PROTEINS: Structure, Function, and Genetics 7:125–155 (1990)

Polyproline, β-Turn Helices. Novel Secondary Structures Proposed for the Tandem Repeats Within Rhodopsin, Synaptophysin, Synexin, Gliadin, RNA Polymerase II, Hordein, and Gluten

Norio Matsushima,[1] Carl E. Creutz,[2] and Robert H. Kretsinger[1]
Departments of [1]Biology and [2]Pharmacology, University of Virginia, Charlottesville, Virginia 22903

ABSTRACT **Seven proteins each contain 8 to 52 tandem repeats of a unique class of oligopeptide. The consensus peptide for each is**

rhodopsin	Tyr Pro Pro Gln Gly
synaptophysin	Tyr Gly Pro Gln Gly
synexin	Tyr Pro Pro Pro Pro Gly
gliadin	Tyr Pro Pro Pro Gln Pro
RNA polymerase II	Tyr Ser Pro Thr Ser Pro Ser
hordein	Phe Pro Gln Gln Pro Gln Gln Pro
gluten	Tyr Pro Thr Ser Pro Gln Gln Gly Tyr

Although there is obvious variation of sequence and of length, the penta- to nonapeptides share an initial Tyr (or Phe) and have high Pro contents and abundant Gly, Gln, and Ser. We have evaluated helical models that both recognize the uniqueness of these sequence repeats and accommodate variations on the basic theme.

We have developed a group of related helical models for these proteins with about three oligopeptide repeats per turn of 10–20 Å. These models share several common features: Most of the ϕ dihedral angles are −54°, to accommodate Pro at all positions except the first (Tyr). Except for the β-turns, most ψ dihedral angles are near +140° as found in polyproline. Each oligopeptide has at least one β-turn; several have two. Some contain a *cis*-Tyr, Pro peptide bond; a few have a *cis*-bond plus one β-turn. Tyr side chains vary from totally exposed to buried within the helices and could move to accommodate either external hydrophobic interactions or phosphorylation. The several related structures seem to be readily interconverted without major change in the overall helical parameters, and therein may lie the key to their functions.

Key words: phosphorylation, tyrosine, *cis*-peptide, exocytosis

INTRODUCTION

During the past 3 years several protein sequences have been noted to contain unique tandem repeats having either a Tyr or Phe and 20–60% Pro with abundant Gly, Gln, and Ser. We have developed a group of related helical models for seven proteins based on the polyproline conformation combined with one or two β-turns. We first review these proteins, then discuss the rationale for developing the models, and finally describe the models.

Creutz et al.[1] noted a repeating hexapeptide motif in synexin and similar tandem repeats in gliadin and in synaptophysin (Table 1). Synexin and synaptophysin are mammalian proteins of relatively low abundance that may play calcium-dependent regulatory roles in the process of exocytosis. Synexin is a soluble, 47 kDa protein that binds to secretory vesicle membranes in the presence of calcium.[2] The protein also has a bivalent action in that it promotes the formation of contacts between two membranes. These membrane contacts are subject to fusion in the presence of physiological concentrations of free arachidonic acid.[3] For these reasons synexin has been postulated as a mediator of membrane fusion in exocytosis.

Synaptophysin, 38 kDa, is a major integral membrane protein of synaptic vesicles.[4,5] The bulk of the protein has been suggested to be arranged as four transmembrane helices, while the domain analyzed here is at the carboxy terminus and is present on the cytoplasmic side of the membrane. This domain is thought to be responsible for the binding of calcium by synaptophysin[6] and is also likely to be the site of Tyr phosphorylation that occurs due to endogenous Tyr kinase(s) in the vesicle membrane.[7] Although the role of the phosphorylation event is not known,

Received May 22, 1989; revision accepted September 18, 1989.

Address reprint requests to Dr. Robert H. Kretsinger, Department of Biology, The University of Virginia, Gilmer Hall, Charlottesville, VA 22903.

Norio Matsushima's permanent address is School of Allied Health Professions, Sapporo Medical College, Chuo-ku, Sapporo, Hokkaido 060, Japan.

TABLE I. Polyproline, β-Turn Sequence Data*

Rhodopsin

Y	P	P	Q	G				
Y	Q	P	P	P	P	P	Q	G
Y	P	P	P	G				
Y	P	P	Q	G	A			
Y	P	P	P	Q	G			
Y	P	P	Q	G				
Y	P	P	Q	G				
Y	P	P	Q	G				
Y	P	P	Q	G	A	P	P	

Synaptophysin

Y	G	P	Q	G				
Y	G	D	A	G				
Y	G	D	G	P	G	G		
Y	G	P	Q	D	S			
Y	G	P	Q	G	G			
Y	Q	P	D					
Y	G	Q	P	A				
S	G	G	G	G				
Y	G	P	Q	G	D			
Y	G	Q	Q	G				
Y	G	Q	Q	G				

Synexin

Y	P	P	P	P	G			
F	P	P	P	G	Q			
Y	P	Y	P	S	G			
F	P	P	M	G	G	G	A	
Y	P	P	A	P	S	S	G	
Y	P	G	A	G	G			
Y	P	A	P	G	G			
Y	P	A	P	G	G			
Y	P	G	A	P	Q			
Y	P							

Gliadin

Gene	Y	P	P	P	Q	P		
A42	F	P	G	Q	Q	Q		Q
A26		L						P
A212		Q						P
A735								P
A1235						E		
A42	F	P	P	Q	Q	P		
A26								
A212								
A735								
A1235								
A42	Y	P	Q	P	Q	P		
A26								
A212								
A735				Q				
A1235			H	Q				
A42	F	P	S	Q	Q	P		
A26								
A212								
A735								
A1235								
A42	Y	L	Q	L	Q	P		
A26								
A212								
A735		M						
A1235		P		P				
A42	F	P	Q	P	Q	P		
A26		L				—		
A212						—		
A735						—		
A1235						—		
A42	F	P	P	Q	L	P		
A26	—	—	—	—				
A212	—	—	—	—				
A735	—	—	—	—				
A1235			—	—				
A42	Y	P	Q	P	Q	—		—
A26		S				—		—
A212						—		—
A735				T		L		P
A1235						—		—
A42	—	—	—	—	—	S		
A26	—	—	—	—	—	P		
A212	—	—	—	—	—	P		
A735	Y	P	Q	P	Q	P		
A1235	—	—	—	—	—	P		
A42	F	P	P	Q	Q	P		
A26		R						
A212		R						
A735		R				S		
A1235								
A42	Y	P	Q	Q	Q	P		Q
A26				P				
A212				P				
A735				P				
A1235						P		
A42	Y	L	Q	P	Q	Q		P
A26		S						
A212		S						
A735		S						
A1235		P						

C-hordein

Y	P	V	Q	P	Q	Q	S
				P	Q	Q	S
Y	P	V	Q	P	Q	Q	P
(F	P	Q	—	P	Q	—	P)
(V	P	Q	Q	P	Q	Q	A)
(S	P	L	Q	P	Q	Q	P)
F	P	Q	G	S	E	Q	I
(I	—	—	—	P	Q	Q	P)
F	P	L	Q	P	Q	—	P
F	P	Q	Q	P	Q	Q	P
L	P	Q	—	P	Q	Q	P
F	R	Q	Q	A	E	L	I
I	P	Q	Q	P	Q	Q	P
L	P	L	Q	P	H	Q	P
Y	T	Q	Q				

B-hordein domain 1

V	S	R	Q	P	Q	Q	I
I	P	Q	Q	P	Q	Q	P
F	P	L	Q	—	—	—	—
—	P	Q	Q	P	Q	—	P
F	—	—	—	P	Q	Q	P
I	P	Q	Q	P	Q	—	P
Y	P	Q	Q	P	Q	—	S
F	—	—	—	P	Q	Q	P
F	P	—	—	S	Q	Q	P
F	P	Q	Q	P	—	—	P
F	W	Q	Q				

RNA Polymerase II

	Y	S	P	T	S	P	S
			Saccharomyces				
	Y	G	G	V	D		
	Y	G	E	A	T	S	P
	F	A	A				

1	Y	G	E	A	P	T	S	P	G
2	F	G	V	S	S	P	G		
3	F	S	P	T	S	P	T		
4	Y	S	P	T	S	P	A		
11	Y	S	P	M	S	A	S		
18	Y	S	P	T	S	P	A		
23	Y	S	P	T	S	P	N		
25	Y	S	P	T	S	P	G		
26	Y	S	P	G	S	P	A		
27	Y	S	P	K	Q	D	E		

Drosophila

a	M	T	P	G	G	P	S					
b	F	S	P	S	A	A	S	D				
c	A	S	G	M	S	P	S					
d	W	S	P	A	H	P	G					
e	S	S	P	S	S	P	G	P	S	M	S	P
f	Y	F	P	A	S	P	S	V	S	P	S	
1	Y	S	P	T	S	P	N					
2	Y	T	A	S	S	P	G	G	A	S	P	N
3	Y	S	P	S	S	P	N					
4	Y	S	P	T	S	P	L	Y	A	S	P	R
5	Y	A	S	T	T	P	N					
6	F	N	P	Q	S		T	G				
7	Y	S	P	S	S	S	G					
8	Y	S	P	T	S	P	V					
9	Y	S	P	T	V		Q					
10	F	Q		S	S	P	S					
11	F	A	G	S	G	S	N	I				
12	Y	S	P	G	N	A						
13	Y	S	P	S	S	S	N					
14	Y	S	P	N	S	P	S					
15	Y	S	P	T	S	P	S					
16	Y	S	P	S	S	P	S					
17	Y	S	P	T	S	P	G					
18	Y	S	P	T	S	P	S					
19	Y	S	P	T	S	P	N					
20	Y	T	P	V	T	P	S					
21	Y	S	P	T	S	P	N					
22	Y	S		A	S	P	Q					
23	Y	S	P	A	S	P	A					
24	Y	S	Q	T	G	V	K					
25	Y	S	P	T	S	P	T					
26	Y	S	P	P	S	P	S					
27	YD	GS	P	G	S	P	Q					
28	Y	T	P	G	S	P	Q					
29	Y	S	P	A	S	P	K					
30	Y	S	P	T	S	P	L					
31	Y	S	P	S	S	P	Q					
32	H	S	P	S	N		Q					
33	Y	S	P	T	G	S	T					
34	Y	S	A	T	S	P	R					
35	Y	S	P	N	M	S	I					
36	Y	S	P	S	S	T	K					
37	Y	S	P	T	S	P	T					
38	Y	T	P	T	A	R	N					
39	Y	S	P	T	S	P	M					
40	Y	S	P	T	A	P	I	H				
41	Y	S	P	T	S	P	A					
42	Y	S	P	S	S	P	T					
	F	E	E	S	E	D	V					
	R	K	G	G	R	G						

Mus

1	Y	S	P	T	S	P	A
2	Y	E	P	R	S	P	G G
3	Y	T	P	Q	S	P	S
6	Y	S	P	T	S	P	N
22	Y	S	P	T	S	P	N
23	Y	S	P	T	S	P	N
24	Y	T	P	T	S	P	S
26	Y	S	P	T	S	P	N
27	Y	T	P	T	S	P	N
31	Y	S	P	S	S	P	R
32	Y	T	P	Q	S	P	T
33	Y	T	P	S	S	P	S
34	Y	S	P	S	S	P	S
35	Y	S	P	T	S	P	K
36	Y	T	P	T	S	P	S
37	Y	S	P	S	S	P	E
38	Y	T	P	A	S	P	K
39	Y	S	P	T	S	P	K
40	Y	S	P	T	S	P	K
41	Y	S	P	T	S	P	T
42	Y	S	P	T	T	P	K
43	Y	S	P	T	S	P	T
44	Y	S	P	T	S	P	V
45	Y	T	P	T	S	P	K
46	Y	S	P	T	S	P	T
47	Y	S	P	T	S	P	K
48	Y	S	P	T	S	P	T
49	Y	S	P	T	S	P	K G S T
50	Y	S	P	T	S	P	G
51	Y	S	P	T	S	P	T
52	Y	S	L	T	S	P	A

*The tandem repeats of oligopeptides for each of the seven proteins are listed beneath their respective consensus sequences. Rhodopsin: Ovchinnikov et al.[8] Synaptophysin: Südhoff et al.[4] Synexin: Creutz et al.[1]. Gliadin: Okita et al.[14] Amino acid sequences are deduced from five cDNA clones (A42 to A1235) for the a-/b- type. Blank indicates the same amino acid as in A42;—indicates deletion. RNA polymerase II: repeats, as 7–21 in *Mus*, identical to the consensus are not listed: Allison et al.,[9] *Saccharomyces;* Corden et al.,[10] *Mus;* Zehring et al.,[12] *Drosophila.* Hordein: Forde et al.[15]—indicates deletion. The sequences in parentheses did not contribute to the consensus but are included to indicate the range of variation that often precedes, or follows, the more strongly conserved regions. Gluten: Halford et al.[18]

it obviously adds an element of interest to the placement and structural role of the Tyr's presented in the models in this communication. The presence of this domain on the cytoplasmic side of the vesicle membrane suggests that it may be involved in contacts with cytoskeletal or regulatory proteins during the life cycle of the vesicle, e.g., in vesicle translocation, fusion with the plasma membrane, or membrane recovery in clathrin coated pits.

The rhodopsin of the octopus, *Paroctopus defleini,* has structural and functional properties similar to the rhodopsins of *Drosophila* and of mammals. From analysis of the amino acid sequence Ovchinnikov et al.[8] suggested that it also contains seven transmembrane helices with two potential glycosylation sites near the N-terminus. The octopus rhodopsin differs from insect and vertebrate rhodopsins in having an extended tail of about a hundred amino acids, which like that of synaptophysin extends into the cytosol. This C-terminus contains eight Tyr·Pro·Pro·Gln·Gly consensus repeats in tandem (Table I).

RNA polymerases contain multiple copies of the consensus heptapeptide Tyr·Ser·Pro·Thr·Ser·Pro·Ser. Allison et al.[9] determined the nucleotide se-

quences two genes, called II and III, from *Saccharomyces cerevisiae* that encode the largest subunits of RNA polymerase. The encoded proteins are homologous to each other as well as to the largest subunit of RNA polymerase, β′, and DNA polymerase I of *E. coli*. The largest subunit, M_r 190,986 of yeast RNA polymerase II, encoded by gene RP021, has 27 tandem repeats of the heptapeptide running to within 10 residues of its C-terminus. Selective proteolysis of RNA polymerase II_A removes this C-terminal domain reducing the M_r of RNA polymerase II_B to *170,620*. The apparent molecular weights based on electrophoretic mobility are 220,000 and 180,000. Allison et al. argue that this C-terminal domain is "required for specific, but not randomly initiated, transcription in vitro." In Table I we list the consensus sequence and only the 10, of total 27 repeats, that differ from the consensus.

Corden et al.[10] inferred the amino acid sequence of the largest or unproteolyzed subunit of mouse RNA polymerase II_α and found 52 repeats of the same Tyr·Ser·Pro·Thr·Ser·Pro·Ser heptapeptide at the C-terminus. They note the low probability of α-helix or β-sheet and suggest, "A possible structure of the C-terminal domain might be comprised of alternating β-turns of 3 and 4 residues." They also note that the Tyr,Ser, and Thr are potential phosphorylation sites. Interestingly, both RNA polymerases have hydrophilic decapeptides at their C-termini following the hydrophilic heptapeptide consensus repeats. Sigler[11] suggested that the regions of transcriptional regulatory proteins rich in anionic groups might interact with the RNA-polymerase "hitching post bristling with hydroxyl groups."

Zehring et al.[12] and Allison et al.[9] determined the sequence of the gene encoding the M_r 215,000 subunit of RNA polymerase II of *Drosophila melanogaster*. The C-terminus contains 42 tandem repeats of the heptapeptide preceded by six less obvious repeats and followed by 13 nonconsensus residues (Table I). Only 2 of the 42 honor the consensus exactly; even so, the consensus is obvious. Zehring et al.[12] constructed lethal mutants that contained only 20 heptapeptide repeats and showed that RNA polymerase functions in vitro without this unique C-terminus. They concluded "that although the unusual repeat structure found at the C-terminus of the largest subunit is essential *in vivo*, partial truncation or complete removal of the structure does not affect overall polymerase activity or eliminate accurate initiation *in vitro*."

In contrast, gliadin is an abundant storage protein found in the wheat kernal.[13] The only known function for this protein is storage of its constituent amino acids as a consequence of concentration and dehydration of the protein in the seed. Okita et al.[14] sequenced cDNAs corresponding to five distinct gliadin genes. Within the repeat regions of these sequences (Table I) some of the genes have small deletions or insertions. It is of note that most of these deletions or insertions are six residues long, as is the repeat length, so that the register is not altered.

The hordeins are found in the starchy endosperm of barley (*Hordeum vulgare*). One of the multiple cDNAs for B hordein and part of a C hordein cDNA have been sequenced.[15] C hordein consists almost entirely, as inferred by the amino acid composition of this 46,000 kD protein, of tandem octapeptide repeats of the consensus sequence—Phe·Pro·Gln·Gln·Pro·Gln·Gln·Pro (Table I). The N-terminus contains at least three copies of the pentapeptide consensus sequence—Tyr·Pro·Gln·Gln·Pro. B hordein consists of a C-terminal domain, about 2/3 of the molecule, Cys rich, and homologous to another plant storage protein, napin. The N-terminal third contains 10 tandem octapeptide repeats whose consensus is identical to that of C hordein. Based on the partial sequence information from C hordein, Tatham et al.[16] had previously proposed that the structure contains little α-helix or β-sheet. They presented CD spectra, measured in 70% ethanol required for solubility, with minima about 203 nm and a shoulder about 225 nm. They noted that "Proteins rich in β-turns may therefore be expected to contain mixtures of turn types." Field et al.[17] measured the viscosity of C hordein and proposed a rod shaped molecule "360 × 17 Å to 265 × 20 Å" . . . "having a helical secondary structure based on repetitive β-turns."

Gluten is the insoluble mass that remains when dough is washed. From the sequences of genes in wheat (*Triticum durum*) Halford et al.[18] have deduced and described the amino acid sequences of four homologous proteins ranging in length from 435 to 687 residues. Both N- and C-termini have unique, possibly globular regions; however, the central 80% consists of tandem and interspersed repeats of a nonapeptide of consensus sequence Tyr·Pro·Thr·Ser·Pro/Leu·Gln·Gln·Gly·Tyr with a few hexapeptides of consensus Pro·Gly·Gln·Gly·Gln·Gln. From a series of viscometric and CD studies Field et al.[19] proposed a model, similar to their model for hordein, in which the central polynonapeptide region has no α-helix or β-sheet, but instead a repeating series of β-turns, which form a β-spiral. The overall dimensions of the rod-shaped molecule are estimated to be 500 by 17.5 Å.

RATIONALE AND METHODS

Initial Observations and Assumptions

The models were developed honoring the following experimental observations and assumptions:

1. The protein segments listed in Table I have an obvious similarity yet vary in both length and sequence. We assume that the initial Tyr (or Phe), 20–60% Pro, and high content of Gln, Gly, and Ser reflect common themes. Some, but obviously not all, characteristics of the seven proteins are similar.

TABLE IIa. Distribution of β-Turns after Wilmot and Thornton[23]

β-Turn type	$i+1$		$i+2$		Ramachandran nomenclature	Number located	Distorted located
	$\phi°$	$\psi°$	$\phi°$	$\psi°$			
I	−60	−30	−90	0	$\alpha_R \rightarrow \alpha_R$	258	69
I′	60	30	90	0	$\alpha_L \rightarrow \gamma$	30	14
II	−60	120	80	0	$\beta \rightarrow \alpha_L$	101	107
II′	60	−120	−80	0	$\epsilon \rightarrow \alpha_R$	20	31
VIa	−60	120	−90	0	$\beta \rightarrow \alpha_R$	2	8
VIb	−120	120	−60	0	$\beta \rightarrow \alpha_R$	3	10
VIII	−60	−30	−120	120	$\alpha_R \rightarrow \beta$	37	55
IV	(none of above)					—	50
					Sum 795	451	344

2. Various groups have noted that secondary structure prediction algorithms, such as those of Chou and Fasman[20] or of Garnier,[21] indicate low contents of α-helix and β-sheet and high contents of β-turns in these proteins.

3. Circular dichroism studies of the hordeins[17] and glutens[19] indicate a high content of β-turns and very little α-helix or β-sheet.

4. Viscometry studies of both the hordeins and the glutens[19] indicate rod-shaped structures about 17 Å in diameter and length appropriate to the number of repeats of oligopeptide.

5. It has been inferred that tandem repeats of Tyr·Ser·Pro·Thr·Ser·Pro·Ser in RNA polymerase II form a protruding rod-like structure.[9,11]

6. Any given structure, for instance an oligopeptide, that is repeated, each monomer having the same relationship to its predecesor, will generate a helix.

Based on these six observations we decided to explore models in which the repeating oligopeptide has one or two β-turns and the remaining residues have $\phi = -54°$ to accommodate Pro, whose ϕ dihedral angle is fixed to $-54° \pm 10°$. In the ϕ, ψ diagram there are only two allowed regions at $\phi = -54°$, $\psi = -45° \pm 15°$ and $\psi = 135° \pm 30°$. The dihedral angles $\phi = -60°$, $\psi = -45°$ are characteristic of the α-helix; we did not use this conformation since it does not have Pro. The angles $\phi = -60°$, $\psi = +140°$ define the polyproline helix and collagen. Hence, we made our initial family of models based on combinations of short lengths of polyproline helices interspersed at regular intervals with β-turns. We used this most restrictive condition initially, since it could accommodate Pro at any position of the polyproline helix and, as will be discussed, at the first two positions of the four involved in β-turns.

Selection of Types of β-Turns

We tested models using type I, II, and III β-turns, as originally defined by Ramachandran and Sasisekharan.[22] These represent a subset of the groups of β-turns recently analyzed by Wilmot and Thornton[23] who located 795 β-turns in 59 nonidentical proteins, "using the standard criterion that the distance between $C^\alpha(i)$ and $C^\alpha(i+3)$ is less than 7 Å" and that "the central residues were not helical." "Ideal ϕ, ψ angles" . . . "were allowed to vary by ± 30°, with the added flexibility of one angle being allowed to deviate by as much as ± 45°." They classified 795 β-turns as indicated in Table IIa; of these 344 are distorted and 451 conform to their definitions. The types I′ and II′ β-turns are mirror images of types I and II and are accessible only to Gly. Types VI_a, VI_b, and VIII occur much less frequently and, when present, show a greater range of variability. They restrict their further analysis to types I and II [(258 + 101) / 451 =] 80% of the total, as did we initially. Subsequently we included in our analyses what was called the type III β-turn in the older literature ($\phi_{i+1} = -60°$, $\psi_{i+1} = -30°$ and $\phi_{i+2} = -60°$, $\psi_{i+2} = -30°$) to permit us to explore a greater range of conformation of the type I β-turn, which occurs in the sampled proteins at greater frequency than all the others. Further, we found type I β-turns to be more frequently acceptable than type II β-turns in our models. The type III β-turn is included within the Wilmot and Thornton analysis as a type I β-turn ± 30°. They confirmed that β-turns frequently occur at the surface of globular proteins and for this reason alone contain a high proportion of hydrophilic residues.

Their further analysis shows the probability of each of the 20 amino acids at the four positions of types I and II, β-turns (Table IIb). It is especially relevant to our analysis that Pro occurs with high frequency at position $i+1$ of both type I and II β-turn but is disfavored at positions $i+2$ and $i+3$. They note that Ser, as well as Asp, Asn, and Cys, shows a significant preference to be at position i of a type I turn. Further 57% of Ser, Asp, Asn, and Cys side chain oxygen or sulfur atoms make a hydrogen bond with the N-H of residue $i+2$. Many of the β-turns in the analysis of Wilmot and Thornton did not form an $i+3$ to i H-bond but are included because they meet the 7 Å ($C^\alpha(i)$ to $C^\alpha(i+3)$ criterion. In our models we demand that an $i+3$ to i H-bond be formed and hence do not permit Pro at position $i+3$. Pro also occurs very infrequently at $i+2$ even if the definition of ϕ_{i+3} is relaxed to −60° as in the type III β-turn.

TABLE IIb. Relative Probability of Occurrence after Wilmot and Thornton[23]

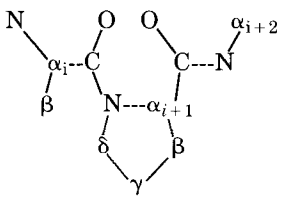
	Type I β-turn				Type II β-turn			
	i	$i+1$	$i+2$	$i+3$	i	$i+1$	$i+2$	$i+3$
Tyr	−1.1	−1.7	−0.8	0.2	−0.4	−0.4	−0.9	1.2
Pro	0.4	5.2*	−3.2*	−3.5*	1.6	9.8*	−2.2*	−2.2*
Gly	−0.7	−2.6*	−1.5	8.1*	−0.7	−1.8	22.9*	−0.3
Ser	2.9*	5.4*	3.6*	−0.7	−0.6	0.8	−2.4*	−1.0
Thr	1.5	2.0	−0.1	0.2	1.7	−2.3*	−2.7*	−0.3
Gln	−0.4	−1.4	−0.7	−0.4	1.3	1.3	−1.4	2.4

*Significant.

Given these occurrences of Pro in β-turns we adopted the rule that β-turns could not have Pro in positions $i + 2$ or $i + 3$. This yielded the patterns summarized in Table III. The consensus sequences are often more restrictive than their constituent individual sequences; however, we initially modeled the consensus sequences searching for a common theme. Variations are treated subsequently. The gliadin consensus is Tyr·Pro·Pro·Pro·Gln·Pro. By our initial criteria no β-turns are allowed. However, no single hexapeptide contains four Pros. The two subconsensus sequences, sequences—Tyr·Pro·Pro·Gln·Gln·Pro and Tyr·Pro·Gln·Pro·Gln·Pro—alternate in tandem repeat. The former can accommodate a β-turn centered at residues 3,4 (i.e., $i + 1$, $i + 2$). The YPQPQP subconsensus sequence can also accommodate a 3,4 β-turn if one relaxes the prohibition against Pro at $i+2$. For the penta-, hexa-, hepta-, and octapeptides we then constructed models trying all three β-turns at each permitted position. As will be discussed the nonapeptide was modeled with only multiple β-turns. In Table III an "x" between residues means that a β-turn, centered at positions $i + 1$ and $i + 2$, is not permitted; "o" indicates permitted.

Cis-Peptide Bonds

In general *cis*-peptide bonds are very rare in proteins. However, in the peptide bond of X·Pro the *trans*-isomer is favored over the *cis* by only 2.0 kcal/mol[24] "because a transition from trans to cis replaces the contact between $C^{\alpha}_{(i)}$ and $C^{\delta}_{(i+1)}$ by a similar contact between $C^{\alpha}_{(i)}$ and $C^{\alpha}_{(i+1)}$.



In the *cis*-peptide bond the dihedral angle, ω, is fixed at 0°. The next dihedral angle, $\phi_{(Pro)}$, is fixed at −54°. By rotating $\psi_{(Pro)}$ one relieves contacts between $O_{(Pro)}$ and $O_{(i)}$ and between $N_{(i+2)}$ O and $O_{(i)}$. These contacts are optimized at $\psi_{(Pro)} = -53°$ and 149°. At these dihedral angles $O_{(Pro)}$–$O_{(i)}$ is 2.98 Å and $N_{(i+2)}$–$O_{(i)}$ is 3.36 Å. By rotating $\psi_{(i)}$ one relieves bad contacts between $C^{\beta}_{(i)}$ and $C_{(Pro)}$ and between $N_{(i)}$ and $C_{(Pro)}$. Only one value of $\psi_{(i)}$ is acceptable; at 144° both of these contacts are 3.68 Å. These two allowed conformations:

$\psi_{(i)}$	$\omega_{(cis)}$	$\phi_{(Pro)}$	$\psi_{(Pro)}$
144°	0°	−54°	149°
144°	0°	−54°	−53°

apply to any amino acid, except Gly, preceding Pro in *cis*-peptide bond.

Huber and Steigemann[25] observed two *cis*-Pros in the Bence–Jones protein Rei—Ser·Pro ($\psi_{(Ser)}$ 139°, $\psi_{(Pro)}$ 174°) and Leu-Pro ($\psi_{(Leu)}$ 150°, $\psi_{(Pro)}$ 156°). They noted the occurrence of seven other *cis*-Pros in globular proteins. Obviously these globular proteins pay some energetic price elsewhere in the molecule for these extra few kilocalories. One cannot predict or understand where these few *cis*-Pro bonds occur.

Grathwohl and Wüthrich[26] determined the equilibrium distribution of X-Pro in oligopeptides in solution by ^{13}C NMR. Although the occurrence of *cis* for both D and L amino acids is low, L-Phe-Pro is unique in that the *cis* conformation is slightly preferred, 76%. Lin and Brandts[27] exploited the specificity of trypsin for substrate bonds that are followed by a *trans*-peptide bond. They showed that in RNase A the Tyr-Pro bond of K_{91}-Y_{92}-P_{93} exists 33% in the *cis* form at equilibrium. Chazin et al.[28] have demonstrated by 2D proton NMR spectroscopy of the 9 kDa intestinal calcium-binding protein that the Gly_{42}, Pro_{43} peptide bond undergoes *trans–cis* isomerization. The resonances resulting from the two forms are present in 4:1 ratio. They suggest that other proteins may have dynamic distributions similar to *cis*-Pro peptide bonds. We evaluated the effect of having a Tyr(Phe), Pro *cis*-peptide in the five proteins in which that doublet occurs in the consensus sequence.

As noted, the conformation of the X·*cis*-Pro dipeptide is highly constrained with only two conformations permitted: $\psi_{(i)}$ 144° $\psi_{(Pro)}$ 149° and $\psi_{(i)}$ 144° $\psi_{(Pro)}$ −53. We can build Ala·Phe·*cis*-Pro·Ala without unacceptable van der Waals contacts for both conformations. In gluten the consensus sequence contains Tyr·Tyr·Pro·Thr and in hordein Pro·Phe·

Table III. Number and Placement of β-Turns Evaluated in the Helical Models*

																		β-turn(s)		cis	cis + one β-turn
																		one	two		
Rhodopsin	Y		P		P		Q		G									6	2	2	12
		x/c		x		o		o		x											
Synaptophysin	Y		G		P		Q		G									9	13	0	0
		x		x		o		o		o											
Synexin	Y		P		P		P		P		G							3	0	1	3
		x/c		x		x		x		o		x									
Gliadin	Y		P		P		P		Q		P							0	0	1	0
		x/c		x		x		x		x		x									
	Y		P		P		Q		Q		P							(3)	0	(1)	(3)
		x/c		x		o		x		x		x									
	Y		P		Q		P		Q		P							(1)	0	(2)	(2)
		x/c/o		x		x/o		x		x/o		x									
RNA polymerase II	Y		S		P		T		S		P		S					9	20	0	0
		x		x		o		x		x		o		o							
Hordein	F		P		Q		Q		P		Q		Q		P			6	9	2	12
		x/c		o		x		x		o		x		x		x					
Gluten	Y		P		T		S		P		Q		Q		G		Y	15	69	2	30
		x/c		o		x		x		o		o		o		o	x				
																		$\overline{52}$	$\overline{113}$	$\overline{11}$	$\overline{62}$
Total, excluding gluten																		37	44	9	32

*Beneath the consensus sequence for each protein are indicated whether a β-turn is permitted "o" or not "x" between the second and third residues. As discussed in the text, a Pro at either the second or third positions generally precludes a β-turn. The gliadin consensus sequence is very restrictive; two less restrictive consensus sequences are listed. The notation "x/o" under the sequence Y·P·Q·P·Q·P indicates that a β-turn is considered; even though, there is Pro at the third position. The possibility of a Tyr(Phe)-*cis*-Pro peptide bond is indicated by "c." The column β-turn, one indicates the total number of models with a single β-turn considered, i.e., three times the number of "o's." The column β-turn, two refers to the pairs of β-turns—$n * (n - 1)/2$—times 9 for the pairwise combinations of β-turn types I, II, and III. When pairs of β-turns occupy adjacent positions, 7 of 9 pairs (except III, I and III, III) are excluded by definition of their respective dihedral angles. For Tyr(Phe)·Pro dipeptides two conformations are considered—ψ_Y 144° ψ_P 149° and ψ_Y 144° ψ_P −53°. The 144/−53 conformation is not permitted for Tyr-*cis*-Pro·Pro. The number of allowed *cis* conformations is indicated in the column headed *cis*. The column *cis* + one β-turn is the product of the columns β-turn, one times *cis*. All of the 122 models, excluding the 116 for gluten, were built. The 41 acceptable models are summarized in Table IV; the remaining 81 have been rejected as indicated in the discussion.

Pro·Gln. Models involving both 144/149 and 144/−53 conformations were explored. However, the 144/−53 conformation cannot be built if the Tyr-*cis*-Pro dipeptide is followed by Pro as is the case for the other three proteins.

rhodopsin	Gly·Tyr·Pro·Pro
synexin	Gly·Tyr·Pro·Pro
gliadin	Pro·Tyr·Pro·Pro

Hence for these three proteins we evaluated only the 144/149 conformation.

Pairs of β-Turns

Pairs of β-turns can occur in various combinations, even in a repeat as short as a pentapeptide. Since there are potentially three types of β-turns—I, II, and III—one must consider nine combinations for each pairing. Seven adjacent pairings can be excluded because of mutually exclusive definitions of the dihedral angles, that is, only the adjacent pairs III, I and III, III are allowed. These possibilities are summarized in Table III. Double β-turn models become more attractive with increasing length of consensus sequence, since with increasing length of the polyproline component the resulting helix becomes less compact. All of the indicated models for penta- through octapeptides were built. There are too many degrees of freedom to permit such a focus on the nonapeptide repeat of gluten.

Placement of Aromatic Side Groups

All seven tandem repeats that we have modeled contain Tyr or Phe, as well as a high content of Pro. At this stage we have restricted our modeling to those oligomers that contain aromatic side chains anticipating that these groups might impart special characteristics beyond their hydrophobicity. Burley and Petsko[29] summarized the occurrence of aromatic side groups in 34 globular proteins. "Aromatic pairs (<7 Å between phenyl ring centroids) were analyzed for" . . . "dihedral angle." "About 60 percent of aromatic side chains in proteins are involved in aromatic pairs, 80 percent of which form networks of three or more interacting aromatic side chains. Phenyl ring centroids are separated by a preferential distance of between 4.5 and 7 Å, and dihedral angles approaching 90° are most common."

TABLE IV. Rhodopsin, Six Pro-β Models*

			Initial		Adjusted				
		rdI34	ϕ°	ψ°	ϕ°	ψ°	Comments		
ΔZ 4.61	A†	Tyr	−54	120	−54	160	Tyr part		
ΔΦ 116°		Pro	−54	120	−54	160	buried		
P 14.3	A	Pro	−54	−30	−54	−30			
n 3.10		Gln	−90	0	−90	0			
Δz 0.92	A	Gly	−54	120	−54	160			
		rdII34							
3.96	A	Tyr	−54	120	−105	120	Tyr out		
−133°		Pro	−54	120	−54	120			
10.7	A	Pro	−60	120	−60	120			
2.70		Gln	80	0	80	0			
0.79	A	Gly	−54	120	−120	145			
		rdIII34							
4.10	A	Tyr	−54	120	−54	160	Tyr out		
−120°		Pro	−54	120	−54	160	of helix		
12.3	A	Pro	−54	−30	−54	−30	open center		
3.00		Gln	−60	−30	−60	−30			
0.82		Gly	−54	120	−54	−160			
		rdIII45							
7.60	A	Tyr	−54	120	−70	145	Tyr part		
−124°		Pro	−54	120	−54	145	buried		
22.1	A	Pro	−54	120	−54	145			
2.90		Gln	−60	−30	−54	−30			
1.52	A	Gly	−60	−30	−60	−30			
		rdIII34I45	Initial		Final		χ°_1	χ°_2	χ°_3
5.05	A	Tyr	−54	120	−106	131	−108	−137	
105°		Pro	−54	120	−54	116			
17.3	A	Pro	−54	−30	−54	−32			
3.43		Gln	−60	−30	−61	−30	Out		
1.01	A	Gly	−90	0	−90	0			
		rdIII34III45							
7.16	A	Tyr	−54	120	−70	160	−38	162	
125°		Pro	−54	120	−54	120			
20.6	A	Pro	−54	−30	−54	−30			
2.88		Gln	−60	−30	−60	−30	−75	−49	22
1.43	A	Gly	−60	−30	−60	−30			

*The characteristics of 41 of the 122 models are summarized in Tables IV–IX. The remaining 81 models were rejected because they either had multiple close van der Waals contacts or because they were so elongated as to provide few stabilizing intrastrand contacts. The type of β-turn—I, II, or III—is followed by the numbers in the oligopeptide corresponding to the i + 1 and the i + 2 positions. Double β-turns are designated by two sequential such designations. *cis* preceding the two letter abbreviation of the protein indicates that the Tyr(Phe)·Pro peptide bond is *cis;* there may be in addition one β-turn per oligomer. The initial model was built assigning to the ϕ, ψ dihedral angles the value of polyproline, or the values for the indicated β-turn, or values necessitated by having the *cis* peptide bond. The model was then adjusted to relieve van der Waals contacts, to make it more compact, or to optimize H-bonding. Comments are included for the 31 adjusted models. The 10 models chosen for further refinement are presented in their "final" forms; when defined side chain dihedral angles are also listed. These 10 models are described in greater detail in the text. For each model we list the helical parameters—ΔZ, rise per oligopeptide; ΔΦ, rotation per oligopeptide; P, pitch; n, number of oligopeptides per turn; and Δz, average rise per amino acid.
†A = Å

Given the dimensions of the helices, this sort of herringbone interaction, if it exists, inevitably occurs near the helix axis.

Construction and Evaluation of Models

Models were built on a Silicon Graphics 3040 interactive graphics system using the program MMS.[30] Initial angles were used for polyproline and for β-turns as described previously. Subsequently slight adjustments were made to optimize van der Waals contacts and hydrogen bonds.

Many of the models of Table III were rejected either because they were elongated with no stabilizing intrachain contacts or because they overlapped themselves after several repeats. For these rejected models neither of these shortcomings could be relieved by small adjustments of ϕ and ψ. Forty-one of the 122 indicated models, for six penta- through

TABLE V. Synaptophysin, Eighteen Pro-β Models*

			Initial		Adjusted		
			φ°	ψ°	φ°	φ°	Comments
		spII34					
ΔZ 4.55	A†	Tyr	−54	120	−54	160	
ΔΦ−116°		Gly	−54	120	−54	130	
P 14.1	A	Pro	−54	−30	−54	30	
n 3.10		Gln	−90	0	−90	0	Tyr buried
Δz 0.91	A	Gly	−54	120	−54	160	
		spII34					
3.16	A	Tyr	−54	120	−105	120	
−144°		Gly	−54	120	−105	120	
7.9	A	Pro	−60	120	−60	120	
2.50		Gln	80	0	80	0	Tyr buried
0.63	A	Gly	−54	120	−105	120	
		spIII34I45					
8.52	A	Tyr	−54	120	−90	160	
133°		Gly	−54	120	−90	160	
23.0	A	Pro	−54	−30	−54	−30	
2.70		Gln	−60	−30	−60	−30	Tyr buried
1.70	A	Gly	−90	0	−90	0	
		spIII34III45					
9.62	A	Tyr	−54	120	−90	160	
124°		Gly	−54	120	−90	160	
27.9	A	Pro	−54	30	−54	30	
2.90		Gln	−60	−30	−60	−30	Tyr buried
1.92	A	Gly	−60	−30	−60	−30	
		spI34I51					
4.60	A	Tyr	−90	0	−90	0	
−120°		Gly	−54	120	−100	140	
13.8	A	Pro	−54	−30	−54	−30	
3.00		Gln	−90	0	−90	0	Tyr part
0.92	A	Gly	−60	−30	−60	−30	buried
		spI34III51					
4.77	A	Tyr	−60	−30	−60	−30	
−120°		Gly	−54	120	−100	140	
14.3	A	Pro	−54	−30	−54	−30	
3.00		Gln	−90	0	−90	0	Tyr part
0.95	A	Gly	−60	−30	−60	−30	buried
		spII34I51					
5.14	A	Tyr	−90	0	−90	0	
−120°		Gly	−54	120	−100	140	
15.7	A	Pro	−54	120	−54	120	
3.00		Gln	80	0	80	0	
1.03	A	Gly	−60	−30	−60	−30	
		spII34II51					
4.76	A	Tyr	80	0	80	0	
−171°		Gly	−54	120	−100	140	
10.0	A	Pro	−54	120	−54	120	
2.10		Gln	80	0	80	0	Tyr part
0.95	A	Gly	−60	120	−60	120	buried
		spII34III51					
4.69	A	Tyr	−60	−30	−60	−30	
−171°		Gly	−54	120	−100	140	
9.84	A	Pro	−54	120	−54	120	
2.10		Gln	80	0	80	0	
0.94	A	Gly	−60	−30	−60	−30	
		spIII34I51					
5.36	A	Tyr	−60	−30	−60	−30	
−136°		Gly	−54	120	−100	140	
14.0	A	Pro	−54	−30	−54	−30	
2.60		Gln	−60	−30	−60	−30	
1.07	A	Gly	−60	−30	−60	−30	

(continued)

TABLE V. Synaptophysin, Eighteen Pro-β Models* (Continued)

			Initial		Adjusted				
			φ°	ψ°	φ°	φ°	Comments		
		spIII34II51							
5.20	A†	Tyr	80	0	80	0			
−120°		Gly	−54	120	−100	140			
15.6	A	Pro	−54	−30	−54	−30			
3.00		Gln	−54	−30	−54	−30			
1.04	A	Gly	−60	120	−60	120			
		spI45							
6.04	A	Tyr	−54	120	−90	140			
−144°		Gly	−54	120	−54	160			
15.1	A	Pro	−54	120	−54	160			
2.50		Gln	−60	−30	−60	−30	Tyr part		
1.21	A	Gly	−90	0	−90	0	buried		
		spII45							
2.47	A	Tyr	−54	120	−135	135			
120°		Gly	−54	120	−135	135			
7.4	A	Pro	−54	120	−135	135			
3.00		Gln	−60	120	−60	120	Tyr out		
0.49	A	Gly	80	0	80	0			
		spIII45							
4.97	A	Tyr	−54	120	−70	180			
−116°		Gly	−54	120	−54	140			
15.4	A	Pro	−54	120	−54	140			
3.10		Gln	−60	−30	−60	−30	Tyr out		
0.99	A	Gly	−60	−30	−60	−30			
		spIII45I51							
6.97	A	Tyr	−90	0	−90	0			
129°		Gly	−54	120	−80	160			
19.5	A	Pro	−54	120	−54	140			
2.80		Gln	−60	−30	−60	−30	Tyr out		
1.39	A	Gly	−60	−30	−60	−30			
		spIII45III51							
7.77	A	Tyr	−60	−30	−60	−30			
120°		Gly	−54	120	−80	160			
23.3	A	Pro	−54	120	−54	120			
3.00		Gln	−60	−30	−60	−30	Tyr out		
1.55	A	Gly	−60	−30	−60	−30			
		spIIIer	Initial		Final		χ°_1	χ°_2	χ°_3
2.84	A	Tyr	−54	120	−51	163	7	127	
−113°		Gly	−54	120	−51	163			
9.1	A	Pro	−54	−30	−54	−31			
3.19		Gln	−60	−30	−59	−30	Out		
0.57	A	Gly	−54	120	−54	163			
		spI34II51							
5.00	A	Tyr	80	0	80	0	−93	180	
−130°		Gly	−54	120	−100	140			
13.8	A	Pro	−54	−30	−54	−30			
2.77		Gln	−90	0	−90	0	−78	144	125
1.00	A	Gly	−60	120	−60	120			

*See Table IV footnote.
†A = Å

octapeptides of Table III, were further adjusted (Tables IV through IX). Ten models, at least one for each of the six oligopeptides, were evaluated in greater detail and are illustrated in Figures 1 through 10. These 10 were tested with the energy minimization option of the program X-PLOR. Water was not included. The conformations did not change greatly. However, when tested as heptaoligomers, the terminal oligomers reached different energy minima than did the central oligomer. A rigorous evaluation of the energetics of these model structures requires an extensive study. Our criteria were acceptable van der Waals contacts and hydrogen bonds.

RESULTS

As discussed under Rationale and summarized in Table III our analysis has indicated that we should evaluate many models—52 with one β-bend, 113 with two β-turns, 11 with a Tyr(Phe)·*cis*-Pro peptide bond, and 62 having one *cis* Tyr, Phe bond, and one β-turn. The nonapeptide of gluten has only two Pros and hence many more degrees of freedom. Our modeling studies with the shorter oligopeptides demonstrate the feasibility of building a model of gluten having two or even three β-turns. Since the goal of our study was not so much to predict the detailed structure(s) of these proteins but to demonstrate the plausibility of polyproline, β-turn models, we did not build most of the gluten models. This reduced the number of models from 52 to 37, 113 to 44, 11 to 9, and 62 to 32. All of the 122 remaining models were built. Either because of unacceptable van der Waals contacts that could not be relieved by adjustment of the available dihedral angles or because the resulting helix was very elongated and could not be adjusted to form any stabilizing intrastrand interactions, 81 of 122 were discarded; the remaining 41 models are summarized in Tables IV–IX. Ten of these 41 were evaluated in greater detail and further refined to optimize hydrogen bonds and van der Waals contacts. We first describe these 10 models then summarize the common or general characteristics of all of the models.

The models are designated by a two-letter abbreviation for the protein, roman numerals for the type of β-turn(s), and arabic numerals for the position(s) of the second and third residues of the turn (Tables IV–IX). The subscript "*j*" refers to the repeating oligopeptide. The "initial" ϕ, ψ angles derived from the definition of the model. For 31 models the "adjusted" dihedral angles indicate a model that is acceptable in that it is relatively compact and has no unacceptable van der Waals contacts. For 10 models the "final" dihedral angles and side chain angles χ_1, χ_2, and χ_3 reflect more detailed refinement. These ten are illustrated and described in greater detail. Each helix can be described by two parameters—ΔZ rise per oligopeptide $\Delta\Phi°$ rotation per oligopeptide. We include the redundant information—*P* pitch, *n* number of oligopeptides per turn, and Δz average rise per amino acid to permit an easy comparison of the compactness of various models.

Rhodopsin Tyr·Pro·Pro·Gln·Gly

Of the six rhodopsin models three are left handed; 18 of the 41 acceptable models are left handed. Only 2 of the 10 more refined models are left handed (Figs. 3 and 4 and Table IV).

The model rdIII34I45, like the following rdIII34III45, is stabilized by the H-bonds of overlapping β-turns—N (Gly-5) to O (Pro-2) and N (Tyr-1) to O [Pro-3 (j − 1)] of the two β-turns. The Tyr side chain is partially buried with only one face exposed to solvent. The rings form a herringbone pattern with closest contact, ε-carbon to δ-carbon, 3.57 Å. The structure might be further stabilized by water molecules forming an H-bond bridge from the η-oxygen of Tyr to either the carbonyl oxygen of Pro-3 (j − 3), 3.94 Å or to the carbonyl-oxygen of Tyr-1 (j − 2), 4.97 Å. As now built, the side chain of Gln-4 sticks out into solution. It might be rotated back to bond the Tyr oxygen via a bridging water. The structure is narrow enough so that the Tyr side chain cannot be fully buried; even so, the structure is compact with relatively few large indentations and no volume for buried water.

The model rdIII34III45 resembles rdIII34I45 in that it too is stabilized by the same overlapping H-bonds. It is, however, more elongated as reflected by the greater rise per pentamer, 7.16 Å vs 5.05 Å for rdIII34I45. The Tyr side chains are no longer in contact and the side chain is more exposed; however, again one face of the ring lies against the chain with no space for buried water. This seemingly small change from type I to type III β-turn facilitates the formation of an H-bond(s) involving the η-oxygen of Tyr—2.92 Å to the carbonyl oxygen of Tyr-1 (j − 1), 3.40 Å to the carbonyl oxygen of Gln-4 (j − 2), 2.54 Å to the carbonyl of Pro-3 (j − 2), and 3.06 Å to the ε-nitrogen of Gln-4 (j − 2). Again we emphasize that the actual amino acid sequences vary slightly relative to their consensus sequences. Over the length of the tandem repeats there could easily exist a mixture of rdIII34I45 and rdIII34III45.

Synaptophysin Tyr·Gly·Pro·Gln·Gly

Since the consensus pentamer of synaptophysin contains only one Pro many more β-turns are permitted. Eighteen models were judged acceptable; 13 of these are left-handed. The extent of exposure of the tyrosyl side chains varied from fully exposed to almost completely buried.

The models spIII34 and spI34II51 (Figs. 3 and 4) are unique among the 10 discussed in detail in that they are the only ones having left-handed pitches, $\Delta\phi = -113°$ and $-130°$. In spIII34 the Tyr side chain oxygen can easily form H-bonds with both carbonyl oxygens of Tyr-1 (j + 2), 2.64 Å, and Gly-2 (j + 2), 2.48 Å. The Tyr side chains point inward toward the axis of the helix with ε- and ξ-carbons as well as the η-oxygen buried. The side chains are in contact with one another with closet approach—3.55 Å ε-carbon to ε-carbon. The α- and β-carbons are exposed to solvent.

As seen from the difference in rise per oligomer, Δz of 2.84 Å for spIII34 verses 5.00 Å for spI34II51, the latter model is more elongated. The Tyr side chains have one face exposed to solvent; they are not in contact with one another. Its η-oxygen can form an H-bond with the carbonyl oxygen of Pro-3 (j − 3), 2.68 Å. The side chain amine of Gln-4 can form H-

TABLE VI. Synexin, Two Pro-β Models*

			Initial		Adjusted		
			φ°	ψ°	φ°	ψ°	Comments
		sx156					
ΔZ 4.76	A†	Tyr	−54	120	−100	117	Tyr part
ΔΦ 144°		Pro	−54	120	−54	120	buried
P 11.9	A	Pro	−54	120	−54	120	
n 2.50		Pro	−54	120	−54	120	
Δz 0.79	A	Pro	−54	−30	−54	−30	
		Gly	−90	0	−94	0	
		sxIII56	Initial		Final		χ°_1 χ°_2
5.10	A	Tyr	−54	120	−101	127	104 163
131°		Pro	−54	120	−54	125	
14.0	A	Pro	−54	120	−54	125	
2.74		Pro	−54	120	−54	120	
0.85	A	Pro	−54	−30	−55	−29	
		Gly	−60	−30	−59	−31	

*See Table IV footnote.
†A = Å.

bonds to the carbonyls of both Gln-4 (*j* + 1), 2.66 Å, and Gly-2 (*j* + 2), 3.22 Å. This combination of β-turns generates an especially tortuous main chain trace.

Synexin Tyr·Pro·Pro·Pro·Pro·Gly

The four Pros of synexin, and similarly of gliadin, place severe constraints on the number of acceptable models and provide the most stringent test of the concept of the pro-β class of models. As expected, sxI56 and sxIII56 are similar.

The model sxIII56 (Fig. 5) is an especially nice example of four Pros in a row interrupted by a β-turn followed by four more Pros. The connecting type III β-turn has a H-bond from nitrogen of Tyr-1 to the carbonyl oxygen of Pro-5 (*j* − 1), 2.81 Å. The Tyr side chains are almost completely buried with only one edge exposed. The rings are layered between the pyrridol rings of two Pros with closest contacts 3.68 and 4.89 Å. These distances could be reduced with further adjustment of free dihedral angles. The structure is compact, although there is a groove following the trace of the helix. This groove provides solvent access to the edge of the Tyr ring as well as its η-oxygen.

Gliadin Tyr·Pro·Pro·Pro·Gln·Pro

No single hexamer has both Pro-3 and Pro-4. In fact the consensus dodecamer is Tyr·Pro·Pro·Gln·Gln·Pro·Tyr·Pro·Gln·Pro·Gln ·Pro.

The strict consensus Tyr·Pro·Pro·Pro·Gln·Pro cannot be built honoring the rules previously described. A Pro can be built at the second position of a type III β-turn as illustrated in glIII34 and glIII56; however, this involves several close contacts and is seen only infrequently in globular proteins.[23] Only 2 of the 41 models contain Tyr·Cis-Pro; *cis*-gl is one of them.

We illustrate one plausible structure, glIII34pq in Figure 6, for Tyr·Pro·Pro·Gln·Gln·Pro. The Tyr side chains are partially buried and in contact, the closest being 3.37 Å, ε-carbon to ε-carbon. The η-oxygen may be protected from the solvent by a H-bond with either the amine or the oxygen of the amido group of Gln-5, shown at 2.68 Å. There is also the possibility of a water bridge to the carbonyl oxygen of this same Gln, 3.78 Å. The helix has a broad diameter, about 19 Å. A relatively deep groove follows the trace of the helix. Both Gln side chains are exposed to solvent.

RNA Polymerase II Tyr·Ser·Pro·Thr·Ser·Pro·Ser

With increasing length of the consensus oligopeptide there is a greater volume to accommodate a buried Tyr side chain. This is illustrated in rpIII34 and rpIII34III67; the Tyr of rpIII34III71 is completely exposed.

Two models—rpIII34 and rpIII34III67—have the Tyr side chains almost completely buried and in contact with one another. As indicated by the helix repeats of 4.35 and 4.38 Å, the Tyr rings tip slightly to form a herringbone pattern. In rpIII34 (Fig. 7) the Tyr η-oxygen might form H-bonds to either or both the amine nitrogen of the preceding (*j* − 1) Tyr (3.03 Å) or to the hydroxyl of Ser-7 (*j* − 2) (2.84 Å) just preceding Tyr-1 (*j* − 1). The III34 β-turn is, of course, stabilized by an H-bond from nitrogen of Ser-5 to the carbonyl oxygen of Ser-2 (2.79 Å). In addition there is the possibility of additional H-bonds between the hydroxyl of Ser-2 and the hydroxyl oxygens of Thr-4 (2.79 Å) and/or Ser-5 (3.53 Å). The structure is compact. Only the edge of the Tyr ring would have minimal contact with the solvent. Without significant change in conformation it could not be phosphorylated. In contrast the side chains of Thr-4 and of Ser-2 and Ser-5 are easily accessible.

TABLE VII. Gliadin, Six Pro-β Models*

			Initial		Adjusted		
			φ°	ψ°	φ°	ψ°	Comments
		glIII34					
ΔZ 4.33	A†	Tyr	−54	120	−70	120	
ΔΦ 120°		Pro	−54	120	−54	120	
P 13.0	A	Pro	−54	−30	−54	−48	
n 3.00		Pro	−54	−30	−54	−30	
Δz 0.72	A	Gln	−54	120	−80	120	
		Pro	−54	120	−54	120	
		glIII56					
3.46	A	Tyr	−54	120	−95	140	
132°		Pro	−54	120	−54	110	
9.7	A	Pro	−54	120	−54	110	
2.80		Pro	−54	120	−54	110	
0.58	A	Gln	−60	−30	−60	−30	
		Pro	−54	−30	−54	−30	
		glI34pg					
4.03	A	Tyr	−54	120	−80	120	Tyr part
120°		Pro	−54	120	−54	120	buried
12.1	A	Pro	−54	−30	−54	−30	
3.00		Gln	−90	0	−90	0	
0.67	A	Gln	−90	120	−90	120	
		Pro	−90	120	−54	120	
		glIII34qp					
4.58	A	Tyr	−54	120	−75	120	Tyr buried
116°		Pro	−54	120	−54	120	
14.2	A	Gln	−60	−30	−60	−30	
3.10		Pro	−54	−30	−54	−30	
0.76	A	Gln	−54	120	−90	120	
		Pro	−54	120	−54	120	
		cisglqp					
6.57		Tyr	−54	144	−120	144	
−170°		Pro	−54	149	−54	149	
13.9		Gln	−54	120	−54	120	
2.12		Pro	−54	120	−54	160	
1.09		Gln	−54	120	−54	160	
		Pro	−54	120	−54	120	

		glIII34	Initial		Final		χ°_1	χ°_2	χ°_3
4.40	A	Tyr	−54	120	−75	120	−162	−132	
128°		Pro	−54	120	−55	120			
12.4	A	Pro	−54	−30	−54	−30			
2.81		Gln	−60	−30	−60	−30	Out		
0.73	A	Gln	−54	120	−90	120	−133	−60	−112
		Pro	−54	120	−55	120			

*See Table IV footnote.
†A = Å.

The compactness is seen in the 4.86 Å distance between γ-carbons of Pro-6 and Pro-3 ($j - 2$). The closest contacts between Tyr rings are 3.63 Å δ-carbon to ε-carbon and 3.08 Å η-oxygen to ε-carbon.

The general shape and characteristics of rpIII34III67 (Fig. 8) are quite similar to those of rpIII34. Both are compact and in both the Tyr side chains are buried. Correspondingly the III34 β-turn (2.81 Å) can be stabilized by an additional H-bond between Ser-2 and Thr-4 (2.48 Å). The III67 β-turn is involved in a complex triad of H-bonds between Ser-5 and Ser-7 hydroxyl oxygen atoms (3.84 Å) and the η-oxygen of Tyr-1 (3.09 and 2.89 Å). The closest contact between rings is 3.40 Å ε-carbon to ε-carbon.

A single change of β-turn from III67 to III71, as seen in rpIII34III71 (Fig. 9) essentially inverts the helix. The helix opens somewhat, ΔZ 4.38–5.72 Å. The most striking feature, though, is that the tyrosyl side chain points out into solution while the hydroxyl groups of Ser-2 and Ser-5 as well as Thr-4 point inward, possibly forming a network of H-bonds. The important point is that a seemingly small change in φ, ψ angles can significantly alter the structure. Such a transition might be involved in

TABLE VIII. RNA Polymerase II, Seven Pro-β Models*

			Initial		Adjusted			
			φ°	ψ°	φ°	ψ°	Comments	
		rpIII67						
ΔZ 4.95	A†	Tyr	−54	120	−82	160		
ΔΦ −150°		Ser	−54	120	−82	160		
P 11.9	A	Pro	−54	120	−54	160		
n 2.40		Thr	−54	120	−82	160		
Δz 0.71	A	Ser	−54	120	−82	160		
		Pro	−60	−30	−60	−30		
		Ser	−60	−30	−60	−30		
		rpIII71						
4.74	A	Tyr	−60	−30	−60	−30		
154°		Ser	−54	120	−82	160		
10.9	A	Pro	−54	120	−54	160		
2.30		Thr	−54	120	−82	160		
0.68	A	Ser	−54	120	−82	160		
		Pro	−54	120	−54	160		
		Ser	−60	−30	−60	−30		
		rpI34						
4.31	A	Tyr	−54	120	−60	90		
124°		Ser	−54	120	−82	160		
12.5	A	Pro	−60	−30	−60	−30		
2.90		Thr	−90	0	−90	0		
0.62	A	Ser	−54	120	−100	160		
		Pro	−54	120	−54	160		
		Ser	−54	120	−82	160		
		rpII34						
7.43	A	Tyr	−54	120	−54	160		
−120°		Ser	−54	120	−54	120		
22.3	A	Pro	−60	−30	−60	−120		
3.00		Thr	−90	0	−90	0		
1.06	A	Ser	−54	120	−100	140		
		Pro	−54	120	−54	120		
		Ser	−54	120	−54	120		
		rpIII34	Initial		Final		χ°_1	χ°_2
4.35	A	Tyr	−54	120	−80	159	−27	−171
140°		Ser	−54	120	−85	157	49	
11.2	A	Pro	−54	−30	−55	−29		
2.57		Thr	−60	−30	−61	−30	45	
0.62	A	Ser	−54	120	−86	156	−35	
		Pro	−54	120	−55	156		
		Ser	−54	120	−83	158	−5	
		rpIII34III67						
4.38	A	Tyr	−54	120	−56	117	−151	−159
160°		Ser	−54	120	−67	118	168	
9.9	A	Pro	−54	−30	−54	−30		
2.25		Thr	−60	−30	−60	−30	72	
0.63	A	Ser	−54	120	−62	117	150	
		Pro	−54	−30	−54	−30		
		Ser	−60	−30	−60	−30	140	
		rpIII34III71						
5.72	A	Tyr	−60	−30	−60	−35	−55	180
164°		Ser	−54	120	−50	120	−56	
12.5	A	Pro	−54	−30	−54	−24		
2.20		Thr	−60	−30	−60	−41	−65	
0.82	A	Ser	−54	120	−54	121	−119	
		Pro	−54	120	−54	121		
		Ser	−60	−30	−60	−25	−166	

*See Table IV footnote.
†A = Å.

TABLE IX. Hordein, Two Pro-β Models*

			Initial		Adjusted				
			φ°	ψ°	φ°	ψ°	Comments		
		cishdI56							
ΔZ 8.78	A†	Phe	−54	144	−120	144			
ΔΦ 277°		Pro	−54	149	−54	149			
P 11.4	A	Gln	−54	120	−120	100			
n 1.30		Gln	−54	120	−130	140			
Δz 1.10	A	Pro	−54	−30	−54	−30			
		Gln	−90	0	−90	0			
		Gln	−54	120	−105	100			
		Pro	−54	120	−80	100			
		hdIII23III56	Initial		Final		χ°_1	χ°_2	χ°_3
8.75	A	Phe	−54	120	−77	101	33	180	
99°		Pro	−54	−30	−56	−30			
31.9	A	Gln	−60	−30	−60	−30	−83	5	−177
3.64		Gln	−54	120	−80	100	−165	−163	−135
1.10	A	Pro	−54	−30	−55	−30			
		Gln	−60	−30	−60	−30	−53	155	−50
		Gln	−54	120	−80	100	−151	83	38
		Pro	−54	120	−54	100			

See Table IV footnote.
†A = Å.

TABLE X. Summary of β-Turns Used in the 41 Accepted Models

	Number of models	Number of β-turns	Types of β-turns			Number *cis*-Pro
			I	III	II	
Rhodopsin	6	8	2	5	1	0
Synaptophysin	18	30	10	12	8	0
Synexin	2	2	1	1	0	0
Gliadin	6	5	1	4	0	1
RNA Polymerase II	7	9	1	7	1	0
Hordein	2	3	1	2	0	1
Totals	41	57	16	31	10	2

some of the functional transformations of RNA polymerase.

Hordein Phe·Pro·Glu·Glu·Pro·Gln·Gln·Pro

Most of the 29 hordein models are quite elongated and appear to have little intrastrand stabilization. Only two seem promising, one of them being the second example of a Phe·*cis*-Pro. Field et al.[17] presented viscosity data consistent with hordein's having a structure based on a single stranded helix. For none of these proteins have we considered interstrand interactions as found in multistrand helices. A third β-turn might generate a more compact helix; but only two sites in the consensus sequence permit β-turns.

The hdIII23III56 model (Fig. 10) is quite elongated, 8.77 Å per octamer, as are all of the hordein models. As illustrated, the Phe side chain is partly exposed to solvent; it could be rotated so that one face approaches the main chain. All four Gln side chains point into solution. Gln-4, -6, and -7 are situated so that their side chains could H-bond with one another.

DISCUSSION

Generalizations

Several generalizations emerge from a survey of the 41 models judged to be reasonable. Only two contain Tyr (or Phe)·*cis*-Pro as peptide bonds, although all nine of Table III were built (Table X). Neither of these two contains the highly constrained Tyr·*cis*-Pro·Pro. Seventeen of the models contain pairs of β-turns. The combinations of types and relative displacements follow no obvious patterns. The types I and III β-turns are quite similar (−60° −30°, −90° 0° and −60° −30°, −60° −30°) and are treated as a single group by Wilmot and Thornston.[23] Not surprisingly, models with β-turn I or β-turn III at the same position are similar. The range of dihedral angle at $i+2$ does provide more range for model building, compare rdIII34I45 and rdIII34III45. Type II β-turns are used less frequently (10 times) than are types I and III (16 + 31). Rhodopsin, Tyr·Pro·Pro·Gln·Gly, has two Pros while the similar synaptophysin, Tyr·Gly·Pro·Gln·Gly, has only one. Not surprisingly, more models are possible for synaptophysin (18) than for rhodopsin (6); all the models

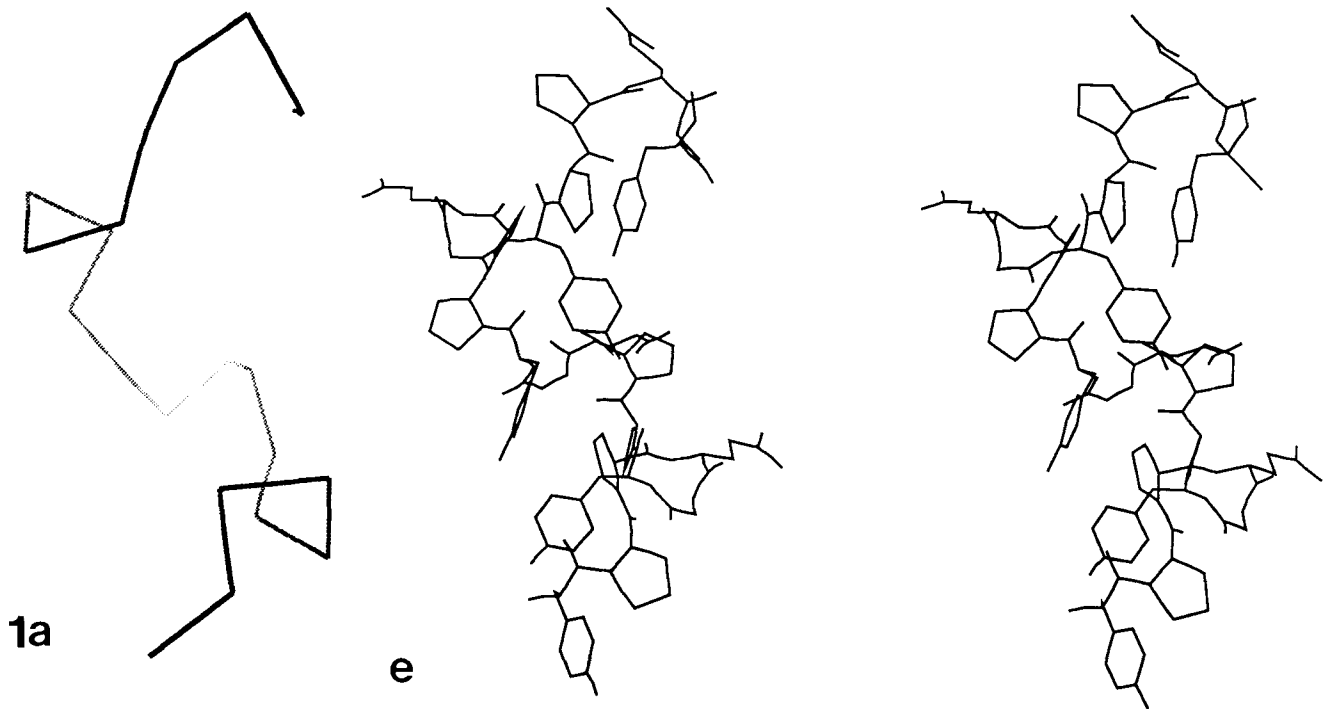


Fig. 1. rdIII34I45. Figs. 1–10. For each of the 10 refined models (Figs. 1–10) we present five standard views. (**a**) α-carbon backbone viewed perpendicular to the helix axis. (**b**) Skeletal model viewed perpendicular to the helix axis with van der Waals radii indicated for the atoms of the Tyr side chains. (**c**) View as in (**b**) with all van der Waals radii indicated. (**d**) Skeletal model viewed along the helix axis with van der Waals radii indicated for the atoms of the Tyr side chains. (**e**) Stereoscopic view of the skeletal model as seen in (**b**).

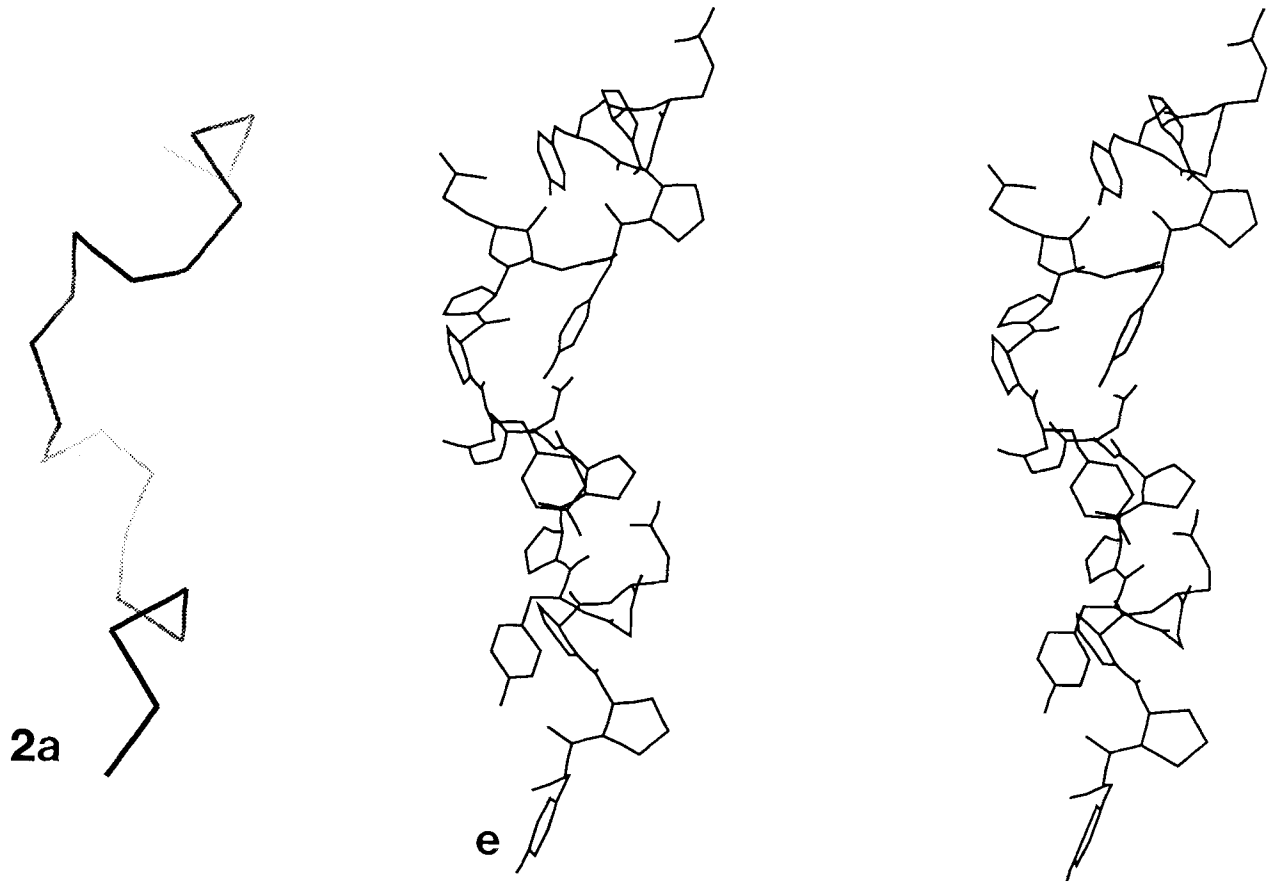


Fig. 2. rdIII34III45.

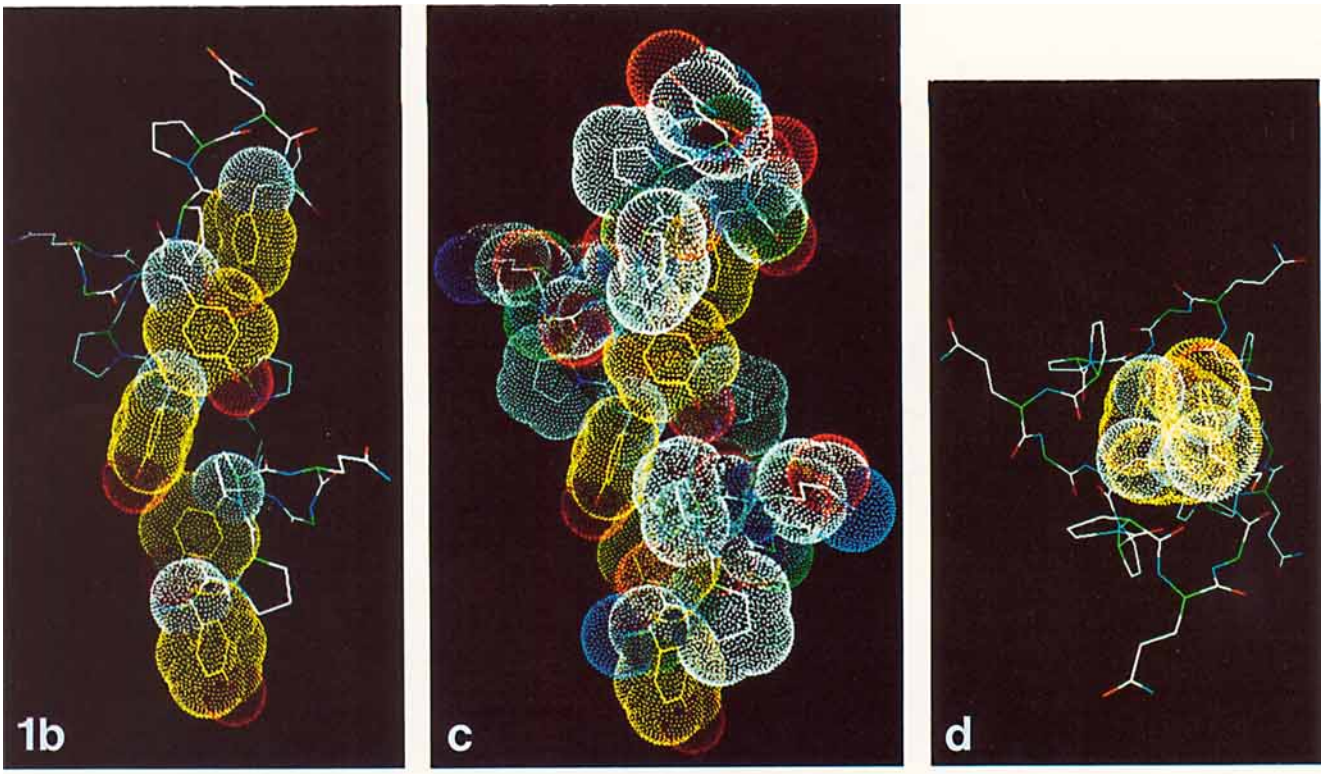


Fig. 1. b,c,d.

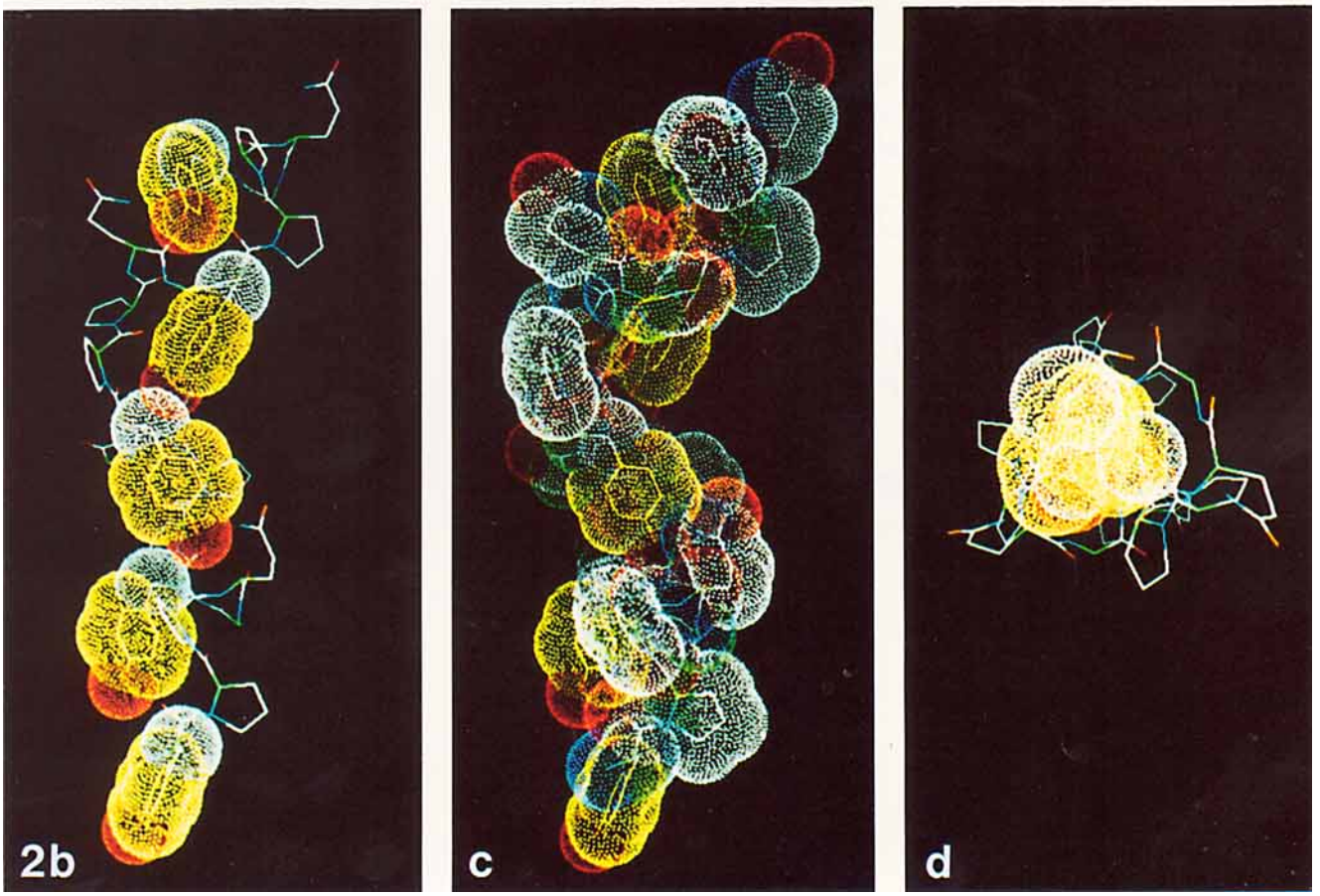


Fig. 2. b,c,d.

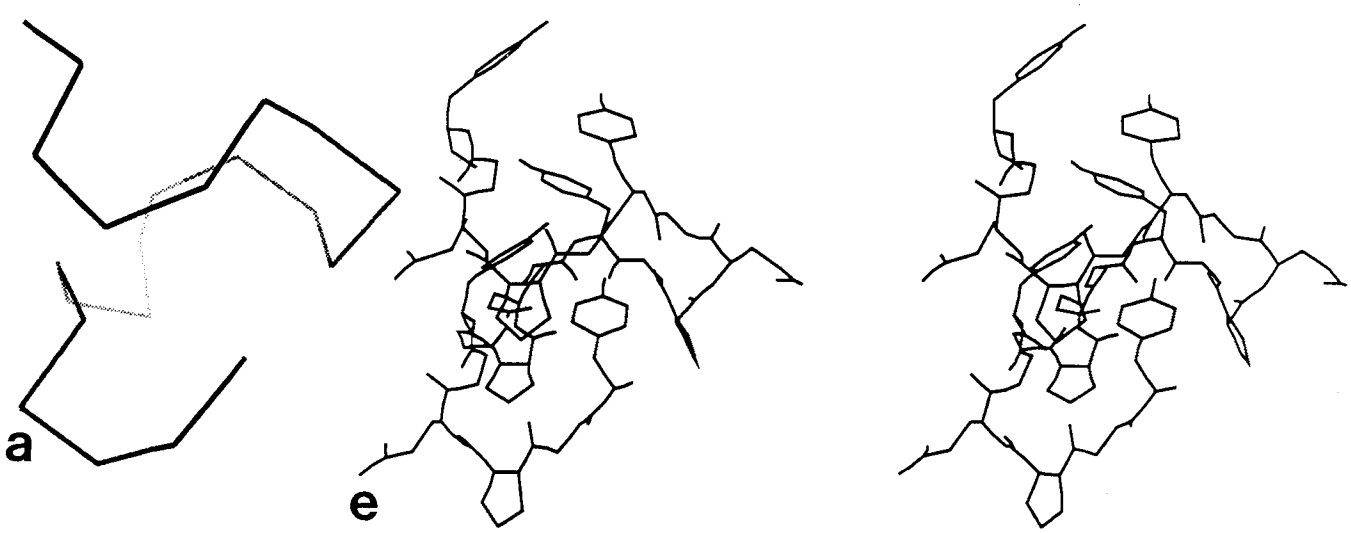


Fig. 3. splII34.

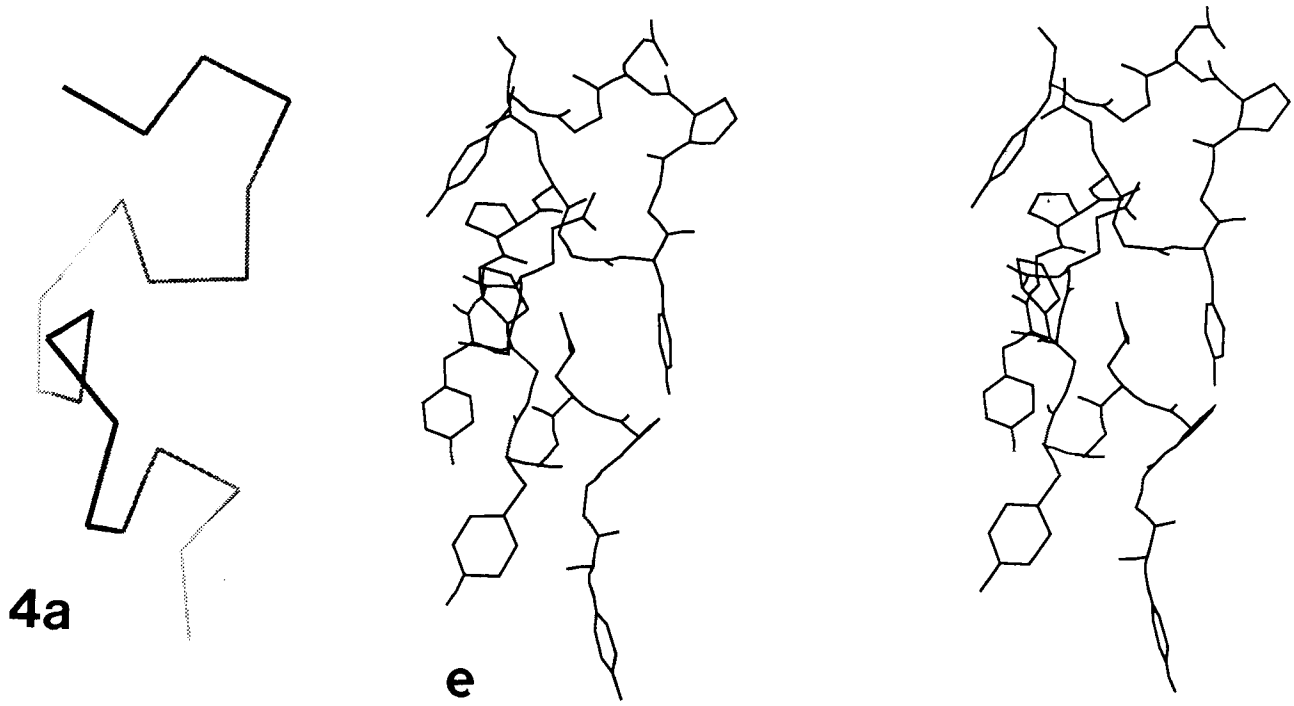


Fig. 4. spl34II51.

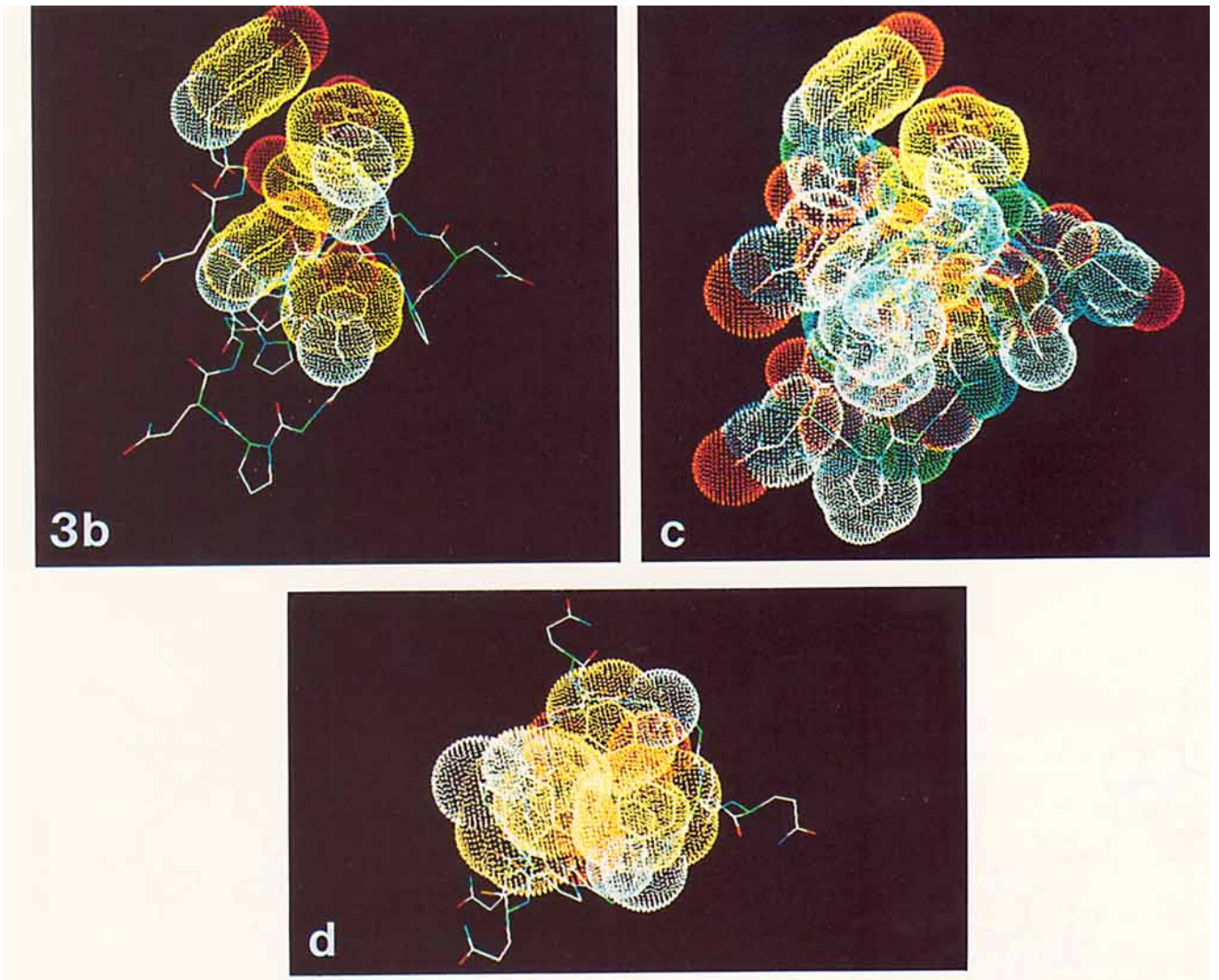


Fig. 3. b,c,d.

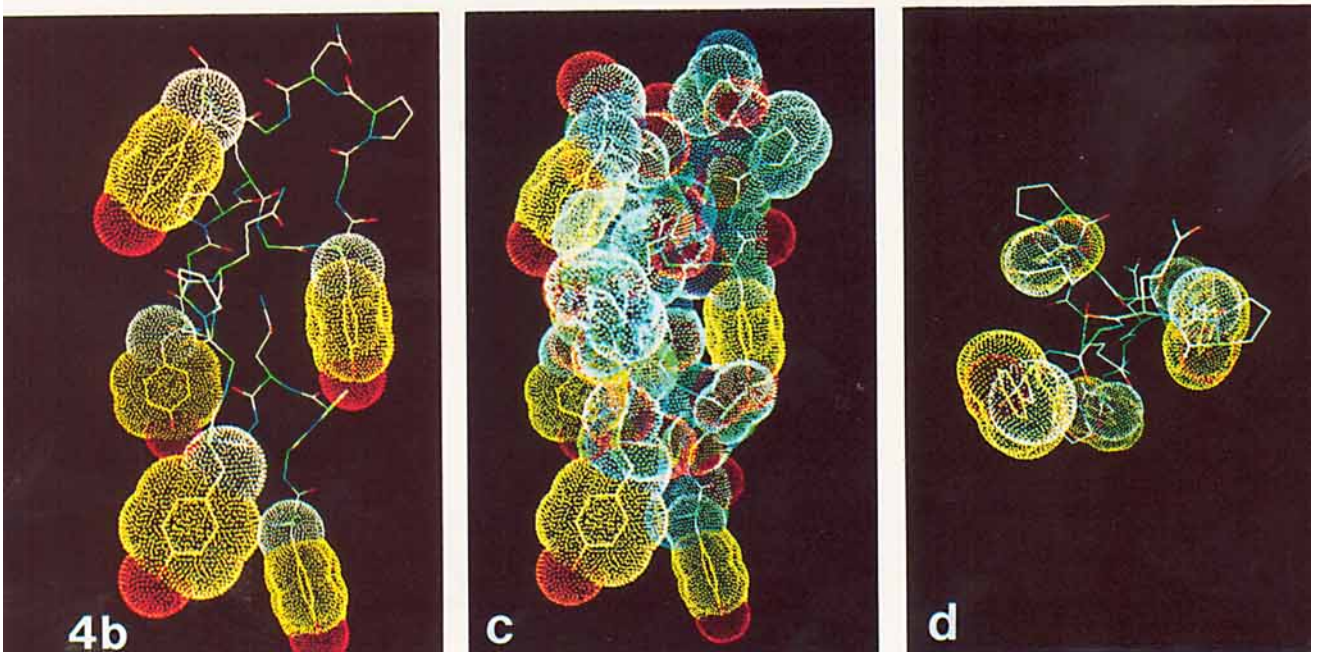


Fig. 4. b,c,d.

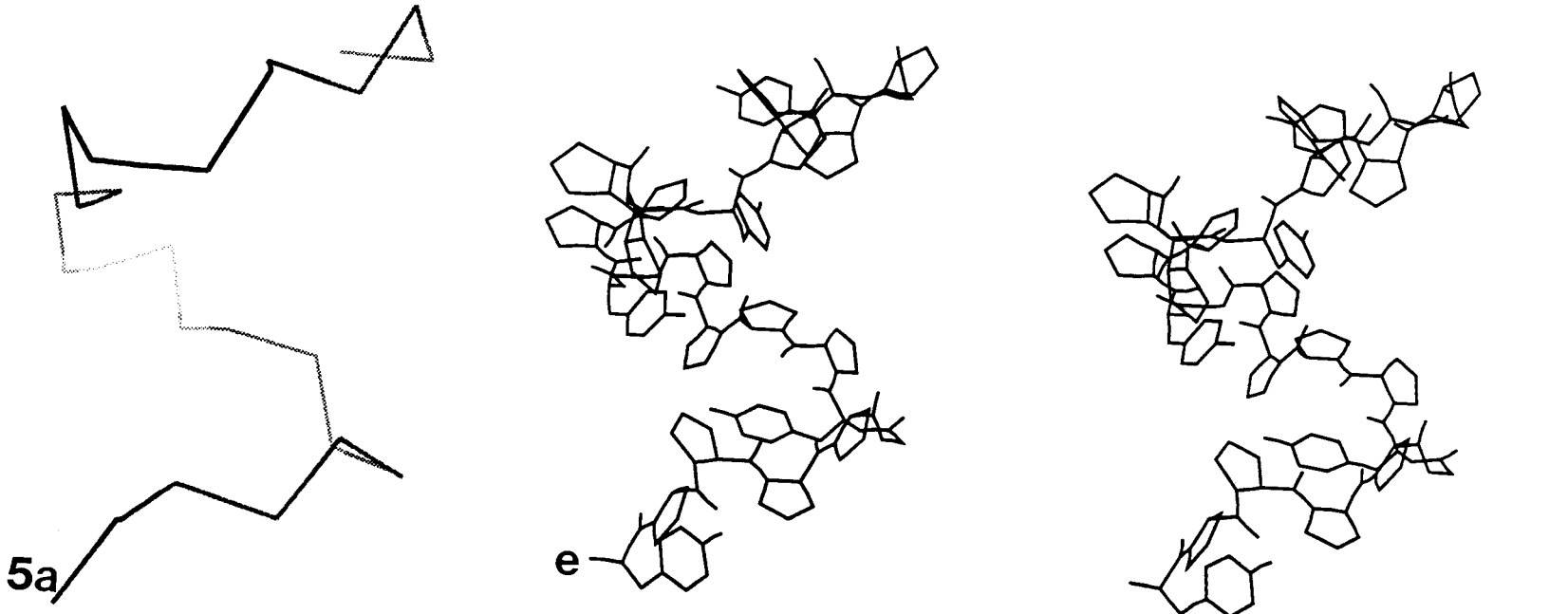


Fig. 5. sxIII56

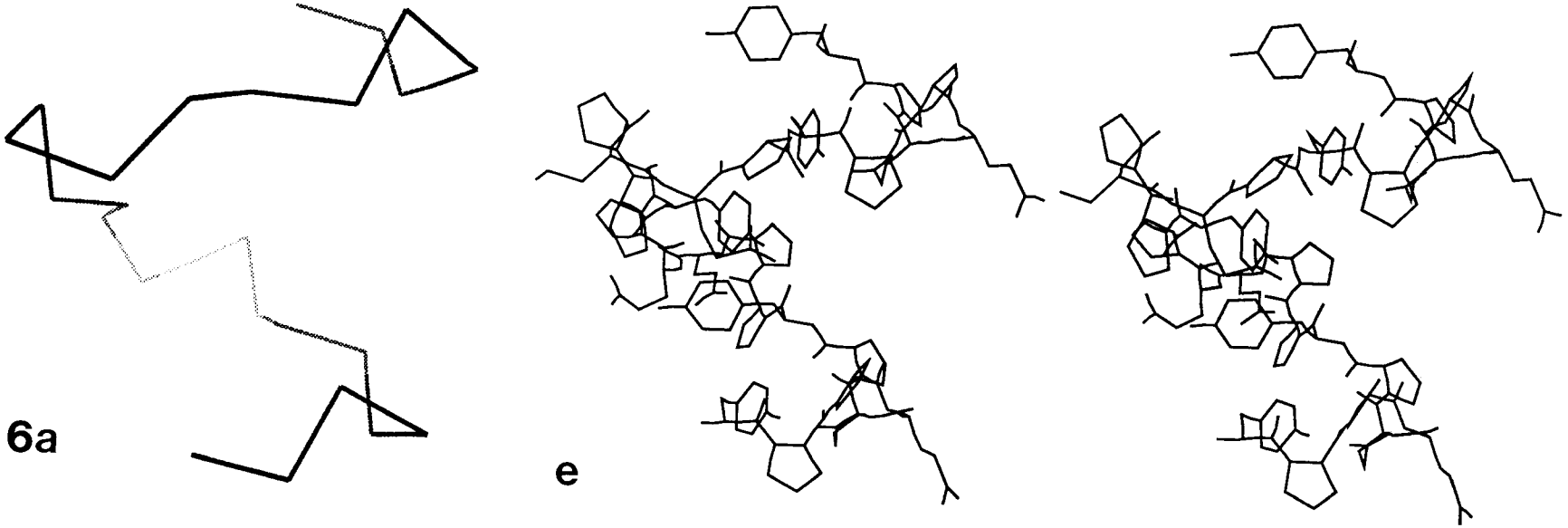
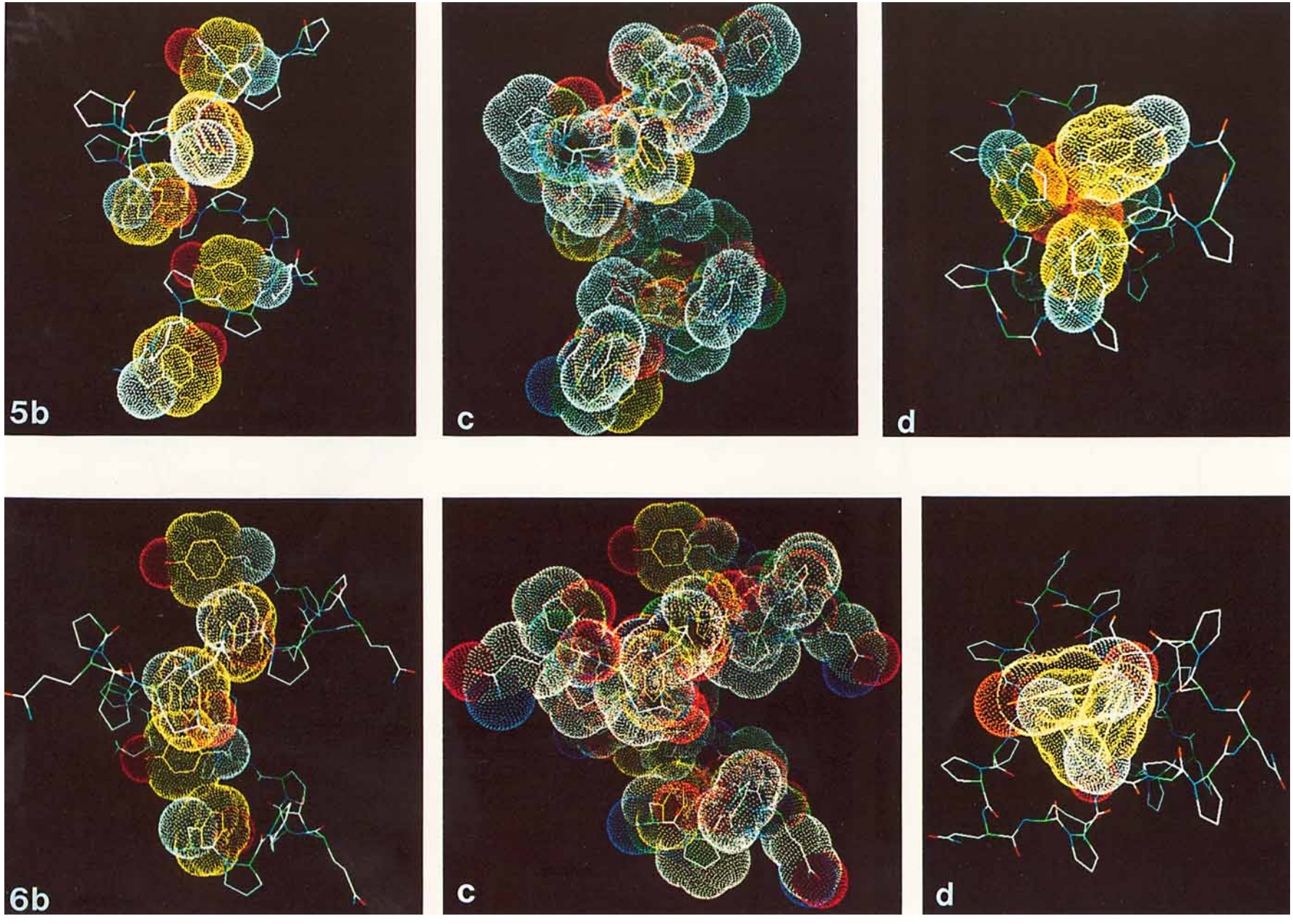
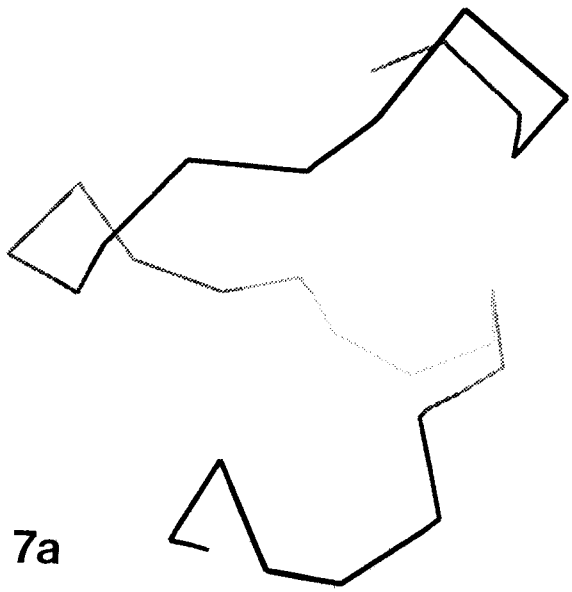


Fig. 6. gIIII34.


5b
c
d
6b
c
d



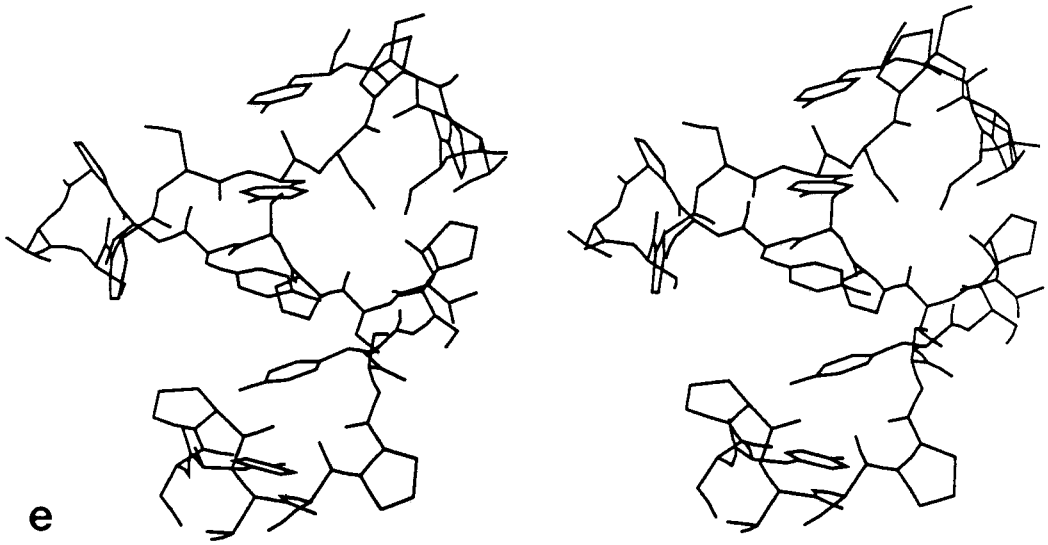


Fig. 7. rpIII34.

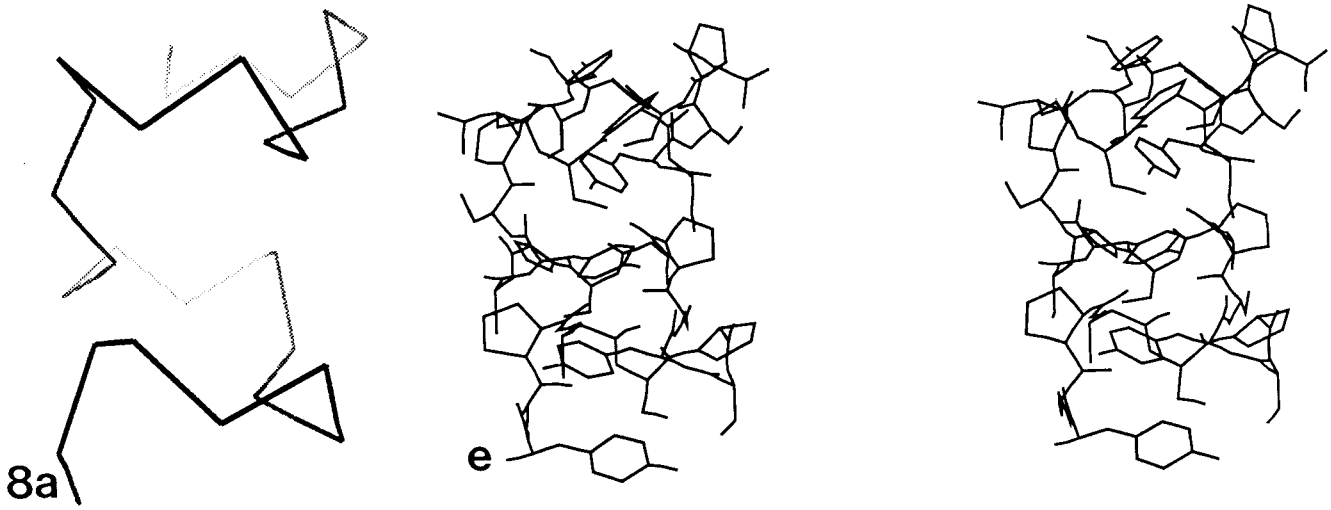
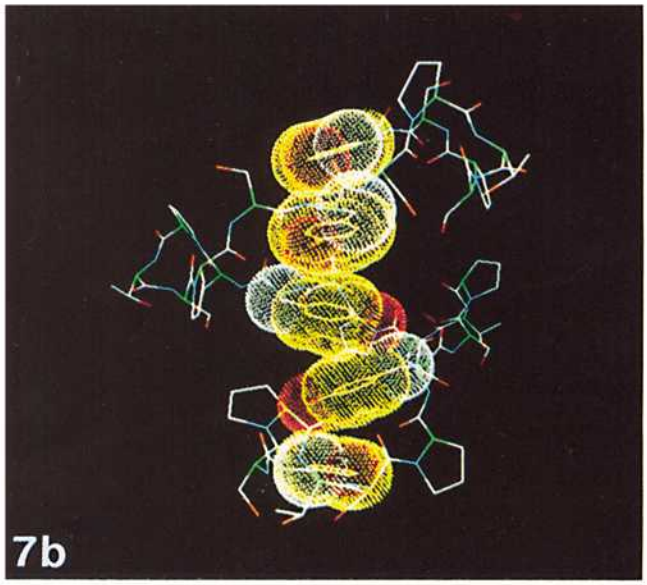
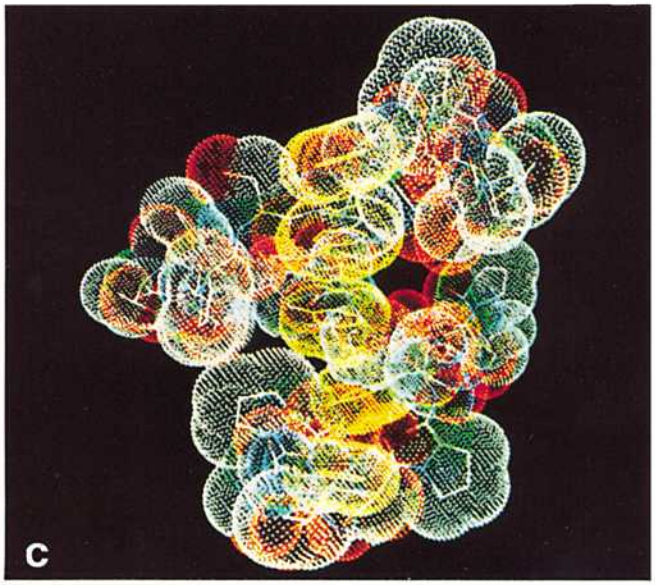


Fig. 8. rpIII34III67.





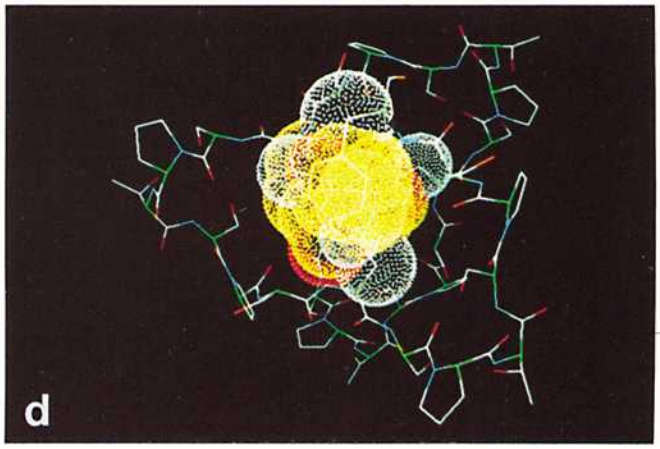
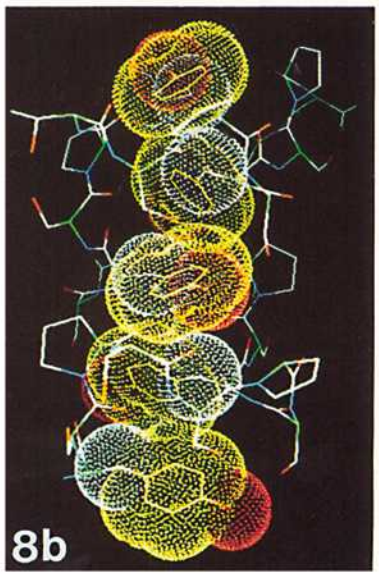
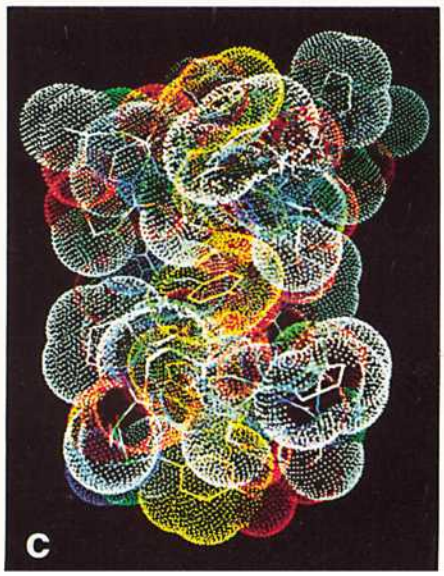


Fig. 7. b,c,d.





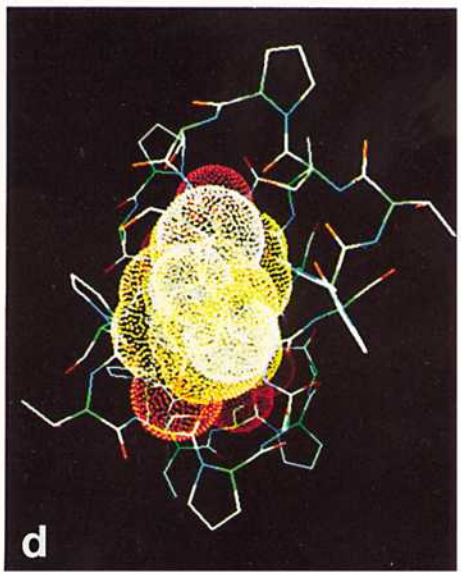


Fig. 8. b,c,d.

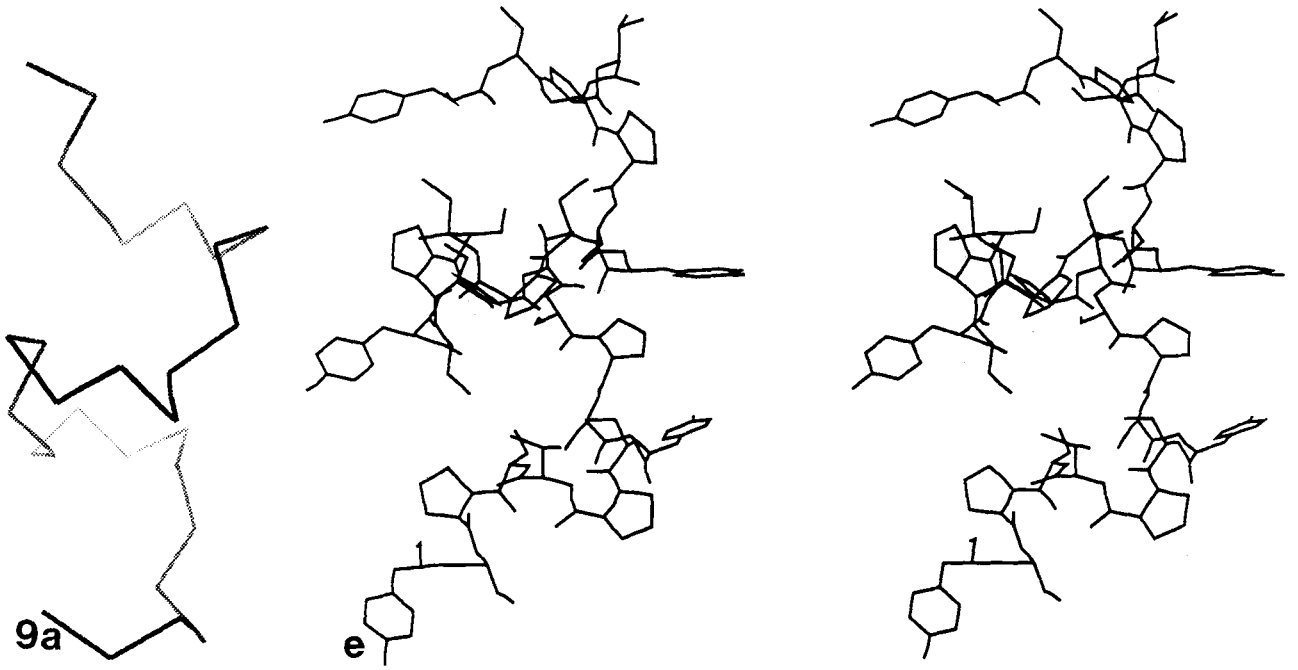


Fig. 9. rpIII34III71.

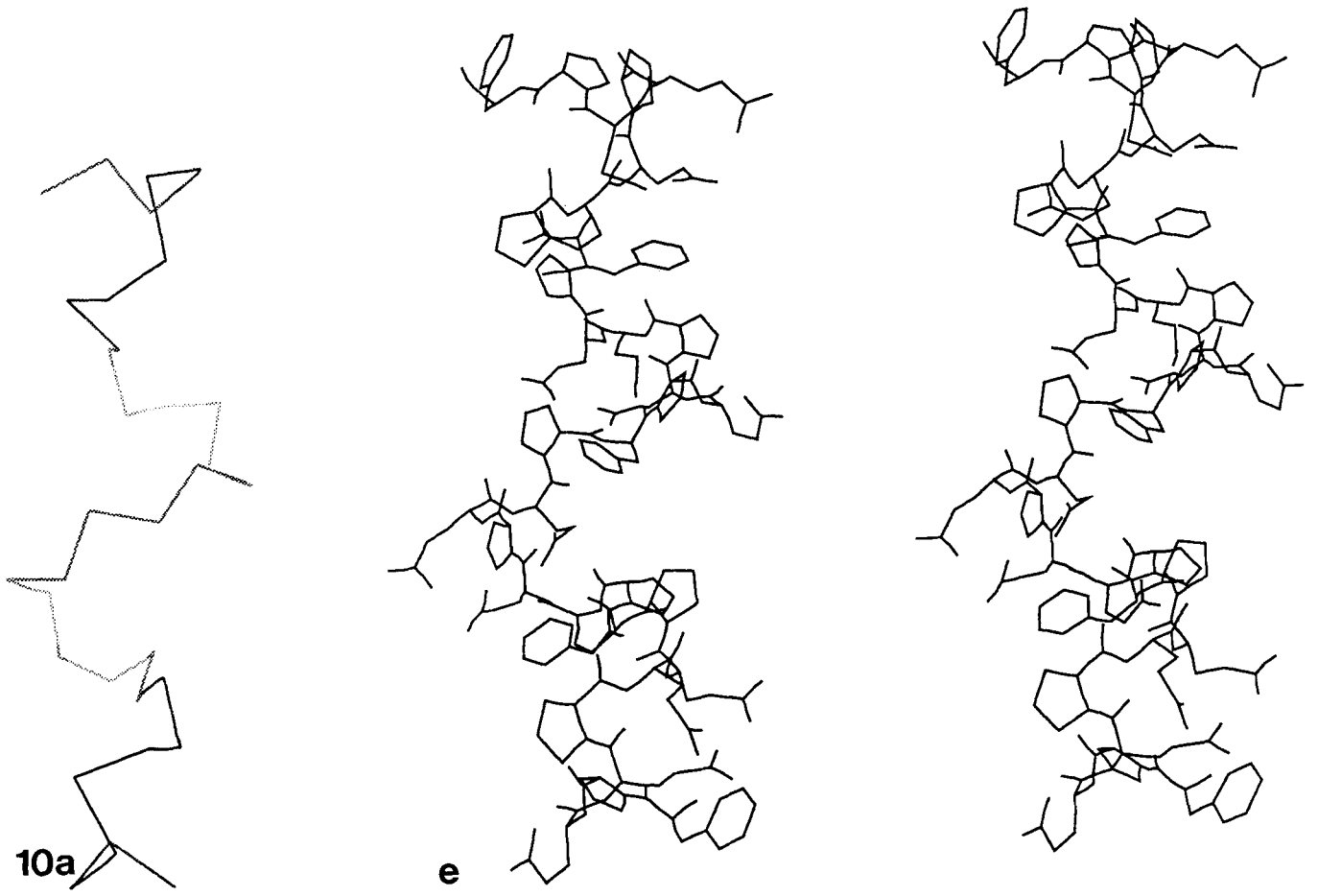
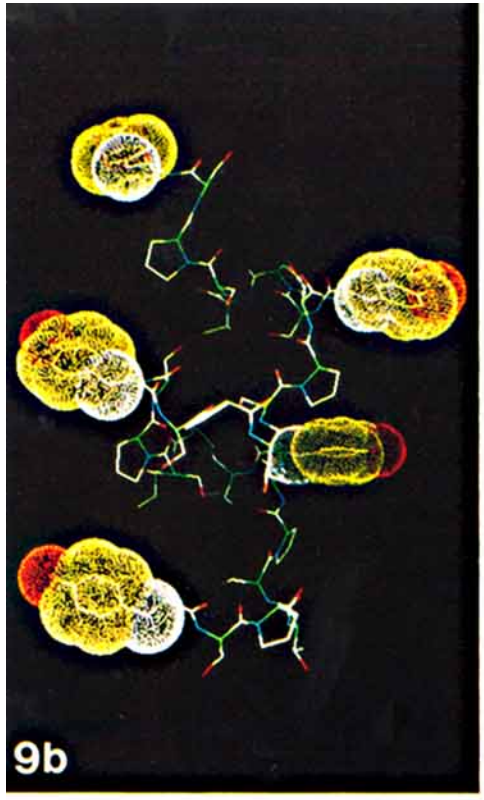


Fig. 10. hdIII23III56.



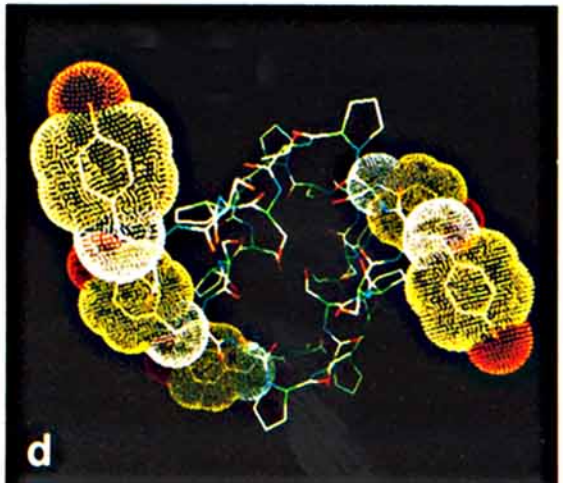
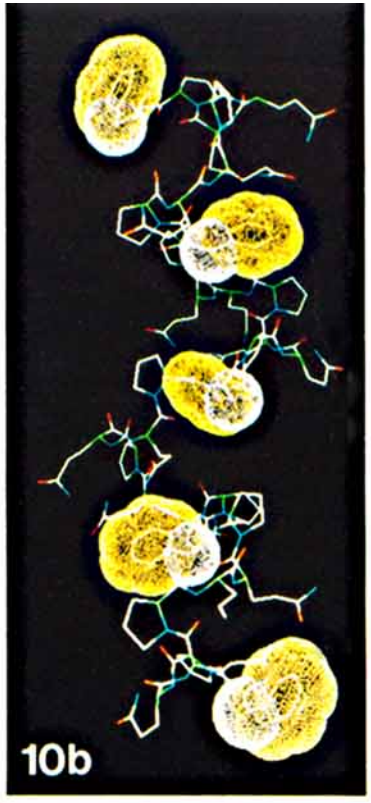
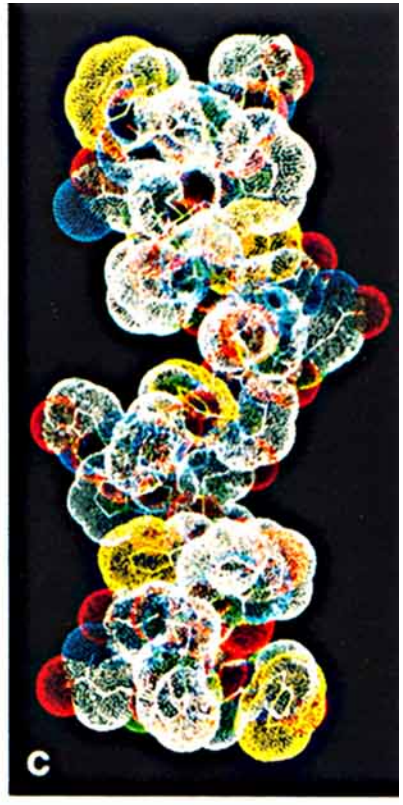


Fig. 9. b,c,d.





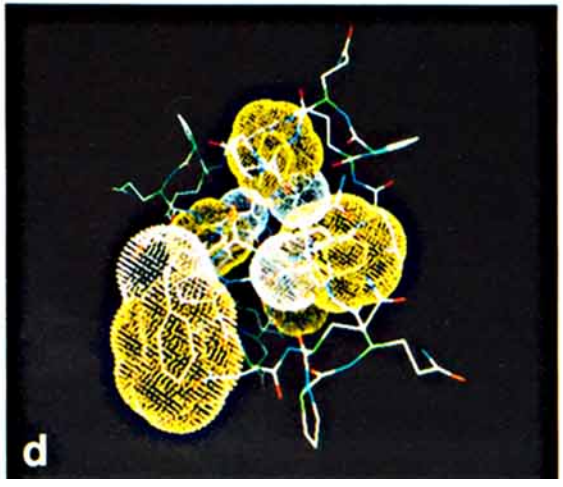


Fig. 10. b,c,d.

acceptable for rhodopsin are acceptable for synaptophysin. Even so, the hydrophobic volume of the side chain of Pro-2 in contrast to Gly-2 leads to a different refinement.

From the 10 more refined models plus the 31 judged to be acceptable we see several general characteristics:

1. Repeating oligopeptides having polyproline dihedral angles ($\phi \sim -60°$, $\psi \sim 140°$) and one or two β-turns, with or without Tyr-*cis*-Pro peptide bonds, form very reasonable models for all seven proteins—rhodopsin, synaptophysin, synexin, gliadin, RNA polymerase II, hordein, and gluten.
2. For each protein there are several acceptable models that, when adjusted, show no bad van der Waals contacts and that do have reasonable compactness and some van der Waals contact between strands of the helix.
3. Among the acceptable models we tried to form additional H-bonds and to optimize hydrophobic van der Waals contacts by burying Tyr, the constant feature of all seven consensus oligopeptides.
4. We have modeled consensus sequences that accommodate Pro at many positions. Further, nearly all positions except $i + 2$ and $i + 3$ in the β-turn and Tyr in the Tyr-cis-Pro can accommodate Pro in a nonconsensus position. The actual sequences of these proteins, as they differ from their respective consensus, can be accommodated by slight changes in the dihedral angles of the model structures or by combining components of two model structures. Changes in length of the consensus can be built maintaining the overall helical course of the model by introducing occasional bulges in the regular structure.

Relation of the Proposed Models to the Characteristics of the Proteins

In most cases we have little functional or structural information by which to evaluate our models. Since many of the naturally occurring sequences have deviations from the consensus sequences we modeled, the actual protein structures may be mosaics of several pro-β helices. Such diversity of structure may also occur even when the consensus sequence is strictly observed if the different conformations possible for a given sequence represent different functional states. The characteristic physical parameters for these domains in proteins could thus be composites of those represented by several models. Nonetheless, there are some interesting general predictions that can be drawn from the models that need to be tested by comparison with the properties of the proteins.

First the models lead to statements about the size of these domains in proteins. The most detailed physical data for these proteins have been obtained with the plant proteins. Unfortunately, little functional information is available for these three—gliadin, hordein, and gluten. It would, however, be premature to relegate them to a form of amino acid storage with reduced osmotic pressure. They have, respectively, 2 of 7, 4 of 8, and 4 of 9 hydrophilic residues (Gln, Ser, or Thr). There is no evidence of aggregation in solution. Hordein and gluten are suggested to be rod shaped with diameters about 18 Å.[17,19] The models for hordein in Table III have diameters 14–16 Å. As noted, multiple polypro, β-turn models can be built for gluten; they have a similar range of diameters.

From our results with the heptapeptide consensus of hordein and even from the hexapeptide repeat of RNA polymerase II we see that an octapeptide with only one β-turn would be very open and elongated and could not interact with adjacent octapeptides to form a stable structure. From the gluten models that we built, not shown in Table III, we concluded that at least two turns are required to form a reasonably compact structure.

One speculation concerning the function of the repeating sequence domain in RNA polymerase II is that it may interact with regulatory proteins that bind to specific sequences on the DNA template.[11] Although no physical data exist for the size of this domain, our models predict that the domain in the murine polymerase would be 276 Å long. This is equivalent to the length of 80 bases of DNA so regulatory proteins binding to the first 100 bases upstream from the transcription start site might readily interact with this domain.

All of these pro-β helices have been built assuming, as is seen in vitro for the hordeins, that the helices are stable as monomers. In fact this situation is seldom observed for α-helices. Although they derive stability from intrastrand hydrogen bonds they are always, with the known exceptions of calmodulin and troponin C, found in contact with other structural components. It is certainly possible that any or all of these pro-β helices interact with other structures for additional stabilization.

A second set of predictions from the models concerns the possible impact of phosphorylation on the structures. Several of the models for the animal proteins include stabilizing hydrogen bonds involving side chains of Ser, Thr, or Tyr. Phosphorylation of the repeat domain in RNA polymerase apparently has important functional consequences. This domain is found to be phosphorylated in vivo when the polymerase is transcriptionally active.[31] The polymerase can also be phosphorylated in this domain in vitro by casein kinase I.[32] Stoichiometric analysis indicates at least 10 to 20 phosphates are incorporated in vitro. Ser is the primary residue phosphorylated, with some Thr, but no Tyr. The phosphorylation leads to a decreased mobility on SDS gels. Interestingly, very little or no intermediate forms are evident on gels of this subunit phosphorylated in vitro: Either no phosphate is incorporated or more

than 10 mol is incorporated in an all-or-none reaction. This suggests that the phosphorylation of the Sers is highly cooperative and could be associated with an unfolding of the pro-β helix as hydrogen bonds involving the Sers are sequentially broken. Since two of the refined models for the polymerase have the Tyr side chain relatively sequestered, it may be that the Tyr would become accessible to Tyr kinases only after serine kinases unfold the structure. This should be examined in both in vitro and in vivo systems.

Some of the synaptophysin models predict the involvement of the Tyrs in helix-stabilizing hydrogen bonds. Phosphorylation of Tyrs in synaptophysin, probably within this domain, occurs in intact synaptosomes and isolated synaptic vesicles.[7] Since this domain is on the cytoplasmic side of the vesicle membrane, conformational changes in the pro-β helix induced by Tyr phosphorylation could regulate interactions of the vesicle with cytoskeletal elements, the presynaptic plasma membrane, or clathrin during the life cycle of the vesicle.

The repeat domain of synaptophysin also appears to be responsible for the calcium-binding properties of this protein,[6] possibly through the four Asps that interrupt the repeating sequence motif. Since the evidence for binding calcium was determined by gel overlay experiments, the magnitude of the affinity is not known. The Asps would seem not to be clustered as they occur at separations of 14, 11, and 16 residues. However, the natural sequence of this domain has some deviations from the pentapeptide repeat length, including one tetrapeptide, three hexapeptides, and one heptapeptide. Perhaps this sequence forms a bent helix. A composite model consisting of variations of pro-β helices needs to be built to determine the possible physical relationships of the Asp residues and the role they might play in binding calcium.

The function of the C-terminal tail of octopus rhodopsin remains unknown; however, it is interesting that the vertebrate rhodopsins lack these pentapeptide repeats. Model rdIII34I45 contains two β-turns. The side chains of Tyr are buried and are in partial contact, thereby forming a continuous hydrophobic core. There are surprising similarities between synaptophysin and octopus rhodopsin. Both are integral membrane proteins anchored in intracellular membranes of neuronal cells by a number of *trans*-membrane helices (four in synaptophysin, seven in rhodopsin). Both have pentapeptide sequence repeats that could form pro-β helices at the carboxy/cytoplasmic termini. It will be important to determine whether the octopus rhodopsin helix can be phosphorylated, and whether the absence of Asp is correlated with an absence of calcium-binding activity.

Synexin, like synaptophysin, has been suggested to play a role in calcium-dependent interactions of the secretory vesicle/synaptic vesicle membrane during exocytosis. However, in the case of synexin, association with the membrane would be transient, associated with increases in intracellular calcium, whereas, synaptophysin is irreversibly anchored to the vesicle membrane. Since Asp is absent from the synexin pro-β helix, the calcium sensitivity of its function might be imparted only by the control calcium has over its presence on the membrane. The recent sequencing of a cDNA for synexin reveals that it is a member of the annexin class of proteins that is characterized by four internally homologous domains involved in binding lipids and calcium and a unique N-terminus.[33] It is the N-terminal domain of each of the seven characterized annexins that imparts distinct properties to each member of this class.[34] In some cases the N-terminal domain is the site of phosphorylation or association with accessory subunits. Alteration of this domain by phosphorylation, subunit association, or proteolysis has dramatic effects on the calcium sensitivity of two members of the annexin class, lipocortin I and calpactin I.[35,36] Only in the case of synexin does the N-terminal sequence suggest the formation of a pro-β helix. It will be important to determine whether the Tyrs in this sequence undergo phosphorylation and the structural/functional changes that this might induce in synexin.

Relevance to Other Proteins

Many other proteins contain or consist of tandem repeats resembling the seven modeled in this report. They were not included because they lack either the signal Tyr (or Phe) or Pro. It is reasonable that the principles and procedures described here might apply to these additional proteins as well.

The structures proposed for the pentapeptide repeat of rhodopsin and synaptophysin can also accommodate the characteristic repeating sequences found in the end domains of epidermal keratin. The central portion of the keratin molecule is believed to form an α-helix that wraps around the central portion of the second molecule of a keratin dimer to form a coiled coil.[37] These coiled-coil structures are assembled to form the core of the intermediate filament. Proteolytic studies of intact filaments suggest that the end domains protude from this core in a regular fashion and may be involved in interaction of the filament with other cell constitutents.[38] The end domains may also play a role in filament assembly since the α-helical core domains, when isolated from proteolytic digests of keratin, do not form filaments. The end domains contain tandem repeats rich in Gly with periodic aromatic residues. In particular, the N-terminal domain of the murine 67 kDa type II keratin contains the sequence Tyr/Phe·Gly·Gly·Gly·Gly/Ser repeated 11 times.[39] Conspicuously, there is no Pro in this sequence domain, so the constraints that were essential to the model

building for synaptophysin and rhodopsin do not apply to keratin. Therefore, if the keratin sequence does adopt a well-defined secondary structure similar to those proposed here, it seems likely that stability of the structure would depend greatly on a unique role of the aromatic residues, such as formation of stabilizing interactions between the aromatic rings. We therefore suggest that the N-terminal end domain of keratin forms a right-handed pro-β helix stabilized by a hydrophobic core of interacting phenyl groups, as illustrated in model rdIII34I45 for rhodopsin. In the case of synaptophysin we suggested that phosphorylation of the repeating Tyr could modulate the structure of the carboxy-terminal pro-β helix. The conformation of the corresponding domain in keratin might also be regulated by phosphorylation since certain (unidentified) serine residues within or near the repeating sequence motif are phosphorylated in vivo.[38] If the end domains do adopt a pro-β helical conformation, they could extend as much as 60 Å from the keratin core to interact with other cytoplasmic elements.

Elastin imparts elasticity to arteries and skin. It has a high content of Pro, Gly, and Val but does not give a well-defined X-ray diffraction pattern similar to collagen, nor to silk with a high β-sheet content, nor to keratin with a high α-helix content. Gray et al.[40] proposed an "oiled-coil" model in which a sequence such as Pro·Gly·Gly·Val would form a type I β-turn. "When several such blocks are linked, they make a broad left-handed helix with bulky side chains in the centre and glycines on the outside." This is similar in principle to the pro-β helix structures that have a core of bulky aromatic sidechains surrounded by spaced β-turns. From their ^{13}C NMR studies Urry and Long[41] proposed a high content of β-turns in elastin. Venkatachalam and Urry[42] subsequently proposed a "β-spiral" model for the Val·Arg·Gly·Val·Gly repeat of elastin.

The A1 protein of the mammalian heterogeneous nuclear ribonucleoprotein complex has a C-terminal region of 120 amino acids with repeats following a consensus sequence Tyr·Gly·Gly·Gly·Arg·Gly·Asn. Cobianchi et al.[43] suggested that the aromatic residues might interact directly with the nucleic acid bases. This might occur if this domain adopted a pro-β helix conformation with the Tyr extended away from the helix axis. However, as our modeling efforts have shown, the Tyr residues might alternatively play an important role in the interior of the protein helix.

Roth and Gall[44] "isolated a cDNA clone that encodes a protein on lampbrush chromosome loops of the newt *Notophthalmus*." The N-terminus of this M_r 68,400 protein has 25 tandem repeats of a dodecamer motif whose consensus is Phe·Pro·Arg·Gly·Met/Leu·Glu·Gly·Arg·Pro·Arg·Gly·Asp. With six residues between Pros an α-helical segment is not precluded. Evenso, at least portions of the oligomer may resemble the pro-β helices described here.

The extensins are the most abundant proteins of cell walls of dicots (see review by Varner and Lin).[45] Several sequences of tandem repeats are available. Soybean extensin has 23 repeats of the pentamer Tyr·Lys·Pro·Pro·Val.[46] This sequence is compatible with the pro-β helix structural motif. However, as a family, the extensins include a number of repeating sequences without Tyr, and the proteins evidently interact extensively in the cell wall with other proteins and polyglycans and may not have a well-defined structure as monomeric molecules. We did not model this family in detail, although preliminary work with soybean extensin suggests a similar approach should be successful, at least in the case of this one member of the family. Examples of sequences include the following:

Tomato
P1 Tyr·Lys·Ser·Hyp·Hyp·Hyp·Hyp·Thr·Hyp·Val
and
Tyr·Lys·Ser·Hyp·Hyp·Hyp·Hyp·Val·Lys·Pro·
Tyr ·His·Pro·Thr·Hyp·Val
P2 Tyr·Lys·Ser·Hyp·Hyp·Hyp·Hyp·Val·Tyr·
Lys (entire)
and Ser·Pro·Pro·Pro·Pro·Val·Ala
and Ser·Pro·Pro·Pro·Pro·Val·His

Carrot
Ser·Pro·Pro·Pro·Pro (×25)
and Ser·Pro·Pro·Pro·Pro·Thr·Pro·Val (×8)
and Tyr·Lys·Tyr·Lys (×12)
and Tyr·Thr·Pro·Pro·Val·His·Lys·Pro·Pro·Val
(×3)

Tobacco
Ser·Pro·Pro·Pro·Pro (×17)
and Ser·Pro·Pro·Pro·Pro·Thr·Pro·Val (×7)
and Tyr·Lys·Ser·Pro·Pro·Pro·ProLys·Lys·Pro·
Tyr·Tyr ·Pro·Pro·His·Thr·Pro·Val (×8)

Bean
Tyr·Tyr·Tyr·Lys·Ser·Pro·Pro·Pro·Pro·Ser·Pro·
Ser·Pro·Pro·Pro
and Tyr·Tyr·Tyr·His·Ser·Pro·Pro·Pro·Lys·His·
Ser·Pro·Pro·Pro·Pro

Soybean
Tyr·Lys·Pro·Pro·Val (×23)
and Pro·Pro·His·Glu·Lys·Pro·Pro (×17)

The basic building block of the pro-β helix, a polyproline-like structure followed by a β-turn, has previously been noted occurring in isolation in a number of proteins. Ananthanasayanan et al.[47] analyzed 40 globular proteins of known structure and found examples of a polyproline like structure eight of which are immediately followed by a β-turn, a so-

called c-βt structure. Three to seven residues have ϕ, ψ values near $-60°$, $+130°$ thereby forming at least one turn of a left-handed helix resembling polyproline. It is interesting that these $-60°$, $+130°$ helices are not preceded by β-turns. Types I, II, and III β-turns are all seen following the $-60°$, $+130°$ helix; in one case there is a double β-turn. In the eight polyproline helixes five Pros occur. Four terminate the helix, once with a double Pro, and immediately precede the β-turn. There are no obvious preferences for residue type or position.

CONCLUSION

The pro-β structures presented in this communication were developed using consensus sequences for the given proteins. Therefore, the models represent only ideal scaffolds for the natural sequences. By accepting Pro in the consensus sequence even at positions where it occurs less frequently in the natural sequence we have attempted to develop models that can accommodate all variations that occur naturally in the composition of the repeats. The helices proposed represent a class of novel secondary structures. As such, one cannot define in a general sense what their functions are, anymore than one can give a general definition for the function of the α-helix. Instead, the structures probably serve a wide variety of functions, reflecting the diversity of the proteins in which they occur. The differences in the lengths of the oligomers are a strong indication that these structures are the result of convergent evolution, and thus it is not surprising to see them occur in such unrelated proteins. We emphasize that these models do not represent energy minima for the naturally occurring sequences; rather that they demonstrate the plausbility of building several related, stable helices using the initial assumptions.

Our analysis has not converged on a unique structure, even for a single protein. This may be an indication of an important functional property: The ability to adopt different, but related, conformational states, perhaps, in some cases, as a consequence of phosphorylation. This is a degree of flexibility and responsiveness not readily available to conventional secondary structures such as the α-helix. This may be the one common attribute of the pro-β helices that is of such value that it explains the repeated evolution of these structures in such diverse contexts.

NOTE ADDED IN PROOF:

Following submission of this manuscript we became aware of six additional proteins that may contain the pro-β helix motif.

Lemaire et al.[48] determined the sequence of cDNA encoding Krox-24, a mouse transcription factor 505 residues long that contains three zinc fingers. They noted a tandem repeat near the C-terminus similar to that in the largest subunit of eukaryotic RNA polymerase.

$_{412}$Tyr Pro Ser Pro Val Ala Thr Ser
Tyr Pro Ser Pro Ala Thr Thr Ser
Phe Pro Ser Pro Val Pro Thr Ser
Tyr Ser Ser Pro Gly Ser Ser Thr
Tyr Pro Ser Pro Ala His Ser Gly
Phe Pro Ser Pro Ser Val Ala Thr Thr
Phe Ala Ser Val Pro Pro Ala
Phe Pro Thr Gln Val Ser Ser
Phe Pro Ser Ala Gly Val Ser Ser Ser
Phe Ser Thr Ser Thr Gly Leu Ser Asp Met Thr Ala Thr
Phe Ser Pro Arg Thr Ile Glu Ile Cys$_{505}$

Eldon et al.[49] have characterized the expression of a mRNA (spec3) from *Strongylocentrotus purpuratus* ectoderm. The encoded protein is 208 residues long. Two strongly hydrophobic helices near the C-terminus may be associated with a membrane. The role of these following seven repeats near the N-terminus remains unknown, but a portion of this region may well assume a pro-β conformation.

$_{9}$Ala Pro Thr Asp Ala Pro Pro Ala
Tyr Pro Pro Pro Pro Gln Gln
Ala Pro Pro Pro Gln Gln Pro Gly
Tyr Gly Gln Pro Gln Leu Gly
Tyr Gly Gln Pro Pro Pro Gln Leu Gly
Tyr Gly Gln Pro Pro Pro Gln Leu Gly
Tyr Gly
Tyr Pro Pro Pro Pro Gln Asn Asn Asn$_{67}$

Antonsson et al (1989)[50] "determined the sequence of a cDNA clone encoding the keratan sulfate-rich domain of the large aggregating cartilage proteoglycan core protein." "It consists of a highly conserved hexapeptide motif Phe Pro Ser Glu Glu/Lys Pro consecutively repeated 23 times." They propose ". . . that the domain has an extended structure," and ". . . forms a space for other cartilage matrix molecules within the aggregate."

Franssen et al (1987)[51] determined the sequence of a cDNA encoding nodulin, a protein found in soybean nodules appearing about ten days after sowing and inoculation with *Rhizobium*. Although its apparent molecular weight by gel mobility is 75,000, that derived from its sequence is 45,000. The C-terminal half of the molecule consists of 22 tandem repeats of Tyr Gln Pro Pro His Glu Lys Pro Pro Pro Glu. Franssen et al noted that nodulin is not homologous to other (hydroxy) Pro rich proteins found in plants and that it is a structural protein.

Hong et al (1987)[52] determined the sequence of a cDNA encoding a soybean protein of 256 amino acids. It consists 43 near identical repeats of Tyr Lys Pro Pro Val. "The presence of a signal sequence on this" . . . "protein is suggestive that it may be a

secreted extracellular, cell wall protein." Hong et al noted that only one base change related this consesus cDNA to that encoding His Lys Pro Pro Val consensus found by Chen and Varner (1985)[53] to be repeated 36 times in a carrot protein. As noted in this MS, Averyhart-Fullard et al (1988)[46] sequenced a similar cDNA from soybean encoding Tyr Lys Pro Pro Val and characterized the corresponding 33 kDa protein. Half of its Pro's are hydroxylated. Even though it is glycosylated, they argue that it differs from the recognized "hydroxyproline-rich glycoproteins"—arabinogalactan proteins, lectins, and extensins. All of these authors have suggested some sort of structural role for the protein.

Keller and Lamb (1989)[54] deduced the amino acid sequence of a novel cell wall hydroxyproline-rich glycoprotein from a genomic library from tobacco. The gene contains no introns; the protein is 620 amino long. From residue 179 to the c-terminus this HRGPnt3 contains 57 tandem repeats, 26 of which are seven residues long with a strong consensus Tyr Ser Pro Pro Pro Pro Thr. The other 31 range from six to twelve long; all have high Pro content and most start with Tyr.

ACKNOWLEDGMENTS

We are pleased to thank J. R. Klein for help in model building and evaluation and S. L. Snyder for patience and expertise in photography. This research has been supported by grants from the Council for Tobacco Research (#1933R1), N.I.H. (DK33151), and N.S.F. (DCB8602705). CEC was supported by an Established Investigator Award from the American Heart Association with funds contributed in part by the Virginia Affiliate.

REFERENCES

1. Creutz, C.E., Snyder, S.L., Husted, L.D., Beggerly, L.K., Fox, J.W. Pattern of repeating aromatic residues in synexin. Similarity to the cytoplasmic domain of synaptophysin. Biochem. Biophys. Res. Commun. 152:1298–1303, 1988.
2. Creutz, C.E., Pazoles, C.J., Pollard, H.B. Identification and purification of an adrenal medullary protein (synexin) that causes calcium-dependent aggregation of isolated chromaffin granules. J. Biol. Chem. 253:2858–2866, 1978.
3. Creutz, C.E. Cis-unsaturated fatty acids induce the fusion of chromaffin granules aggregated by synexin. J. Cell Biol. 91:247–256, 1981.
4. Südhoff, T.C., Lottspeich, F., Greengard, P., Mehl, E., Jan, R.A. A synaptic vesicle protein with a novel cytoplasmic domain and four transmembrane regions. Science 238: 1142–1144, 1987.
5. Buckley, K.M., Floor, E., and Kelly, R.B. Cloning and sequence analysis of cDNA encoding p38, a major synaptic vesicle protein. J. Cell Biol. 105:2447–2456, 1987.
6. Rehm, H., Wiedenmann, B., Betz, H. Molecular characterization of synaptophysin, a major calcium-binding protein of the synaptic vesicle membrane. EMBO J 5:535–546, 1986.
7. Pang, D.T., Wang, J.K.T., Valtorta, F., Benfenati, F., Greengard, P. Protein Tyr phosphorylation in synaptic vesicles. Proc. Natl. Acad. Sci. U.S.A. 85:762–766, 1988.
8. Ovchinnikov, Y.A., Abdulaev, N.G., Zolotarev, A.S., Artamonov, I.D., Bespalov, I.A., Dergachev, A.E., Tsuda, M. Octopus rhodopsin. Amino acid sequence deduced from cDNA. FEBS Lett. 232:69–72, 1988.
9. Allison, L.A., Moyle, M., Shales, M., Ingles, C.J. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell 42:599–610, 1985.
10. Corden, J.L., Cadena, D.L., Ahearn, J.M., Dahmus, M.E. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc. Natl. Acad. Sci. U.S.A. 82:7934–7938, 1985.
11. Sigler, P.B. Transcriptional activation acid blobs and negative noodles. Nature (London) 333:210–212, 1988.
12. Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S., and Greenleaf, A.L. The C-terminal repeat domain of RNA polymerase II largest subunit is esesntial in vivo but is not required for accurate transcription invitation in vitro. Proc. Natl. Acad. Sci. U.S.A. 85:3698–3702, 1988.
13. Rafalski, J.A., Scheets, K., Metzler, M., Peterson, D.M., Hedgcoth, C., Soll, D.G. Developmentally regulated plant genes: The nucleotide sequence of a wheat gliadin genomic clone. Embo J 3:1409–1415, 1984.
14. Okita, T.W., Cheesbrough, V., Reeves, C.D. Evolution and heterogeneity of the α-/β-type and α-type gliadin DNA sequences. J. Biol. Chem. 260:8203–8213, 1985.
15. Forde, B.G., Kreis, M., Williamson, M.S., Fry, R.P., Pywell, J., Shewry, P.R., Bunce, N., Miflin, B. Short tandem repeats shared by β- and C-hordein cDNAs suggest a common evolutionary origin for two groups of cereal storage protein genes. EMBO. J. 4:9–15, 1985.
16. Tatham, A.S., Drake, A.F., Shewry, P.R. A conformational study of a glutamine- and Pro-rich cereal seed protein, C Hordein. Biochem. J. 226:557–562, 1985.
17. Field, J.M. Tatham, A.S., Baker, A.M., Shewry, P.R. The structure of C hordein. FEBS Lett. 200:76–80, 1986.
18. Halford, N.G., Forde, J., Anderson, O.D., Greene, F.C., Shewry, P.R. The nucleotide and deduced amino acid sequences of an HMW glutenin subunit gene from chromosome 1B of bread wheat (*Triticum aestivum* L.) and comparison with those of genes from chromosomes 1A and 1D. Theor. Appl. Genet. 75:117–126, 1987.
19. Field, M., Tatham, A.S., and Shewry, P.R. The structure of a high-M_r subunit of durum-wheat (*Triticum durum*) gluten. Biochem. J. 247:215–221, 1987.
20. Chou, Y., Fasman, G.D. β-turns in proteins. J. Mol. Biol. 115:135–175, 1977.
21. Garnier, J., Osguthorpe, D.J., Robson, B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120:97–120, 1978.
22. Ramachandran, G.N., Sasisekharan, V. Conformation of polypeptides and proteins. Adv. Protein Chem. 23:283–247, 1968.
23. Wilmot, C.M., Thornton, J.M. Analysis and prediction of the different types of β-turn in proteins. J. Mol. Biol. 203: 221–232, 1988.
24. Schultz, G.E., Schirmer, R.H. "Principles of Protein Structure. New York: Springer-Verlag, 1979: 25–26.
25. Huber, R., Steigemann, W. Two cis-Pro's in the Bence-Jones protein Rei and the cis-Pro-turn. FEBS Lett. 48: 235–237, 1984.
26. Grathwohl, C., Wüthrich, K. The X-Pro peptide bond as an NMR probe for conformational studies of flexible linear peptides. Biopolymers 15:2025–2041, 1976.
27. Linn, L-N., Brandts, J.F. Role of cis-trans isomerism of the peptide bond in protease specificity. Kinetic studies on small Pro-containing peptides and on polyproline. Biochemistry 18:5037–5042, 1979.
28. Chazin, W.J., Kördel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundström, T., Forsén, S. Proline isomerism leads to multiple folded conformations of calbindin D_{9K}: direct evidence from two-dimensional ^1H NMR spectroscopy. Proc. Natl. Acad. Sci. U.S.A. 86:2195–2198, 1989.
29. Burley, S.K., Petsko, G.A. Aromatic-aromatic interaction: A mechanism of protein structure stabilization. Science 229:23–28, 1985.
30. Dempsey, S. Molecular modeling system (MMS). Department of Chemistry Computers Facility, Univ. of California at San Diego, 1987.
31. Cadena, D.L., Dahmus, M.E. Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. J. Biol. Chem. 262:12468–12474, 1987.
32. Dahmus, M.E. Phosphorylation of eukaryotic DNA-depen-

dent RNA polymerase. J. Biol. Chem. 256:3332–3339, 1981.
33. Pollard, H.B., Burns, A.L., Rojas, E. A molecular basis for synexin-driven, calcium-dependent membrane fusion. J. Exp. Biol. 256:3332–3339, 1988.
34. Crompton, M.R., Moss, S.E., Crumpton, M.J. Diversity in the lipocrotin/calpactin family. Cell 55:1–3, 1988.
35. Schlaepfer, D.D., Haigler, H.T. Characterization of Ca^{2+}-dependent phospholipid binding and phosphorylation of lipocortin I. J. Biol. Chem. 262:6931–6937, 1987.
36. Drust, D.S., Creutz, C.E. Aggregation of chromaffin granules by calpactin at micromolar levels of calcium. Nature (London) 331:88–91, 1988.
37. Steinert, P.M., Parry, D.A.D., Racoosin, E.L., Idler, W.W., Steven, A.C., Trus, B.L., Roop, D.R. The complete cDNA and deduced amino acid sequence of a type II mouse epidermal keratin of 60,000 Da: Analysis of sequence differences between type I and type II keratins. Proc. Natl. Sci. U.S.A. 81:5709–5713, 1984.
38. Steinert, P.M., Rice, R.H., Roop, D.R., Trus, Benes, L., Steven, A.D. Complete amino acid sequence of a mouse epidermal keratin subunit and implications for the structure of intermediate filaments. Nature (London) 302:794–800, 1983.
39. Steinert, P.M., Parry, D.A.D., Idler, W.W., Johnson, L.D., Steven, A.C., Roop, D.R. Amino acid sequences of mouse and human epidermal type II keratins of Mr 67,000 provide a systematic basis for the structural and functional diversity of the end domains of keratin intermediate filament subunits. J. Biol. Chem. 260:7142–7149, 1985.
40. Gray, W.R., Sandberg, L.B., Foster, J.A. Molecular model for elastin structure and function. Nature (London) 246: 461–466, 1973.
41. Urry, D.W., Long, M.M. Conformations of the repeat peptides of elastin in solution: An application of proton and carbon-13 magnetic resonance to the determination of polypeptide secondary structure. CRC Crit. Rev. Biochem. 4:1–46, 1976.
42. Venkatachalam, C.M., Urry, D.W. Development of a linear helical conformation from its cyclic correlate. β-spiral model of the elastin poly(pentapeptide) $(VRGVG)_n$. Macromolecules 14:1225–1229, 1981.
43. Cobianchi, F., Karpel, R.L., Williams, K.R., Notario, V., Wilson, S.H. Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. Large-scale overproduction in *Escherichia coli* and *cooperative* binding to single stranded nucleic acids. J. Biol. Chem. 263:1063–1071, 1988.
44. Roth, M.B., Gall, J.G. Targeting of a chromsomal protein to the nucleus and to lampbrush chromosome loops. Proc. Natl. Acad. Sci. U.S.A. 86:1269–1272, 1989.
45. Varner, J.E., Lin, L.-S. Plant cell wall architecture. Cell 56:231–239, 1989.
46. Averyhart-Fullard, V., Datta, K., Marcus A. A hydroxyproline-rich protein in the soybean cell wall. Proc. Natl. Acad. Sci. U.S.A. 85:1082–1085, 1988.
47. Ananthanarayanan, V.S., Soman, K.V., Ramakrishnan, C. A novel supersecondary structure in globular proteins comprising the collagen-like helix and β-turn. J. Mol. Biol. 198:705–709, 1987.
48. Lemaire, P., Revelant, O., Bravo, O. and Charnay, P. Two mouse genes encoding potential transcription factors with identical DNA-binding domains are activated by growth factors in cultured cells. Proc. Natl. Acad. Sci. U.S.A. 85: 4691–4694, 1988.
49. Eldon, E.D., Angerer, L.M., Angerer, R.C., and Klein, W.H. Spec3: Embryonic expression of a sea urchin gene whose product is involved in ectodermal ciliogenesis. Genes Dev. 1:1280–1292, 1987.
50. Antonsson, P., Heinegard, D., Oldberg, A. "The Keratan Sulface-enriched Region of Bovine Cartilage Proteoglycan Consists of a Consecutively Repeated Hexapeptide Motif" J. Biol. Chem. 264:16170–16173, 1989.
51. Franssen, H.J., Nap, J.-P., Gloudemans, T., Stiekema, W., van Dam, H., Govers, F., Louwerse, J., van Kammen, A., Bisseling, T. Characterization of cDNA for Nodulin-75 of Soybean. A Gene Product Involved in Early Stages of Root Nodule Development Proc. Natl. Acad. Sci. 84:4495–4499, 1987.
52. Hong, J.C., Nagao, R.T., Key, J.L. Characterization and sequence analysis of a developmentally regulated putative cell wall protein gene isolated from soybean J. Biol. Chem. 262:8367–8376, 1987.
53. Chen, J., Varner, J.E. Isolation and Characterization of cDNA Clones for Carrot Extensin and a Proline-rich 33-kDa Protein Proc. Natl. Acad. Sci. 82:4399–4403, 1985.
54. Keller, B., Lamb, C.J. Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation Genes & Devel. 3:1639–1646, 1989.