Module **BCH 304**

Protein Biophysics

(Part I)

Amedeo Caflisch

Biochemisches Institut der
Universität Zürich
Winterthurerstrasse 190
CH-8057 Zürich

email: caflisch@bioc.uzh.ch

[http://www.biochem-caflisch.uzh.ch](http://www.biochem-caflisch.uzh.ch)

2016
1 Introduction

Proteins are fascinating macromolecules. They fulfill or control most of the functions necessary for life. Examples of protein functions are: transport of oxygen from the lungs to muscles (hemoglobin), muscle stretching (titin), catalysis of biochemical reactions (enzymes), recognition of viral agents (antibodies), signal transduction (receptor tyrosine kinases), potentiation of synaptic connection for long-term memory (CREB, cyclic-AMP response element-binding protein), packaging of DNA into chromatin (histones), regulation of transcription and translation. Some of these functions (e.g., regulation of transcription) are common to all living beings. Importantly, protein folding as well as protein/protein and protein/ligand recognition are governed by non-covalent interactions and the hydrophobic effect which are the main topics of part I of this course.

1.1 Schedule and Contents

Monday 8:15-10:00
Tuesday 8:15-10:00

I Non-covalent interactions; protein structure, flexibility, folding and aggregation
- Electrostatics, van der Waals, hydrogen bonds
- Properties of water and hydration
- Hydrophobic effect
- Patterns of folding and association of polypeptide chains
- Protein misfolding and aggregation
- The "folding problem"
- Energy landscapes and folding funnels
- Protein flexibility and conformational transitions
- Theoretical and computational approaches
- Molecular dynamics simulations

II Protein dynamics and folding: NMR spectroscopy (Oliver Zerbe)

III Thermodynamics and kinetics of protein folding (Ilian Jelesarov)

IV Single molecule biophysics and stochastic processes (Ben Schuler)

IMPORTANT: Proficiency certificate (= Leistungsnachweis):

Assessment for part I: fourth week of semester.
1.2 Suggested reading

The following books and papers are just suggestions. Note that the present script contains all of the material you need to know for part I (Caflisch) of the Protein Biophysics lecture. Moreover, you should not hesitate to ask the teachers for research papers on specific topics.

Protein structure


Methods for determining protein structures


Molecular dynamics simulations


Protein folding problem

- A. Caflisch and P. Hamm, Complexity in protein folding: Simulation meets experiment, Current Physical Chemistry 2, 4-11, 2012. [link](#)
Contents

1 Introduction 1
   1.1 Schedule and Contents ........................................... 1
   1.2 Suggested reading ............................................... 2

2 Interactions between atoms 5
   2.1 Classification of interactions .................................. 5
   2.2 Electrostatic interactions in molecules: an overview .......... 7
   2.3 Dipole-dipole interactions ...................................... 10
   2.4 Dipole-induced dipole interactions .............................. 11
   2.5 Van der Waals interactions (induced dipole-induced dipole) 12
   2.6 Hydrogen bond ................................................... 14
   2.7 Properties of water .............................................. 18
   2.8 Hydration ....................................................... 21
   2.9 Hydrophobic effect ............................................... 23
   2.10 Micelles and membranes (movies) ............................... 26
   2.11 Hydrophobic effect and membrane proteins .................... 27

3 Protein structure 29
   3.1 The 20 natural amino acids and their role in proteins ....... 29
   3.2 Dihedral angles and Ramachandran plot ........................ 30
   3.3 Secondary structure ............................................. 35
       3.3.1 Regular and irregular secondary structure ............... 35
       3.3.2 Supersecondary structure ................................... 38
   3.4 Tertiary structure ............................................... 41
       3.4.1 Anatomy of protein structures .............................. 41
       3.4.2 Taxonomy of protein structures ............................ 42
   3.5 Quaternary structure ............................................ 44
   3.6 Experimental approaches for structural studies: Overview .... 46

4 Ordered aggregation and amyloid fibrils 49
   4.1 Historical note and future perspectives ........................ 49
   4.2 Amyloid fibrils .................................................. 50
   4.3 Kinetics of fibril formation ..................................... 53
   4.4 A unique fibrillar structure? .................................... 57
   4.5 Amyloid related diseases ........................................ 60
       4.5.1 Prion protein ................................................. 61
       4.5.2 Alzheimer's disease ....................................... 62
   4.6 Amyloid fibrils: toxic or functional? .......................... 64
   4.7 Monte Carlo simulations of monomeric Aβ40 and Aβ42 (Excursus) 65
5 Protein folding

5.1 The funnel model of folding: a simple picture of a complex process 68
5.2 Protein folding mechanisms 69
5.3 Folding kinetics 70
5.4 Two-state folding: An approximation of a complex process 71
5.5 Is folding really two-state? The unfolded state is complex! 72
5.6 Protein denaturation in vitro 73
5.7 Experimental and computational approaches to folding 74

6 Molecular dynamics

6.1 Biophysical techniques 75
6.2 Simulations 76
2 Non-covalent interactions

... all things are made of atoms, little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another.


2.1 Classification of interactions

Since atoms consist of charged particles, the interatomic forces are mainly of electrostatic nature. They depend on: (1) charge state (i.e., neutral atoms or ions) and (2) electronic structure, i.e., outermost electron shell (Table 2.1).

Closed-shell atoms:
He, Li⁺, F⁻, Ne, Na⁺, Cl⁻, Ar, Zn²⁺, Cd²⁺, Ca²⁺, Mg²⁺.
Open-shell atoms:

Table 2.1 Overview of chemical bonds

<table>
<thead>
<tr>
<th>outer electron shell</th>
<th>charge state</th>
<th>example</th>
<th>type of bond</th>
</tr>
</thead>
<tbody>
<tr>
<td>open-shell</td>
<td>neutral</td>
<td>H–H</td>
<td>covalent (H–I is 5% ionic)</td>
</tr>
<tr>
<td>closed-shell</td>
<td>neutral</td>
<td>He·····He</td>
<td>van der Waals</td>
</tr>
<tr>
<td>open-shell M²⁺</td>
<td>charged</td>
<td>M²⁺(X⁻)₆</td>
<td>transition-metal complex</td>
</tr>
<tr>
<td>(closed-shell X⁻)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>closed-shell</td>
<td>charged</td>
<td>Li⁺F⁻</td>
<td>ionic 92% (covalent 8%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Na⁺Cl⁻</td>
<td>ionic 67% (covalent 33%)</td>
</tr>
</tbody>
</table>

This classification is useful but not all bonds can be uniquely assigned. For example, the bond between H and F (both open-shell atoms but with electronegativity values of 2.1 and 4.0, respectively) has a covalent and ionic character of 55% and 45%, respectively.

Pairs of atoms in (macro)molecules interact by local interactions (called also bonding interactions, e.g., covalent bonds), and non-local (non-bonding) interactions, viz., van der Waals and Coulombic interactions. The two latter interactions govern also (macro)molecular recognition, e.g., substrate (or inhibitor) binding to an enzyme, antigen/antibody association, binding of ligands to receptors. Thus, (macro)molecular recognition is mainly non-covalent (Figure 2.1).
Examples

Fig. 2.1 The binding mode of the drugs ruxolitinib (left) and SAR302503 (right) to the JAK2 kinase domain (Leukemia 28, 404, 2014).

Click on the following links to visualize examples of binding/unbinding processes.
Movie of binding of natural ligand (acetyl-lysine) to a human bromodomain: http://www.biochem-caflisch.uzh.ch/movie/22/

Movie of unbinding of small molecule from the FKBP protein: http://www.biochem-caflisch.uzh.ch/movie/7/
2 Non-covalent interactions

2.2 Electrostatic interactions in molecules: an overview

Most interactions in molecules are electrostatic. In a medium with dielectric constant $\epsilon$ the Coulombic energy $E$ (in kcal/mol) of a point charge $q_i$ (in electronic charge units $e_0$) at distance $r$ (in Å) from a point charge $q_j$ is

$$E = \frac{332 q_i q_j}{\epsilon r}$$

Energy units: 1 Electronvolt = 1 eV = 23 kcal/mol.

An electric dipole consists of two charges $q$ and $-q$ separated by a distance $l$. The dipole is represented by a vector $\vec{\mu}$ which points from the negative to the positive charge, $-q \rightarrow q$. Dipole moment unit: 1 Debye = 1 D = 0.208 e$_0$Å.

A polar molecule (e.g., water) has an electric dipole moment originating from the partial charges on its atoms. Apolar molecules do not have a permanent electric dipole moment, but a dipole can be induced by an external electric field upon distortion of the electron distribution with respect to the atomic nuclei.

**Excursus on electric field:** A field is a physical quantity that assumes different values at different points in space. An important equation is field = −gradient(potential). From this equation it follows that the field vectors are orthogonal to the equipotential surfaces (which you can draw in each of the illustrations of Figure 2.2). The electrostatic potential has a physical meaning: It is the potential energy that a unit charge has if brought to the specified point in space from infinite. In the equation (1) above the electrostatic potential $\phi$ of the charge $q_j$ is $\phi(r) = \frac{332q_j}{\epsilon r}$ where $r$ is the distance. Note that the field of a monopole is proportional to $q/r^2$ because of the radial symmetry and the definition of gradient (vector of first derivatives).

![Fig. 2.2. Illustration of electric fields of monopoles, dipoles and charged surfaces.](image-url)
**Multipole expansion:** A multipole expansion is a mathematical series that represents a function. These series are truncated if the first few terms are a good approximation to the original function.

The multipole expansion of the electric field provides an exact description of the electrostatic potential and converges if the sources (e.g., charges) are localized close to the origin. In this case, the coefficients of the series expansion are called multipole moments. The zeroth-order term in the expansion is called the monopole moment ($2^0$-pole), the first-order term is denoted as the dipole moment ($2^1$-pole), and the second-order and third-order terms are called quadrupole ($2^2$-pole) and octupole ($2^3$-pole), respectively (Figure 2.3a). Each multipole moment is determined by the charge distribution, i.e., the positions and magnitudes of the (partial) charges.

Some interesting considerations on molecules, electric fields, and dipoles.

- In an external electric field the dominant molecular multipole is the dipole (Figure 2.3b).
- Permanent and induced dipole moments are important in biochemistry. (Many examples will be presented in the next sections.)
- All heteronuclear two-atomic molecules are polar, because the difference of electronegativity of both atoms results in partial charges (e.g., 1.08 D in HCl).
- Depending on the symmetry, a multiatom molecule can be polar. Water has a large electric dipole moment (1.85 D), whereas CO$_2$ does not because its three atoms are on a line and the two dipoles point in opposite directions.

**Table 2.2 Distance dependence of multipole interactions**

<table>
<thead>
<tr>
<th>Type of interaction</th>
<th>Distance-dependence</th>
<th>Typical energy values [kcal/mol]</th>
<th>Examples in proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monopole–monopole</td>
<td>$1/r$</td>
<td>-50 to -5</td>
<td>salt bridges</td>
</tr>
<tr>
<td>Monopole–dipole</td>
<td>$1/r^2$</td>
<td>-3.5</td>
<td>Lys-ammonium and α-helix dipole</td>
</tr>
<tr>
<td>Dipole–dipole</td>
<td>$1/r^3$</td>
<td>-0.5</td>
<td>backbone C=O in α-helix</td>
</tr>
<tr>
<td>Dispersion</td>
<td>$1/r^6$</td>
<td>-0.1</td>
<td>all atoms</td>
</tr>
</tbody>
</table>

In general, for the interaction between a permanent $2^n$-pole and a permanent $2^m$-pole at distance $r$ one can write $E \propto \frac{1}{r^{n+m+1}}$ (examples are listed in Table 2.2).

Interactions involving induced dipoles have a different distance dependence (see below).

**Exercise 1:** Why is the electric dipole moment of benzene equal zero? Does chlorobenzene have a dipole moment? And 1,4-dichlorobenzene? Is the dipole moment of 1,2-dichlorobenzene higher than the one of 1,3-dichlorobenzene?
Fig. 2.3a Schematic representation of electric multipoles.

Fig. 2.3b Schematic representation of alignment of dipoles in an external electric field.
2.3 Dipole-dipole interactions

In a medium with relative dielectric constant $\epsilon$ the interaction energy (kcal/mol) between two aligned dipoles $\mu_i = q_i l$ and $\mu_j = q_j l$ at distance $r$ with $r \gg l$ is

$$E = -664 \frac{\mu_i \mu_j}{\epsilon r^3}$$

Derivation:

The sum of the four contributions is

$$E = \frac{332}{\epsilon} \left( -\frac{q_i q_j}{r + l} + \frac{q_i q_j}{r} + \frac{q_i q_j}{r - l} \right) = -332 \frac{q_i q_j}{\epsilon r} \left( \frac{1}{1 + x} - 2 + \frac{1}{1 - x} \right)$$

with $x = l/r$. For $r \gg l$ one has that $x \rightarrow 0$:

$$\frac{1}{1 + x} = (1 + x)^{-1} = 1 - x + x^2 + O(x^3)$$

It follows that

$$E = -332 \frac{q_i q_j 2x^2}{\epsilon r} = -664 \frac{q_i l q_j l}{\epsilon r^3}$$

For a given fixed orientation the interaction energy between dipoles $\vec{\mu}_i$ and $\vec{\mu}_j$ at distance $r_{ij}$ (much larger than the dipole length) is

$$E = \frac{332}{\epsilon} \left[ \frac{(\vec{\mu}_i \cdot \vec{\mu}_j)}{r_{ij}^3} - 3(\vec{\mu}_i \cdot \vec{r}_{ij})(\vec{\mu}_j \cdot \vec{r}_{ij}) \right]$$

(3)

For $\epsilon = 4$ the energy between two fixed dipoles of 1 D each at distance 5.0 Å is

$\leftrightarrow$ $E = -1.32$ kcal/mol; $\rightarrow\leftarrow$ $E = +1.32$ kcal/mol;

$\downarrow \uparrow$ $E = -0.66$ kcal/mol; $\uparrow \uparrow$ $E = +0.66$ kcal/mol.

**Exercise 2:** Use the above equation to calculate $E$ for the four configurations of the dipole pairs shown above. Make a schematic drawing of the dipole moment of a peptide bond $\text{CH}_3$-$\text{CO}$-$\text{NH}$-$\text{CH}_3$.

**Exercise 3:** Which structural elements in proteins are rich in dipole–dipole interactions?
2 Non-covalent interactions

2.4 Dipole-induced dipole interactions

Even if a molecule does not have a permanent electric dipole moment (e.g., CH$_4$), an external electric field can induce a dipole moment on it. A polar molecule with dipole moment $\mu_i$ can induce a dipole moment on a polarizable molecule with polarizability $\alpha_j$ at distance $r$ and the two are attracted by a favorable energy

$$E \propto -\frac{\mu_i^2 \alpha_j}{r^6}$$  \hspace{1cm} (4)

The dipole–induced dipole interaction energy is independent of the temperature because thermal motion has no effect on the averaging process. In other words, the induced dipole always follows the direction of the inducing dipole; dipoles remain aligned however fast the molecules tumble.

The distance dependence stems from the $1/r^3$ dependence of the field of the permanent dipole and the $1/r^3$ dependence of the potential energy of interaction between the permanent dipole and induced dipole:

$$E \propto -\frac{\mu_i \mu_j}{r^3} = -\frac{\mu_i \alpha_j |D_i|}{r^3}$$  \hspace{1cm} (5)

where $|D_i| \propto \mu_i/r^3$ is the magnitude of the field of the polar molecule $i$.

Analogously, considering that the magnitude of the field of a monopole is proportional to $q/r^2$ one can derive the interaction energy between a ion of charge $q$ and a molecule with polarizability $\alpha$ which is the monopole-induced dipole interaction $E \propto q^2 \alpha/r^4$ (using $n=0$ and $m=1$ in the equation $E \propto \frac{q^2 \alpha}{r^{n+m+1}}$).
2.5 Van der Waals interactions (induced dipole-induced dipole)

The dispersion interaction (also called van der Waals attraction or London) is proportional to \( r^{-6} \), where \( r \) is the distance between two atoms. The attractive dispersion interaction is balanced by the electron repulsion (Pauli) and the latter dominates at short distances. For proteins one can use the \((6,12)\) potential as an approximation of the van der Waals energy. The sum of repulsive and attractive interactions is approximated by

\[
E_{vdW} = E_{\text{min}} \left[ \left( \frac{r_{\text{min}}}{r} \right)^{12} - 2 \left( \frac{r_{\text{min}}}{r} \right)^{6} \right]
\]

where \( E_{\text{min}} \) is the minimum of the energy at \( r = r_{\text{min}} \) (Fig. 2.4).

![van der Waals Energy](image)

Fig. 2.4 van der Waals energy for the dispersion interaction and Pauli-repulsion.

**Exercise 4.1:** Calculate the first derivative of the van der Waals energy and show that the minimum of \( E_{vdW} \) is at \( r = r_{\text{min}} \). At what distance \( r \) is the favorable energy term equal to the unfavorable one, i.e., \( E_{vdW}(r) = 0 \)? What is the qualitative behavior of the van der Waals energy at \( r = 2r_{\text{min}} \) and \( r = 0.8r_{\text{min}} \)?

**Exercise 4.2:** Why is the van der Waals energy important for protein stability? (Hint: compare protein-protein and protein-water interactions.)

**Exercise 4.3:** Butter is solid whereas olive oil is liquid at room temperature. Why? (Hint: Why do saturated fatty acids have higher melting points than unsaturated?)
**Exercise 4.4:** The side chains are densely packed, i.e., pairs of atoms are at optimal van der Waals distance, in the hydrophobic cores of globular proteins (top, left). Explain the correlation between protein destabilization upon mutation ($\Delta \Delta G_{D-N}$) and the number of neighbors ($N$) of a given side chain. [Hints: The $\Delta \Delta G_{D-N}$ value originates mainly from the reduced stability of the folded state upon mutations that deteriorate the optimal packing. The change in free energy of the unfolded state upon mutation can be neglected for these single-point and double-point mutants.] From Cota et al., J. Mol. Biol. 302, 713, 2000. TNfn3 = Third fibronectin (fn) domain of human tenascin (TN), which is an extracellular matrix glycoprotein.

(Top, right) The black bars refer to TNfn3 data while the white bars refer to a closely related fibronectin type domain which shows a different behavior.
2.6 Hydrogen bond

The hydrogen bond consists of a hydrogen atom between two strong electronegative atoms. One can describe it as a consequence of the Coulombic interaction between the positive partial charge of the H atom that is bound to an electron-withdrawing "donor" (D), and the lone pair of electrons at the "acceptor" atom (A).

\[ \delta^{-}D-H^{\delta+} \cdots \delta^{-}A \]

where D, A = N, O and F.

Note that the D (donor) and A (acceptor) atoms are at shorter distance than the optimal van der Waals distance (Table 2.3). The hydrogen bond energy depends on the geometry of D, H, and A. The optimal D–H⋯A angle is 180°, while the optimal H⋯A–AA angle (AA = "anterior acceptor") depends on the D, A and AA elements and the hybridization state. For instance, the most common H⋯A–AA angle in >N–H⋯O=C< (amide-carbonyl hydrogen bond) is 135°.

Table 2.3 Hydrogen bonds in proteins

<table>
<thead>
<tr>
<th>Type</th>
<th>donor⋯acceptor van der Waals distance [Å]</th>
<th>donor⋯acceptor van der Waals distance reduction [%]</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>amide-carbonyl</td>
<td>&gt;N–H⋯O=C&lt;</td>
<td>2.9 ± 0.1</td>
<td>20</td>
</tr>
<tr>
<td>hydroxyl-carbonyl</td>
<td>−O–H⋯O=C&lt;</td>
<td>2.8 ± 0.1</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>−O–H⋯O=C&lt;</td>
<td>2.8 ± 0.1</td>
<td>Ser, Thr, Tyr</td>
</tr>
<tr>
<td>hydroxyl-hydroxyl</td>
<td>−O–H⋯O/H</td>
<td>2.8 ± 0.1</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>−O–H⋯O/H</td>
<td>2.8 ± 0.1</td>
<td>Ser, Thr, Tyr</td>
</tr>
<tr>
<td>amide-hydroxyl</td>
<td>&gt;N–H⋯O/H</td>
<td>2.9 ± 0.1</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>&gt;N–H⋯O/H</td>
<td>2.9 ± 0.1</td>
<td>Ser, Thr, Tyr</td>
</tr>
<tr>
<td>amide-imidazole</td>
<td>&gt;N–H⋯N≤</td>
<td>3.1 ± 0.2</td>
<td>15</td>
</tr>
<tr>
<td>ammonium-carbonyl</td>
<td>−NH\textsubscript{3}⋯−OOC−</td>
<td>2.7 ± 0.1</td>
<td>30</td>
</tr>
<tr>
<td>guanid.-carbonyl</td>
<td>−NH\textsubscript{3}⋯−OOC−</td>
<td>2.7 ± 0.1</td>
<td>30</td>
</tr>
</tbody>
</table>

Typical distances between atom pairs:
- covalent bond between two carbon atoms, 1.5 Å;
- hydrogen bond between amide nitrogen and carbonyl oxygen, 3.0 Å;
- optimal van der Waals distance between two carbon atoms, 4.5 Å (e.g., in methylene or methyl groups in the interior of a protein).

Exercise 5: The interaction between a methylene or methyl group and a carbonyl (i.e., C–H⋯O=C) is much weaker than the hydrogen bond N–H⋯O=C. Why? (Hint: Electronegativity)
<table>
<thead>
<tr>
<th>A-H-B</th>
<th>compounds</th>
<th>average bond length $^b$ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O-H-O</td>
<td>carboxylic acids</td>
<td>0.26 ± 0.01$^c$</td>
</tr>
<tr>
<td>O-H-O</td>
<td>phenols</td>
<td>0.27 ± 0.01$^c$</td>
</tr>
<tr>
<td>O-H-O</td>
<td>alcohols</td>
<td>0.27 ± 0.01$^c$</td>
</tr>
<tr>
<td>O-H-N</td>
<td>all O-H</td>
<td>0.28 ± 0.01$^c$</td>
</tr>
<tr>
<td>N-H-O</td>
<td>ammoniums</td>
<td>0.29 ± 0.01$^c$</td>
</tr>
<tr>
<td>N-H-C</td>
<td>amides</td>
<td>0.28 ± 0.01$^c$</td>
</tr>
<tr>
<td>N-H-N</td>
<td>amines</td>
<td>0.50 ± 0.01$^c$</td>
</tr>
<tr>
<td>N-H-O</td>
<td>all N-H</td>
<td>0.31 ± 0.01$^c$</td>
</tr>
</tbody>
</table>

$^a$The values in this table are reproduced directly from tables in ref 29. $^b$These are the distances between the heteroatoms nitrogens or oxygen. $^c$These standard deviations may be standard deviations of actual lengths or standard deviations of the measurement or both.
From X-ray data on small molecules (crystals of)

(A) One donor

Distribution of values for the angle $b$ (Figure 5-11B) and the angle $c$ (Figure 5-11C) over a population of hydrogen bonds between nitrogen–hydrogen donors and carbonyl oxygen acceptors or acyl oxygen acceptors observed in crystallographic molecular models of small molecules.\textsuperscript{29} (A) Hydrogen bonds involving a carbonyl oxygen or acyl oxygen in which the oxygen atom accepts one other hydrogen bond. In each of the four panels, the number of bonds falling within a range of values is plotted as the value of the ordinate. In the two left panels, the values on the abscissa defining the ranges are values of the sine of angle $c$ ($\sin c$). In the two right panels, the values on the abscissa defining the ranges are values of the angle $b$ in degrees. Adapted with permission from ref 29. Copyright 1968 American Chemical Society.

(B) Two donors

Distribution of values for the angle $b$ (Figure 5-11B) and the angle $c$ (Figure 5-11C) over a population of hydrogen bonds between nitrogen–hydrogen donors and carbonyl oxygen acceptors or acyl oxygen acceptors observed in crystallographic molecular models of small molecules.\textsuperscript{29} (A) Hydrogen bonds involving a carbonyl oxygen or acyl oxygen in which the oxygen atom accepts one other hydrogen bond. In each of the four panels, the number of bonds falling within a range of values is plotted as the value of the ordinate. In the two left panels, the values on the abscissa defining the ranges are values of the sine of angle $c$ ($\sin c$). In the two right panels, the values on the abscissa defining the ranges are values of the angle $b$ in degrees. Adapted with permission from ref 29. Copyright 1968 American Chemical Society.
### 2 Non-covalent interactions

<table>
<thead>
<tr>
<th>Type of Interaction</th>
<th>Model</th>
<th>Example</th>
<th>Dependence of Energy on Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(a) Charge–charge</strong></td>
<td><img src="image" alt="Model" /></td>
<td><img src="image" alt="Example" /></td>
<td>$1/r$</td>
</tr>
<tr>
<td>Longest-range force, nondirectional</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **(b) Charge–dipole**    | ![Model](image) | ![Example](image) | $1/r^2$ |
| Depends on orientation of dipole |

| **(c) Dipole–dipole**    | ![Model](image) | ![Example](image) | $1/r^3$ |
| Depends on mutual orientation of dipoles |

| **(d) Charge–induced dipole** | ![Model](image) | ![Example](image) | $1/r^4$ |
| Depends on polarizability of molecule in which dipole is induced |

| **(e) Dipole–induced dipole** | ![Model](image) | ![Example](image) | $1/r^6$ |
| Depends on polarizability of molecule in which dipole is induced |

| **(f) Dispersion**        | ![Model](image) | ![Example](image) | $1/r^6$ |
| Involves mutual synchronization of fluctuating charges |

| **(g) van der Waals repulsion** | ![Model](image) | ![Example](image) | $1/r^{12}$ |
| Occurs when outer electron orbitals overlap |

| **(h) Hydrogen bond**     | ![Model](image) | ![Example](image) |  |
| Charge attraction + partial covalent bond |

from Mathews, van Holde, Ahern, *Biochemistry.*

Quiz: One of the terms in the column ”Dependence of Energy on Distance” is wrong and another one is an approximation. Which?
2.7 Properties of water

- polar (dipole moment 1.85 D);
- high melting and boiling points;
- high dielectric constant (liquid water: $\epsilon = 78.5$ at 25 °C; $\epsilon = 88$ at 0 °C. Ice Ih: $\epsilon = 100$ at 0 °C, 1 atm);
- the temperature dependence of the density of liquid water has a maximum at 4 °C;
- high viscosity ($\eta = 10^{-2}$ g cm$^{-1}$ s$^{-1}$ at 20 °C, compare with acetonitrile $\eta = 0.36 \times 10^{-2}$ g cm$^{-1}$ s$^{-1}$ at 20 °C); the application of pressure decreases the viscosity of liquid water rather than increasing it as it does for the viscosities of other liquids;
- high heat capacity indicates high degree of structure. (High heat capacity: a heat supply generates only a minor increase in temperature).
  Molar heat capacity at constant pressure: $c_P = (\delta H/\delta T)_P$:
  - ice 8.8 cal/K mol,
  - liquid phase 18.0 cal/K mol (excess or configurational heat capacity),
  - vapor 8.0 cal/K mol.
  The high and essentially constant value of the heat capacity at temperatures between 0 and 100 °C is consistent with a gradual deterioration of the network of hydrogen bonds.
- Hydrogen bonds:
  the shape of the radial distribution function (obtained by Röntgen diffraction) at 100 °C indicates that the intermolecular forces, mainly hydrogen bonds, are strong enough to determine the local structure up to the boiling point.
- the structure of liquid water is not easy to describe (experiments and simulations).
- hydration; hydrophobic effect.

**Exercise 6.1:** The hydrogen bonds are responsible for the high dielectric constant of water. Why does the dielectric constant of liquid water decrease by increasing temperature? $[\epsilon_w(20^\circ C) \cong 80, \epsilon_w(50^\circ C) \cong 70, \epsilon_w(80^\circ C) \cong 60]$

**Exercise 6.2:** Why do (liquid) acetone and methanol have a much lower dielectric constant than water? Explain the order of boiling points: water (100 °C), methanol (65 °C), and acetone (55 °C).
2 Non-covalent interactions

Microwave spectroscopy:

Dimer of water molecules in vapor

Schematic drawing of a dimer of two water molecules in the gas phase. The distances and angles were determined by microwave spectroscopy.

Ice Ih: in equilibrium with liquid H₂O at 0°C and 1 atm

A fragment of the crystal structure of ice Ih. Each O-atom is bonded to four H-atoms at a distance of 0.298 nm. The central O-atom is bonded to two short O-H bonds and two relatively long O-H bonds.

Ice Ih: tetrahexagonal diamond lattice of O-atoms

O-O distance: 2.76 Å

There is a significant amount of empty space in ice Ih.

⇒ less dense than liquid water
2 Non-covalent interactions

Melting points and boiling points of hydrides of nonmetallic elements, showing abnormally high values for hydrogen fluoride, water, and ammonia, caused by hydrogen-bond formation.

Note discontinuity
2.8 Hydration

The Gibbs free energy of hydration of a spherical ion $\Delta G_{\text{hydr}}$ (transfer from vacuum to water) was approximated by Max Born in 1920 as follows. Let $Q$ and $R_{\text{ion}}$ be the charge (in electronic units) and radius (in Å) of an ion, respectively, and $\epsilon_{\text{water}}$ the dielectric constant of water. It follows

$$\Delta G_{\text{hydr}} = -\frac{332}{2R_{\text{ion}}} Q^2 \left(1 - \frac{1}{\epsilon_{\text{water}}} \right) \text{[kcal/mol]} \quad (7)$$

Derivation:
The ion is a sphere of radius $R_{\text{ion}}$ in a medium of dielectric constant $\epsilon$. If the charge of the sphere is $q$, the electrostatic potential at the surface is

$$\phi = \frac{332 \, q}{\epsilon \, R_{\text{ion}}}$$

The work to charge the sphere by $dq$ is $\phi dq$. The total work $w$ to charge the sphere from 0 to $Q$ is

$$w = \int_0^Q \phi \, dq = \frac{332}{\epsilon \, R_{\text{ion}}} \int_0^Q q \, dq = \frac{332}{\epsilon \, R_{\text{ion}}} \frac{Q^2}{2}$$

The Gibbs free energy of hydration $\Delta G_{\text{hydr}}$ corresponds to the total work. Since vacuum has $\epsilon = 1$ and water $\epsilon = \epsilon_{\text{water}}$, the Born formula follows by substituting twice in the above equation.

Salt bridges. In general they do not contribute to the stability of a protein if they are on the surface. (Exception: networks of multiple salt bridges on the surfaces of proteins from thermophilic organisms). The energetic gain due to electrostatic interaction is balanced by the desolvation penalty and the entropic penalty (freezing of the side chain degrees of freedom).

Example: The stability of the heterodimeric (positively- and negatively-charged peptides) "coiled-coil" is not influenced by the ionic strength of the solution. On the other hand, the stability of the homodimeric "coiled-coil" decreases by reducing the ionic strength because of the reduced salt-screening effect.

Exercise 7: For the transfer process water ($\epsilon_{\text{water}} = 80$) $\rightarrow$ protein interior ($\epsilon_{\text{protein}} = 4$), calculate $\Delta G_{\text{transfer}}$ for an ion ($Q = 1$) of radius 2 Å. [Ionizable side chains are in equilibrium between charged and uncharged species so that the calculated value of $\Delta G_{\text{transfer}}$ represents an upper bound for the penalty of Asp/Glu or Arg/Lys/His burial into a protein hydrophobic core.]
Protein hydration. In the first hydration layer, water molecules are involved in hydrogen bonds with hydrophilic groups on the protein surface (Figure 2.5).

Fig. 2.5  X-ray structure of chymotrypsin inhibitor 2 (stick model with carbon atoms in green) and water molecules (red spheres) of first hydration layer.

Exercise 8: Draw schematically the possible interactions between water molecules and backbone polar groups. Which elements of secondary structure are accessible to water molecules? Discuss the differences in the water-water interactions in the first hydration layer with respect to the bulk.
2.9 Hydrophobic effect

- At physiological temperature the hydrophobic effect is mainly entropic (reduced entropy of the water molecules around hydrophobic solute).

- It has been shown that for saturated alkanes the logarithm of the solubility in water is inversely related to the solvent accessible surface area of the solute (Hermann, R.B., J. Phys. Chem. 76, 2745-2759, 1972).

- The apolar side chains are buried in hydrophobic cores in water-soluble proteins. The surface of a water-soluble protein is mainly hydrophilic whereas a membrane protein has a hydrophobic surface. Why? (The lipidic tails of the bilayer are hydrophobic).

- Most of the buried polar groups are involved in hydrogen bonds.

- Charged side chains (Asp, Glu, Arg, Lys) are usually on the surface of water-soluble proteins. Salt bridges are relatively seldom because of desolvation penalty.
2 Non-covalent interactions

Typical thermodynamics of transfer of a nonpolar molecule from the gas to aqueous solution at two temperatures: (A) T, approximately 20°C, where ΔH = 0 for transfer between liquid and water; and (B) T, approximately 140°C, where ΔH = 0. The values of ΔH, TΔS, and ΔG are in units of kcal/mol, that of ΔG in units of cal/(K·mol).

Typical thermodynamics of the free energy of transfer of a hydrocarbon from the liquid to aqueous solution, using pentane as an example. The strong temperature dependence of both the enthalpy and entropy difference between the two phases is a result of the different heat capacities of the two phases. The free-energy difference is the net difference between the enthalpic ΔH and entropic TΔS contributions. It reaches a maximum where ΔS = 0, whereas the equilibrium constant (which is proportional to -ΔG/ΔT) reaches a maximum where ΔH = 0. (Adapted from P. L. Privatov and S. J. Gill, Adv. Protein Chem. 32:191-274, 1989.)

Relationships between the solubility of hydrocarbons in water and their cavity surface areas. The negative logarithms of the solubilities (log S) of the liquid hydrocarbons in water in moles liter⁻¹ were used directly and were plotted as a function of their cavity surface areas (nm²).
Table 11.5  Free energies of transfer of amino acid side chains to water from cyclohexane or the gas phase

<table>
<thead>
<tr>
<th>Side chain of</th>
<th>Cyclohexane→H₂O kcal/mol</th>
<th>kJ/mol</th>
<th>Vapor→H₂O kcal/mol</th>
<th>kJ/mol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu</td>
<td>4.92</td>
<td>20.59</td>
<td>2.28</td>
<td>9.54</td>
</tr>
<tr>
<td>Ile</td>
<td>4.92</td>
<td>20.59</td>
<td>2.15</td>
<td>9.00</td>
</tr>
<tr>
<td>Val</td>
<td>4.04</td>
<td>16.91</td>
<td>1.99</td>
<td>8.33</td>
</tr>
<tr>
<td>Pro</td>
<td>3.58</td>
<td>14.98</td>
<td>1.50</td>
<td>6.28</td>
</tr>
<tr>
<td>Phe</td>
<td>2.98</td>
<td>12.47</td>
<td>-0.76</td>
<td>-3.18</td>
</tr>
<tr>
<td>Met</td>
<td>2.35</td>
<td>9.83</td>
<td>-1.48</td>
<td>-6.19</td>
</tr>
<tr>
<td>Trp</td>
<td>2.33</td>
<td>9.75</td>
<td>-5.88</td>
<td>-24.61</td>
</tr>
<tr>
<td>Ala</td>
<td>1.81</td>
<td>7.57</td>
<td>1.94</td>
<td>8.12</td>
</tr>
<tr>
<td>Cys</td>
<td>1.28</td>
<td>5.36</td>
<td>-1.24</td>
<td>-5.19</td>
</tr>
<tr>
<td>Gly</td>
<td>0.94</td>
<td>3.93</td>
<td>2.39</td>
<td>10.00</td>
</tr>
<tr>
<td>Tyr</td>
<td>-0.14</td>
<td>-0.59</td>
<td>-6.11</td>
<td>-25.57</td>
</tr>
<tr>
<td>Thr</td>
<td>-2.57</td>
<td>-10.75</td>
<td>-4.88</td>
<td>-20.42</td>
</tr>
<tr>
<td>Ser</td>
<td>-3.40</td>
<td>-14.23</td>
<td>-5.06</td>
<td>-21.18</td>
</tr>
<tr>
<td>His</td>
<td>-4.66</td>
<td>-19.50</td>
<td>-10.27</td>
<td>-42.98</td>
</tr>
<tr>
<td>Gln</td>
<td>-5.54</td>
<td>-23.18</td>
<td>-9.38</td>
<td>-39.25</td>
</tr>
<tr>
<td>Lys</td>
<td>-5.55</td>
<td>-23.23</td>
<td>-9.52</td>
<td>-39.84</td>
</tr>
<tr>
<td>Asn</td>
<td>-6.64</td>
<td>-27.79</td>
<td>-9.68</td>
<td>-40.51</td>
</tr>
<tr>
<td>Glu</td>
<td>-6.81</td>
<td>-28.50</td>
<td>-10.24</td>
<td>-42.85</td>
</tr>
<tr>
<td>Asp</td>
<td>-8.72</td>
<td>-36.49</td>
<td>-10.95</td>
<td>-45.82</td>
</tr>
<tr>
<td>Arg</td>
<td>-14.92</td>
<td>-62.44</td>
<td>-19.92</td>
<td>-83.36</td>
</tr>
</tbody>
</table>

2.10 Micelles and membranes (movies)

In aqueous solution, amphiphilic molecules (Figure 2.6) aggregate because of the hydrophobic effect. The formation of micelles or bilayers (e.g., cell membranes) depends on the conical or cylindrical shape, respectively, of the amphiphilic molecule (Figure 2.7).

- MD simulation of micelle formation starting from 54 monodisperse dodecylphosphocholine (DPC) molecules at 300 K
  Movie: [http://www.biochem-caflisch.uzh.ch/movie/13/](http://www.biochem-caflisch.uzh.ch/movie/13/)

- MD simulation of membrane formation starting from 78 monodisperse molecules of dimyristoyl phosphatidylcholine (DMPC) at 330K
  Movie: [http://www.biochem-caflisch.uzh.ch/movie/14/](http://www.biochem-caflisch.uzh.ch/movie/14/)

![Fig. 2.6 Amphiphilic molecules.](image)

![Fig. 2.7 The shape of lipid molecule determines the type of aggregate.](image)
2.11 Hydrophobic effect and membrane proteins

[Diagram of Glycophorin]

[Diagram of Bacteriorhodopsin]
Rhodopsin.
+ charged, − charged, and hydrophobic side chains are in blue, red, and green, respectively.
3 Protein structure

3.1 The 20 natural amino acids and their role in proteins

Amino acids are the basic structural units of proteins. An α-amino acid consists of an amino group, a carboxyl group, a hydrogen atom and a residue R that are all bound to a so-called Cα-atom. The tetrahedral configuration of four different groups at the Cα-atom confers an optical activity to amino acids (except for glycine); the two mirror image forms are termed L- and D-isomers (Fig. 3.1). Only L-amino acids are components of naturally occurring proteins.

Fig. 3.1  Stereo representation (wall-eye) of the “stick-and-ball” model for L-alanine

Usually 20 types of side chains are found in proteins: they differ in size, shape, charge, hydrogen bonding ability, hydrophilicity, and chemical reactivity.

Gly:  R = H; flexible.
Ala Val Leu Ile:  aliphatic. All these residues are hydrophobic.
Pro:  Aliphatic with a ring structure. The Pro side chain is bound to both Cα and N atoms of the main chain. Proline is an imino acid, it has a secondary amino group. It is often found in the bends of folded protein chains.
Phe Tyr Trp:  aromatic. The aromatic rings have clouds of delocalized π-electrons that allow them to interact with other π-systems (π-stacking) and to transfer electrons.
Cys Met:  have a sulfur atom (Cys, R = −CH2–SH; Met, R = −CH2–CH2–S–CH3). The −SH group of cysteine is reactive. Disulfide bridges are important for the shape and stability of the tertiary structure.
Ser Thr Asn Gln:  polar but not charged. These residues are often involved in hydrogen bonds through their −OH groups (Ser, Thr) and their −CO−NH2 groups (Asn, Gln).
Asp Glu:  acidic. Almost always negatively charged under physiological pH values (except aspartic proteases). In cases where these residues are located in the interior of the protein, they are almost always involved in salt bridges with Arg or Lys.
Arg Lys:  basic. Almost always positively charged under physiological pH values. In cases where these residues are located in the interior of the protein, they are almost always involved in salt bridges with Asp or Glu.
His:  basic and aromatic. The pK value of His lies in the physiological pH range. Histidine is present in the active center of serine proteases, where the imidazole ring can switch between both ionization forms.
3.2 Dihedral angles and Ramachandran plot

Under the assumption that all bond lengths and bond angles correspond to the standard dimensions, the conformation of a polypeptide chain can be uniquely determined through the specification of the dihedral angles (also often called “torsion angles”) at the individual bonds. For each residue three backbone dihedral angles $\phi$, $\psi$ and $\omega$ have to be specified.
Table 3.1  Definitions of protein dihedral angles

<table>
<thead>
<tr>
<th>Angle</th>
<th>Bond</th>
<th>Atoms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main chain</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_i$</td>
<td>$N^i-C^i_a$</td>
<td>$C'_{i-1}-N^i-C^i_a-C^i$</td>
</tr>
<tr>
<td>$\psi_i$</td>
<td>$C^i_a-C'^i$</td>
<td>$N^i-C^i_a-C'^i-N^{i+1}$</td>
</tr>
<tr>
<td>$\omega_i$</td>
<td>$C'^i-N^{i+1}$</td>
<td>$C^i_a-C'^i-N^{i+1}-C^i_{a+1}$</td>
</tr>
<tr>
<td><strong>Side chains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi_1$</td>
<td>$C^i_a-C^i_{\beta}$</td>
<td>$N-C^i_a-C^i_{\beta}-C^i_{\gamma}$</td>
</tr>
<tr>
<td>$\chi_2$</td>
<td>$C^i_{\beta}-C^i_{\gamma}$</td>
<td>$C^i_a-C^i_{\beta}-C^i_{\gamma}-C^i_{\delta}$</td>
</tr>
<tr>
<td>$\chi_3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(...</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Note:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi_1$</td>
<td>$C^i_a-C^i_{\beta}$</td>
<td>$N-C^i_a-C^i_{\beta}-O_{\gamma}$</td>
</tr>
<tr>
<td>Methionine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi_2$</td>
<td>$C^i_{\beta}-C^i_{\gamma}$</td>
<td>$C^i_a-C^i_{\beta}-C^i_{\gamma}-S_{\delta}$</td>
</tr>
</tbody>
</table>

Due to the partial double bond character, peptide groups are planar and the $\omega$ angle is trans. The $\omega$ angle involving the nitrogen atom of Pro can also be cis.

![Diagram of dihedral angles](image)

Fig. 3.2  Dipeptide unit of residues 86-88 in barnase. Add labels of dihedral angles according to Table 3.1.
Ramachandran plot: two dimensional representation for the conformation of individual dipeptide units in the \( \phi-\psi \)-plane.

![Ramachandran plot](image)

**Fig. 3.3a** +, barnase; o, immunoglobulin fragment Fab (McPC603, PDB code 1MCP).

**Remarks**

- **\( \beta \)-sheet region**: higher density on the right side of the diagonal because of the right-handed chirality of the individual strands.

- **\( \alpha \)-helical region**: higher density on the left side of the diagonal because of the right-handed chirality of the helix.

- **Glycine**: allowed regions are larger because glycine has no \( C_\beta \) atom; the Ramachandran plot is symmetrical because glycine is not chiral at \( C_\alpha \).

**Exercise S1**: What is the difference between the Ramachandran maps of glycine and alanine? What is the difference between the Ramachandran maps of \( \alpha \)-amino-isobutyric acid (a non-natural amino acid with two \( \beta \)-carbons) and alanine?
MolProbity Ramachandran analysis

IRNB, model 1

96.3% (103/107) of all residues were in favored (98%) regions.
100.0% (107/107) of all residues were in allowed (>99.8%) regions.

There were no outliers.

http://kinemage.biochem.duke.edu


Fig. 3.3b PDB code 1rnb (X-ray structure of barnase at 1.9 Å resolution);
Click on the header menu “Geometry” — “MolProbity Ramachandran Plot”
MolProbity Ramachandran analysis

86.5% (379/438) of all residues were in favored (98%) regions.
95.9% (420/438) of all residues were in allowed (>99.8%) regions.

There were 18 outliers (phi, psi):
L 15 ALA (-55.3, 137.3)
L 132 THR (-56.1, 39.5)
L 133 SER (-168.5, 9.8)
L 158 GLY (-67.2, -101.2)
L 161 ARG (155.3, 156.7)
L 196 ASN (-163.5, -9.7)
L 218 ASN (158.3, -35.9)
H 30 SER (-54.6, 19.6)
H 31 ASP (179.0, 35.7)
H 54 LYS (57.0, 68.3)
H 57 LYS (66.4, 124.9)
H 123 GLU (-56.3, -173.2)
H 137 PRO (-54.3, -9.7)
H 142 ASP (66.0, 125.0)
H 168 LYS (-58.1, -92.2)
H 170 ILE (167.5, 109.9)
H 199 PRO (-57.9, -119.2)
H 203 SER (112.6, 149.3)

http://kinemage.biochem.duke.edu


Fig. 3.3c  PDB code 1mcp (X-ray structure of phosphocholine binding immunoglobulin Fab at 2.7 Å resolution);
Click on the header menu "Geometry” → "MolProbity Ramachandran Plot"
3.3 Secondary structure

3.3.1 Regular and irregular secondary structure

**Definition:** in regular secondary structures, all dipeptide units show the same combination of φ-ψ-values, i.e., the same position in the Ramachandran plot.

**Table 3.2 Geometrical characteristics of the regular secondary structures**

<table>
<thead>
<tr>
<th>Structure</th>
<th>Average values of φ</th>
<th>Residues per turn</th>
<th>Rise per residue [Å]</th>
<th>Helix radius [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antiparallel β-sheet</td>
<td>-139</td>
<td>2.0</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>Parallel β-sheet</td>
<td>-119</td>
<td>2.0</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>Right-handed α-helix (i + 4)</td>
<td>-57</td>
<td>3.6</td>
<td>1.5</td>
<td>2.3</td>
</tr>
<tr>
<td>3_{10}-helix (i + 3)</td>
<td>-49</td>
<td>3.0</td>
<td>2.0</td>
<td>1.9</td>
</tr>
<tr>
<td>π-helix (i + 5)</td>
<td>-57</td>
<td>4.4</td>
<td>1.1</td>
<td>2.8</td>
</tr>
</tbody>
</table>

**Helices (3_{10}, α, π)**

**Exercise S2:** Thermodynamics of helix folding: Although in the α-helix the \( i \cdots i + 4 \) hydrogen bond occurs, peptides usually do not assume α-helical structure at room temperature and neutral pH, even if the same sequence is helical in the context of a protein. Why? (Hint: Role of water molecules and entropic penalty of peptide.) Kinetics of helix folding: Theoretical and computational studies indicate that helix initiation (formation of first turn) is endergonic, while helix elongation is only marginally exergonic. Also, the barrier is located at the formation of the first helical turn. Why?

**Exercise S3 Why is the π-helix extremely sporadic?** Hint: compare the optimal van der Waals distance of two carbon atoms (see Fig. 2.2) with the helical radius in the last column of Table 3.2.
**β-sheets (antiparallel, parallel)**

**Antiparallel**

Aufsicht

Seitenansicht

![Diagram](image)

0,7 nm

(a)

**Parallel**

Aufsicht

Seitenansicht

![Diagram](image)

0,65 nm

(b)
About one quarter of the residues in globular proteins are involved in turns and loops in order to change the direction of the polypeptide chain at the surface.

**Table 3.3 Geometrical characteristics of the tight turns**

<table>
<thead>
<tr>
<th>Type</th>
<th>$\phi_{i+1}$</th>
<th>$\psi_{i+1}$</th>
<th>$\phi_{i+2}$</th>
<th>$\psi_{i+2}$</th>
<th>Residues per turn</th>
<th>Residue at $i + 1$</th>
<th>Residue at $i + 2$</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ I</td>
<td>-60</td>
<td>-30</td>
<td>-90</td>
<td>0</td>
<td>4</td>
<td>Pro</td>
<td></td>
<td>35%</td>
</tr>
<tr>
<td>$\beta$ I'</td>
<td>60</td>
<td>30</td>
<td>90</td>
<td>0</td>
<td>4</td>
<td>Gly</td>
<td>Gly</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>$\beta$ II</td>
<td>-60</td>
<td>120</td>
<td>80</td>
<td>0</td>
<td>4</td>
<td>Gly</td>
<td></td>
<td>15%</td>
</tr>
<tr>
<td>$\beta$ II'</td>
<td>60</td>
<td>-120</td>
<td>-80</td>
<td>0</td>
<td>4</td>
<td>Gly</td>
<td></td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>$\beta$ III</td>
<td>-60</td>
<td>-30</td>
<td>-60</td>
<td>-30</td>
<td>4</td>
<td></td>
<td></td>
<td>15%</td>
</tr>
<tr>
<td>$\beta$ III'</td>
<td>60</td>
<td>30</td>
<td>60</td>
<td>30</td>
<td>4</td>
<td></td>
<td></td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>80</td>
<td>-65</td>
<td></td>
<td></td>
<td>3</td>
<td>Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma$ inverse</td>
<td>-80</td>
<td>65</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.3.2 Supersecondary structure

Supersecondary structures show a higher degree of structure than secondary structures, but they do not form entire structural domains.

Combinations of polypeptide strands in β-structures:

- hairpin motif in antiparallel sheet. Example: β-meander, i.e., antiparallel β-sheet of three strands connected through two turns.
- right-handed “cross-over” motif in parallel β-sheet. Example: The Rossmann fold, i.e., a βαβαβ-unit, is common and may contain an hydrophobic core between β-sheet and α-helix.
- left-handed “cross-over” motif in parallel sheet (very rare).

Exercise S4: Explain the helical propensity for the three peptides listed below. Results were obtained with the program AGADIR which predicts the helical behaviour of monomeric peptides by considering only short range interactions. Explain the reason for the differences between peptides 1 and 2. Explain the reason for the low helical propensity of the barnase segment 6-18 which is helical in the native structure.

Thanks for using the EMBL WWW Gateway to AGADIR
http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html
Copyright ©1997-2002, Lacroix E., Munoz V., Petukhov M. & Serrano, L.

<table>
<thead>
<tr>
<th>pH</th>
<th>7.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>278</td>
</tr>
<tr>
<td>Ionic Strength</td>
<td>0.100</td>
</tr>
<tr>
<td>Nterm</td>
<td>acetylated</td>
</tr>
<tr>
<td>Cterm</td>
<td>amidated</td>
</tr>
</tbody>
</table>

Peptide 1 EAAAAAAAHH
res, aa, Hel
01, E, 43.7
02, A, 51.0
03, A, 54.4
04, A, 56.4
05, A, 57.1
06, A, 57.1
07, A, 56.4
08, A, 54.9
09, A, 52.4
10, A, 48.5
11, A, 42.4
12, A, 33.3
13, H, 16.1
Percentage helix 47.97
========================================================================

Peptide 2 EAAAAAGAAAAH
res, aa, Hel
01, E, 13.4
02, A, 15.5
03, A, 16.4
04, A, 16.9
05, A, 16.4
06, A, 15.3
07, A, 13.4
08, G, 9.7
09, A, 8.4
10, A, 7.2
11, A, 6.9
12, A, 5.7
13, H, 2.8
Percentage helix 11.39
========================================================================

Barnase 1-26 AQVINTFDGVADYLQTYHKLPNYIT
res, aa, Hel
01, A, 0.2
02, Q, 0.2
03, V, 0.2
04, I, 0.2
05, N, 0.1
06, T, 0.2
07, F, 0.1
08, D, 0.1
09, G, 0.4
10, V, 1.0
11, A, 1.1
12, D, 1.1
13, Y, 1.2
14, L, 1.2
15, Q, 0.9
16, T, 0.9
17, Y, 0.7
18, H, 0.2
19, K, 0.1
20, L, 0.0
21, P, 0.0
22, D, 0.0
23, N, 0.0
24, Y, 0.0
25, I, 0.0
26, T, 0.0
Percentage helix 0.40
============================================================================
3.4 Tertiary structure

Arrangement of regular secondary structure and loops into folded three-dimensional structure. The topology is the overall backbone arrangement; the details of the side chain packing are very important, in particular in the hydrophobic core(s).

3.4.1 Anatomy of protein structures

The ribonuclease of *Bacillus amyloliquefaciens* (barnase) consists of three α-helices and an antiparallel β-sheet.

![Schematic representation of barnase tertiary structure](image)

Fig. 3.4 Schematic representation of barnase tertiary structure. The side chains of the hydrophobic core are depicted as “ball-and-stick”. The Asp, Glu, Lys, Arg and His side chains are in the “stick” representation. The sequence numbering of Asp and Glu is shown.

**Exercise S5.1:** In barnase, helix1 is docked to the β-sheet. Which effect is responsible for this?

**Exercise S5.2:** The double salt bridge Asp8-Arg110-Asp12 lies at the surface of barnase and does not contribute to the thermodynamic stability. Why?

**Exercise S5.3:** The $pK_a$ of His18 is 7.9, i.e., 1.6 units higher than in model peptides. Why? Which interactions exist between His18 and α-helix1 in barnase? (Answer: Šali et al., *Nature* **335**, 740-743, 1988.)
3.4.2 Taxonomy of protein structures

(α)–proteins contain mainly α-helices; 50% to 95% of the amino acid residues are in α-helical structure. Example: myoglobin (Fig. 3.5).

(β)–proteins contain mainly β-sheets. Two β-sheets, usually antiparallel, are packed on top of each other. Example: antibody domains (Fig. 3.6).

(α + β)–proteins contain helices and sheets in separate parts of the sequence. There is often a unique antiparallel β-sheet and the helices dock to its end. Example: barnase (Fig. 3.4).

(α/β)–proteins consist of alternating helices and sheets that show significant interactions. There is often a large parallel β-sheet whose strands are connected through α-helices. Examples: pyruvate kinase, domain 1 (singly wound β-barrel); phosphorylase, domain 2 (doubly wound β-sheet).

Fig. 3.5 Stereo representation of the backbone of cetacean myoglobins (deoxy).
PDB code 5MBN.

Fig. 3.6 Stereo representation of the backbone of McPC603 Fab antibody fragments. The H-chain is shown in bold.
PDB code 2MCP.
3.5 Quaternary structure

Proteins that consist of more than one polypeptide chain show a further structural organization level. Each polypeptide chain of such a protein is referred to as subunit. The quartenary structure describes the spatial arrangement of such subunits and the nature of their contacts. The individual chains of a protein with several subunits may be identical (HIV-1 aspartic protease homodimer, Fig. 3.7) or different (antibody L- and H-chains, α- and β-chains of hemoglobin, Fig. 3.8). The spherical membrane of tomato bushy stunt virus, a plant virus, is composed of 180 identical membrane protein molecules.

The contact surfaces of the subunits are often of functional significance. For instance in the case of hemoglobin, which consists of four chains (Fig. 3.8), the contact surfaces of the four subunits take part in the information transfer between the binding sites for O₂, CO₂ and H⁺. Similarly, in the antibody shown in Fig. 3.6, parts of two different chains are involved in the antigen binding site.

Fig. 3.7 Stereo representation of HIV-1 aspartic protease. PDB code 3HVP.

Fig. 3.8 (Top) Stereo representation of human deoxyhemoglobin. PDB code 4HHB. (Bottom) Ribbon and surface rendering.
3 Protein structure

[a] Antigenbindungsstelle

[b] gebundenes Kohlenhydrat

C = konstante Domäne
V = variable Domäne
H, L = schwere, leichte Ketten

### 3.6 Experimental approaches for structural studies: Overview

#### Table 3.4 High resolution methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray crystallography</td>
<td>Atomic level description of 3D structure in a crystal, (but: water molecules, enzymatic activity)</td>
</tr>
<tr>
<td>Nuclear magnetic resonance</td>
<td>Atomic level description of 3D structure in solution, analysis of conformational flexibility</td>
</tr>
<tr>
<td>Electron microscopy</td>
<td>CryoEM + X-ray/NMR</td>
</tr>
<tr>
<td>Free-electron laser crystallography</td>
<td>very new</td>
</tr>
</tbody>
</table>

"Free-electron laser crystallography is without question the most spectacular new technical development in the field. A jet of droplets with nanocrystals is sent past a beam of femtosecond x-ray laser pulses, and diffraction is recorded before the crystal has time to explode. This has the potential for determining otherwise intractable molecular structures and for accessing rapid time steps."

From Richardson and Richardson, Biophysical Journal, 106, 510, 2014. [link](#)

#### Table 3.5 Single molecule methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atomic force microscopy</td>
<td>Micromanipulation of single molecules, determination of the force for dissociating a complex</td>
</tr>
<tr>
<td>Laser tweezers</td>
<td></td>
</tr>
<tr>
<td>Förster resonance energy transfer</td>
<td>Free energy surface of protein folding</td>
</tr>
<tr>
<td>(FRET)</td>
<td></td>
</tr>
</tbody>
</table>

#### Table 3.6 Further methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circular dichroism (CD)</td>
<td>Main chain conformation (secondary structure)</td>
</tr>
<tr>
<td>UV-spectroscopy</td>
<td>Investigation of molecular conformation details in solution and partly in crystals,</td>
</tr>
<tr>
<td>Fluorescence spectroscopy</td>
<td></td>
</tr>
<tr>
<td>IR-spectroscopy</td>
<td></td>
</tr>
<tr>
<td>Raman-spectroscopy</td>
<td></td>
</tr>
<tr>
<td>Electron spin resonance</td>
<td></td>
</tr>
<tr>
<td>Mössbauer spectroscopy</td>
<td></td>
</tr>
<tr>
<td>Sedimentation (ultracentrifuge)</td>
<td>Determination of the molecule size and form</td>
</tr>
<tr>
<td>Light diffraction</td>
<td></td>
</tr>
<tr>
<td>X-ray small angle diffraction</td>
<td></td>
</tr>
<tr>
<td>Fluorescence depolarization</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 3.9 Stereoview of twenty NMR conformers of barnase (PDB file 1bnr.pdb, α-helices in red).

Fig. 3.10 Stereoview of the X-ray structure (PDB file 1rnb.pdb, green and yellow) superposed to five of the 20 NMR conformers. The root mean square deviation of the $C_\alpha$ carbons is smaller than 1.0 Å. Interestingly, the N-terminus and loop3 of the X-ray structure (yellow) are the segments with the highest B-factors (see Fig. 3.11B), which is consistent with the large deviations in the corresponding regions of the NMR conformers.
**Exercise S6:** Describe the secondary structure of chymotrypsin inhibitor 2 and assign its fold to one of the four classes mentioned above. [Hint: Visualize the 3D structure on the PDB http://www.rcsb.org/pdb/ by entering the code 2ci2 and clicking on "View in 3D: JSmol" in the box with the structure on the right.] What are the two most flexible segments of chymotrypsin inhibitor 2 in its native structure? [Hint: Check the profile of B-factors in Fig. 3.11A]

Ribbon diagram of the structure of chymotrypsin inhibitor 2.

Fig. 3.11. Crystallographic B-factors as a function of residue number: (A) Chymotrypsin inhibitor 2, (B) barnase.
4 Ordered aggregation and amyloid fibrils

4.1 Historical note and future perspectives

"Ich habe mich verloren" (Auguste Deter, 1901)
4.2 Amyloid fibrils

- Microscopic techniques (left) and X-ray diffraction (right) have been used to study amyloid fibril structures. **Problem: no atomic detail!**

- Amyloid fibrils consist of $\beta$-sheet structures with individual strands perpendicular to the fibril axis and backbone hydrogen bonds parallel to the axis.

(A) Amyloid fibrils are composed of long filaments that are visible in negatively stained transmission electron micrographs.

(B) Schematic diagram of the cross-$\beta$ sheets in a fibril, with backbone hydrogen bonds represented by dashed lines.

(C) Fiber diffraction pattern with a meridional reflection at $\sim 4.7$ Å (black dashed rectangles) and an equatorial reflection at $\sim 6-11$ Å (white dashed rectangles). From Greenwald and Riek, Structure 18, 1244, 2010.
- Fibrils are not crystalline, yet they are too large to be studied by solution NMR spectroscopy.
- Insight into fibril structure at the atomistic scale is obtained by solid-state NMR (e.g., for Aβ_{40}).

(Left) Structural model for Aβ_{1-40} fibrils, consistent with solid state NMR constraints on the molecular conformation and intermolecular distances, and incorporating the cross-motif common to all amyloid fibrils. The N-terminal residues 1 to 9 are considered fully disordered and are omitted. Petkova et al. PNAS 99, 16742, 2002.

(Right) Profile of aggregation propensity (π, shown on y-axis) along the sequence of the full length Alzheimer’s polypeptide precursor, abbreviated as APP (top), and the Aβ segment (bottom). Note that the highest aggregation propensity is predicted for the Aβ segment L_{17}VFFA_{21}, which corresponds to the central region of the strand in the bottom of the left panel.
For short peptides, X-ray diffraction of microcrystals has also been used to obtain atomic resolution data (e.g., visualize the PDB structure 2OMM using JSmol).

Fiber formed by the heptapeptide GNNQQNY from the yeast prion-like protein Sup35. The pair-of-sheets structure shows the backbone of each $\beta$-strand as an arrow, with side chains protruding. The dry interface is between the two sheets, with the wet interfaces on the outside surfaces. Side chains Asn2, Gln4 and Asn6 point inwards, forming the dry interface. Nature 447, 453, 2007.

Exercise A1. Why are hydrogen bonds in the interior of the fibrils very stable?

Exercise A2. Discuss the energy contributions that stabilize the GNNQQNY fibrils. Do you expect fibrils of GNNQQN to be more or less stable than those of GNNQQNY and why?

Exercise A3. The peptide NNQGQNY has a much smaller aggregation tendency than GNNQQNY although their amino acid composition is identical. Why? [Hint: Think about both enthalpic and entropic terms.]
4.3 Kinetics of fibril formation

- Several water-soluble proteins and peptides can form insoluble fibrils.

- Amyloid fibril formation requires partial unfolding of globular proteins (e.g., antibody light chain-associated renal disorders) or formation of $\beta$-extended structure in the case of disordered peptides (e.g., the 42-residue Alzheimer’s $A_\beta$ peptide).

Fibrillogenesis requires nucleation, is accelerated by seeding and occurs via prefibrillar intermediates and it is irreversible.

The pathways of fibril formation can include an amorphous aggregate. Protofibrils may intertwine and form mature fibrils (fibers). The seeding can accelerate the kinetics overcoming a lag-phase.
- Concentration (left) and sequence (right) strongly affect amyloid formation kinetics.


Kinetic traces of mutants of Aβ40 (Christopeit et al., Protein Sci. 14, 2125, 2005).

Exercise A4. Explain the correlation between length of lag phase and length of elongation phase (which emerges clearly from the figure panel on the left).

Exercise A5. Explain why lower concentrations of insulin result in lower plateaus (left panel).

Exercise A6. Explain the tendency to form parallel or anti-parallel β-sheet aggregates of these peptides (in one-letter code, with/without neutral capping groups): (a) capped LVFFA, [parallel]; (b) uncapped KLVFFAE, [antiparallel]; (c) uncapped ELVFFK, [parallel] (d) capped YNQQNY, [both]; (e) uncapped YNQQNY, [antiparallel]. Discuss the energy contributions that stabilize the β-sheet aggregates formed by each of these peptides.

Exercise A7. Discuss the temperature-dependence of the fibril formation kinetics. Explain the reasons for the slower kinetics of aggregation below and above a temperature range at which kinetics are fastest.
- The lag phase and elongation rate depend on the physico-chemical properties of the sequence (i.e., $\beta$-propensity, distribution of charged side chains, aromatic side chains) and environmental conditions (i.e., concentration and temperature).

Prediction of elongation rates ($\nu$) using an analytical approach based on physico-chemical properties (Tartaglia et al., Protein Science 14, 2723, 2005).

Exercise A8. In the figure above, glucagon shows the slowest kinetics of aggregation while the immunoglobulin-like domain of titin (about 90 residues) shows the fastest. Why? [Hint: Think about the differences in secondary structure propensity of glucagon (a peptide hormone of 30 residues, PDB code 1GCN) and the immunoglobulin domain.]
• Computational prediction of a Ure2p (prion) mutant with slower aggregation kinetics

Details of prediction based on MD simulations and experimental validation by thioflavin T binding are in Cecchini et al., J. Mol. Biol. 357, 1306, 2006.
4.4 A unique fibrillar structure?

- Some amyloids assume a unique fibrillar structure under physiological conditions, e.g., the β-solenoid fold of the prion forming domain HET-s.

![The β-Solenoid Motif in HET-s. The left view is a ribbon representation of the HET-s PFD fibril with the peptide chains alternating blue and white and a ball on the N-termini (residue 225). The four β strands are labeled. Wasmer et al. Science 319, 1523, 2008.]

- Several amyloids, including disease-associated amyloids, can polymerize into different fibrillar structures (polymorphism).

![TEM images of amyloid fibrils formed by the Aβ1−40 peptide. Parent fibrils were prepared by incubation of Aβ1−40 solutions either under quiescent dialysis conditions or in a closed polypropylene tube with gentle agitation. Petkova et al. Science 307, 262, 2005]
• Fibril structures are influenced by environmental factors (ionic strength, pH, temperature, agitation, etc). In contrast, globular proteins have been evolutionary selected to fold in a unique 3D structure at a relatively broad range of external conditions (e.g., pH range of 2 to 10 and temperature up to 70-80 °C).

• Protein folding is (almost always) under thermodynamic control, i.e., the folded state has the lowest free energy. In contrast, fibril formation and amyloid polymorphism is under kinetic control. Pellarin et al. JACS 132, 14960, 2010.
• Molecular "recycling" during final fibril/monomer equilibrium: see movie at [http://www.biochem-caflisch.uzh.ch/movie/12/]
4.5 Amyloid related diseases

- Several diseases have been associated with formation of extracellular amyloid deposits or intracellular inclusions with amyloid-like characteristics, like Alzheimer’s disease, spongiform encephalopathies, Huntington’s disease and type II diabetes (Chiti, Dobson, Annu. Rev. Biochem. 75, 333, 2006).

<table>
<thead>
<tr>
<th>Disease</th>
<th>Aggregating protein or peptide</th>
<th>Number of residues</th>
<th>Native structure of protein or peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurodegenerative diseases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alzheimer’s disease&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Amyloid β peptide</td>
<td>40 or 42&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Natively unfolded</td>
</tr>
<tr>
<td>Spongiform encephalopathies&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Prion protein or fragments thereof</td>
<td>253</td>
<td>Natively unfolded (residues 1–120) and α-helical (residues 121–230)</td>
</tr>
<tr>
<td>Parkinson’s disease&lt;sup&gt;c&lt;/sup&gt;</td>
<td>α-Synuclein</td>
<td>140</td>
<td>Natively unfolded</td>
</tr>
<tr>
<td>Dementia with Lewy bodies&lt;sup&gt;c&lt;/sup&gt;</td>
<td>α-Synuclein</td>
<td>140</td>
<td>Natively unfolded</td>
</tr>
<tr>
<td>Frontotemporal dementia with Parkinsonism&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Tau</td>
<td>352–441&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Natively unfolded</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Superoxide dismutase 1</td>
<td>153</td>
<td>All-β, Ig like</td>
</tr>
<tr>
<td>Huntington’s disease&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Huntingtin with polyQ expansion</td>
<td>3144&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Largely natively unfolded</td>
</tr>
<tr>
<td>Spinocerebellar ataxia&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Ataxins with polyQ expansion</td>
<td>816&lt;sup&gt;h&lt;/sup&gt;</td>
<td>All-β, AXH domain (residues 562–694); the rest are unknown</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 17&lt;sup&gt;d&lt;/sup&gt;</td>
<td>TATA box-binding protein with polyQ expansion</td>
<td>339&lt;sup&gt;f&lt;/sup&gt;</td>
<td>α+β, TBP like (residues 159–339); unknown (residues 1–158)</td>
</tr>
<tr>
<td>Spinal and bulbar muscular atrophy&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Androgen receptor with polyQ expansion</td>
<td>919&lt;sup&gt;f&lt;/sup&gt;</td>
<td>All-α, nuclear receptor ligand-binding domain (residues 669–919); the rest are unknown</td>
</tr>
<tr>
<td>Hereditary dentatorubral-pallidolysian atrophy&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Arthropin-1 with polyQ expansion</td>
<td>1185&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Unknown</td>
</tr>
<tr>
<td>Familial British dementia&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Aβ11</td>
<td>23</td>
<td>Natively unfolded</td>
</tr>
<tr>
<td>Familial Danish dementia&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Aβ12</td>
<td>23</td>
<td>Natively unfolded</td>
</tr>
</tbody>
</table>
4.5.1 Prion protein

- A prion is an infectious agent composed of protein in a misfolded state. Prions are responsible for transmissible spongiform encephalopathies.

- The folded structure (mainly helical) can convert to an amyloid prone structure that is able to act as a template to guide the misfolding of more protein into prion form and to build amyloid fibrils.

(Left) Ribbon view of the prion protein. (Right) Amyloid fibril structure obtained by cryo-electron microscopy with a cross-$\beta$ structure modeled into the electron density map. J.L. Jimenez et al., EMBO J. 18, 815, 1999

Stereoview of 20 NMR conformers of the bovine prion protein (abbreviated as PrP; PDB file 1dx1.pdb, $\alpha$-helices in red and $\beta$-sheet in yellow). Note that the PrP (monomeric) structure is mainly $\alpha$-helical. Note also the flexibility at the loops.
4.5.2 Alzheimer’s disease

- Alzheimer’s disease (AD) is the most common form of dementia.
- Amyloid-β protein accumulation is one of the major hallmarks of AD.
- It is not clear what are the toxic species. Accumulating evidence indicates that the early (soluble) oligomers are toxic.
- Alzheimer’s Aβ peptide (40 or 42 residues) originates from abnormal cleavage of the 770-residue amyloid precursor protein (in neurons) by β-secretase (BACE) and γ-secretase.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Gene Defect</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>β-Amyloid precursor protein mutations</td>
<td>Increased production of all β-amyloid proteins or β-amyloid protein 42</td>
</tr>
<tr>
<td>19</td>
<td>Apolipoprotein E4 polymorphism</td>
<td>Increased density of β-amyloid plaques and vascular deposits</td>
</tr>
<tr>
<td>14</td>
<td>Presenilin 1 mutations</td>
<td>Increased production of β-amyloid protein 42</td>
</tr>
<tr>
<td>1</td>
<td>Presenilin 2 mutations</td>
<td>Increased production of β-amyloid protein 42</td>
</tr>
</tbody>
</table>
Therapeutic strategies: inhibition of enzymes that generate the Aβ peptide by cleavage of the precursor (BACE and γ-secretase), small-molecule "modulators" of Aβ aggregation (Figure top), specific targeting of soluble oligomers and stabilization of less toxic fibril products by antibodies (Figure bottom).

Antonio Todde (1889-2002)
His secret: "Wine, pasta, and genes"
4.6 Amyloid fibrils: toxic or functional?

- Recently, it has been highlighted that living systems can utilize the amyloid structure as the functional state of some specific proteins (Chiti, Dobson, Annu. Rev. Biochem. 75, 333, 2006).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Organism</th>
<th>Function of the resulting amyloid-like fibrils</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curlin</td>
<td><em>Escherichia coli</em> (bacterium)</td>
<td>To colonize inert surfaces and mediate binding to host proteins</td>
</tr>
<tr>
<td>Chaplins</td>
<td><em>Streptomyces mediciolor</em> (bacterium)</td>
<td>To lower the water surface tension and allow the development of aerial hyphae</td>
</tr>
<tr>
<td>Hydrophobin[a] EAS</td>
<td><em>Neurospora crassa</em> (fungus)</td>
<td>To lower the water surface tension and allow the development of aerial hyphae</td>
</tr>
<tr>
<td>Proteins of the chorion of the eggs[b]</td>
<td><em>Bombyx mori</em> (silkworm)</td>
<td>To protect the oocyte and the developing embryo from a wide range of environmental hazards</td>
</tr>
<tr>
<td>Spidrin</td>
<td><em>Nephila edulis</em> (spider)</td>
<td>To form the silk fibers of the web</td>
</tr>
<tr>
<td>Intraluminal domain of Pmc17</td>
<td><em>Homo sapiens</em></td>
<td>To form, inside melanosomes, fibrous striations upon which melanin granules form</td>
</tr>
<tr>
<td>Ure2p (prion)</td>
<td><em>Saccharomyces cerevisiae</em> (yeast)</td>
<td>To promote the uptake of poor nitrogen sources ([URE3])</td>
</tr>
<tr>
<td>Sup35p (prion)</td>
<td><em>Saccharomyces cerevisiae</em> (yeast)</td>
<td>To confer new phenotypes ([PSI+]) by facilitating the readthrough of stop codons on mRNA</td>
</tr>
<tr>
<td>Rnq1p (prion)</td>
<td><em>Saccharomyces cerevisiae</em> (yeast)</td>
<td>Not well understood ([RNQ+], also known as [PIN+], phenotype)</td>
</tr>
<tr>
<td>HET-s (prion)</td>
<td><em>Podophora anserina</em> (fungus)</td>
<td>To trigger a complex programmed cell death phenomenon (heterokaryon incompatibility)</td>
</tr>
<tr>
<td>Neuron-specific isoform of CPEB (prion)</td>
<td><em>Aplysia californica</em> (marine snail)</td>
<td>To promote long-term maintenance of synaptic changes associated with memory storage</td>
</tr>
</tbody>
</table>

- Amyloid formation can be physiologically useful, provided it is regulated and allowed to take place under highly controlled conditions.

From E. Kandel’s recent publication in Cell Reports, 11, 1752, 2015:
CPEB3 is a functional prion with a tripartite prion domain architecture.
CPEB3 interacts with the actin cytoskeleton in yeast and at mouse synapses.
Actin filaments are essential for CPEB3 prionization and actin mRNA is a CPEB3 target.
Actin and CPEB3 constitute a positive feedback loop important for CPEB3 function.
4.7 Monte Carlo simulations of monomeric A\(_{\beta40}\) and A\(_{\beta42}\) (Excursus)

Central structures (large ribbon) of most populated clusters at 280 K and backbone of nine other structures from the same cluster (thin ribbons).

- Collapsed structure for both A\(_{\beta40}\) and A\(_{\beta42}\) with fluid hydrophobic core; see movie at [http://www.biochem-caflisch.uzh.ch/movie/8/](http://www.biochem-caflisch.uzh.ch/movie/8/)

- The N-terminal 11-residue segment is completely unstructured;

- All charged side chains, particularly Glu22 and Asp23 are exposed to solvent.
4 Protein aggregation

(a) Monte Carlo simulations

(b,c) Experimentally determined structures (NMR)

β-propensity at 310 K
Vitalis and A.C., J. Mol. Biol. 403, 148, 2010

Fibril model of Aβ1−42
Riek, PNAS 2005 (PDB 2BEG)

Aβ1−40 bound to an affibody

- Low β-sheet propensity but sequence specific and consistent with solid-state NMR

- The sequence partially encodes fibril structure, but fibril elongation must be thought of as a templated assembly step.
5 Protein folding

- The folding/unfolding process is important for:
  - action of certain proteins (e.g., titin, see Figure top left),
  - transfer across membranes (see Figure top right),
  - protein degradation (e.g., by the proteasome).

- There are many diseases related to protein misfolding, aggregation, and instability. They are caused by:
  - inherent properties of the wild-type protein,
  - changes in the protein environment,
  - mutation(s).

- Small proteins (up to the size of 100-150 residues or so) fold in the same way in vitro as in vivo, i.e., they do not require chaperones.
5.1 The funnel model of folding: a simple picture of a complex process

Fig. 5.1 Single pathway on a 'golf course' (left) and multiple pathways on 'funnel' (middle and right).

Fig. 5.2 Two-state approximation (left) and funnel (right). This plot shows that the folding barrier originates from the compensation of enthalpy and configurational entropy of the protein chain. The free energy of folding (horizontal axis on the left plot) shows enthalpy/entropy compensation which is worst at the transition state. During folding, the effective energy (vertical axis on the right plot) becomes more favorable while the configurational entropy (horizontal axis on the right plot) decreases, i.e., $-T\Delta S_{\text{protein}}$ becomes less favorable.
5.2 Protein folding mechanisms

- Proteins fold by diverse pathways.

- Variations of the nucleation-condensation mechanism describe the overall features of folding of most domains.

Fig. 5.3 Folding mechanisms (from Daggett and Fersht, Trends in Biochemical Sciences 28, 18, 2003)

- Secondary structure is inherently unstable and is stabilized by tertiary interactions (see Exercise on helical propensity of α helix 1 in barnase).

- Higher propensity for stable secondary structure \(\rightarrow\) framework (diffusion-collision) mechanism.

- Lower propensity for stable secondary structure \(\rightarrow\) hydrophobic collapse.
5.3 Folding kinetics

- **Folding**: The temperature dependence of the folding rate \( k_f \) shows a maximum, i.e., Arrhenius behavior at low temperatures and anti-Arrhenius behavior at high temperatures.

- **Unfolding**: The temperature dependence of the unfolding rate \( k_u \) shows Arrhenius behavior. It is a simpler "reaction" than folding.

![Graph showing ln(kf) and ln(ku) vs. 1/T (10^3 K^-1)](image)

- The contact order is the separation along the primary structure (i.e., the sequence) averaged over residues that are close in space in the tertiary structure. There is a weak anticorrelation between contact order and folding rate, i.e., mainly helical proteins fold faster than proteins with \( \beta \)-sheets, particularly parallel \( \beta \)-sheets (figure from Daggett and Fersht, NRCMB 4, 497, 2003).

![Graph showing log(k) vs. Contact order](image)
5.4 Two-state folding: An approximation of a complex process

- Several proteins with less than about 100 amino acids seem to populate only the folded and unfolded states at equilibrium.

- Two small proteins have been studied in detail during the last ten years: CI2 (chymotrypsin inhibitor 2 which consists of 64 residues) and barnase (ribonuclease of *B. amyloliquefaciens*, 110 residues).

- CI2 is a two-state folder (with folding time of 10 ms) whereas barnase has at least one folding intermediate (folding time of 50 ms).

- The transition state is an ensemble of (relatively) similar structures.

- Transition state for folding is often structurally similar to native structure for single-domain proteins.
5.5 Is folding really two-state? The unfolded state is complex!

- The unfolded state is an ensemble of heterogeneous conformations.

![Diagram of protein folding model](image)

Fig. 5.4 The two-state model (left) is a simplifying approximation while the conformational space (right) is very broad, especially in the denatured state [Caflisch and Hamm, 2012]. See also movie [http://www.biochem-caflisch.uzh.ch/movie/1/](http://www.biochem-caflisch.uzh.ch/movie/1/)

- In general, the high-T unfolded state is rather compact whereas the chemically (urea or guanidinium) induced denatured state is expanded.

- NMR spectroscopy is the most appropriate method to study the (partially) denatured states in solution.

- The ensemble is not completely random. There are:
  - some native and non-native interactions
  - transient hydrophobic clusters
  - residual but unstable regular secondary structure and turns.

**Exercise F1:** Why does a globular protein fold in water? Analyze the different enthalpic and entropic contributions (intraprotein, protein-water and water-water).

**Exercise F2:** A polypeptide sequence consisting of hydrophilic residues only (i.e., no hydrophobic side chains) does not fold at physiological temperature. Why?
5.6 Protein denaturation in vitro

- The low conformational stability of a folded protein (between $-15$ and $-5$ kcal/mol) allows to denaturate them by a variety of techniques that alter the balance of the weak nonbonding energy contributions.

- **Heating** modifies the entropy/enthalpy balance. One usually observes cooperative transitions.

- **pH variations** alter the ionization states of amino acid side chains. Low pH: excess of positive charges.

- **Detergents** (e.g. SDS) associate with the nonpolar residues of a protein.

- **Guanidinium ion or urea (chaotropic agents)** are the most commonly used chemical denaturants. Typical denaturant concentration ranges from 1 to 10 M. Chaotropic agents bind to both polar and nonpolar protein groups and “solvate” them better than pure water. Their mechanism of denaturation is not completely understood.

---

**Exercise F3:** There are some proteins that do not unfold at low pH. Why? Same question for those proteins that do not unfold at high pH.

**Exercise F4:** High temperature can be used together with a chemical denaturant to unfold a protein. Why? Explain the individual effects of temperature and urea as well as the effect of combining them.
5.7 Experimental and computational approaches to folding

- Folding at atomic-level resolution is investigated by experimental and computational approaches. One example of the latter is the simulation of reversible folding of a simplified-sequence variant of protein G consisting of only three different types of residues, i.e., Gly, Ala, and Thr (Guarnera et al., Biophys. J. 97, 1737, 2009). Its unfolded state is rather complex:

- Timescale problem (only few µs are accessible by molecular dynamics simulations).

- The synergy between experiments and simulations is increasing because of the discovery of ultrafast folding proteins (engrailed homeodomain, 61-residue α-helical protein which folds in 20-30 µs, see Nature 421, 863, 2003) and faster computers (Science 330, 341, 2010).
6 Molecular dynamics

6.1 Biophysical techniques

Fig. 6.1 Biophysical techniques vary in their temporal (x-axis) and spatial (y-axis) resolution. Molecular dynamics simulations have the highest resolution in time and space. Reprinted from Dror et al., *J. Gen. Physiol.* 135, 555, 2010.
6 Molecular dynamics

6.2 Simulations

Table 6.1 Classification of molecular systems

<table>
<thead>
<tr>
<th></th>
<th>Crystalline solid body</th>
<th>Liquid, macromolecule</th>
<th>Gas phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduction to fewer</td>
<td>Reduction to fewer</td>
<td>Many-body system</td>
<td>Reduction to few particles</td>
</tr>
<tr>
<td>degrees of freedom</td>
<td>degrees of freedom</td>
<td></td>
<td>through <em>dilution</em></td>
</tr>
<tr>
<td>due to <em>symmetry</em></td>
<td>due to <em>symmetry</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>properties</td>
<td>properties</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantum mechanics</td>
<td>Possible</td>
<td>Still impossible</td>
<td>Possible</td>
</tr>
<tr>
<td>Classical mechanics</td>
<td>Simple</td>
<td>Molecular dynamics</td>
<td>Trivial</td>
</tr>
<tr>
<td>QM/MM</td>
<td></td>
<td>Ex.: DNA-repair enzyme; Nature 434, p.612 (2005)</td>
<td></td>
</tr>
</tbody>
</table>

In computer simulations of (macro)molecular systems in the condensed phase, two fundamental problems arise:

- **the huge size of the configuration space** accessible to the molecular system
  
  \[ \rightarrow \text{statistical error}, \]

- **the accuracy** of the molecular model (e.g., explicit solvent or implicit solvent, atomistic or coarse-grained representation) and force field (e.g., fixed partial charges or polarizable model)

  \[ \rightarrow \text{systematic error}. \]