say 100.

So we could estimate that these simulations may still be 10.000-fold off from real convergence, and the simulation should be extended to about 1 second. (The slide assumes 1000 conformations were visited in the RMSD plot A, which yields 0.1s as a minimum.)



The Shaw et al. paper shows that ten years later, in 2010 also for a small (70 residue) protein, we can now obtain reversible folding. The timescale needed is much larger than that for the peptide; here we see the simulation goes up to 200 μs (note, that this is a major investment of computational power, of the order of many months in parallel on a large number of CPU's).

Nevertheless, in panel A we only see a handful of folding and unfolding transitions (perhaps about 10-20 in total in the two simulations). We will go more into the issue of completeness of sampling, and convergence, in the next two slides.

Panel B shows a close-up of a single folding and an unfolding transition. Interestingly, the order of sub-events appears to be consistent (also across the other events). First, the first turn (blue) folds, then the two adjacent strands (orange) – think of this a a sort of ‘zipper’ motion. Then the third strands joins the first two. Both termini ‘fold’ last.

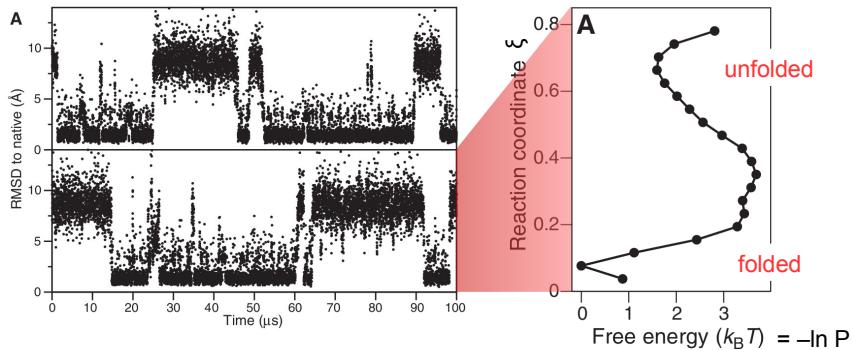
Panel C shows a cluster of conformations from the folded state (below about 5 \AA the RMSD plots in panel B). We can recognize a three-stranded β -sheet shape, but also lots of flexibility mainly in the termini. Only the backbone is drawn – the sidechains are (even) much more flexible.

In Panel D the probability of transitioning from the intermediate (barrier) position halfway folded and unfolded; there it is equally likely to go either way.

(Panel E was not discussed.)

Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wriggers W. Atomic-level characterization of the structural dynamics of proteins. *Science*. 2010 Oct 15;330(6002):341-6. doi: 10.1126/science.1187409.

From Statistics to Thermodynamics



- How does one go from sampled conformations to free energy?
- ξ is the reaction coordinate
- F_ξ is the free energy at ξ
- P_ξ is the probability at ξ

$$P_\xi = \frac{1}{Z} e^{-F_\xi/k_B T}$$

$$F_\xi = -k_B T \ln P_\xi$$



29 | 2 Mar 2016 (Thermo)dynamics and Interactions

Protein Folding & Dynamics

- Introduction to folding
- Dynamics



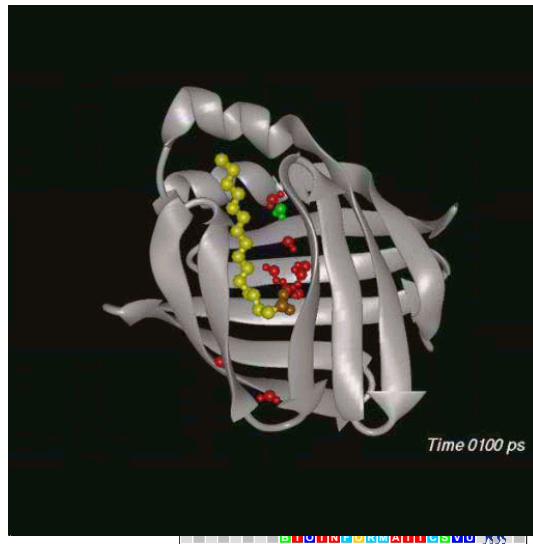
[30] 9 feb 2017 Protein Folding & Dynamics

This lecture will give a short introduction on the physical process by which chains of amino acids adopt a three-dimensional structure that we call a ‘protein’: the protein folding process. Also, we will touch on the related topic of protein (native) dynamics, i.e. the flexibility and movability that a protein structure has in its native state. Finally, we will start the introduction of thermodynamics as it relates to dynamics at the molecular scale.

Protein flexibility - The native state is heterogeneous

- Also a correctly folded protein is dynamic
 - Crystal structure yields average position of the atoms
 - 'Breathing' overall motion possible

(Fatty-acid binding protein – FABP)



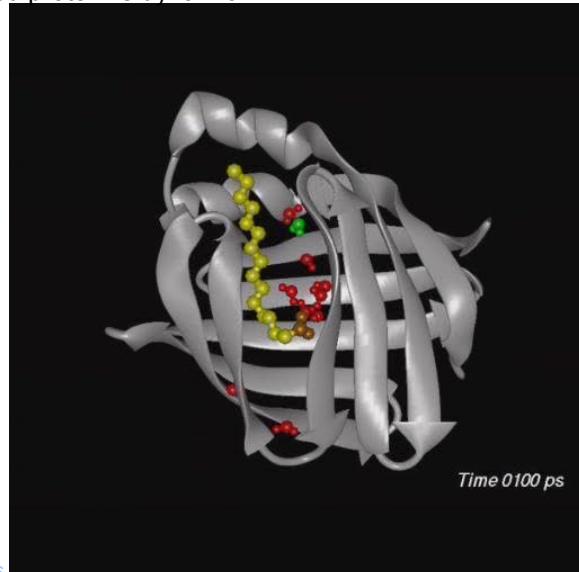
[31] 9 feb 2017 Protein Folding & Dynamics

We will now diverge via native protein dynamics (flexibility) to thermodynamics and in the end come back again to protein folding.

Protein flexibility - The native state is heterogeneous

- Also a correctly folded protein is dynamic
 - Crystal structure yields average position of the atoms
 - 'Breathing' overall motion possible

(Fatty-acid binding protein – FABP)



[32_9 feb 2017 Protein Folding & Dynamics](#)

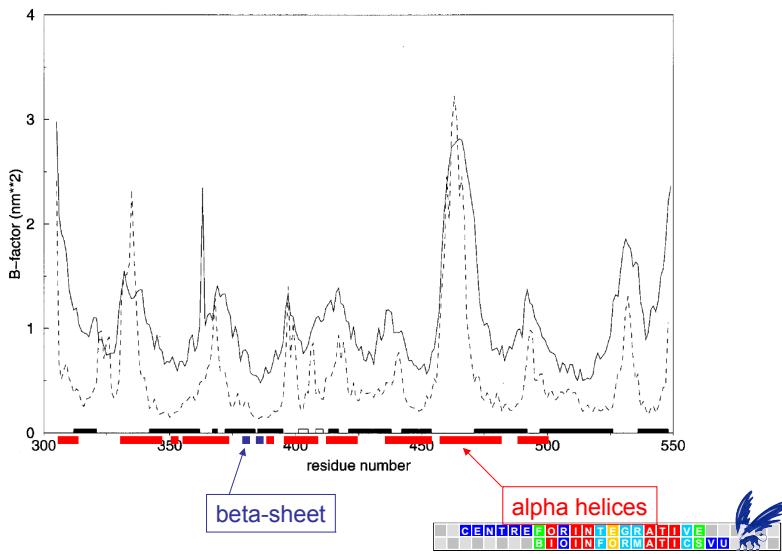
In spite of the rigid-sounding name, protein crystal structures are actually quite dynamic. For example, most enzymes are still active in the crystal form, although sometimes much slower than in solution. Some enzymes, however, require large scale motions of the protein structure. These are known to physically break the crystal when a substrate is added! (Shows how strong molecular motions can be.)

The movie shows the native dynamics of a fatty acid binding protein. The fatty acid is visible in the interior as the large green molecule. (Also visible are water molecules. The red ones are in the interior at the start of the simulation, the green ones are initially at the outside with the rest of the water. You see that most of the red ones 'escape' and that green ones, eventually, find their way into the interior. There are many other water molecules in this simulation that are not shown here.)

During the simulation you see that everything moves, all the time, but the overall structure of the protein is the same; it remains a folded β -barrel. Also the bound fatty acid molecule remains in its place, even though it wiggles back and forth a lot. (Only the small molecules (water) really move around, but the average location of water is also stable, you can see sometimes one water molecule being replaced by another.)

B-factors - experimental evidence for dynamics

- The average motion of an atom around the average position



[33] 9 feb 2017 Protein Folding & Dynamics

Even though protein crystal structures give us initially a rigid picture of the protein, the data does contain some information that may also be interpreted in a dynamic sense. The B-factors, or temperature factors as they are also sometimes called, indicate the how well the local electron density fits the atoms placed in it. A low B-factor means the fit is very good, a high B-factor means it is poor.

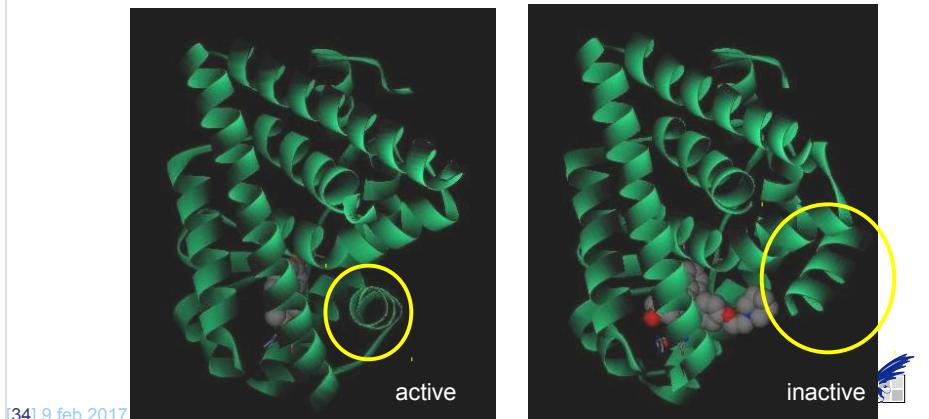
There are two contributions to this goodness of fit. One is variations between the molecules in the crystal. The refinement process of X-ray crystallography assumes all molecules to have identical conformation and orientation throughout the crystal. For proteins this does not strictly hold.

The second contribution is dynamics. Even if, in principle, all protein molecules are identically oriented, there will still be motion going on in them. This leads to a blurring of the densities observed (similar to what you get when taking a picture of a fast moving object), and hence higher B-factors. This also is the reason that crystal structures are often recorded at low temperature; this slows down the molecular motions and diminishes the ‘blurring’ effect.

From high-resolution low-temperature structures we now know that, typically, the first effect is minor. So, it is relatively safe to interpret high B-factors as indicating high mobility in the structure.

Conformational changes

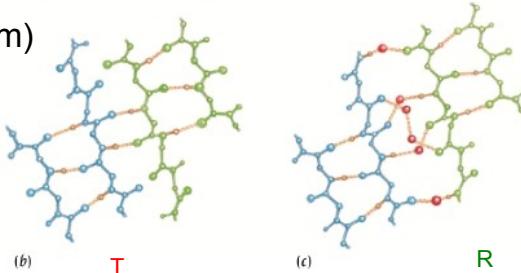
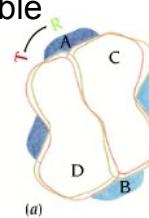
- Often conformational changes play an important role for the function of the protein
- Estrogen receptor
 - With activator (agonist) bound: active
 - With inactivator (antagonist) bound: not active



Of course, we all know that proteins have functional motions. Shown here is an example of the activation of the estrogen receptor. Note the orientation of the helix at the bottom right. Binding of an agonist (activator) rotates it by about 90 degrees.

Allosteric control

- Often two conformations possible
 - active T(ense) and inactive R(elaxed)
- Modulators change the conformation in the active form (or the inactive form)
- Not bound to active site: allosteric control



phosphofructokinase

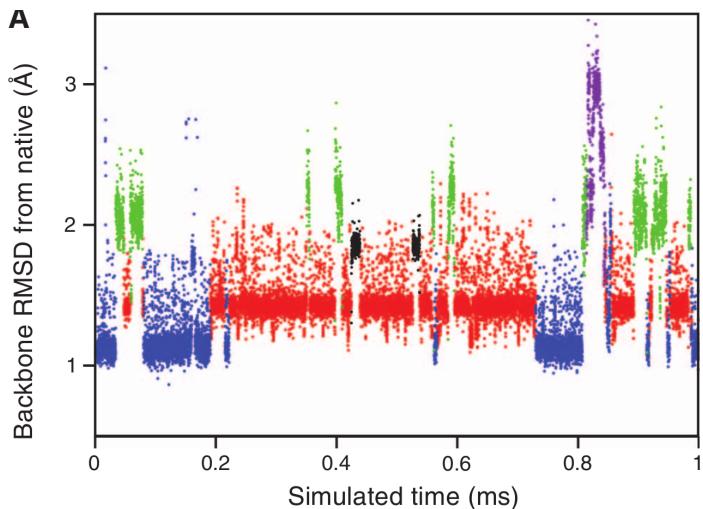
[35] 9 feb 2017 Protein Folding & Dynamics

© 1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group



The classical text-book example of functional motion is allostery. Here, also, ligand binding induces a conformational change, but in this case this results in altered (enhanced or reduced) affinity for a second binding site.

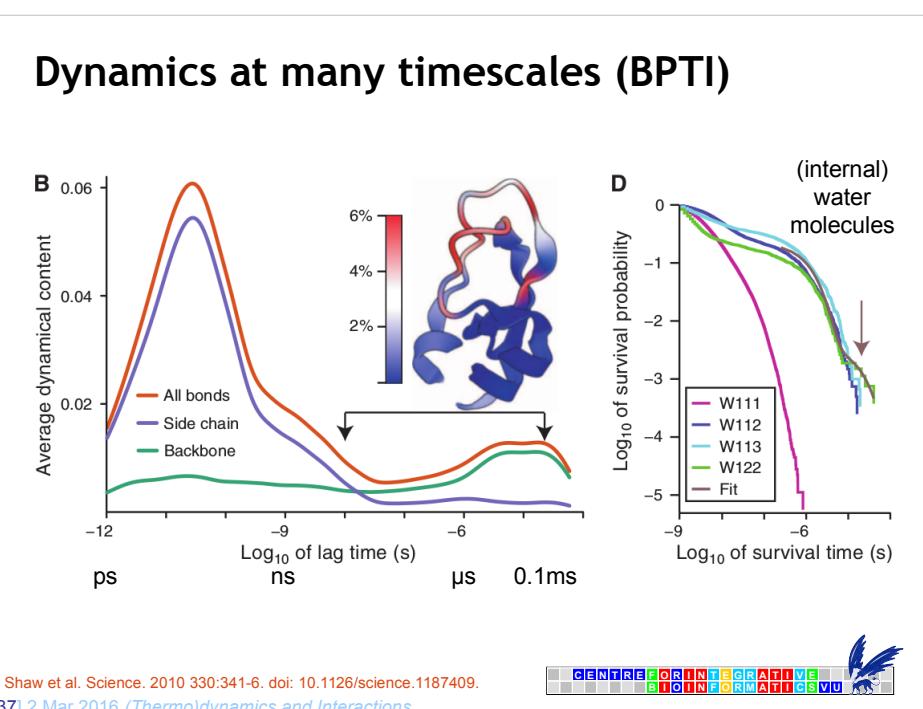
Native state dynamics of BPTI



Shaw et al. Science. 2010 330:341-6. doi: 10.1126/science.1187409.

36 | 2 Mar 2016 (Thermo)dynamics and Interactions

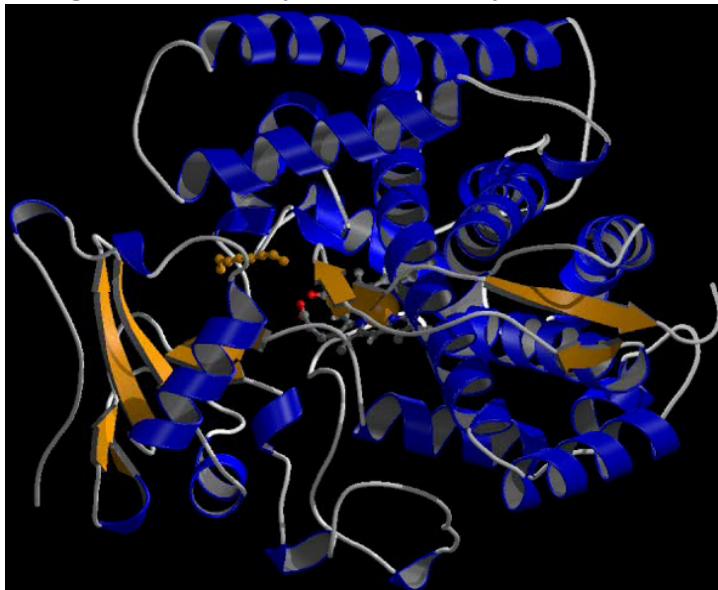




From the millisecond simulations of BPTI (bovine pancreatic trypsin inhibitor) in panel A (shown on previous slide), one can look at dynamic motions at different timescales, panel B. Here, we see that the fastest motions (ps-ns) are predominantly sidechain, while slow motions (above μs) are involves the backbone.

Inside the protein structure there are a few holes that are usually occupied by a water molecule. Panel D shows that three of these can remain there for many μs (microseconds), except for one which escapes within several 100 ns.

Analyzing Protein Dynamics: Cytochrome P450



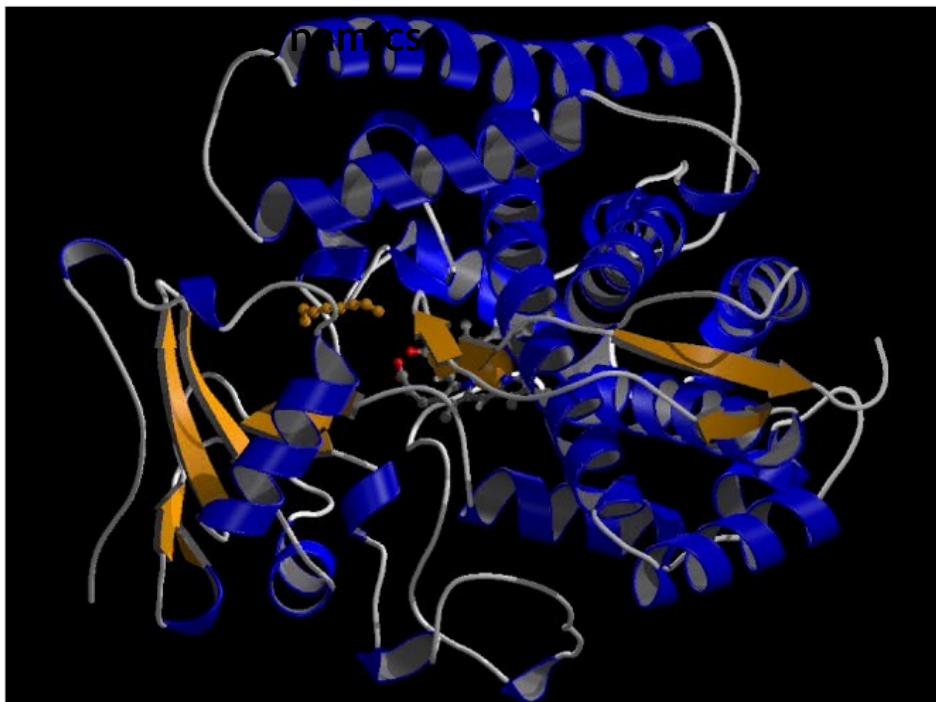
Feenstra, et al. *Prot. Sci.*, 16: 420-431 (2007).

38 | 2 Mar 2016 (Thermo)dynamics and Interactions

CENTRE FOR INTEGRATIVE
BIOMINFORMATICS VU



This is another example of enzyme dynamics, but here we could start from a crystal structure. The interest was in how and where ligands would bind, or rather, where they would bind most often.



Main points

- Anfinsen's theorem: proteins fold reversibly!
 - also means that *sequence dictates structure*
- Levinthal's paradox: how can proteins fold so fast?
 - Hydrophobic collapse, local rearrangement, folding funnel (energy landscape)
- Proteins are flexible
 - Everything is (always) moving; Flexibility is function!
 - The crystal structure is (only) an average
 - Reversible folding from simulation (peptides and small proteins)

[43] 9 feb 2017 Protein Folding & Dynamics





Protein Folding & Dynamics

Protein Folding & Dynamics
9 feb 2017