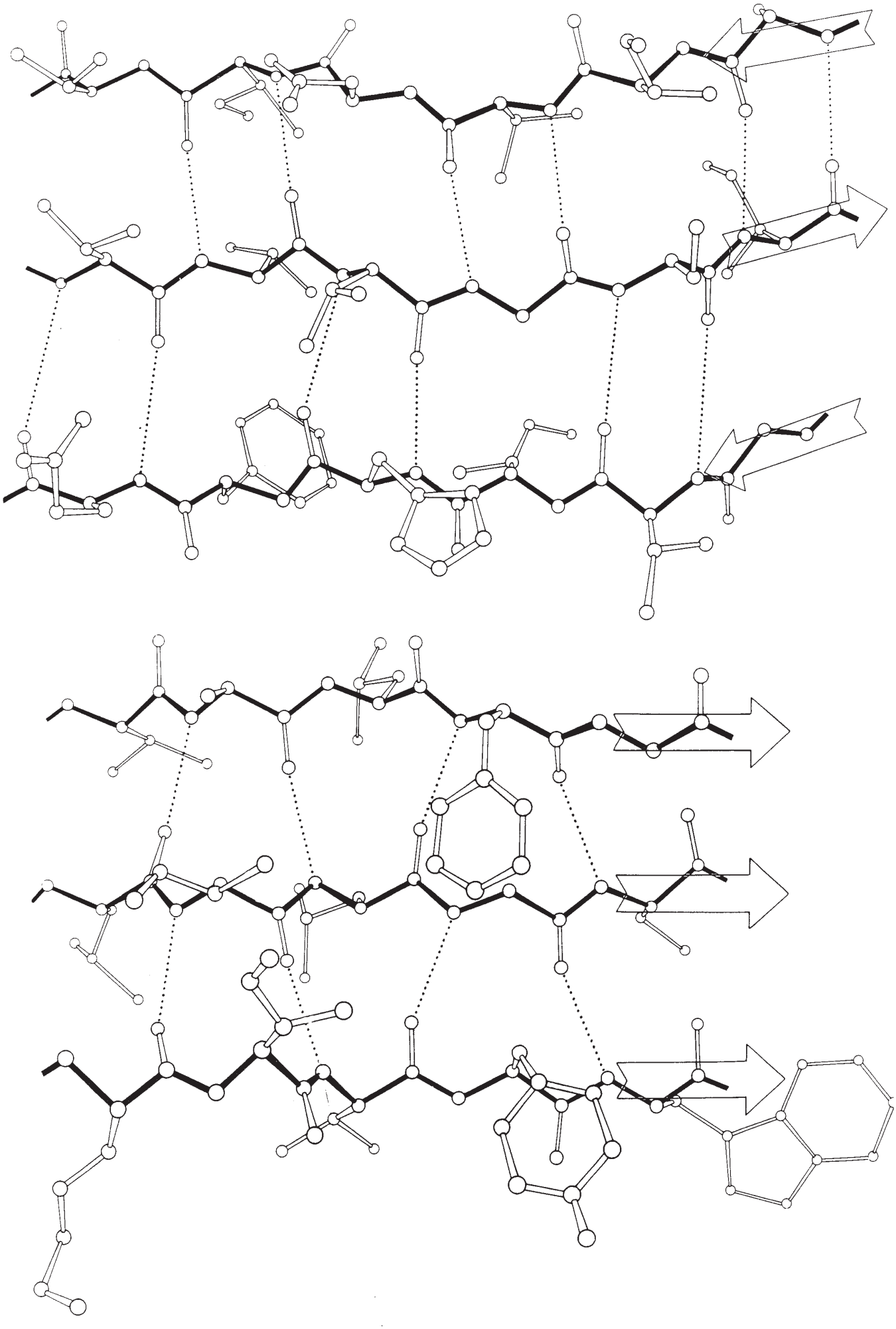
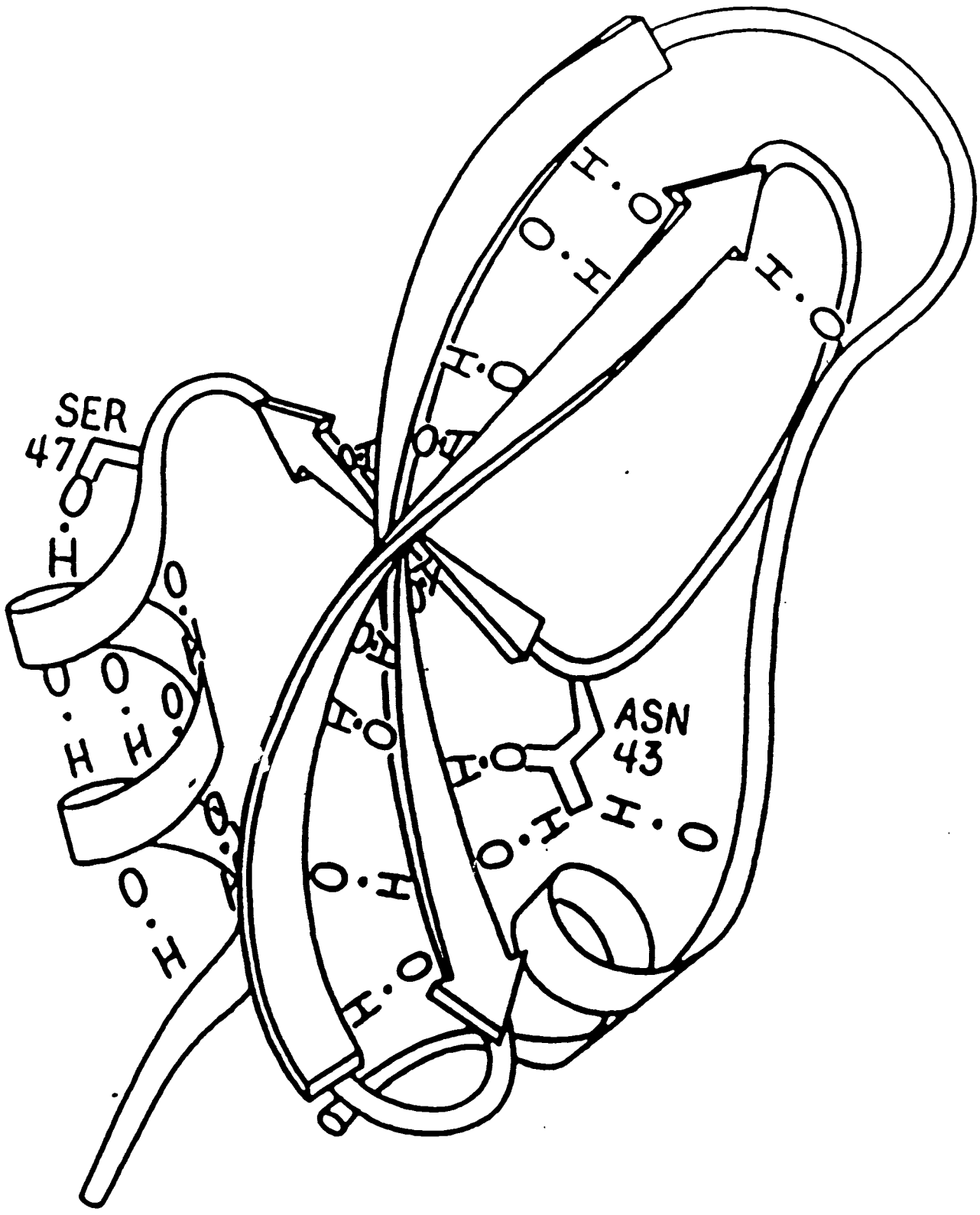
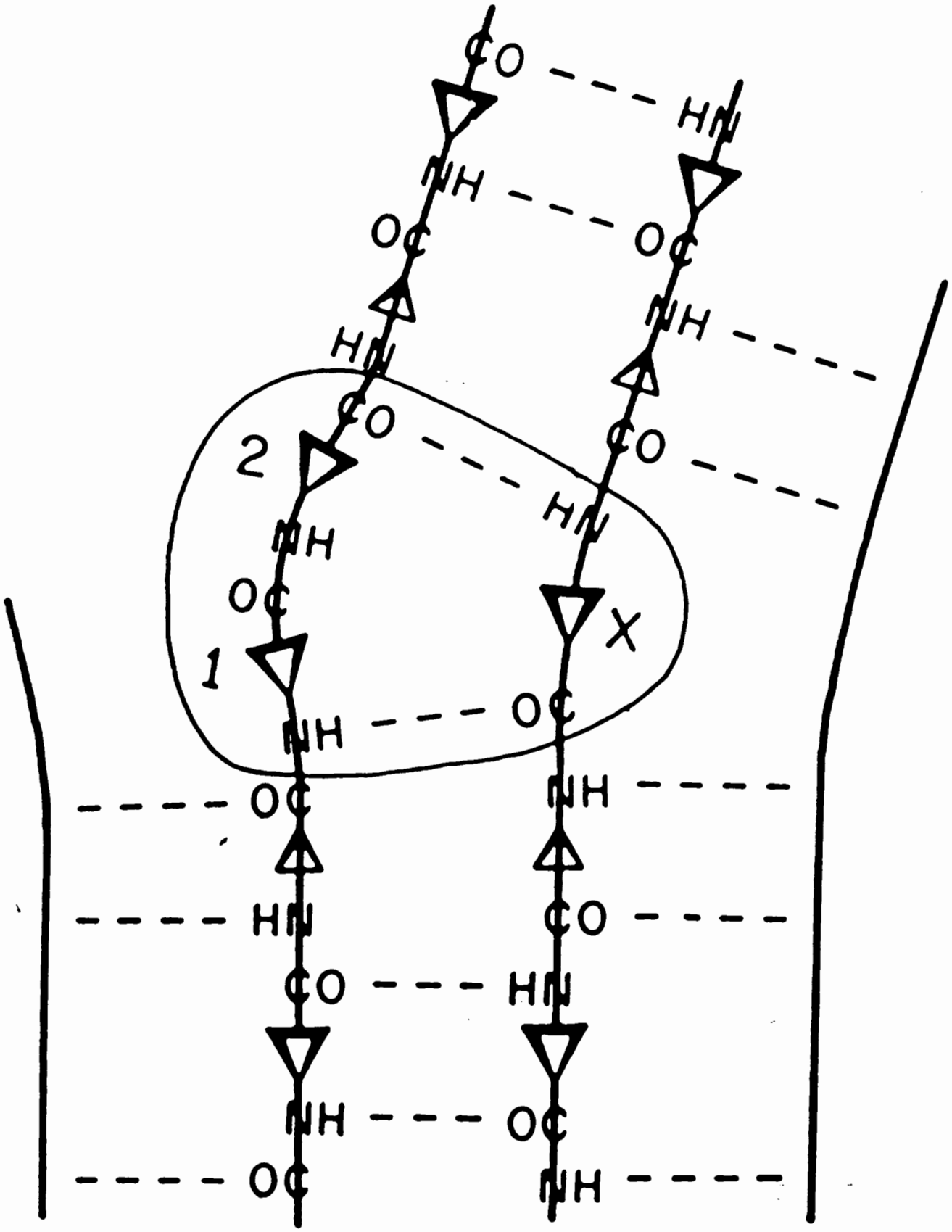
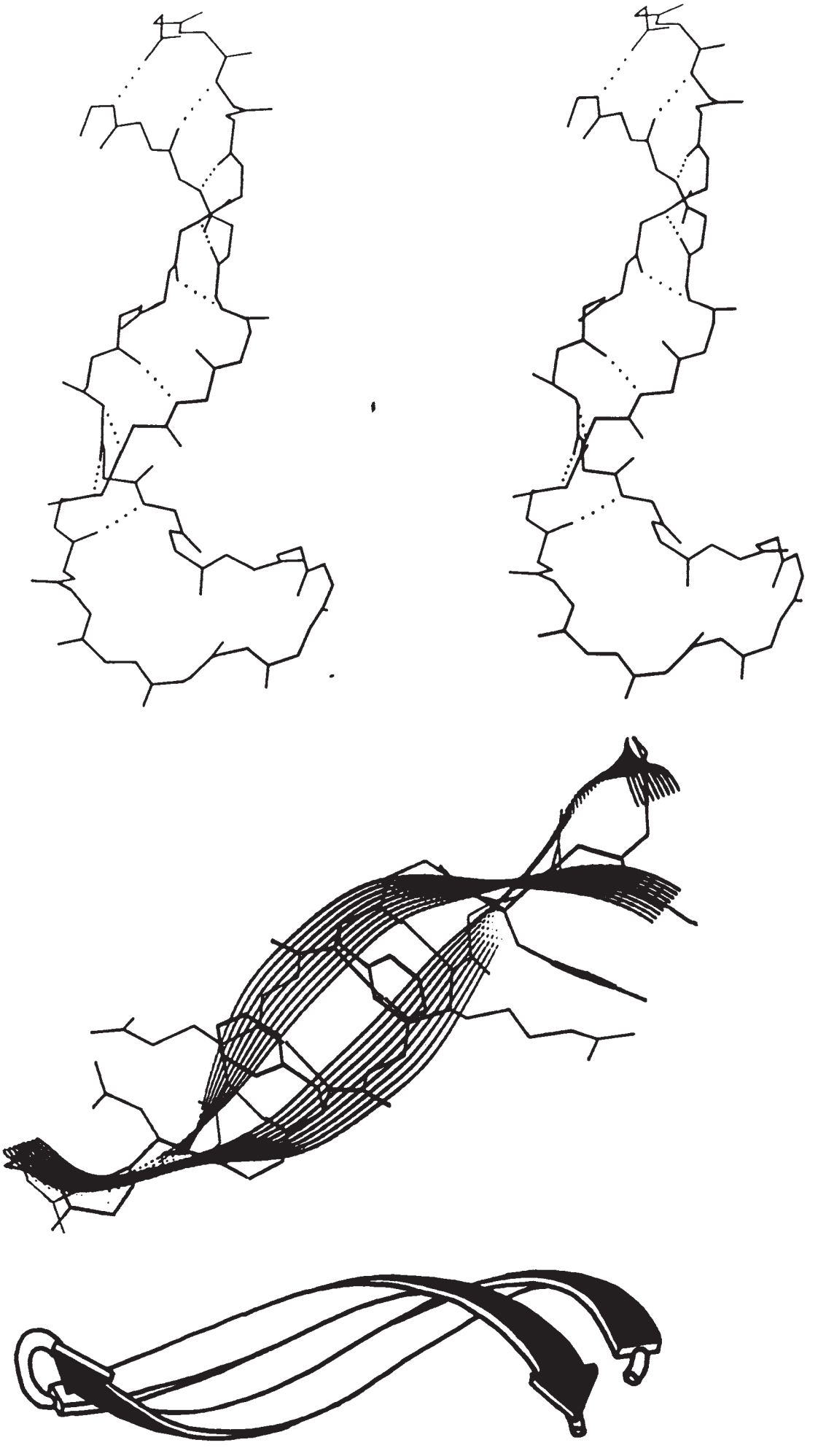



SER
47
ASN
43

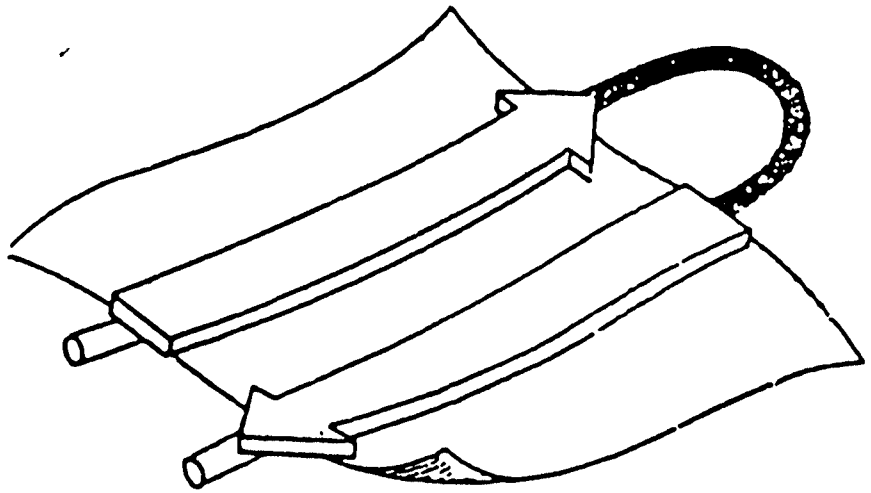

2
1
X

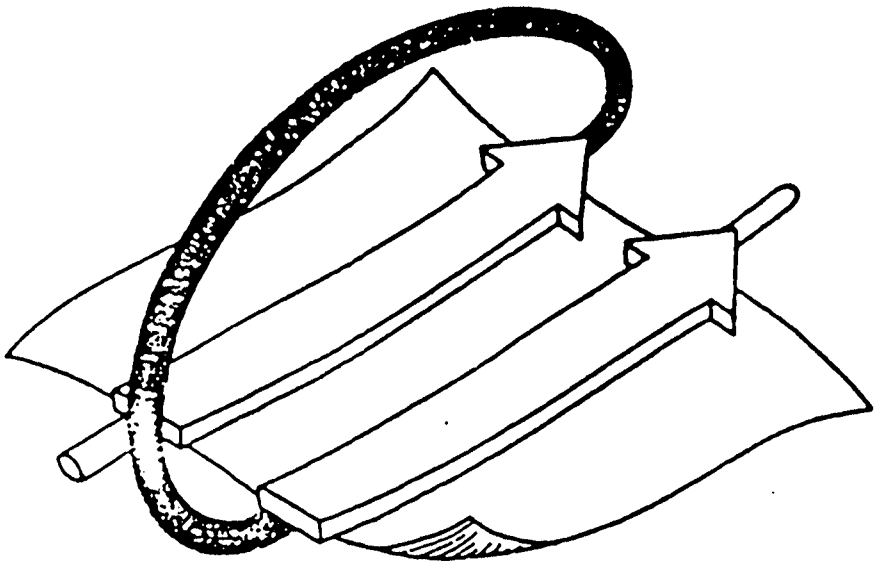


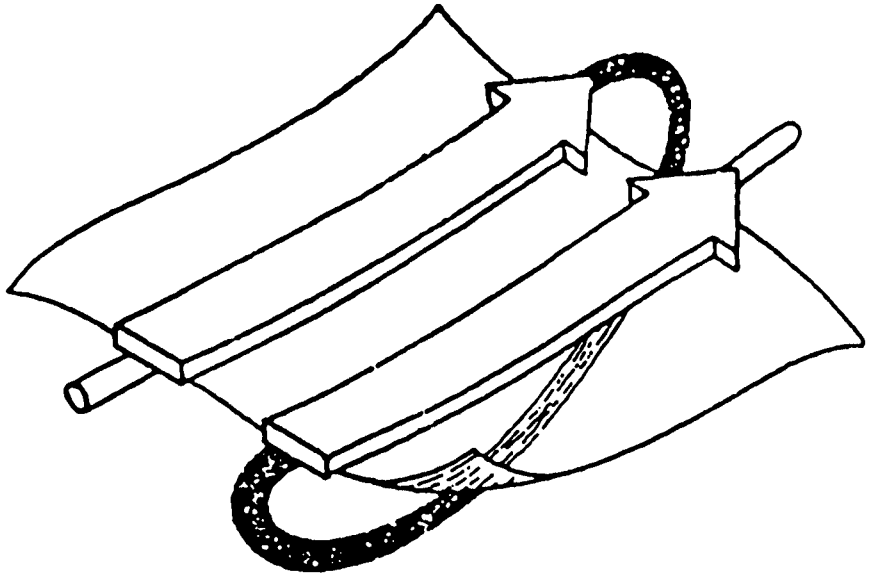
Specific recognition between individual or minimally grouped residues. Pair contact counts N_{XY} (top) and pair correlations $g_{XY} \pm \sigma_{XY}$ (bottom)

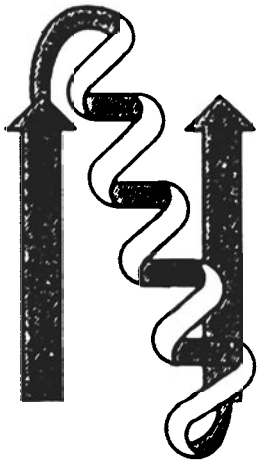
A. *Antiparallel strands*

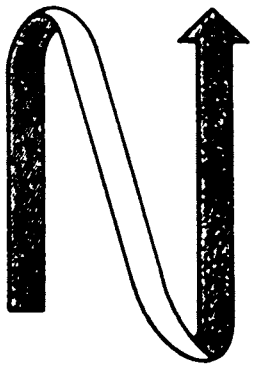
	1 V	2 L	3 I	4 Y	5 FW	6 MC	7 T	8 S	9 AP	10 G	11 KR	12 DNH	13 QE	
1 V	26	25	23	11	18	7	6	19	14	12	13	8	6	
	1·2±0·3	1·5±0·2	1·7±0·3	0·9±0·3	1·3±0·3	0·9±0·3	0·4±0·2	1·3±0·3	0·8±0·2	1·2±0·3	0·9±0·3	0·5±0·3	0·5±0·3	
2 L	25	8	12	12	14	11	12	5	13	10	5	10	4	
	1·5±0·2	0·6±0·4	1·2±0·3	1·4±0·3	1·4±0·3	1·8±0·4	1·0±0·3	0·5±0·3	1·0±0·3	1·4±0·4	0·4±0·3	0·9±0·3	0·4±0·3	
3 I	23	12	8	5	9	0	9	4	20	2	8	9	3	
	1·7±0·3	1·2±0·3	1·0±0·5	0·7±0·4	1·1±0·3	0·0±0·5	0·9±0·3	0·5±0·3	1·9±0·3	0·3±0·4	0·9±0·3	1·0±0·3	0·4±0·4	
4 Y	11	12	5	12	8	6	9	1	7	6	7	6	8	
	0·9±0·3	1·4±0·3	0·7±0·4	2·0±0·6	1·1±0·4	1·4±0·5	1·0±0·3	0·1±0·4	0·8±0·3	1·2±0·4	0·9±0·4	0·8±0·4	1·2±0·4	
5 FW	18	14	9	8	8	6	3	4	9	7	12	10	7	
	1·3±0·3	1·4±0·3	1·1±0·3	1·1±0·4	1·0±0·5	1·2±0·4	0·3±0·3	0·4±0·3	0·8±0·3	1·2±0·4	1·3±0·3	1·1±0·3	0·9±0·4	
6 MC	7	11	0	6	6	8	5	6	8	4	2	2	4	
	0·9±0·3	1·8±0·4	0·0±0·5	1·4±0·5	1·2±0·4	2·8±0·8	0·8±0·4	1·1±0·4	1·2±0·4	1·1±0·5	0·4±0·4	0·4±0·4	0·8±0·5	
7 T	6	12	9	9	3	5	18	21	18	4	13	7	14	
	0·4±0·2	1·0±0·3	0·9±0·3	1·0±0·3	0·3±0·3	0·8±0·4	1·5±0·4	1·9±0·3	1·4±0·3	0·6±0·4	1·2±0·3	0·6±0·3	1·4±0·3	
8 S	19	5	4	1	4	6	21	14	13	5	11	13	7	
	1·3±0·3	0·5±0·3	0·5±0·3	0·1±0·4	0·4±0·3	1·1±0·4	1·9±0·3	1·5±0·5	1·1±0·3	0·8±0·4	1·1±0·3	1·3±0·3	0·8±0·3	
9 AP	14	13	20	7	9	8	18	13	10	7	11	7	11	
	0·8±0·2	1·0±0·3	1·9±0·3	0·8±0·3	0·8±0·3	1·2±0·4	1·4±0·3	1·1±0·3	0·7±0·4	0·9±0·4	0·9±0·3	0·6±0·3	1·1±0·3	
10 G	12	10	2	6	7	4	4	5	7	0	7	9	8	
	1·2±0·3	1·4±0·4	0·3±0·4	1·2±0·4	1·2±0·4	1·1±0·5	0·6±0·4	0·8±0·4	0·9±0·4	0·0±0·7	1·1±0·4	1·4±0·4	1·4±0·4	
11 KR	13	5	8	7	12	2	13	11	11	7	12	11	16	
	0·9±0·3	0·4±0·3	0·9±0·3	0·9±0·4	1·3±0·3	0·4±0·4	1·2±0·3	1·1±0·3	0·9±0·3	1·1±0·4	1·2±0·4	1·1±0·3	1·8±0·3	
12 DNH	8	10	9	6	10	2	7	13	7	9	11	20	12	
	0·5±0·3	0·9±0·3	1·0±0·3	0·8±0·4	1·1±0·3	0·4±0·4	0·6±0·3	1·3±0·3	0·6±0·3	1·4±0·4	1·1±0·3	2·0±0·5	1·4±0·3	
13 QE	6	4	3	8	7	4	14	7	11	8	16	12	10	
	0·5±0·3	0·4±0·3	0·4±0·4	1·2±0·4	0·9±0·4	0·8±0·5	1·4±0·3	0·8±0·3	1·1±0·3	1·4±0·4	1·8±0·3	1·4±0·3	1·3±0·5	
N_x	188	141	112	98	115	69	139	123	148	81	128	124	110	Total 1576
%	11·9	8·9	7·1	6·2	7·3	4·4	8·8	7·8	9·4	5·1	8·1	7·9	7·0	100

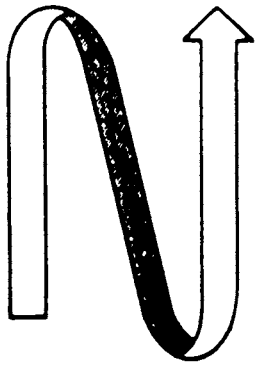




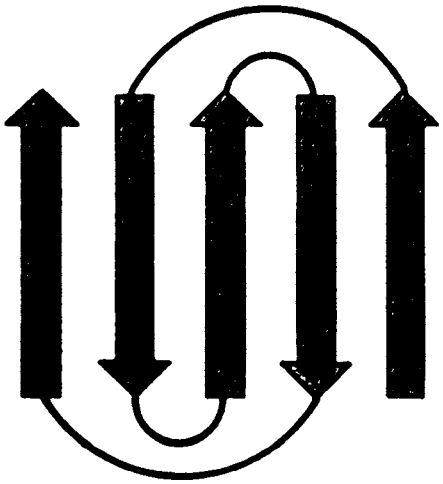




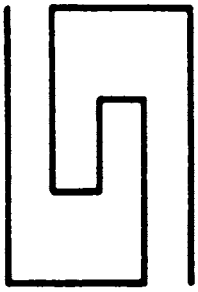


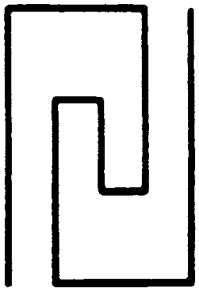


Very
rare

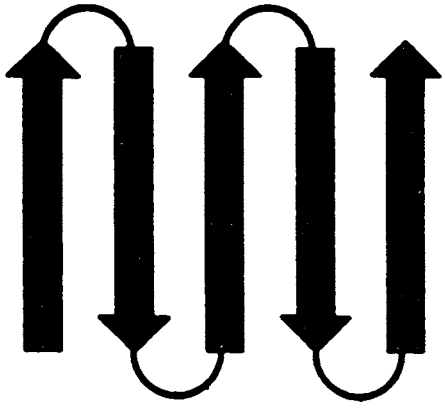


5 2 3 4 1

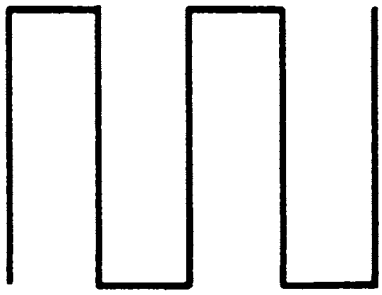




Not
observed



1 2 3 4 5



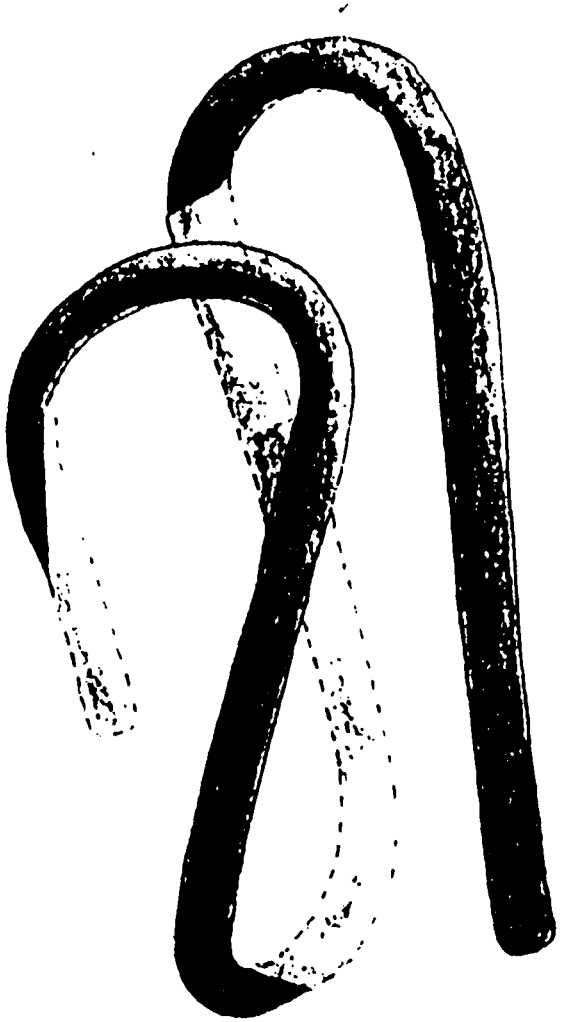


Figure 5.16 A schematic diagram of a piece of string wrapped around a barrel to illustrate the basic pattern of a jelly roll motif.

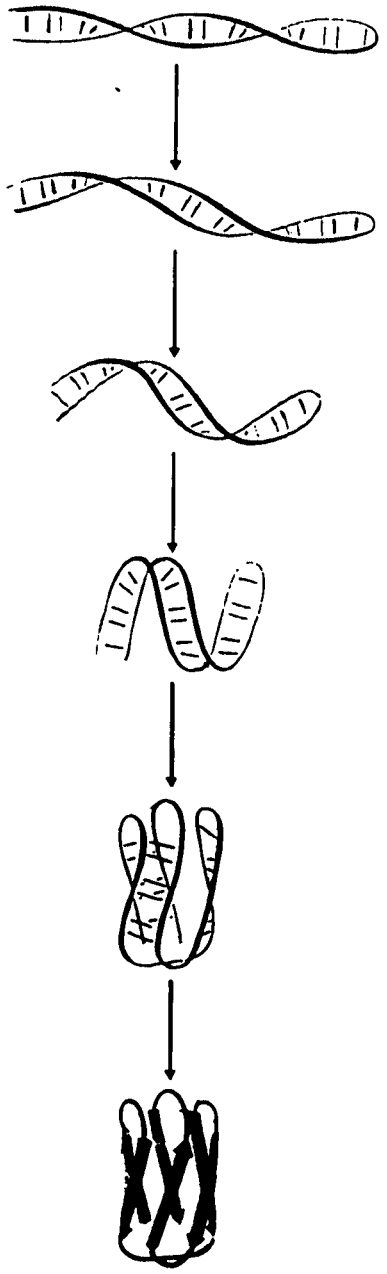
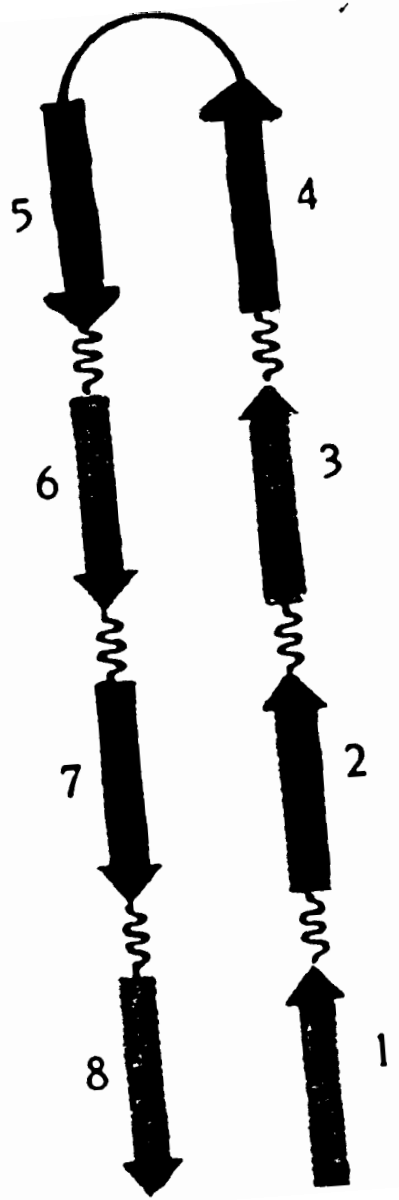
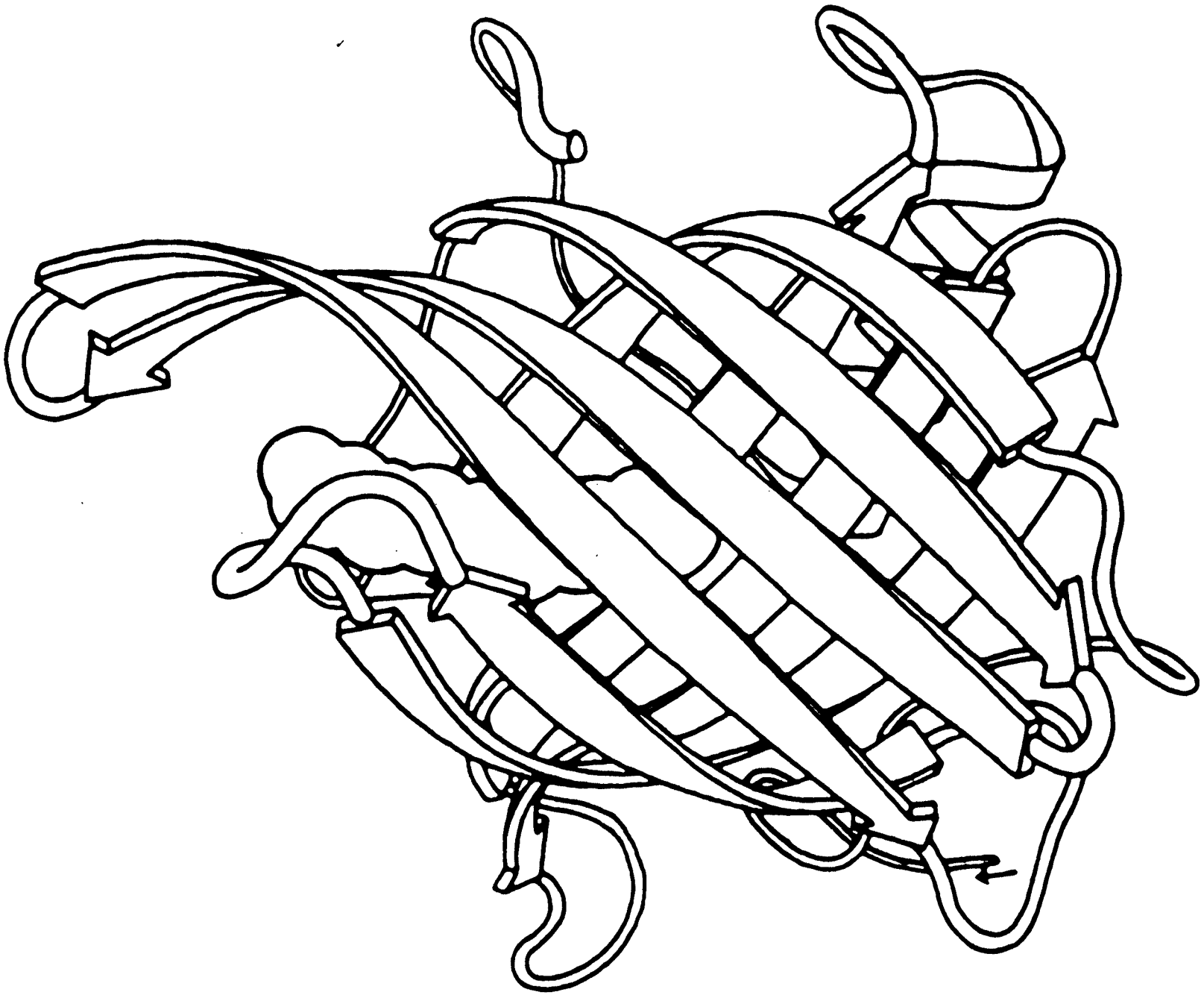


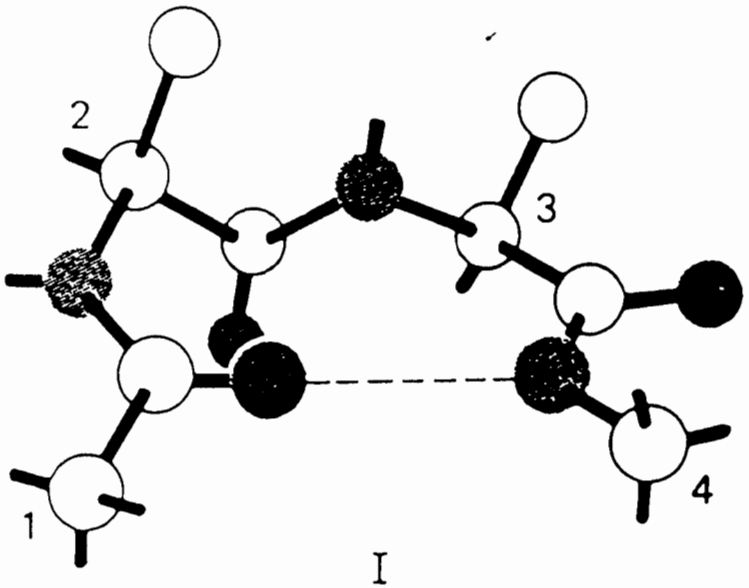
Figure 5.19 The jelly roll motif may fold from a long antiparallel hairpin structure that serves as a relatively stable folding intermediate. This hairpin structure can then twist and curl as shown in the diagram to form the jelly roll structure of six strands in a way similar to the illustration in Figure 5.17. (Adapted from J. Richardson.)

(a)

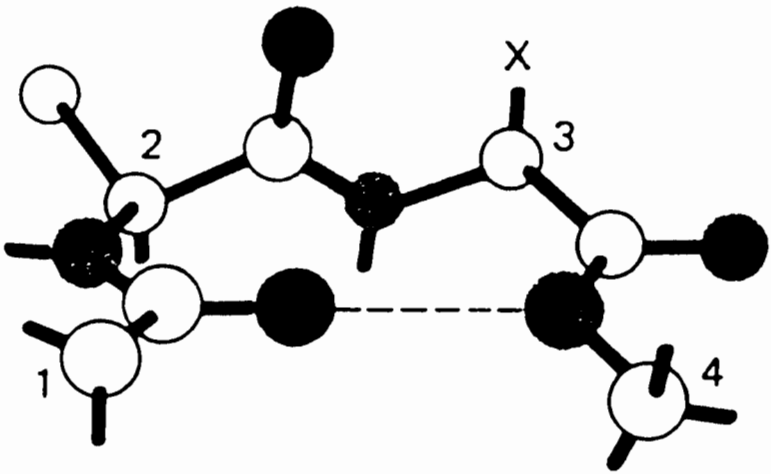


(b)

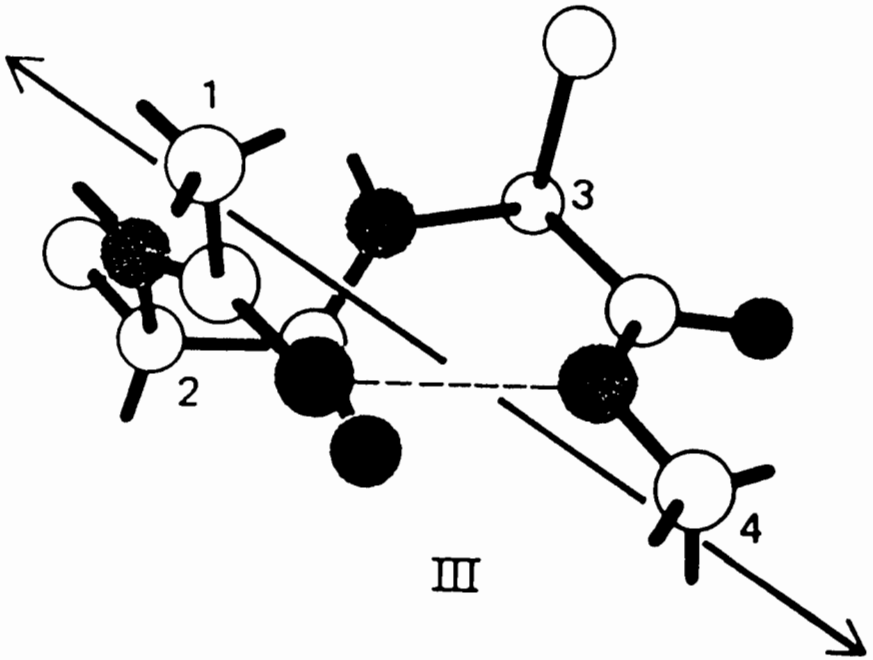




I



II



III

Turn

	<----i+1---->		<----i+2---->	
	ϕ	ψ	ϕ	ψ
β-turns				
Type I	**-60**	**-30**	**-90**	**0**
Type I'	**60**	**30**	**90**	**0**
Type II	**-60**	**120**	**80**	**0**
Type II'	**60**	**-120**	**-80**	**0**
Type III	**-60**	**-30**	**-60**	**-30**
Type III'	**60**	**30**	**60**	**30**
Type VIa *(cis)*	**-60**	**120**	**-90**	**0**
Type VIb *(cis)*	**-120**	**120**	**-60**	**0**

γ turns

turn	**70 to 85**	**-60 to -70**
inverse turn	**-70 to -85**	**60 to 70**

I common; C_α are non planar
II Gly in residue 3 (~60% have Gly); C_α are planar
III compare to 3_{10} helix
VI *cis* Pro turn

•buried turns

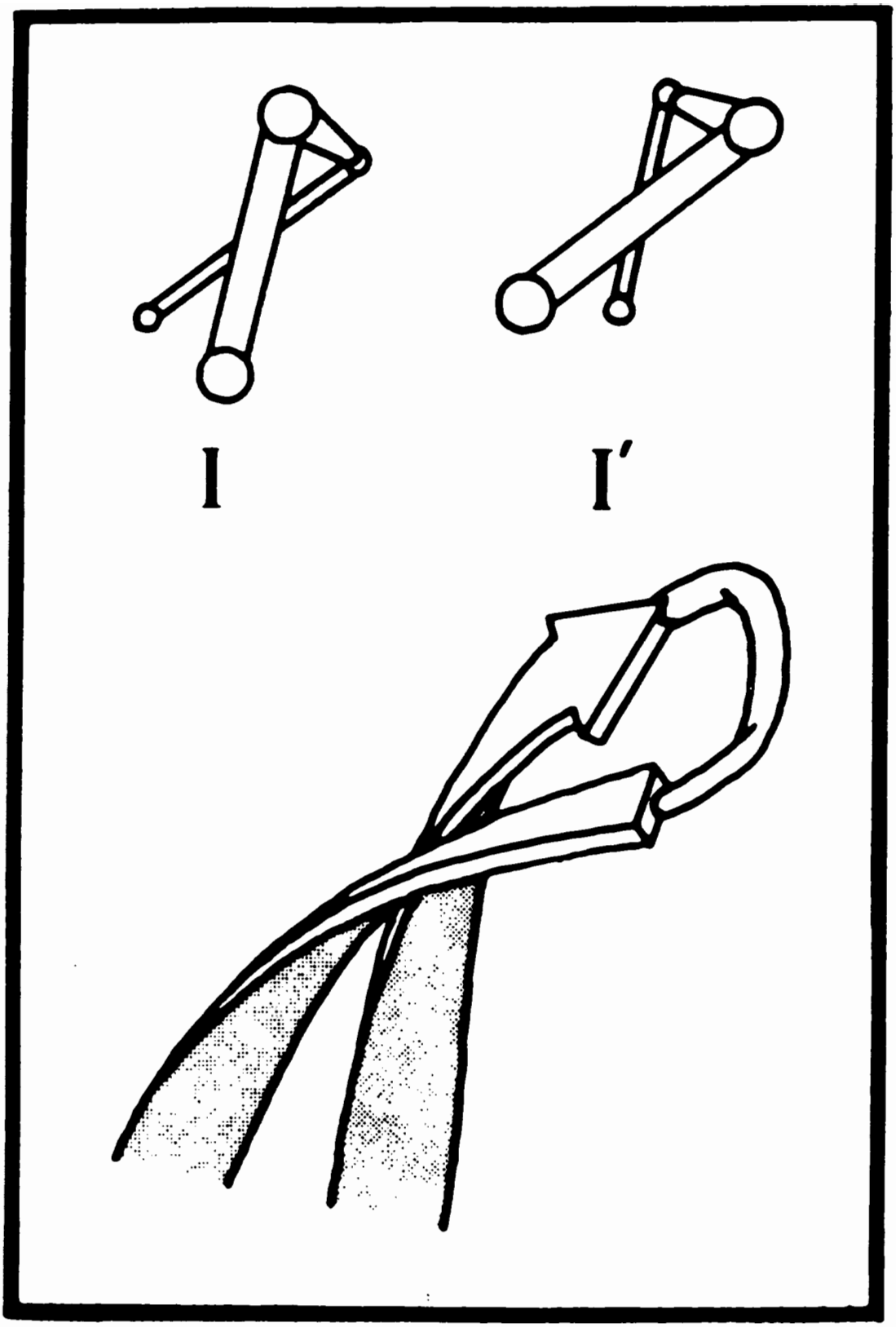
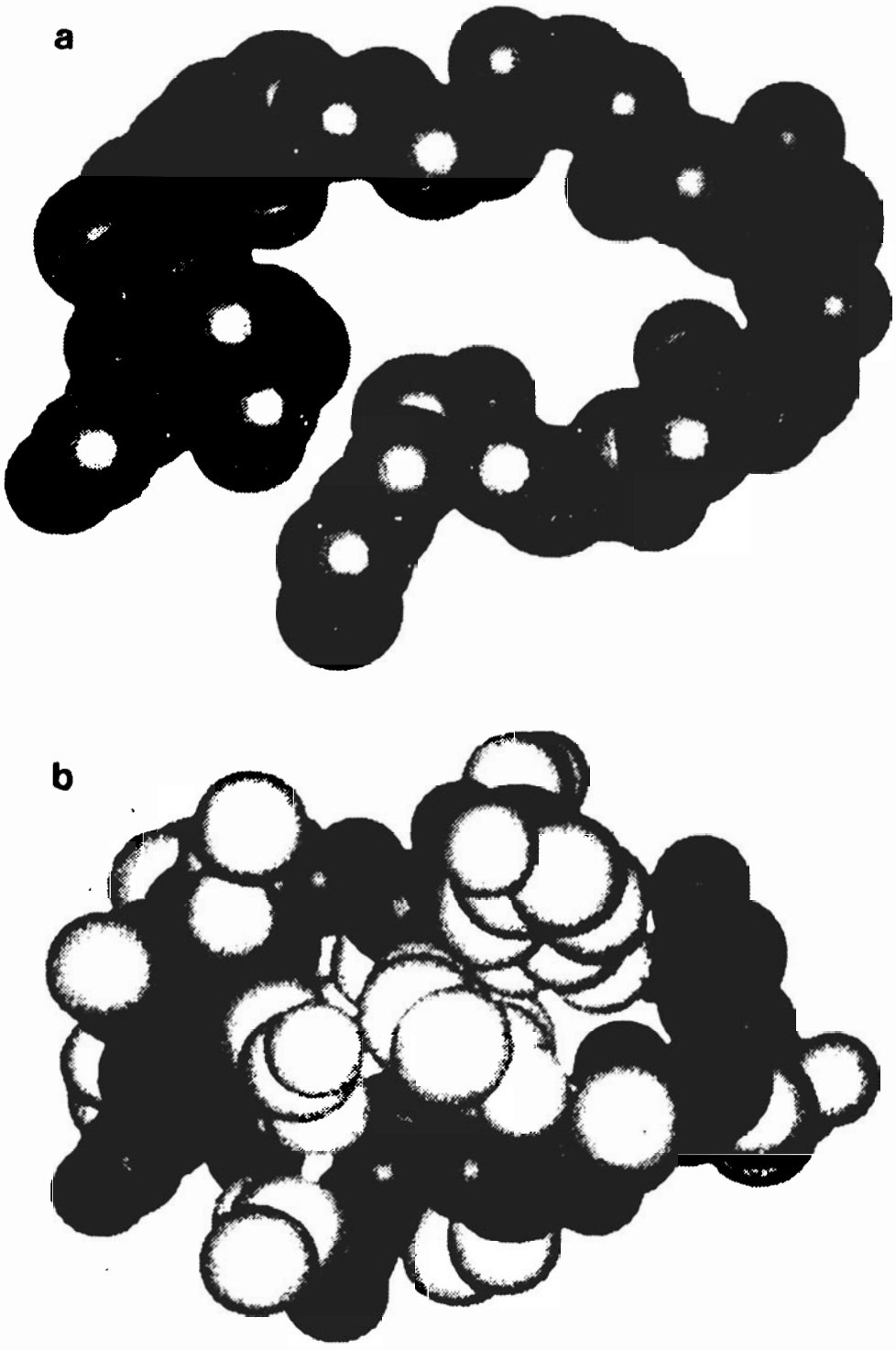


FIGURE 16 Simplified drawings of Type I vs. Type I′ tight turns, showing that Type I has a twist incompatible with the twist direction of β hairpins.

Compact Ω–Loops in Proteins



Cytochrome C, Residues 40–54

Frequency of helical, β, and coil residues in 15 proteins with their conformational parameters P_α, P_β, and P_c

Amino acid	f_α	P_α	f_β	P_β	f_c	P_c
Ala	0.522	1.45	0.167	0.97	0.311	0.66
Arg	0.282	0.79	0.154	0.90	0.564	1.20
Asn	0.263	0.73	0.113	0.65	0.624	1.33
Asp	0.351	0.98	0.137	0.80	0.514	1.09
Cys	0.278	0.77	0.222	1.30	0.500	1.07
Gln	0.421	1.17	0.211	1.23	0.368	0.79
Glu	0.549	1.53	0.044	0.26	0.407	0.87
Gly	0.190	0.53	0.138	0.81	0.668	1.42
His	0.446	1.24	0.122	0.71	0.432	0.92
Ile	0.358	1.00	0.274	1.60	0.368	0.78
Leu	0.480	1.34	0.209	1.22	0.311	0.66
Lys	0.383	1.07	0.126	0.74	0.491	1.05
Met	0.429	1.20	0.286	1.67	0.286	0.61
Phe	0.402	1.12	0.219	1.28	0.378	0.81
Pro	0.212	0.59	0.106	0.62	0.682	1.45
Ser	0.282	0.79	0.124	0.72	0.594	1.27
Thr	0.295	0.82	0.205	1.20	0.494	1.05
Trp	0.409	1.14	0.203	1.19	0.386	0.82
Tyr	0.220	0.61	0.220	1.29	0.560	1.19
Val	0.409	1.14	0.282	1.65	0.309	0.66

Classification of amino acid residues as formers, breakers, or indifferent for α-helical and β-sheet regions

Helical residues	P_α	Classification	β-Sheet residues	P_β	Classification
$Glu^{(-)}$	1.53	H_α	Met	1.67	H_β
Ala	1.45	H_α	Val	1.65	H_β
Leu	1.34	H_α	Ile	1.60	H_β
$His^{(+)}$	1.24	h_α	Cys	1.30	h_β
Met	1.20	h_α	Tyr	1.29	h_β
Gln	1.17	h_α	Phe	1.28	h_β
Trp	1.14	h_α	Gln	1.23	h_β
Val	1.14	h_α	Leu	1.22	h_β
Phe	1.12	h_α	Thr	1.20	h_β
$Lys^{(+)}$	1.07	I_α	Trp	1.19	h_β
Ile	1.00	I_α	Ala	0.97	I_β
$Asp^{(-)}$	0.98	i_α	$Arg^{(+)}$	0.90	i_β
Thr	0.82	i_α	Gly	0.81	i_β
Ser	0.79	i_α	$Asp^{(-)}$	0.80	i_β
$Arg^{(+)}$	0.79	i_α	$Lys^{(+)}$	0.74	b_β
Cys	0.77	i_α	Ser	0.72	b_β
Asn	0.73	b_α	$His^{(+)}$	0.71	b_β
Tyr	0.61	b_α	Asn	0.65	b_β
Pro	0.59	B_α	Pro	0.62	b_β
Gly	0.53	B_α	$Glu^{(-)}$	0.26	B_β

NOTE: H = strong former; h = former; I = weak former; i = indifferent; b = breaker; B = strong breaker.

SOURCE: P. Y. Chou and G. D. Fasman, *Biochemistry* 13:222 (1974).

Table II. Directional Information Values: $I(S_j = X:\bar{X};R_{j+m})$ for Helical Conformation $X = H$

Amino acid	Residue position																
	j−8		*j*−6		*j*−4		*j*−2		*j*		*j*+2		*j*+4		*j*+6		*j*+8
Gly	0	0	−5	−10	−20	−30	−50	−70	−100	−70	−50	−30	−20	−10	−5	0	0
Ala	0	5	15	20	30	40	45	55	60	55	45	40	30	20	15	5	0
Val	0	0	0	0	−5	−15	−10	−5	0	−5	−10	−10	−5	0	0	0	0
Leu	0	0	5	5	5	10	15	20	25	20	15	10	5	5	5	0	0
Ile	0	0	0	0	0	0	5	10	15	10	5	0	0	0	0	0	0
Ser	0	0	−5	−10	−15	−20	−25	−30	−35	−33	−30	−25	−20	−15	−10	−5	−5
Thr	0	−5	−10	−15	−20	−25	−35	−40	−45	−40	−35	−25	−20	−15	−10	−5	0
Asp	0	0	0	0	10	15	10	5	0	−10	−25	−35	−25	−10	−5	0	0
Glu	25	28	32	35	44	50	60	65	70	55	40	20	10	5	5	0	0
Asn	0	0	0	0	−5	−10	−20	−30	−42	−30	−20	−10	−5	0	0	0	0
Gln	0	0	0	0	0	0	0	0	15	20	25	25	20	10	0	0	0
Lys	5	5	5	5	5	5	10	15	28	40	55	58	50	45	42	40	40
His	0	0	0	0	0	0	0	0	5	15	25	32	36	40	42	43	44
Arg	0	0	−10	−18	−20	−7	10	20	30	36	36	30	20	10	0	0	0
Phe	0	0	0	0	5	15	22	28	30	28	22	15	5	0	0	0	0
Tyr	0	−3	−10	−15	−22	−30	−35	−47	−35	−20	−10	−10	−10	−12	−17	−25	−45
Trp	25	40	30	20	10	42	42	33	20	12	5	3	0	0	0	0	0
Cys	−15	−15	−15	−15	−15	−12	−10	−5	0	0	−10	−25	−35	−43	−46	−47	−37
Met	5	5	8	10	20	30	38	45	50	48	45	42	35	30	20	15	10
Pro	0	−5	−12	−20	−25	−35	−50	−66	−96	−177	−118	−90	−70	−52	−40	−30	−20

Table III. Directional Information Values $I(S_j = X:\bar{X};R_{j+m})$ for Extended Conformation $X = E$

Amino acid	Residue position																
	$j-8$		$j-6$		$j-4$		$j-2$		j		$j+2$		$j+4$		$j+6$		$j+8$
Gly	5	12	22	35	40	33	22	−10	−45	−10	22	33	40	35	22	12	5
Ala	0	0	0	−5	−10	−15	−20	−30	−40	−30	−20	−15	−10	−5	0	0	0
Val	0	0	0	5	10	20	45	70	90	70	45	20	10	5	0	0	0
Leu	0	0	0	0	0	0	10	25	33	25	5	−15	−25	−25	−20	−10	−5
Ile	−15	−25	−40	−15	0	15	35	60	70	60	35	15	0	−15	−40	−25	−15
Ser	25	25	25	22	18	15	5	0	−5	0	5	15	18	22	25	25	25
Thr	5	5	5	7	10	13	21	35	28	22	18	15	15	15	15	15	15
Asp	0	0	0	0	0	−15	−55	−105	−55	−15	0	0	0	0	0	0	0
Glu	−10	−10	−15	−20	−27	−35	−45	−55	−65	−77	−63	−50	−40	−30	−20	−15	−10
Asn	5	12	20	20	15	−10	−45	−90	−60	−30	−5	10	25	30	35	30	20
Gln	15	20	20	10	5	0	−5	−15	−20	−35	−20	−15	0	10	15	20	20
Lys	0	0	0	0	−5	−12	−23	−35	−53	−70	−58	−45	−37	−30	−25	−20	−15
His	0	−5	−15	−5	0	0	0	0	0	0	5	15	25	15	5	0	0
Arg	0	0	0	0	0	0	0	0	−5	−10	−20	−25	−28	−25	−10	−5	0
Phe	−20	−35	−60	−60	−45	−30	0	25	40	25	0	−30	−45	−60	−60	−35	−20
Tyr	0	0	0	0	7	15	27	40	40	27	15	0	0	0	0	0	0
Trp	−15	−25	−40	−45	−80	−15	0	10	15	17	20	20	20	20	20	20	20
Cys	0	0	−20	−60	−55	−40	−20	5	15	17	10	5	0	0	0	0	0
Met	−20	−65	−90	−80	−60	−30	−5	15	30	15	−15	−45	−50	−45	−40	−30	−25
Pro	20	20	20	20	10	0	−30	−65	−110	−65	−30	0	10	20	20	20	20

Table IV. Directional Information Values $I(S_j = X:\bar{X};R_{j+m})$ for β Turn X = T

Amino acid	Residue position																
	$j-8$		$j-6$		$j-4$		$j-2$		j		$j+2$		$j+4$		$j+6$		$j+8$
Gly	0	0	0	−5	−10	−20	−30	0	70	80	10	0	0	0	0	0	0
Ala	0	0	0	0	0	−3	−5	−10	−15	−22	−28	−32	−20	−7	−3	0	0
Val	0	0	0	0	−5	−10	−30	−50	−95	−50	−30	−10	−5	0	0	0	0
Leu	0	0	0	0	0	−3	−10	−15	−50	−15	−10	−3	0	0	0	0	0
Ile	0	0	0	−5	−20	−30	−50	−80	−140	−80	−30	−15	0	0	0	0	0
Ser	0	0	0	0	0	0	15	30	40	30	5	−15	−30	−15	−5	0	0
Thr	0	0	−5	−8	−10	−15	−18	−25	−42	−15	10	20	10	5	0	0	0
Asp	0	0	0	0	5	10	30	35	30	15	5	0	0	0	0	0	0
Glu	0	0	0	0	−5	−10	−10	25	20	−5	−10	20	20	5	0	0	0
Asn	0	0	−5	−20	−5	10	30	45	55	55	30	15	0	−10	−20	−5	0
Gln	0	0	0	0	0	0	0	0	0	0	0	−25	−60	−25	−15	−10	−15
Lys	0	0	0	0	0	3	5	18	30	25	10	−10	−25	−20	−12	−5	0
His	0	0	0	3	10	25	38	5	−20	15	15	−5	10	35	15	0	0
Arg	0	0	0	5	10	25	35	20	0	−35	−5	0	0	0	0	0	0
Phe	0	10	20	25	20	10	−15	−55	−50	−40	−30	−15	0	10	20	25	0
Tyr	0	0	0	5	10	20	30	5	−25	−10	0	15	30	40	30	15	0
Trp	0	−15	−45	15	35	25	10	−15	−57	−20	−10	−5	0	0	0	0	0
Cys	0	0	5	30	50	60	40	5	10	25	25	10	45	30	10	0	0
Met	0	0	0	0	0	−5	−25	−55	−85	−55	−25	−5	0	0	0	0	0
Pro	0	0	0	0	0	0	20	54	40	−154	40	50	20	0	0	0	0

Table V. Directional Information Values $I(S_j = X:\bar{X};R_{j+m})$ for Coil X = C

Amino acid	Residue position																
	$j-8$		$j-6$		$j-4$		$j-2$		j		$j+2$		$j+4$		$j+6$		$j+8$
Gly	0	0	0	0	0	10	35	70	45	20	10	5	0	0	0	0	0
Ala	0	0	0	0	−5	−10	−15	−20	−25	−20	−15	−10	−5	0	0	0	0
Val	0	0	0	0	−5	−10	−15	−25	−45	−25	−15	−10	−5	0	0	0	0
Leu	0	0	0	0	−5	−20	−30	−37	−35	−25	−15	−5	0	0	0	0	0
Ile	0	0	0	0	0	−10	−25	−32	−32	−25	−10	−5	0	0	0	0	0
Ser	0	0	0	0	0	0	5	10	20	10	5	0	0	0	0	0	0
Thr	0	0	0	0	0	0	5	15	25	15	5	0	0	0	0	0	0
Asp	0	0	0	0	0	0	5	20	35	45	40	30	15	5	0	0	0
Glu	−5	−7	−10	−20	−25	−35	−40	−45	−40	−15	−5	0	0	0	0	0	0
Asn	0	0	0	0	0	5	20	35	40	35	20	5	0	0	0	0	0
Gln	0	0	0	0	5	10	20	0	0	0	0	0	0	0	0	0	0
Lys	0	−7	−12	−18	−7	0	8	0	−8	−15	−30	−20	−12	−10	−8	−5	0
His	0	0	0	0	0	0	0	−5	−10	−15	−25	−40	−55	−40	−30	−25	−20
Arg	0	0	0	0	−5	−10	−20	−30	−35	−30	−20	−10	−5	0	0	0	0
Phe	0	0	0	0	−5	−10	−20	−30	−35	−30	−20	−10	−5	0	0	0	0
Tyr	0	0	10	23	10	0	0	0	0	0	0	0	5	18	5	0	0
Trp	0	0	−5	−20	−35	−43	−40	−35	−30	−23	−15	−5	0	0	0	0	0
Cys	15	15	20	30	20	10	5	−5	−15	−25	−10	0	10	20	30	40	25
Met	0	0	0	0	−5	−10	−15	−25	−40	−55	−30	−20	−10	−5	0	0	0
Pro	0	0	0	5	15	25	40	60	90	202	90	40	25	15	10	5	0

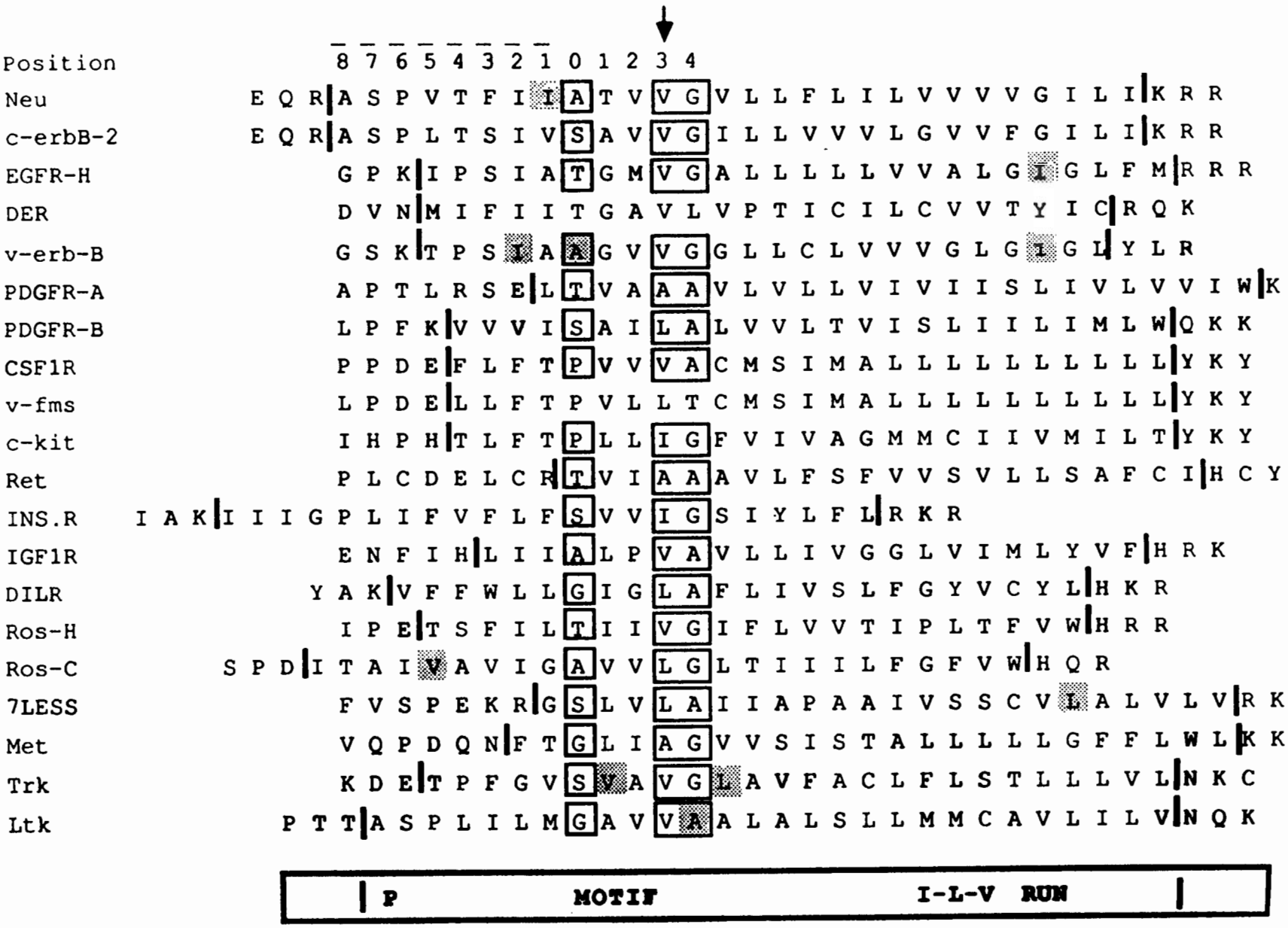
```
position 1 2 3 4 5 6 7 8

         G L S G S G K S  E.coli UvrA
         A G A G S G K T  E.coli UvrD
         G P H G M G K T  HSV Thymidine Kinase
         G P E S S G K T  E.coli RecA
         P S P G T G K T  Phage T4 Gene 44
```

Fig. 1. Octapeptides from five out of about 120 known examples of this motif, known as the Walker A-consensus, have been aligned. Many of these proteins are known to bind nucleoside triphosphates (Walker *et al.*, 1982; Gorbalenya *et al.*, 1985; Higgins *et al.*, 1988). The sequence positions have been marked above the alignment, and the names of the parent proteins on the right.

```
EBV            YVGVLTDGKTL-MKGV
HSV            YIGVIYGGKML-IKGV
VZV            YIGVIYGGKVL-MKGV
CMV            YIGKVEGASGLSMKGV
Phage Phi29    YLRQKTYIQDIYMKEV
```

Fig. 3. The above alignment is from a family of herpes virus DNA polymerases (Davison and Scott, 1986; Kouzarides *et al.*, 1987), and falls between conserved regions which are also found in other virus and cellular DNA polymerases (Wong *et al.*, 1988) – a bacteriophage segment is shown for comparison. The underlined part of the Epstein–Barr Virus sequence matches the Walker A-consensus. However, it is not conserved. The equivalent regions suggest this is merely a loop between two functionally significant parts of the molecule, and is even seven residues longer in the polymerases not shown.

Table I. The sequence pattern in the transmembrane domain of tyrosine kinase growth factor receptors



Rosetta Uses a Fragment Library + Monte Carlo Search

Examples of the best-center cluster found by *Rosetta* for some test proteins. In many cases the overall fold is predicted well enough to be recognizable. However, relative positions of the secondary structure elements are almost always shifted somewhat from their correct values.