Chapter 1
Macromolecular Structure and Dynamics

1.1 Physical properties

Biological macromolecules, protein, RNA, DNA & polysaccharides

Provides a description of their structures at various levels, from the atomic level to large multisubunit assemblies.

Their behavior in electric, magnetic, or centrifugal fields

Basic principles of structure and structural complexity found in biological macromolecules.

1.1.1 Macromolecules

What is a molecule?

Chemistry:

covalent bonded in specific proportions according to weight or stoichiometry and with unique geometry.

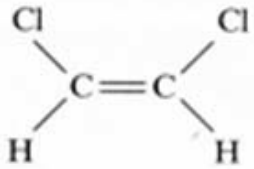
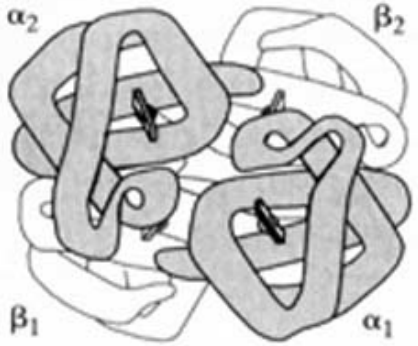
Molecule	Stoichiometry	Geometry
cis-Dichloroethylene	$C_2H_2Cl_2$	
Hemoglobin	$\alpha_2\beta_2$	

Figure 1.1 Examples of molecules in chemistry and macromolecules in biochemistry. The simple compound *cis*-dichloroethylene is uniquely defined by stoichiometry of its atomic components and the geometry of the atoms. Similarly, the structure of a biological macromolecule such as hemoglobin is defined by the proportions of the two subunits (the α- and β-polypeptide chains) and the geometry by the relative positions of the subunits in the functional complex.

What is a molecule?

Biochemist:
not necessary covalent bonded
but noncovalently associated polymers.

Ex: Hemoglobin
4 subunits/ monomer units/$\alpha_2\beta_2$
Large and complexity

atoms $\Rightarrow$ functional groups $\Rightarrow$ monomer/multimer etc.,

What is considered to be large?

The DNA of human chromosome/ tens of billions of atoms

25 residues/ oligomers

DNA condensing j-protein of the virus g4/24 aa

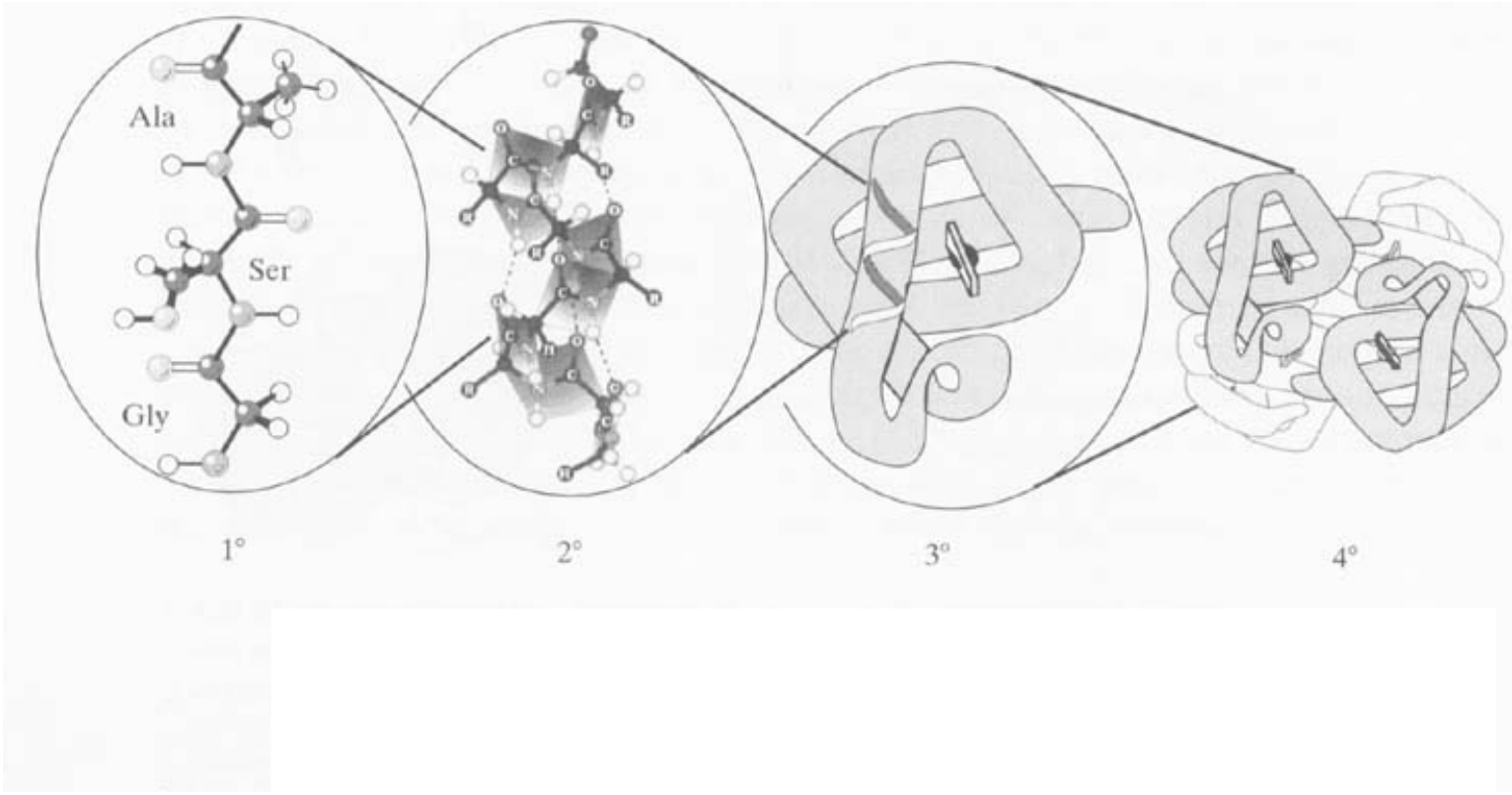
Monomers: building blocks (aa/sugars) polymerized to a macromolecule.

Primary structure (1°): linear arrangement/ colvent linked polymer

Secondary structure (2°): local regular structure, helical structures

Tertiary structure (3°): 3-D topology of the molecule, functional molecule structure. domain, motif etc.

Quaternary structure (4°): multiple distinct polymers (or subunit) that form a functional comples. Tetramer, dimer etc.



1° ⇒ 2° ⇒ 3° ⇒ 4° (if present)

How molecule folds into its functional form?
Not clear

•"**molten globule state**"
less compact 3° structure; must occur to form the environment to stabilize 2° structure.

•Protein-folding problem

•Model: described the atoms and the **positions** of the atoms in the 3D space

•**Atomic coordinates** (x,y,z) in space.

1.1.2 Configuration and Conformation

The **arrangement** of atoms or groups of atoms in a molecule is described by the terms configuration and conformation.

- **Configuration**

- **Conformation**

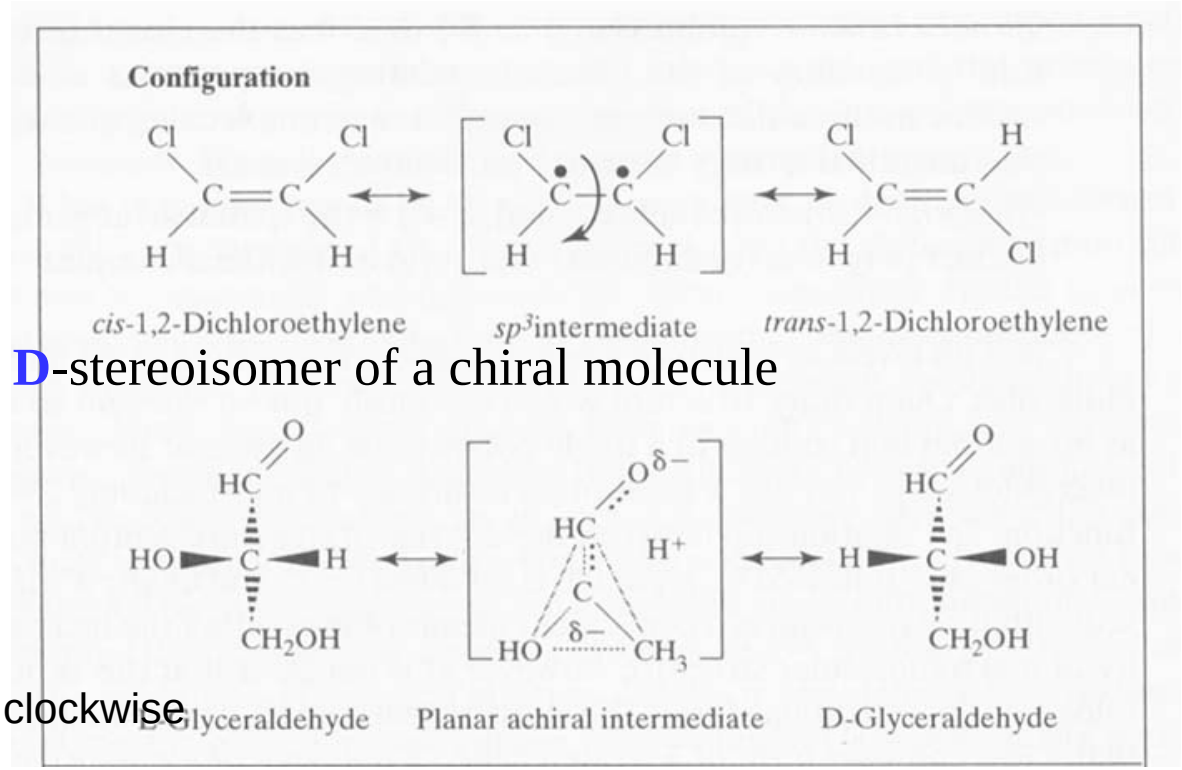
•Configuration

•The position of groups around one or more nonrotating bonds or around chiral centers

•defined as an atom having no plan or center of symmetry.

•To change the configuration of a molecule, chemical bonds must be broken and remade.

•Ex. **cis**- or **trans**-configurations



Ex: **L**- & **D**-stereoisomer of a chiral molecule

anti clockwise

Conformation

•The arrangement of groups about one or more freely rotating bonds.

•A molecule does not require any changes in chemical bonding to adopt a new conformation, but may require a new set of properties that are specific for that conformation.

Ex: gauche and anti structural isomers

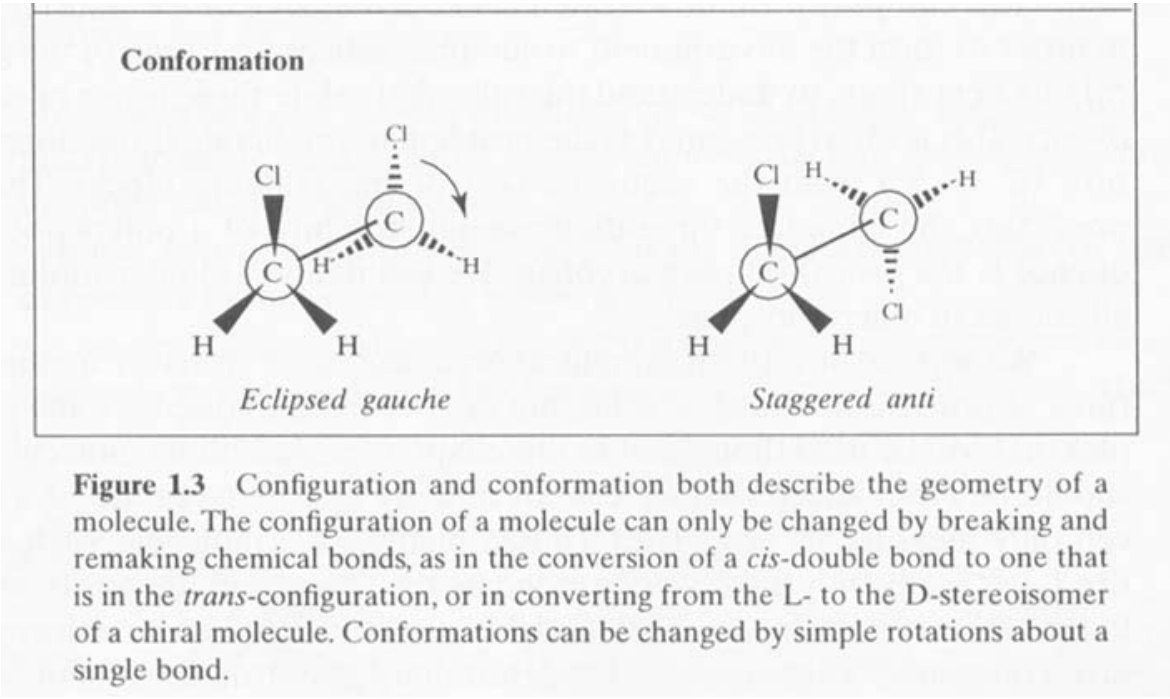


Figure 1.3 Configuration and conformation both describe the geometry of a molecule. The configuration of a molecule can only be changed by breaking and remaking chemical bonds, as in the conversion of a *cis*-double bond to one that is in the *trans*-configuration, or in converting from the L- to the D-stereoisomer of a chiral molecule. Conformations can be changed by simple rotations about a single bond.

The Stereochemistry of monomers

Most biological macromolecules are chrial molecules

L- and D-glyceraldehyde
L-: rotate in an anti-clockwise direction around the chrial carbon.
D-: rotate in a clockwise direction around the chrial carbon.

Biopolymers are typically constructed from only one enantimer (L form) of the monomer building block.

Amino acid / the chiral center is the carbon directly adjacent to the carboxylic acid (the **C**α-carbon)

The Stereochemistry of monomers

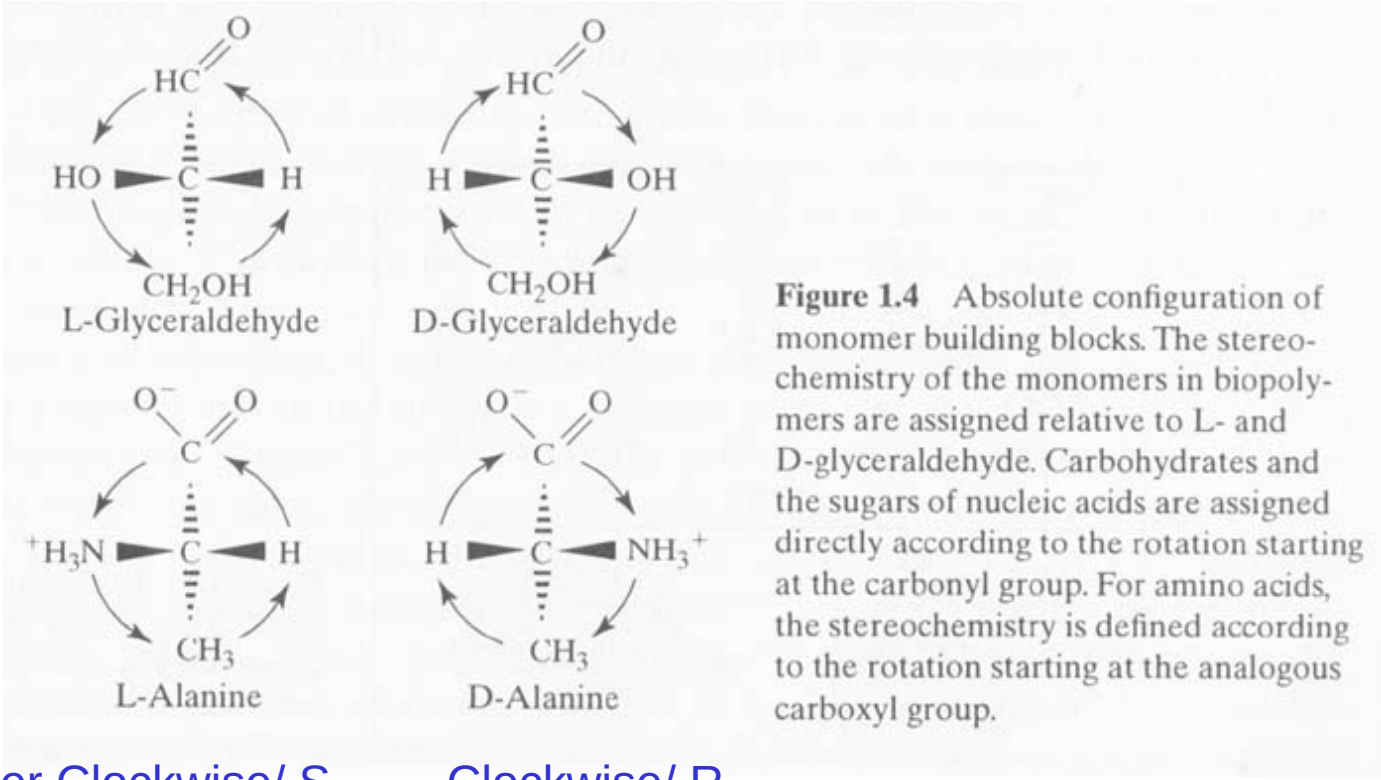


Figure 1.4 Absolute configuration of monomer building blocks. The stereochemistry of the monomers in biopolymers are assigned relative to L- and D-glyceraldehyde. Carbohydrates and the sugars of nucleic acids are assigned directly according to the rotation starting at the carbonyl group. For amino acids, the stereochemistry is defined according to the rotation starting at the analogous carboxyl group.

Counter Clockwise/ S Clockwise/ R

Conformation of molecules

Torsion angle: "θ" (-180° to +180°)
The angle between two groups on either side of freely rotating chemical bond.

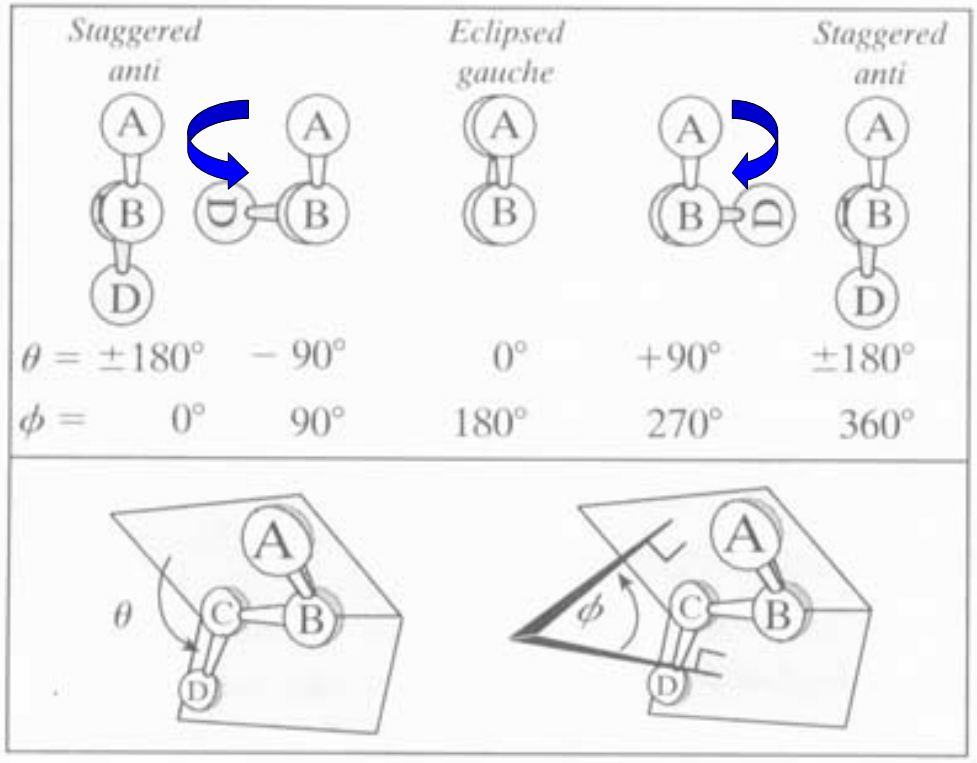
Dihedral angle "ϕ" (0° to +360°)
The angle between the normals of the planes formed by the atoms A-B-C and that of atoms B-C-D.

Changed the conformation of a molecules does not make a new molecule, but can change its properties

Torsion angle and Dihedral angle are Complement $\phi = \theta + 180°$

Conformation of molecules

The rotation around a single bond is described by torsion angle θ of the **4** atoms around the bound A-B-C-D



Torsion angle: " θ "
(-180° to +180°)

Dihedral angle " ϕ "
(0° to +360°)

$$\phi = \theta + 180°$$

Properly folded conformation of a protein

⇒native conformation

⇒functional form

Unfolded or denatured conformation

⇒nonfunctional

⇒ proteolysis by the cell

1.2 Molecular interaction in Macromolecular Structure

Configuration is fixed by covalent bonding

Conformation is highly variable and dependent on a number of factors

Folding of macromolecules depends on a number of interactions, including the interactions between atoms in the molecule and between the molecule and its environment

1.2.1 Weak interactions

Conformation of a macromolecule is stabilized by *weak interactions* with energies of formation that are at least one order of magnitude less than that of a covalent bond.

Distance-dependent interactions

Inversely proportional to the distance r (or r2, r3 etc)

1.2.1 Weak interactions

TABLE 1.1 RELATIONSHIP OF NONCOVALENT INTERACTIONS TO THE DISTANCE SEPARATING THE INTERACTING MOLECULES, *r*

Type of Interaction	Distance Relationship
Charge-charge	$1/r$
Charge-dipole	$1/r^2$
Dipole-dipole	$1/r^3$
Charge-induced dipole	$1/r^4$
Dispersion	$1/r^6$

Longer range interactions

charge-charge α $1/r$
charge-dipole α $1/r^2$
dipole-dipole α $1/r^3$

Short range interactions

dipole-induced dipole interaction (dispersion) α $1/r^4$
dispersion (very short-range interaction ~1nm) α $1/r^6$
steric repulsion α $1/r^{12}$

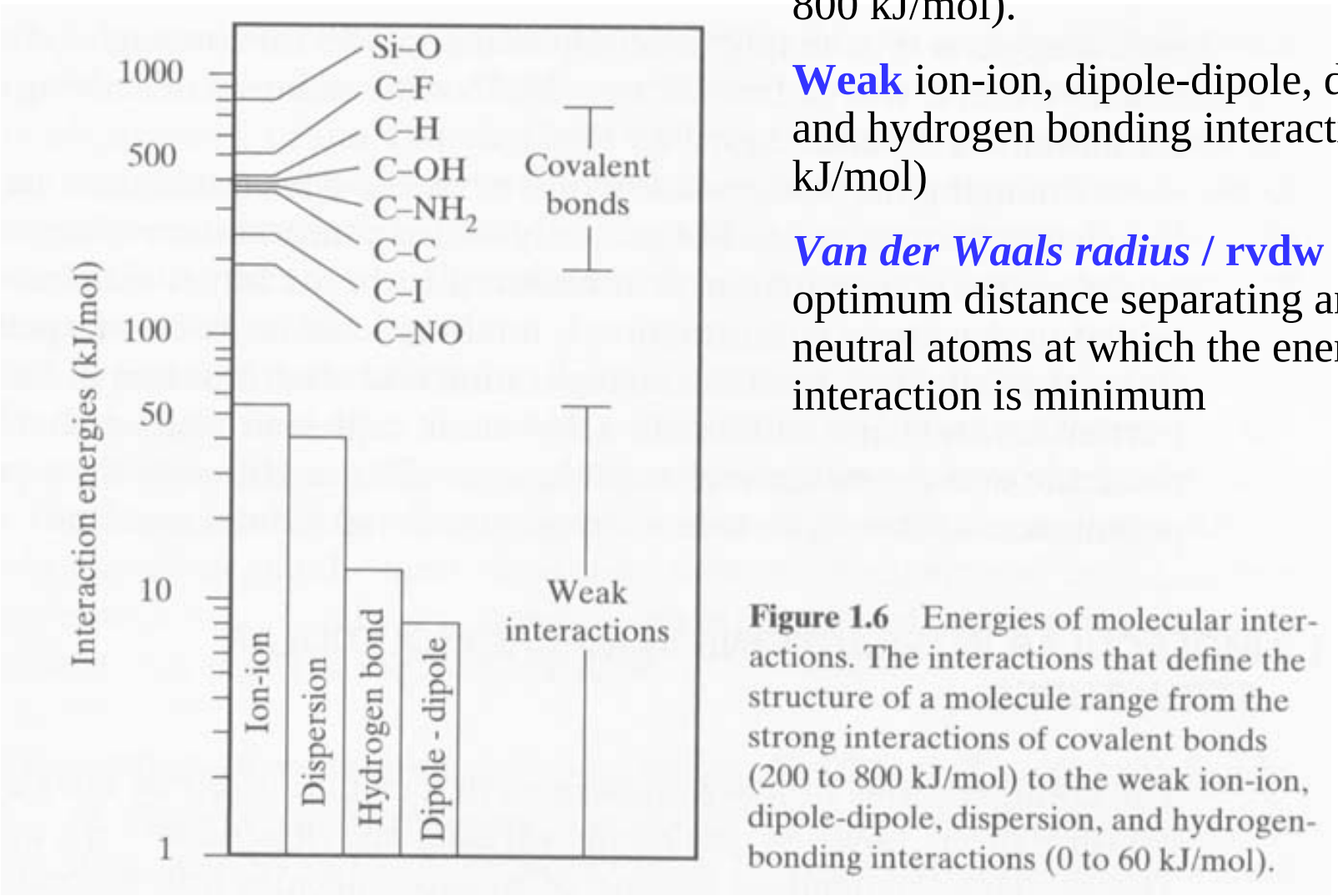


Figure 1.6 Energies of molecular interactions. The interactions that define the structure of a molecule range from the strong interactions of covalent bonds (200 to 800 kJ/mol) to the weak ion-ion, dipole-dipole, dispersion, and hydrogen-bonding interactions (0 to 60 kJ/mol).

Strong interaction of covalent bind (200-800 kJ/mol).

Weak ion-ion, dipole-dipole, dispersion and hydrogen bonding interactions (0-60 kJ/mol)

***Van der Waals radius* / rvdw** : an optimum distance separating any tow neutral atoms at which the energy of interaction is minimum

Longer-range interaction

Longer-range interactions (charge-charge, charge-dipole and dipole-dipole) are dependent on the intervening medium, shielded in a polar medium and weakened.

The least polarizable medium is **vacuum**,
dielectric constant of $k\varepsilon o = 4\ \pi\ 8.85 \times 10^{-12}\ C^2/J\ m$ (D)=1

Inversely related to the dielectric of the medium

Weakened in a highly polarizable medium such as **water** (~80D).

Dielectric constant/ the environment factor in stabilizing the conformation of a macromolecule.

How the environment affects the weak interactions

2 additional interactions (hydrogen bonds & hydrophobicity)

1.3 The Environment in the Cell

Biological system
70% water, aqueous solution, dilute aqueous solution

Membranes
Nonaqueous environment,
For protein that are integral parts of the bilayer of the membranes

ex: TATA-binding protein
An important aromatic interaction between a **Phe** of protein and the nucleotide base of the bound DNA.
Represent an important nonaquous enviroment

Solvent molecules
Water (ex: between protein and its bound DNA)
often helps to mediate interaction,
treated as part of the macromolecule rater than part of the bulk solvent.

1.3.1 Water structure

Intramolecular interaction /within molecules

Intermolecular interaction /between molecules

H_2O molecule

Tetrahedral, SP^3 Oxygen atom

Two hydrogens and two pairs of nonbonding electrons

“O” is more electronegative than “H”

1.3.1 Water structure

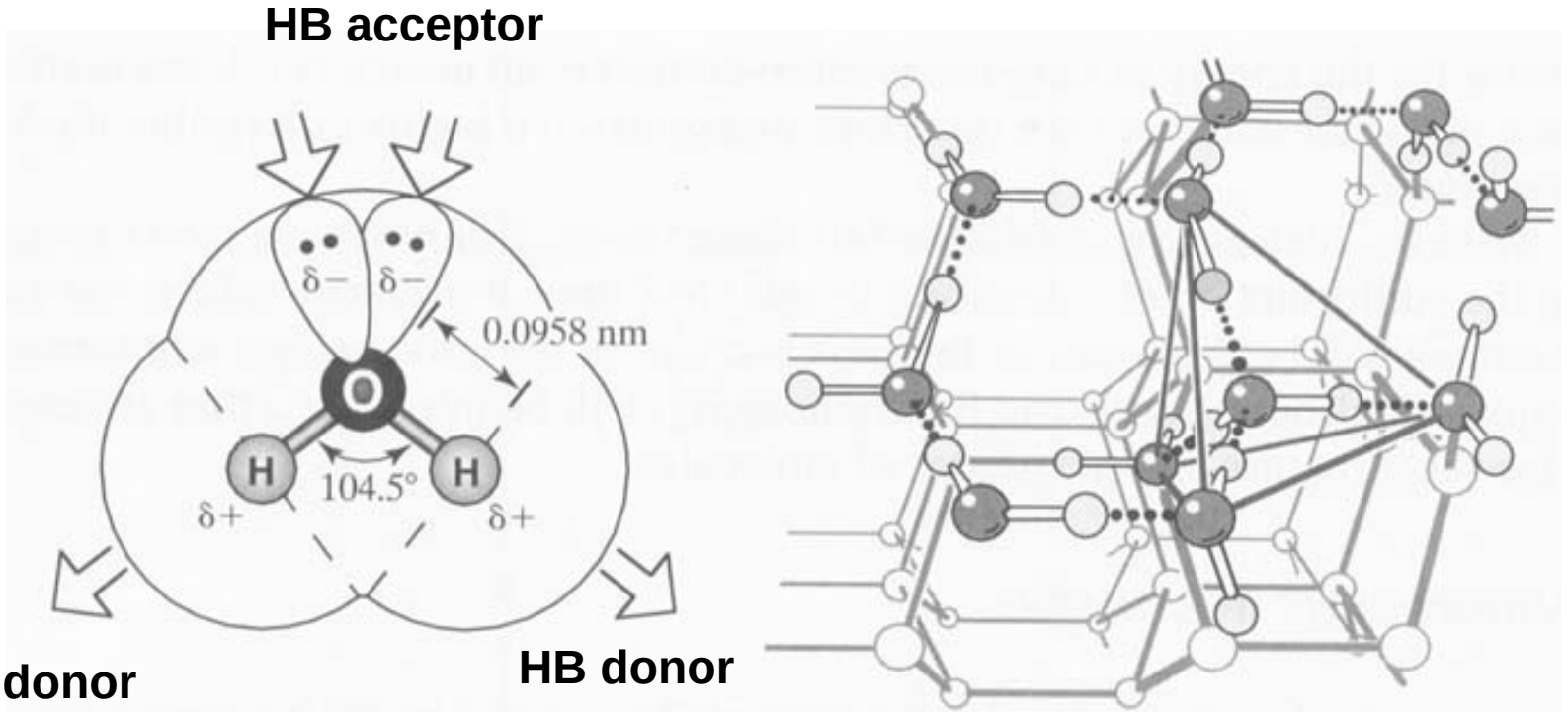


Figure 1.7 The structure of water. Each H_2O molecule has two hydrogens and two lone pairs of unbonded electrons at each oxygen. In ice, the hydrogens act as hydrogen-bond donors to the lone pairs of the oxygens, which act as hydrogen-bond acceptors. This results in a hexagonal lattice of hydrogen-bonded water molecules, with each H_2O molecule having four neighbors arranged in a tetrahedron. [Adapted from Mathews and van Holde (1996), *Biochemistry,* 2d ed., Benjamin-Cummings Publishing Co., Menlo Park, CA, p. 33.]

Higher **electronegativity** higher electron affinity

TABLE 1.2 ELECTRONEGATIVITIES OF ELEMENTS TYPICALLY FOUND IN BIOLOGICAL MOLECULES

Element	Electronegativity
O	3.5
Cl	3.0
N	3.0
S	2.5
C	2.5
P	2.1
H	2.1
Cu^{2+}	1.9
Fe^{2+}	1.8
Co^{2+}	1.8
Mg^{2+}	1.2
Ca^{2+}	1.0
Na^{+}	0.9
K^{+}	0.8

Higher values indicate a higher electron affinity.

The **dipole moment** of O-H is from the "H" (+ end) to the "O" (- end)

1.86 debye (deye=3.336 x10^{-30} C/m) (Isolated water)

2.6 debye (a cluster of 6 waters or more)

3 debye (ice)

Water is highly polarizable, as well as being polar.

has a high dielectric constant relative to a vacuum (D ~80 D, kεo)

water-water hydrogen bond, hydrogen-bond donor & hydrogen-bond acceptor, form a **hydrogen bond network**.

H-bond: 2.5 ~ 3.2 / 3.5A

TABLE 1.3 HYDROGEN-BOND DONORS AND ACCEPTORS IN MACROMOLECULES

Donor	Acceptor	*r* (nm)
	r	
N—H···	O	0.29
N—H···	O—H	0.29
N—H···	N	0.31
N—H···	S	0.37
O—H···	O	0.28
O—H···	O—H	0.28

Equation 1.1 is reduced to the standard equation for self-dissociation of water

Water freezes to different ice forms

Crystalline ice form of water/tetrahedral and hexagonal arrays
Liquid water freezes to different ice forms dep on **Temp** & **Pressure**
"H" can only be ordered precisely at pressure>20Kbars and temp<0ºC

Ice VIII
more ordered form, hexagonal arrays (low T & high P)

Ice IX
more compact form, pentagonal arrays

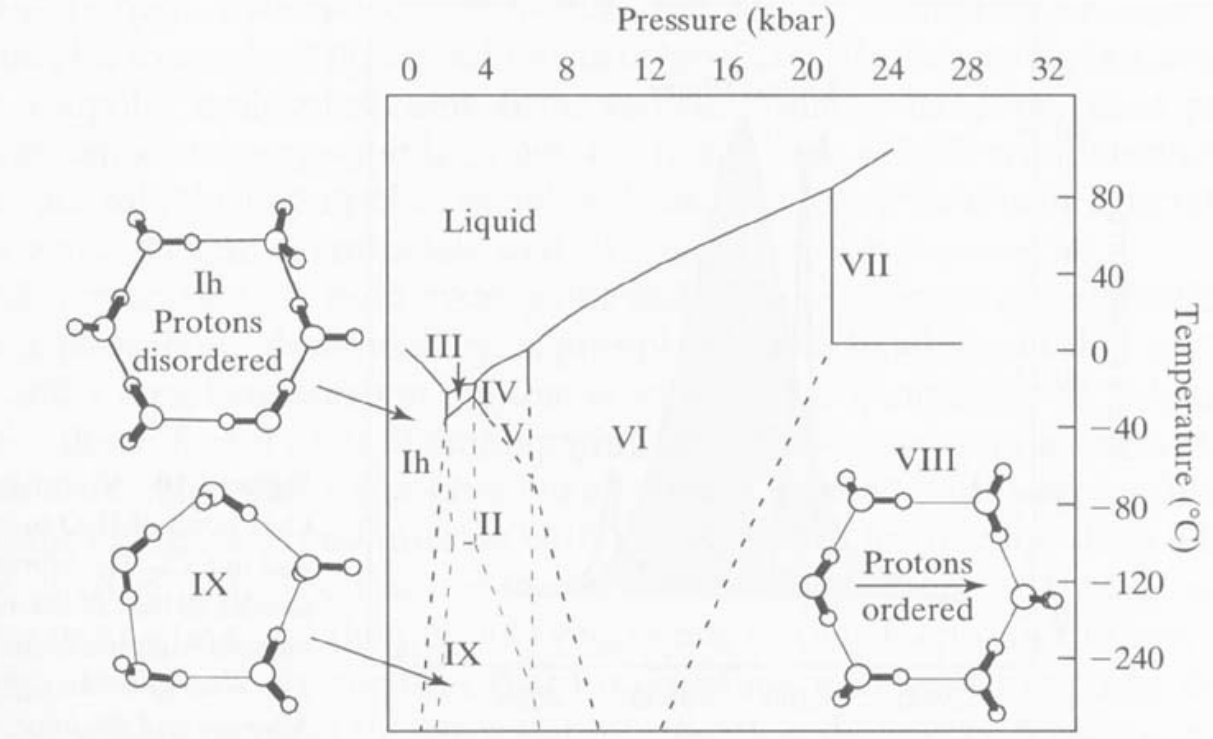


Figure 1.8 Phase diagram for water. Liquid water freezes to different ice forms, depending on the temperature and pressure. Under normal conditions, ice is a hexagonal network in which the protons of the hydrogen bonds are equally shared and cannot be assigned to a specific oxygen center (ice Ih). More compact forms (e.g., ice IX) or more ordered forms (e.g., ice VIII) are observed at low temperatures and high pressures. [Adapted from H. Savage and A. Wlodawer (1986), *Meth. Enzymol.*, **127**: 162–183.]

The solvent structure is composed of regular Hexagonal & pentagonal faces

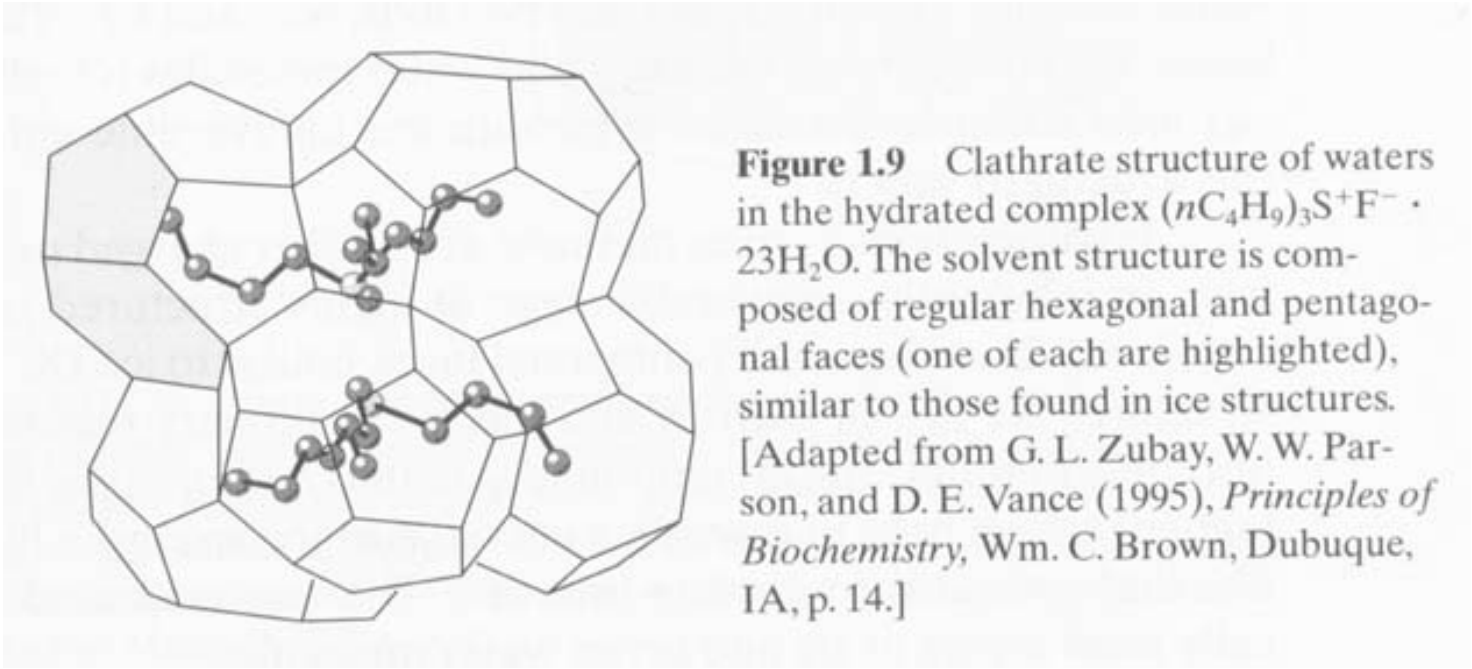


Figure 1.9 Clathrate structure of waters in the hydrated complex $(nC_4H_9)_3S^+F^- \cdot 23H_2O$. The solvent structure is composed of regular hexagonal and pentagonal faces (one of each are highlighted), similar to those found in ice structures. [Adapted from G. L. Zubay, W. W. Parson, and D. E. Vance (1995), *Principles of Biochemistry,* Wm. C. Brown, Dubuque, IA, p. 14.]

Cage-like clathrate structure

Vibration Frequency of O-H bond of H2O

The structure of liquid water is very similar to that of ice I (0°C/ 1 atm)

Vibration frequency : O-H bond / ice < liq water <CCl_4

The structure of liquid water is more dynamic than ice the pattern of H-bonds changing about every p-sec.

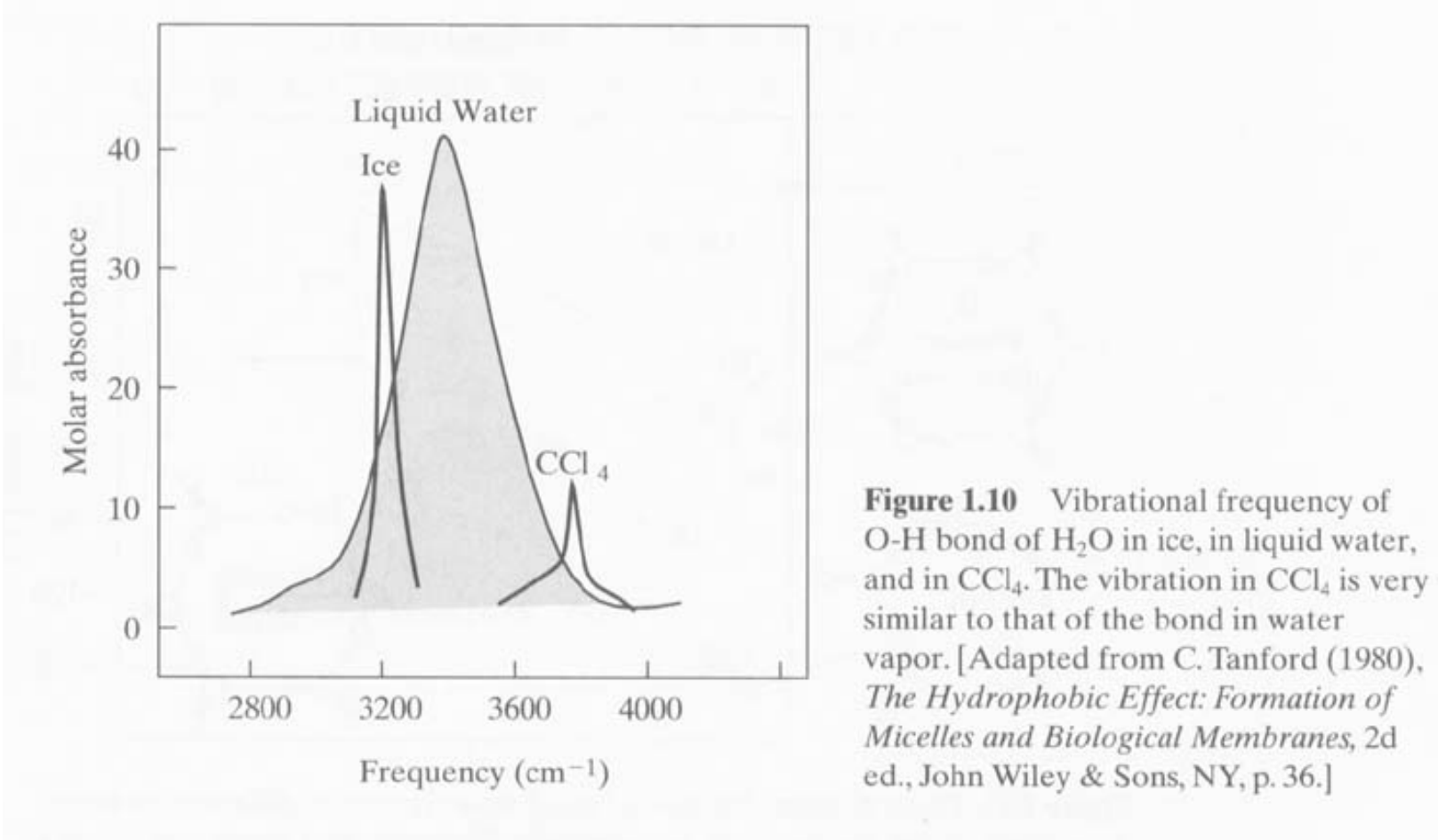


Figure 1.10 Vibrational frequency of O-H bond of H_2O in ice, in liquid water, and in CCl_4. The vibration in CCl_4 is very similar to that of the bond in water vapor. [Adapted from C. Tanford (1980), *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2d ed., John Wiley & Sons, NY, p. 36.]

1.3.2 The interaction of Molecules with water

Water

- Polarizability

- Affect the interaction between charged, polar but uncharged and uncharged and nonpolar groups

- Solvent form an envelop

- Form a Cage-like clathrate structure around ion, ex (fig1.9)/ Overcome low entropy

1.3.2 The interaction of Molecules with water

Hydrophilic compounds/water-loving (ex. NaCl)
The strong interaction between the charged ions and the polar water molecules in highly favorable, even unfavorable entropy.
Ice IX-rigid ice-like cage
The waters around hydrophilic atoms typically form arrays of 6 or 7 water molecules.

Hydrocarbon /Hydrophobic or water heating **(**ex. Methane**)**
Highly soluble in organic solvent

Like dissolve like
Polar charged compounds – soluble in polar solvent, water
Nonpolar compounds-soluble in nonpolar organic solvent, chloroform

Amphipathic molecules

Amphipathic molecules are both *hydrophilic* and *hydrophobic*
Ex. Phospholipid

a charged phosphoric acid head group/solube in water

a long hydrocarbon tail /soluble in organic solvents

Different parts of amphipathic molecules sequester themselves into different environments.

The types of the structures of phoslipids

Micelles: formed by dilute dispersions

Monolayer: at the air-water interface

Bilayer vesicle: useful in biology as a membrane barrier to distinguish between interior and exterior environment of a cell or organelle.

Structures formed by amphipathic molecules in water

Monolayer: at the air-water interface

Micelles: formed by dilute dispersions
Bilayer vesicle: useful in biology as a membrane barrier to distinguish between interior and exterior environment of a cell or organelle.

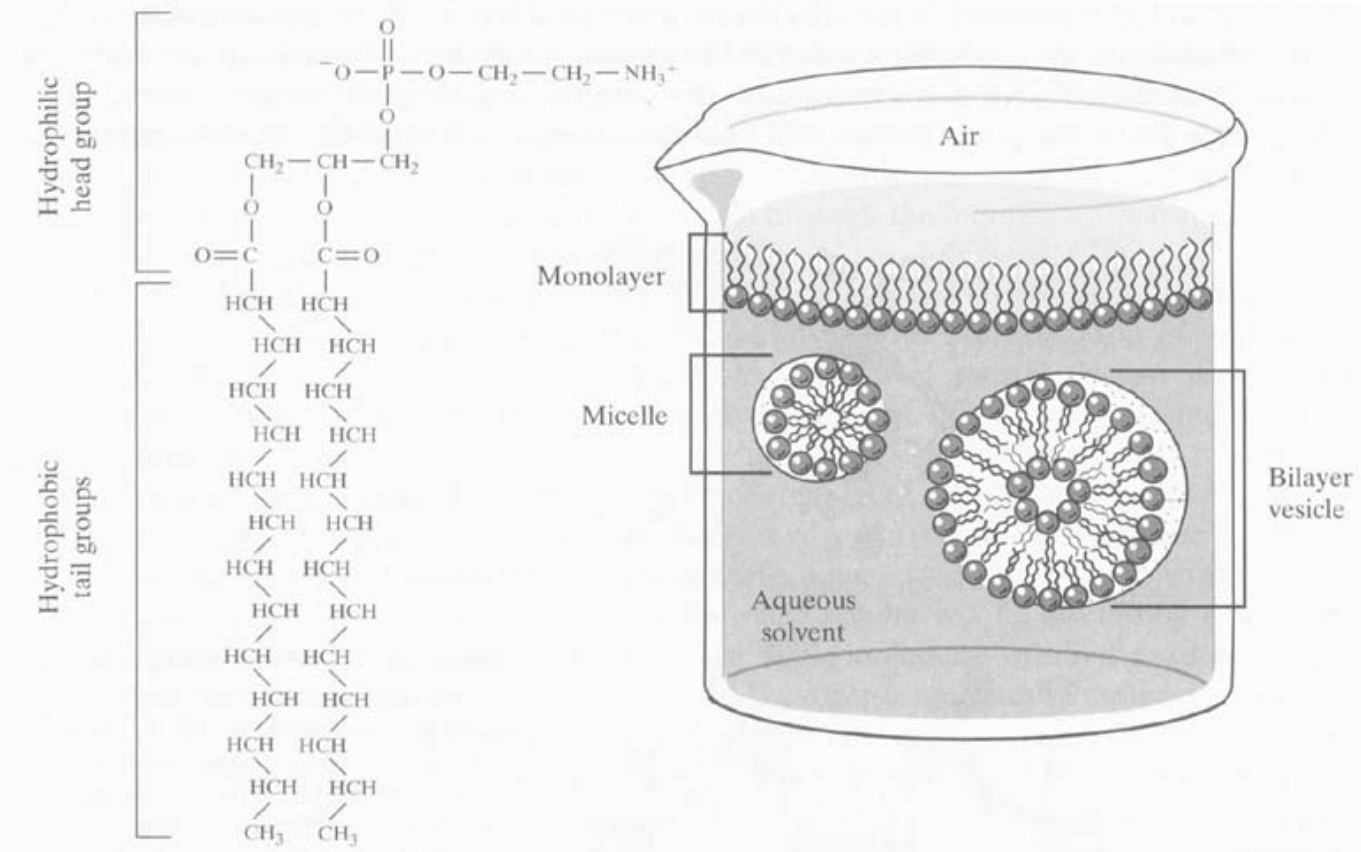


Figure 1.11 Structures formed by amphipathic molecules in water. An amphipathic molecule, such as phosphatidyl choline, has a head group that is hydro-

Protein and nucleic acid are amphipathic molecules

Protein and nucleic acid are amphipathic molecules

Nucleic acids are composed of hydrophobic bases and negatively charged phosphates.

The **Hydrophobic effect** directs the folding of macromolecules and stabilizes the macromolecular structure

1.3.3 Nonaquous Enviroment of Biological Molecules

Significant differences between a cell membrane and the aqueous solution in a cell

Water-soluble protein the hydrophobic group exposed to the solvent and the hydrophilic atoms form the internalized core.

An **integral membrane protein** can be though of as being inverted relative to the structure of a water-soluble protein, with the hydrophobic groups now exposed to the solvent, while the hydrophilic atoms from the internalized core

Ion channel: The polar groups that line the internal surface of the channel **mimic** the polar water solvent, thus allowing charged ions to pass readily through an otherwise impenetrable bilayer.

An **integral membrane protein** can be though of as being inverted relative to the structure of a water-soluble protein, with the hydrophobic groups now exposed to the solvent, while the hydrophilic atoms from the internalized core

Ion channel: The polar groups that line the internal surface of the channel **mimic** the polar water solvent, thus allowing charged ions to pass readily through an otherwise impenetrable bilayer.

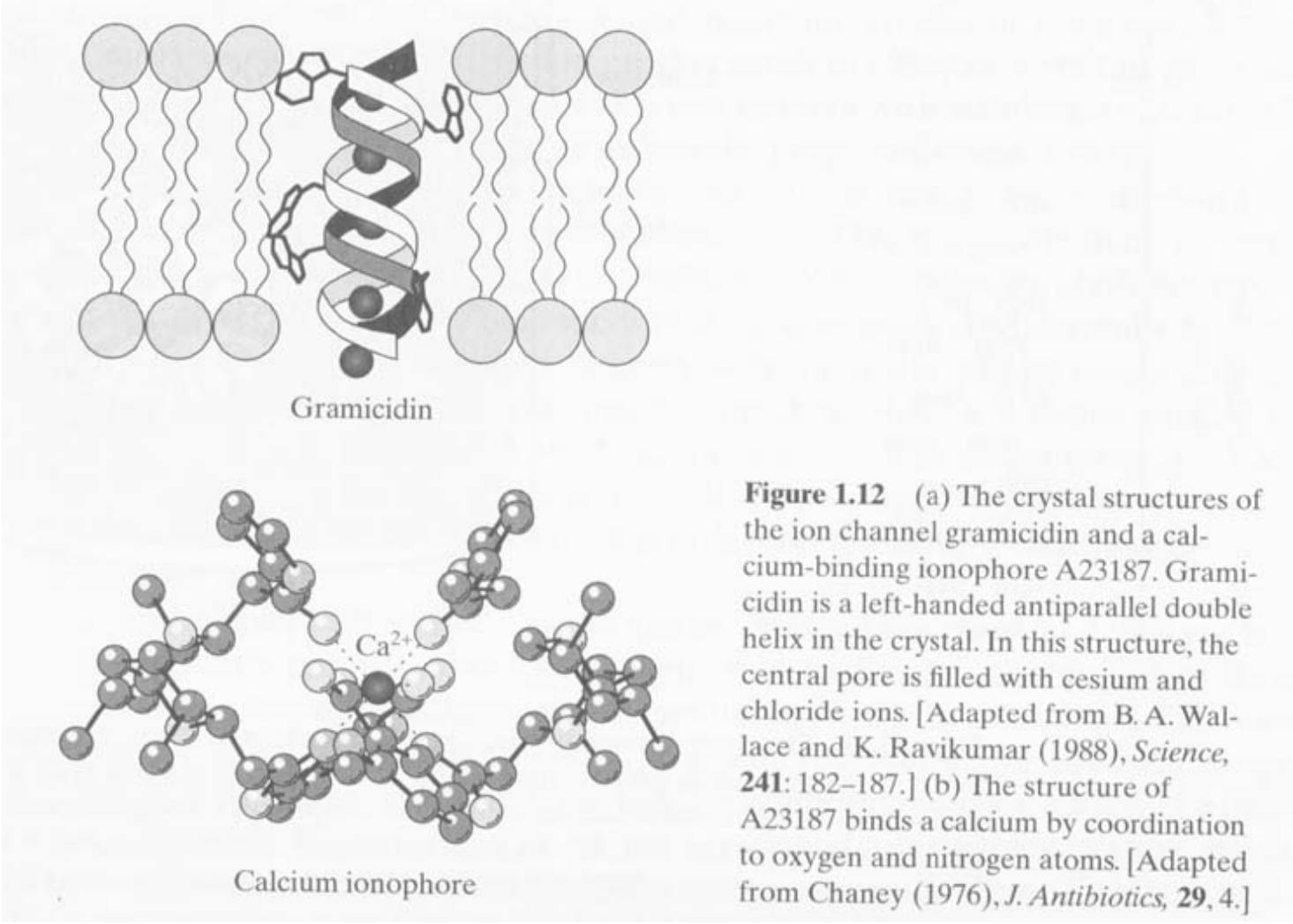


Figure 1.12 (a) The crystal structures of the ion channel gramicidin and a calcium-binding ionophore A23187. Gramicidin is a left-handed antiparallel double helix in the crystal. In this structure, the central pore is filled with cesium and chloride ions. [Adapted from B. A. Wallace and K. Ravikumar (1988), *Science*, **241**: 182–187.] (b) The structure of A23187 binds a calcium by coordination to oxygen and nitrogen atoms. [Adapted from Chaney (1976), *J. Antibiotics*, **29**, 4.]

Self-energy (Es)

$$E_s = q^2 / 2 D r_s$$

The E_s of an ion in water is **40**-times lower that in a lipid bilayer.
Ion in Membrane is 10^{-18} times lower than that in water/ efficient barriers

Ion in H_2O: E_s small, translate faster
Ion in membrane: E_s large, translate slower

Self-energy (Es)

$$\mathbf{E_s = q^2 / 2\ D\ r_s}$$

Ex. Lysine Pka=9.0 for the side chain would be protonated and positively charged in water

If we transfer this charged aa (r=0.6 nm) into a protein interior (D~3.5), the difference in $\mathbf{E_s}$ in the protein versus water is $\Delta E \sim 40$ kJ/mol.

$Es = q^2 / 2\ D\ r_s = (q_{lys})^2 / 2\ k\varepsilon o \times 3.5 \times (0.6\ nm) = 40 KJ/mol$

$$\mathbf{\Delta\ Pk_a = \Delta Es / 2.303\ k_B\ T}$$

Pka<1, for a lysine buried in the hydrophobic core of a globular protein and therefore would be **uncharged** unless it is paired with a **counterion** such as aspartic acid residue.

1.4 Symmetry relationships between molecules

Biological systems tend to be symmetry
Building blocks (aa) are always asymmetry

Symmetry/in composition, shape and relative position of parts that are on opposite sides of a dividing line or median plane of that are distributed about a center or axis

Mirror: relates 2 motifs on opposite sides of a dividing line or plane
Rotation: relates motifs distributed about a point or axis
Screw symmetry: Rotation + Translation

Symmetry element, symmetry operator **O**, ***O (m)= m'***

Point symmetry/point group: a point, line or plane passes through the center of the mass of the motifs.

Symmetry/in composition, shape and relative position of parts that are on opposite sides of a dividing line or median plane of that are distributed about a center or axis

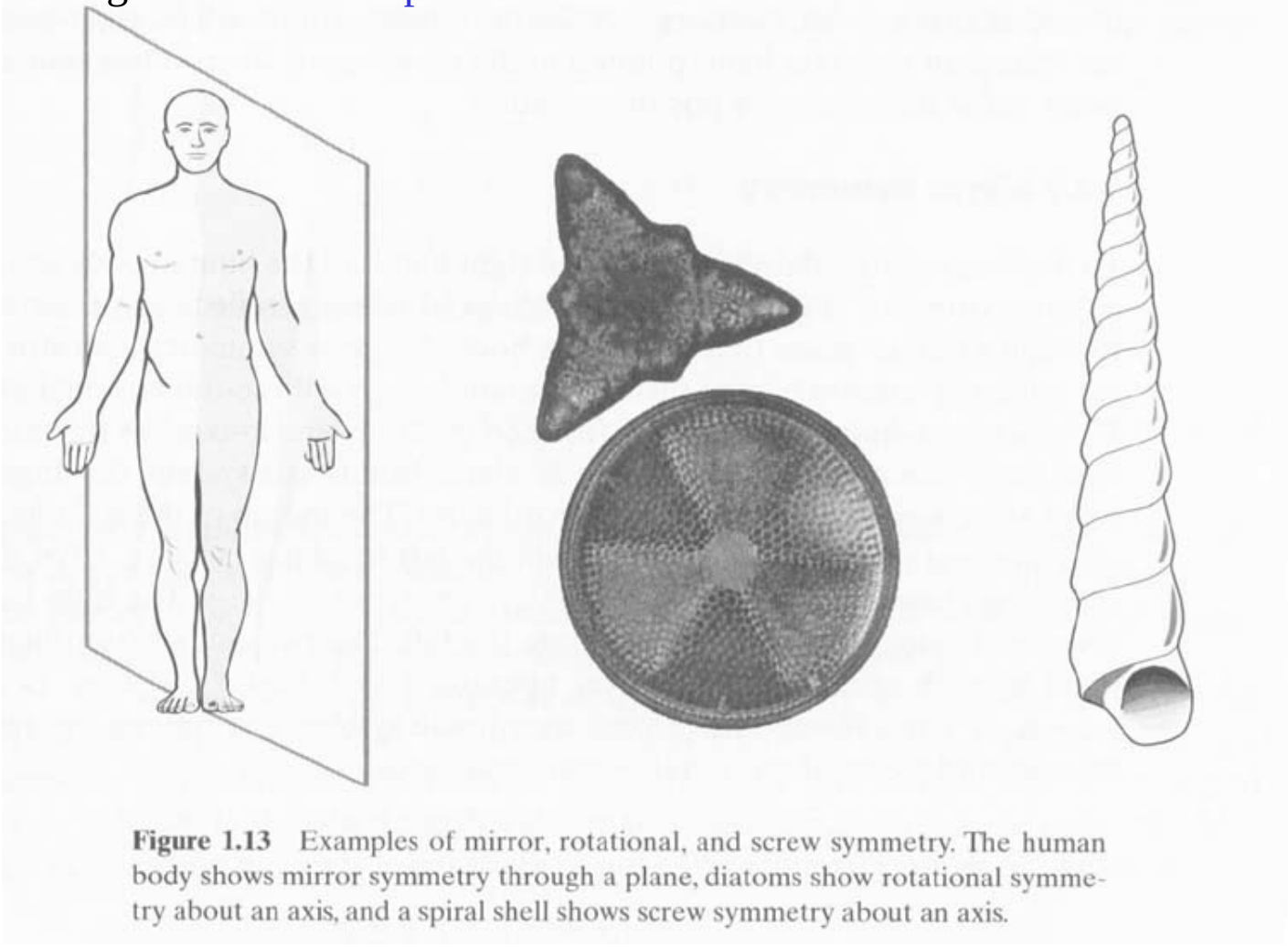


Figure 1.13 Examples of mirror, rotational, and screw symmetry. The human body shows mirror symmetry through a plane, diatoms show rotational symmetry about an axis, and a spiral shell shows screw symmetry about an axis.

Right-handed Cartesian coordinates

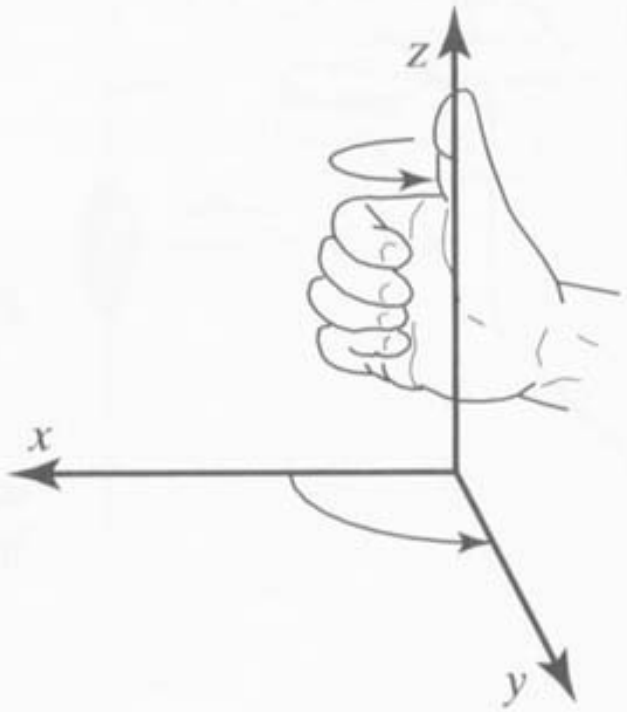


Figure 1.14 Right-handed Cartesian coordinates and right-handed rotations. In a right-handed axis system, the fingers point from the x-axis towards the y-axis when the thumb is aligned along the z-axis. This same system describes a right-handed rotation, where the fingers of the right hand represent a positive rotation about a particular axis, in this case the z-axis.

1.4.1 Mirror Symmetry

For 3D coordinate system, a symmetry operator can be represented by

$$a_1x + b_1y + c_2 z = x'$$
$$a_2 x + b_2y + c_2 z = y'$$
$$a_3 x + b_3y + c_3 z = z'$$

Matrix form (x,y,z) to (x', y',z')

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \times \begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} x' \\ y' \\ z' \end{vmatrix}$$

Plane mirror symmetry / reflection plane

xz plane mirror symmetry: (x, y, z) $\Rightarrow$ (x, -y, z)

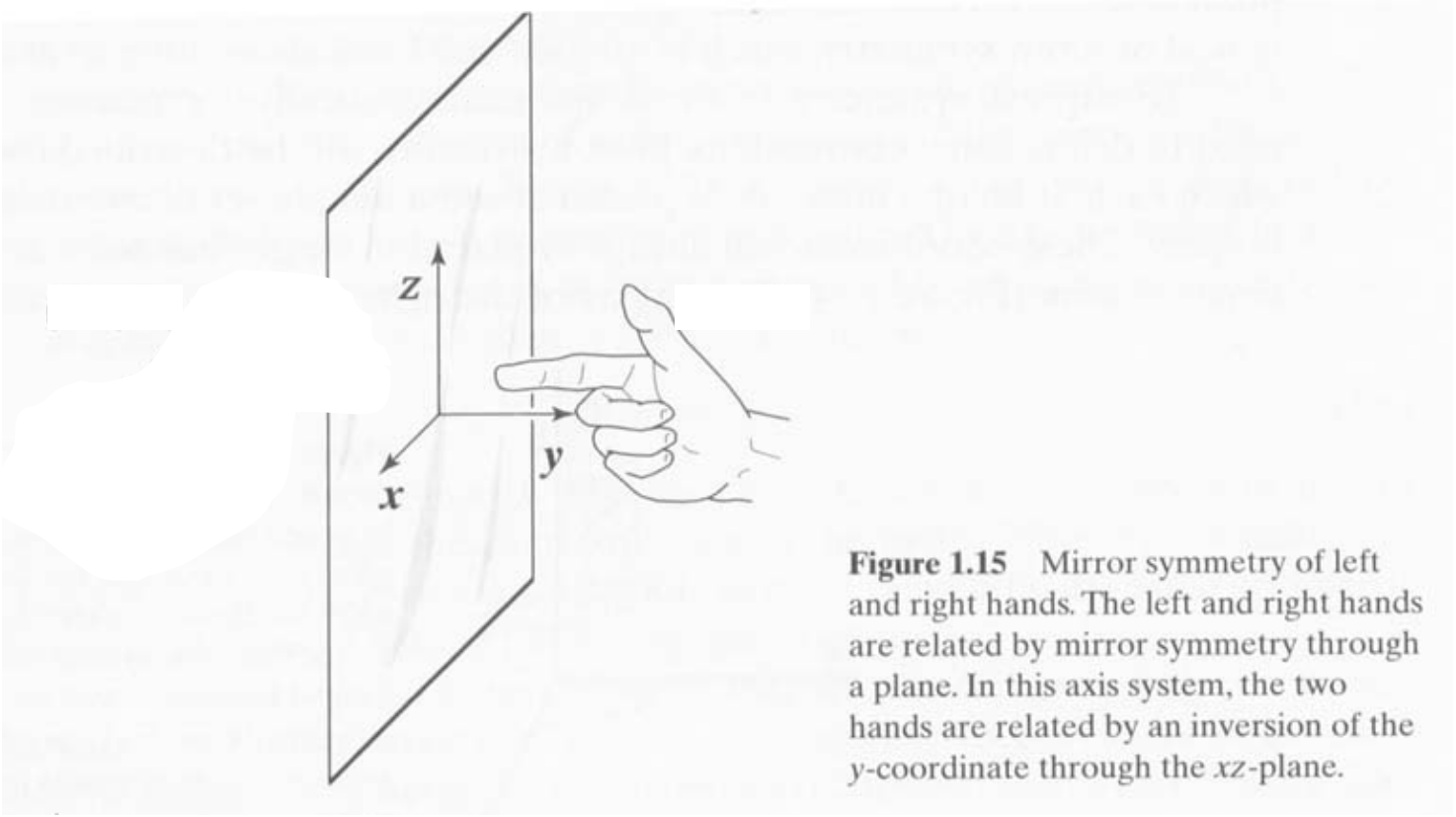


Figure 1.15 Mirror symmetry of left and right hands. The left and right hands are related by mirror symmetry through a plane. In this axis system, the two hands are related by an inversion of the *y*-coordinate through the *xz*-plane.

1.4.2 Rotational Symmetry

Two-fold rotational axis/ Two-fold symmetry/ dyad symmetry

2-fold, z-axis is perpendicular to the page of the figure (x,y,z) $\Rightarrow$ (-x, -y z)

For the rotational angle θ, the symmetry is said to be related by n-fold rotation

C_n - symmetry n=360°/θ

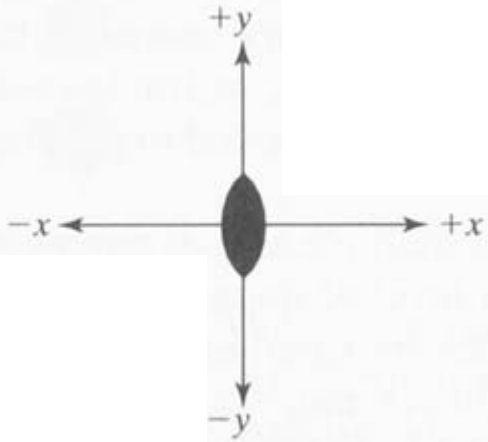


Figure 1.16 Rotational symmetry. The two hands in this figure are related by two-fold rotational symmetry about the *z*-axis. In this example, both the *x*- and *y*-coordinates are inverted.

1.4.2 Rotational Symmetry

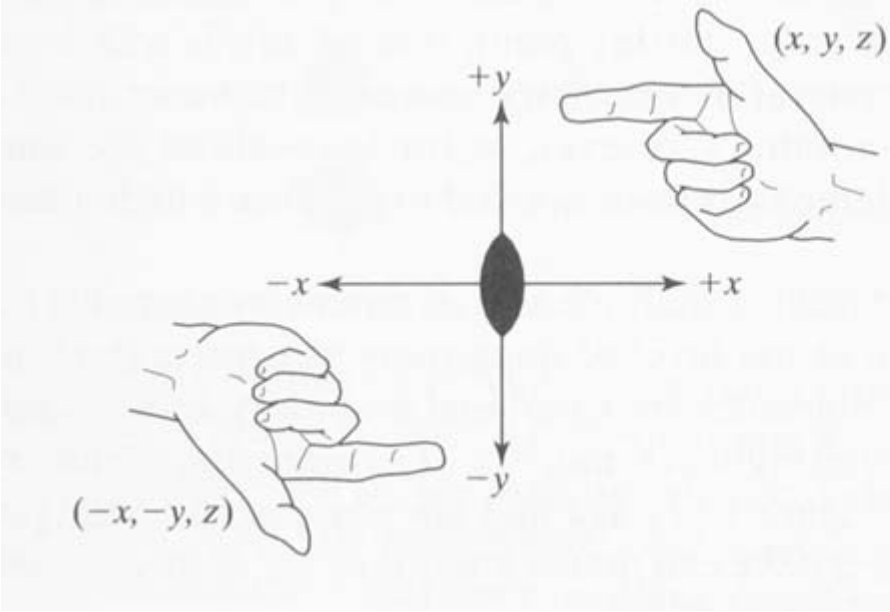


Figure 1.16 Rotational symmetry. The two hands in this figure are related by two-fold rotational symmetry about the z-axis. In this example, both the x- and y-coordinates are inverted.

C_n symmetry, $n = 360^o/\theta$

For rotation about the z-axis by any angle θ , the general operator in matrix form is

$$\begin{vmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

TABLE 1.4 SYMBOLS FOR SYMMETRY

Symbol	Symmetry	Motif
	C_2 (two-fold)	Monomer
	2_1 (two-fold screw)	Monomer
	C_3 (three-fold)	Monomer
	3_1 (right-handed three-fold screw)	Monomer
	3_2 (left-handed three-fold screw)	Monomer
	C_4 (four-fold)	Monomer
	4_1 (right-handed four-fold screw)	Monomer
	4_2 (four-fold screw)	Dimer
	4_3 (left-handed four-fold screw)	Monomer
	C_6 (six-fold)	Monomer
	6_1 (right-handed six-fold screw)	Monomer
	6_2 (right-handed six-fold screw)	Dimer
	6_3 (six-fold screw)	Trimer
	6_4 (left-handed six-fold screw)	Dimer
	6_5 (left-handed six-fold screw)	Monomer

Pseudo symmetry: Motifs appear symmetric, but not truly symmetric

1.4.3 Multiple Symmetry Relationships and Point Groups

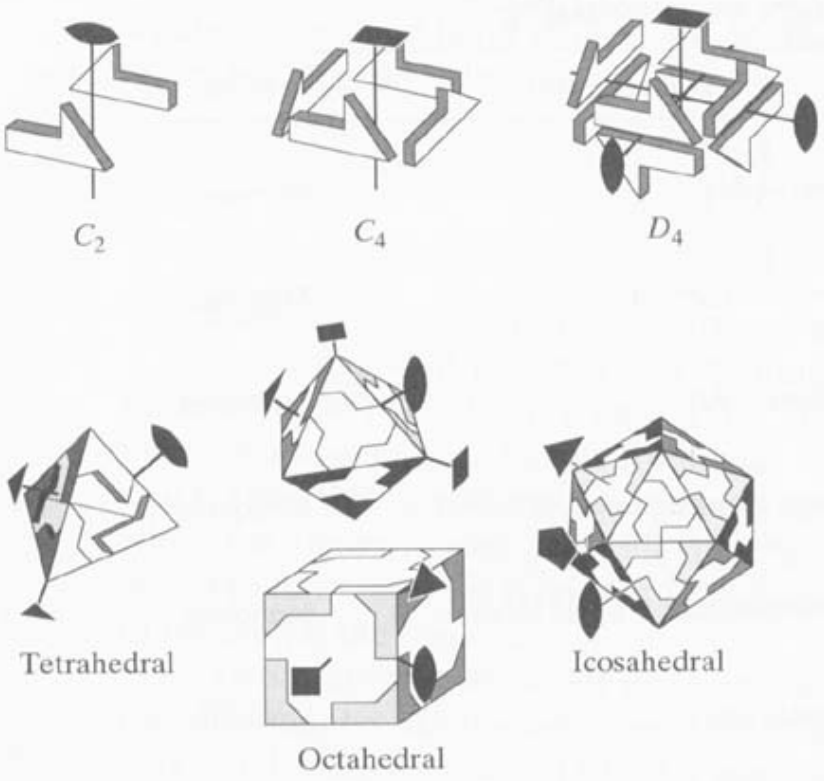


Figure 1.17 Point groups. The repeating motif in each figure is represented by an arrow. The C_2 point group shows two-fold rotational symmetry, C_4 shows four-fold symmetry, and D_4 shows both four-fold and two-fold symmetry. In the D_4 point group, two of the four two-fold axes exactly overlap, leaving only two unique two-fold axes perpendicular to the four-fold axis. Tetrahedral symmetry has two-fold axes at the edges and three-fold axes at the corners and faces of the four sides. Octahedral symmetry has two-fold, three-fold, and four-fold symmetry. This defines both a standard octahedron and a cube. Icosahedral symmetry is defined by two-fold, three-fold, and five-fold symmetry axes.

Point Group

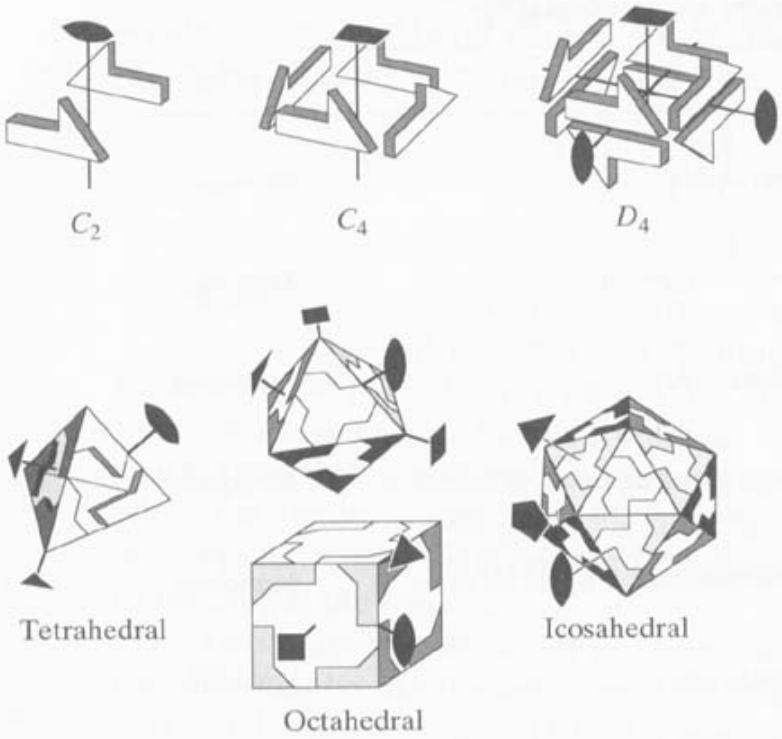


Figure 1.17 Point groups. The repeating motif in each figure is represented by an arrow. The C_2 point group shows two-fold rotational symmetry, C_4 shows four-fold symmetry, and D_4 shows both four-fold and two-fold symmetry. In the D_4 point group, two of the four two-fold axes exactly overlap, leaving only two unique two-fold axes perpendicular to the four-fold axis. Tetrahedral symmetry has two-fold axes at the edges and three-fold axes at the corners and faces of the four sides. Octahedral symmetry has two-fold, three-fold, and four-fold symmetry. This defines both a standard octahedron and a cube. Icosahedral symmetry is defined by two-fold, three-fold, and five-fold symmetry axes.

N (total repeating motifs) = 3m

C_n, n-fold rotational axis is the C_n-axis, value n also refers to the # of motifs that are related by the C_n axis.

C_1: a single motif that has no rotational relationships. This can only be found in asymmetric molecules that have no plan or center of symmetry (ex. Chiral molecules)

m: m-hedral symmetry, there are m faces on the solid shape, the total # of C_3

Symmetry in each point group

N: the # of repeating motifs, N=3m **N= m x n,** m number of C_n symmetry axes

Ex: icosahedral, N=3x20=60, 60 repeating motifs

12 of C_5 symmetry axis (12 x 5 = 60),

30 of C_2 symmetry axis

no true C_6 axis in icosahedral

	Tetrahedral	**Octahedral**	**Icsohedral**
m	**4**	**8**	**20**
N (Total motifs)	**12**	**24**	**60**
#2-fold	**6**	**12**	**30**
#3-fold	**4**	**8**	**20**
#4-fold	**3**	**6**	**15**

1.4.4 Screw Symmetry

Translation: simply moves a motif from one point to another, without changing its orientation.

T: translation operator

$$(x, y, z) + T = (x+T_x, y+T_y, z+T_z)$$

Screw symmetry/Helical symmetry: translation + rotation

$$C_n (x,y,z) + T = (x', y', z'),$$

the translation resulting form a 360° rotation of the screw is its pitch.

right-handed (clockwise) & left-handed (counterclockwise)

A helix n-fold screw symmetry, n times to give one complete turn of the helix.

1.4.4 Screw symmetry

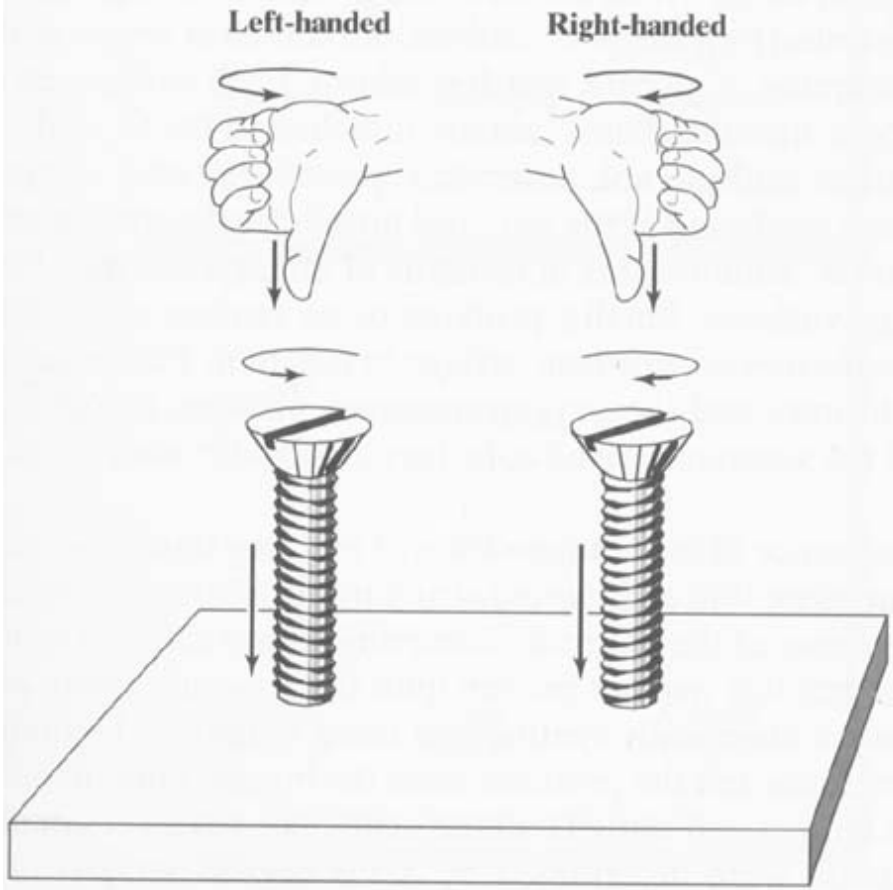


Figure 1.18 Screw symmetry. The rotation of a screw becomes translational motion by screw-symmetry. Screws can be either left-handed or right-handed, depending on which direction the screw must be turned to drive it into the board.

1.5.1 Amino Acids

Proteins are polymers built from amino acid.

20 amino acids, α amino acid with amino and carboxylic acid groups separate by a single Cα carbon.

L-amino acids: natural aa

D-amino acid: unnatural aa, are found in the antiviral protein valinomycin and gramicidin, produced by bacteria.

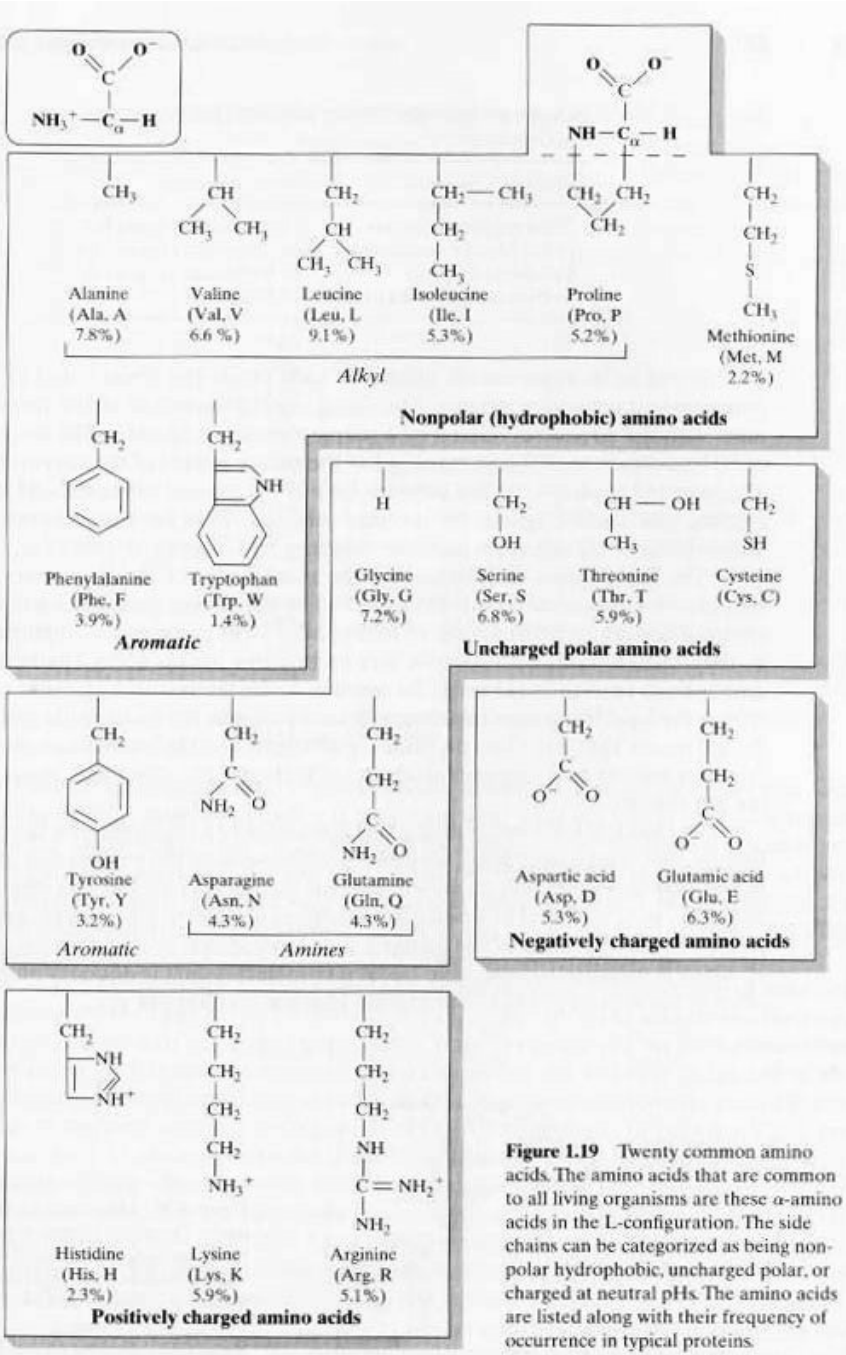


Figure 1.19 Twenty common amino acids. The amino acids that are common to all living organisms are these α-amino acids in the L-configuration. The side chains can be categorized as being non-polar hydrophobic, uncharged polar, or charged at neutral pHs. The amino acids are listed along with their frequency of occurrence in typical proteins.

TABLE 1.5 ENZYME COFACTORS AND THEIR DIETARY PRECURSORS

Coenzyme	Precursor
Thiamine pyrophosphate	Thiamine (vitamin B_1)
Flavin adenine dinucleotide	Riboflavin (vitamin B_2)
Pyridoxal phosphate	Pyridoxine (vitamin B_6)
5′-Deoxyadenosylcobalamine	Vitamin B_{12}

2 examples: rubredoxin and HIV protease from synthesized D-aa and determined by X-ray (fig1.20).

Rubredoxin crystal structure

Racemic mixture/equal proportions of the 2 steroisomers
related by a mirror symmetry operate

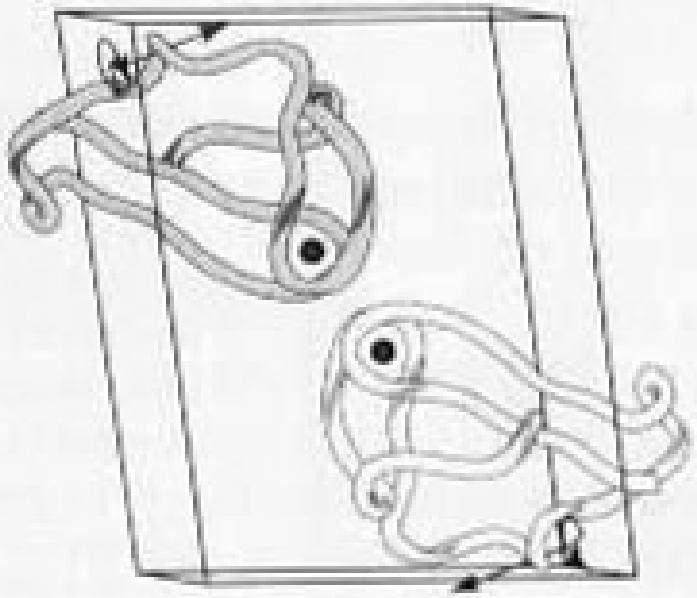


Figure 1.20 Mirror images of rubredoxin. The native protein (shaded) and its mirror image were solved from the same crystal. The two proteins are related by mirror symmetry and rotational symmetry. [Adapted from Zawadzke and Berg (1993), *Proteins*, **16**: 301.]

HIV protease

Inverted protein did not cat. the nature substrate, lysine
Native enzyme was inactive against the inverted substrate.

Hydropathy partition (P) = X_{aq}/ X_{nonaq}

X_{aq} : the mole fraction in aqueous X_{aq}
X_{nonaq} : the mole fraction in organic phase.

+: nonpolar SC, -: polar and charged SC

TABLE 1.6 HYDROPATHY INDEX OF AMINO ACIDS

Amino Acid	Hydropathy
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	−0.4
Thr	−0.7
Ser	−0.8
Trp	−0.9
Tyr	−1.3
Pro	−1.6
His	−3.2
Asn	−3.5
Gln	−3.5
Asp	−3.5
Glu	−3.5
Lys	−3.9
Arg	−4.5

Source: From J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157**: 105–132 (1982).

Hydrophobic aromatic SC

The interaction tend to place of the **aromatic rings perpendicular** to each other when buried in the interior of a globular protein, or within the hydrophobic region of the membrane bilayer in membrane protein.

Similar to the perpendicular arrangement of benzene molecules in solution.

Faces stacked parallel is **entropically favored**

The structures of nucleotide bases in DNA and RNA form **parallel stacks** to **reduce** their exposure to the solvent.

Do not need to minimize their exposed surface to water.

Isoelectric point (pI): total charge is zero

Charged density of a protein (ρ_c)

Estimated as the ration of the effective charge of a protein at any pH relative to its molecular weight

The charge density of a protein ρ_c **= (pI-pH)/MW**

1.5.2 The Unique Protein Sequence

The sequence of a protein is the covalent linkage of aa by peptide bonds

2 freely rotating bonds for each aa

" ϕ " torsion angle: the rotation **N-Cα** bond

" φ " torsion angle: the rotation **Cα-C** bond

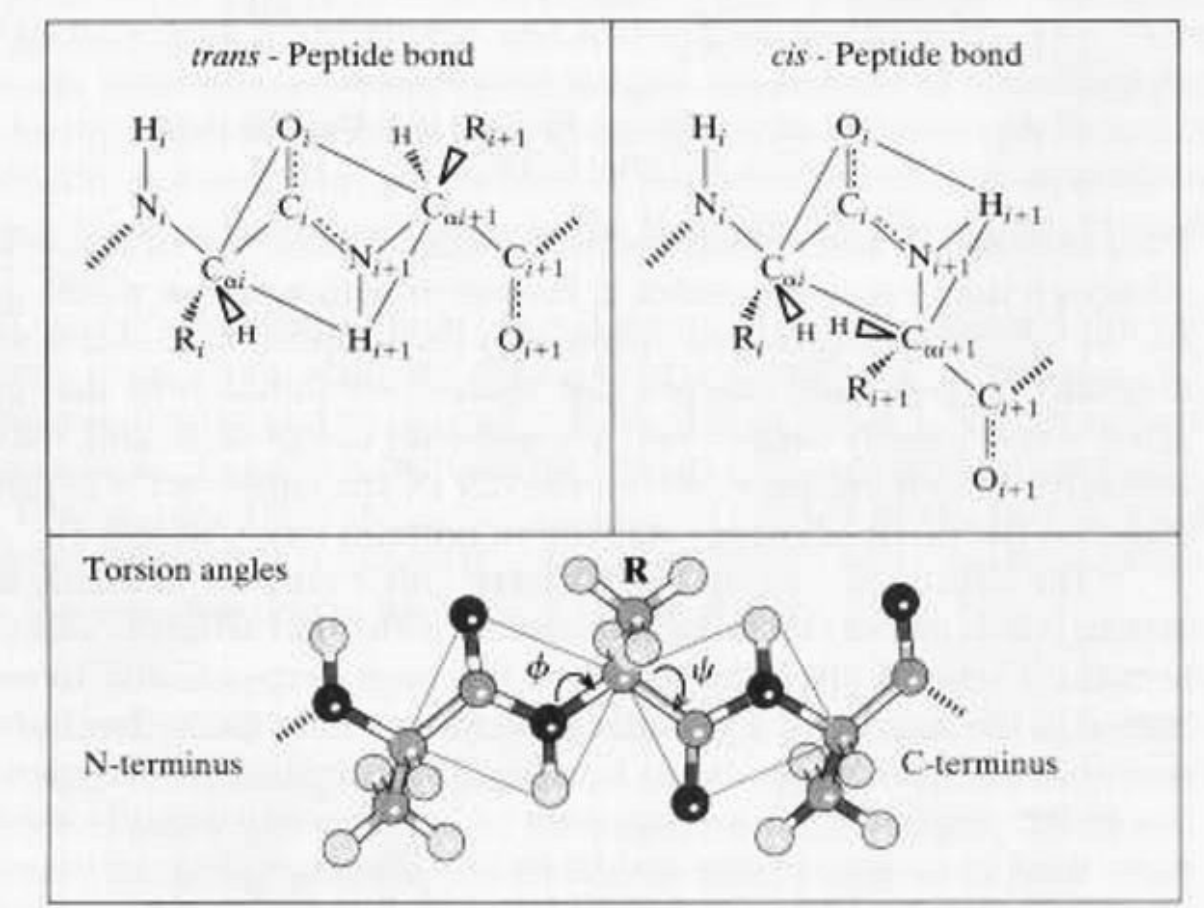


Figure 1.21 The peptide bond. The peptide bond that links two amino acid residues along a polypeptide chain is an C-N bond of an amide linkage. The

For example, tripeptide,

20x20x20=8000 different possible sequences

a small protein size N≈100,

$20^{100} \approx \mathbf{10^{130}}$ different possible aa sequences

More than the # of particles currently thought to be **in the universe**

Equal probability of p;acing each of the 20 aa at any particular position along the peptide chain

This assumption **may not be correct**

the simple analysis could lead to **incorrect conclusions**

The most accurate estimate for the **frequency** of the occurrence of a polypeptide sequence must take into account the statistical probability for the occurrence

Basic statistical approach

Application 1.1 ” Musical Sequences”

“**ELVIS**” occurred 4 times in 25,814 protein sequence

It is significantly higher that we would expect from the random occurrence of any 5 aa (once in roughly 3 million random aa)

“**HAYDN**” did not occur in the same data base, sequence bias
“**LIVES**” was not represented in the data set.

How many different sequences have identical composition of residues?

"G2A" composition

3 different tripeptide sequences

How many different ways to arrange 2 particles in 3 boxes

GGA/GAG/AGG

g=3 & n=2

$W = \Pi [g!/n!\,(g-n)!] = 3!/2!\,1! = 3$

The general form for determining the # ways to arrange a set of particles into identical (degenerate) positions.

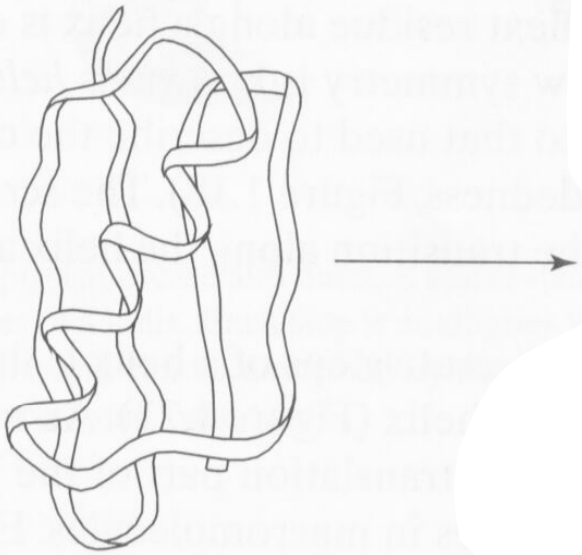
- Sequence directs structure, 2 molecules with homologous sequence can be assumed to have similar structures.

- unique structure requires a unique sequence
- 25-30% homology

- Molecules with similar structure need not be homologous sequence
- Example: nucleoportein H5 and HNF-3/fork-head protein
- 9 identical aa of 72 aa
- 12.5 % only

Application 1.2 ” Engineering A New fold” Protein G & Rop

Without altering more than 50% of the sequence, design a protein with a completely different fold



protein G

Figure A1.2 Replacement of 50% of the amino acids changes the tertiary structure of the mostly β-sheet protein G to a helical bundle similar to the RNA-binding protein Rop. The shaded regions are the amino acids in protein G that were replaced by helix forming resideus from Rop.

1.5.3 Secondary Structure of Protein

The regular and repeating structure of a polypeptide is its secondary structure (2^{o}).

"**Regular**" defines these are **symmetric** structure.

3_{10} helix,
10 atoms separating the amino hydrogen and carboxyl oxygen atoms that are hydrogen-bonded together to form three complete turns of the helix.

β-sheet
Hydrogen bonds are the significant interaction in secondary structure. hydrogen bond between residue i and i+4.
Hydrogen bonds between strands

1.5.4 Helical Symmetry

Helix: a structure in which residues rotate and rise in a repeating manner along an axis.

Generate by the rotation operate “ C “ and translation operate “ T ”

Screw symmetry / Helical symmetry

$$(x,y,z)_i + T = (x,y,z)_{i+1}$$

Helix axis

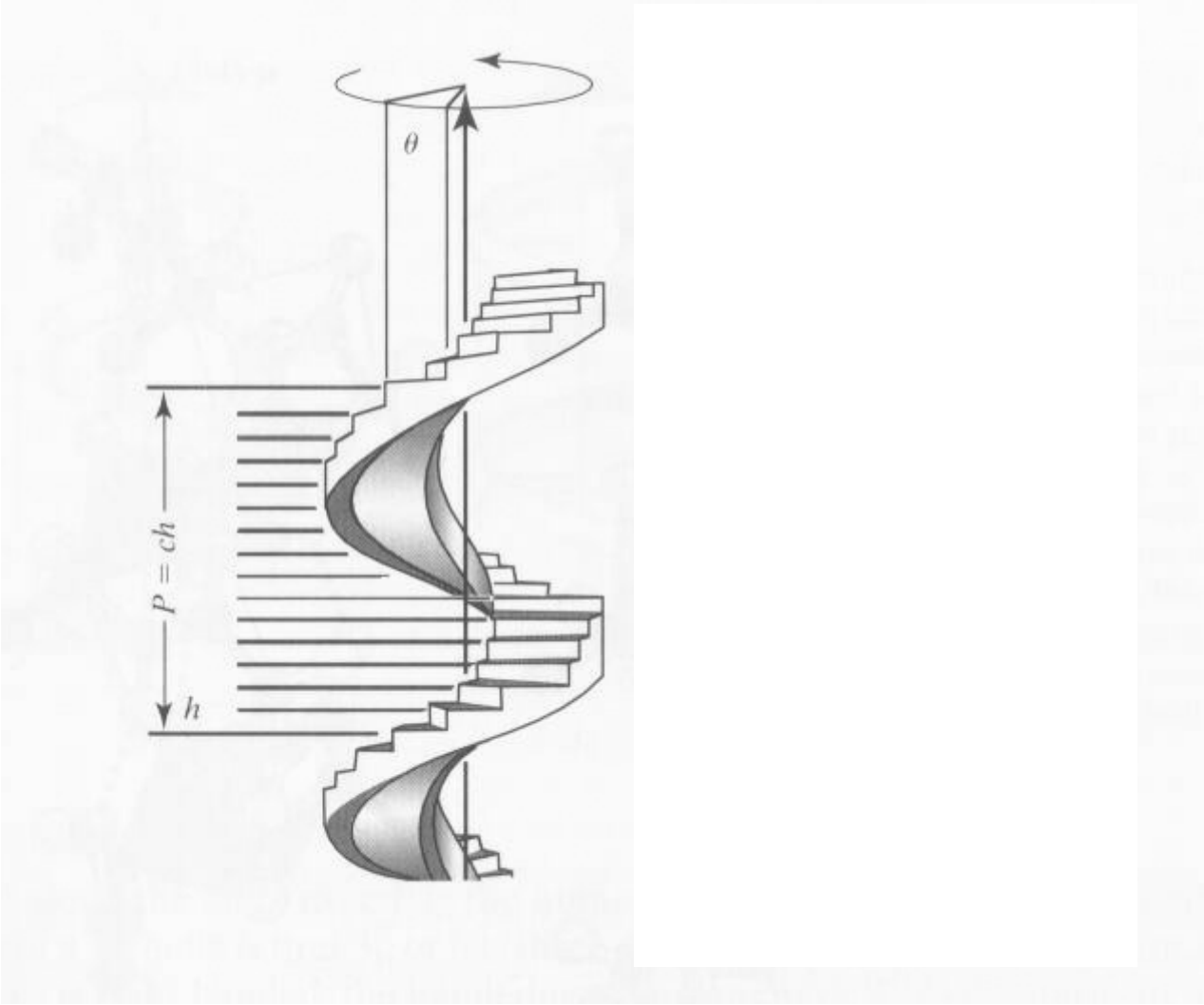
Pitch (P): the translation along the helix axis in one complete turn of the helix.

$$P = c\,h$$

Helical rise (h): the steps rise by a specific vertical distance h and rotate by "θ"

repeat (c) : the # of steps repeat required to reach "P"

helical angle/helical twist (θ): $\theta = 2\pi / c$



Handedness

Positively rotation of the helical angle $\theta > 0°$ right handed helix

Negative rotation of the helical angle $\theta < 0°$ left handed helix

Mathematically generate the 3-D coordinates of a helical structure

"N_T ": N represents the N-fold rotation operator and T the translation in fractions of a repeat for the symmetry operator

Helix

"**3_{10}**" : 3 residues per turn, 1/3 of this repeat h= (1/3)P along the axis, the helical symmetry is **3_1,** three fold screw symmetry.

α-helix : c=3.6 per turn,
360 °/3.6 = 100 °/residue,
h=0.15nm/ residue,
P=3.6x1.5=0.54nm
helical symmetry "**3.6_1**"= "**18_5**": 18 residues in 5 full turns of the helix.

β-sheet
Trans conformation:
two-fold screw symmetry,
a fully extend straight backbone
β-sheet: twist two-fold screw symmetry, like the folds of a curtain
antiparallel & parallel

α-helix

c=3.6 per turn,
360 °/3.6 = 100 °/residue,
h=0.15nm/ residue
P=3.6x1.5=0.54nm

helical symmetry
"**3.6_1**"= "**18_5**": 18 residues in 5 full turns of the helix

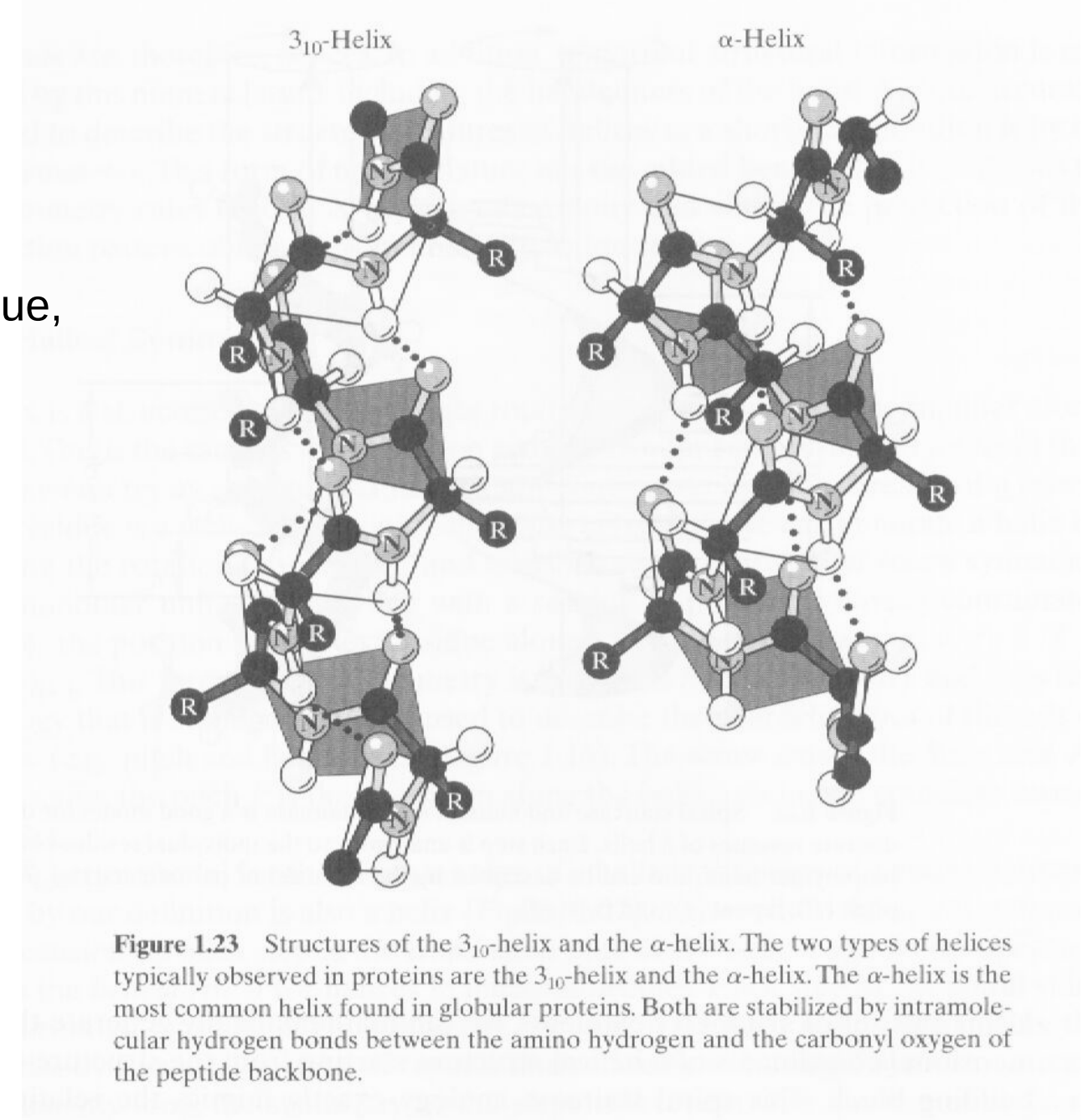


Figure 1.23 Structures of the 3_{10}-helix and the α-helix. The two types of helices typically observed in proteins are the 3_{10}-helix and the α-helix. The α-helix is the most common helix found in globular proteins. Both are stabilized by intramolecular hydrogen bonds between the amino hydrogen and the carbonyl oxygen of the peptide backbone.

TABLE 1.7 HELICAL SYMMETRY OF MACROMOLECULAR HELICES

Structure Type	Residues per turn	Rise (nm)	Helical Symmetry
Trans-conformation (polypeptides)	2.0	0.36	2_1
β-sheet	2.0	0.34	2_1
3_{10} helix	3.0	0.20	3_1
α-helix	3.6	0.15	18_5
π-helix	4.4	0.12	22_5
A-DNA	11.0	0.273	11_1
B-DNA	10.0–10.5	0.337	10_1
C-DNA	9.33	0.331	28_3
Z-DNA	−12.0	−0.372	6_5*

*The repeating unit of Z-DNA is two base pairs in a dinucleotide. This gives an average repeat of −12 base pairs per turn of the left-handed double helix.

Right-handed and left-handed helical symmetry

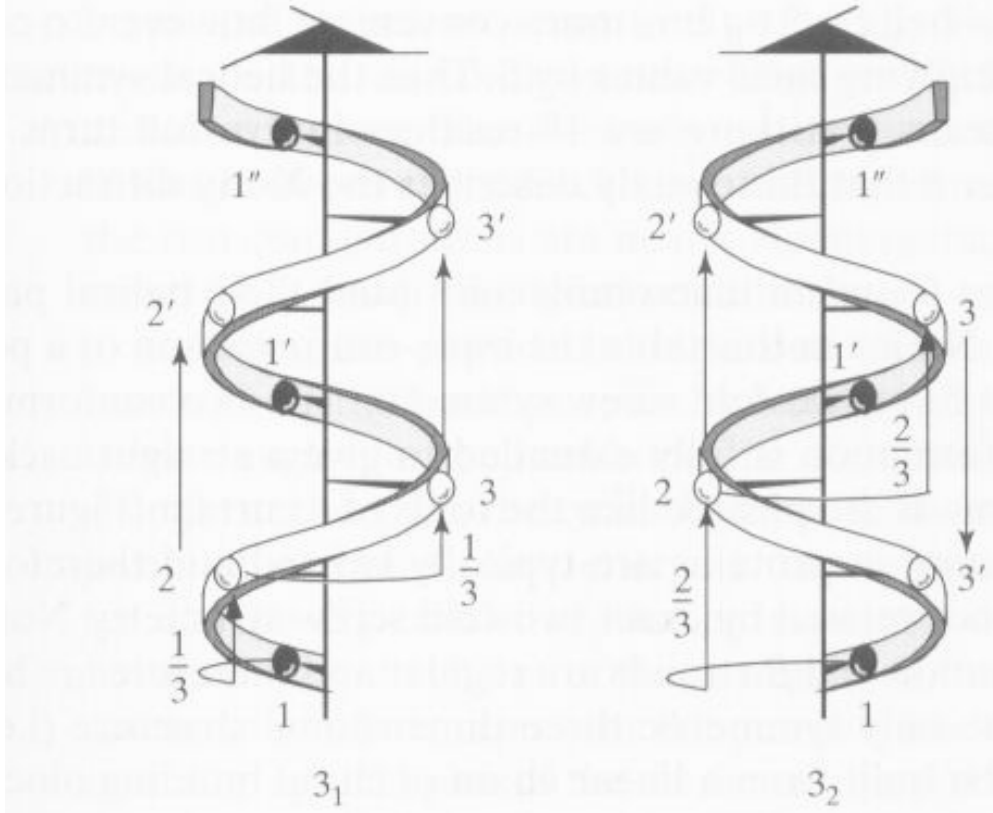


Figure 1.24 Right-handed and left-handed helical symmetry. A helix with 3_1 helical symmetry has each residue rotated +120° and translated 1/3 of the each repeat ($h = P/3$). This generates a right-handed helix. In contrast, 3_2 helical symmetry rotates each residue by the same angle, but translates the residues by 2/3 of a repeat ($h = 2P/3$). When each of the repeating units are filled by the symmetry related residues, the connections from residue 1′ to 2 to 3′ results in a left-handed helix.

1.5.5 Effect of the Peptide Bond on Protein Conformation

Polypeptides: the N-C bond of the peptide linkage is a partial double bond and therefore is not freely rotating.

cis-configuration & *trans*-configuration

***cis*-configuration** is energetically unfavorable, collisions between the SC of adjacent residues.
Proline

***trans*-configuration** is slightly favored (4:1) under biological conditions.

2 torsion angles for the backbone:
"ϕ" angle at the amino nitrogen-Cα bond
"ψ " angle at the Cα -carboxyl bond

Ramachadran plot: the sterically allowed conformations of a polypeptide chain

Ramachandran Plot

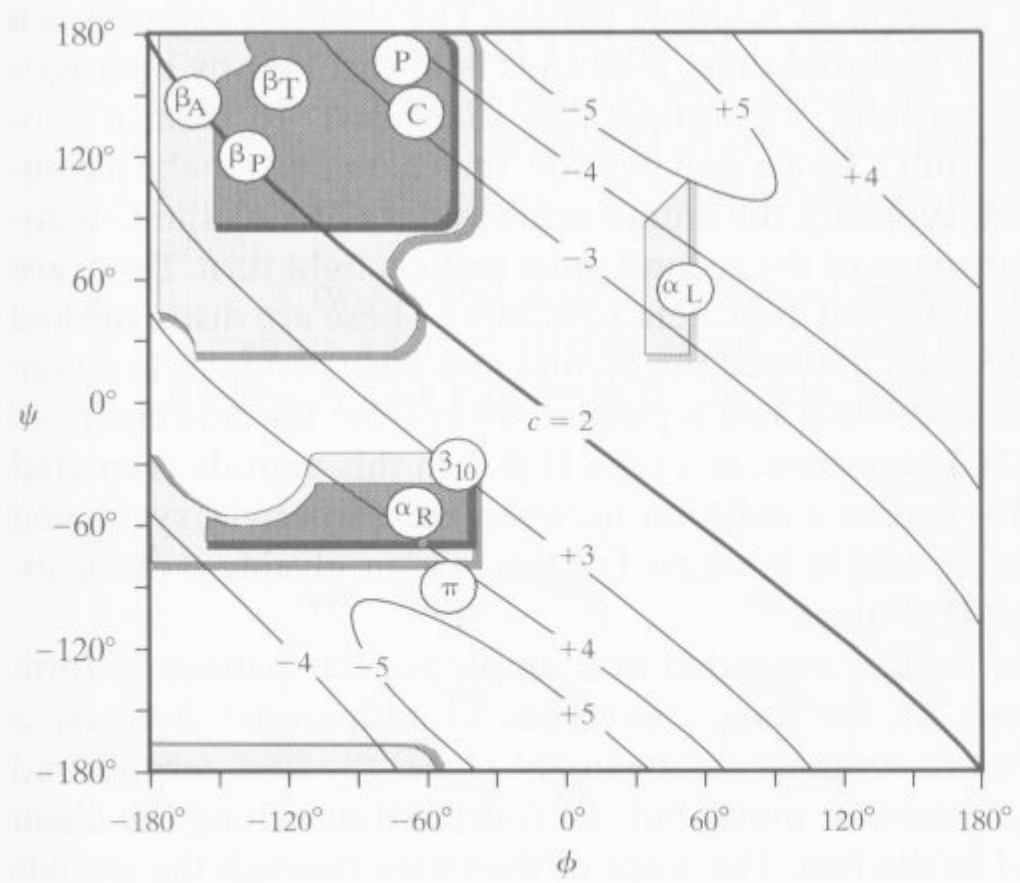


Figure 1.26 Ramachandran plot. The van der Waals interaction energies of an internal Ala residue in a polypeptide chain represents the values of ϕ and ψ-angles that can be adopted by a typical amino acid residue in the chain. A ϕ, ψ-plot of the energies shows conformations that are sterically allowed (dark-shaded regions), moderately allowed (light-shaded regions), and disallowed (open regions). The curves through the plot represent the angles for helices with a particular repeat c. The secondary structures that are found in proteins have torsion angles that lie within or near allowed and moderately allowed regions of the ϕ, ψ-plot. These include the structures of the right-handed α-helix (α_R), left-handed α-helix (α_L), 3_{10}-helix, π-helix, parallel β-sheets (β_P), antiparallel β-sheets (β_A), twisted β-sheets (β_T), polyproline (P), and collagen (C). The helical repeat for protein secondary structures follow the contours through the plot (positive values of c are for right-handed helices and negative values are for left-handed helices).

1.5.6 The structure of Globular Proteins

Protein can often be segregated into **domains** that have distinct structures and functions supersecondary structures.

motifs formed by the association of 2 or more helices are categorized as supersecondary structures because they are regularly occurring patterns of multiple helices.

β-turn

- Type I β-turn
- Type II β-turn

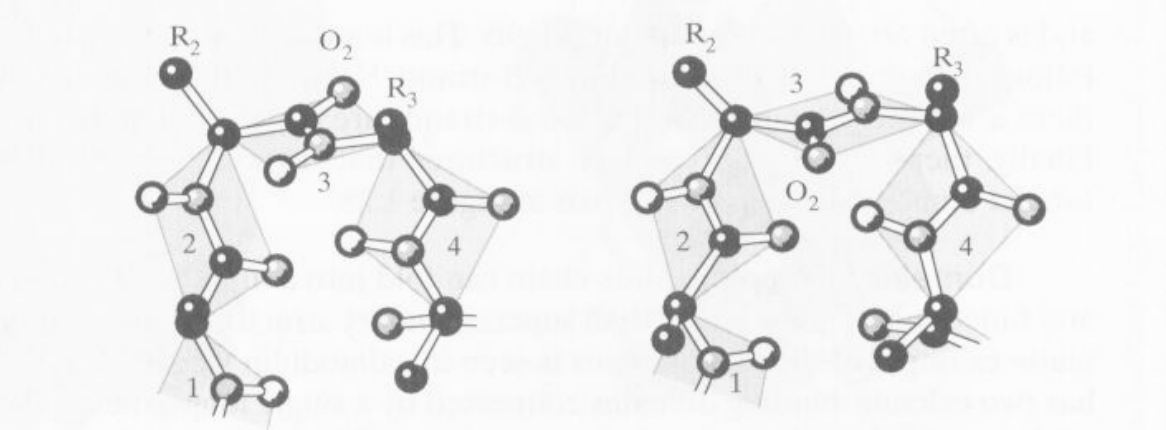


Figure 1.27 Type I and type II β-turns. The tight turns formed by four amino acids residues can have the keto oxygen (O_2) of the second residue in the turn pointing away from the flanking side chains R_2 and R_3 (type I β-turn) or in the same direction as the side chains (type II β-turn).

1.5.6 The structure of Globular Proteins

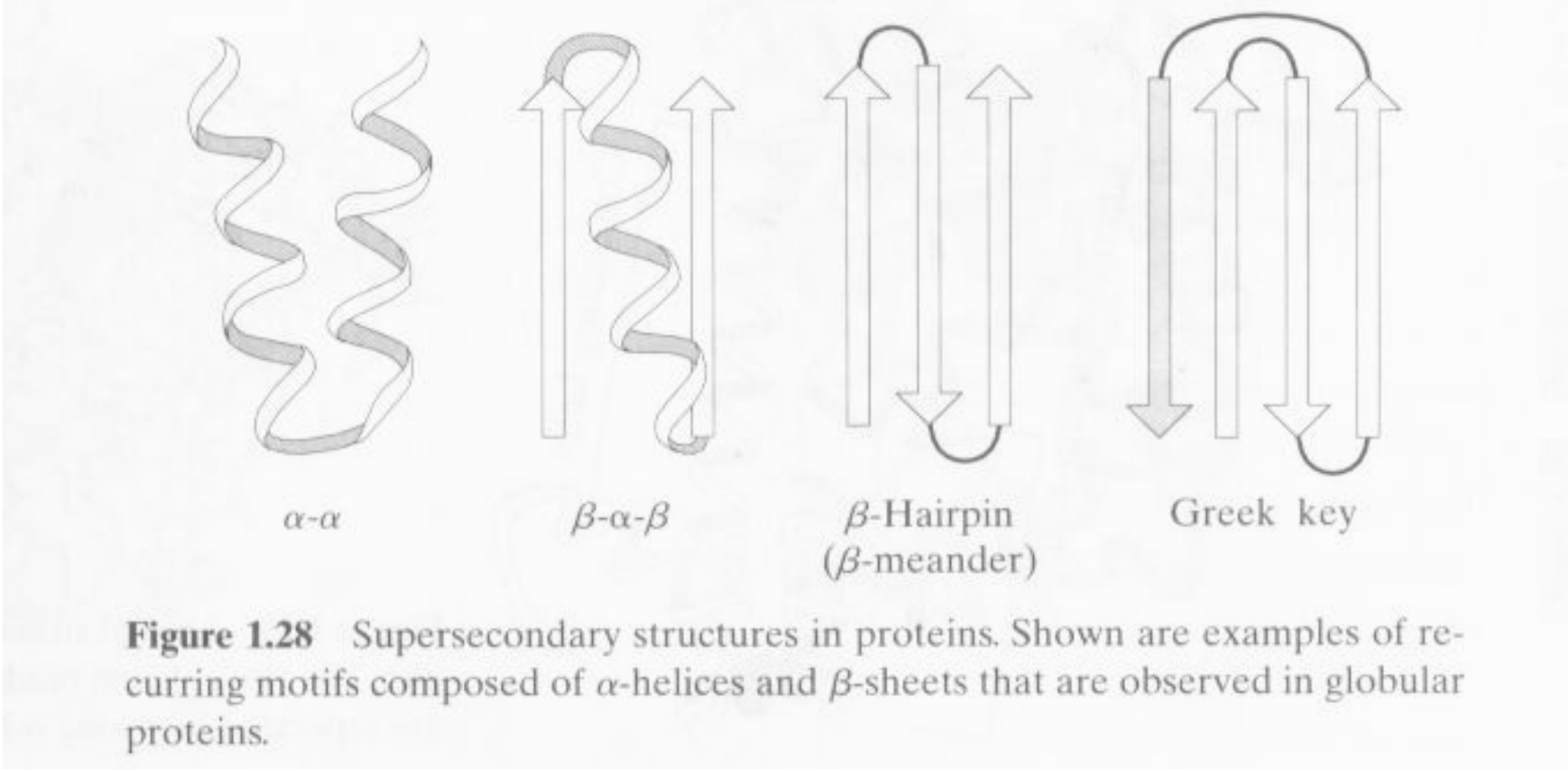


Figure 1.28 Supersecondary structures in proteins. Shown are examples of recurring motifs composed of α-helices and β-sheets that are observed in globular proteins.

Domains: a number of distinct structural and functional regions larger than supersecondary structures.

Ex calmodulin:2 calcium binding domains

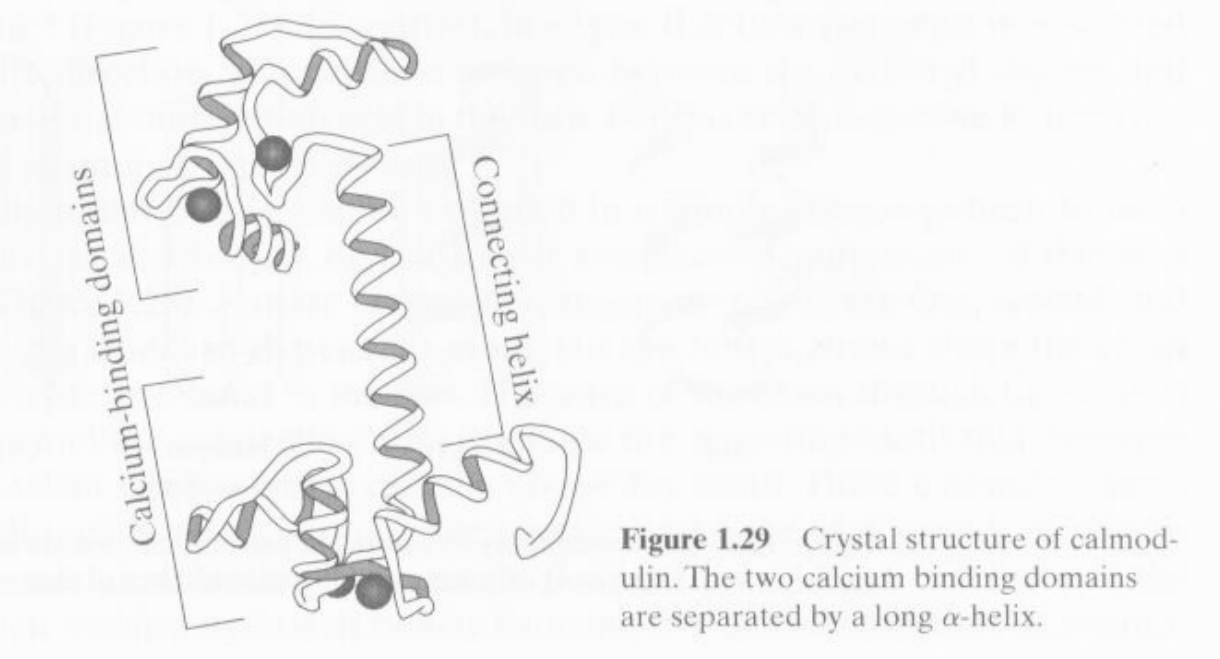


Figure 1.29 Crystal structure of calmodulin. The two calcium binding domains are separated by a long α-helix.

Tertiary Structure

The overall 3-D conformational of the polypeptide chain.

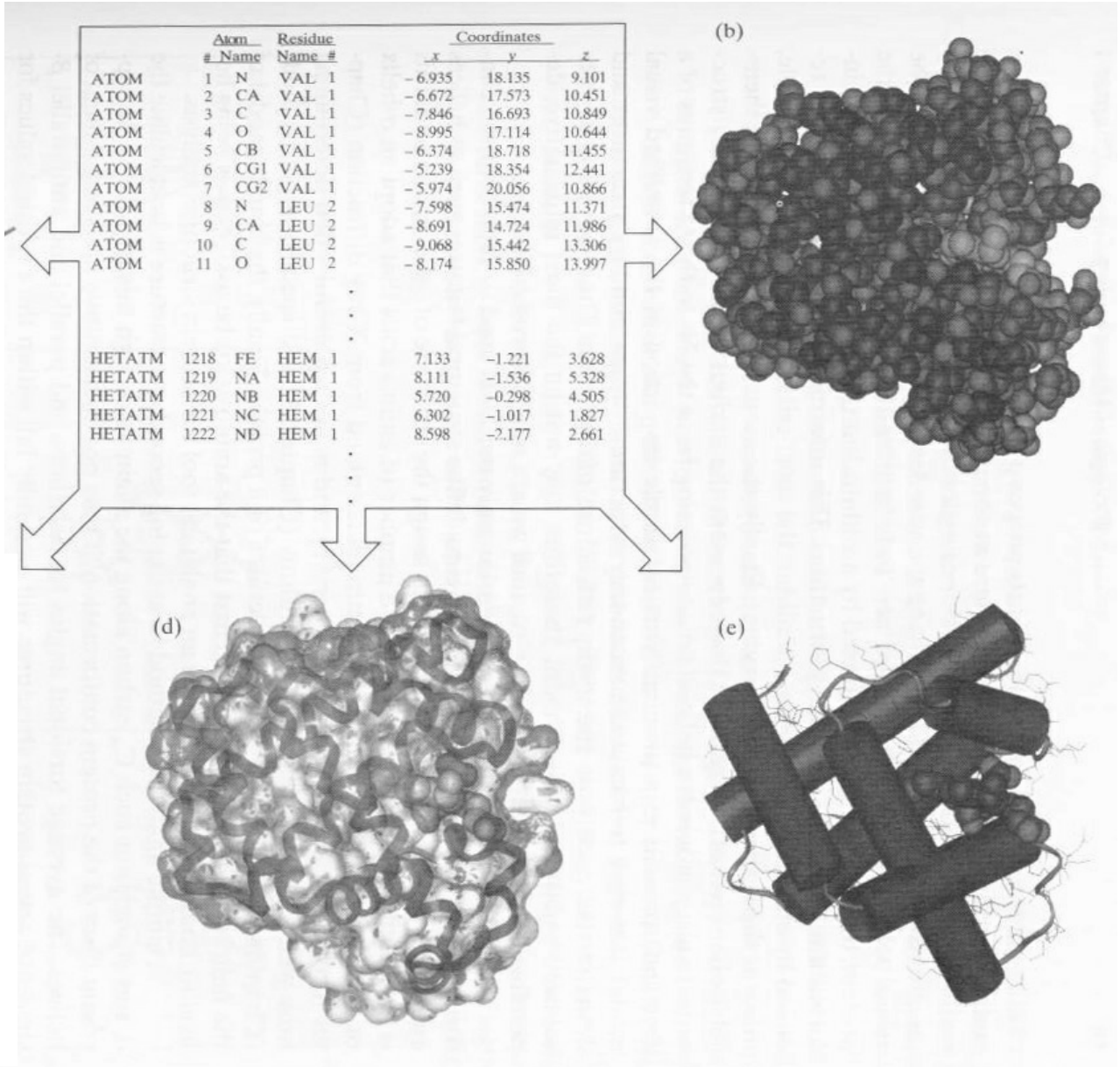


Figure 1.30 Representing the structure of a protein molecule. The structure of a macromolecule is a list of atoms and their (x, y, z) coordinates. This set of *atomic coordinates* can be interpreted to give a stick model (a), a CPK or van der Waal's surface model (b), a ribbon model (c), a solvent accessible surface model (d), or a simple caricature of the molecule (e).

Contact plots provide some information on the 3D structure of a protein

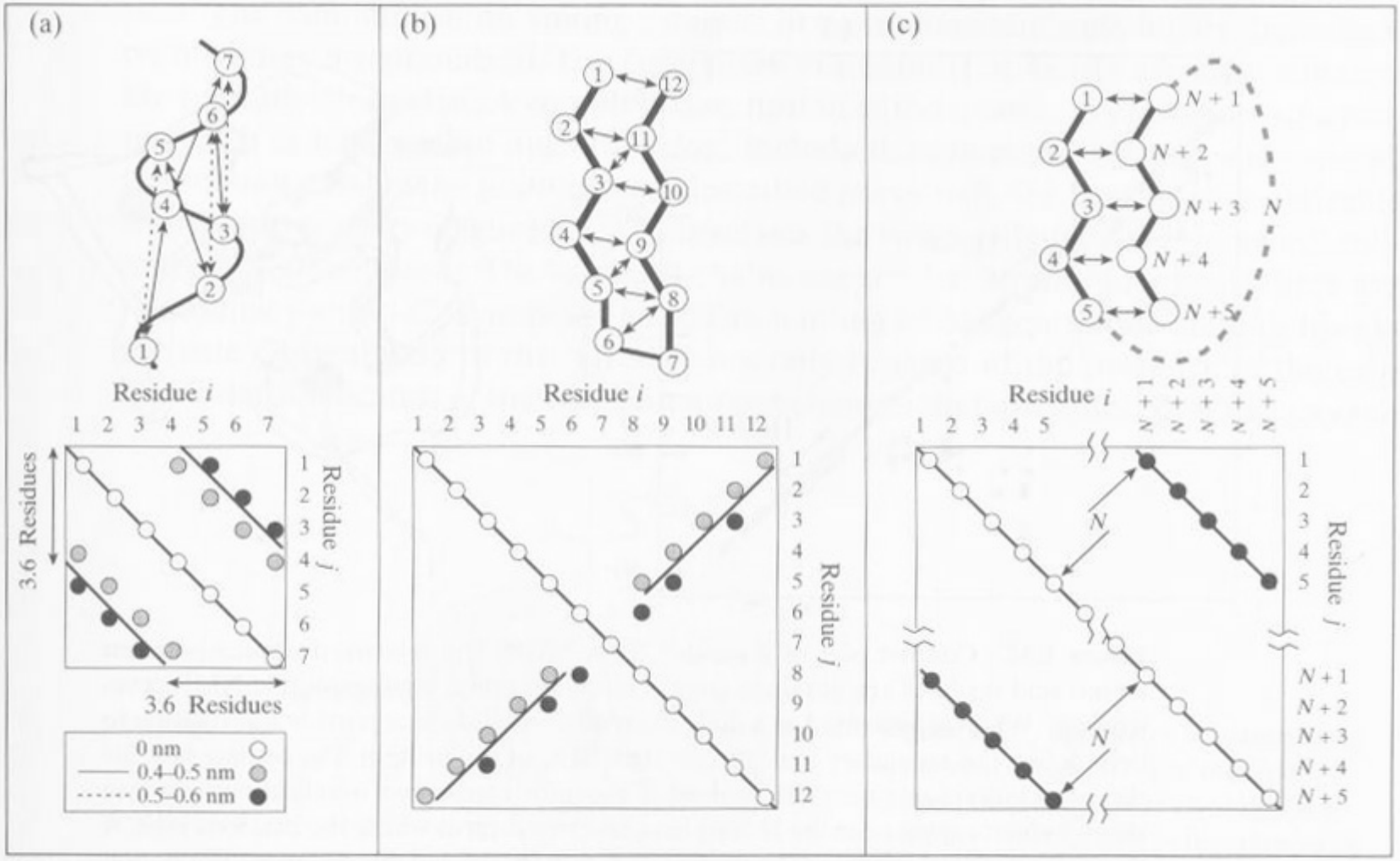


Figure 1.31 Contact plots for α-helix and β-sheets. The points of contact for the residues in (a) an α-helix, (b) antiparallel β-sheet, and (c) parallel β-sheet give characteristic patterns that are the signatures of each form of secondary structure.

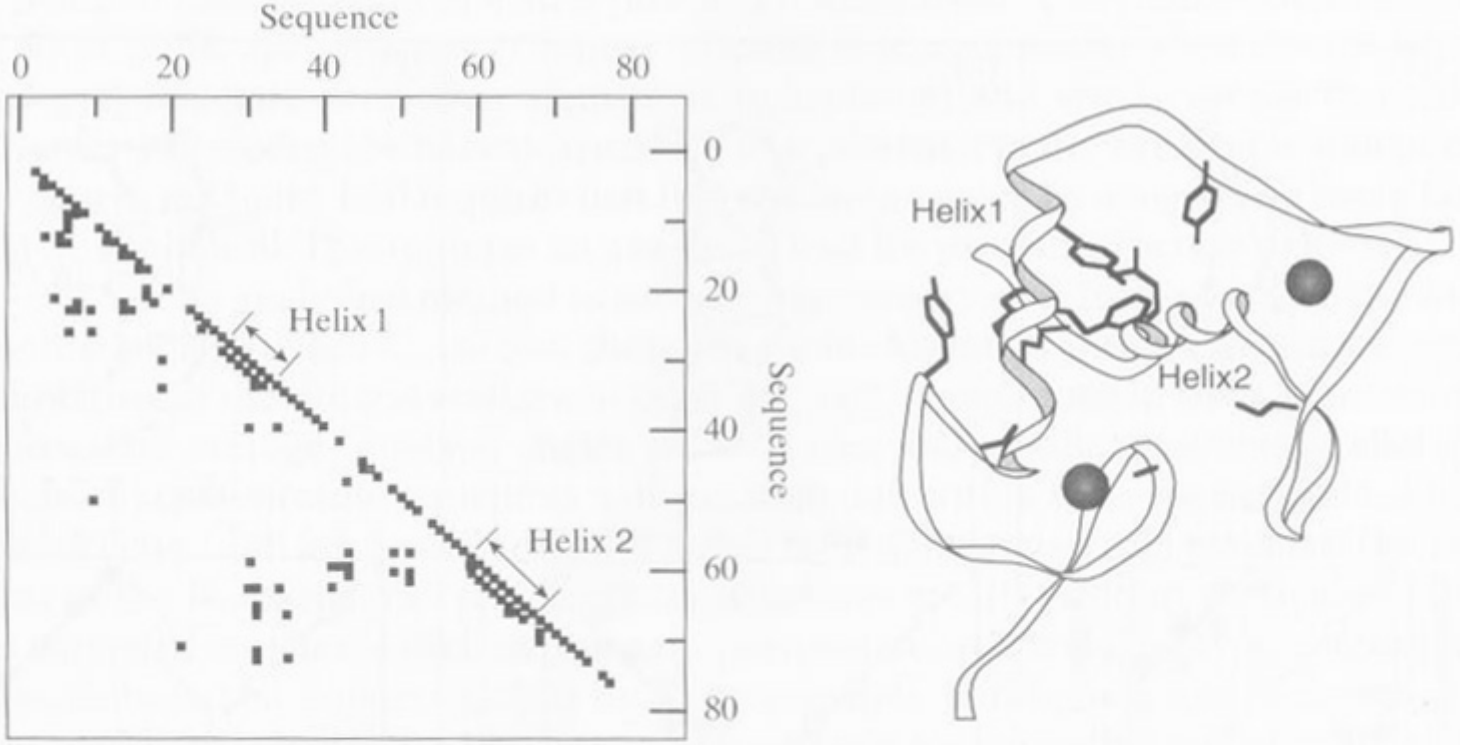


Figure 1.32 Contact plot of a protein from NMR. The relative distances between amino acid residues are obtained from multidimensional high-resolution NMR spectroscopy. When represented as a diagonal plot, these distances provide information to help define the secondary and tertiary structures of the protein. The contact plot for the *zinc-finger* domain of the oestrogen receptor shows two α-helical regions, one short β-sheet region near the N-terminus, and two β-turns where the zinc ions bind. A three-dimensional model of the protein is constructed using this distance information and the allowed geometries of amino acids. The structure apparently is stabilized by the zinc ions, as well as the packing of aromatic and alkyl side chains to form a hydrophobic core. [Adapted from Schwabe et al. (1990), *Nature,* **348**, 458–461.]

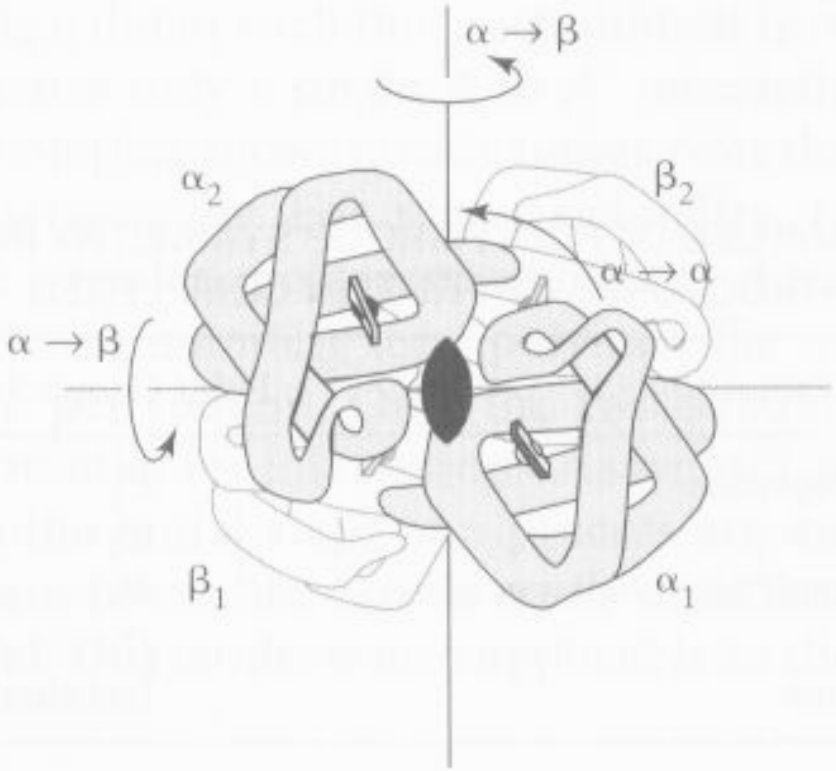


Figure 1.33 Symmetry in hemoglobin quaternary structure. The tetramer of hemoglobin shows true C_2-symmetry, with the two-fold axis perpendicular to the plane of the page. This relates α-subunit to α-subunit, and β-subunit to β-subunit. There are two pseudo-two-fold symmetry axes relating α- to β- subunits.

TABLE 1.8 EXAMPLES OF POINT GROUP SYMMETRY IN THE QUATERNARY STRUCTURES OF PROTEIN COMPLEXES

Protein	Point Group Symmetry
Alcohol dehydrogenase	C_2
Prealbumin	C_2
Hemerythrin from *Phascolopsis gouldii*	D_4
Hemocyanin	D_5
Viral coat proteins	Icosahedral

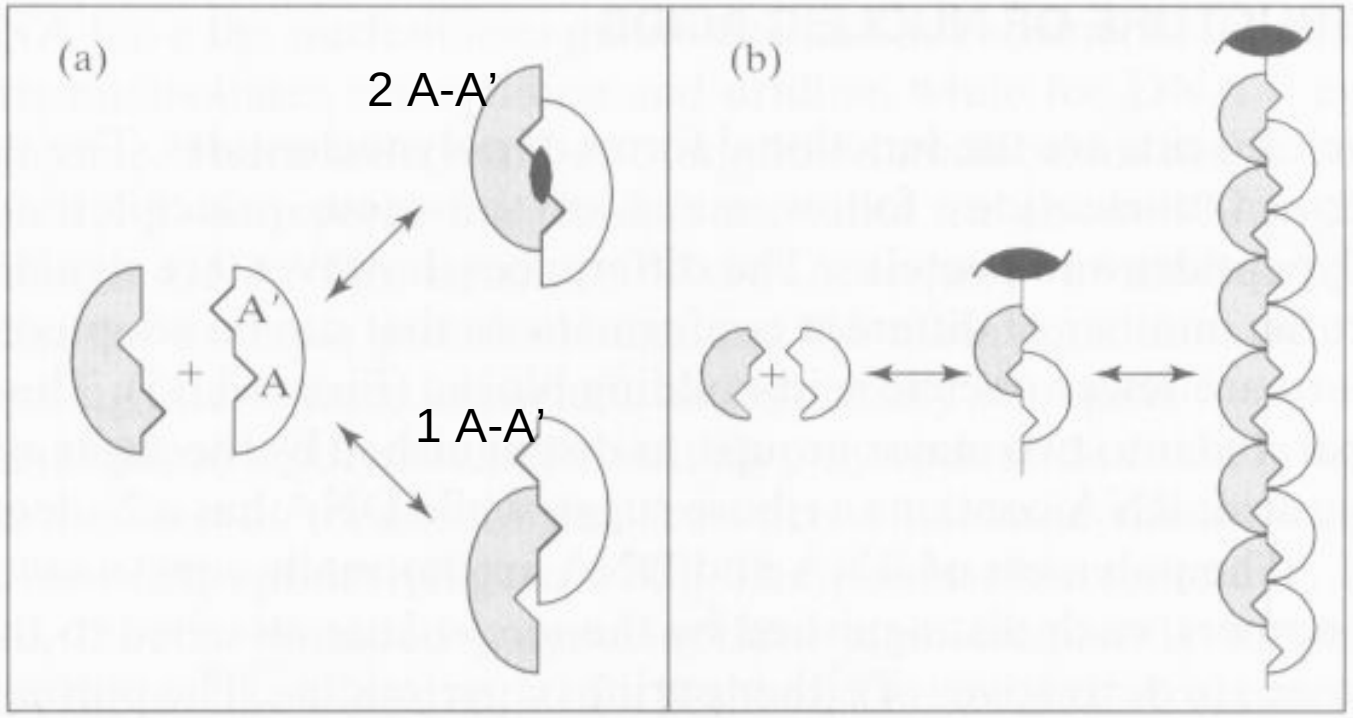


Figure 1.34 Symmetric versus nonsymmetric association of subunits. Panel (a) shows a possible symmetric and nonsymmetric association of two subunits, each having an interaction site A and its complement A′. In the symmetric association, there are two A-A′ interactions, while the nonsymmetric complex has only a single A-A′ interaction. (b) Subunits related by two-fold screw symmetry would each have two A-A′ interactions, except for the subunits at the two ends, which have only a single such interaction.

1.6 The structure of Nucleic Acids

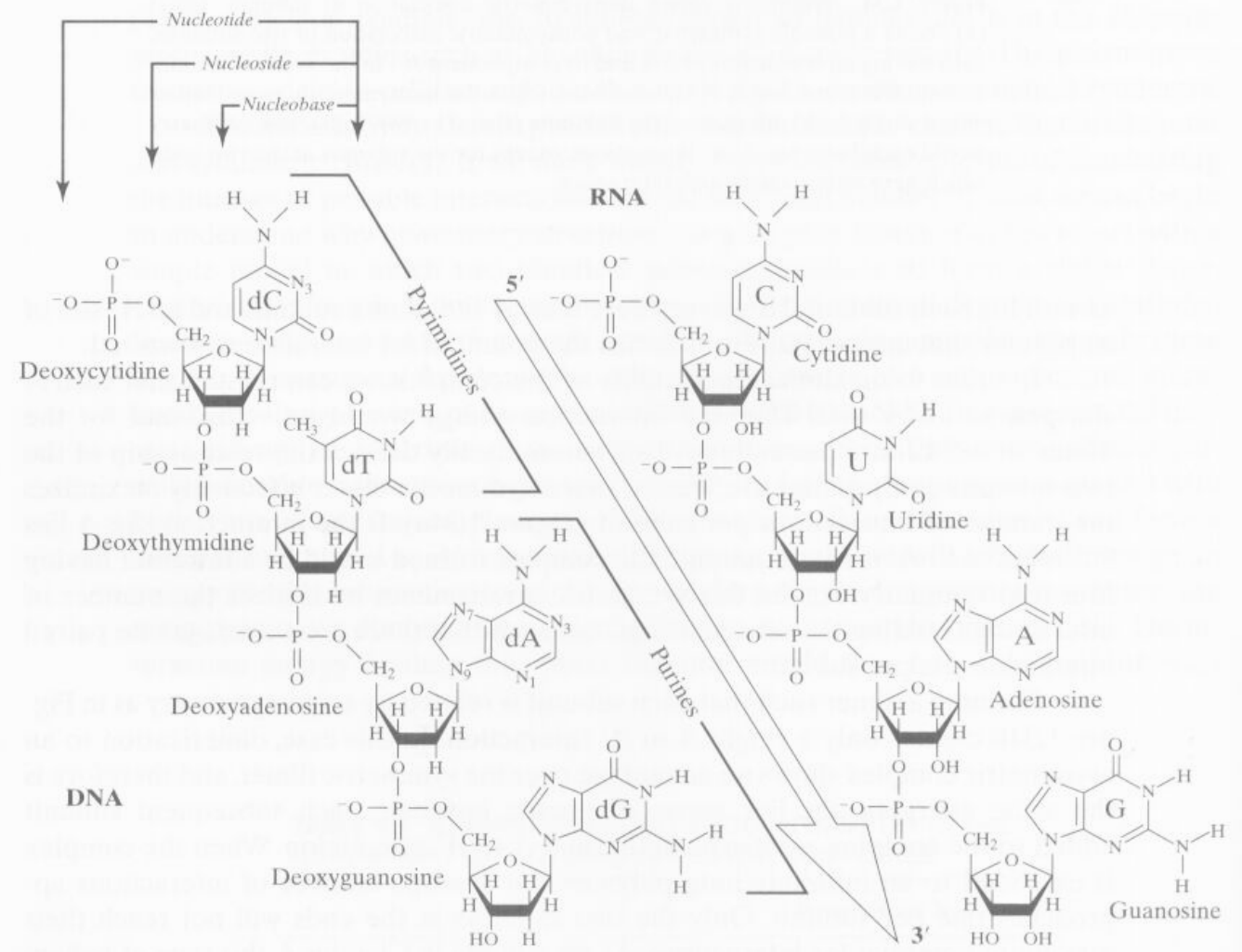


Figure 1.35 Nucleic acids. Ribonucleic acids (RNA) and 2′-deoxyribonucleic acids (DNA) are polymers constructed from nucleotide monomers. The nucleotides are distinguished by the nucleobases, which can be either of the pyrimidine or purine type. The polynucleotide sequence is a chain that extends from the 5′-terminus to the 3′-terminus.

1.6.1 Torsion Angles in the Polynucleotide Chain

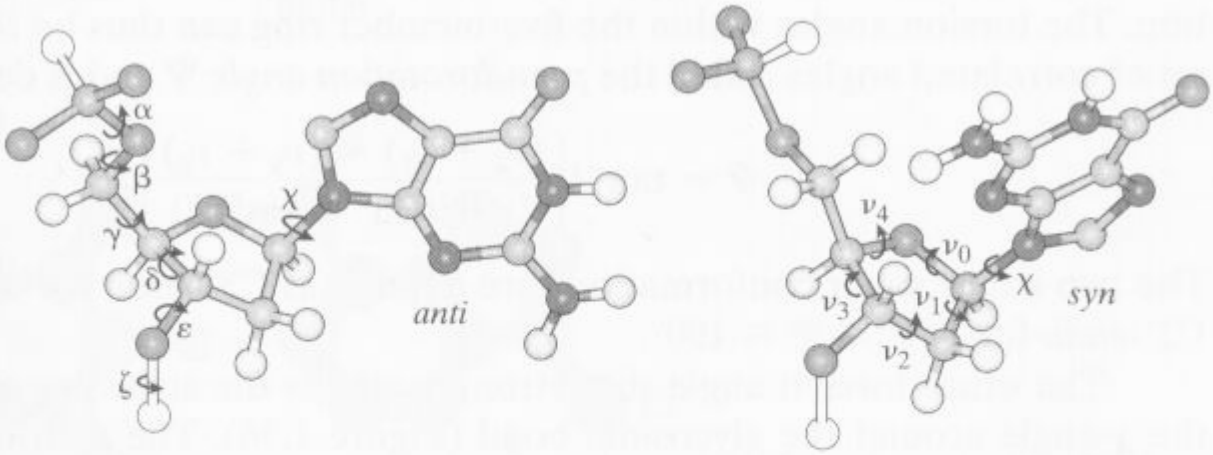


Figure 1.36 Torsion angles in nucleic acids. The structures of nucleic acids are defined by the torsion angles along the phosphoribose backbone (α to ζ), the torsion angles within the sugar ring (ν_0 to ν_4), and the rotation of the nucleobase relative to the sugar (χ). Rotation about χ places the base either extended from the ribose (*anti*-conformation) or sitting above the ring (*syn*-conformation).

Sugar conformation of nucleic acids

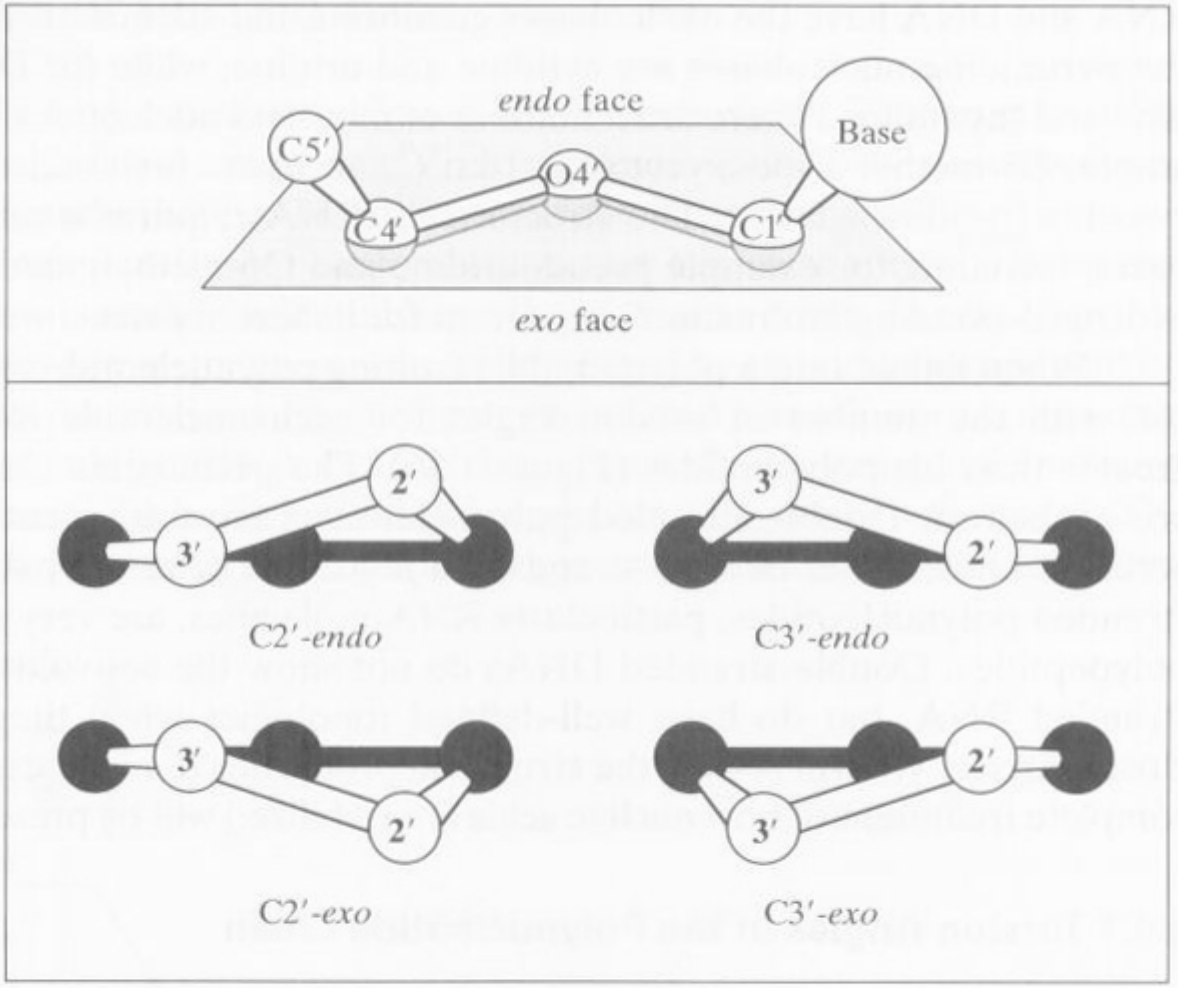


Figure 1.37 Sugar conformations of nucleic acids. The pucker of the sugar ring in RNA and DNA is defined relative to the plane formed by the C1′-carbon, C4′-carbon and O4′-oxygen of the five-member ring. The *endo* face lies above the plane, towards the nucleobase, while the *exo* face lies below the plane.

1.6.2 The Helical Structures of Polynucleic Acids

B-DNA
A-DNA
Z-DNA

H-DNA: triple-stranded

Cruciform DNA: inverted repeat sequences, with a dyad axis of symmetry between the 2 strands of the duplex.

G-quartet structure: 4 strands of polydeoxyganines, telomere end of chromsomal DNAs

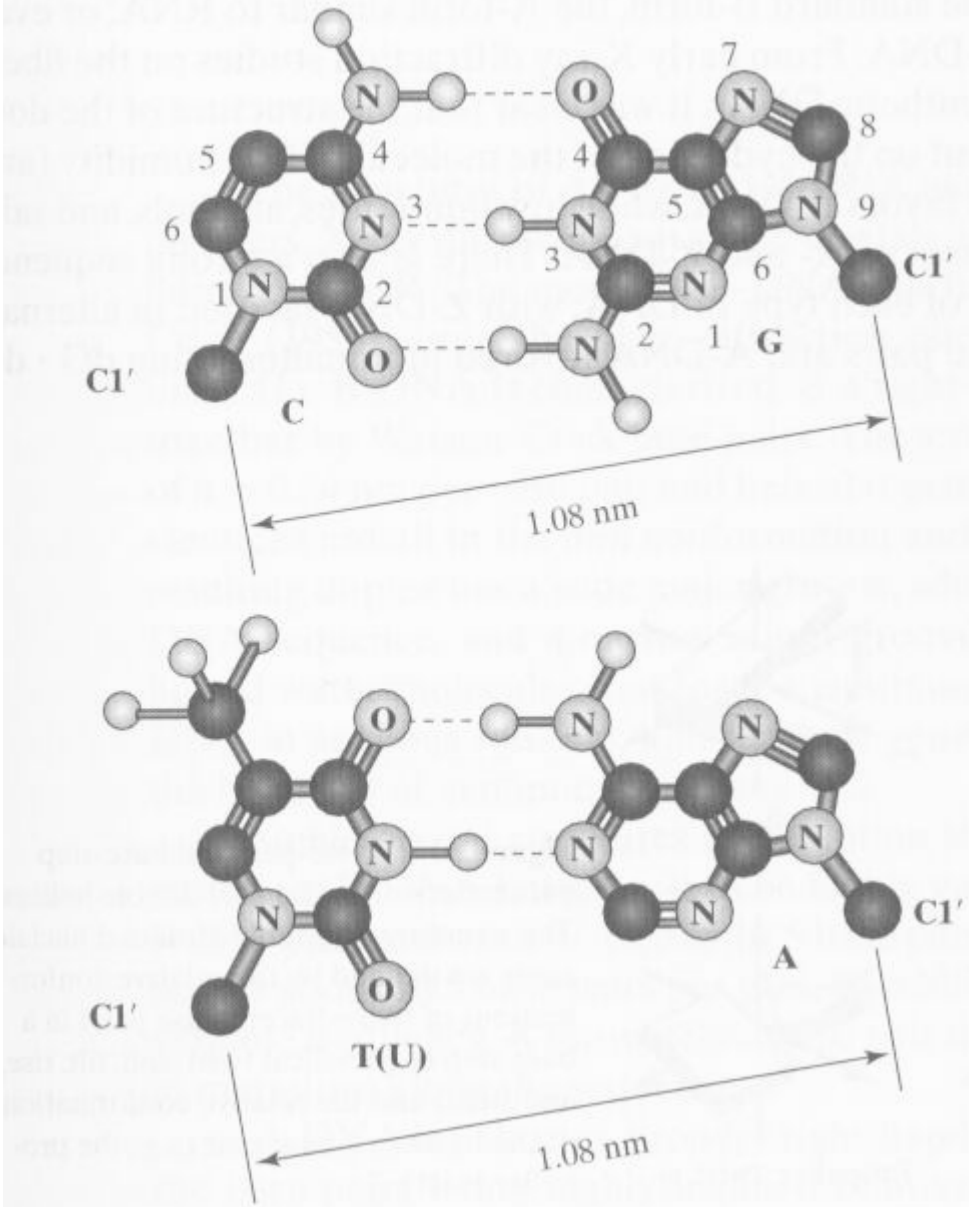


Figure 1.38 Watson-Crick base pairs. Bases interact by hydrogen bonds to form base pairs. The standard base pairs in double-stranded nucleic acids are the C·G and T·A Watson-Crick-type base pairs.

Base-pair and base-step parameters of nuclei acid double helices

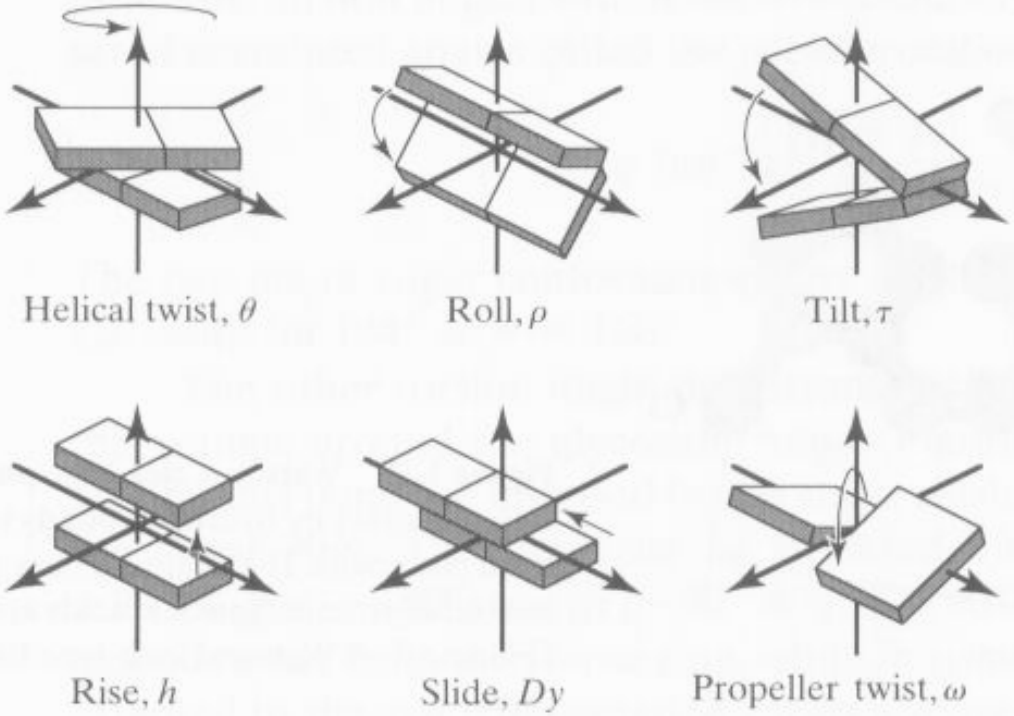


Figure 1.39 Base-pair and base-step parameters of nuclei acid double helices. The structure of double-stranded nucleic acids are defined by the relative conformations of two adjacent base pairs in a base step (e.g., helical twist, roll, tilt, rise, and slide) and the relative conformations of the bases in a base pair (e.g., the propeller twist).

Structure of DNA

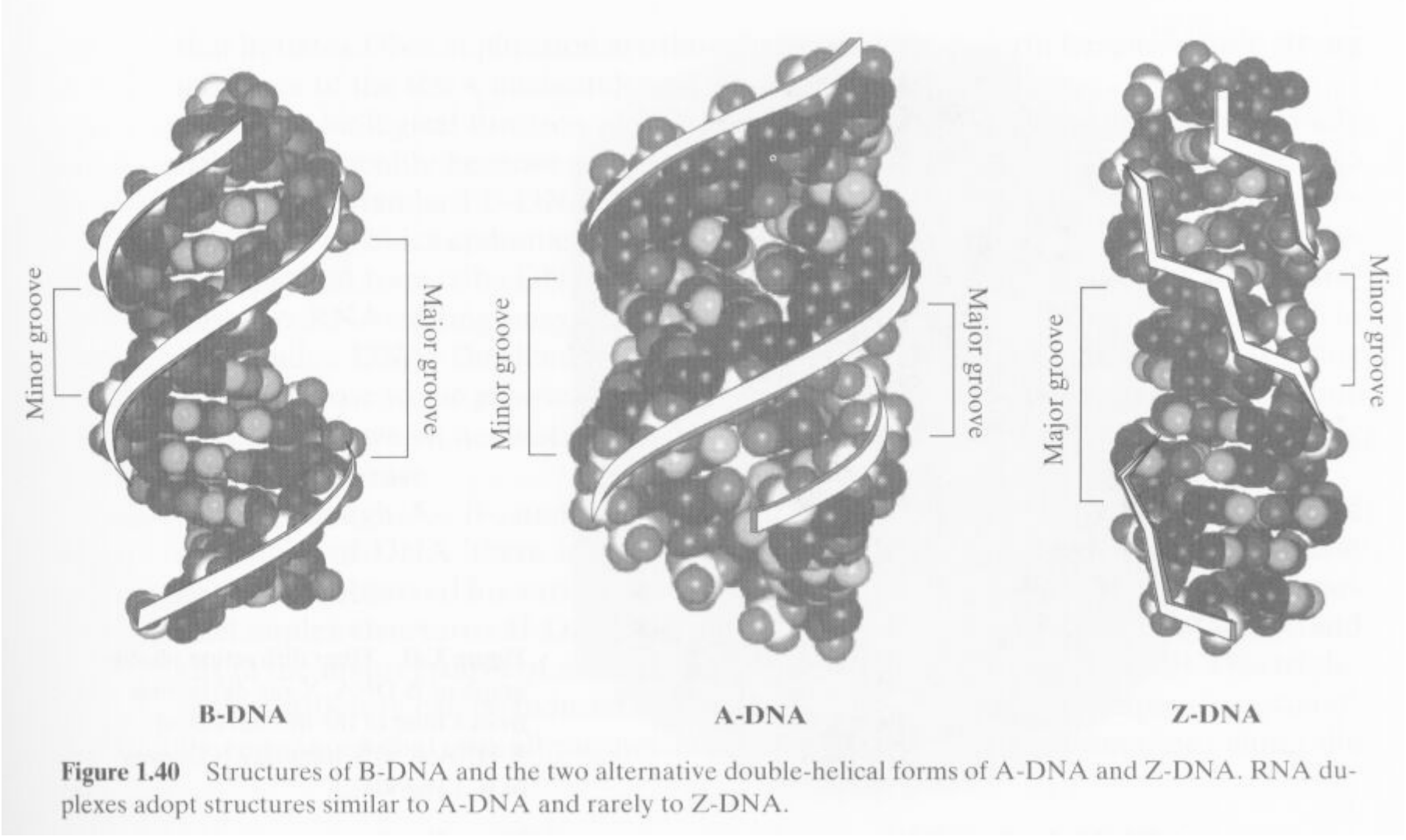


Figure 1.40 Structures of B-DNA and the two alternative double-helical forms of A-DNA and Z-DNA. RNA duplexes adopt structures similar to A-DNA and rarely to Z-DNA.

Fiber diffraction of B-DNA

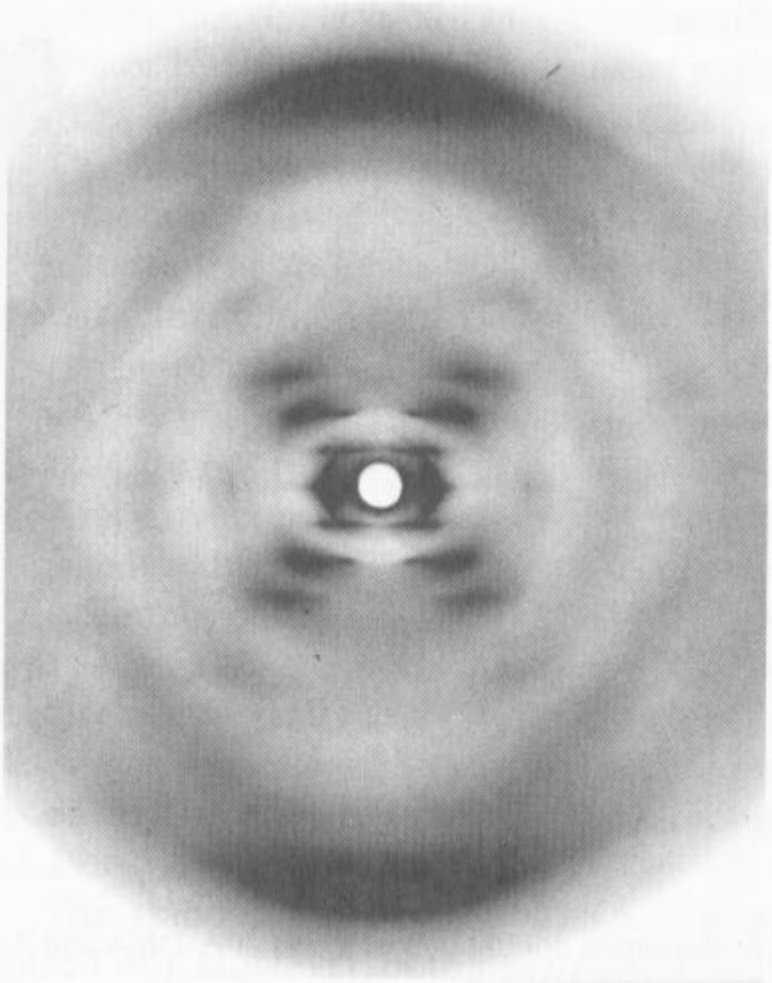


Figure 1.41 Fiber diffraction photograph of B-DNA. X-ray diffraction from a fiber of the lithium salt of B-DNA at 90% humidity. [Courtesy of R. Langridge.]

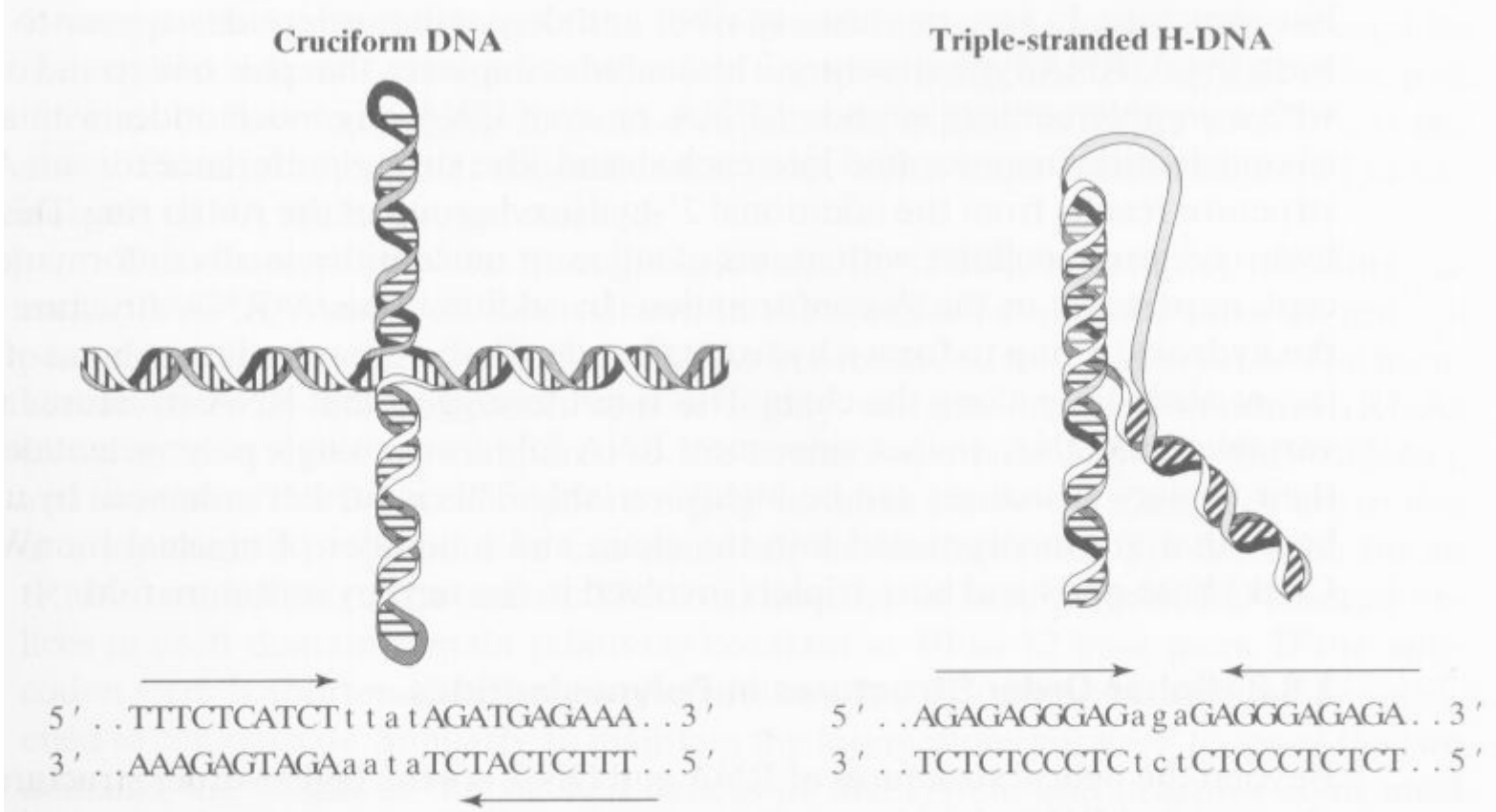


Figure 1.42 Cruciform DNA and triple-stranded H-DNA. Cruciform DNA is formed by inverted repeat sequences, with a dyad axis of symmetry between the two strands of the duplex. Triple-strand H-DNA is stabilized by hydrogen bonding between three nucleobases to form *base triplets*. These are typically formed by sequences that are rich in purines along one strand and pyrimidines in the complementary strand. When drawn as a duplex, the sequence shows mirror symmetry along the stands. [Adapted from Schroth and Ho (1995), *Nucleic Acids Res.*, **23**: 1977–1983.]

1.6.3 Higher-Order Structures in Polynucleotides

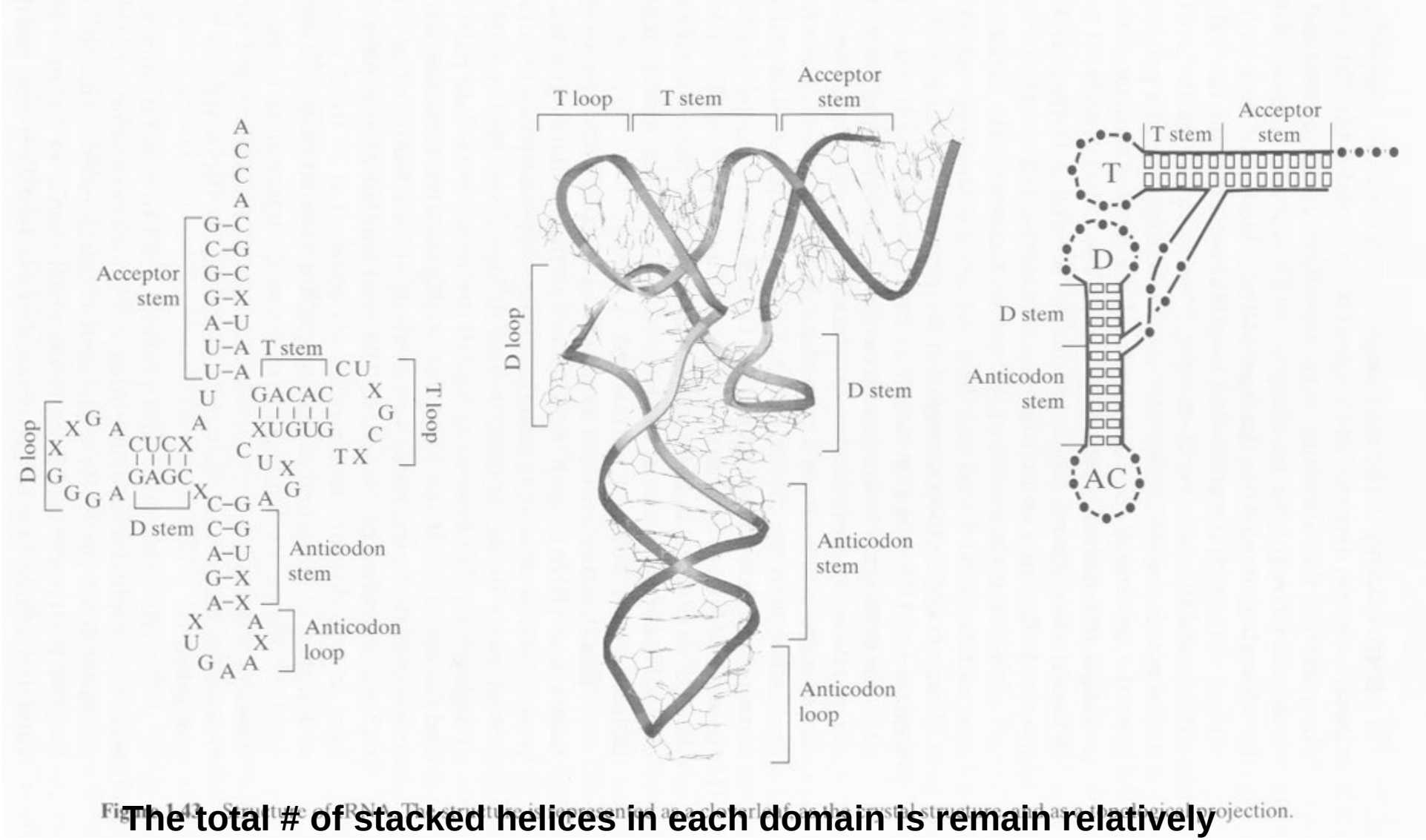


Figure 1.42 Structure of tRNA. The structure is represented as a cloverleaf, the most stable structure, and has two-dimensional projection.

The total # of stacked helices in each domain is remain relatively constant at 10-12 base pairs/ Acceptor stem↑, T stem ↓

The overall shape of tRNA 3° structure is largely conserved

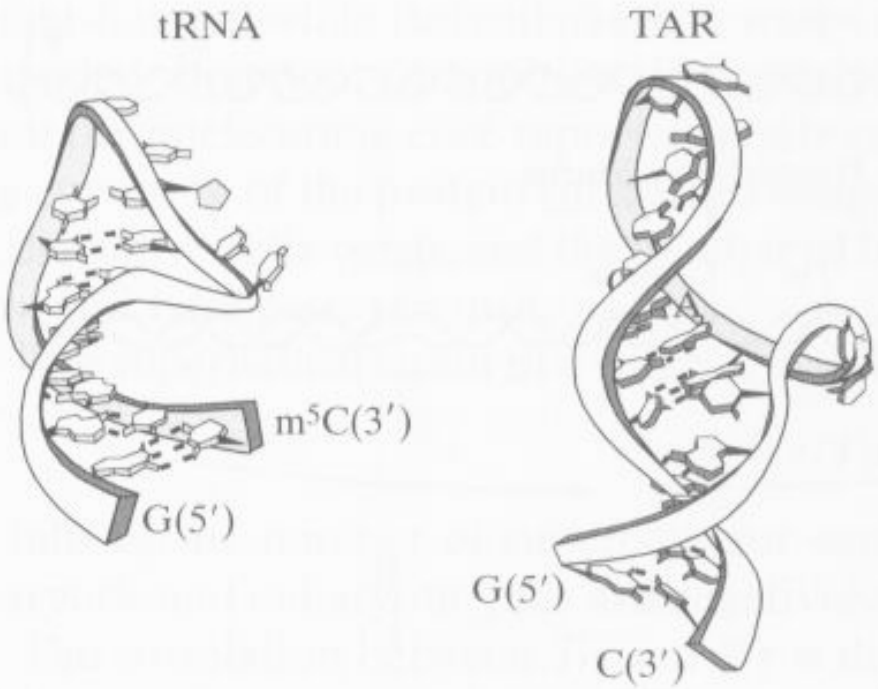


Figure 1.44 Structure of TAR. The consensus transacting RNA (TAR) sequence found at the 5′-ends of the genes encoded by the HIV virus is a stem-loop structure with a bulge in the middle of the stem. A molecular model of this sequence was constructed using an analogous bulged stem-loop structure found in the elbow of aspartyl tRNA (left) as a template. [Adapted from Loret et al. (1992), *Proc. Natl. Acad. Sci., USA*, **89**: 9734–9738.]

DNA topology

N = the length 147 base pairs
<c> average repeat
twist $\mathbf{T_w}$ = N / <c>
<c>~ 10.5 for B-DNA, $\mathbf{T_w}$ = 14

Supercoil/ writhe (**Wr**)
supercoil positive Wr>0,
Supercoil negative Wr<0

Closed circular DNA(ccDNA):
absorbed by a writhing or supercoiling of the circle

Superhelical density (σ):
the # supercoil for each turn of DNA,
σ~ -0.006

$$\sigma = \mathbf{W_r}/\mathbf{T_w}$$

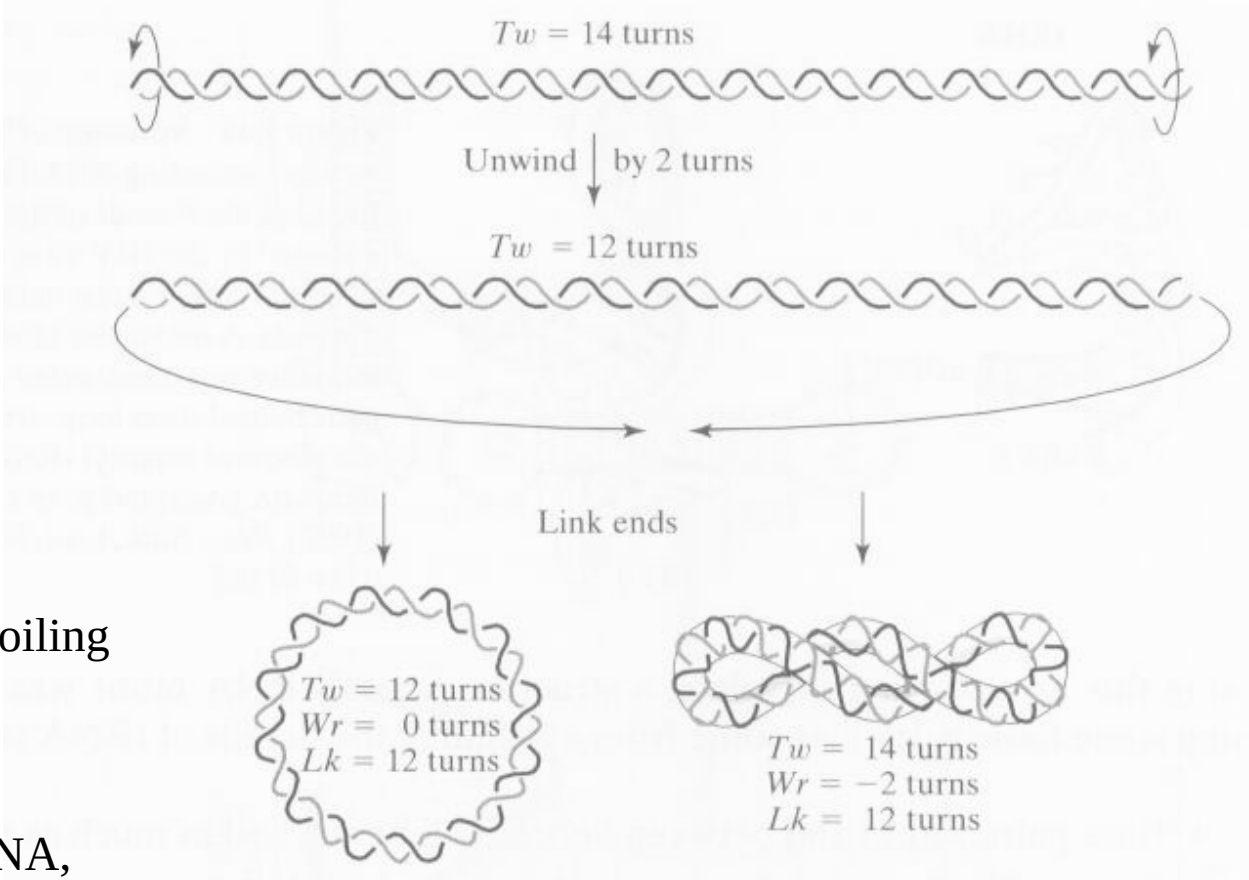


Figure 1.45 Supercoiling DNA. Unwinding 14 turns of B-DNA by 2 turns results in a loss in helical twist and generation of 2 negative supercoils.

Linking number (**Lk**): the overall conformation or topology of the ccDNA according to the degree to which torsional strain is partitioned between T_w and W_r $\mathbf{Lk = T_w + W_r}$

$$\mathbf{Lk = T_w + W_r}$$

LK is fixed and can only be changed by **breaking** the bonds of the phosphodiester backbone of one or both strand of the duplex.

Topoisomers

Lowest energy reference state is relaxed closed circular B-DNA

$T_w° = N/ 10.5$ turns, $W_r° = 0$ and $LK° = T_w°$

Thermodynamic need higher energy

$\Delta T_w = T_w - T_w°$, $\Delta W_r = W_r - W_r°$, $\Delta LK = LK - LK°$,

$\Delta LK = \Delta T_w + \Delta W$

more compact supercoiled form of the plasmid migrate faster than the relaxed ccDNA.

Ex: 1050 bp B-DNA ⇒ A-DNA

$\Delta T_w = T_{A\text{-}DNA} - T_w°$ = 1050/11 turns- 1050/10.5 turns = -4.5

$\Delta LK = \Delta T_w + \Delta W$

For a topoisomer with $\Delta LK = 0$, $\Delta W_r = +4.5$ turn,

Either $\Delta T_w = 0$ turns , $\Delta W_r = +4.5$ supercoils

Or $\Delta T_w = -4.5$ turns , $\Delta W_r = 0$ supercoils

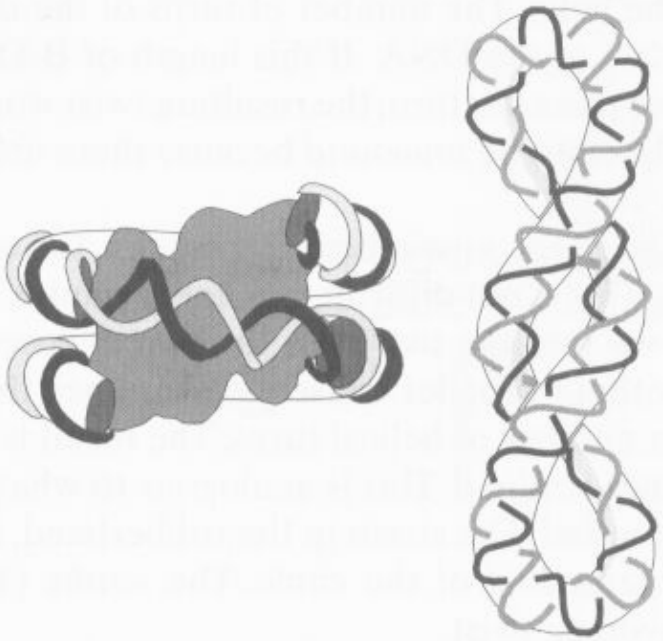


Figure 1.46 Two forms of supercoiled DNA. Negatively supercoiled DNA found in the chromatin structure wraps twice around the nucleosome core proteins in a left-handed direction. Negatively supercoiled DNA in the absence of a core forms right-handed crossovers. [Adapted from Arents and Moudrianakis (1993), *Proc. Natl. Acad. Sci., USA,* **90**: 10489.]

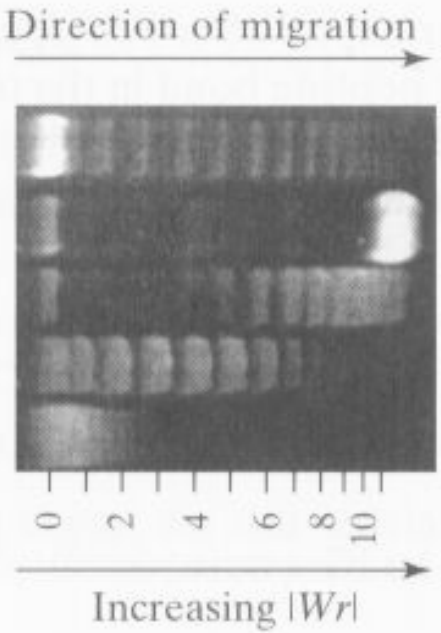


Figure 1.47 Topoisomers of a bacterial plasmid. The plasmid pBR322 can exist as a relaxed closed circle ($|Wr| = 0$ turns), as a highly supercoiled closed circle ($|Wr| > 20$ turns), or as a mixed population with the writhe distributed over a broad range of $|Wr|$ values. These can be resolved by agarose gel electrophoresis with the more compact supercoiled form of the plasmid migrating faster in the electric field than the relaxed closed circular form. [Courtesy of M. N. Ho.]