3.2. General data valid for all possible 3D structures

3.2.1. Geometric constraints

Three-dimensional structures or conformations are synonyms for a complete data set defining a molecule in the 3 dimensions of space and time.

A structure can be sufficiently described by intramolecular distances between atoms A and B. This distances are dependent on:

- **bonds** contributing parameter is the bond length A-B, (if A and B are connected directly);
- **angles** parameters are the bond length A-X, X-B and the angle A-X-B, (if A and B are in geminal position and connected to a common atom X);
- **torsions** parameters are the bond length A-X, X-Y, Y-B, the angles A-X-Y, X-Y-B and the dihedral angle, (if A and B are in vicinal position and connected via two common atoms X,Y).

The **bond length** r is a function of the special atom and thereby, of the atom radius or atom orbitals which contribute to a bonding. Bond length are varying in time around their local minimum-energy values since oscillations are caused by temperature. A deformation of 3-5% of the bond length (~ 5 pm) needs around 4 $kJmol^{-1}$ which is a quite high energy compared to room temperature and therefore can be neglected.

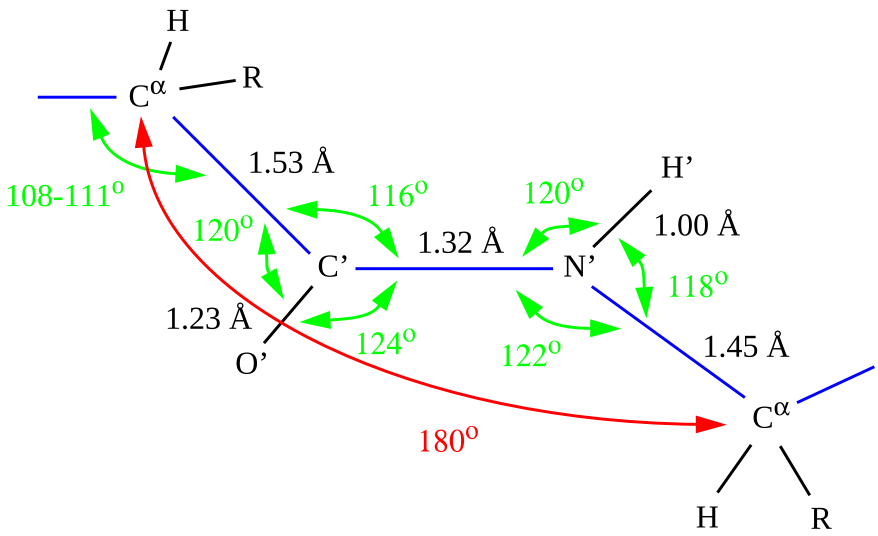
The **bond angle** α is influenced by the hybridization of the contributing atom orbitals, electronic effects of the participating atoms and steric repulsions of the substituents (known as 'Thorpe-Ingold-effect'). This can be seen in differences between H_2S (92^o) and H_2O (104^o) angles. The steric repulsion e.g. leads to a distortion by 5^o for the two angles CH_3-C-CH_3 (112^o) and H-C-H (107^o) in propane. The deformation energy for 8^o is about 2.4 $kJmol^{-1}$ which is appr. equal to RT.
[For carbohydrates: $E = 0.04\ kJmol^{-1}grad^{-2}*(\Delta\alpha)^2\ grad^2$].

The **dihedral angle** is the most important structural parameter since changes in torsions can influence the global molecular structure, while the two previously mentioned parameters act only on a local level.

Note the different definitions of torsion angle and dihedral angle (Nowadays torsion angle and dihedral angle are used synonymously)!
The torsion angle Θ_{AB} is the angle between the two projected bonds A-X and Y-B when looking along the central bond X-Y or Y-X. It can also be defined as angle between the two planes defined by the aforementioned bond angles. The sign is positive for clockwise rotation.
The dihedral angle Φ_{AB} is defined as the angle between normals to the planes containing A, X, Y and X, Y, B, respectively. Hence, $180^o = \Theta + \Phi$.

Although dihedral angles can be defined between every atom (also hydrogens), in biopolymers the torsion angles are as standard defined for the mainchain or sidechain heavy atoms. This causes still problems in X-ray structure determination (no hydrogens detectable).

The figure gives numerical values for the bond length (backbone = blue), the angles (green) and the torsion angle ω of the peptidic bond (red) of a peptide.



Through-bond interactions depend on the electronic nature of the bound atoms. They determine the rotational freedom around a bond, but are usually only defining the local conformation.

Nonbonded interactions between spatially separated atoms can be divided in repulsive and electrostatic contributions. These effects become important when we deal with macromolecules whose sizes allow to fold back. In this case numerous simultaneous interactions between different parts of the molecule occur.
The electrostatic forces or **Coulomb interactions** depend on the distance r_{AB} between two charges (q_A, q_B) in a solvent of polarizability D:

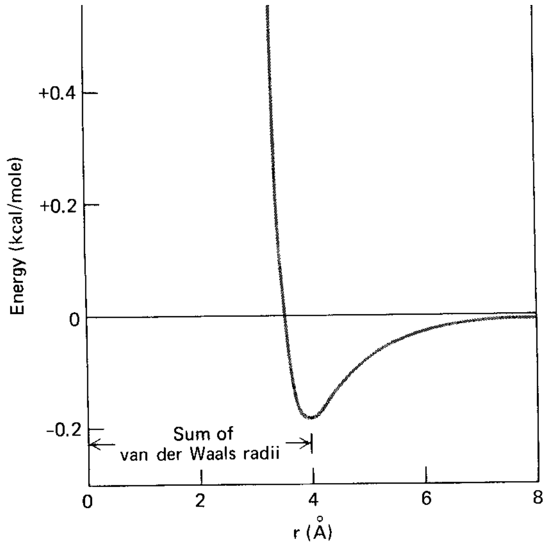
$$\Delta E = \frac{q_A \cdot q_B}{D \cdot r_{AB}}$$

Special electrostatic interactions are hydrogen bonds as an intermediate between covalent and ionic interactions. They need a special geometry between the hydrogen donor and the hydogen acceptor (linear=strong H-bond, N-O distance about 2.9 Å).
Van der Waals forces are an example for dipolar interactions. The energy contribution can be approximated using the **Lennard-Jones potential** function consisting of an attractive r^{-6}-dependent term and a repulsive r^{-12}-dependent term:

$$\Delta E = -\frac{A}{r^6} + \frac{B}{r^{12}}$$

The minimum of the function is the van der Waals contact distance which can be seperated into two atom characteristic radii, the **van der Waals radii**.

List of van-der-Waals radii and energy profile of C,C-van-der-Waals interaction calculated with A=1425 kcalÅ^{6}mole^{-1} and B=2.75 10^{6} kcalÅ^{12}mole^{-1}.



ATOM	COVALENT BONDING	VAN DER WAALS RADIUS (Å)
Hydrogen		1.00
Carbon	double bonds	1.60
	aromatic	1.70
	amide	1.50
Oxygen		1.35
Nitrogen	aromatic	1.70
	amide	1.45
Sulfur		1.70

From Handbook of Biochemistry. Cleveland, Chemical Rubber Publishing Co., 1968.

3.2.2. Description of 3D structures

Once the topology of the building blocks or fragments of a biopolymer is defined in a library, the structure of a molecule can be defined in two ways:

- by the cartesian coordinates of atoms;
- by internal coordinates, the torsion angles.

Internal and cartesian coordinate set for a dipeptid fragment.

Dihedral angle description (internal coordinates)

1	LEU	PHI	-130.949
1	LEU	CHI1	-150.614
1	LEU	CHI2	72.597
1	LEU	CHI31	-127.183
1	LEU	CHI32	-168.829
1	LEU	PSI	160.768
2	LYS	PHI	40.956
2	LYS	CHI1	-80.130
2	LYS	CHI2	-89.254
2	LYS	CHI3	100.713
2	LYS	CHI4	100.519
2	LYS	CHI5	-113.212
2	LYS	PSI	63.871

Cartesian coordinates in Brookhaven Protein Data Bank (PDB) format

ATOM	1	N	LEU	1	1.325	0.000	0.000
ATOM	2	HN	LEU	1	1.884	0.000	0.829
ATOM	3	CA	LEU	1	2.073	0.000	-1.245
ATOM	4	HA	LEU	1	1.375	-0.241	-2.046
ATOM	5	CB	LEU	1	2.630	1.395	-1.536
ATOM	6	HB2	LEU	1	1.967	2.131	-1.082
ATOM	7	HB3	LEU	1	3.597	1.489	-1.042
ATOM	9	CG	LEU	1	2.807	1.753	-3.013
ATOM	10	HG	LEU	1	3.268	0.903	-3.517
ATOM	13	CD1	LEU	1	1.453	1.995	-3.684
ATOM	14	HD11	LEU	1	1.375	1.377	-4.579
ATOM	15	HD12	LEU	1	0.653	1.735	-2.992
ATOM	16	HD13	LEU	1	1.368	3.046	-3.961
ATOM	17	CD2	LEU	1	3.751	2.945	-3.180
ATOM	18	HD21	LEU	1	3.706	3.303	-4.208
ATOM	19	HD22	LEU	1	3.450	3.744	-2.503
ATOM	20	HD23	LEU	1	4.770	2.636	-2.947
ATOM	22	C	LEU	1	3.145	-1.091	-1.192
ATOM	23	O	LEU	1	3.508	-1.557	-0.112
ATOM	24	N	LYS	2	3.622	-1.466	-2.369
ATOM	25	HN	LYS	2	3.321	-1.082	-3.242
ATOM	26	CA	LYS	2	4.645	-2.493	-2.470
ATOM	27	HA	LYS	2	4.600	-2.900	-3.480
ATOM	28	CB	LYS	2	6.036	-1.883	-2.288
ATOM	29	HB2	LYS	2	6.012	-1.145	-1.486
ATOM	30	HB3	LYS	2	6.740	-2.658	-1.987
ATOM	32	CG	LYS	2	6.518	-1.223	-3.582
ATOM	33	HG2	LYS	2	5.666	-0.821	-4.129
ATOM	34	HG3	LYS	2	7.170	-0.383	-3.344
ATOM	36	CD	LYS	2	7.269	-2.225	-4.461
ATOM	37	HD2	LYS	2	8.154	-1.750	-4.886
ATOM	38	HD3	LYS	2	7.617	-3.060	-3.853
ATOM	40	CE	LYS	2	6.373	-2.746	-5.587
ATOM	41	HE2	LYS	2	6.521	-3.819	-5.710
ATOM	42	HE3	LYS	2	5.326	-2.596	-5.325
ATOM	44	NZ	LYS	2	6.677	-2.048	-6.856
ATOM	45	HZ1	LYS	2	5.906	-1.466	-7.164
ATOM	46	HZ2	LYS	2	7.486	-1.443	-6.773
ATOM	48	C	LYS	2	4.332	-3.617	-1.480
ATOM	49	O	LYS	2	5.098	-3.860	-0.549

Cartesian coordinates in DIANA format (only 1 amino acid)

1	N	1	LEU	1.3249	0.0000	0.0000
2	HN	1	LEU	1.8841	0.0000	0.8290
3	CA	1	LEU	2.0733	0.0000	-1.2455
4	HA	1	LEU	1.3746	-0.2415	-2.0464
5	CB	1	LEU	2.6299	1.3952	-1.5361
6	HB2	1	LEU	1.9668	2.1313	-1.0816
7	HB3	1	LEU	3.5968	1.4891	-1.0418
8	QB	1	LEU	2.7818	1.8102	-1.0617
9	CG	1	LEU	2.8067	1.7525	-3.0133
10	HG	1	LEU	3.2683	0.9034	-3.5173
11	QD1	1	LEU	1.1321	2.0527	-3.8438
12	QD2	1	LEU	3.9753	3.2278	-3.2194
13	CD1	1	LEU	1.4534	1.9951	-3.6844
14	HD11	1	LEU	1.3754	1.3769	-4.5788
15	HD12	1	LEU	0.6526	1.7352	-2.9920
16	HD13	1	LEU	1.3680	3.0461	-3.9607
17	CD2	1	LEU	3.7510	2.9447	-3.1798
18	HD21	1	LEU	3.7062	3.3029	-4.2083
19	HD22	1	LEU	3.4495	3.7443	-2.5031
20	HD23	1	LEU	4.7702	2.6363	-2.9466
21	QQD	1	LEU	2.5536	2.6402	-3.5316
22	C	1	LEU	3.1450	-1.0906	-1.1915
23	O	1	LEU	3.5076	-1.5567	-0.1125

3.2.3. Sources for structural data

A reliable set of standard geometries for biopolymer building blocks is critical in the refinement of NMR or X-ray structures of macromolecules as well as in molecular modelling.
In principle, two ways are available to determine this basis sets:
- an empirical survey of high-resolution small molecule crystal structures based on statistics;
- theoretical calculations.

The main source of conformational data for amino acids, nucleotides and nucleosides is the X-ray crystallography. Therefore, the set of standards is time-dependent and increases with the number of available atomic resolution structures allowing to reinvestigate the basic geometrical parameters.
The examples a.) and b.) show that because of the increasing number of available structures it is not only possible to probe more reliable model structures but also to use more stringent structure selection criteria for the statistical analysis.
a.) Earlier studies of nucleic acid fragments were forced to include furanose rings other than ribose or 2'deoxyribose in the determination of parameters. But even then, there were not sufficient data available to describe the continuum of states associated with the interconversion of the common ring conformations.
b.) A recent statistics on nitrogen bases was based on the following criteria:
1. the structural resolution had to be better than 1 Å,
2. the R-value of the structures had to be better than 6%,
3. the C-C bond length had to have an average estimated standard deviation of less than 0.01 Å.

However, there still might be differences to solution geometries since all parameters were determined from molecules in a crystal lattice.

Parameters for protein secondary structure elements (torsion angles in degree).

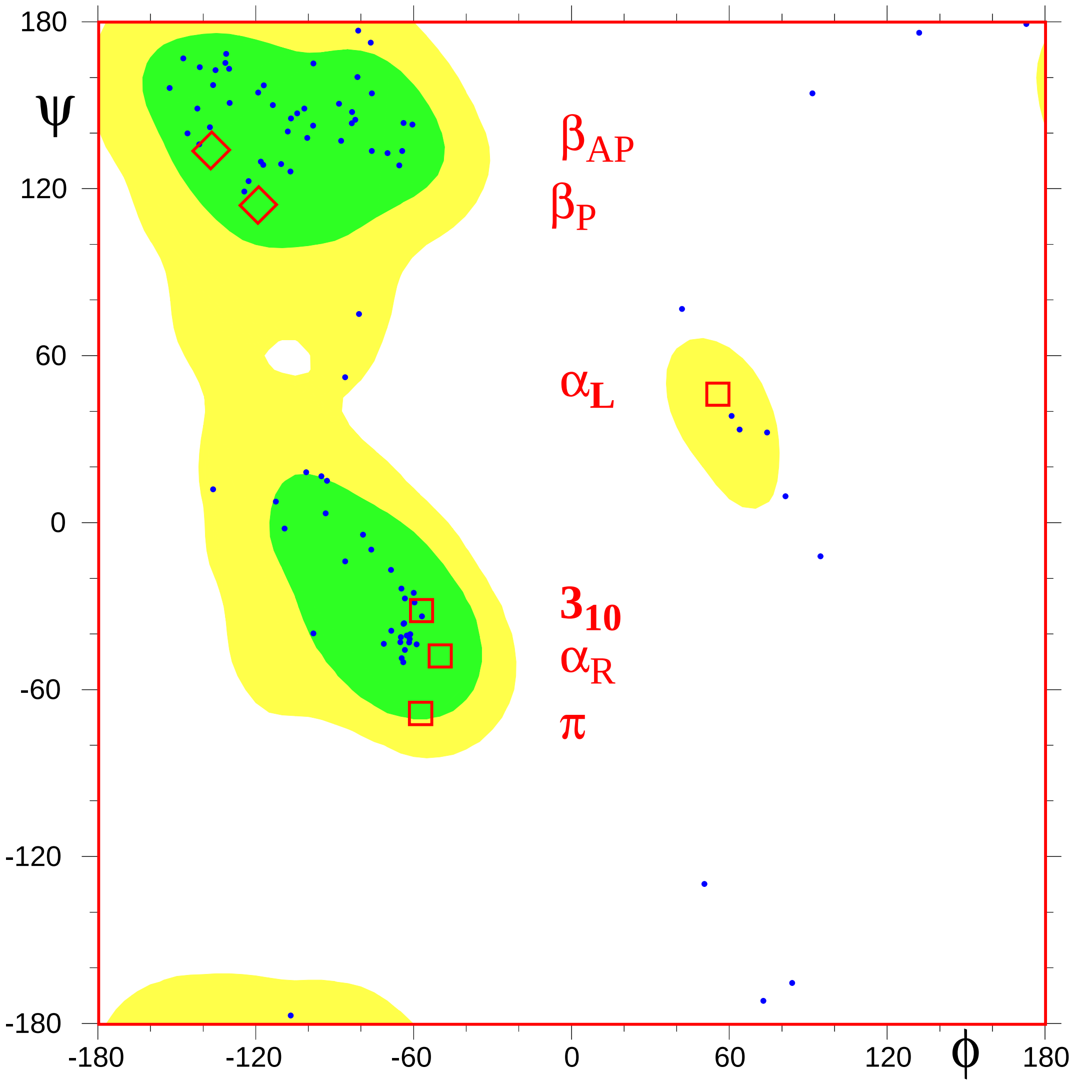
secondary structure element	ϕ	ψ	ω	residues per turn	translation per residue (Å)
antiparallel β-sheet	-139	+135	-178	2.0	3.4
parallel β-sheet	-119	+113	180	2.0	3.2
right-handed α-helix	-57	-47	180	3.6	1.5
3_{10} helix	-49	-26	180	3.0	2.0
π helix	-57	-70	180	4.4	1.2

3.2.4. Consequences of geometric constraints

Proteins

Since the description of a peptide backbone is virtually complete by giving ϕ- and ψ-torsions, a two-dimensional plot of them, called Ramachandran plot, provides an overview about the favourable amino acids conformations.

Ramachandran plot for the pdb-structure of RNase T1. (The allowed regions (green=95%, yellow=80% of all amino acids) were determined from 109 different crystal structures with a resolution of 2.5 Å or better. Gly and Pro residues were excluded.)



For most L-amino acids the fraction of the at least partially allowed regions to the total area is about 23%. As given in the figure all conformations are found at negative ϕ-torsions.
For glycine this ratio increases to 61% since no steric restrictions are introduced by a sidechain interaction and also positive ϕ-torsions can be adopted. (Indeed, the points on the right side and not allowed regions of the above figure are the 12 glycines.)
Ramachandran maps may also give information on the relative energies of each conformation mainly based on empirical energy functions.
For geometries of typically allowed structures compare the figures shown in the secondary structure elements section.

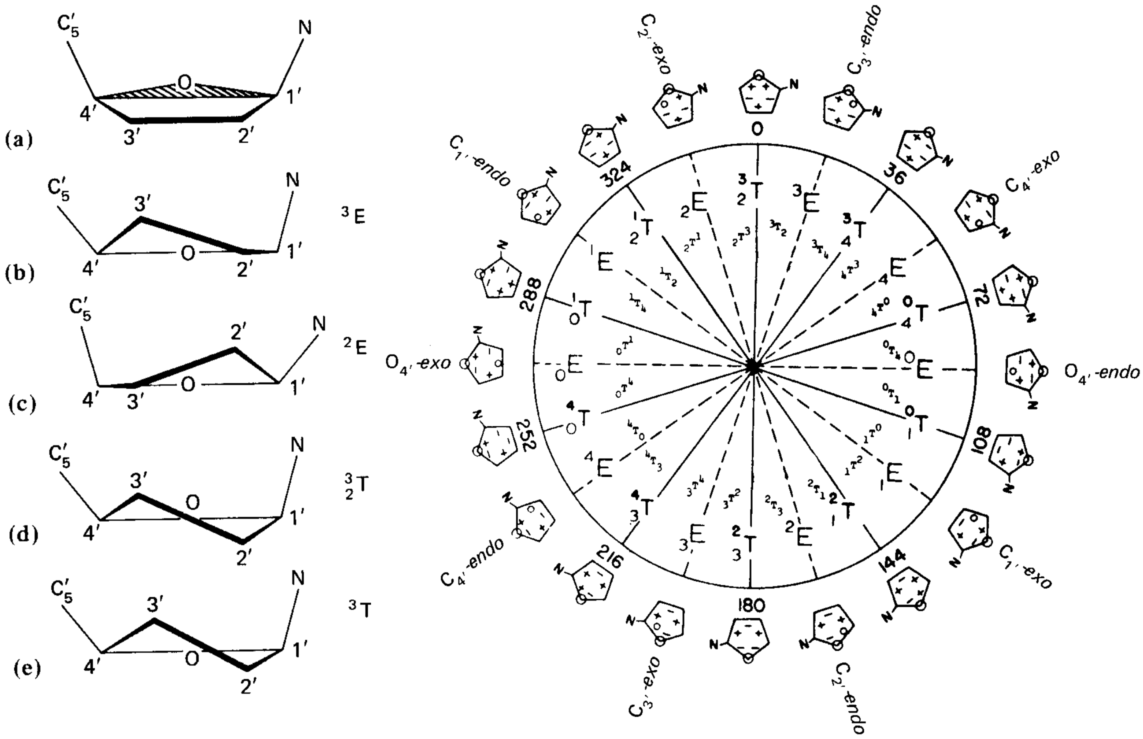
Nucleic acids & proteins

In proteins one amino acid, the proline exhibits a specific geometry. The sidechain is covalently bonded to the nitrogen resulting in a fairly rigid 5-membered ring.

A similar geometry is observed for the ribose units of nucleic acids in which one ring atom is an oxygen.

These five-membered rings, although restricted in the accessible conformational space, undergo at room temperature local conformational changes between either an envelope or a half-chair form. This phenomenon is called *pucker* and is a concerted process of the ring torsions, thereby avoiding a totally eclipsed and energetically unfavourable conformation. The maximum seperation between intermediate conformations is an energy barrier of over 5 kcalmol^{-1}. Because of the different energy levels of intermediates arising from influences of ring substituents, some pucker states are preferred. For RNA A-form helices the $C_{3'}$-endo state is stronger populated, while B-form helices more often adopt the $C_{2'}$-endo conformation.

Pseudorotation cycle and sugar pucker modes (Saenger, 1983).



Nucleic acids

One exceptional feature of nucleic acids is their ability to form pairs between adjacent bases by hydrogen bonding. This requires complementary sequences of bases in one or two strands satisfying the stereochemical requirements and a special geometry to establish the H-bonds. This geometry is a double helix in which the bases with their hydrophobic surfaces are stacking together inside while the negatively charged phosphates form a spiral outside. Several types of helix geometries are observed.

- **A-form helix** (standard for RNA)
- **B-form helix** (standard for DNA)
- **Z-form helix** (DNA)