Intrinsically disordered proteins (IDPs)

Although the functions of many proteins clearly fit within this structure−function paradigm, where a unique amino acid sequence encodes a unique energetically stable 3D fold associated with conformational fluctuations that allow for unique biological function, recent studies have revealed that many functional proteins or functional protein regions do not have unique 3D structures under functional conditions.

In fact, contrarily to the ordered proteins and domains, such biologically active **intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs)** have no single, well-defined equilibrium structure and exist as highly dynamic, heterogeneous ensembles of conformers resulting from their relatively flat free-energy surface
dx.doi.org/10.1021/cr500288y | Chem. Rev. 2014, 114, 6557−656

A very rough scheme of protein functioning looks as follows:

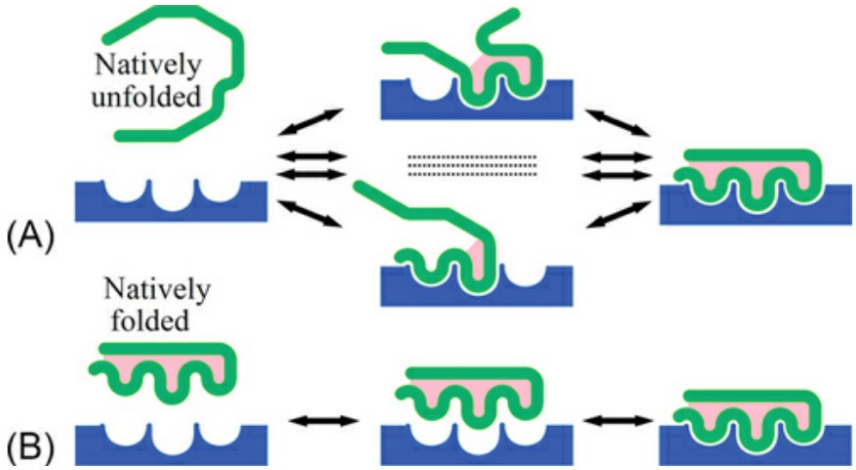
Remember that some proteins may perform only some of these actions; that the words “BIND” and “RELEASE” may imply binding and releasing a few different molecules; and that the word “TRANSFORM” may mean some chemical transformation, a change in conformation (of both the protein and the substrate), and/or movement of the protein or the substrate in space.

BINDING

Among these there are many **“natively disordered” proteins.** They acquire their unique spatial structure only when bound to a ligand (another protein, or DNA, or RNA or a small molecule).

The hard structure provides the binding specificity, but is in demand only when there is a binding; while the protein is unbound, it may be disordered.

The well-known example of disorder is presented by unfolded, in separately taken proteins, histone tails (Peng et al., 2012), which wrap around DNA to form a nucleosome. Many other biologically active proteins contain long disordered regions, such as histones, or are wholly disordered. These intrinsically disordered proteins/regions are common in nature and abundantly found in all organisms, where they carry out important biological functions that use the initial disorder of the chain.



The most obvious advantage is that the intrinsically unfolded chain **can bind to a large area**, which can lead to a large contact energy, and therefore to a high specificity, but the binding will be not too strong (because a great deal of the contact energy will be spent to fold the unfolded chain).

As a result, the binding can be very specific but reversible. The induced folding, that is, binding + folding (as well as unbinding + unfolding) can also be faster than binding of the rigid protein because it can start from any point of the unfolded chain, and this can lead to formation of the “energy funnel,” where the binding energy compensates the entropy loss step-by-step.
(A): Step-by-step binding + folding of the intrinsically unfolded chain. It does not require a precise preliminary positioning of the chain, because the binding can start from its any point, and then the growing contact energy is spent on compensating the binding and folding entropy. (B) An abrupt binding of a rigid protein requires its precise preliminary positioning (not compensated by the energy), and unbinding of such a protein requires an abrupt energy increase that is not compensated by a simultaneous energy increase.

The above described effect can lead to **acceleration of molecular recognition**—if the folding is fast (and it can be very fast indeed).

Therefore, no wonder that natively disordered proteins are widely involved in recognition, signaling, and regulation (Uversky, 2013), and frequently enough, one protein can be involved in different functions, because its conformation is affected by the interaction partner. The functions of these proteins complement the functional repertoire of “normal” ordered proteins. But they are not used as enzymes: effective catalysis requires hardness, in fact.

IDPs are characterized by their **biased amino acid composition** and **low sequence complexity**, as well as by their low proportions of bulky hydrophobic amino acids and high proportions of charged and hydrophilic amino acids.

Biased aminoacid composition:
Rich in polar, charged, and low-complexity amino acids like**: serine (Ser), proline (Pro), glutamine (Gln), alanine (Ala), and glycine (Gly)** (proline-rich regions and serine/threonine phosphorylation sites are critical for regulatory functions like signal transduction).

Deficient in hydrophobic residues like: **phenylalanine (Phe), tyrosine (Tyr), leucine (Leu), and valine (Val)**, which are important for stabilizing ordered structures.

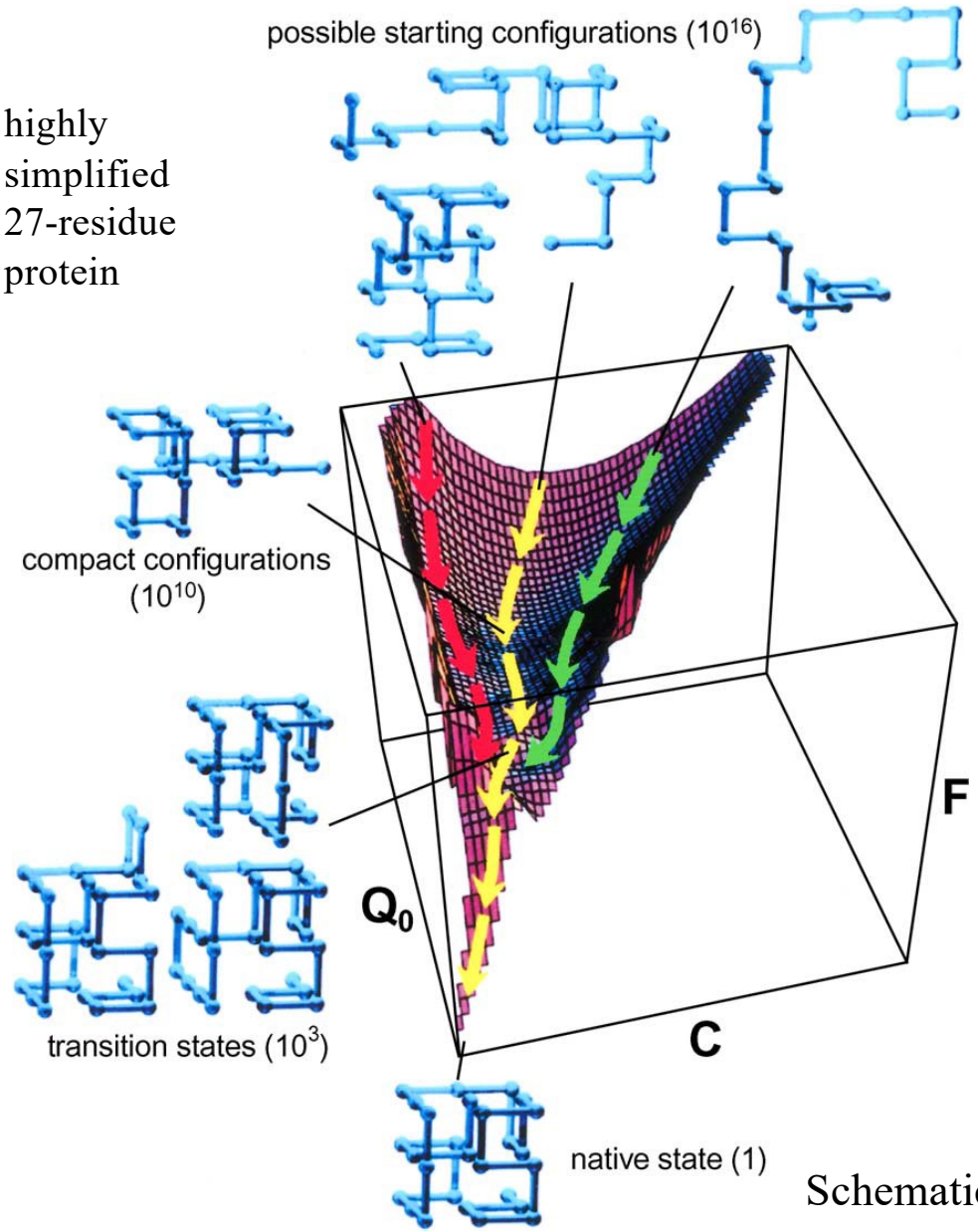
This **amino acid bias** contributes to their **flexibility** and **lack of stable tertiary structure**, making IDPs more prone to adopting multiple conformations.

Low sequence complexity:
refers to sequences in which there is **repetition of certain amino acids** or **motifs** that make the sequence relatively simple or homogeneous. In contrast to highly variable sequences, **low-complexity sequences** tend to have patterns of repetitive residues or short motifs (e.g., **polyglutamine (polyQ)** repeats or **serine-threonine-rich** regions).

Despite being perfectly functional, such protein sequences are unable to fold spontaneously into stable, well-defined, globular three-dimensional structures; instead, they are **dynamically disordered and fluctuate rapidly through a range of conformations**, which cover a continuum of conformational space ranging from extended statistical coils to collapsed globules.

Play a fundamental role in:

- ordered assembly of macromolecular machines such as the ribosome
- organization of chromatin
- assembly and disassembly of microfilaments and microtubules
- transport through the nuclear pore
- binding and transport of small molecules
- functioning of protein and RNA chaperones
- separation of functional protein domains (as flexible ‘entropic’ linkers)



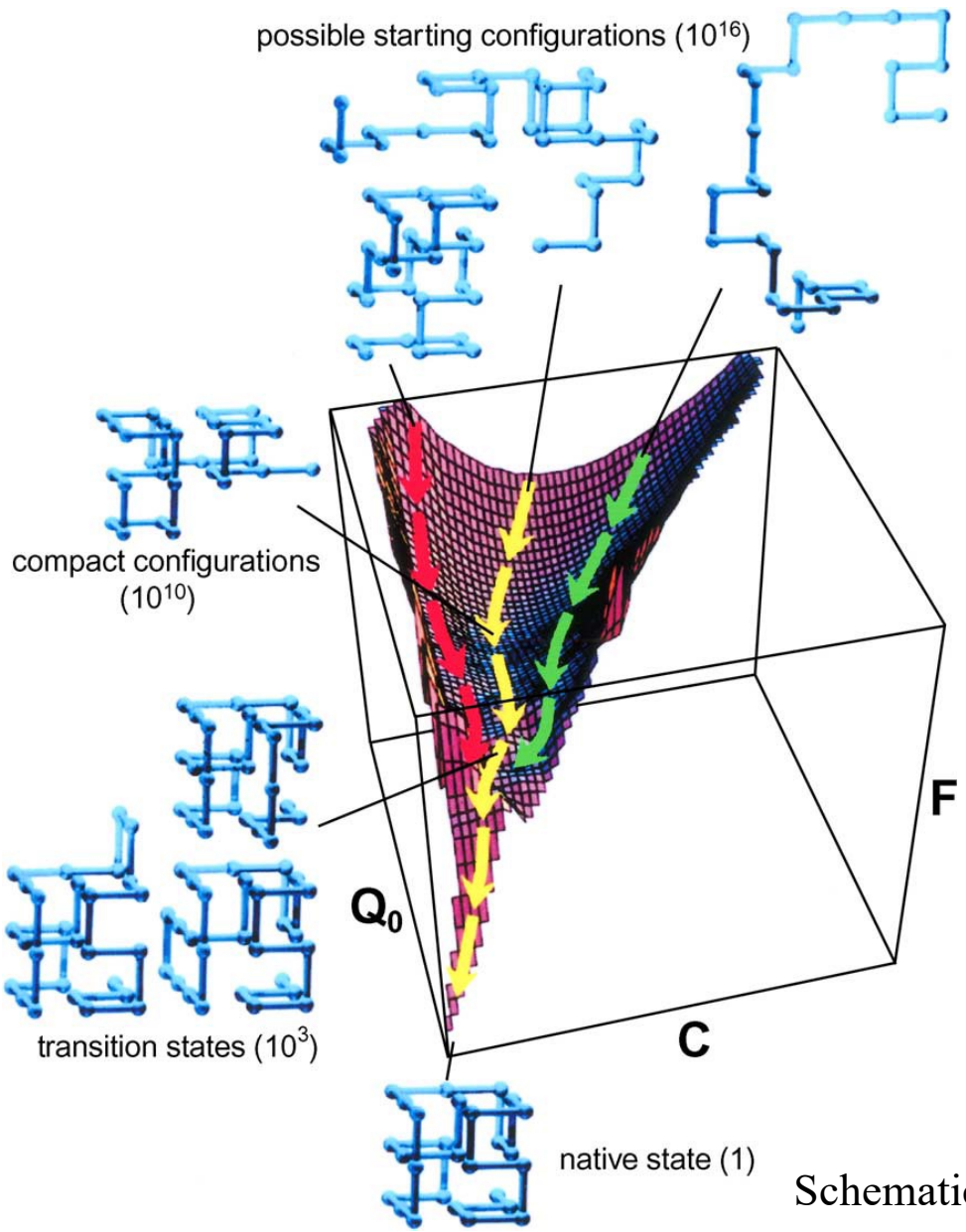
The **free energy (F)** of the system is shown as a function of the **total number of contacts between residues (C)** and the **number of contacts that correspond to those of the most stable structure, denoted the native state (Q0).**

The abundance of IDPs (40% of eukaryotes proteins) in the cell is tightly regulated to ensure precise signalling in time and space, and mutations in IDPs or changes in their cellular abundance are associated with disease.

Mutation in IDP is related to **protein misfolding.** We should first understand **protein folding.**

The native state of a protein corresponds to the structure that is most stable under physiological conditions.
Nevertheless, the total number of possible conformations of a polypeptide chain is so large that it would take an astronomical length of time to find this particular structure by means of a systematic search of all conformational space.

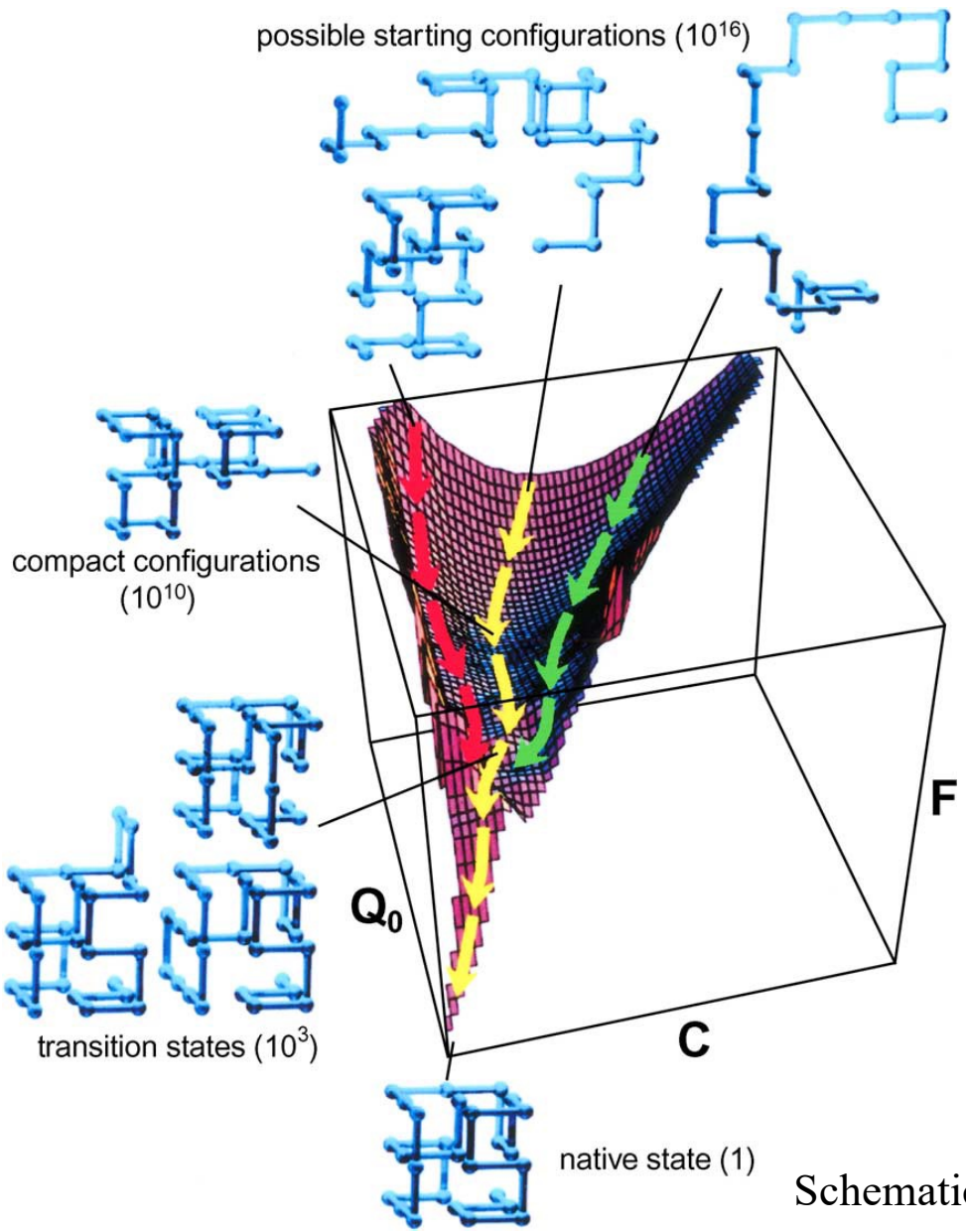
Schematic energy landscape for protein folding (**Folding "funnel"**)



It is now evident that the folding process does not involve a series of mandatory steps between specific partially folded states, but rather a **stochastic search of the many conformations accessible to a polypeptide chain.**

In essence, the **inherent fluctuations in the conformation** of an incompletely folded polypeptide chain enable even residues at very different positions in the amino acid sequence to come into contact with one other. Because the correct (native-like) interactions between different residues are on average more stable than the incorrect (nonnative) ones, such a search mechanism is in principle able to find the lowest energy structure. It is evident that this process is extremely efficient for those special sequences that have been selected during evolution to fold to globular structures, and indeed only a very small number of all possible conformations needs be sampled during the search process.

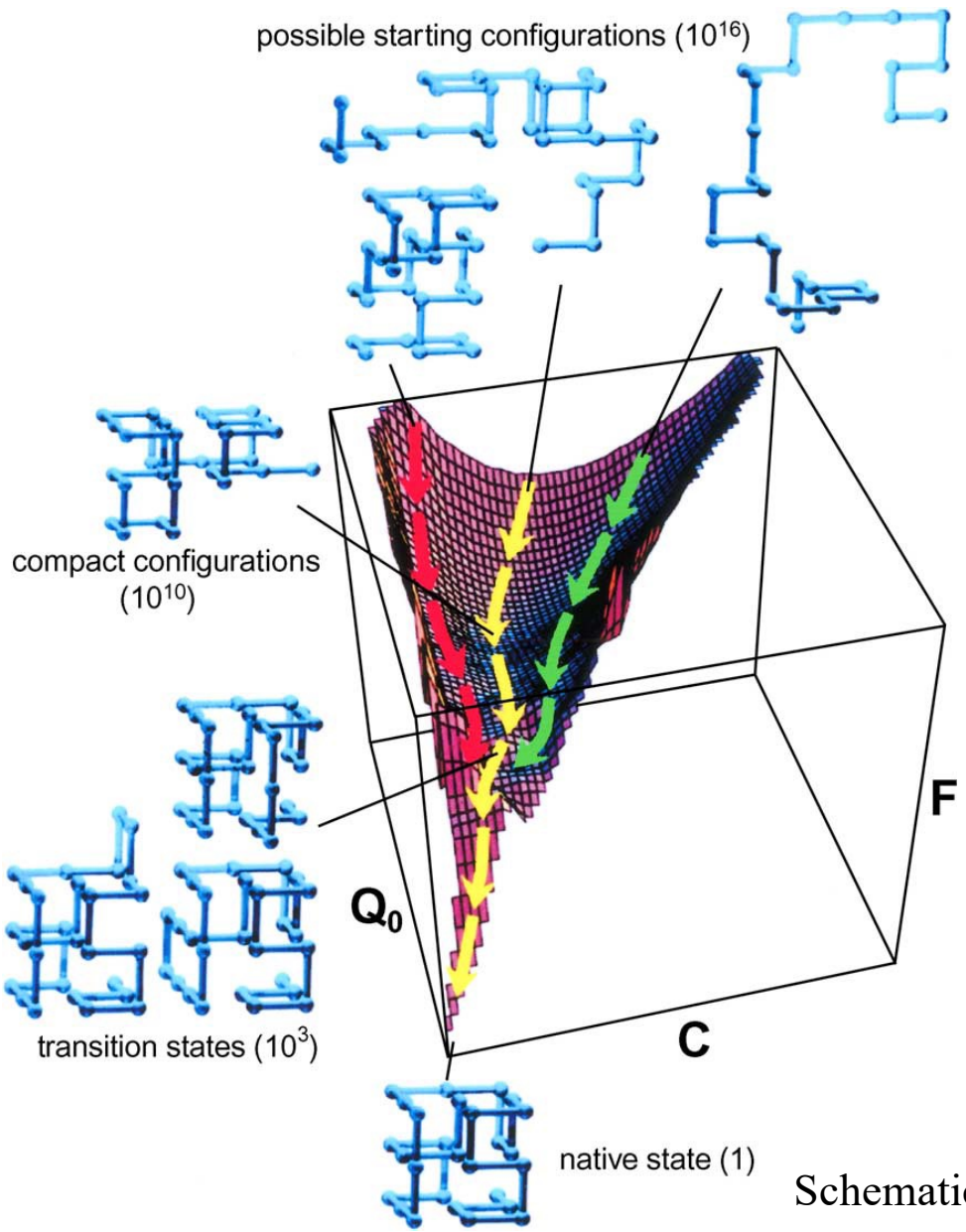
Schematic energy landscape for protein folding (**Folding "funnel"**)



This **stochastic description of protein folding** involves the concept of an **"energy landscape"** for each protein, describing the free energy of the polypeptide chain as a function of its conformational properties. To enable a protein to fold efficiently, the landscape required has been likened to a **funnel** because the conformational space accessible to the polypeptide chain is reduced as the native state is approached.

In essence the high degree of disorder of the polypeptide chain is reduced as folding progresses, as the **more favourable enthalpy** associated with stable native-like interactions **can offset the decreasing entropy** as the structure becomes more ordered.

Schematic energy landscape for protein folding (**Folding "funnel"**)



The exact manner in which the correct overall fold can be achieved through such a process is emerging primarily from studies of a group of small proteins—most having less than 100 residues—that fold to their native states without populating significantly any intermediate states (use of site-directed mutagenesis to probe the roles of individual residues in the folding process).

The results of a wide range of studies suggest that the fundamental mechanism of folding can be described as **"nucleation–condensation"** in which a folding nucleus of a small number of key residues forms, about which the remainder of the structure can then condense.

Schematic energy landscape for protein folding (**Folding "funnel"**)

How to determine the folding pathway?

Typically, a protein is placed in a solution of a chemical denaturant, such as urea, so that it unfolds. By returning the protein to conditions under which the native state is again stable (this can sometimes be achieved simply by diluting the solution) it is often possible to monitor the complex process of the recovery of the native structure over time by using biophysical techniques:

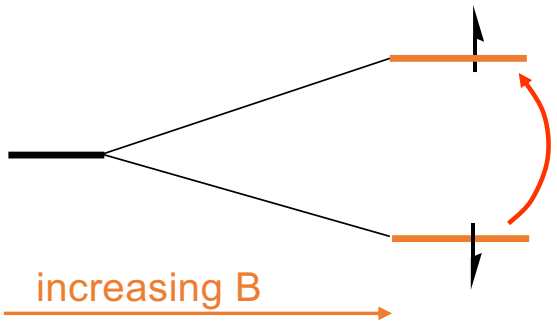
- Circular Dichroism (CD)
- NMR spectroscopy
- SAXS (small angle x-ray scattering)

coupled with stopped or quenched flow methodologies. Moreover:

- theoretical approaches, particularly based on computer simulations of the events occurring during folding/unfolding

Nuclear magnetic resonance (NMR)

Certain nuclei (^{1}H, ^{13}C, ^{15}N, ^{31}P) have angular momentum. In high magnetic fields we can separate the energy levels of nuclei with different spin angular momentum.



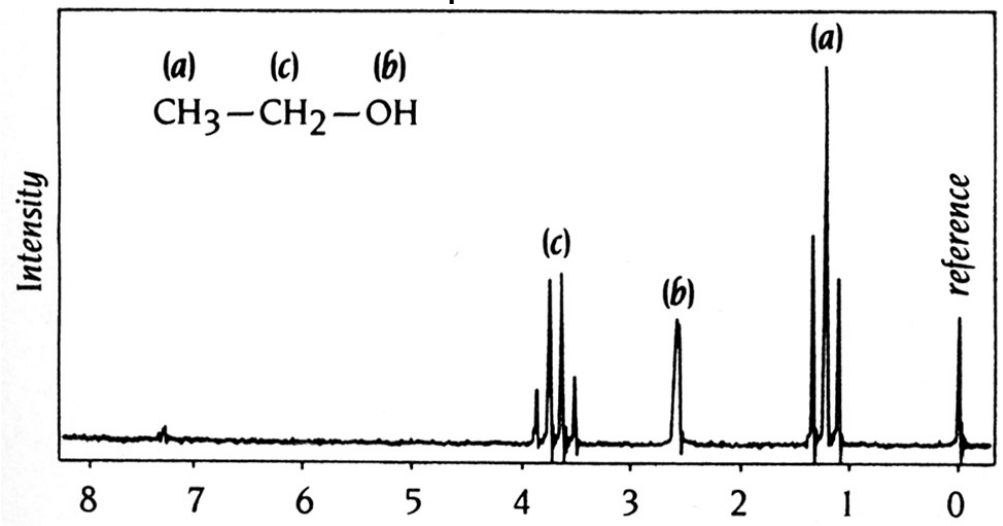
the spin will align along the field, but absorption of electromagnetic radiation of appropriate frequency (radio) induces a transition.

When the nuclei revert to their equilibrium state they emit radiation that can be measured. The exact frequency of the emitted radiation depends on the environment of the nucleus.

NMR: chemical shifts

The exact frequency of the emitted radiation depends on the environment of the nucleus. Nuclei in different chemical environment will absorb at slightly different frequencies. These different frequencies (relative to a reference signal) arethe chemical shifts.

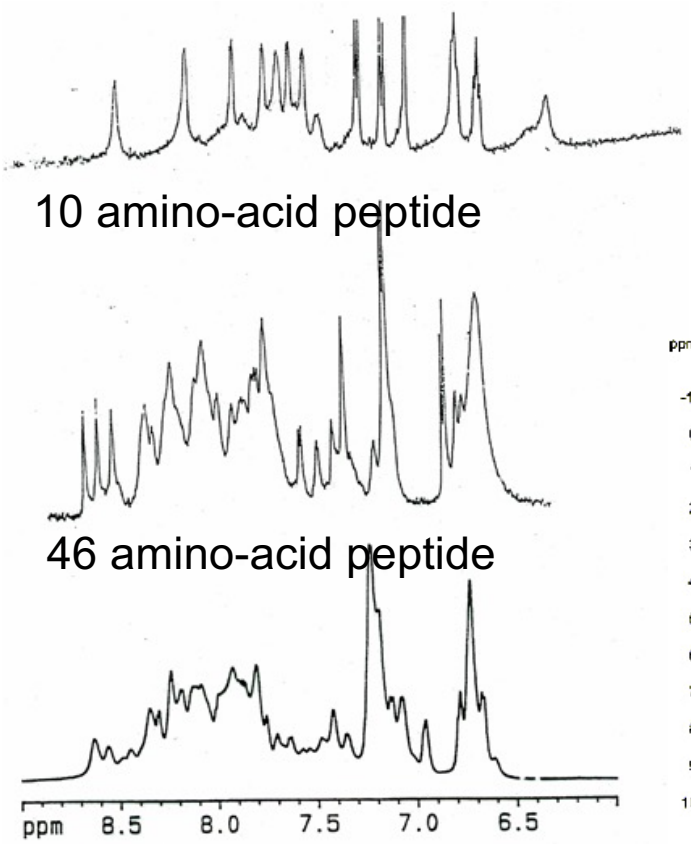
^{1}H-NMR spectrum of ethanol



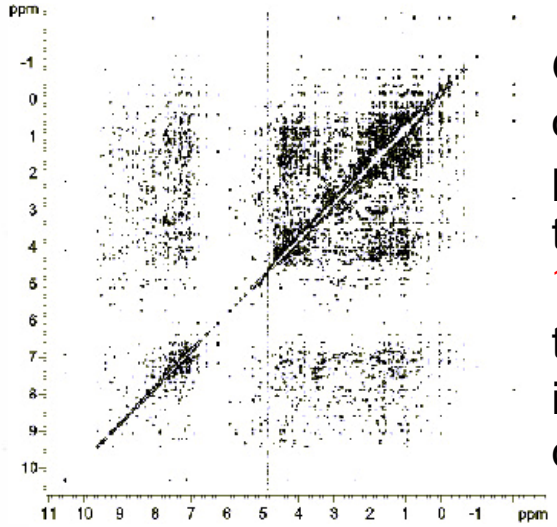
Chemical shifts for all the hydrogens in this small molecule: hydrogen that are part of the CH_3, or CH_2, or OH group have different shifts.

NMR and peptides

1D NMR spectra of proteins contain such a large number of overlapping peaks as to be uninterpretable.



We need to use *pulse sequences* with specific shapes and time intervals to allow magnetization transfer between nuclei and, therefore the detection of nuclear-nuclear interactions through space (NOESY).

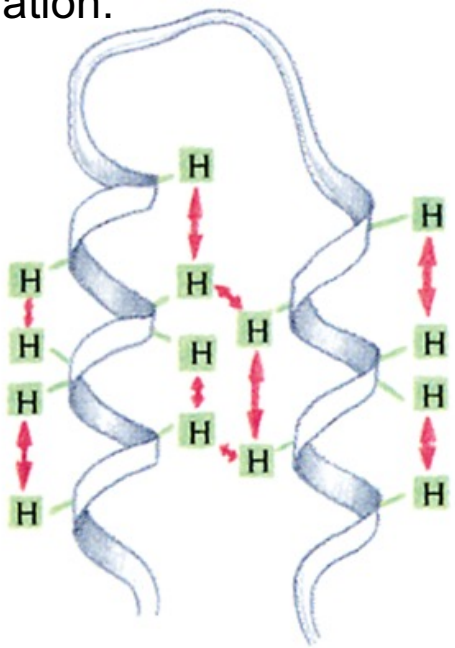


One problem is overlap between peaks: labelling the protein with ^{13}C and ^{15}N help to resolve overlap in the proton dimension.

NMR: mapping distances

NOESY is a 2D NMR experiment that provides information about proton-proton proximities via the nuclear Overhauser effect (NOE). NOE is an interaction between nuclei that are close in space (usually within 5 Å), and it can be used to determine spatial relationships between atoms in a molecule, which is very useful in protein and nucleic acid structure determination.

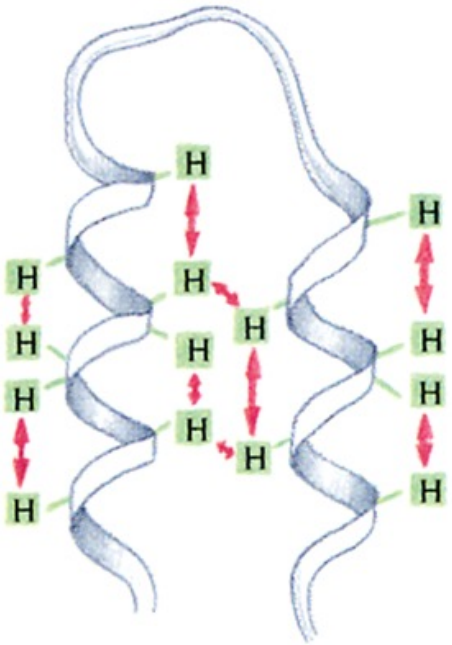
The NOE occurs when the relaxation of one nucleus is influenced by the presence of another nucleus in close proximity. For a proton to produce an observable NOE effect, it must be within a certain distance (usually <5 Å) of the second proton. This allows NOESY to detect spatial proximity rather than chemical bonding.



NMR: mapping distances

The intensity of a NOESY peak is $\propto r^{-6}$, so the distance is determined according to intensity of the peak. The intensity-distance relationship is not exact, so usually a distance range is used.

We can basically map distances between pairs of atoms: the NMR spectra by specifying which pairs of atoms are close together in space, contain three-dimensional information about the protein molecules.



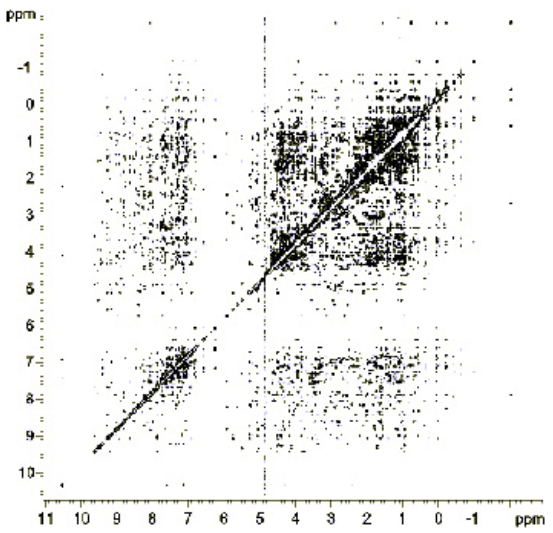
NMR: mapping distances

In a typical NOESY experiment:

•**Diagonal peaks** represent **single proton signals**.

•**Off-diagonal peaks** (cross-peaks) represent **NOE interactions** between different protons.

•The presence of these cross-peaks is used to infer **3D spatial proximity** between the protons.



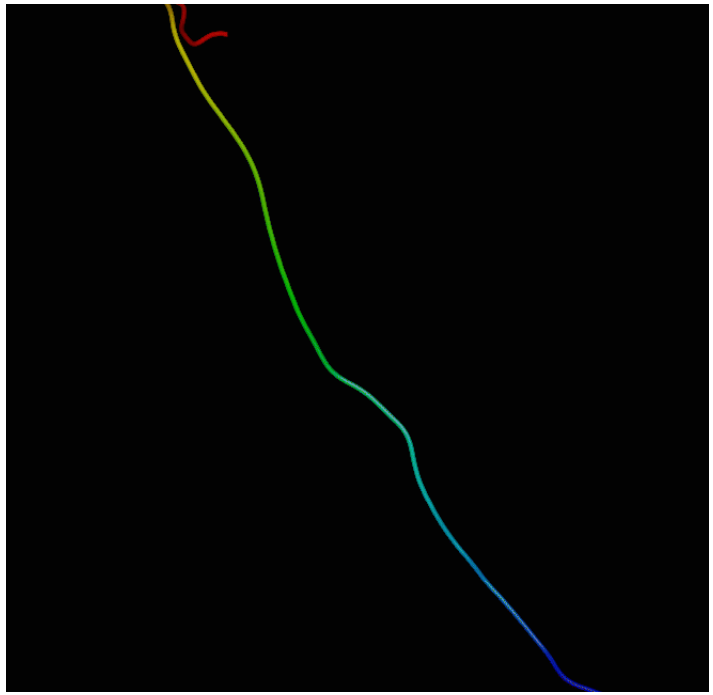
The **position** of the cross-peaks corresponds to the **chemical shifts** of the two interacting protons.
The **intensity** of the cross-peak is related to the **distance** between the interacting protons. The intensity of an NOE interaction typically decays as $1/r^6$

NMR refinement

The experimentally determined distance restraints (together with the protein stereochemical constraints) can be used as input for the structure determination.

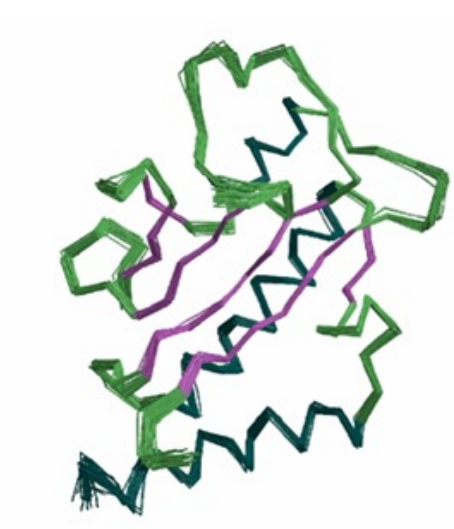
Algorithms start from an unfolded polypeptide and attempt to satisfy as many of the restraints as possible, in addition to general properties of proteins such as bond lengths and angles.
As in crystallography, this is done by converting them into energy terms and thus minimizing the energy.

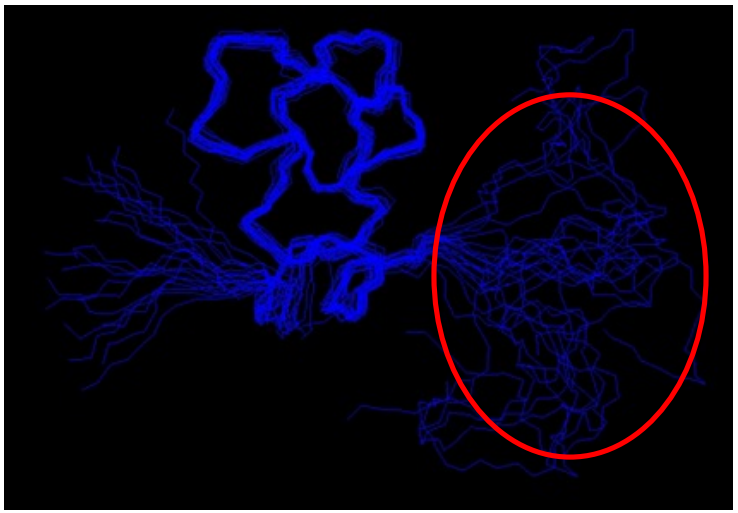
Repeat many times and compare results.



NMR structure

The process is repeated many times and results in an ensemble of structures that, if the data were sufficient to dictate a certain fold, will converge.

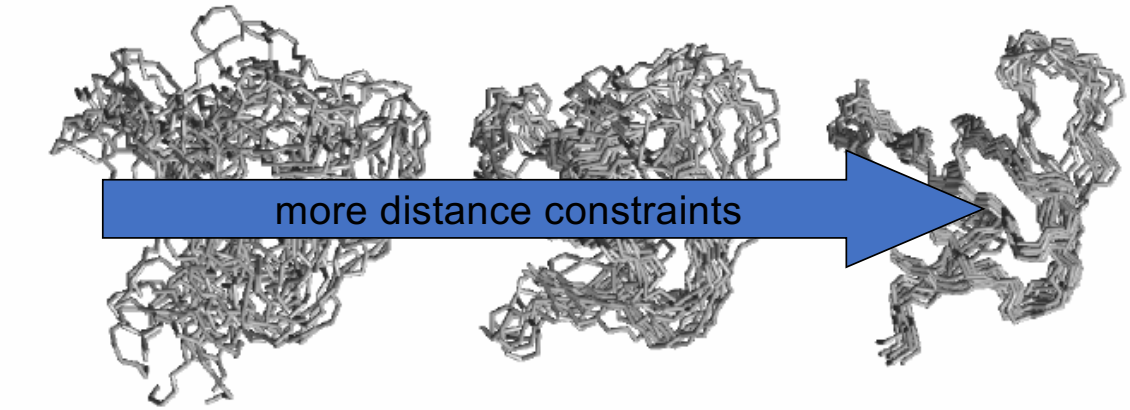




intrinsic disorder or lack of data?

Often it is claimed that NMR structures visualise the “dynamic” nature of proteins: however it is difficult to discriminate between flexibility and lack of data... (there are sophisticated ways of addressing this, but they are not used on the “average” NMR structure).

Accuracy of the structure?



There is a direct correlation between the quality of the NMR structure and the number of distance constraints:
more constraints → higher the precision of the structure

Accuracy is measured by the RMS deviation of the ensemble:
- “high resolution” : RMSD < 1A
- “medium resolution” : RMSD < 2A

Warning! strongly depends on which programme was used for the refinement!

NMR and protein size

Traditionally NMR is limited to relatively small proteins or protein domains.

Problems resolving overlapping peaks - alleviated by the introduction of isotope labelling and multidimensional experiments.

A more serious problem is the fact that in large proteins the magnetization relaxes faster, which means there is less time to detect the signal, causing the peaks to become broader and weaker, and eventually disappear.

Two techniques have been introduced to attenuate the relaxation: transverse relaxation optimized spectroscopy (TROSY) and deuteration of proteins.

No labelling:	< 15 kDa
$^{13}C/^{15}N$:	< 20 kDa
$^{13}C/^{15}N$ and deuteration:	< 40 kDa
Deuteration + TROSY:	up to 100 kDa??

Small angle X-rays scattering (SAXS)

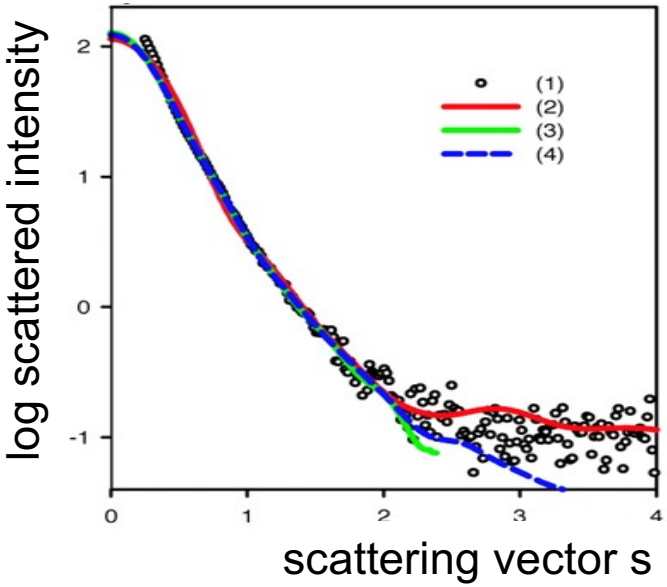
Crystals:

- diffracted intensity concentrates around specific directions
- high resolution (~ λ)

Molecules in solution:

- intensity is isotropic and concentrates around the beam (small-angle)
- low resolution (>>>λ)
- scattering is proportional to that of a single particle averaged over all orientations
- allows to determine size and shape to 1-10 nm resolution (?)
- not a lot of data (1D scattering pattern, radially averaged)
- no unique solution: multiple 3D structures can give rise to same pattern
- need concentration of ~ 1mg/ml, with no aggregation
- need accurate estimate of conc.

Small angle X-rays scattering

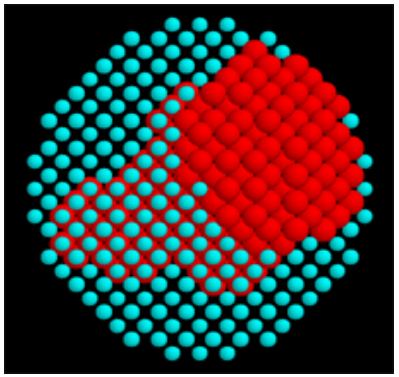


o experimental scattering curve

curve calculated from PX model (models used for phasing X-ray diffraction data)

curves calculated from ab initio models

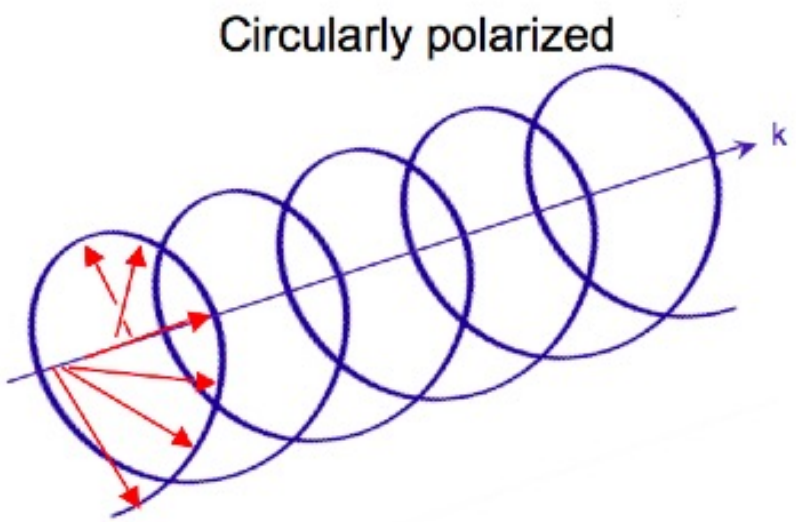
- compare data with known structure/s, to analyse:
  - conformational changes
  - shape/position of unknown parts
- *ab initio* reconstruction by modelling spheres and compare with experimental curve (?)



Circular Dichroism

Circular Dichroism (CD) is a spectroscopic technique used to investigate the **secondary structure** of proteins and other biomolecules. CD measures the differential absorption of left- and right-handed circularly polarized light, which provides information about the **conformation** of chiral molecules like proteins and nucleic acids.

- **Determining the secondary structure** content of proteins (e.g., alpha-helices, beta-sheets, random coils).
- **Monitoring conformational changes** in proteins (e.g., unfolding, ligand binding).
- **Quantifying protein folding** and **stability**.



Circular Dichroism

When circularly polarized light passes through an absorbing optically active medium, the speeds between right and left polarizations differ:

$$C_L \neq C_R$$

$$\lambda_L \neq \lambda_R$$

$$\varepsilon_L \neq \varepsilon_R$$

Circular dichroism is the difference $\Delta = \varepsilon_L - \varepsilon_R$

The electric field excite a molecule causing linear displacement of charge (electric dipole), its magnetic field causes a circulation of charge (magnetic dipole), and overall an elliptical path. If the molecule is chiral, the two types of circularly polarized light are absorbed to different extents.
In a CD experiment, equal amounts of left and right circularly polarized light of a selected wavelength are alternately radiated into a (chiral) sample. One of the two polarizations is absorbed more than the other one, and this wavelength-dependent difference of absorption is measured, yielding the CD spectrum of the sample.

The amyloid state and its association with protein misfolding diseases

Tuomas P. J. Knowles, Michele Vendruscolo and Christopher M. Dobson

There are now approximately 50 disorders, with a multitude of disparate symptoms, which are associated with the **misfolding** of normally soluble, functional peptides and proteins, and their subsequent conversion into **intractable aggregates**, of which the archetypal examples are **amyloid fibrils,** thread-like structures, the formation of which is associated both with a loss of function of the proteins involved and with the generation of often **toxic intermediates** in the process of **self-assembly.**

Table 1 | **Some human diseases associated with protein misfolding and amyloid aggregation***

Disease	Aggregating protein or peptide	Polypeptide length (number of residues)	Structure of protein or peptide
Neurodegenerative diseases			
Alzheimer's disease	Amyloid-β peptide	37–43	Intrinsically disordered
Spongiform encephalopathies	Prion protein or its fragments	230	Intrinsically disordered and α-helical
Parkinson's disease	α-synuclein	140	Intrinsically disordered
Amyotrophic lateral sclerosis	Superoxide dismutase 1	153	β-sheet and Ig-like
Huntington's disease	Huntingtin fragments	Variable	Mostly intrinsically disordered
Familial amyloidotic polyneuropathy	Transthyretin mutants	127	β-sheet
Non-neuropathic systemic amyloidosis			
Amyloid light chain (AL) amyloidosis	Immunoglobulin (Ig) light chains or its fragments	~90	β-sheet and Ig-like
Amyloid A (AA) amyloidosis	Serum amyloid A1 protein fragments	76–104	α-helical and unknown fold
Senile systemic amyloidosis	Wild-type transthyretin	127	β-sheet
Haemodialysis-related amyloidosis	β_2-microglobulin	99	β-sheet and Ig-like
Lysozyme amyloidosis	Lysozyme mutants	130	α-helical and β-sheet
Non-neuropathic localized amyloidosis			
Apolipoprotein A1 (Apo A-1) amyloidosis	Apo A-1 fragments	80–93	Intrinsically disordered
Type II diabetes	Amylin	37	Intrinsically disordered
Injection-localized amyloidosis	Insulin	21 and 30	α-helical and insulin-like

*A selection of diseases associated with extracellular amyloid deposits or intracellular inclusions with amyloid-like characteristics. See REF. 5 for a more comprehensive list of the approximately 50 human protein misfolding diseases and their associated proteins.

The amyloid state and its association with protein misfolding diseases

Tuomas P. J. Knowles, Michele Vendruscolo and Christopher M. Dobson

From a wide range of *in vitro* experiments on peptides and proteins we now know that the formation of amyloid structures is not a rare phenomenon associated with a small number of diseases but rather **that it reflects a well-defined structural form of the protein that is an alternative to the native state** — a form that may in principle be adopted by many, if not all, polypeptide sequences

Table 1 | **Some human diseases associated with protein misfolding and amyloid aggregation***

Disease	Aggregating protein or peptide	Polypeptide length (number of residues)	Structure of protein or peptide
Neurodegenerative diseases			
Alzheimer's disease	Amyloid-β peptide	37–43	Intrinsically disordered
Spongiform encephalopathies	Prion protein or its fragments	230	Intrinsically disordered and α-helical
Parkinson's disease	α-synuclein	140	Intrinsically disordered
Amyotrophic lateral sclerosis	Superoxide dismutase 1	153	β-sheet and Ig-like
Huntington's disease	Huntingtin fragments	Variable	Mostly intrinsically disordered
Familial amyloidotic polyneuropathy	Transthyretin mutants	127	β-sheet
Non-neuropathic systemic amyloidosis			
Amyloid light chain (AL) amyloidosis	Immunoglobulin (Ig) light chains or its fragments	~90	β-sheet and Ig-like
Amyloid A (AA) amyloidosis	Serum amyloid A1 protein fragments	76–104	α-helical and unknown fold
Senile systemic amyloidosis	Wild-type transthyretin	127	β-sheet
Haemodialysis-related amyloidosis	β_2-microglobulin	99	β-sheet and Ig-like
Lysozyme amyloidosis	Lysozyme mutants	130	α-helical and β-sheet
Non-neuropathic localized amyloidosis			
Apolipoprotein A1 (Apo A-1) amyloidosis	Apo A-1 fragments	80–93	Intrinsically disordered
Type II diabetes	Amylin	37	Intrinsically disordered
Injection-localized amyloidosis	Insulin	21 and 30	α-helical and insulin-like

*A selection of diseases associated with extracellular amyloid deposits or intracellular inclusions with amyloid-like characteristics. See REF. 5 for a more comprehensive list of the approximately 50 human protein misfolding diseases and their associated proteins.

How do proteins either adopt and maintain the specific states that are needed to carry out given functions or instead misfold and form potentially pathogenic aggregates such as amyloid fibrils? Importance of investigating the nature and properties of the various states in which these molecules can be found.

How is the conversion of proteins into the amyloid state generally avoided in living systems? Importance of define the specific mechanisms by which aggregation occurs and the manner in which it can induce pathogenic behaviour.

Progress from:

- chemical kinetics, nanotechnology and microfluidics.
- comparative studies of the behaviour of a wide range of peptides and proteins, also non pathogenic

The multiplicity of protein states

We discussed how the varied and intricate structures of globular proteins are encoded by their amino acid sequences, and that these molecules have an intrinsic ability to fold spontaneously.

From these studies we now know that the folding process can be described as a **diffusional search on a free energy surface**. A combination of experimental and theoretical studies has clarified how **given sequences define specific free energy surfaces that enable folding**.

However, **free energy surface depends on a very large number of relatively weak interactions**, including hydrogen bonds, electrostatic interactions, dispersion forces and **interactions with solvent molecules** that underlie the hydrophobic effect.

Moreover, in some cases the native state of a given peptide or protein may not be structured in a globular form but disordered.

The functional native state is likely to only reflect a local free energy minimum at physiological concentrations, as self-association into aggregated protein species may in many cases lower the global free energy. Indeed, the maintenance of protein solubility has emerged as a central aspect of the more general topic of protein homeostasis.

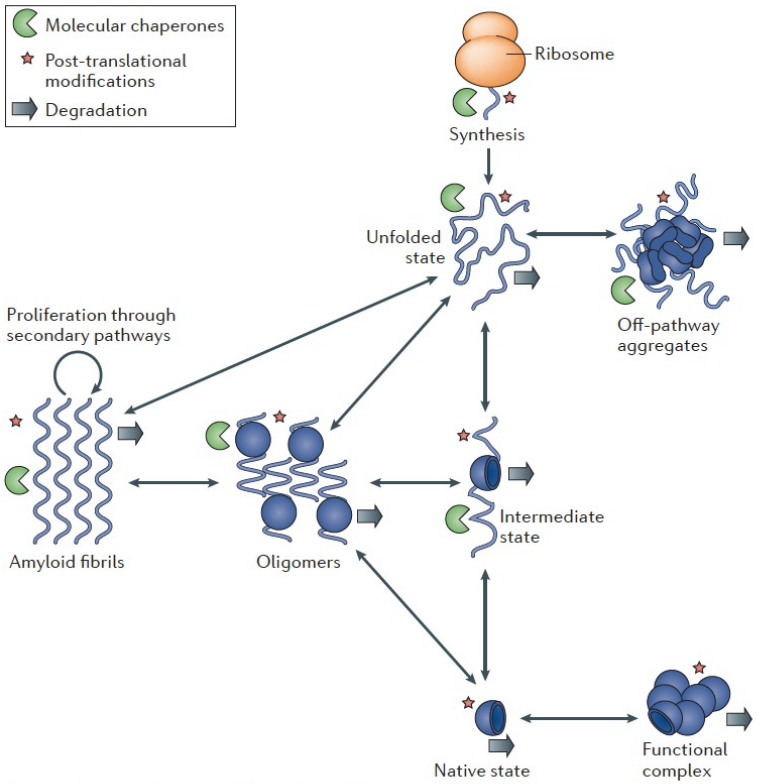


Figure 1 | **A protein can exist in various different states.** The populations of the different states and their interconversion rates are determined by their different thermodynamic stabilities, by the free energy barriers that are associated with the corresponding transitions as well as by the rates of synthesis and degradation, the propensity to interact with chaperones and to undergo post-translational and other chemical modifications. The amyloid state of a protein is a highly ordered form of aggregate in which the polypeptide chains adopt a fibrillar structure, which is capable of self-replication, for example through secondary processes. Amyloid fibrils are rich in β-sheet structure and typically form from unfolded or partially unfolded conformations of proteins and peptides, some of which are fragments of larger proteins. The amyloid state is 'generic' in that its characteristic architecture is not encoded by specific amino acid sequences.

Intrinsically disordered proteins are not necessarily prone to aggregation, as their sequences have usually evolved to maintain the **level of solubility that is required for their optimal function**; the existence of extensive regions that are highly abundant in charged and polar groups disfavour intermolecular association from a thermodynamic point of view.

Strong **kinetic barriers to aggregation** to overcome.

Partially folded protein species are particularly vulnerable to misfolding and aggregation from which they must be protected in living systems. Partially folded states may, however, also be required for functional reasons, including trafficking to specific cellular locations that require translocation across mitochondrial and even nuclear membranes.

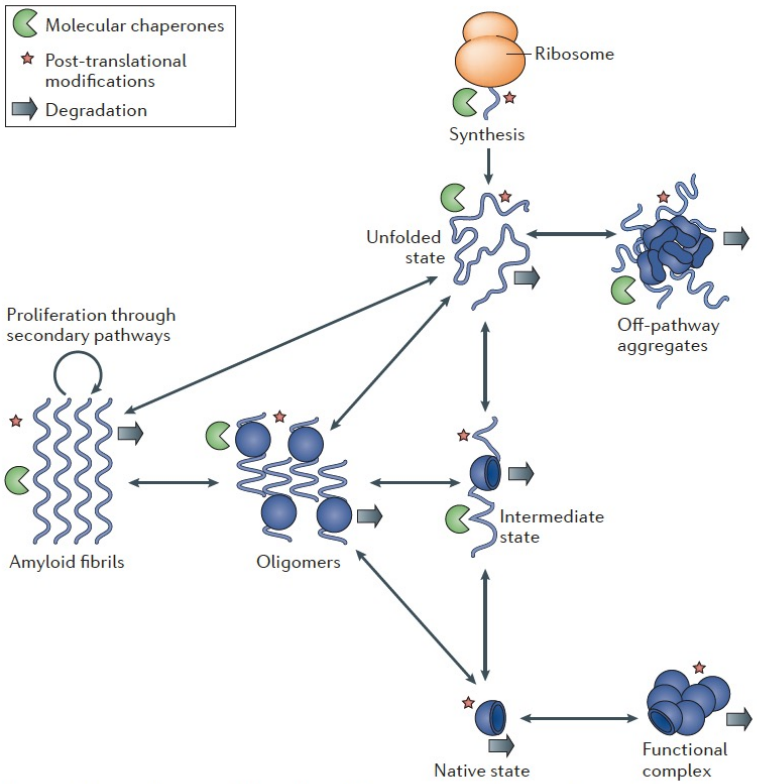


Figure 1 | **A protein can exist in various different states.** The populations of the different states and their interconversion rates are determined by their different thermodynamic stabilities, by the free energy barriers that are associated with the corresponding transitions as well as by the rates of synthesis and degradation, the propensity to interact with chaperones and to undergo post-translational and other chemical modifications. The amyloid state of a protein is a highly ordered form of aggregate in which the polypeptide chains adopt a fibrillar structure, which is capable of self-replication, for example through secondary processes. Amyloid fibrils are rich in β-sheet structure and typically form from unfolded or partially unfolded conformations of proteins and peptides, some of which are fragments of larger proteins. The amyloid state is 'generic' in that its characteristic architecture is not encoded by specific amino acid sequences.

Examples:

proteins that normally function in globular states often adopt intermediate conformations (corresponding to local minima on the free energy surface) before becoming fully folded and incorporated into their biological environments (for example, in the cytosol or within membranes) following their synthesis on the ribosome, particularly those proteins that are large and have complex folds.

proteins that have folded correctly can subsequently unfold, at least locally, often simply as a consequence of the dynamical fluctuations that occur within protein molecules

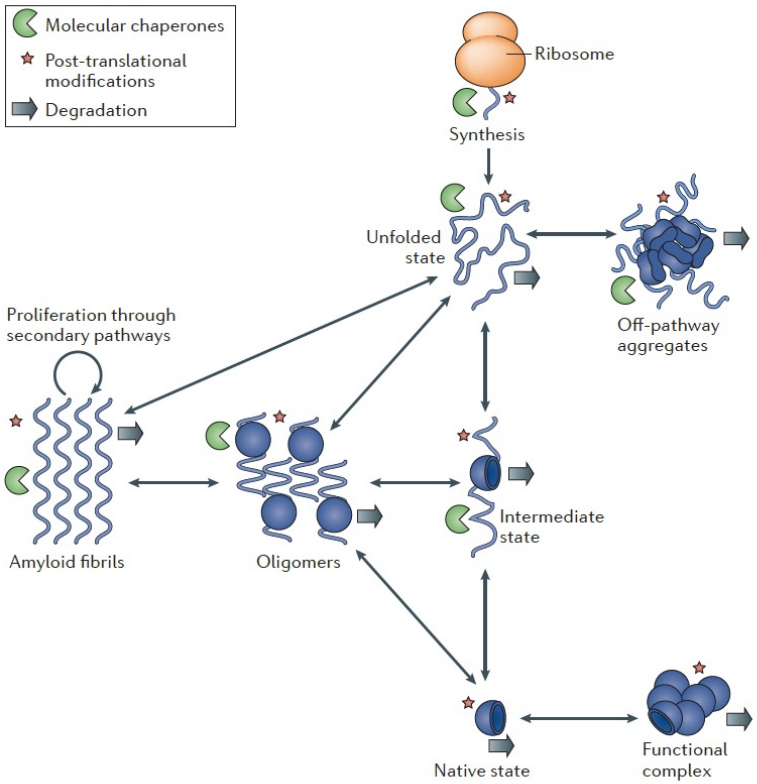


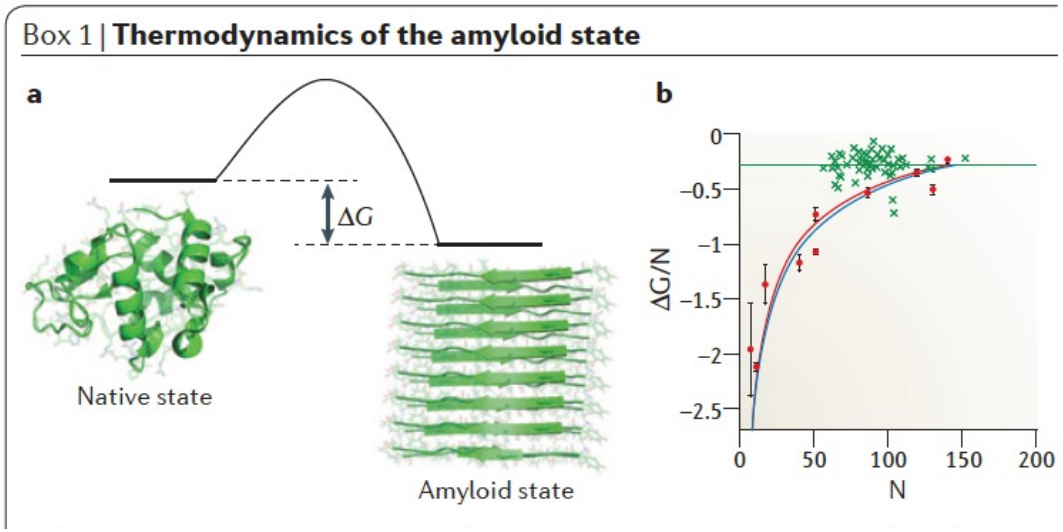
Figure 1 | **A protein can exist in various different states.** The populations of the different states and their interconversion rates are determined by their different thermodynamic stabilities, by the free energy barriers that are associated with the corresponding transitions as well as by the rates of synthesis and degradation, the propensity to interact with chaperones and to undergo post-translational and other chemical modifications. The amyloid state of a protein is a highly ordered form of aggregate in which the polypeptide chains adopt a fibrillar structure, which is capable of self-replication, for example through secondary processes. Amyloid fibrils are rich in β-sheet structure and typically form from unfolded or partially unfolded conformations of proteins and peptides, some of which are fragments of larger proteins. The amyloid state is 'generic' in that its characteristic architecture is not encoded by specific amino acid sequences.

Although most common forms of proteins are soluble, some functional states are insoluble, for example the fibrillar assemblies that form the cellular cytoskeleton or spider silk.

Proteins exhibit generic polymeric and colloidal patterns of behaviour, and many non-biological and synthetic polymers show similarly condensed phases, including, for example, filamentous and particulate gel states.

Aggregated forms of proteins can be generally amorphous on an ultrastructural level, consisting of more or less disordered assemblies of interacting chains of the same or of different sequences.

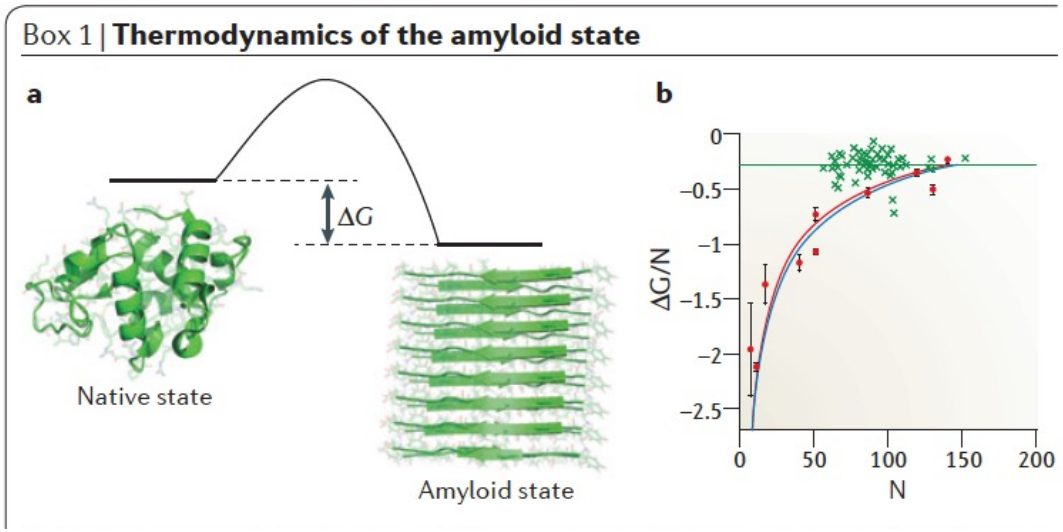
Amyloid form of protein structure



Of particular fascination, because of their remarkable structures and properties, are **highly ordered self-associated species of peptides and proteins, notably amyloid fibrils** and closely related prion-like states.

In the native state of a protein, most of the interactions between amino acid residues are intramolecular, whereas in the amyloid state intermolecular interactions generally dominate. The native state is thermodynamically stable relative to the amyloid state if the free energy (G) of the peptide or protein molecule is lower in the native state than in the amyloid state. As a protein will not spontaneously transition from a state of lower free energy to a state of higher free energy, the conversion into the amyloid state will only take place when its free energy is lower than that of the native state. As the stability of the amyloid state (ΔG) is dependent on the protein concentration, whereas that of the native state (unless it exists in a functional complex) is to a good approximation independent of the protein concentration, there is a concentration at which the stability of the amyloid state is the same as that of the native state; this is the critical concentration. At concentrations exceeding this critical value, a protein is more stable in the amyloid state than in its native state. In such situations, the native state can only persist if there are high free energy barriers that hinder the transition into the more stable amyloid state (see the figure, part **a**). Under such conditions, the native state is then said to be kinetically metastable[41,42]. See the figure, part **b**, for a depiction of experimentally determined standard free energies per residue (ΔG/N, red dots) for a set of peptides and proteins as a function of the length, N, of the polypeptide chain[41]; for reference, the free energy differences between native and denatured (unfolded) states of proteins with sequence lengths in the same range are also shown (green crosses)[41]. Topological constraints associated with the packing of a long polypeptide chain decrease the free energy gain from the conversion to the amyloid form and for polypeptide chains longer than ~150 residues. The standard free energy associated with amyloid formation is on average comparable or more negative than that associated with protein folding. Part **b** adapted with permission from Baldwin, A. J. *et al.* Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* **133**, 14160–14163 (2011). © (2011) American Chemical Society.

Amyloid form of protein structure

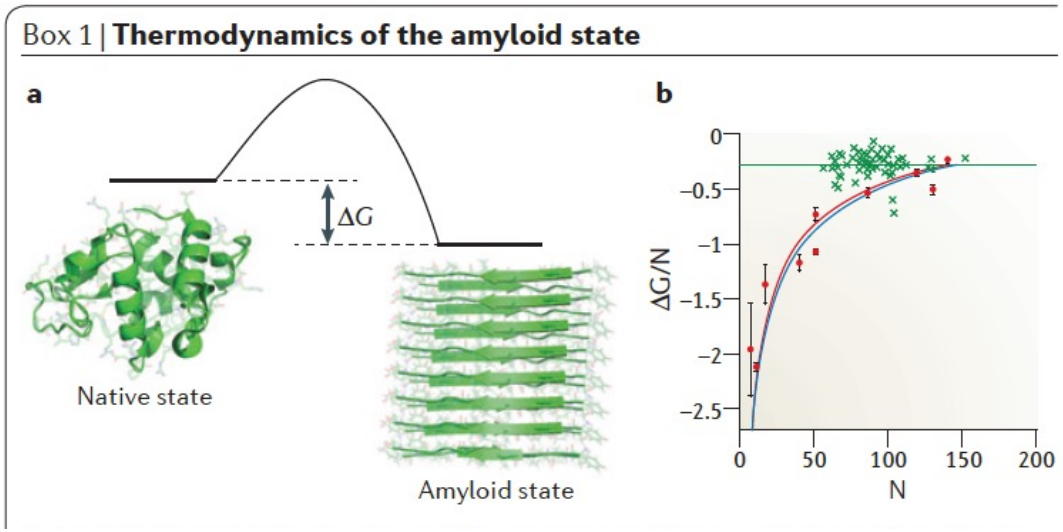


The amyloid state was first observed in the context of systemic amylodidosis more than 150 years ago, and indeed the name ‘amyloid’ means ‘starch-like’, as the deposits observed in the tissues and organs of patients who died from these conditions contained deposits that stained with iodine, which is used to detect starch

In the native state of a protein, most of the interactions between amino acid residues are intramolecular, whereas in the amyloid state intermolecular interactions generally dominate. The native state is thermodynamically stable relative to the amyloid state if the free energy (G) of the peptide or protein molecule is lower in the native state than in the amyloid state. As a protein will not spontaneously transition from a state of lower free energy to a state of higher free energy, the conversion into the amyloid state will only take place when its free energy is lower than that of the native state. As the stability of the amyloid state (ΔG) is dependent on the protein concentration, whereas that of the native state (unless it exists in a functional complex) is to a good approximation independent of the protein concentration, there is a concentration at which the stability of the amyloid state is the same as that of the native state; this is the critical concentration. At concentrations exceeding this critical value, a protein is more stable in the amyloid state than in its native state. In such situations, the native state can only persist if there are high free energy barriers that hinder the transition into the more stable amyloid state (see the figure, part **a**). Under such conditions, the native state is then said to be kinetically metastable[41,42]. See the figure, part **b**, for a depiction of experimentally determined standard free energies per residue (ΔG/N, red dots) for a set of peptides and proteins as a function of the length, N, of the polypeptide chain[41]; for reference, the free energy differences between native and denatured (unfolded) states of proteins with sequence lengths in the same range are also shown (green crosses)[41]. Topological constraints associated with the packing of a long polypeptide chain decrease the free energy gain from the conversion to the amyloid form and for polypeptide chains longer than ~150 residues. The standard free energy associated with amyloid formation is on average comparable or more negative than that associated with protein folding. Part **b** adapted with permission from Baldwin, A. J. *et al.* Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* **133**, 14160–14163 (2011). © (2011) American Chemical Society.

Amyloid form of protein structure

The amyloid structure can in principle be adopted by any polypeptide chain



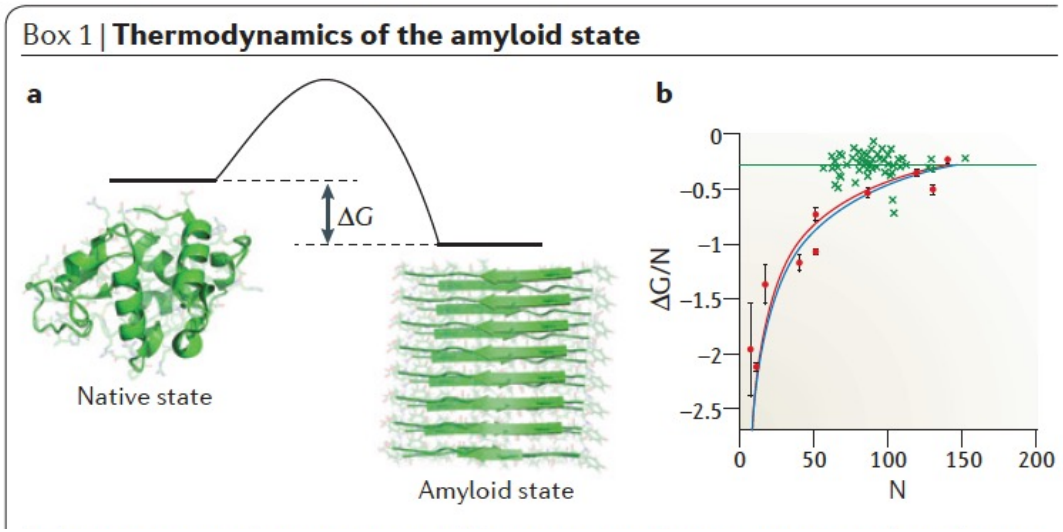
In the native state of a protein, most of the interactions between amino acid residues are intramolecular, whereas in the amyloid state intermolecular interactions generally dominate. The native state is thermodynamically stable relative to the amyloid state if the free energy (G) of the peptide or protein molecule is lower in the native state than in the amyloid state. As a protein will not spontaneously transition from a state of lower free energy to a state of higher free energy, the conversion into the amyloid state will only take place when its free energy is lower than that of the native state. As the stability of the amyloid state (ΔG) is dependent on the protein concentration, whereas that of the native state (unless it exists in a functional complex) is to a good approximation independent of the protein concentration, there is a concentration at which the stability of the amyloid state is the same as that of the native state; this is the critical concentration. At concentrations exceeding this critical value, a protein is more stable in the amyloid state than in its native state. In such situations, the native state can only persist if there are high free energy barriers that hinder the transition into the more stable amyloid state (see the figure, part **a**). Under such conditions, the native state is then said to be kinetically metastable[41,42]. See the figure, part **b**, for a depiction of experimentally determined standard free energies per residue (ΔG/N, red dots) for a set of peptides and proteins as a function of the length, N, of the polypeptide chain[41]; for reference, the free energy differences between native and denatured (unfolded) states of proteins with sequence lengths in the same range are also shown (green crosses)[41]. Topological constraints associated with the packing of a long polypeptide chain decrease the free energy gain from the conversion to the amyloid form and for polypeptide chains longer than ~150 residues. The standard free energy associated with amyloid formation is on average comparable or more negative than that associated with protein folding. Part **b** adapted with permission from Baldwin, A. J. *et al.* Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* **133**, 14160–14163 (2011). © (2011) American Chemical Society.

Such deposits are usually primarily composed of **one protein**, although they are typically associated *in vivo* with various other molecules.

Remarkably, there is **no evident similarity in the sequences**, native structures or functions of the group of disease-associated proteins.

Despite such differences, the corresponding **amyloid fibrils all contain a common 'cross-β' pattern in X-ray fibre diffraction studies that is indicative of the component β-strands being oriented perpendicularly to the fibril axis.**

Amyloid form of protein structure



Such deposits are usually primarily composed of **one protein**, although they are typically associated *in vivo* with various other molecules.

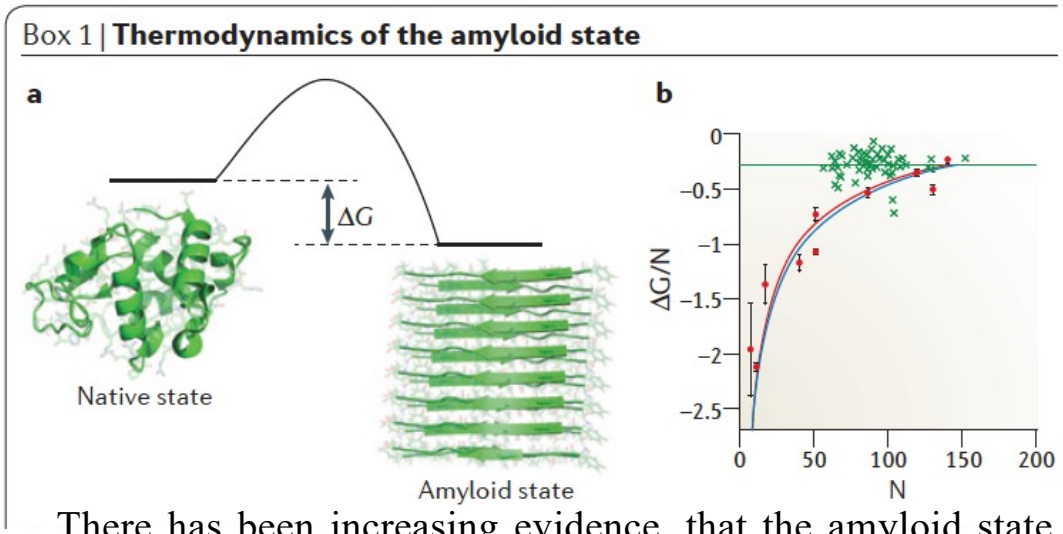
Remarkably, there is **no evident similarity in the sequences**, native structures or functions of the group of disease-associated proteins.

Despite such differences, the corresponding **amyloid fibrils all contain a common 'cross-β' pattern in X-ray fibre diffraction studies that is indicative of the component β-strands being oriented perpendicularly to the fibril axis.**

The amyloid state essential architecture is not encoded by the amino acid sequence, although the details of its structure and stability can be markedly sequence-dependent.

In the native state of a protein, most of the interactions between amino acid residues are intramolecular, whereas in the amyloid state intermolecular interactions generally dominate. The native state is thermodynamically stable relative to the amyloid state if the free energy (G) of the peptide or protein molecule is lower in the native state than in the amyloid state. As a protein will not spontaneously transition from a state of lower free energy to a state of higher free energy, the conversion into the amyloid state will only take place when its free energy is lower than that of the native state. As the stability of the amyloid state (ΔG) is dependent on the protein concentration, whereas that of the native state (unless it exists in a functional complex) is to a good approximation independent of the protein concentration, there is a concentration at which the stability of the amyloid state is the same as that of the native state; this is the critical concentration. At concentrations exceeding this critical value, a protein is more stable in the amyloid state than in its native state. In such situations, the native state can only persist if there are high free energy barriers that hinder the transition into the more stable amyloid state (see the figure, part **a**). Under such conditions, the native state is then said to be kinetically metastable[41,42]. See the figure, part **b**, for a depiction of experimentally determined standard free energies per residue (ΔG/N, red dots) for a set of peptides and proteins as a function of the length, N, of the polypeptide chain[41]; for reference, the free energy differences between native and denatured (unfolded) states of proteins with sequence lengths in the same range are also shown (green crosses)[41]. Topological constraints associated with the packing of a long polypeptide chain decrease the free energy gain from the conversion to the amyloid form and for polypeptide chains longer than ~150 residues. The standard free energy associated with amyloid formation is on average comparable or more negative than that associated with protein folding. Part **b** adapted with permission from Baldwin, A. J. *et al.* Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* **133**, 14160–14163 (2011). © (2011) American Chemical Society.

Amyloid form of protein structure



There has been increasing evidence, that the amyloid state might be **thermodynamically more stable than the functional native states** of many protein molecules even under physiological conditions. As amyloid involves formation of intermolecular contacts, the thermodynamic stability is favoured at higher concentrations. The critical concentration above which the stability of the amyloid state exceeds the soluble state may be lower than the physiological concentration of a given protein *in vivo*, which suggests the possibility that some, and perhaps many, proteins might routinely function at concentrations higher than for their thermodynamic solubility .

In the native state of a protein, most of the interactions between amino acid residues are intramolecular, whereas in the amyloid state intermolecular interactions generally dominate. The native state is thermodynamically stable relative to the amyloid state if the free energy (G) of the peptide or protein molecule is lower in the native state than in the amyloid state. As a protein will not spontaneously transition from a state of lower free energy to a state of higher free energy, the conversion into the amyloid state will only take place when its free energy is lower than that of the native state. As the stability of the amyloid state (ΔG) is dependent on the protein concentration, whereas that of the native state (unless it exists in a functional complex) is to a good approximation independent of the protein concentration, there is a concentration at which the stability of the amyloid state is the same as that of the native state; this is the critical concentration. At concentrations exceeding this critical value, a protein is more stable in the amyloid state than in its native state. In such situations, the native state can only persist if there are high free energy barriers that hinder the transition into the more stable amyloid state (see the figure, part **a**). Under such conditions, the native state is then said to be kinetically metastable[41,42]. See the figure, part **b**, for a depiction of experimentally determined standard free energies per residue (ΔG/N, red dots) for a set of peptides and proteins as a function of the length, N, of the polypeptide chain[41]; for reference, the free energy differences between native and denatured (unfolded) states of proteins with sequence lengths in the same range are also shown (green crosses)[41]. Topological constraints associated with the packing of a long polypeptide chain decrease the free energy gain from the conversion to the amyloid form and for polypeptide chains longer than ~150 residues. The standard free energy associated with amyloid formation is on average comparable or more negative than that associated with protein folding. Part **b** adapted with permission from Baldwin, A. J. *et al.* Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* **133**, 14160–14163 (2011). © (2011) American Chemical Society.

Amyloid fibrils from different proteins seem to be remarkably similar at the nanometre length scale. Indeed, under electron microscopy or atomic force microscopy (AFM), amyloid fibrils tend to appear as unbranched filamentous structures only a few nanometers in diameter but often micrometres in length. They are typically observed to consist of multiple protofilaments that twist around each other to form mature fibrils. X-ray fibre diffraction studies indicate **that the core of each protofilament adopts a cross-β structure, in which β-strands form effectively continuous hydrogen-bonded β-sheets that run along the length of the fibril**

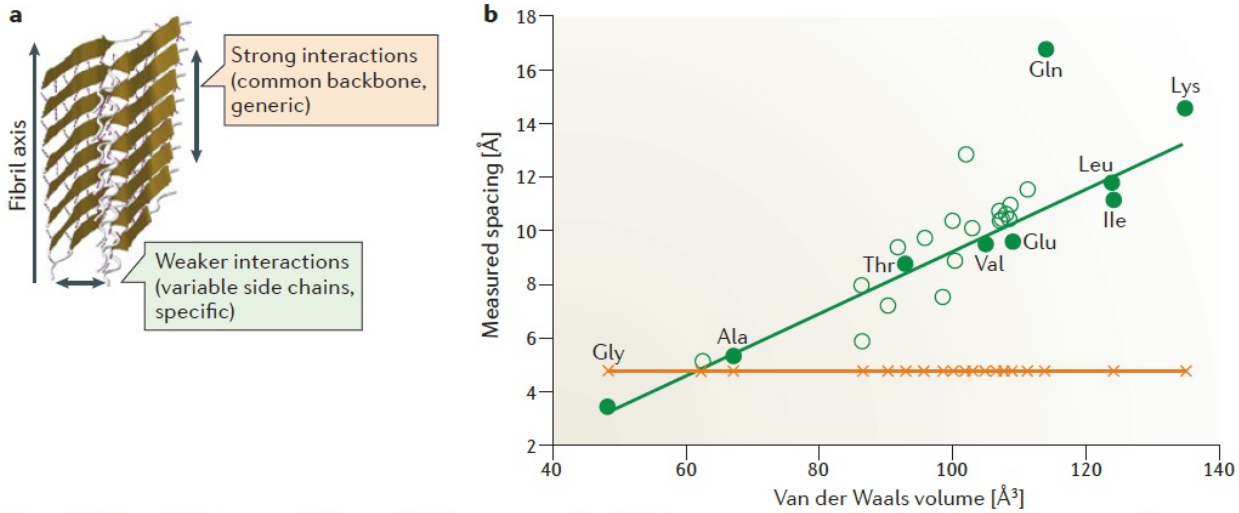


Figure 2 | **Generic features of the amyloid structure.** Amyloid structures exhibit a range of specific features and possess common characteristics. Representation of the 'cross-β' structure common to amyloid fibrils (part **a**). The spacing between polypeptide chains along the fibril axis is constant to a good approximation even for very different polypeptide sequences, a generic property arising from the common inter-main chain hydrogen bonding constraints (orange line in part **b**). By contrast, the spacing between the β -sheets in the direction perpendicular to the fibril axis is highly dependent on the nature of the side chains and originates from variable packing constraints. Correlation of the inter-sheet spacing (green circles; green line indicates specific properties) with the average van der Waals volumes of the side chains for a range of different amyloid fibrils (part **b**); filled green circles indicate fibrils that are formed by homopolymers (that is, polypeptide repeats of the same type of amino acid residue)[31]. Part **b** reprinted with permission from REF. 31, Wiley.

NMR and Cryo-EM studies confirm the generic nature of the overall structures of different fibrils, which can be attributed to the common properties of the polypeptide backbone that support the hydrogen bonding pattern in the fibril core and to variations that result from the manner in which the different sets of side chains are incorporated into the common fibrillar architecture.

The cross-β architecture provides very great stability to the fibrils, as it allows the formation of a **continuous array of hydrogen bonds**

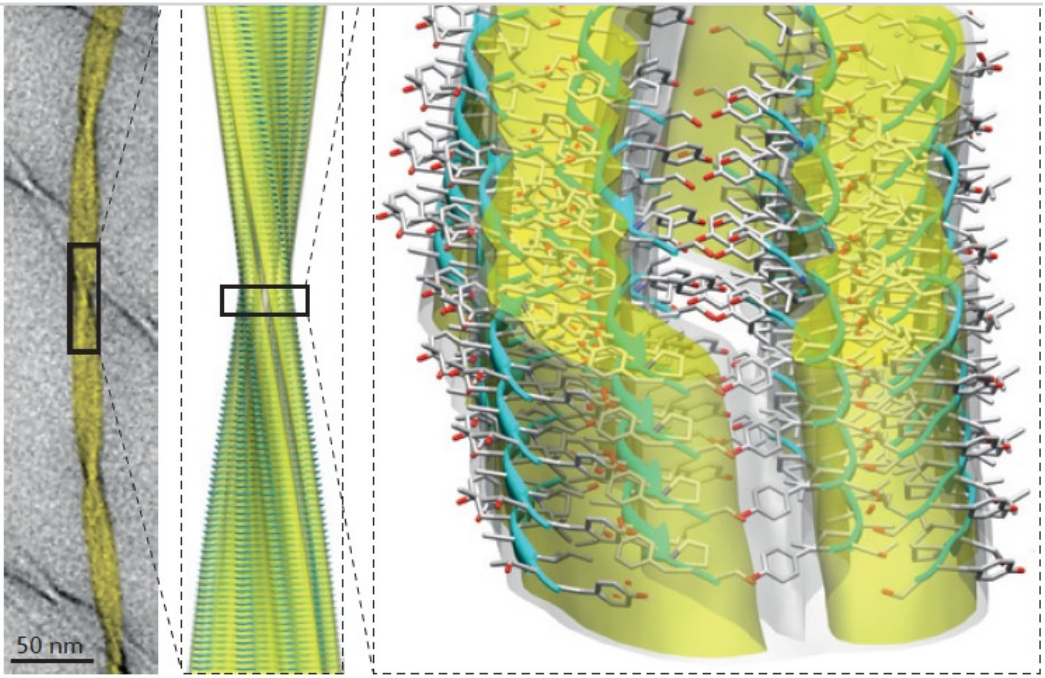


Figure 3 | **Structure of an amyloid fibril at atomic resolution.** The structure shown is of one of several polymorphs of the amyloid fibrils that are formed from an 11-residue fragment of transthyretin[75]. The combination of cryo-electron microscopy imaging (part **a**) with solid-state NMR analysis has enabled the determination of an atomic-level structure (part **b**). A more detailed view (part **c**) shows the hierarchical organization of this amyloid fibril in which the three filaments that form the mature fibril illustrated here are in turn formed by pairs of cross-β protofilaments, which are each composed of pairs of β-sheets[75]. The fibril surfaces are shown as electron density maps, and the constituent β-sheets are shown in a ribbon representation; oxygen, carbon, and nitrogen atoms are shown in red, gray and blue, respectively. Adapted with permission from Fitzpatrick, A. W. P. *et al.* Atomic structure and hierarchical assembly of a cross-β

The common cross-β architecture of amyloid structures originates from the **universal propensity of polypeptide chains to form backbone hydrogen bonding** — a propensity **that**, in a condensed state, **can be most readily accommodated through extended intermolecular β-sheets**. The lateral packing of such β-sheets, however, relies on specific patterns of interactions between side chains that depend on the amino acid sequences of the component proteins.

Fibrillar structures have remarkable mechanical properties, including a high Young's modulus and tensile strength, which arise in large part from the hydrogen bonding network of the cross-β structure (intermediate between carbon nanotubes, and actin- tubulin)

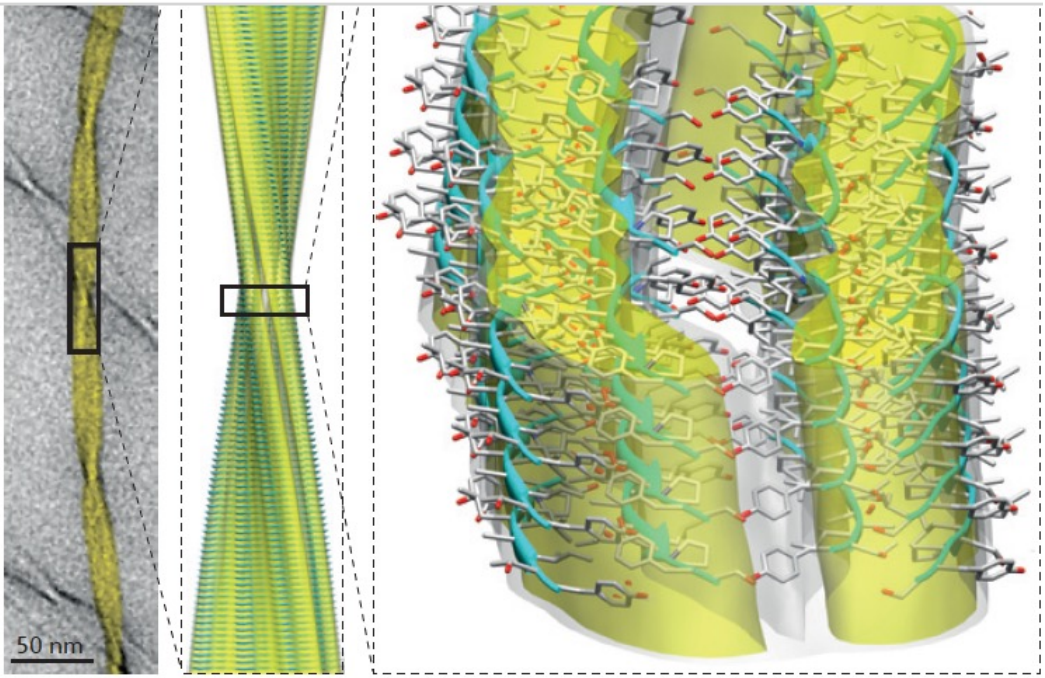
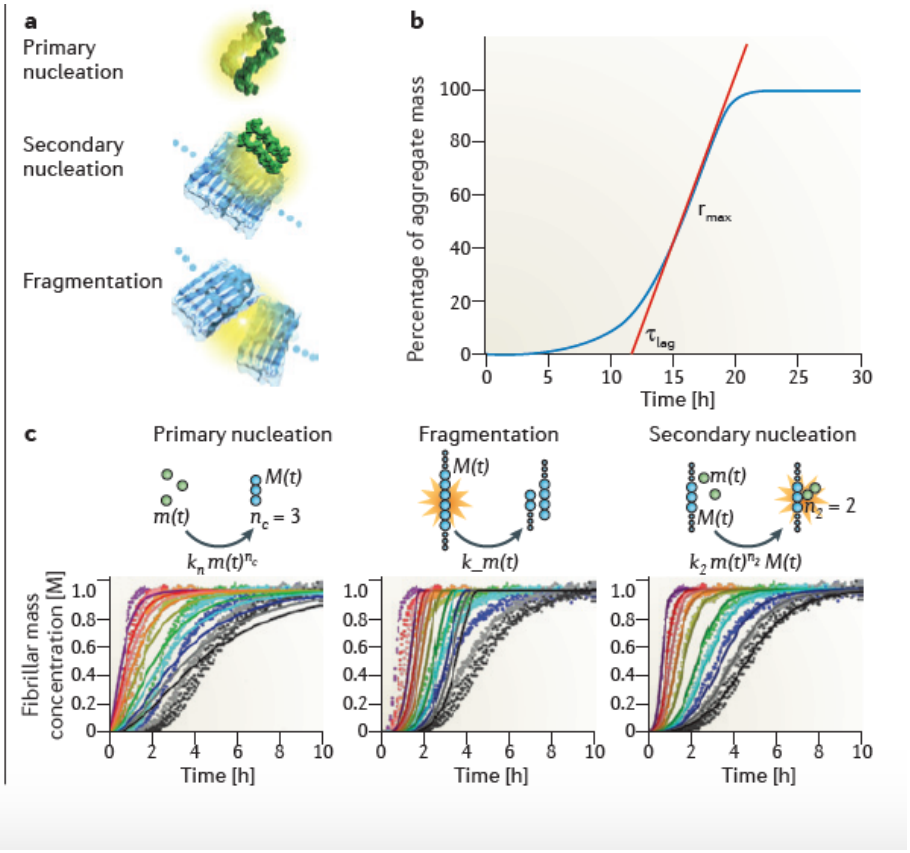


Figure 3 | **Structure of an amyloid fibril at atomic resolution.** The structure shown is of one of several polymorphs of the amyloid fibrils that are formed from an 11-residue fragment of transthyretin[75]. The combination of cryo-electron microscopy imaging (part **a**) with solid-state NMR analysis has enabled the determination of an atomic-level structure (part **b**). A more detailed view (part **c**) shows the hierarchical organization of this amyloid fibril in which the three filaments that form the mature fibril illustrated here are in turn formed by pairs of cross-β protofilaments, which are each composed of pairs of β-sheets[75]. The fibril surfaces are shown as electron density maps, and the constituent β-sheets are shown in a ribbon representation; oxygen, carbon, and nitrogen atoms are shown in red, gray and blue, respectively. Adapted with permission from Fitzpatrick, A. W. P. *et al.* Atomic structure and hierarchical assembly of a cross-β



Box 2 | Kinetics of amyloid formation

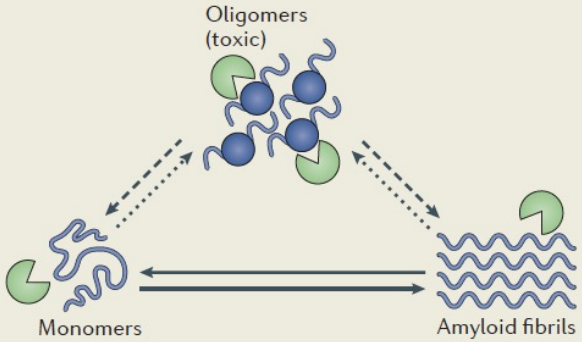
Whereas thermodynamics describes whether or not a transition from one state to another is spontaneous and can occur without an external driving force, kinetics addresses the question of how fast such a transformation will take place. Kinetic measurements have historically been one of the main tools for elucidating reaction mechanisms for both small-molecule reactions and for enzymology[174]. As aggregate populations observed during amyloid formation are heterogeneous, a convenient tool for describing the kinetics of amyloid formation is the master equation formalism[99,100]. This formalism connects the microscopic steps that underlie amyloid formation with their macroscopic manifestations by considering the different ways in which new aggregates can be formed: for example, from monomer through primary nucleation (see the figure, part **a**; top; monomers are shown in green); from existing fibrils through fragmentation (see the figure, part **a**; bottom; fibrils are shown in blue); or from a combination of both monomeric and aggregated species through secondary nucleation (see the figure, part **a**; centre). Integrated rate laws that are obtained for filament growth processes that occur under the action of these various microscopic processes commonly take the form of sigmoidal functions (see the figure, part **b**; blue line), which are characterized by a lag time (τ_{lag}) and a maximal growth rate (r_{max}; red line). Comparison of such integrated rate laws with experimental kinetic measurements enables the relative importance of specific microscopic processes to be tested. In this example (see the figure, part **c**) the aggregation kinetics for increasing concentrations (see the figure, part **c**; coloured lines) of the amyloid-β peptide are compared with integrated rate laws that contain primary nucleation (see the figure, part **c**; left), fragmentation (see the figure, part **c**; centre) and monomer-dependent secondary nucleation (see the figure, part **c**; right). The data show that secondary nucleation is the dominant process under these conditions (n_c and n_2 indicate the sizes of the critical nuclei for primary and secondary nucleation, respectively; k_n and k_+ indicate the rates of primary nucleation and fragmentation, respectively; and m and M indicate the concentrations of the monomers and the fibril mass, respectively), at time t. To avoid overfitting, global analysis approaches have emerged as powerful tools[99,100]. In this strategy, a known parameter, most commonly the concentration of a monomeric protein or peptide at the beginning of the aggregation reaction is varied in a known manner, and then the entire data set in terms of both time and concentration is fitted to a single rate law. More details can be found in the literature[99,100]. Parts **a** and **c** adapted with permission from Cohen, S. I. A. *et al.* Proliferation of amyloid-β 42 aggregates occurs through a secondary nucleation mechanism. *Proc. Natl Acad. Sci. USA* **110**, 9758–9763 (2013). Part **a** © (2005) National Academy of Sciences, USA. Lührs, T. *et al.* 3D structure of Alzheimer's amyloid-β(1-42) fibrils. *Proc. Natl Acad. Sci. USA* **102**, 17342–17347 (2005). Part **b** from Knowles, T. P. J. *et al.* An analytical solution to the kinetics of breakable

Typical sigmoidal reaction time course.

This type of behaviour, in which a lag phase is observed before a rapid growth phase, is a feature of nucleated polymerization; in cases in which the total quantity of protein is limited, the growth phase is followed by a plateau phase in which the reaction rate declines as a result of the depletion of the soluble species that is being monitored as it converts into fibrils. When monomers add to the ends of the fibrils, they adopt the cross-β conformation to match that of the peptides already present in the aggregate that hence function as templates. Indeed, when pre-formed aggregates are added to a solution during its lag phase, rapid fibril formation can take place as a result of seeding, a process analogous to that familiar from studies of crystallization phenomena. Thus, the conversion of a protein molecule from its soluble state into the amyloid form can be triggered by nucleation, as well as by templating or seeding from existing aggregates.

The maintenance of biological functionality. As proteins in their functional forms can be thermodynamically and chemically metastable, mechanisms have evolved to maintain their solubility *in vivo* for prolonged periods of time and to avoid their conversion into non-functional amyloid states. Exceptions are cases in which the amyloid state is functiona. Thus, for example, Pmel17 is a highly aggregation-prone protein that forms functional amyloid structures that are involved in melanosome biogenesis, and it has been found that certain peptide and protein hormones in secretory granules of the endocrine system are stored in cross-β conformations. Rather few such examples have, however, been observed, particularly in mammalian systems, and functional amyloid formation is undoubtedly very carefully regulated.

Box 3 | **Strategies for therapeutic intervention**



The maintenance of protein homeostasis and the avoidance of protein metastasis can be achieved by regulating the concentrations of the different states of all the proteins that are contained within a living system and by controlling the rates of conversion between them. To this end, it is particularly important to reduce the population of oligomeric species (pre-fibrillar species) by disrupting the processes of their formation (see the figure; dashed arrows) or by promoting the pathways of their removal (dotted arrows). These strategies can be implemented in various ways[166]; for example, through the modulation of production processes (targeting synthesis or proteolysis)[175], degradation (targeting the ubiquitin–proteasome[126,127] and the autophagy[129,130] systems) or stability (usually targeting the native state)[167,168]. For example, one can use the ability of antibodies[176,177], or artificially-generated analogues such as affibodies[178], to selectively bind to the native states of aggregation-prone proteins, as binding generally results in increased stability and hence to a reduction in aggregation propensity. In some cases, it might be possible to use antibody-based immunotherapy approaches to reduce the level of highly aggregation-prone species (such as amyloid-β) by stimulating their clearance[179,180]. Antibodies and their analogues also offer other possibilities, one of which could be to mimic the action of natural molecular chaperones by targeting the aberrant misfolded species that give rise to cellular damage. If such 'artificial chaperones' can be developed and can be targeted to the appropriate location (for example, by enhancing their ability to cross the blood–brain barrier), then they could represent a highly effective therapy[3,43]. Furthermore, it is also becoming evident that small molecules can also function to suppress the early stages of protein aggregation, for example by binding to specific amyloidogenic species and by reducing the risk of nucleation and proliferation of pathogenic agents[168–172].