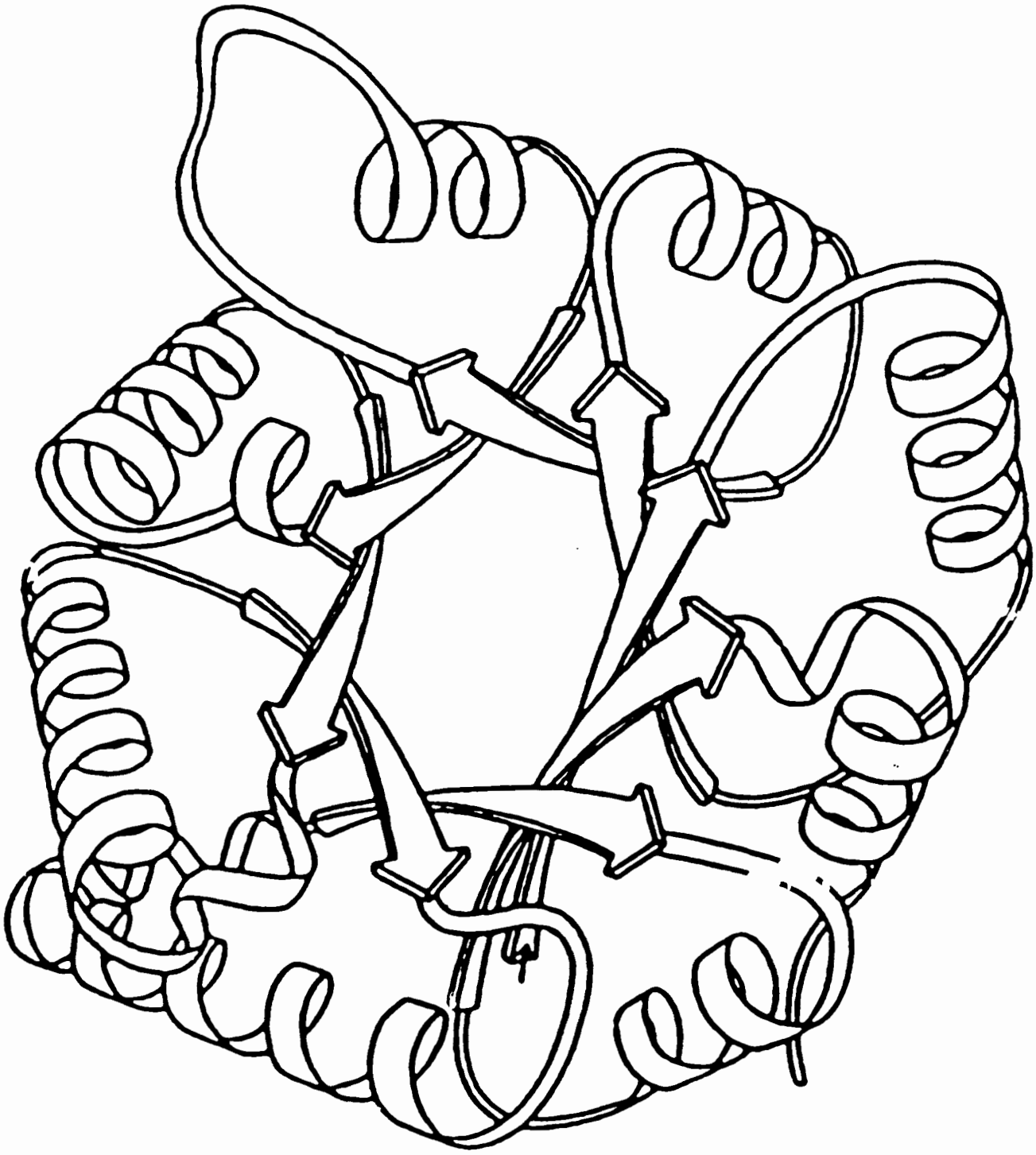


Triose phosphate isomerase. Reproduced from Richardson (1981).

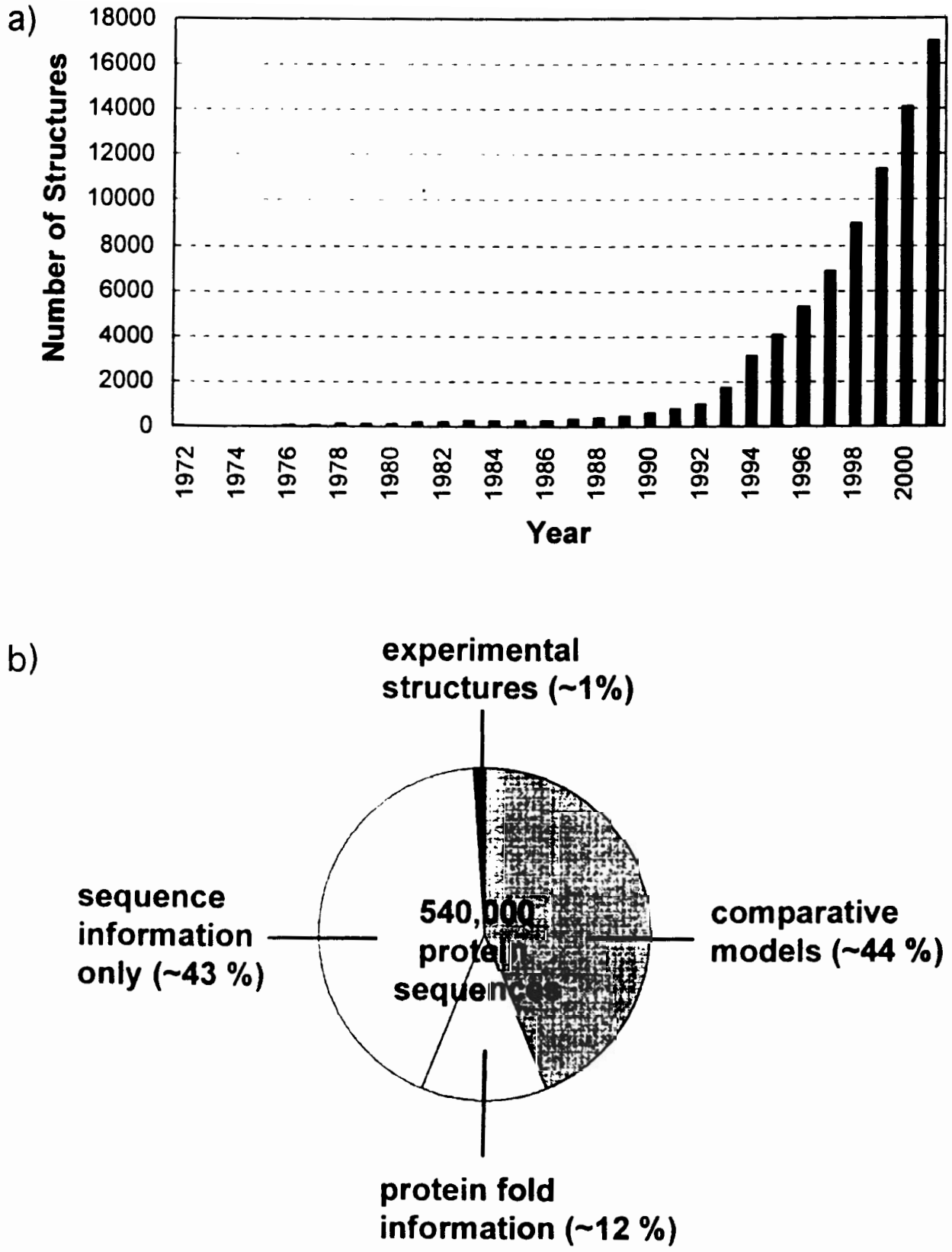


Figure 2. a) Growth in the number of experimental biopolymer structures deposited in the Protein Data Bank (PDB). By early 2002, about 17,000 structures of biopolymers were publically accessible in the PDB. b) Structure information on 540,000 protein sequences in the Swiss-Prot and TrEMBL databases. For about 1% of these sequences (~6200 different proteins), experimental structure information (x-ray, NMR) is available. Comparative models can be built for up to 44% (237,000) and protein fold information can be obtained for an additional ~12% of the sequences. The remaining 43% of the proteins actually cannot be assigned any fold or other structure information yet.

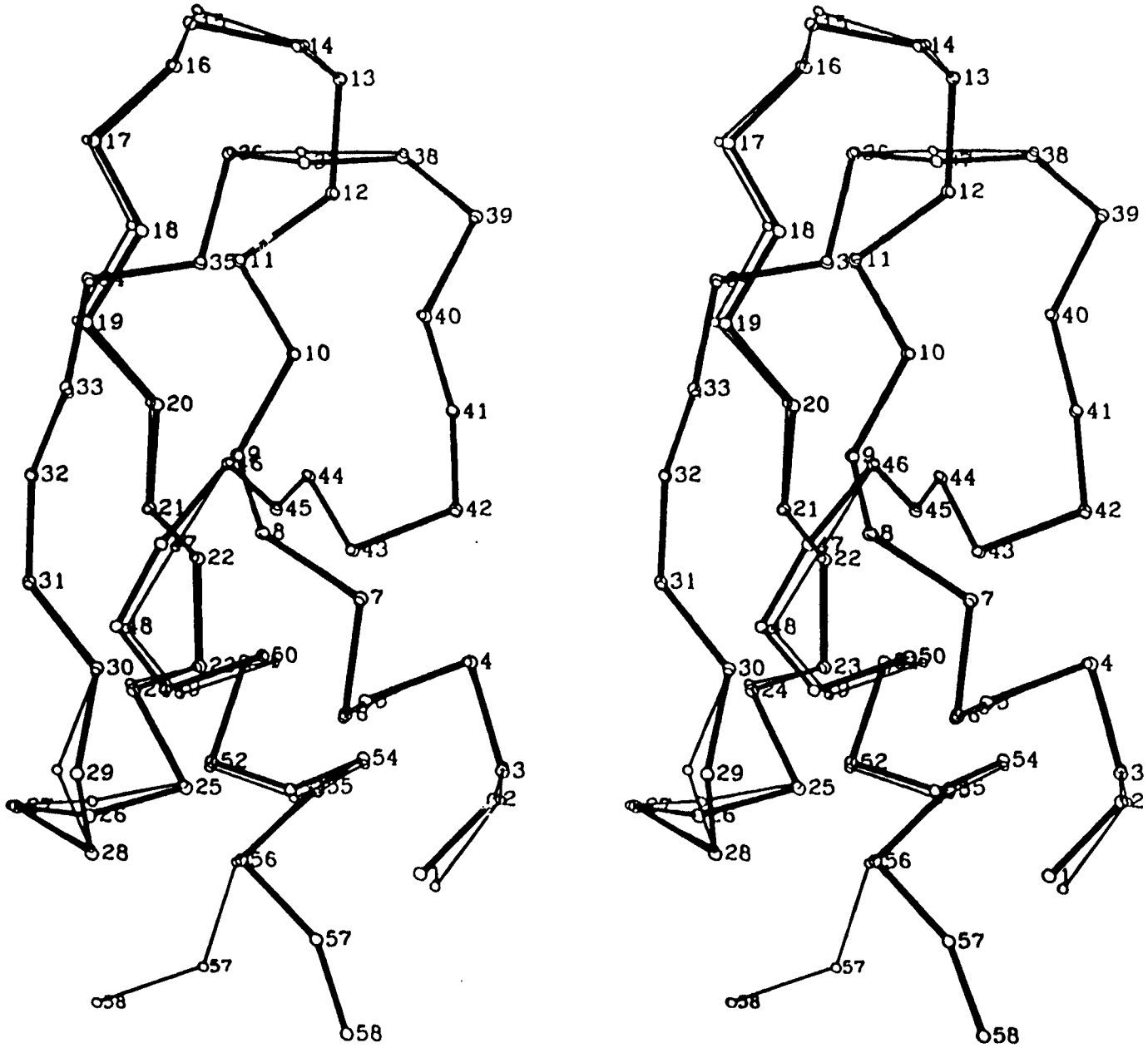
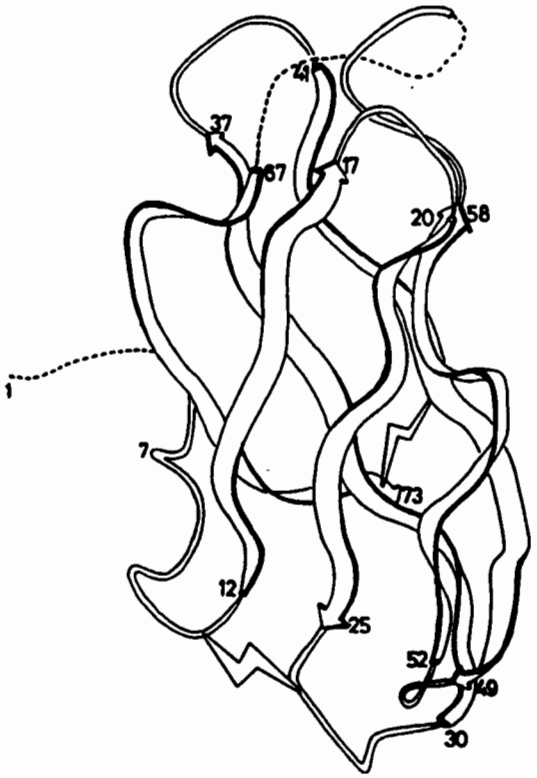
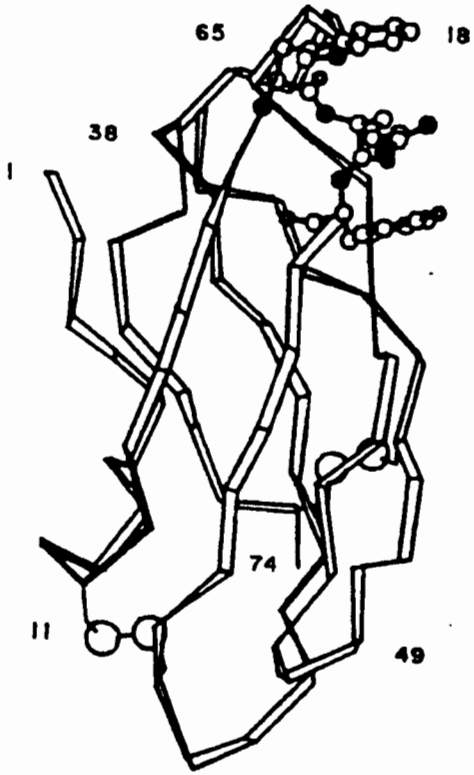


Figure 1. Stereo superposition of the positions of C^{α} backbone atoms in crystal forms I (thick lines) and II (thin lines)

AMYLASE INHIBITOR STRUCTURES



NMR



X-RAY

Table 3.1. *Typical features of some internal motions of proteins and nucleic acids*

Motion	Spatial extent (nm)	Amplitude (nm)	Log_{10} of characteristic time (s)
Relative vibration of bonded atoms	0.2 to 0.5	0.001 to 0.01	−14 to −13
Longitudinal motions of bases in double helices (nucleic acids)	0.5	0.01	−14 to −13
Lateral motions of bases in double helices (nucleic acids)	0.5	0.1	−13 to −12
Global stretching (nucleic acids)	1 to 30	0.03 to 0.3	−13 to −11
Global twisting (nucleic acids)	1 to 30	0.1 to 1.0	−13 to −11
Elastic vibration of globular region	1 to 2	0.005 to 0.05	−12 to −11
Sugar repuckering (nucleic acids)	0.5	0.2	−12 to −9
Rotation of sidechains at surface (protein)	0.5 to 1	0.5 to 1	−11 to −10
Torsional libration of buried groups	0.5 to 1	0.05	−11 to −9
Relative motion of different globular regions (hinge bending)	1 to 2	0.1 to 0.5	−11 to −7
Global bending (nucleic acids)	10 to 100	5 to 20	−10 to −7
Rotation of medium-sized sidechains in interior (protein)	0.5	0.5	−4 to 0
Allosteric transitions	0.5 to 4	0.1 to 0.5	−5 to 0
Local denaturation	0.5 to 1	0.5 to 1	−5 to +1

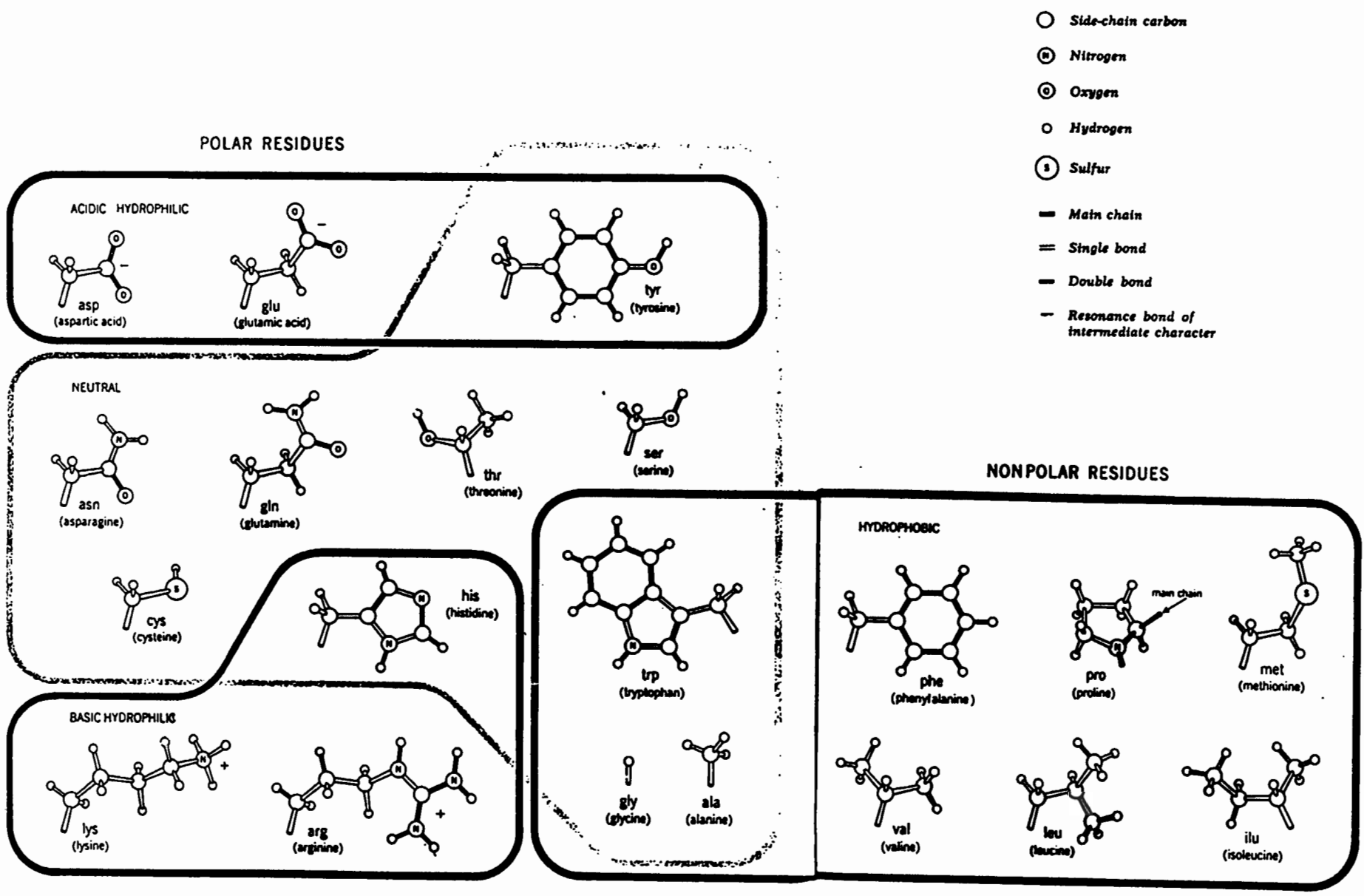

POLAR RESIDUES
ACIDIC HYDROPHILIC
asp
(aspartic acid)
glu
(glutamic acid)
tyr
(tyrosine)
NEUTRAL
asn
(asparagine)
gln
(glutamine)
thr
(threonine)
ser
(serine)
cys
(cysteine)
his
(histidine)
BASIC HYDROPHILIC
lys
(lysine)
arg
(arginine)
NONPOLAR RESIDUES
HYDROPHOBIC
trp
(tryptophan)
gly
(glycine)
ala
(alanine)
phe
(phenylalanine)
main chain
pro
(proline)
met
(methionine)
val
(valine)
leu
(leucine)
ilu
(isoleucine)
Side-chain carbon
Nitrogen
Oxygen
Hydrogen
Sulfur
Main chain
Single bond
Double bond
Resonance bond of intermediate character

Table 1.1 Properties of Individual Amino Acid Residues

Residue	Residue mass[a] (daltons)	Van der Waals volume[b] (Å^3)	Frequency in proteins[c] (%)
Ala (A)	71.09	67	8.3
Arg (R)	156.19	148	5.7
Asn (N)	114.11	96	4.4
Asp (D)	115.09	91	5.3
Cys (C)	103.15	86	1.7
Gln (Q)	128.14	114	4.0
Glu (E)	129.12	109	6.2
Gly (G)	57.05	48	7.2
His (H)	137.14	118	2.2
Ile (I)	113.16	124	5.2
Leu (L)	113.16	124	9.0
Lys (K)	128.17	135	5.7
Met (M)	131.19	124	2.4
Phe (F)	147.18	135	3.9
Pro (P)	97.12	90	5.1
Ser (S)	87.08	73	6.9
Thr (T)	101.11	93	5.8
Trp (W)	186.21	163	1.3
Tyr (Y)	163.18	141	3.2
Val (V)	99.14	105	6.6
Weighted average[d]	119.40	161	

Table 1.2 Intrinsic pK_a Values of Ionizable Groups Found in Proteins

Group	Observed pK_a[a]
α-Amino	6.8–8.0
α-Carboxyl	3.5–4.3
β-Carboxyl (Asp)	3.9–4.0
γ-Carboxyl (Glu)	4.3–4.5
δ-Guanido (Arg)	12.0
ϵ-Amino (Lys)	10.4–11.1
Imidazole (His)	6.0–7.0
Thiol (Cys)	9.0–9.5
Phenolic hydroxyl (Tyr)	10.0–10.3

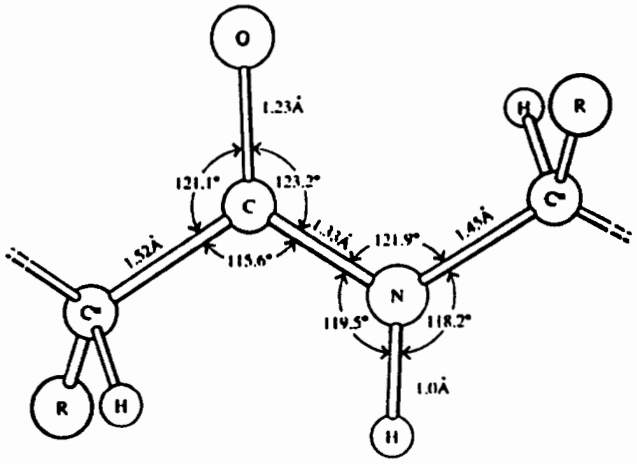
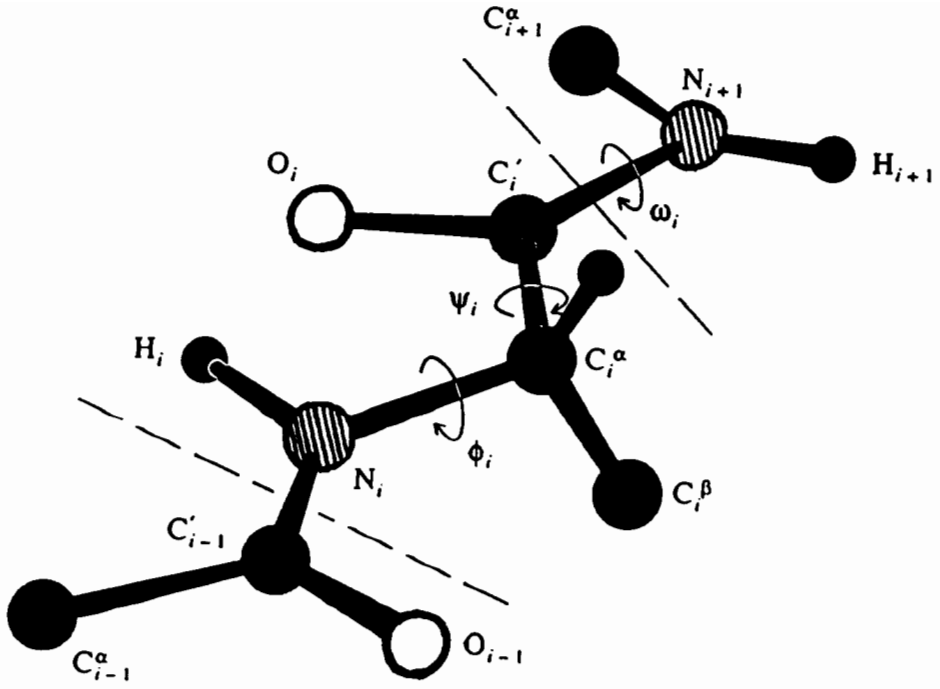
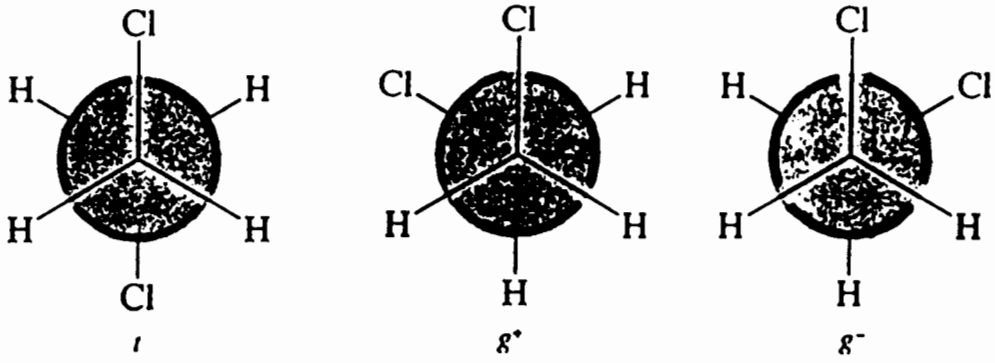


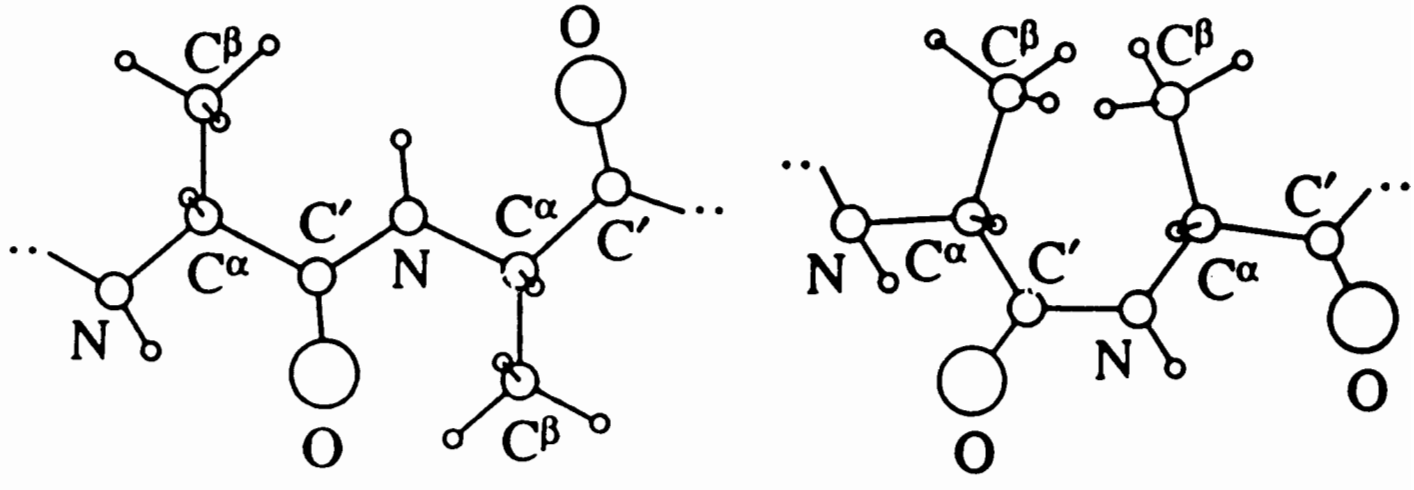
Table 1.3 Spectroscopic Properties of the Aromatic Amino Acids at Neutral pH

	Absorbance[a]		Fluorescence Emission[b]	
	λ_{max} (nm)	Molar absorbance ($M^{-1}cm^{-1}$)	λ_{max} (nm)	Quantum yield
Phenylalanine	257.4	197	282	0.04
Tyrosine	274.6	1420	303	0.21
Tryptophan	279.8	5600	348	0.20

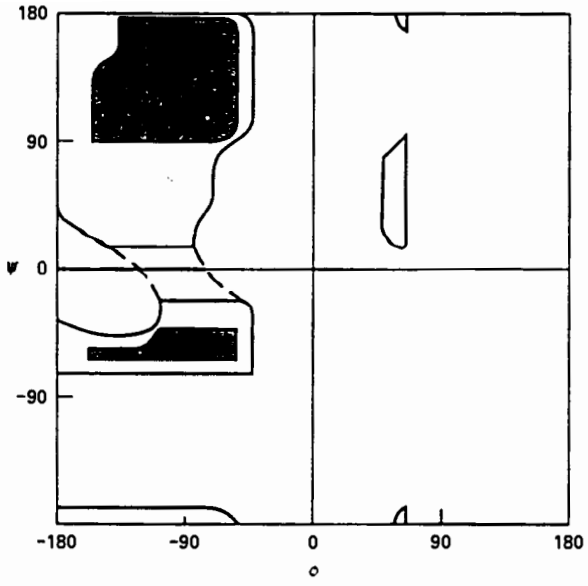


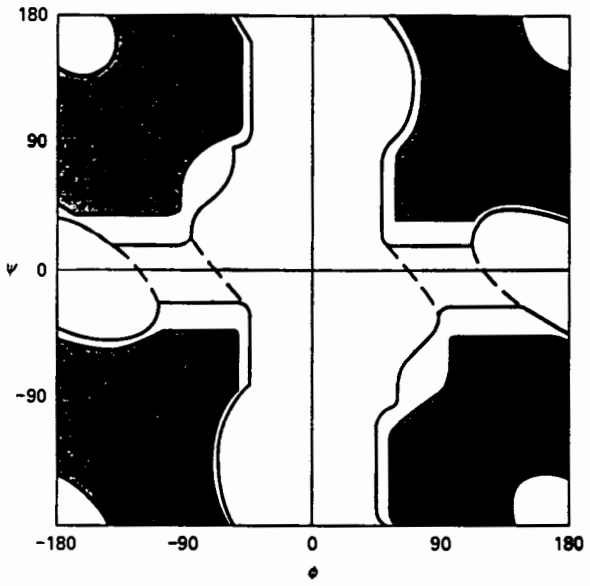


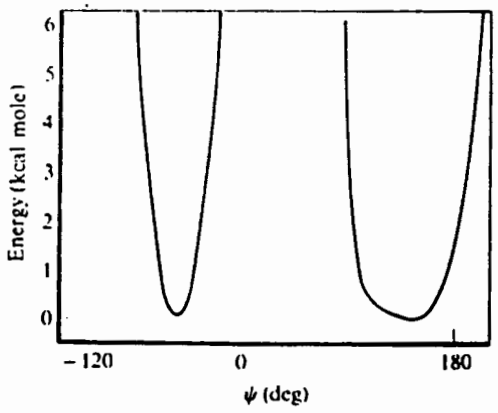
t g^+ g^-



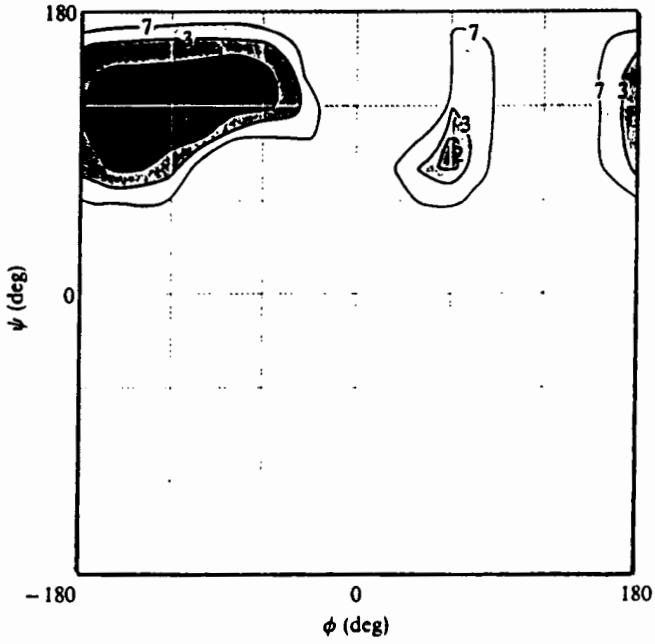
trans *cis*



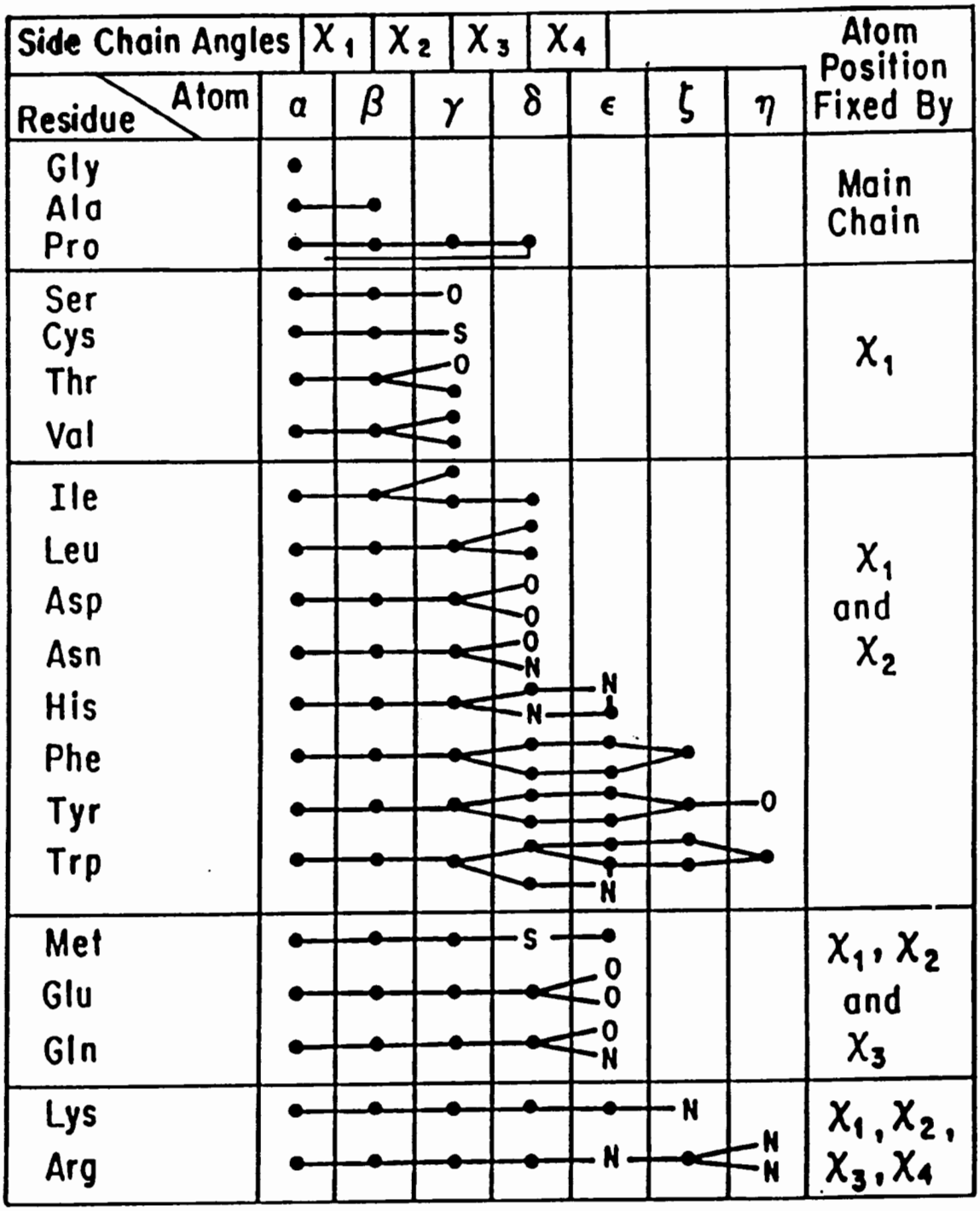




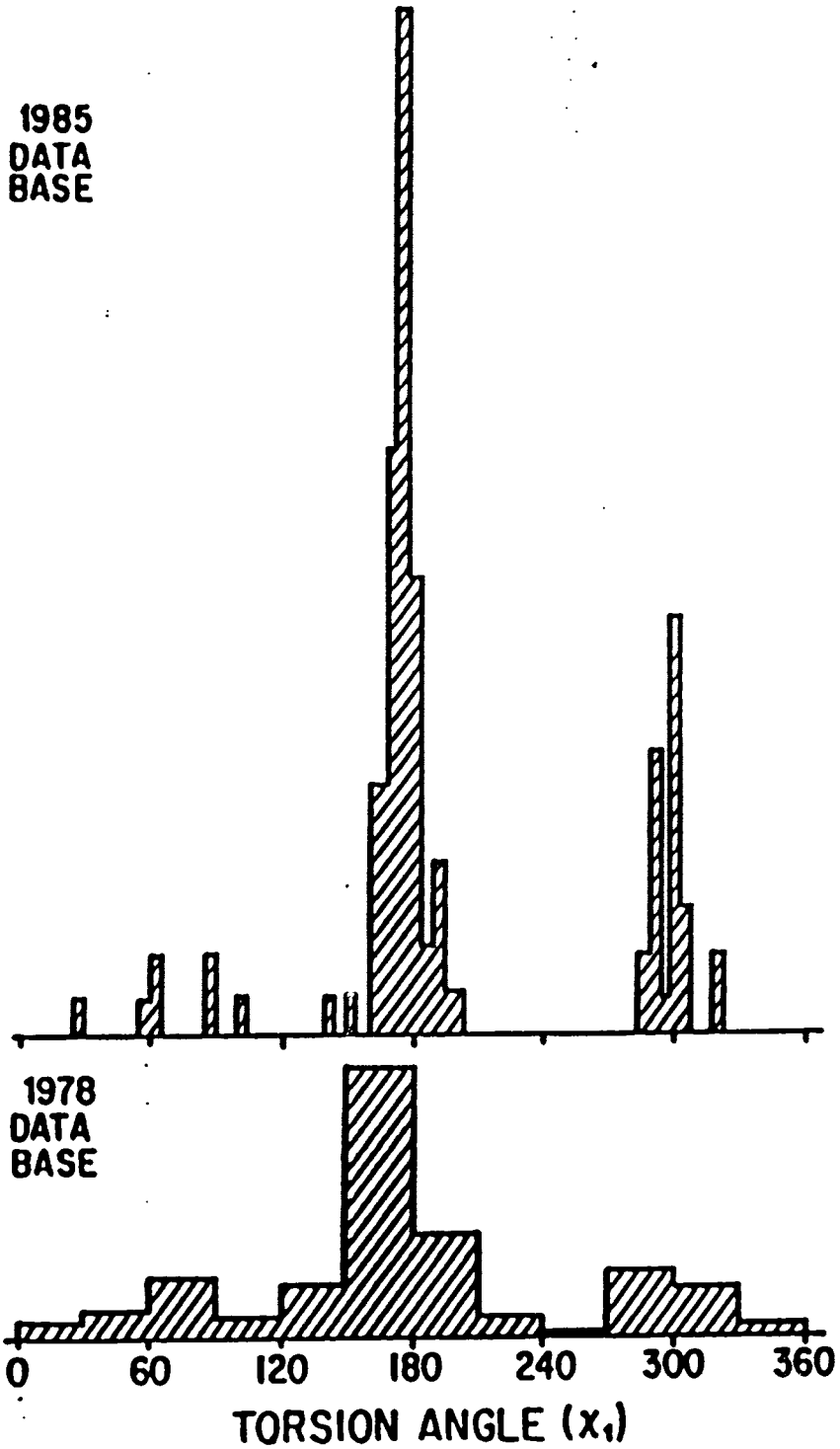
Plot of conformational energy versus ψ for an isolated proline within a polypeptide



Energy contour diagram for an L-alanyl residue that is succeeded by proline.


Side Chain Angles
X1
X2
X3
X4
Atom Position Fixed By
Residue
Atom
α
β
γ
δ
ε
ζ
η
Gly
Ala
Pro
Main Chain
Ser
Cys
Thr
Val
O
S
O
X1
Ile
Leu
Asp
Asn
His
Phe
Tyr
Trp
O
O
O
N
N
N
O
N
X1 and X2
Met
Glu
Gln
S
O
O
O
N
X1, X2 and X3
Lys
Arg
N
N
N
N
X1, X2, X3, X4

VALINE ROTAMER DISTRIBUTION



GENERAL STATEMENTS ON STRUCTURE OF PROTEINS

(1) Severe Steric Overlap and Strain is Rare or Non-Existent

(2) Mean Packing Density is Equivalent to Molecular Crystals

(3) Most Polar Groups are involved in Hydrogen Bonds

(4) Most Formally Charged Groups are in Contact with Water

(5) Minimization of Solvent Contact is a Strong Force leading to Compact Structures

Table 2 Volume occupied by residues in the interior of nine proteins[a](10)

Residue	Total no. in proteins	No. buried[b]	Average volume (V_R) of buried residues (Å^3)	Standard deviation of V_R Å^3	Standard deviation of V_R % V_R	Residue crystal volume equal to amino acid volume less than 11.1 Å^3[c]
Val	163	91	141.7	8.4	5.9	143.4
Ala	186	71	91.5	6.7	7.3	96.6
Ile	106	69	168.8	9.8	5.8	169.7
Gly	160	60	66.4	4.7	7.1	66.5
Leu	138	57	167.9	10.2	6.1	—
Ser	190	46	99.1	7.4	7.5	102.2
Thr	128	32	122.1	6.7	5.5	124.3
Phe	60	29	203.4	10.3	5.1	—
Asp	117	17	124.5	7.7	6.2	122.0
Cys	34	16	105.6	6.0	5.7	108.7
Pro	67	16	129.3	7.3	5.6	124.4
Met	28	14	170.8	8.9	5.2	176.1
Tyr	98	13	203.6	9.6	4.7	201.7
Glu	65	13	155.1	11.4	7.4	143.9
Asn	116	12	135.2	10.1	7.5	—
Trp	39	9	237.6	13.6	5.3	—
His	43	8	167.3	7.4	4.4	166.3
Lys	119	5	171.3	6.8	4.0	—
Gln	80	5	161.1	13.0	8.1	148.0
Cyh	10	4	117.7	4.9	4.2	123.1
Arg	63	0	—	—		—

[a]The nine proteins are calcium binding protein, ribonuclease S, lysozyme, papain, α-chymotrypsin, subtilisin, carboxypeptidase, thermolysin, and lactate dehydrogenase.

[b]A residue is defined as buried if 5% or less of its potential accessible surface area is available to solvent contact.

[c]11.1 Å is the volume lost by an amino acid on becoming a residue. The value used here was found by comparing the crystal volumes of glycerine and glycylglcine, 11.0 Å^3, determined by solution studies. No accurate unit cell dimensions were found for Leu, Phe, Asn, Trp, and Lys.

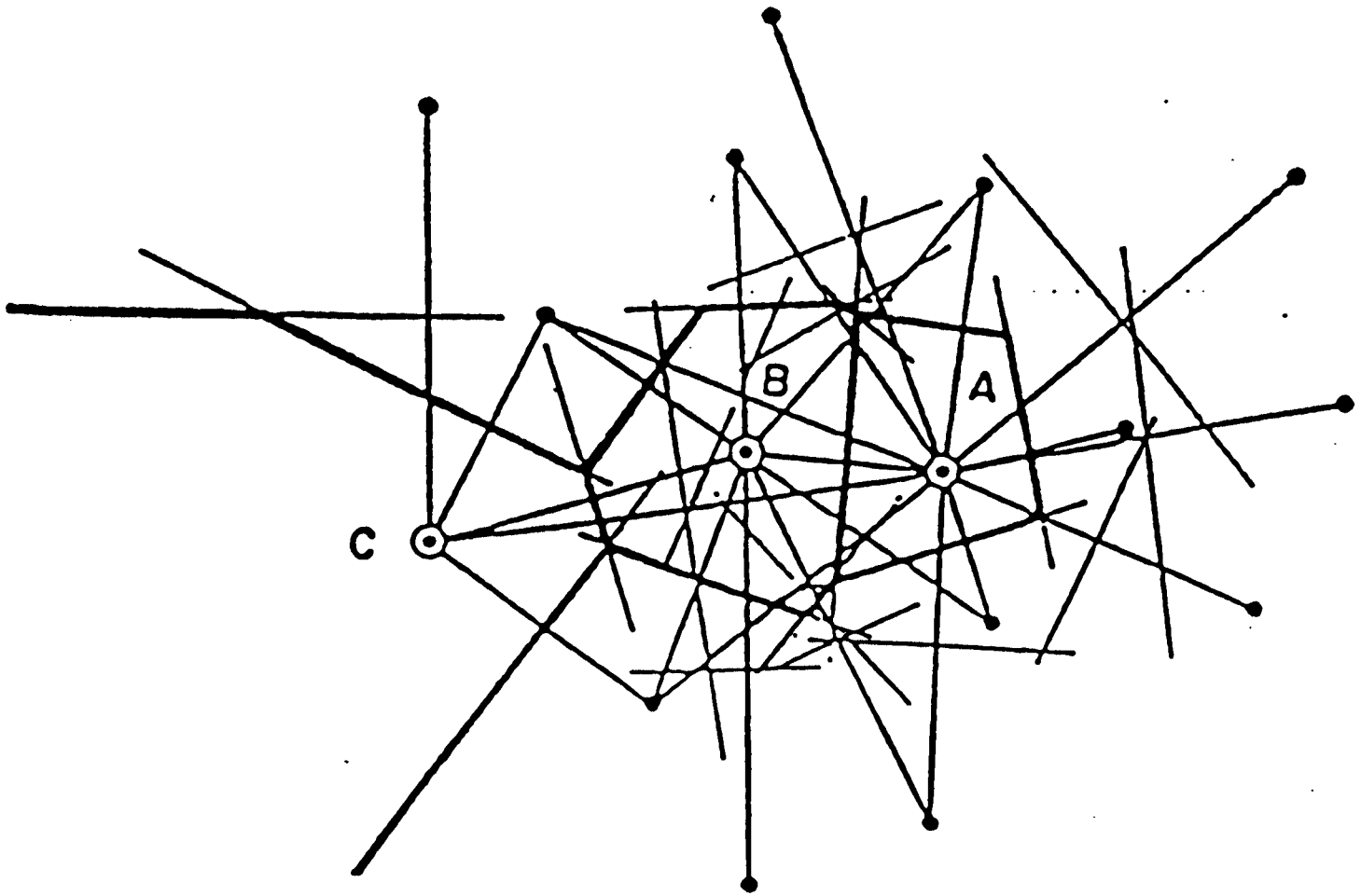


FIG. 1. An example of the Voronoi construction in two dimensions. The heavy lines outline the limiting polygons. The middle weight lines are interatomic vectors between the atom centers represented by dots. The fine lines are the perpendicular bisectors of the vectors. The circled points labeled A and B are interior points surrounded by unique closed polygons. The point labeled C is on the surface of the set of points and a closed polygon cannot be defined. (Adapted from Richards.[16])

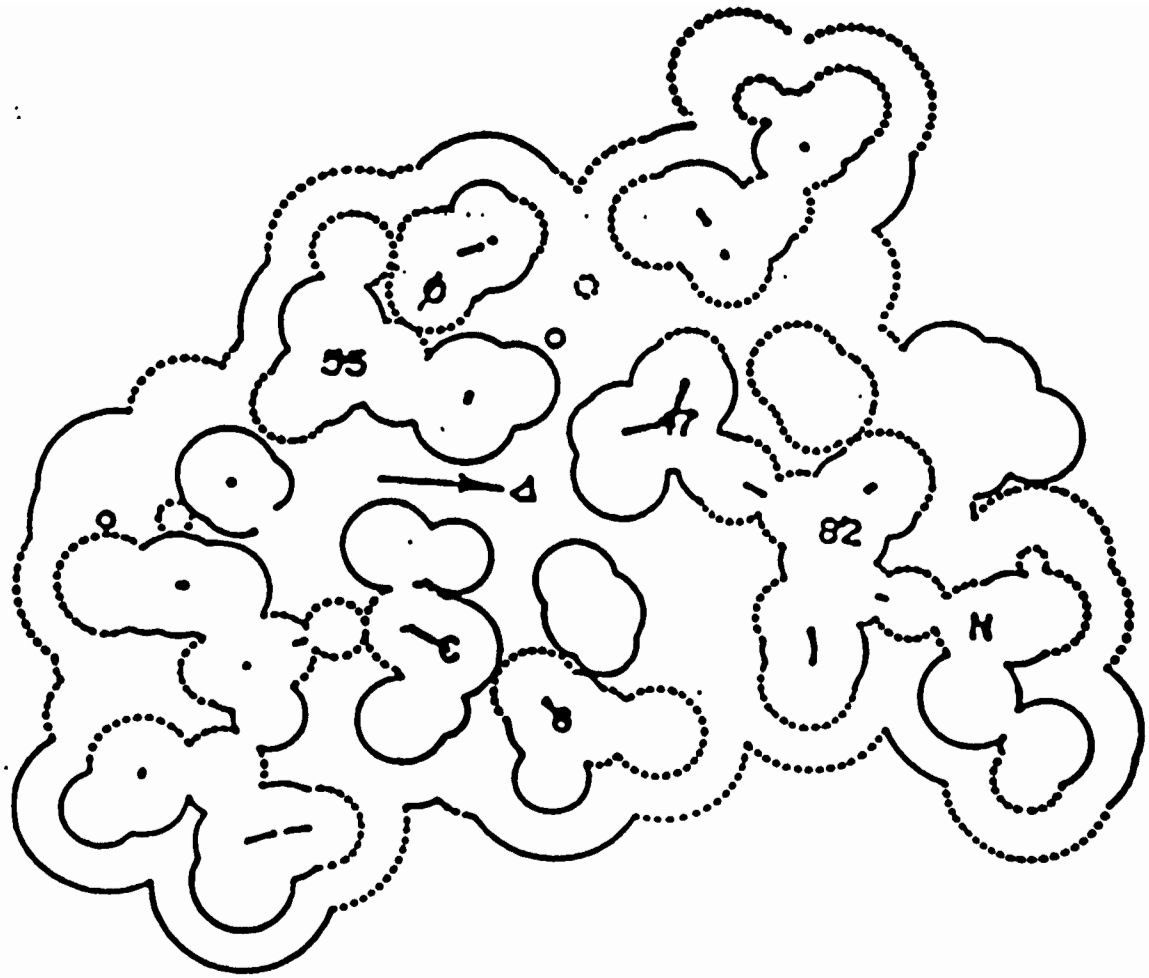


FIG. 8. Superposition of sections through the van der Waals envelope and the accessible surface of ribonuclease S. The arrow indicates a cavity inside the molecule large enough to accommodate a solvent molecule with a radius of 1.4 Å, although it appears to be unfilled in the electron density map. In places the accessible surface is controlled by atoms above or below the section shown. The dashed outline is the surface of N or O atoms, the solid outline C or S atoms. (Reprinted with permission from Lee and Richards.[25])

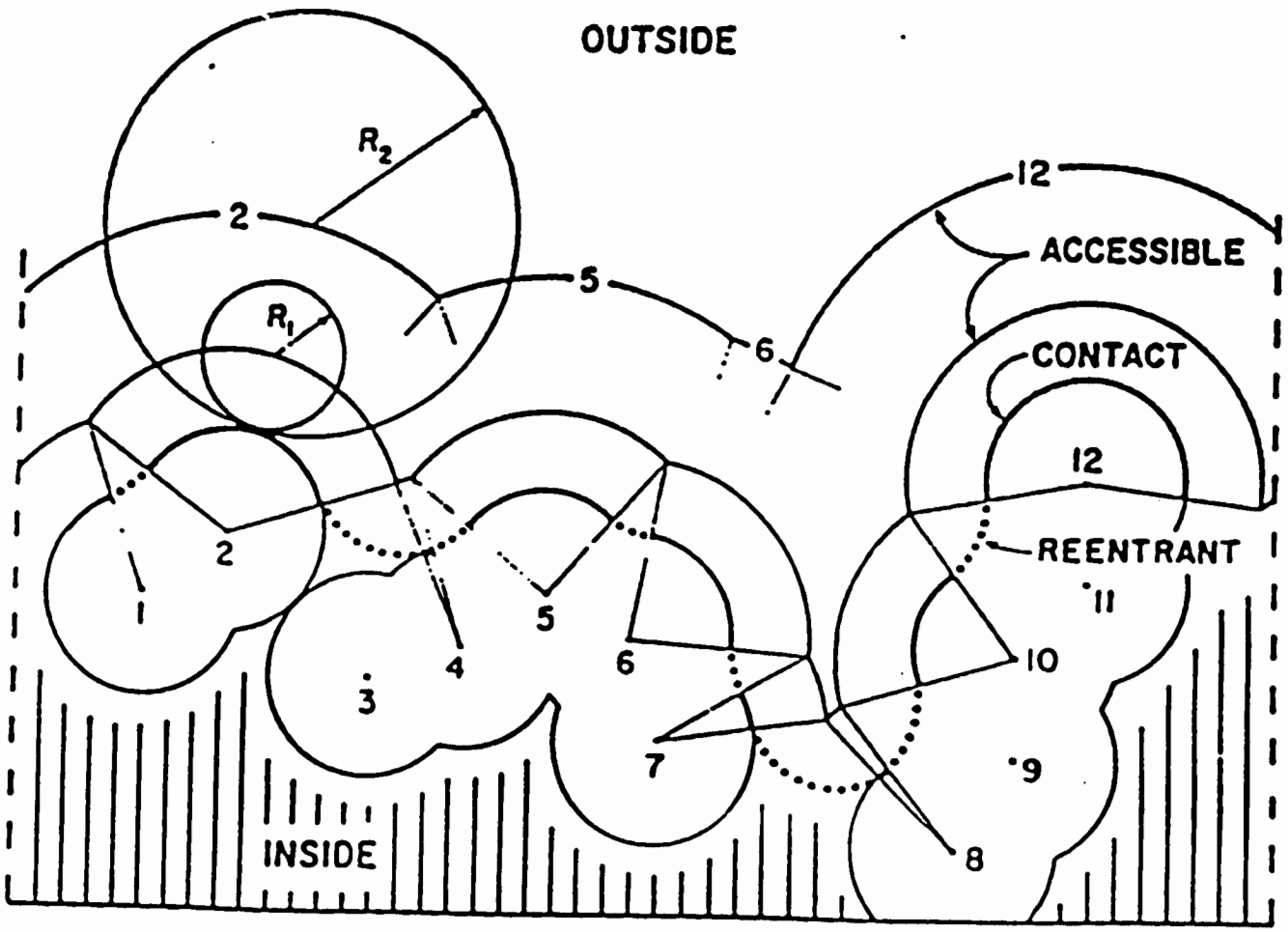
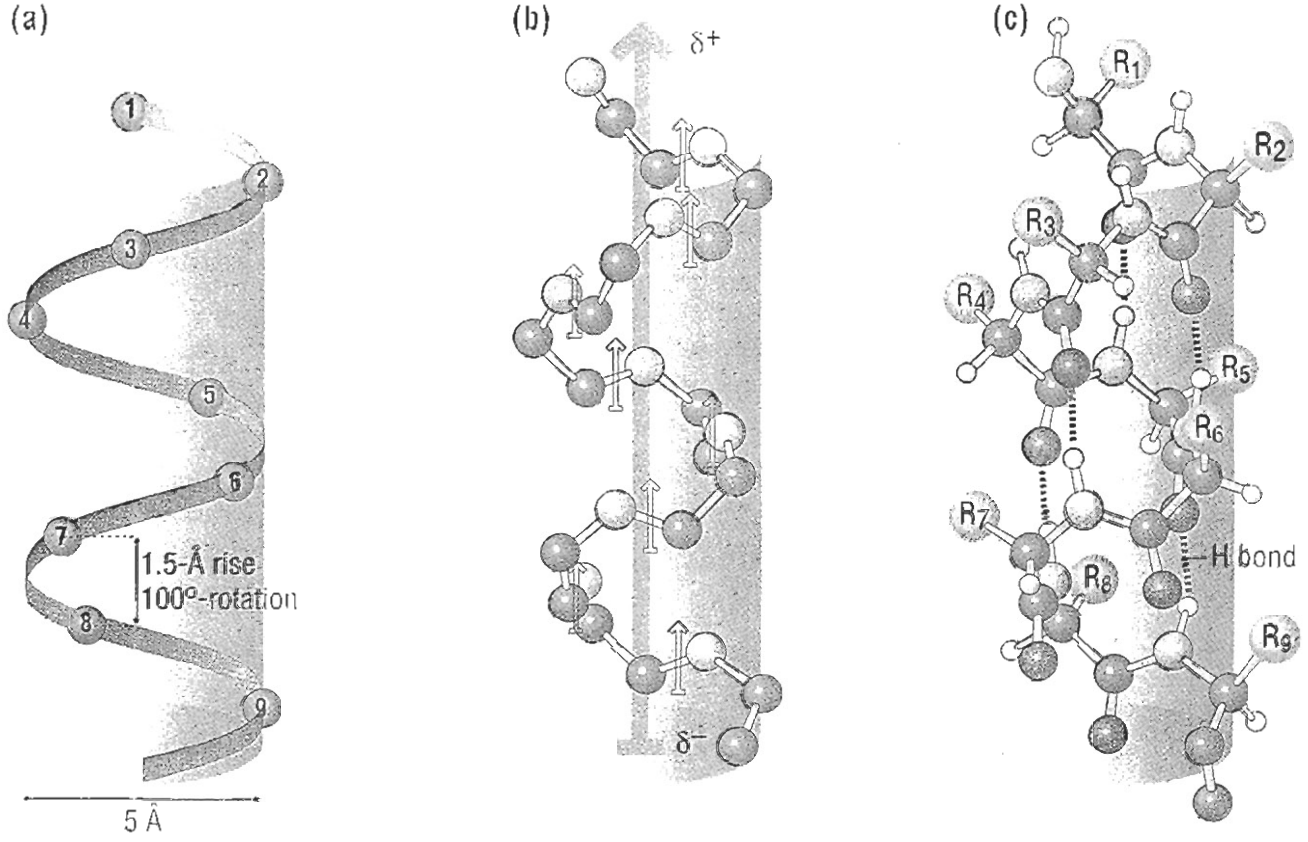


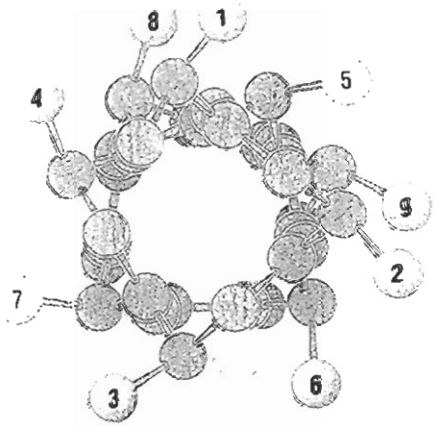
FIG. 7. Schematic representation of possible molecular surface definitions. A section through part of the van der Waals envelope of a hypothetical protein is shown with the atom centers numbered. The accessible surfaces generated by two probes of difference size, R_1 and R_2, and the geometrical definition of contact and reentrant surfaces are shown. (Reproduced with permission from Richards.[22])

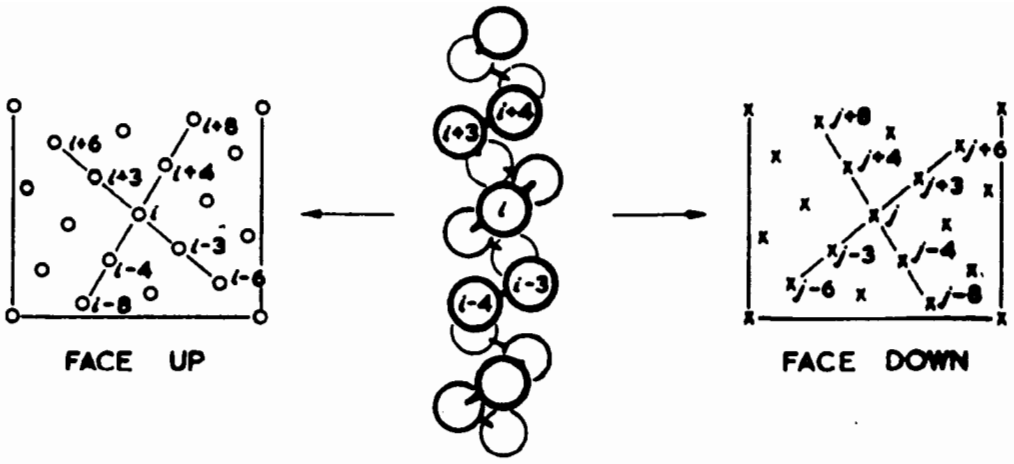
Average Conformational Parameters of Helical Elements

Conformation	Phi	Psi	Omega	Residues per turn	Translation per residue
Alpha helix	–57	–47	180	3.6	1.5
3-10 helix	–49	–26	180	3.0	2.0
Pi-helix	57	–70	180	4.4	1.15
Polyproline I	–83	+158	0	3.33	1.9
Polyproline II	–78	+149	180	3.0	3.12
Polyproline III	–80	+150	180	3.0	3.1

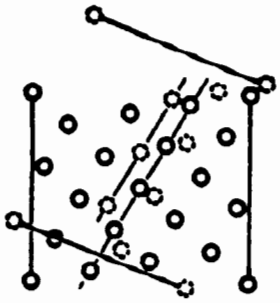


View along the axis of an idealized alpha-helical polypeptide The view is from the amino-terminal end. Side chains project outward from the helical axis at 100° intervals. Note that side chains four residues apart in the sequence tend to cluster on the same face of the helix, for shorter helices. For long helices any such pattern would slowly coil about the helix axis, so if two long helices had a pattern of hydrophobic groups four residues apart they would interact by forming a coiled coil (see Figure 1-67).



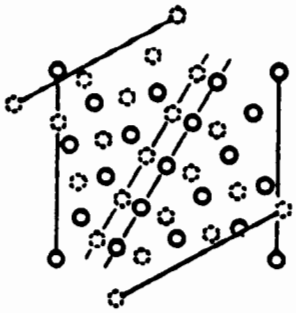


CLASS 1-4
$i \pm 4n \quad j \pm 1n$



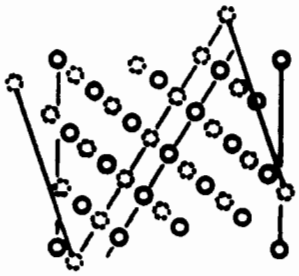
$\Omega = -105^\circ$

CLASS 4-4
$i \pm 4n \quad j \pm 4n$

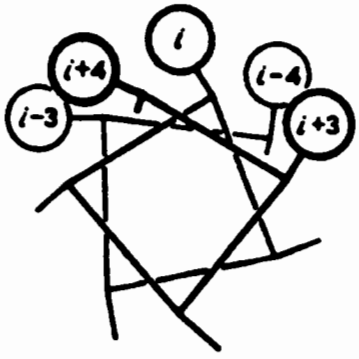


$\Omega = -52^\circ$

CLASS 3-4
$i \pm 4n \quad j \pm 3n$



$\Omega = +23^\circ$



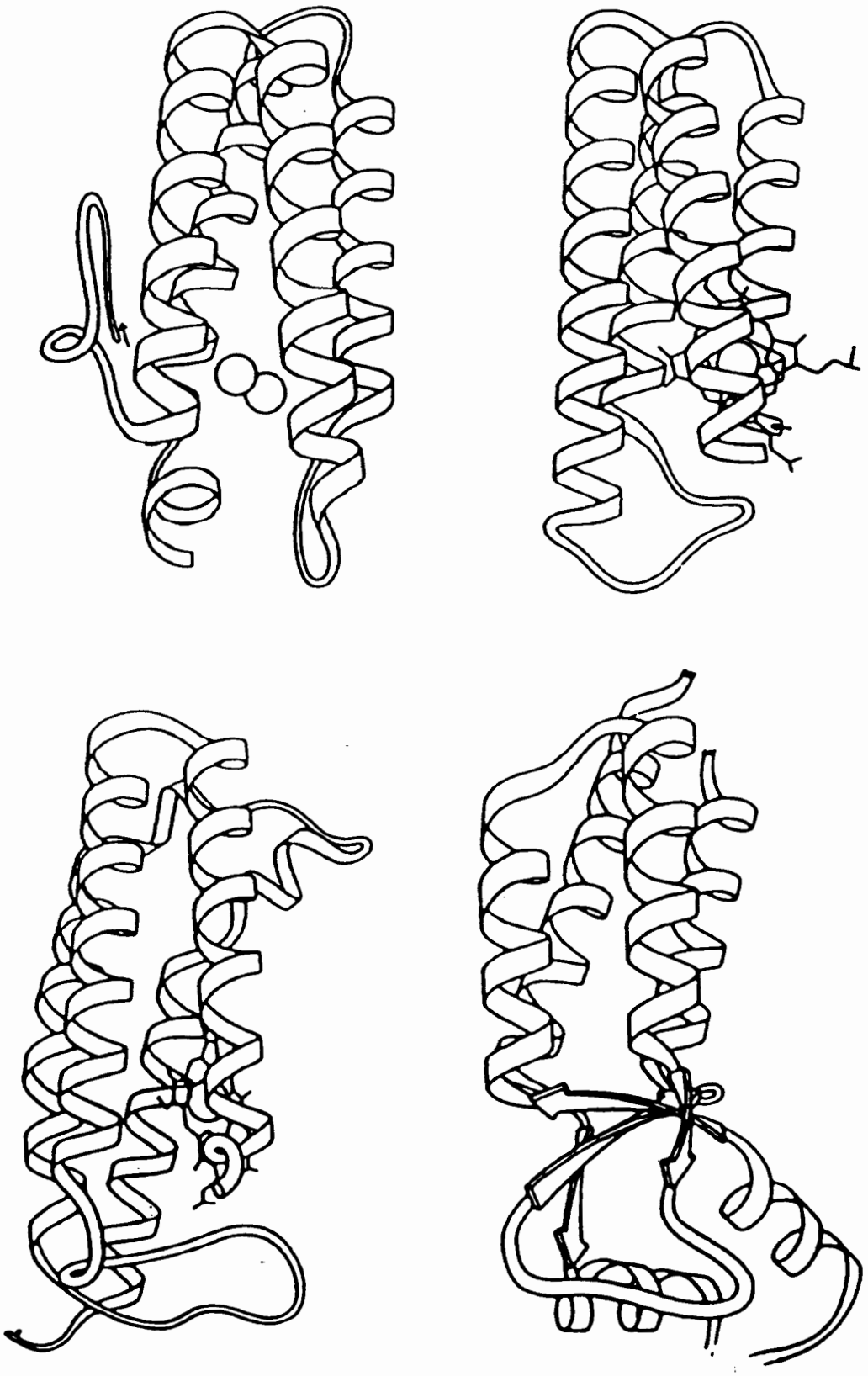
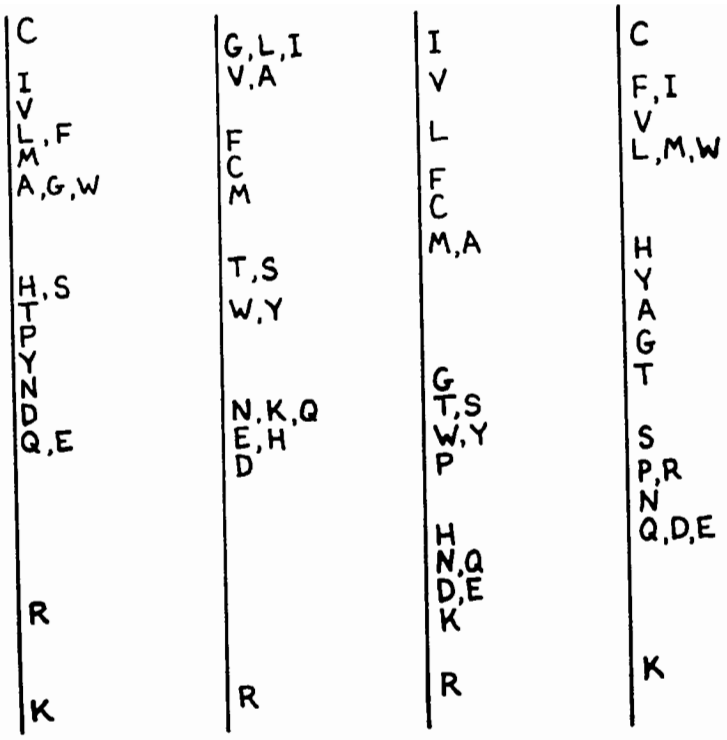
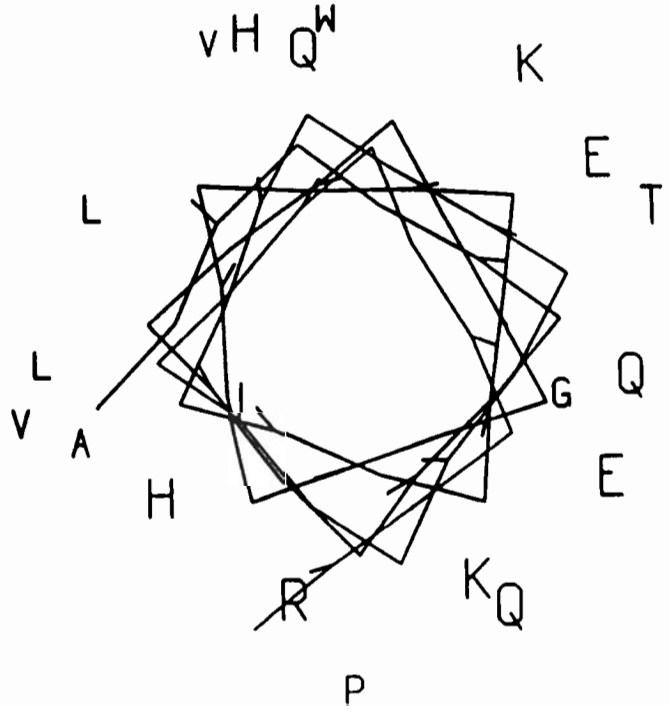


Figure 36. Ribbon drawings of four examples of up-and-down four-helix bundles (MHR, 56B, CCY, TMV).



Janin (1979) | **Wolfenden *et al.* (1981)** | **Kyte and Doolittle (1982)** | **Rose *et al.* (1985)**



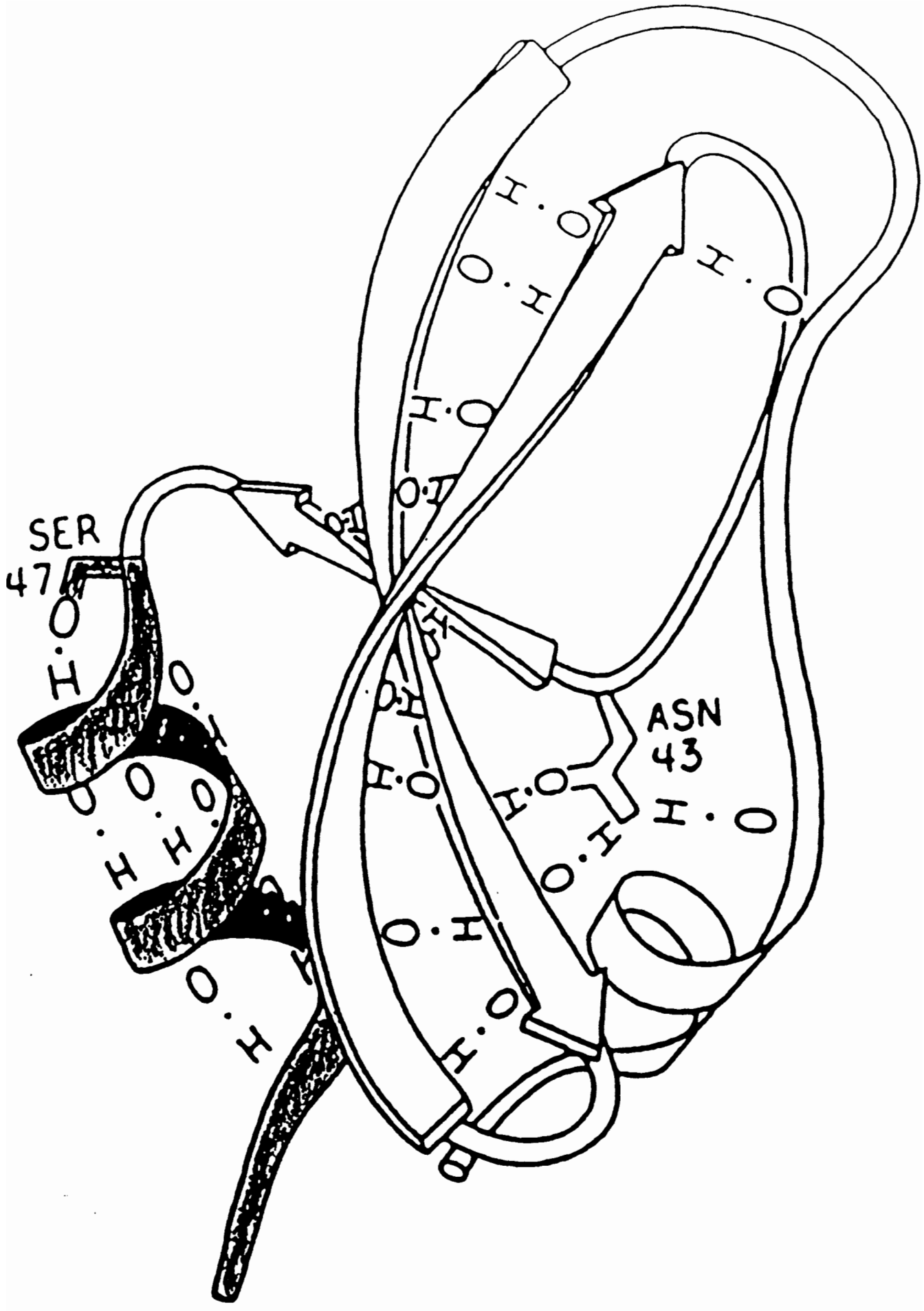

SER
47
ASN
43

Table 1. Helix propensity values. T4L, T4 lysozyme. All free-energy scales have been normalized so that the helix propensity of Gly is 0. The symbol P_{mid} represents the relative frequency with which the amino acid occurs in the middle of α helices in known protein structures; s is the Zimm-Bragg parameter and the correlation is with RT ln$s$. The sources of data are indicated by the reference numbers at the top of their respective columns.

Amino acid	ΔΔG (kcal mol^{-1})					Percent helix (*5*)	P_{mid} (*2*)	s (*3*)
	T4L site 44*	T4L site 131†	(*6*)	(*4*)	Barnase site 32 (*9*)			
Ala	0.96	0.94	0.79	0.77	0.91	78	1.41	1.07
Leu	0.92	0.77	0.62	0.62	0.56	80	1.34	1.14
Met	0.86	0.81	0.57	0.50	0.60		1.30	1.20
Ile	0.84	0.84	0.39	0.23	0.10	41	1.09	1.14
Gln	0.80		0.48	0.33	0.43		1.27	0.98
Arg	0.77			0.68	0.77		1.21	1.03
Lys	0.73			1.23	0.72		1.23	0.94
Tyr	0.72			0.17	0.09		0.74	1.02
Val	0.63	0.69	0.34	0.14	0.03	17	0.98	0.95
Phe	0.59			0.41	0.22	23	1.16	1.09
Trp	0.58			0.45	0.07		1.02	1.11
His	0.57			0.06	0.13		1.05	0.69
Thr	0.54	0.56	0.23	0.11	0.12		0.76	0.82
Glu	0.53	0.88‡		0.27	0.36		1.18	1.35
Ser	0.53	0.64	0.28	0.35	0.50		0.57	0.76
Asp	0.42	0.77‡		0.15	0.20		0.99	0.78
Cys	0.42			0.23	0.09		0.66	0.99
Asn	0.39		0.18	0.07	0.25		0.76	0.78
Gly	0.00	0	0.00	0.00	0.00		0.43	0.59
Pro§	−2.50			−3.00			0.19	0.19

*The conditions for optimal reversibility of unfolding and two-state behavior were 0.025 M KCl, 0.003 M H_3PO_4, and 0.017 M KH_2PO_4 (pH 3.01). The difference between the free energy of unfolding of each mutant and that of the Gly variant used as the reference was determined by van't Hoff analysis (*17*, *33*). The estimated uncertainty in ΔΔG for all mutants is less than 0.10 kcal mol^{-1}. †Data are from (*8*) and this work. The values of ΔΔG determined under conditions identical to those used for the site 44 replacements. ‡Glu and Asp were not included in the determination of the correlation because of presumed electrostatic stabilization. §Correlations with the site 44 data from T4 lysozyme, excluding Pro: T4L site 131, 0.97, (*6*), 0.93, (*4*), 0.71, barnase site 32, 0.62, percent helix, 0.93, P_{mid}, 0.81; and s, 0.69.

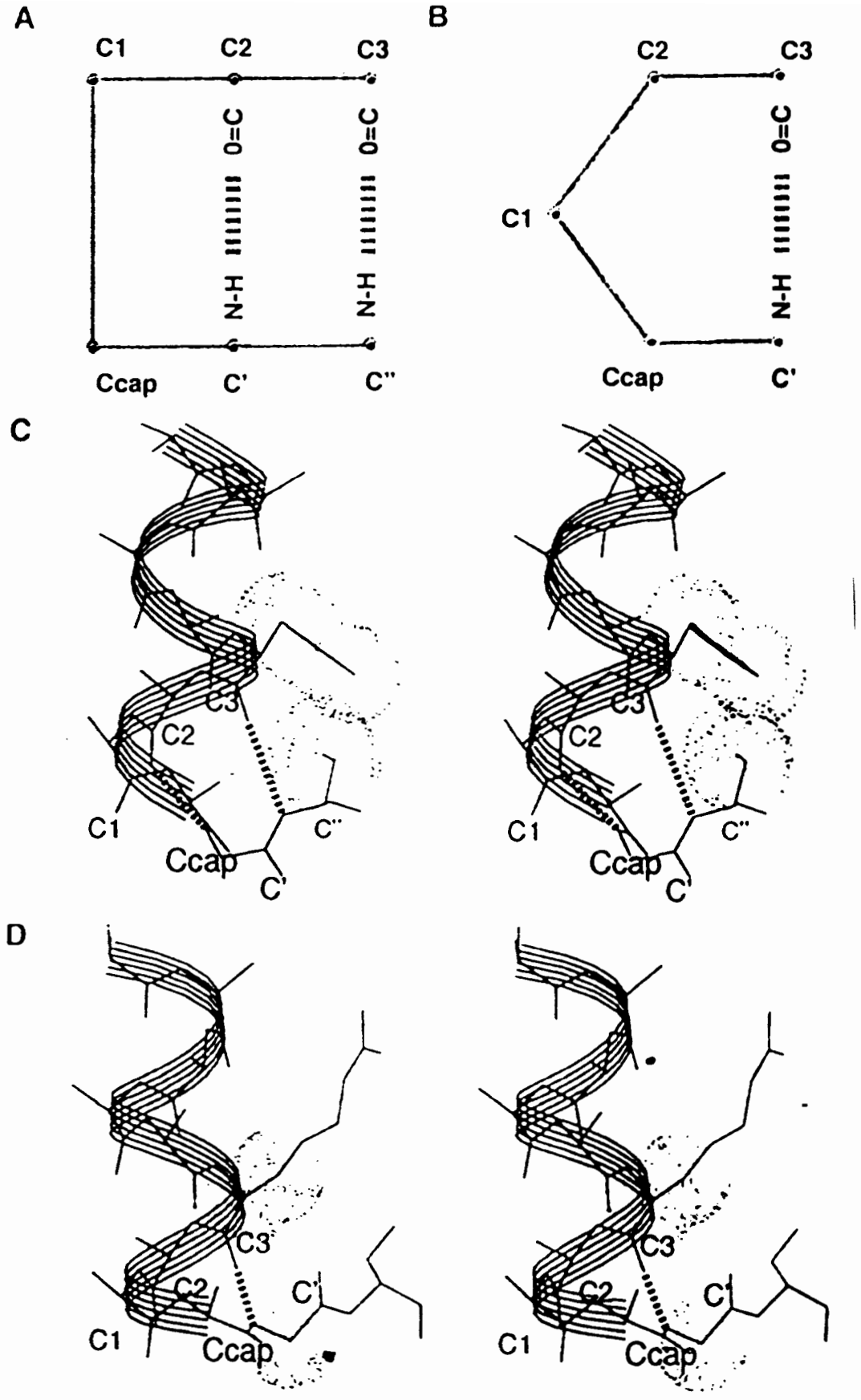


Fig. 1. An example of each Gly motif, from cytochrome C551, designated as 351c in the PDB (*14*). (A and B) Schematic diagrams of the structures shown in (C) and (D), respectively, in which only the hydrogen bonding pattern is emphasized. Molecular representations (C) and (D) include all backbone (N, Cα, C, and O) and Cβ atoms, with hydrogen bonds represented as green dashed lines. The stippled surface represents the van der Waals radius surrounding selected atoms. (C) Helix from residues 2 (Ncap) to 10 (Ccap), terminating in a Schellman motif. The interaction between side chains of C" (Cys) and C3 (Phe) and backbone hydrogen bonds from C" to C3 ($i \rightarrow i$-5) and C' to C2 ($i \rightarrow i$-3) is shown. (D) Helix from residues 39 (Ncap) to 50 (Ccap), terminating in an α_L motif. The hydrogen bond from C' to C3 ($i \rightarrow i$-4) is shown.

Table 2. The stereochemical rules for Gly motifs. Rules for helix continuation or termination are summarized. $C^{\bullet}$ is the key position. If Gly is followed by an apolar residue at $C^{\bullet}$, an interaction with the apolar residue at C3 (or C4 if C3 is polar, or C2 if both C3 and C4 are polar) terminates the helix in a Schellman motif, exposing C1 to solvent. If Gly is followed by a polar residue, the helix terminates in an α_L motif unless a favorable side chain–side chain interaction between $C^{\bullet}$ and C2 prevents helix termination. When $C^{\bullet}$ is Gly (that is, Gly-Gly), the polarity of $C^{\bullet\bullet}$ (in lieu of $C^{\bullet}$) selects between the Schellman and α_L motifs; apolar: Ala, Val, Ile, Leu, Met, Phe, Trp, Cys; polar: Ser, Thr, Asn, Asp, Gln, Glu, Arg, Lys; omitted: His, Tyr, Pro. The long alkyl side chain moieties of Lys and Arg can function as hydrophobic sites (*19*) and are observed to interact with apolar residues at the indicated positions. Histidine can be polar or apolar, depending on its protonation state, which cannot be determined in x-ray structures. Tyrosine acts ambiguously at $C^{\bullet}$, partitioning to both motifs. Glycine-proline terminates a helix with a (Ccap) N–H . . . O=C (C3) hydrogen bond; unlike the Schellman and α_L motifs, Gly in Gly-Pro structures has a negative value of ϕ.

Residue					Structural motif
C3 > C4 > C2	C1	$C^{\cdot}$	$C^{\bullet}$	$C^{\bullet\bullet}$	
Apolar Lys or Arg	Polar or Ala	Gly	Apolar Lys or Arg	Not "bulky"†	Schellman
Polar* not Lys not Arg	Apolar* not Ala	Gly	Apolar Lys or Arg		Helix continuation
Apolar Lys or Arg preferred		Gly only	Polar* not Lys not Arg		α_L

*The helix continues if C1 is apolar or if C3, C4, and C2 are all polar and interaction with $C^{\bullet}$ is not possible. †In our polyalanine model, Trp was strongly disfavored for steric reasons, although smaller aromatic residues were allowed. More realistic sequences may impose further steric constraints.