The folded conformations of globular proteins

Chapter 6

Proteins

TE Creighton

Conformations

- All bioloical active proteins belong to the class of globular proteins.
- They differ drastically from the conformations of random peptides or the simple structures of the structural proteins.
- A folded protein does not change its structure gradually, but it tries to keep its structure as long as possible, and then collapses.
- This unfolded structure is also known as denatured state.

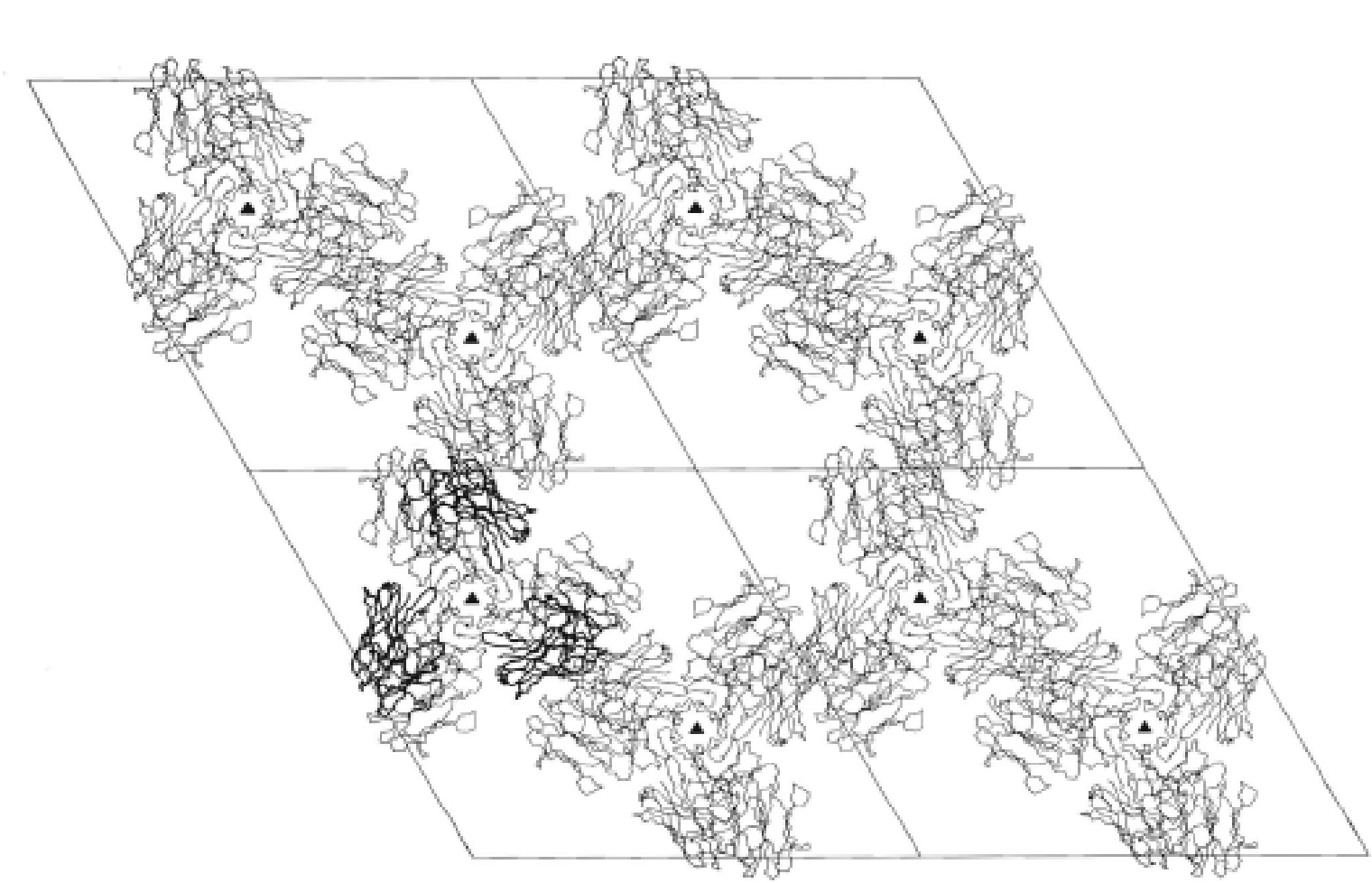
Structures

- To determine the structure of proteins
- Light is needed with a wavelength comparable to the size of the structure that one wants to visualise
- Therefore to visualize something in the Å size range, one needs X-rays, electrons or neutrons.
- X-rays and neutrons are scattered by proteins in solution, but they can not be focused to reconstruct an image of a single protein.
- This technique provides a mean to determine the size and radius of gyration, and shape. Resolution 20-40Å

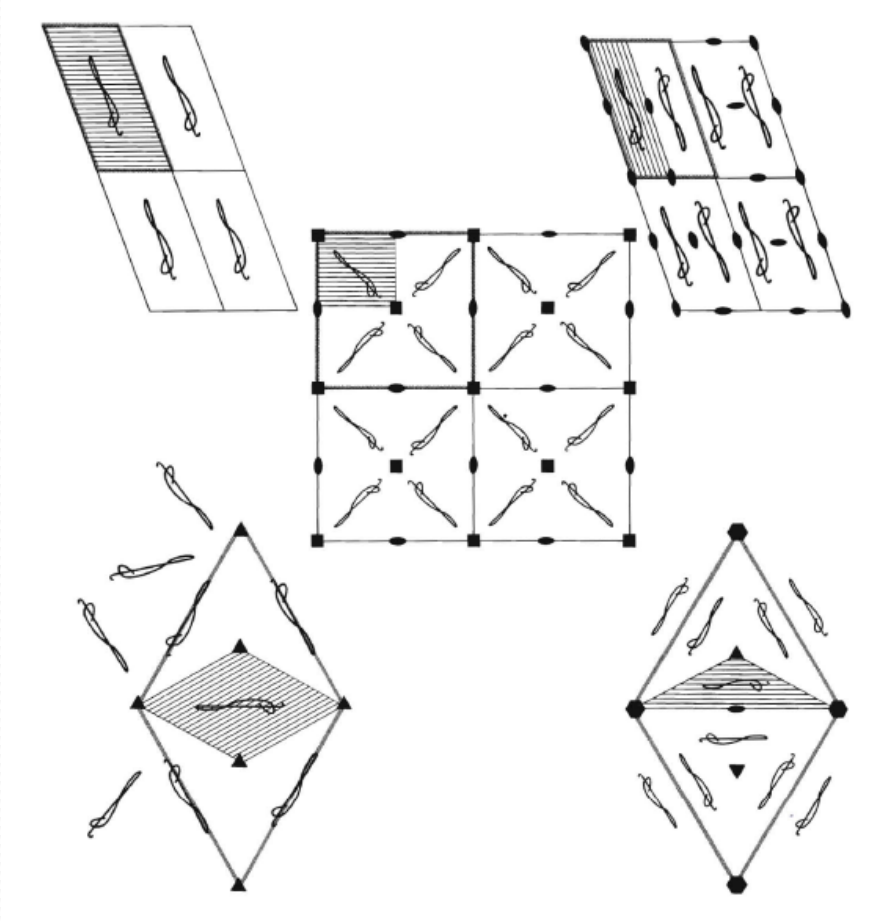
X-ray determination of structures

- One prerequisite for this is that the protein can be crystallised
- The crysdtals must be large.
- Individual
- Well ordered
- Problem:
- Proteins have been selected by evolution to be not crystalline
- Protein crystallisation today has still a touch of black magic

X-ray determination of structures



The unit cell in protein crystals



Unit cells and symmetry operations

- Not all space groups for proteins with L-amino acids are possible
- The X-rays used today stem from Synchrotron radiation and are of continous wavelength.
- The X-rays are scattered by the electrons of the object
- The scattering is proportional to the atomic number of the atom.
- The structure determined by X-ray diffraction analysis is that of the electron density of molecules in the asymetric unit.

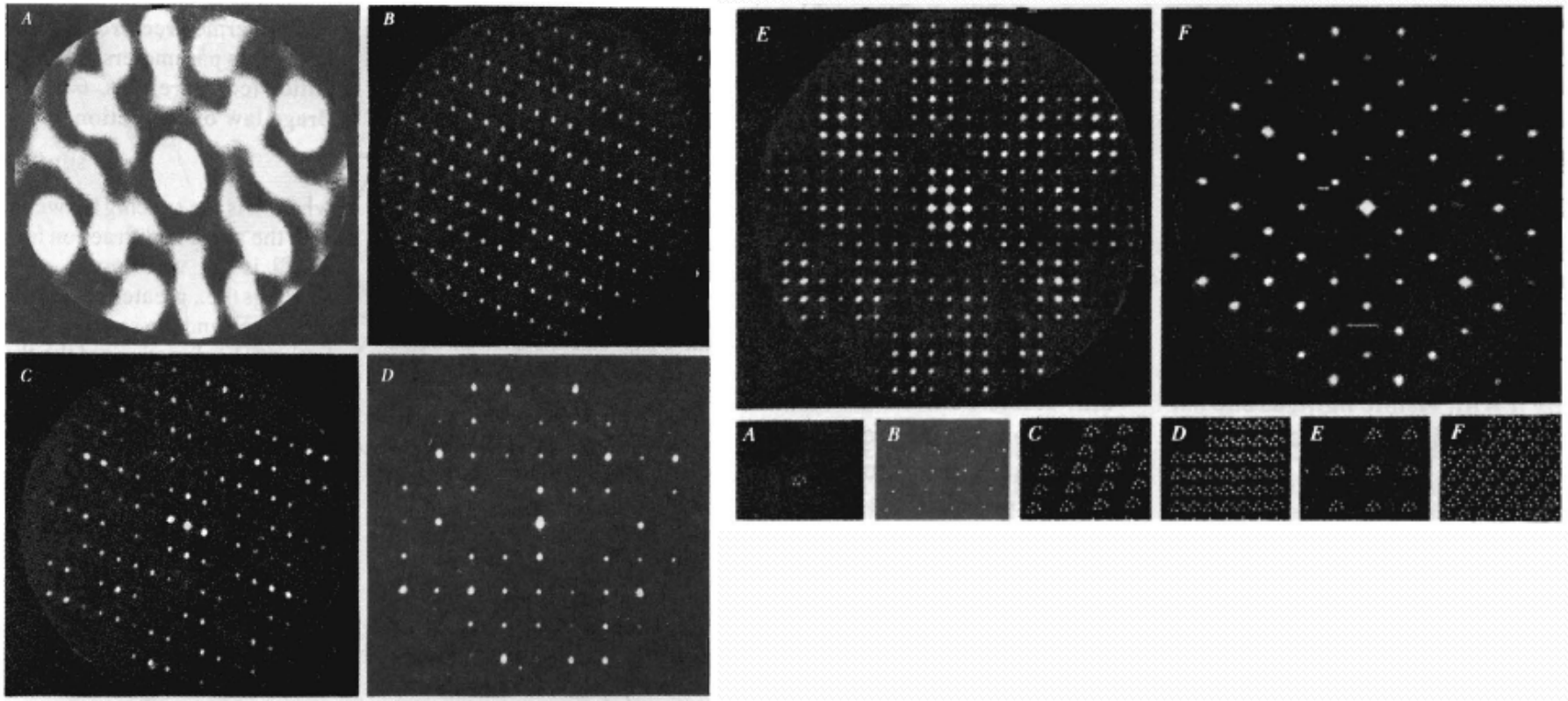
Unit cells and symmetry operations

Table 6.1 *Amplitudes of X-ray and Neutron Waves Scattered by the Atoms Comprising Proteins*

Element or isotope	Number of electrons	Amplitude (10^{-12} cm/atom)[a]	
		X-ray	Neutron
H	1	0.28	−0.374
^{2}H (deuterium)	1	0.28	0.667
C	6	1.69	0.665
N	7	1.97	0.94
O	8	2.25	0.58
S	16	4.5	0.28

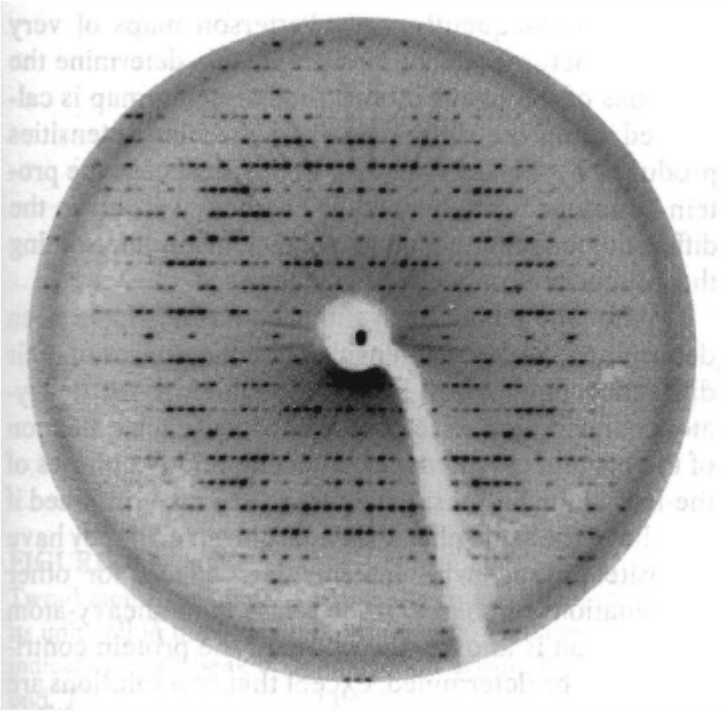
[a] All amplitudes are of positive sign unless otherwise noted. Only coherent scattering is considered, which predominates with these atoms.

X-ray diffraction



The effect of the crystal lattice on the diffraction pattern. illustrated with optical transforms. The masks (bottom row) were used to generate the six optical transforms at the top. The holes in the masks represent the atoms of one or more molecules; the optical transform gives the diffraction pattern of the mask. *A shows the continuous transform of a single sixatom molecule. B shows the pattern from a lattice only. which gives just a uniform series of reciprocal lattice points. The consequences of incorporating the "molecule" of A into various lattices (C-F) is to sample the continuous transform of A at the points of the corresponding reciprocal lattice. (From C. A. Taylor and H. Lipson. Optical Transforms, Bell & Hyman Ltd.• London. 1964.)*

X-ray diffraction

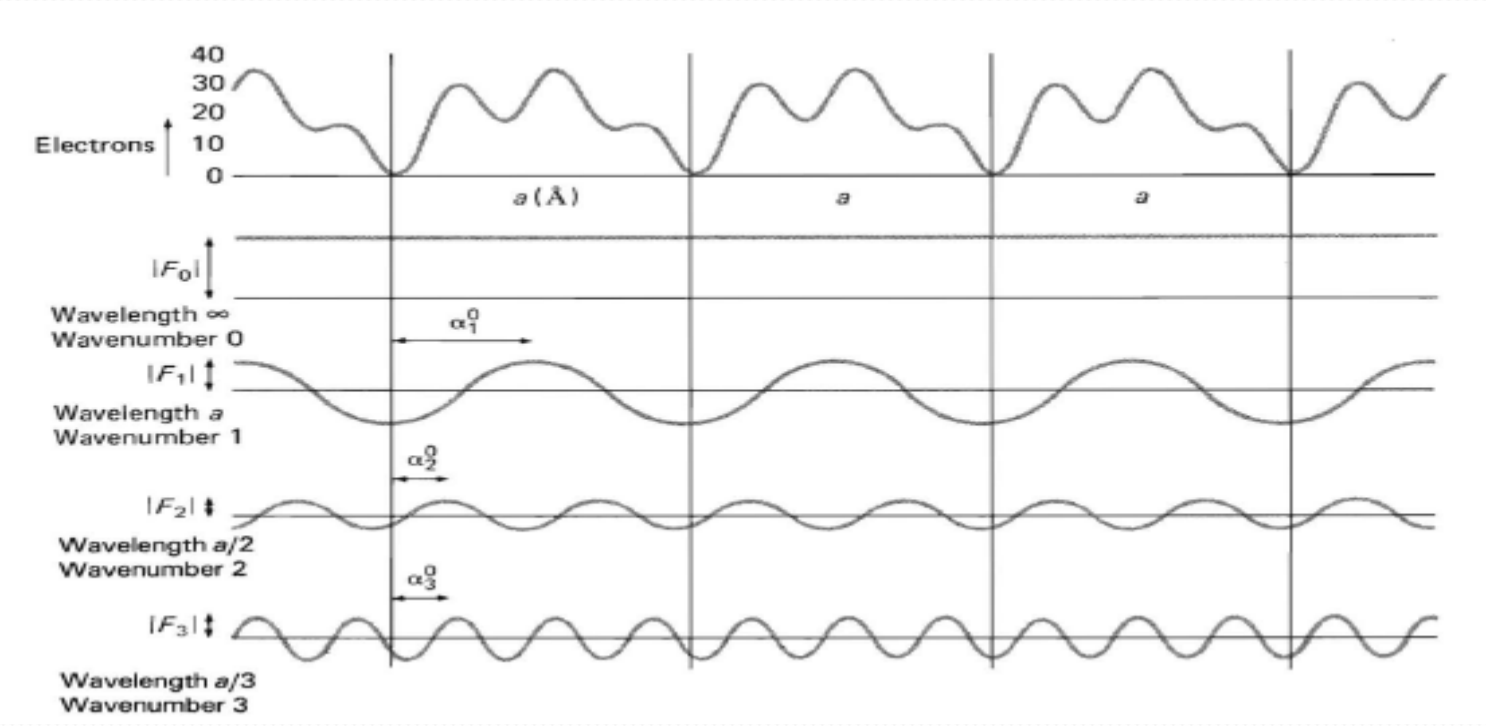


Example of an X-ray diffraction precession photograph. The *hOI plane from the diffraction pattern of a crystal of the pore-forming fragment of colicin A is illustrated. The crystal was tetragonal. with space group P432t 2 and unit cell dimensions a = b = 73.0 Aand c = 171.6 A. (Photograph kindly provided by Dmitri Tsernoglou.)*

Phase determination

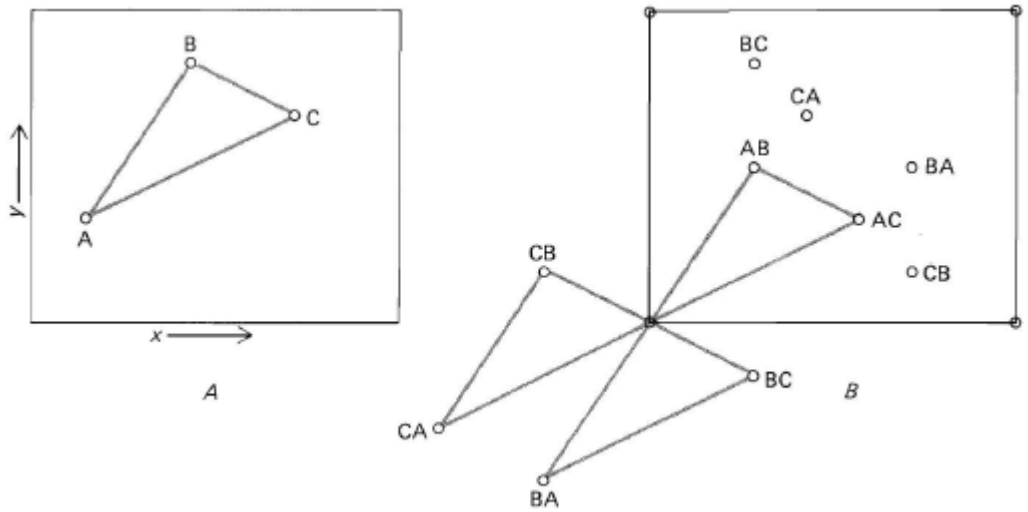
- The phase problem of crystallography
- X-rays can not be focused
- The diffraction pattern has to be recorded
- The intensities of the reflections are recorded
- Isomorphous replacement:
- Incorporation of heavy atoms: Uranium, Mercury, Platinum)
- Selenomethionine

Phase determination



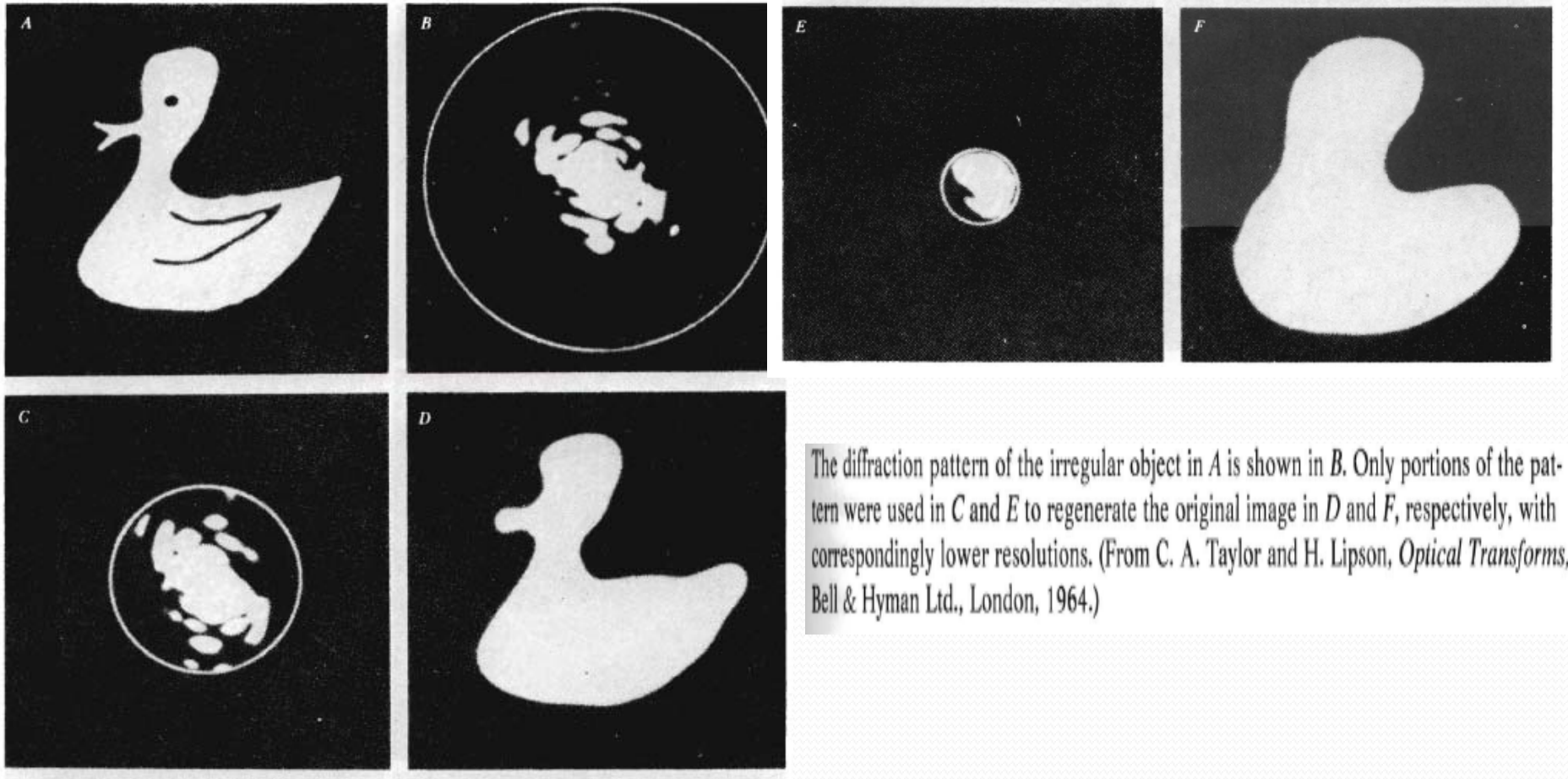
Fourier synthesis of a one-dimensional electron density profile *(top) using four terms, F⁰ through F3 • The unit cell dimension is a. IFil and lYi are, respectively, the amplitude and the phase of the ith term. The sum of these four terms gives the electron density at the top. (Adapted from C. C. F. Blake, Essays Biochem.* ***11:37, 1975.)***

Difference Patterson maps

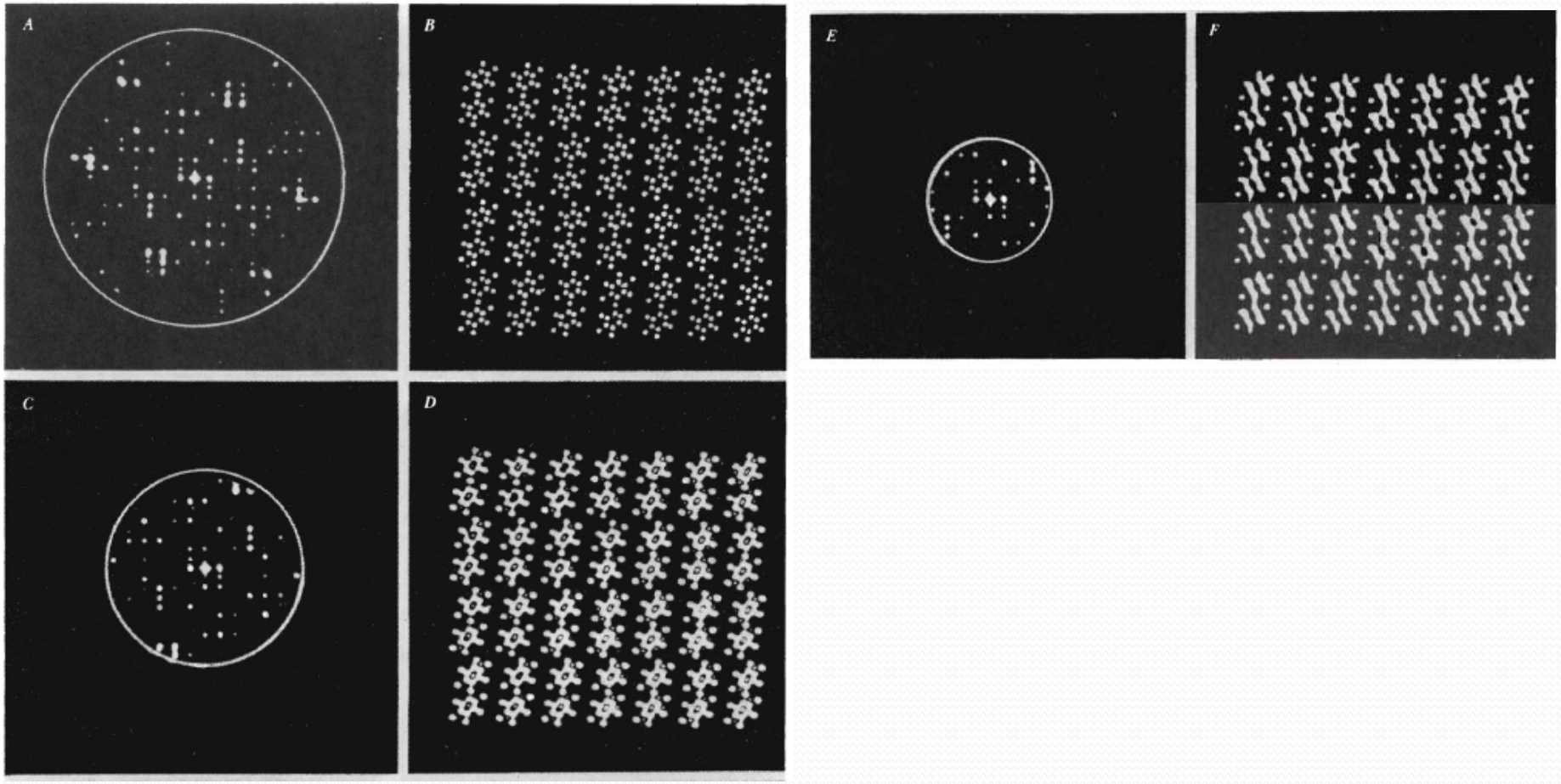


- Two-dimensional Patterson map. *A: Hypothetical molecule with three atoms (A, B, and C) in its unit cell in the crystal. B: Theoretical Patterson map of A with the positions of the peaks indicated. Each peak is designated as arising from the vector from the first atom to the second. The map can be generated by placing each of the three atoms, in turn, at the origin with the molecule in the correct orientation and marking the positions of the other two atoms; the resulting three representations of the molecule are indicated by the lines joining the appropriate peaks. The Patterson map is repetitive, just as the crystal is, and one unit cell is indicated by the rectangle. Large peaks at each of the unit cell origins are due to the self-vectors (AA, BB, and CC).*

The calculation of the electron density map

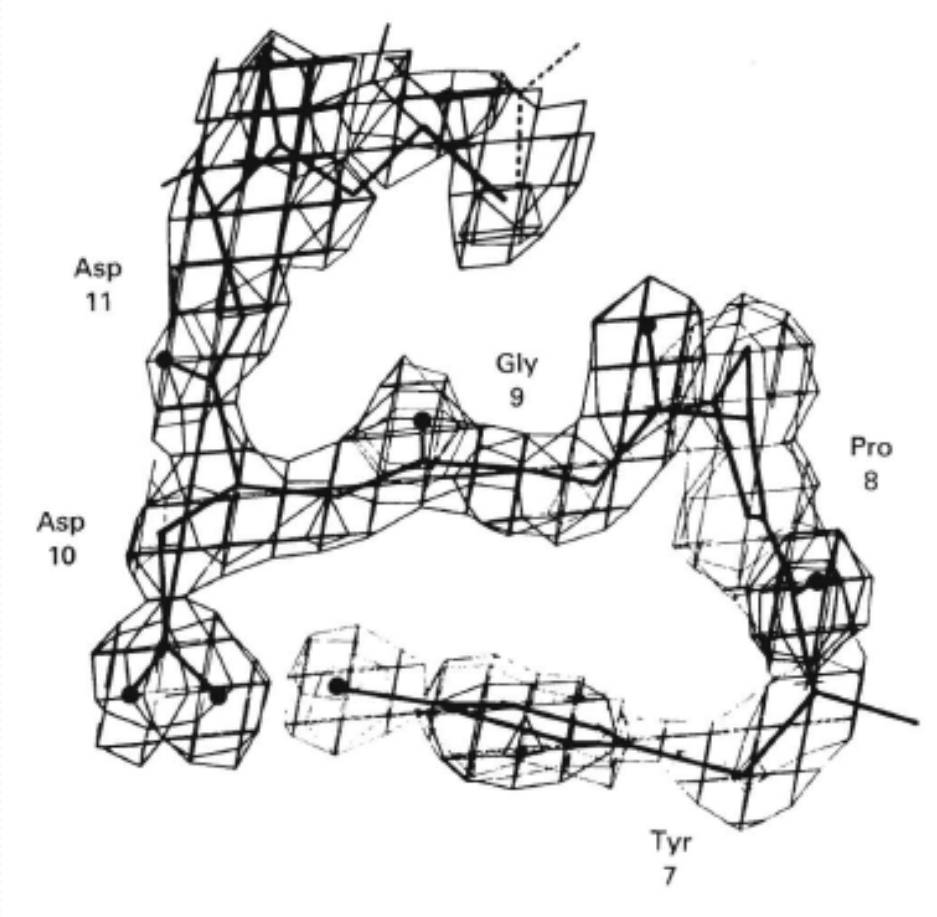


The diffraction pattern of the irregular object in *A* is shown in *B*. Only portions of the pattern were used in *C* and *E* to regenerate the original image in *D* and *F*, respectively, with correspondingly lower resolutions. (From C. A. Taylor and H. Lipson, *Optical Transforms*, Bell & Hyman Ltd., London, 1964.)



The effects of using fewer data of lower resolution in structure analysis, illustrated with optical transforms. The diffraction patterns enclosed in circles in *A*, *C*, and *E* were used to regenerate the original crystal lattice in *B*, *D*, and *F*, respectively; the clarity is correspondingly decreased upon using less of the diffraction pattern. (From C. A. Taylor and H. Lipson, *Optical Transforms*, Bell & Hyman Ltd., London, 1964.)

Interpretation of the electron density map



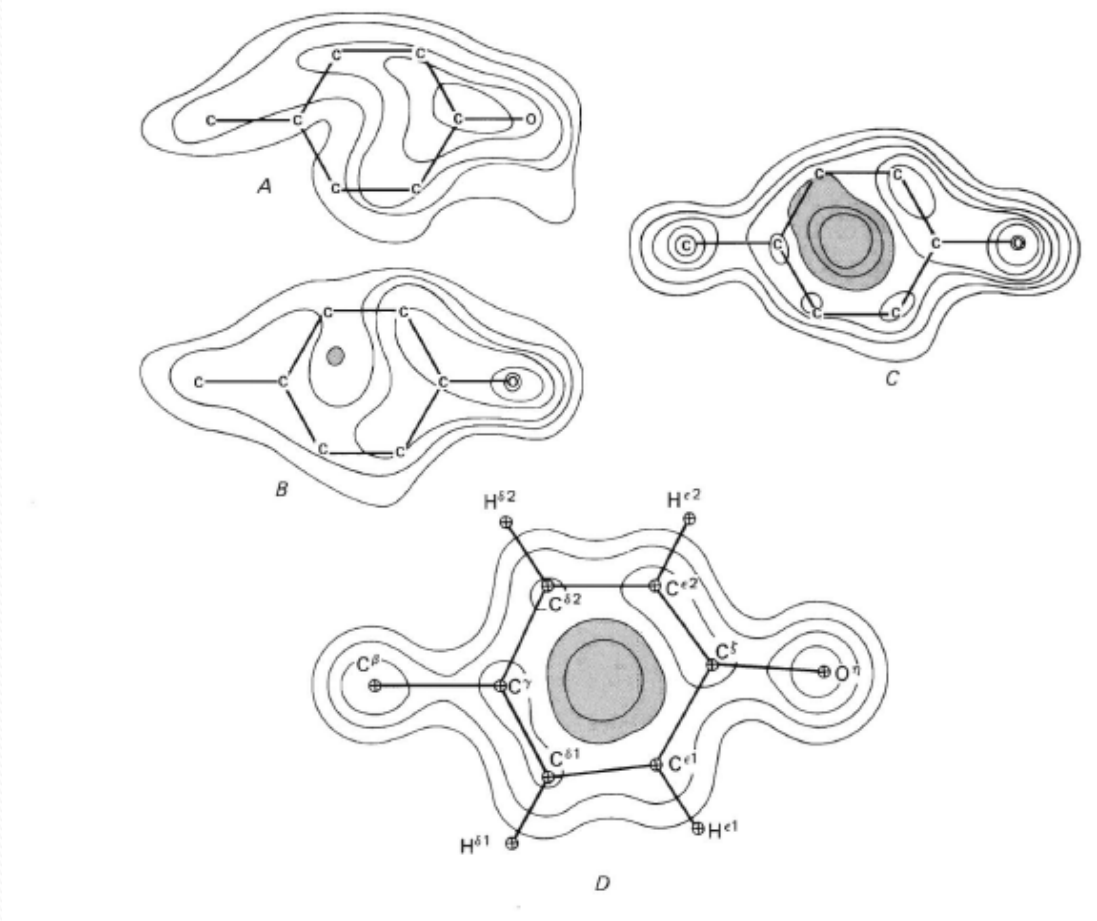
Refinement model

$$R = \frac{\Sigma_{h,k,l}|F_{obs}(h,\ k,\ l) - F_{calc}(h,\ k,\ l)|}{\Sigma_{h,k,l}F_{obs}(h,\ k,\ l)}$$

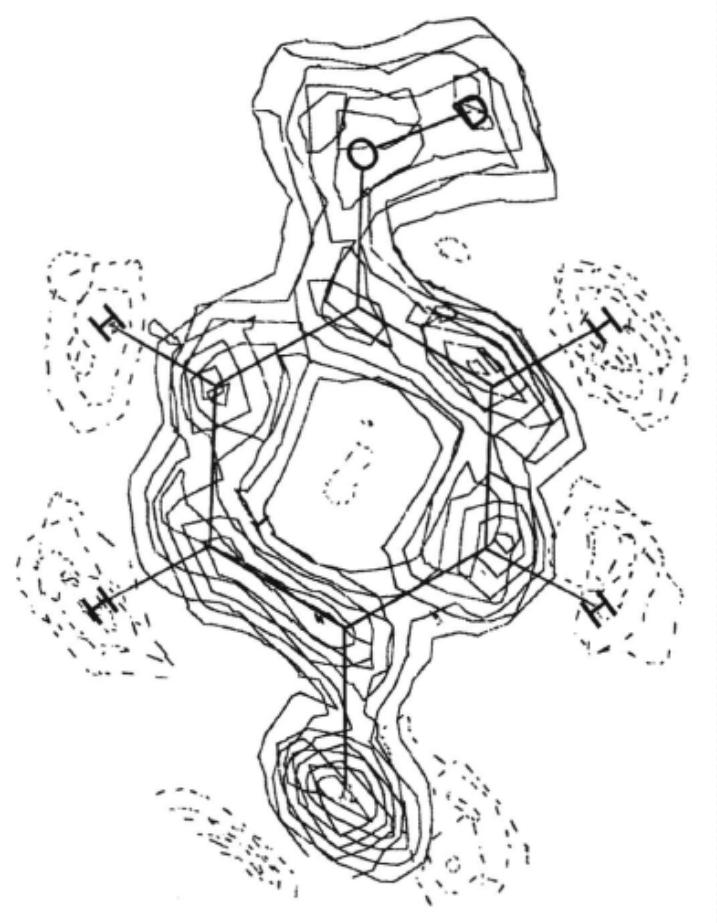
$$\exp\left(-B_i \frac{\sin^2\theta}{\lambda^2}\right)$$

$$B_i = 8\pi^2 U_i^2 \cong 79 U_i^2$$

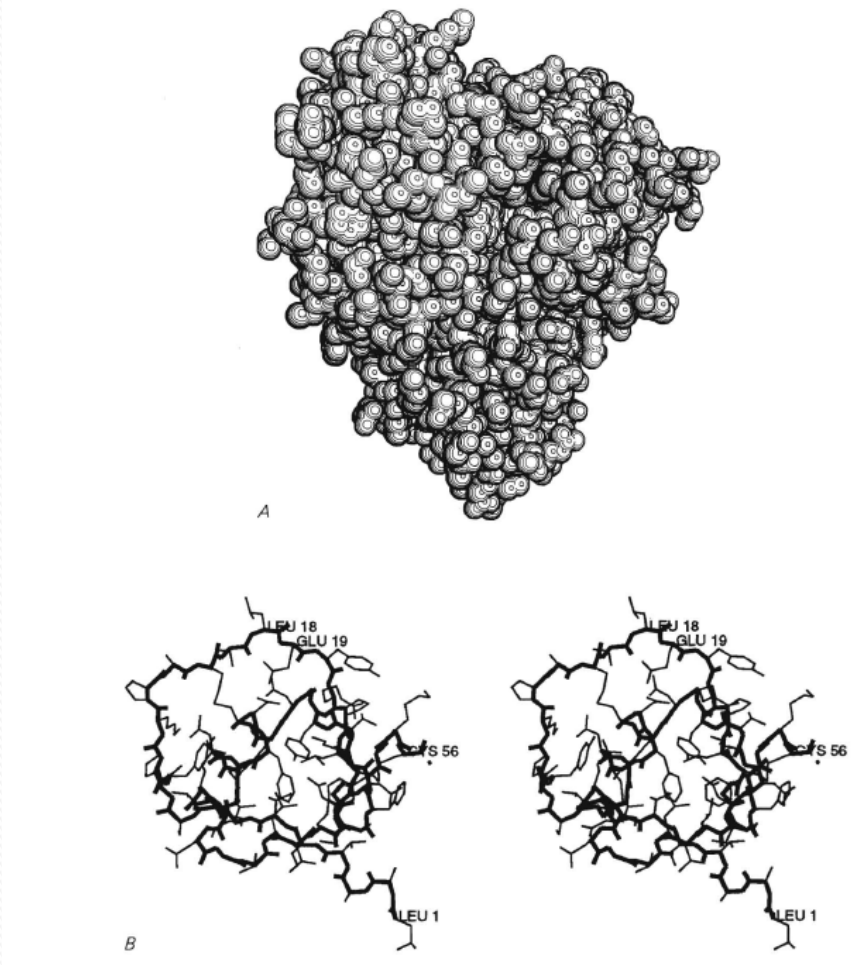
Model refinement



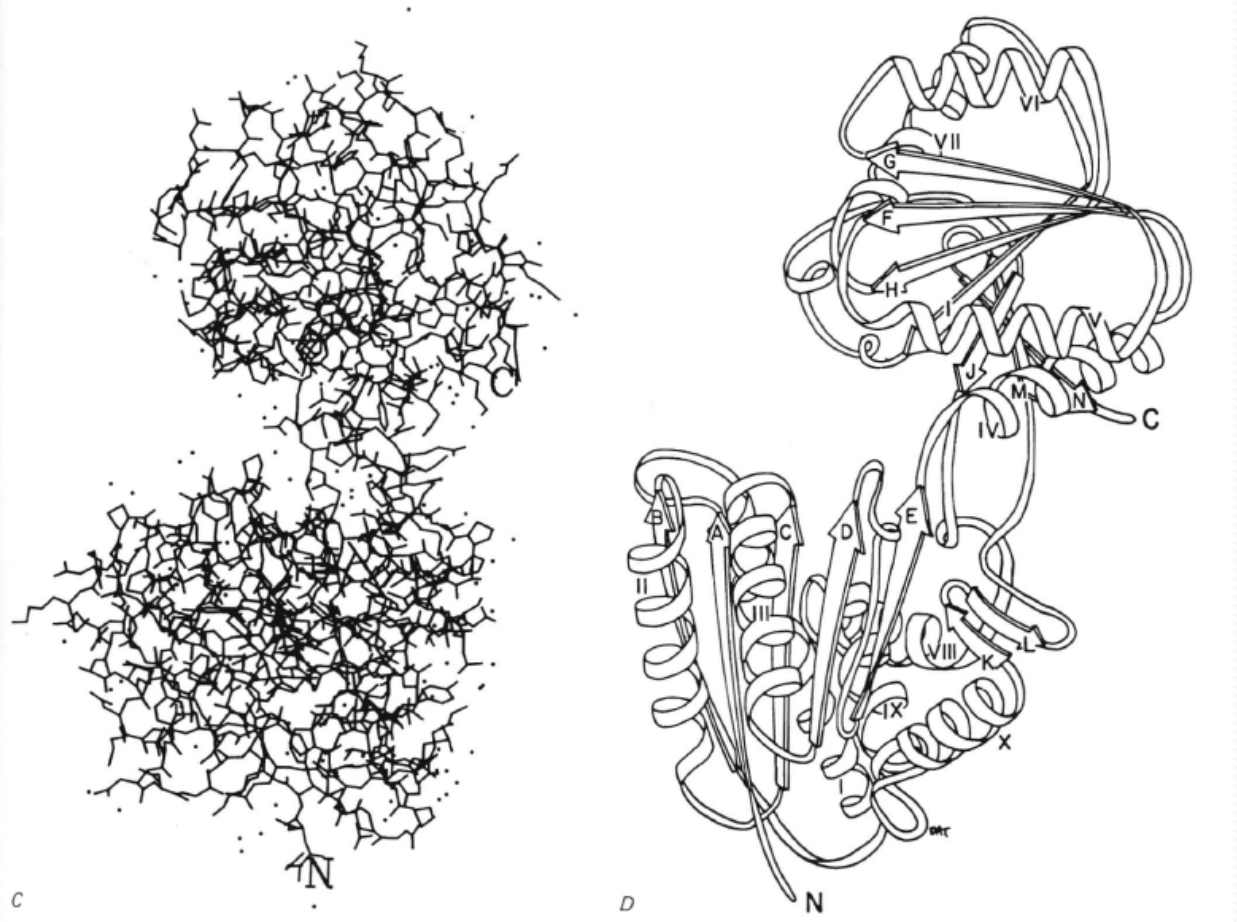
Neutron diffraction



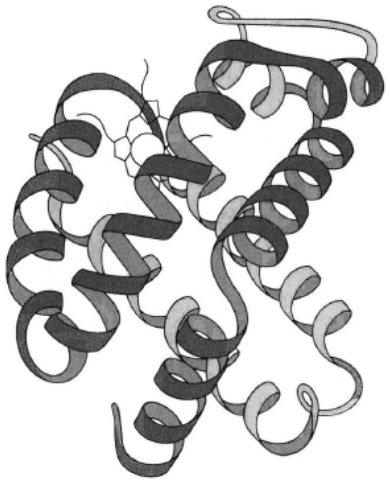
General properties of protein structures



Domain organisation

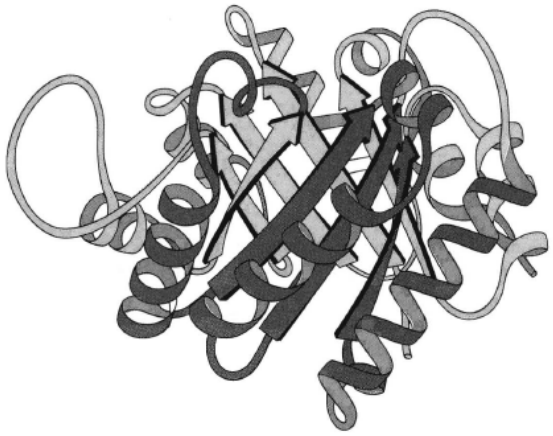


Different folds of proteins

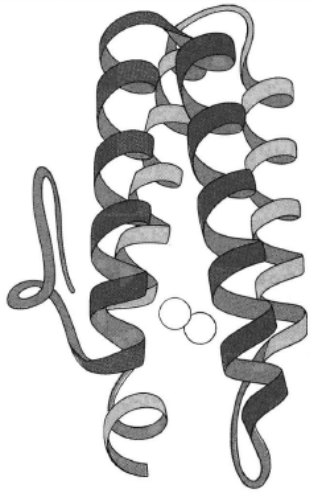


Hemoglobin β subunit

Lactate dehydroge

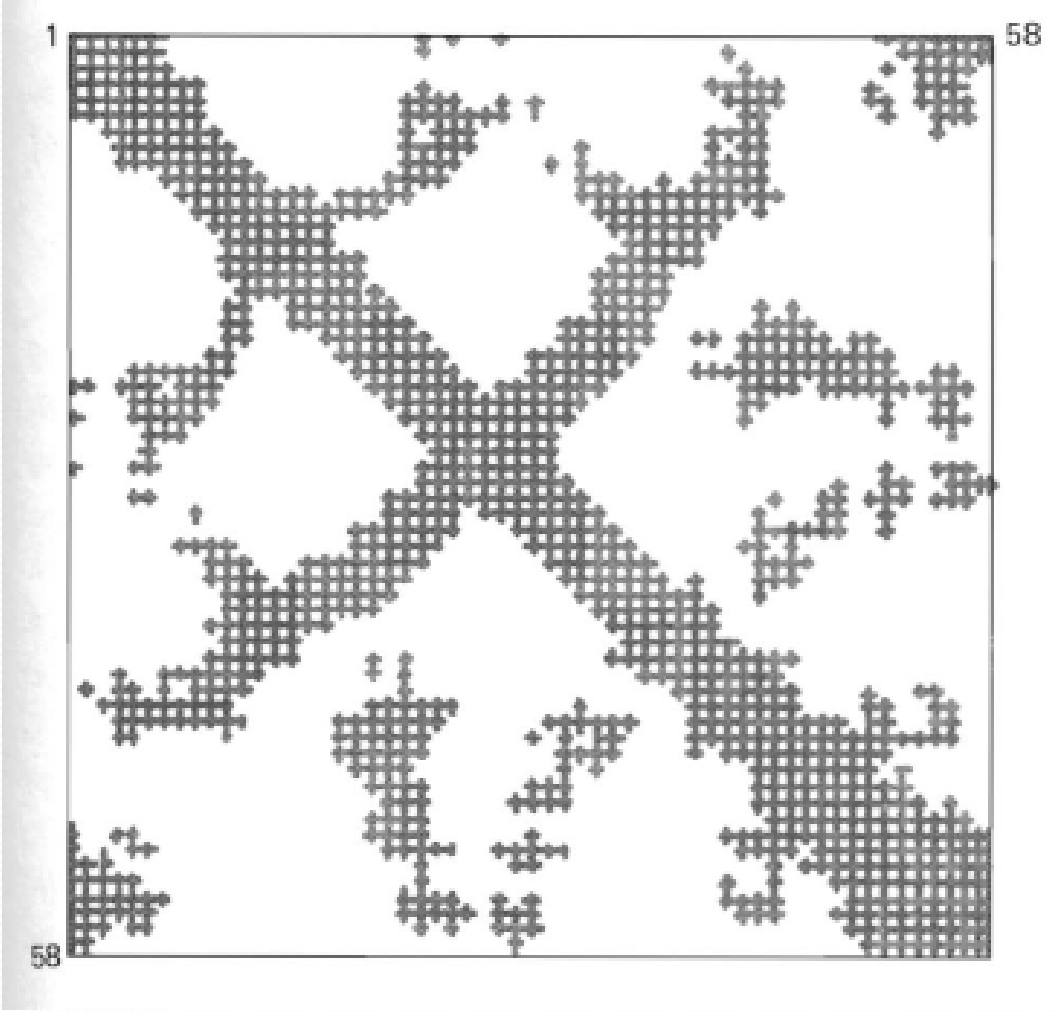


Triose phosphate isomerase

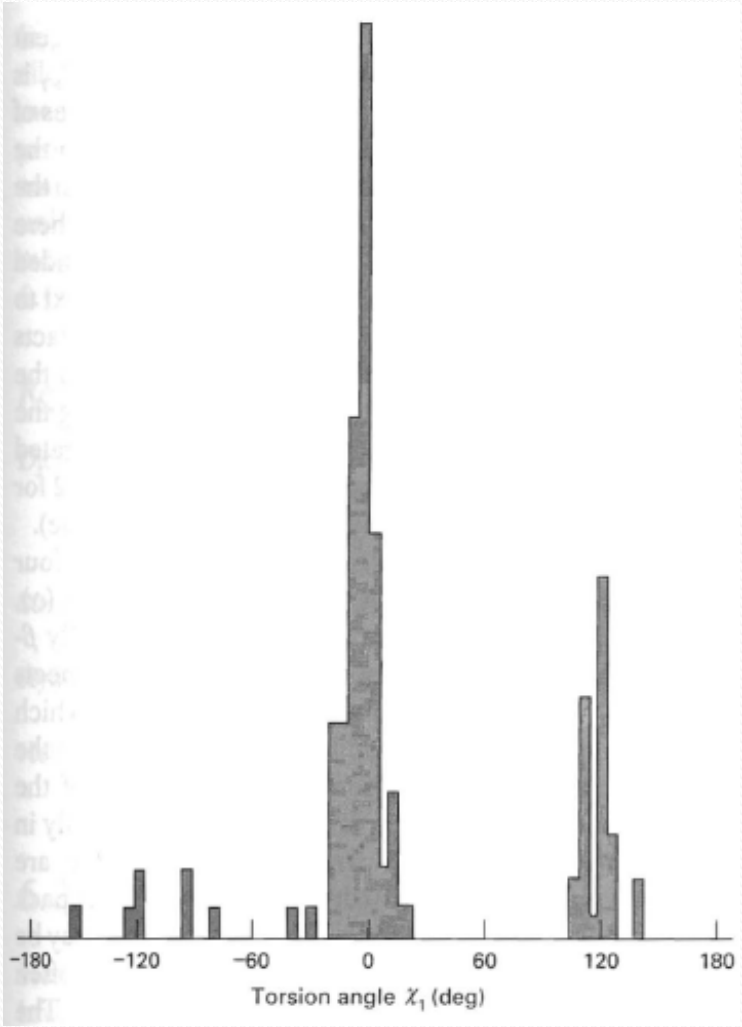


Hemerythrin

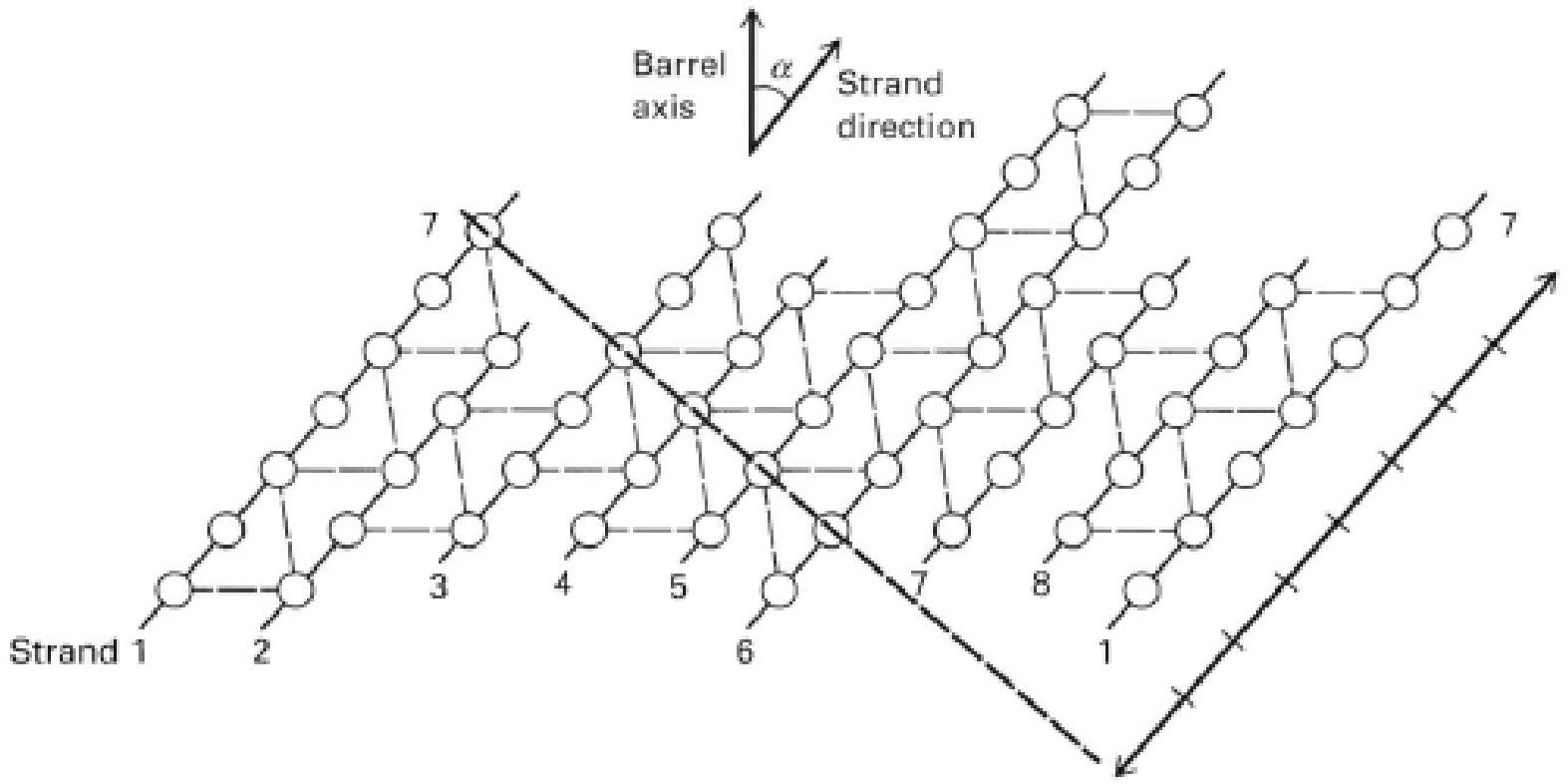
The contact map



Intrinsic rates of Bond rotation



Beta barrel of TIM



Turns

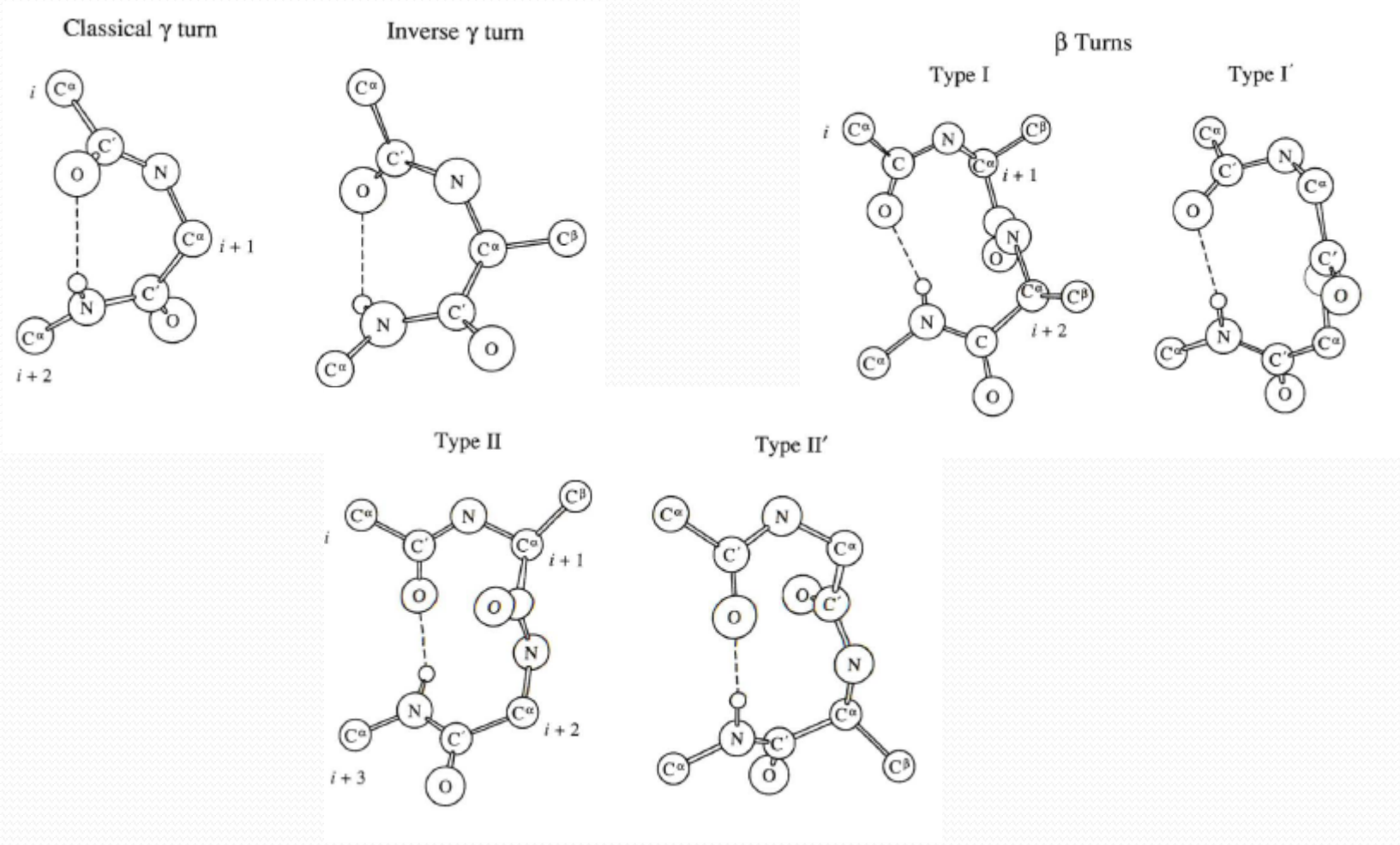

Classical γ turn
Inverse γ turn
β Turns
Type I
Type I′
Type II
Type II′
i
i + 1
i + 2
i + 3

Table 6.2 *Structural Features of γ and β Turns*

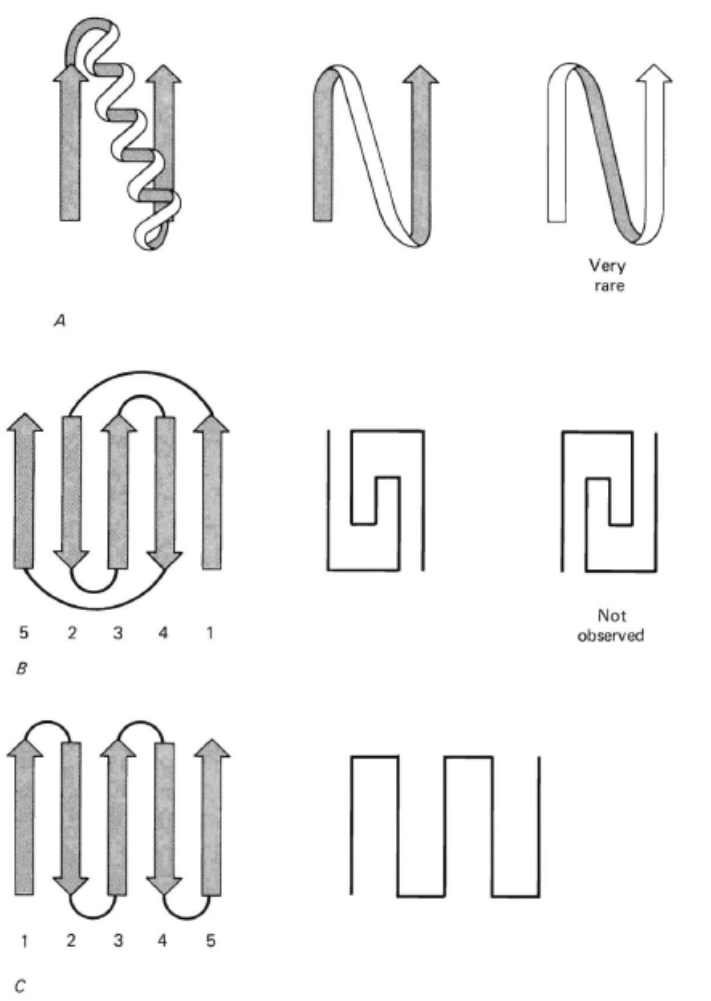
Bend type	Dihedral Angles of Central Residues (deg)[a]			
	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
γ				
classical	70 to 85	−60 to −70		
inverse	−70 to −85	60 to 70		
β				
I	−60	−30	−90	0
I′	60	30	90	0
II	−60	120	80	0
II′	60	−120	−80	0
III	−60	−30	−60	−30
III′	60	30	60	30
IV	Any bend with two or more angles differing by >40° from those given here			
V	−80	80	80	−80
V′	80	−80	−80	80
VIa[b]	−60	120	−90	0
VIb[b]	−120	120	−60	0
VII	Kink in chain created by $\psi_2 \approx 180°$, $\lvert\phi_3\rvert < 60°$; or $\lvert\psi_2\rvert < 60°$, $\phi_3 \approx 180°$			
VIII	−60	−30	−120	120

[a] The central residue of a γ turn is numbered $i + 2$; the two central residues of a β turn are $i + 2$ and $i + 3$.

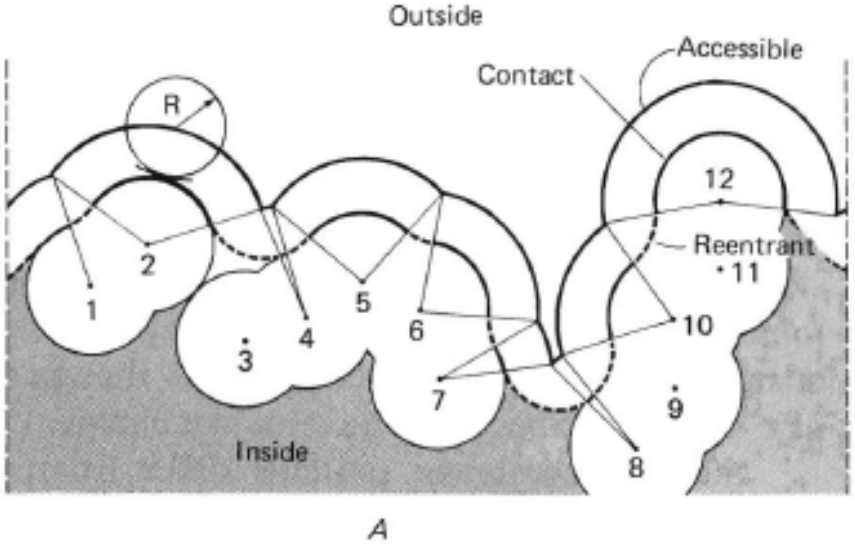
[b] The peptide bond between residues $i + 1$ and $i + 2$ is *cis*, and residue $i + 2$ is Pro.

Data from P. Y. Chou and G. D. Fasman, *J. Mol. Biol.* 115:135–175 (1977); C. M. Wilmot and J. M. Thornton, *J. Mol. Biol.* 203:221–232 (1988); J. S. Richardson, *Adv. Protein Chem.* 34:167–339 (1981).

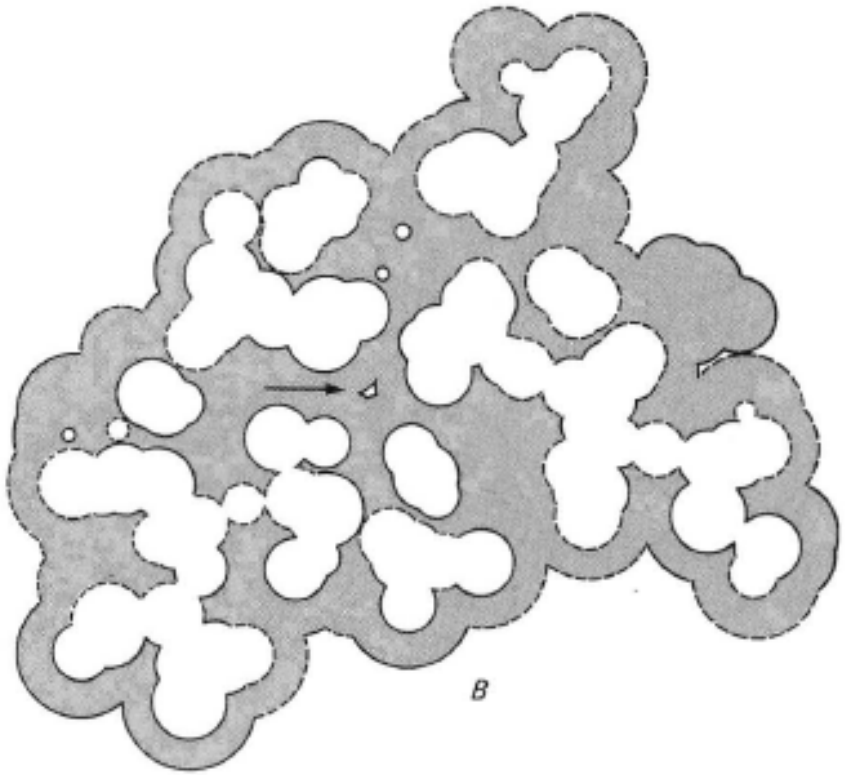
Supersecondary structures



The protein surface

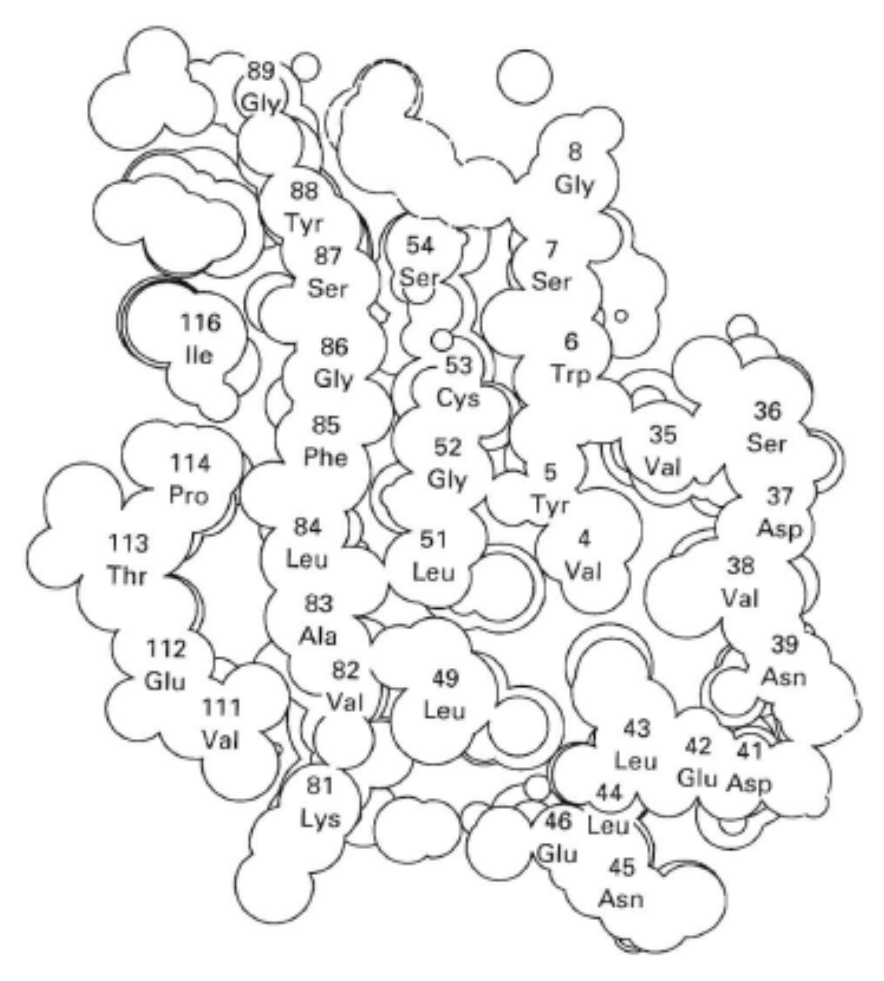


A

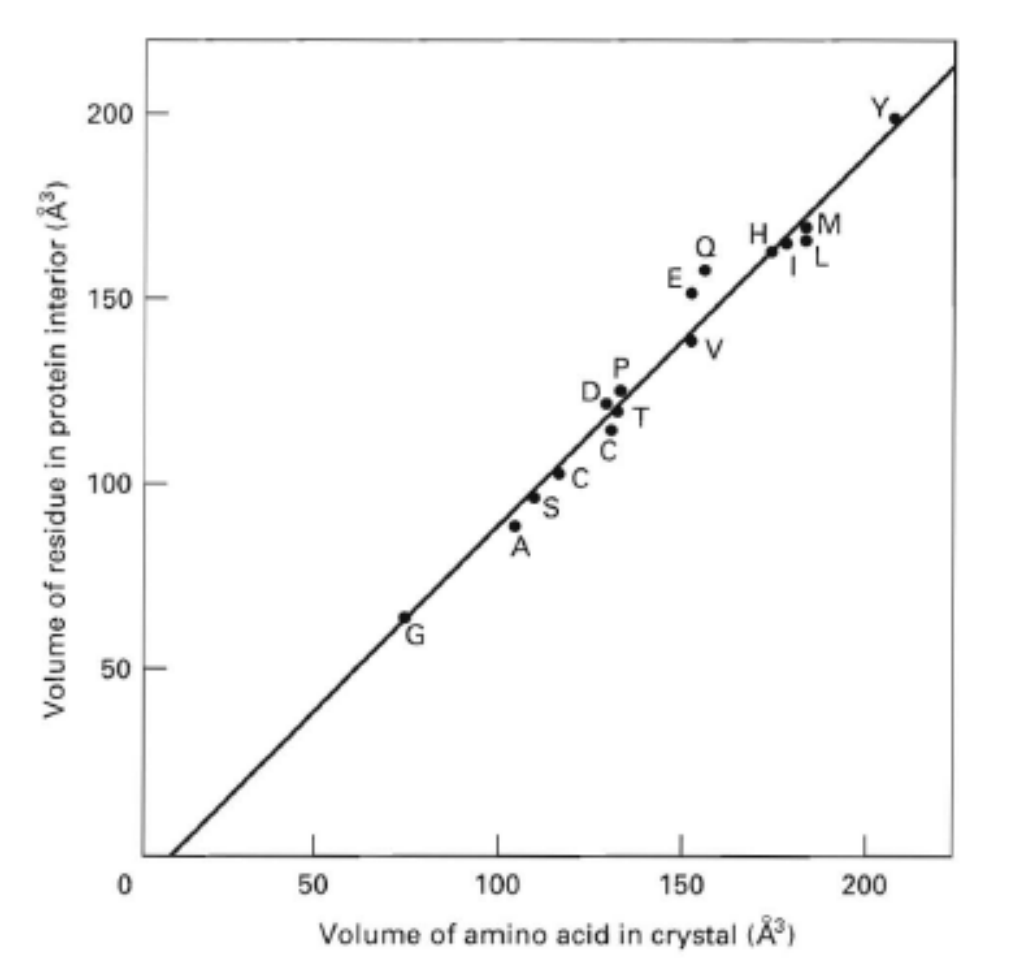


B

Protein packing



Protein packing



Protein packing

Table 6.3 *The Packing of Residues in the Interior of Proteins*

Residue	Average volume of buried residues (Å³)[a]	Fraction of residues at least 95% buried[b]	Relative free energy of residue in interior to that on surface (kcal/mol)[c]
Gly	66	0.36	0
Ala	92	0.38	−0.14
Val	142	0.54	−0.55
Leu	168	0.45	−0.59
Ile	169	0.60	−0.68
Ser	99	0.22	0.40
Thr	122	0.23	0.32
Asp	125	0.15	0.78
Asn	125	0.12	0.75
Glu	155	0.18	1.15
Gln	161	0.07	0.80
Lys	171	0.03	2.06
Arg	202	0.01	1.40
His	167	0.17	0.02
Phe	203	0.50	−0.61
Tyr	204	0.15	0.28
Trp	238	0.27	−0.39
Cys	106[d]	0.40[d]	−0.61[d,e]
	118[e]	0.50[e]	—
Met	171	0.40	−0.65
Pro	129	0.18	0.50

[a] From C. Chothia, *Nature* 254:304–308 (1975).

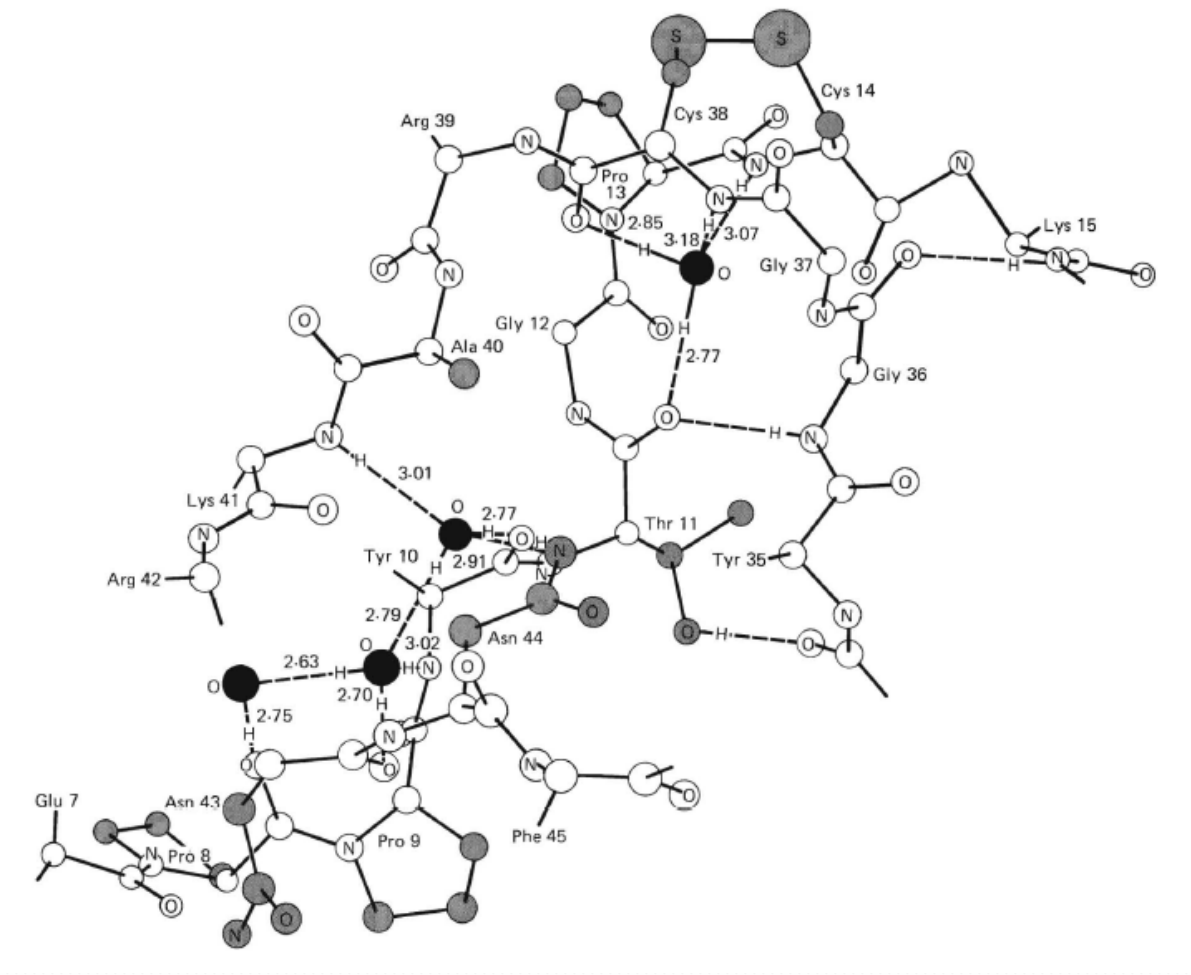
[b] Average for 12 proteins. From C. Chothia, *J. Mol. Biol.* 105:1–14 (1976).

[c] Calculated as $-RT \log_e f$, where f is the ratio of the occurence of this amino acid residue on the interior to that on the surface. The values were normalized with that for Gly set to zero. From S. Miller et al., *J. Mol. Biol.* 105:641–656 (1987).

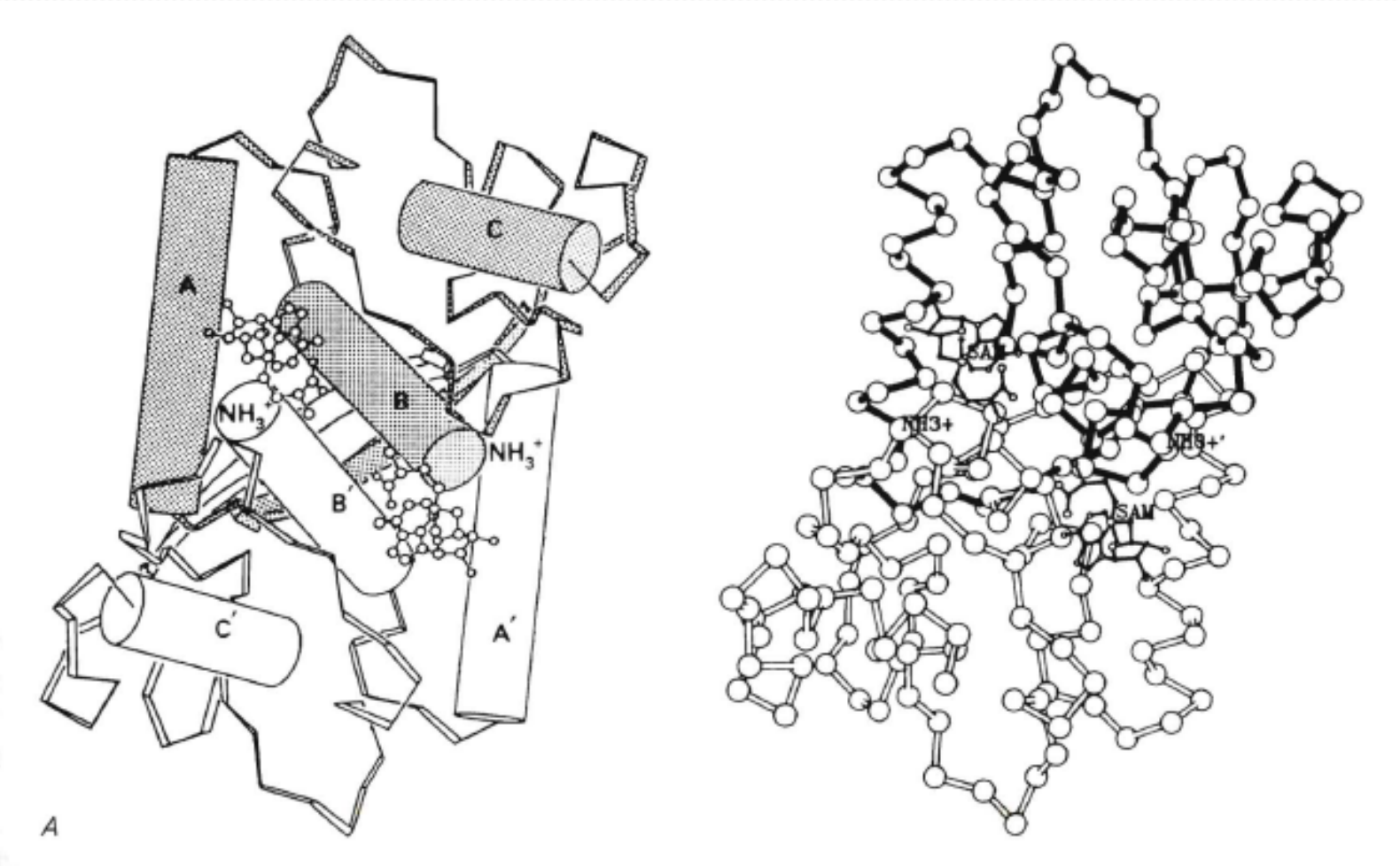
[d] When in disulfide form.

[e] When in thiol form.

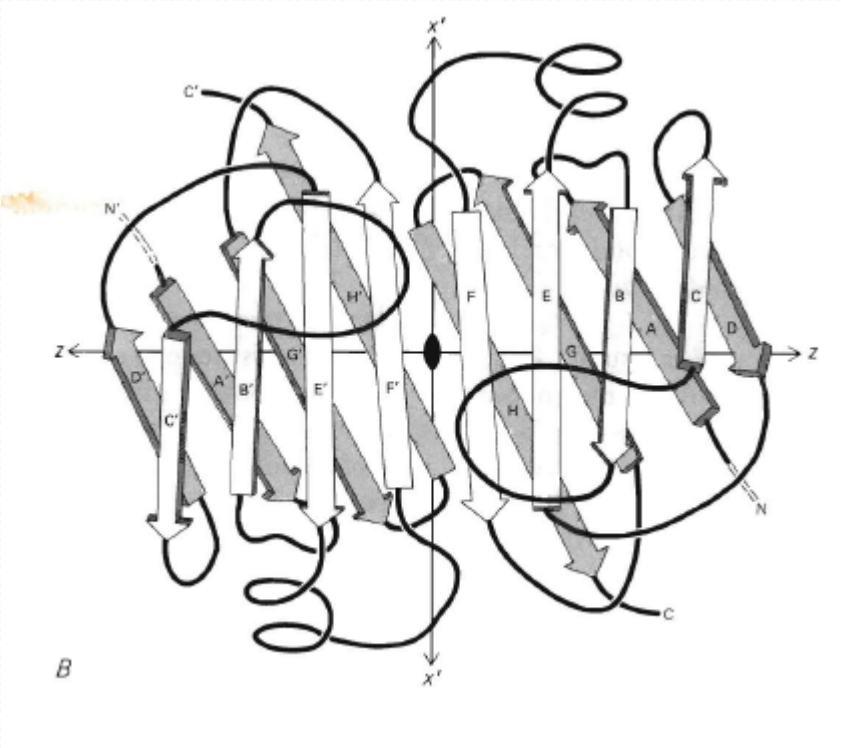
Internal water molecules

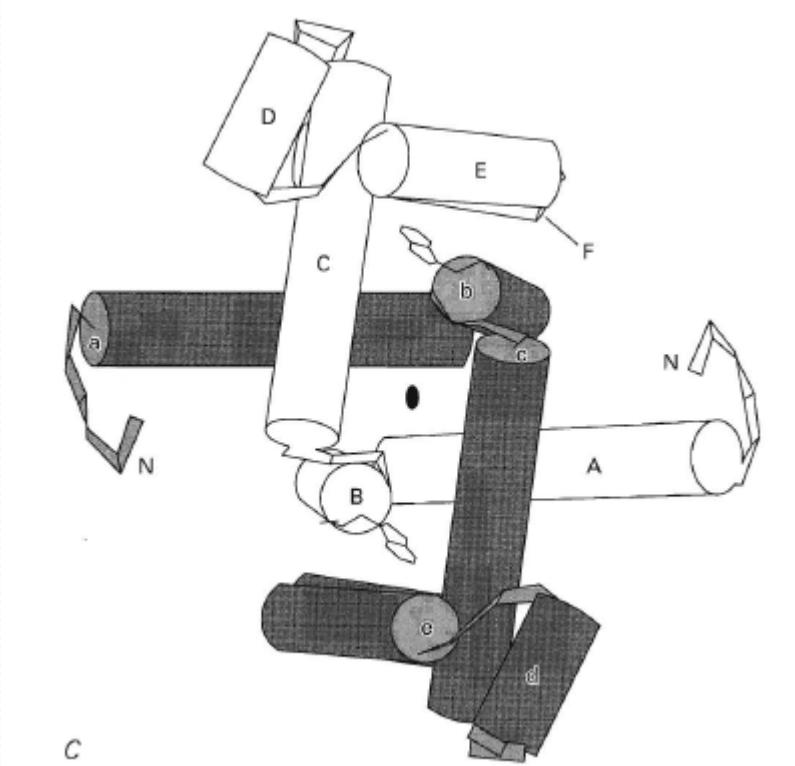


Quaternary structure

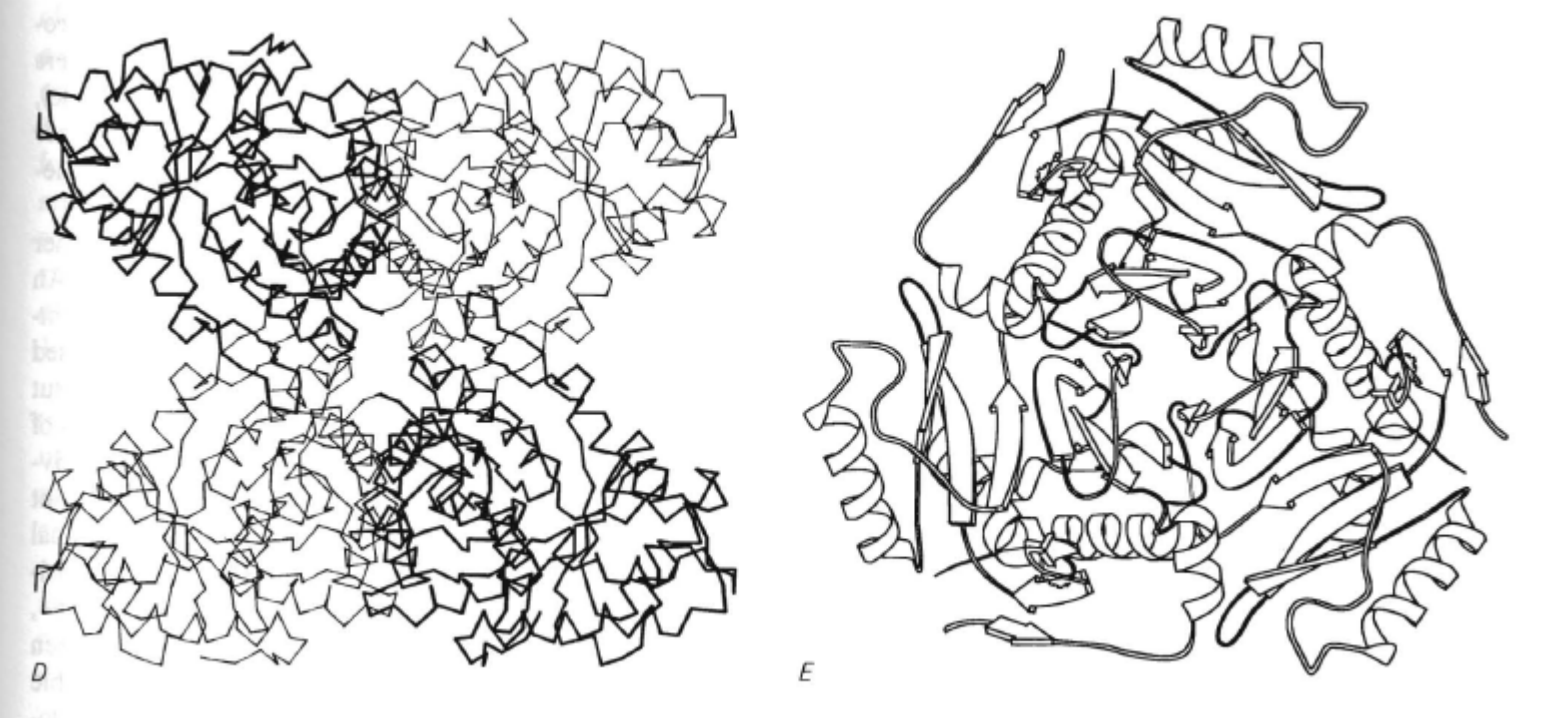


Quarternary Structure

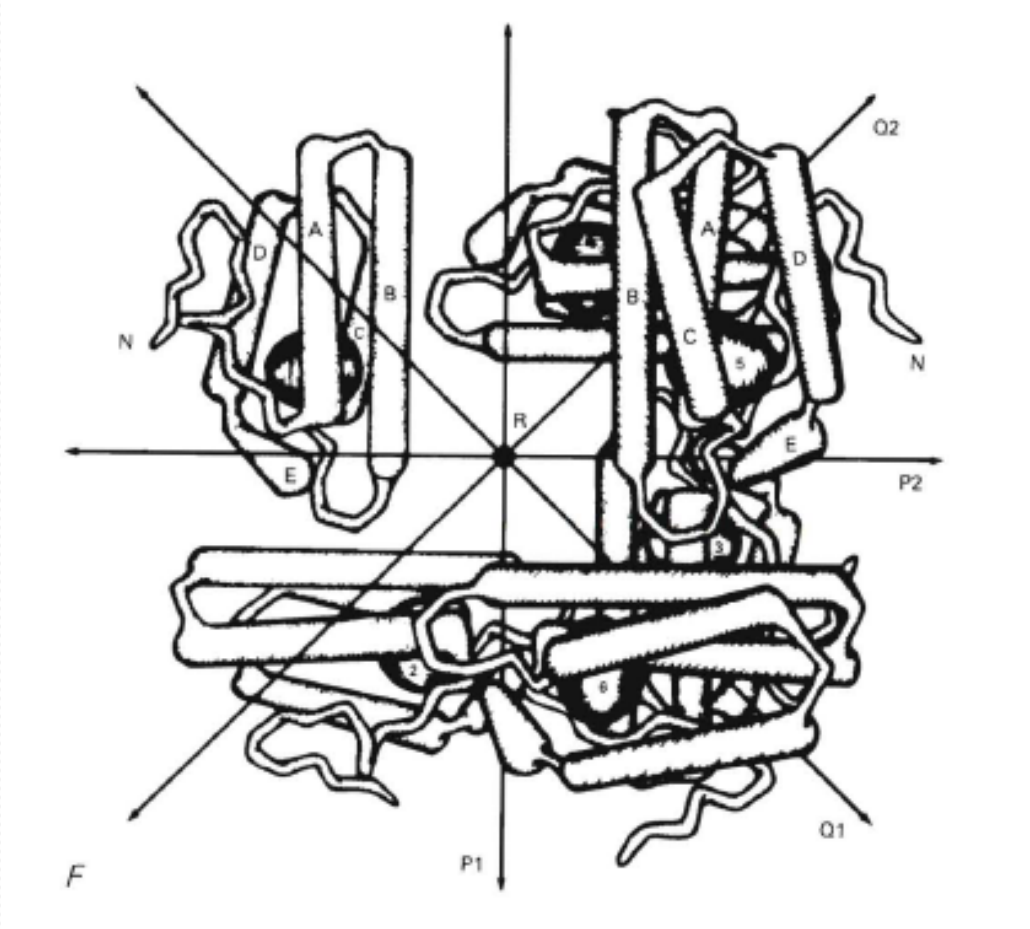




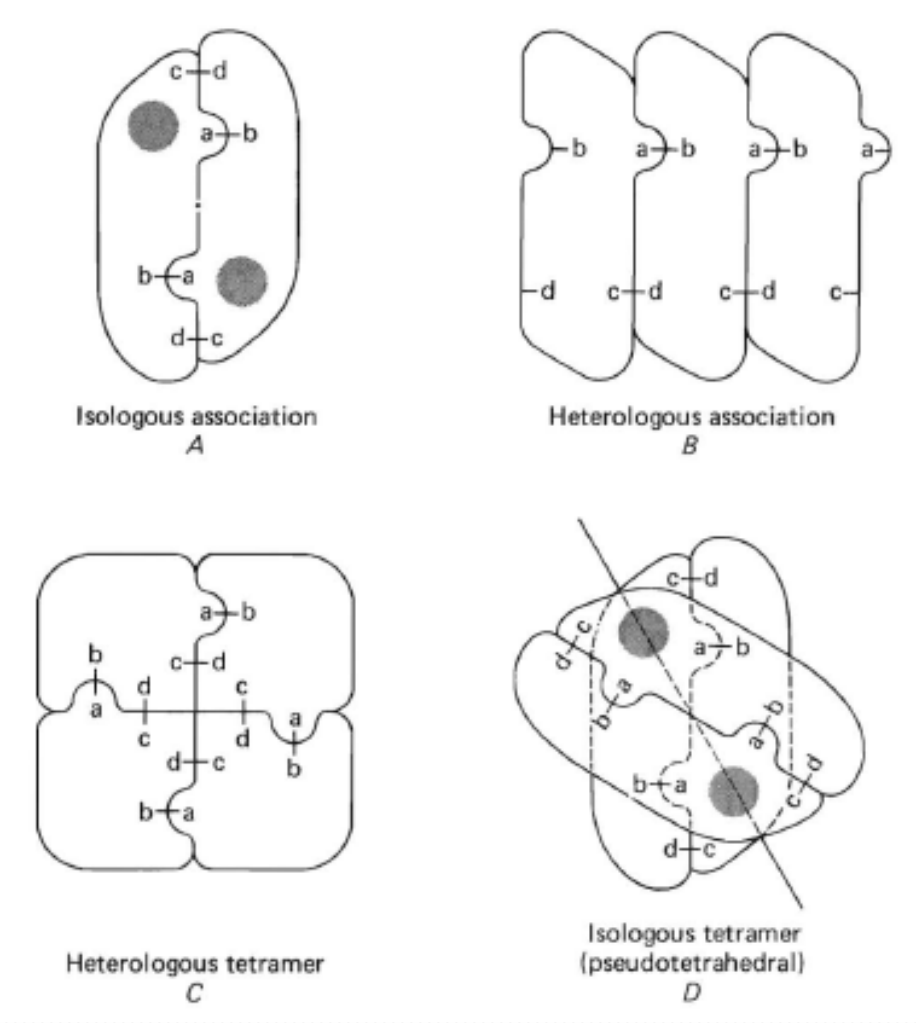
Isologous tetrameric structures



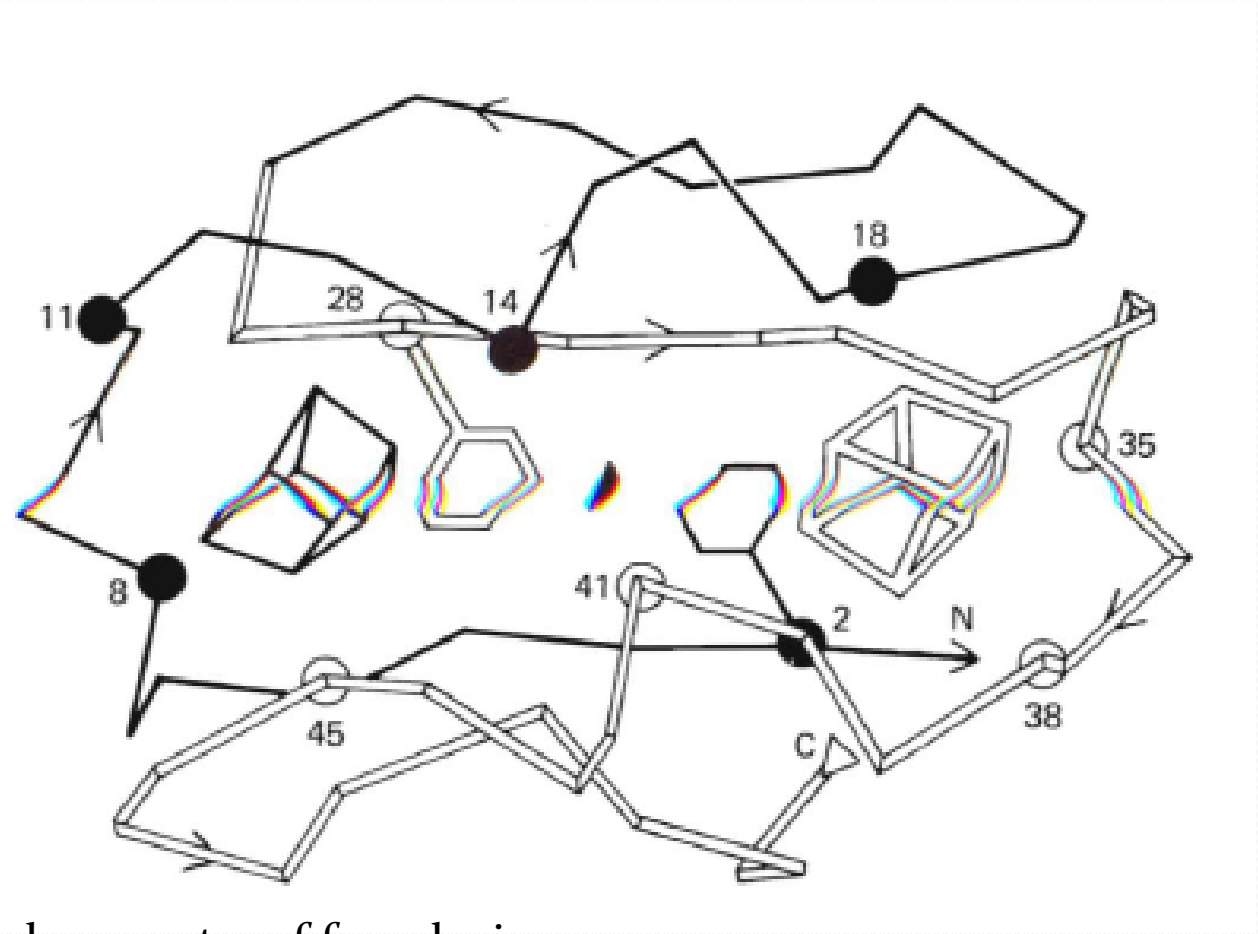
Isologous tetrameric structures



Isologous and heterologous structure organisation

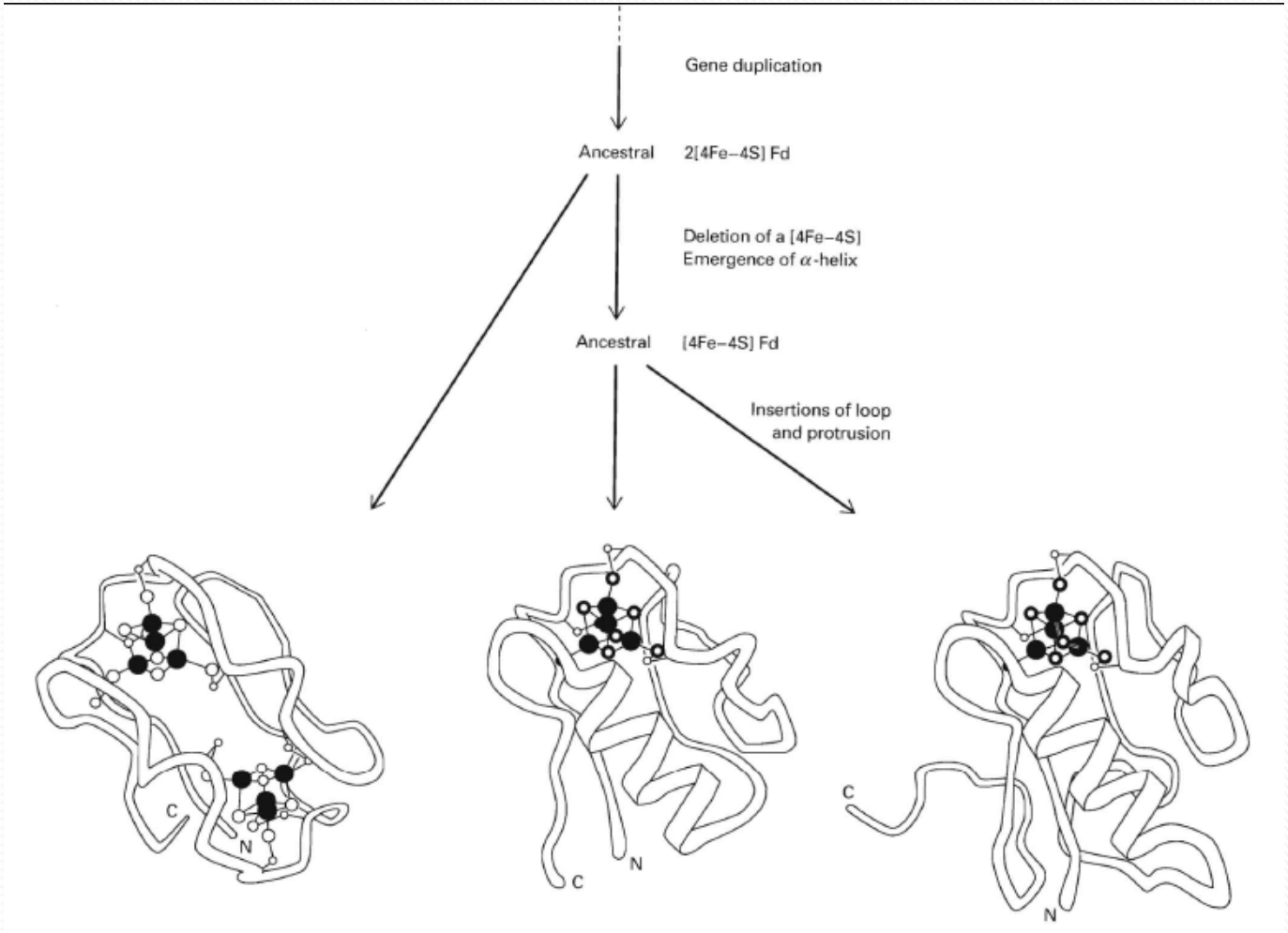


Structural Hommology within a polypeptide chain

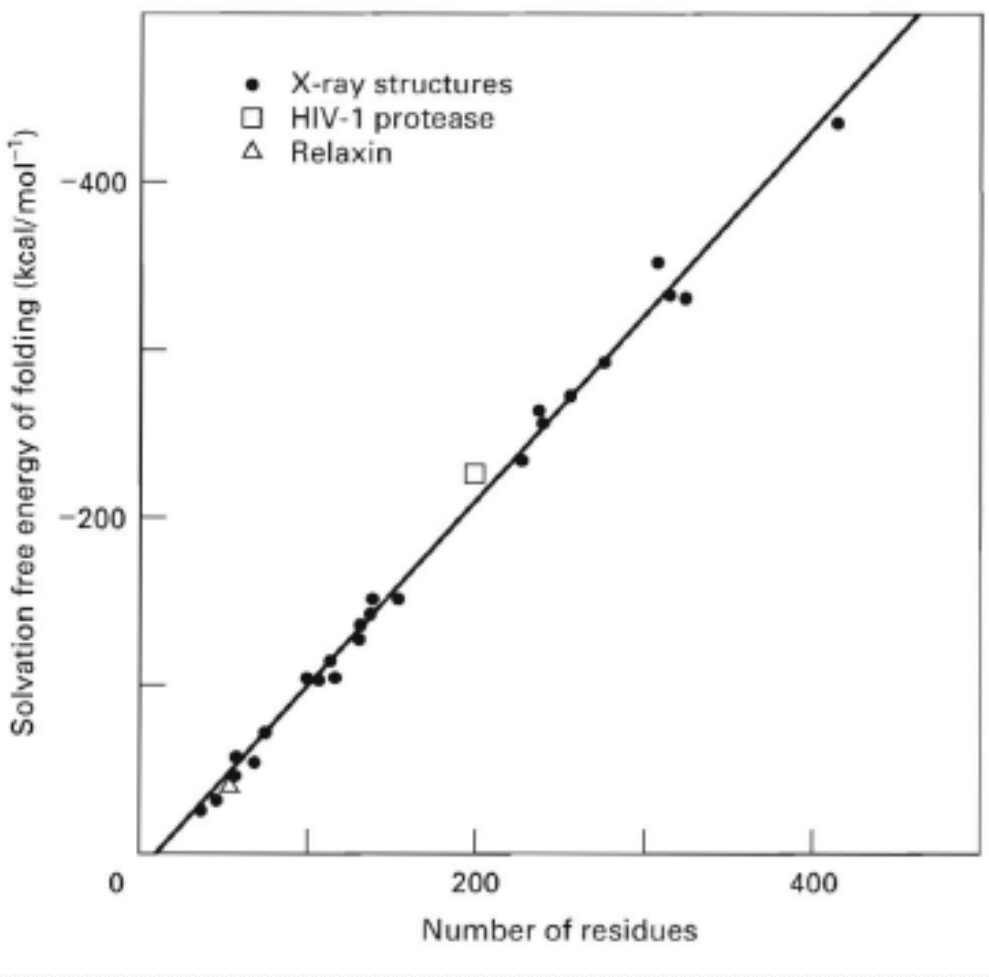


Internal symmetry of ferredoxin

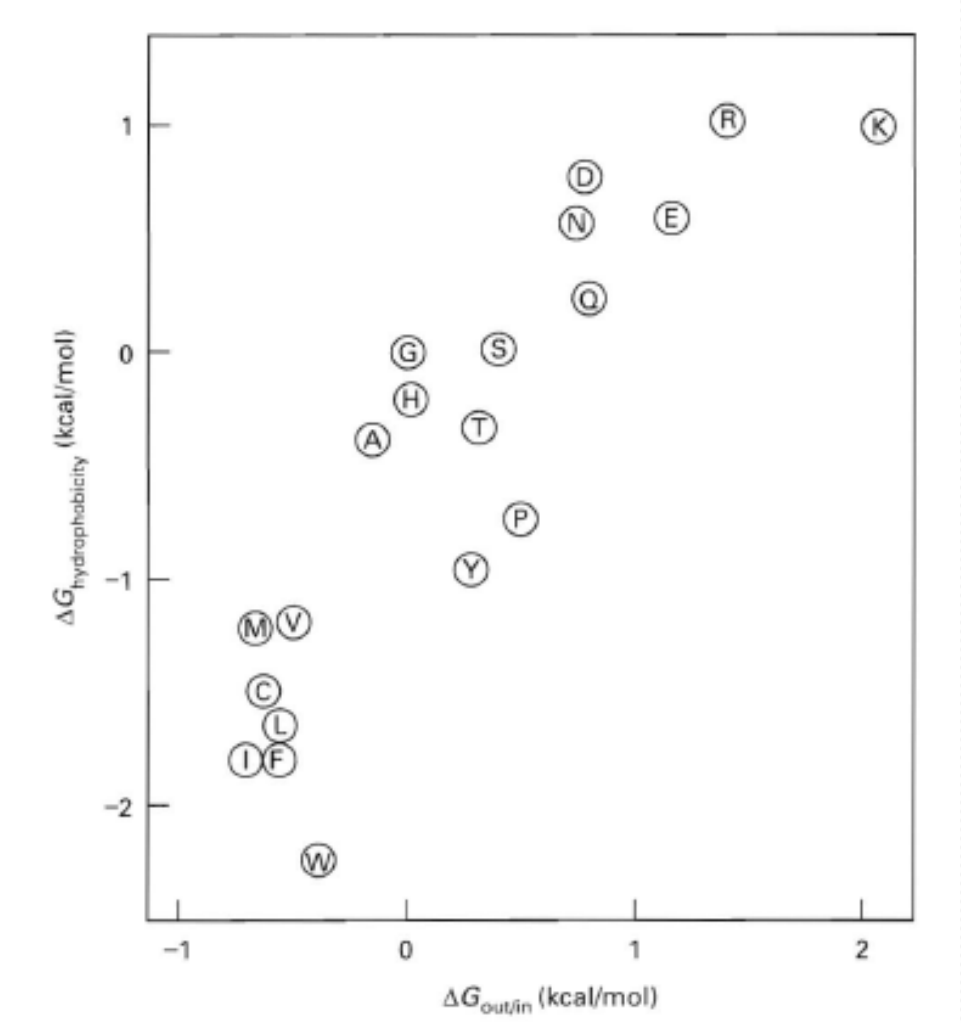
Structural Hommology within a polypeptide chain



Rationalisation and prediction of protein structure



Hydrophobicity vs free energy of transfer



Conformational preferences of AA

Table 6.5 *Conformational Preferences of the Amino Acids*

Amino acid residue	Preference[a]			α-Helix Preference[b]			Turn Preference		
	α-helix (P_α)	β-strand (P_β)	Reverse turn (P_t)	N-term	Middle	C-term	Type I	Type II	Other
Glu	**1.59**	0.52	1.01	2.12	1.18	1.21	1.12	0.84	1.06
Ala	**1.41**	0.72	0.82	1.33	1.60	1.46	0.74	0.94	0.58
Leu	**1.34**	1.22	0.57	1.03	1.50	1.46	0.61	0.53	0.75
Met	**1.30**	1.14	0.52	0.75	1.44	1.92	0.66	0.73	0.96
Gln	**1.27**	0.98	0.84	1.39	1.22	1.24	0.79	1.45	1.02
Lys	**1.23**	0.69	1.07	0.98	1.05	1.68	0.70	0.73	1.04
Arg	**1.21**	0.84	0.90	1.26	1.25	1.23	0.88	1.22	0.84
His	**1.05**	0.80	0.81	0.68	0.97	1.57	0.78	0.64	1.00
Val	0.90	**1.87**	0.41	1.00	1.09	1.08	0.39	0.61	0.48
Ile	1.09	**1.67**	0.47	0.96	1.31	0.99	0.39	0.43	0.93
Tyr	0.74	**1.45**	0.76	0.63	0.61	1.00	0.71	0.91	0.97
Cys	0.66	**1.40**	0.54	0.78	0.66	0.56	1.38	0.99	0.78
Trp	1.02	**1.35**	0.65	1.20	1.34	0.78	1.35	0.15	0.52
Phe	1.16	**1.33**	0.59	0.94	1.45	1.20	0.77	0.76	0.53
Thr	0.76	**1.17**	0.90	0.75	0.87	0.80	1.25	0.67	0.93
Gly	0.43	0.58	**1.77**	0.60	0.47	0.31	1.14	2.61	1.38
Asn	0.76	0.48	**1.34**	0.80	0.80	0.75	1.79	0.99	1.37
Pro	0.34	0.31	**1.32**	0.90	0.19	0.06	0.95	1.80	1.51
Ser	0.57	0.96	**1.22**	0.67	0.44	0.73	1.47	0.76	1.49
Asp	0.99	0.39	**1.24**	1.35	1.03	0.67	1.98	0.71	1.28

[a] The normalized frequences for each conformation (e.g., P_α, P_β, P_t) were calculated from the fraction of residues of each amino acid that occurred in that conformation, divided by this fraction for all residues. Random occurrence of a partricular amino acid in a conformation would give a value of unity.

[b] *N-term* and *C-term* include the four helical residues at the ends of a helical segment eight or more residues long, and three residues at the ends of segments six or seven residues long. *Middle* includes all helical residues between N-term and C-term.

From R. W. Williams et al., *Biochim Biophys. Acta.* 916:200–204 (1987); C. M. Wilmot and J. M. Thornton, *J. Mol. Biol.* 203:221–232 (1988).

Modeling homologous Protein structures