

TURNS IN PEPTIDES AND PROTEINS

By GEORGE D. ROSE,* LILA M. GIERASCH,† and JOHN A. SMITH‡

* Department of Biological Chemistry, Hershey Medical Center,
Pennsylvania State University, Hershey, Pennsylvania

† Department of Chemistry, University of Delaware, Newark, Delaware

‡ Departments of Molecular Biology and Pathology, Massachusetts General Hospital,
Harvard Medical School, Boston, Massachusetts

I.	Introduction	1
A.	Goal of Review	1
B.	Terminology	2
II.	Turns in Peptides	4
A.	Model Peptides	6
B.	Naturally Occurring Peptides	36
III.	Turns in Proteins	49
A.	Identification of Turns from X-Ray Coordinates	50
B.	Location of Turns in Proteins	55
C.	Prediction of Turns	58
D.	Larger Loops	71
E.	Turns and Protein Folding	71
F.	Possible Roles of Turns as Recognition Sites in Proteins	74
	Appendix	89
	References	97

I. INTRODUCTION

A. Goal of Review

The goal of this article is to examine structural and functional roles of turns in peptides and proteins. Turns are a fundamental class of polypeptide structure, and are defined as sites where the peptide chain reverses its overall direction.

In the past 20 years, the peptide field has witnessed major development, stimulated by the discovery of a host of bioactive peptides. Turn structures have been proposed and implicated in the bioactivity of several of these naturally occurring peptides. In addition, many structural details of turns have been derived from conformational studies of model peptides.

During this same period, more than 100 complete protein structures have been elucidated in single-crystal X-ray studies. These examples document the rich diversity of structural patterns in the chain folds of native proteins. Attempts to resolve this molecular composite into basic

units of structure began with α helix and β sheet; each of these classes is apparent in physical models because each has backbone torsion angles with repeating values. By contrast, turns are not as conspicuous because their backbone torsion angles are nonrepeating. Nonetheless, they are regular structures which, once characterized, are observed to be abundant in globular proteins, accounting for about one-third of the total molecule (Richardson, 1981).

The idea that turns have functions related to their structural characteristics appears throughout the literature. Turns are intrinsically polar structures with backbone groups that pack together closely and side chains that project outward (Fig. 1). Such an array of atoms may constitute a site for molecular recognition, and indeed, the literature abounds with suggestions that turns serve as loci for receptor binding, antibody recognition, and posttranslational modification. This familiar idea—that function follows structure—must be evaluated with particular care in the case of turns. In proteins, intermolecular recognition takes place upon surfaces that are already densely populated by turns. Hence, the appearance of a turn at or near a recognition site does not necessarily imply a relationship of cause and effect. In peptides, turns are the conformations of choice for simultaneously optimizing both backbone-chain compactness (intramolecular nonbonded contacts) and side-chain clustering (to facilitate intermolecular recognition). Presence of turns in bioactive conformations may in fact reflect the lack of alternative conformational possibilities. These topics will be further elaborated in the sections that follow.

B. Terminology

In this review, a *turn* is defined as a site where the polypeptide chain reverses its overall direction. The terms β and γ turn have more restricted definitions and describe turns of four or three residues, respectively. These turns may or may not be stabilized by an intraturn hydrogen bond; in β turns, the C=O of residue i may be hydrogen bonded to the NH of residue $i + 3$, while in γ turns, the C=O of residue i may be hydrogen bonded to the NH of residue $i + 2$ (Fig. 2).

Venkatachalam (1968) used hard-sphere calculations to determine possible types of hydrogen-bonded β turns. His set has since been augmented (Chandrasekaran *et al.*, 1973; Lewis *et al.*, 1973). The possible ϕ, ψ angles for hydrogen-bonded β turns were determined by Némethy and Printz (1972). Table I lists these turn classifications for both β and γ turns. Note that type I and III β turns are very similar; they occupy contiguous regions of ϕ, ψ space and are not distinct types (Némethy and Scheraga, 1980). However, it is convenient to distinguish type III as a

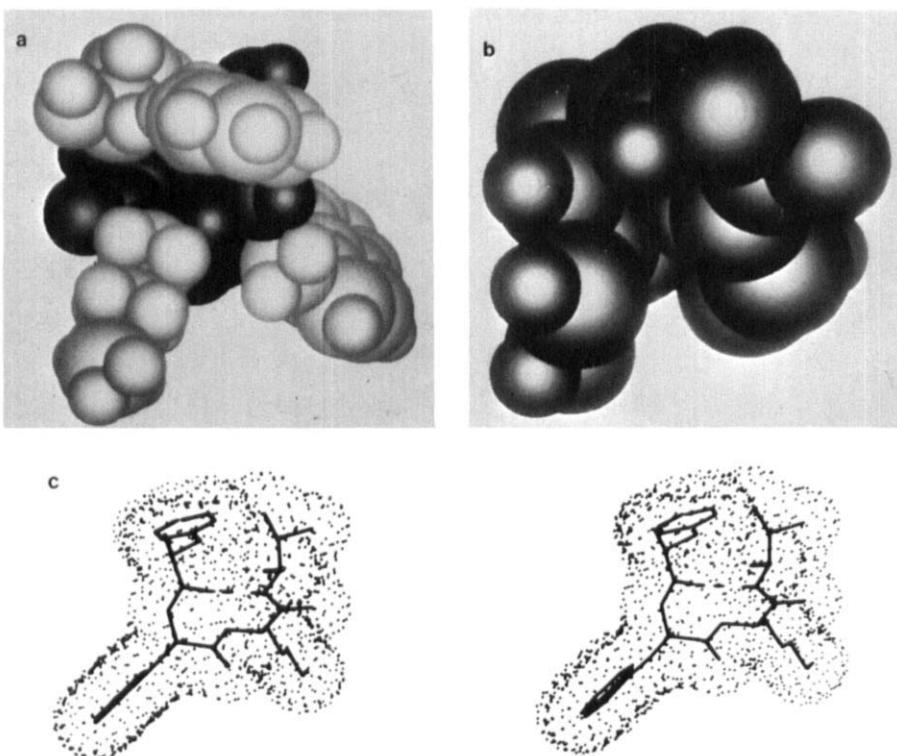


FIG. 1. Illustration of a turn showing the compact folding of backbone elements and the outward projection of side chains. The turn shown is a type I β turn, with the tetrapeptide sequence Phe-Trp-Lys-Thr, as occurs in the hormone somatostatin. (a), (b) Computer-generated space-filling models with atomic van der Waals radii from Sober (1970); backbone elements are shaded dark gray, and side chains light gray. The same turn is shown in the two views, but the side chains are omitted in (b) so that the backbone can be seen more easily. (c) A stereoview of the same turn as a stick diagram with no hydrogens. The stippled surface is the water-accessible surface calculated according to the method of Lee and Richards (1971), as modified by Connolly (1983). Note that the turn leads to a highly accessible disposition of the side chains of the "corner" residues (i.e., those in positions $i + 1$ and $i + 2$).

separate category since it describes a helical turn (actually one loop of a 3_{10} helix) and may form part of a larger repeating structure. Detailed descriptions of β turns appeared in a recent review in this series (Richardson, 1981). The terms "open" β or γ turns will be used for situations in which no hydrogen bond exists, and the ϕ, ψ angles are within 30° of the ones cited.

Turns, defined more generally as sites of overall chain reversal, do not

lend themselves to classification by dihedral angles. Nevertheless, inspection of protein-folding patterns clearly reveals their loci (Fig. 2c). In Section III,C,2, the problem of recognizing and defining turns from X-ray coordinates is discussed.

II. TURNS IN PEPTIDES

In the first part of this review, the structural and functional roles of turns in peptides will be considered. Although turns in peptides and in

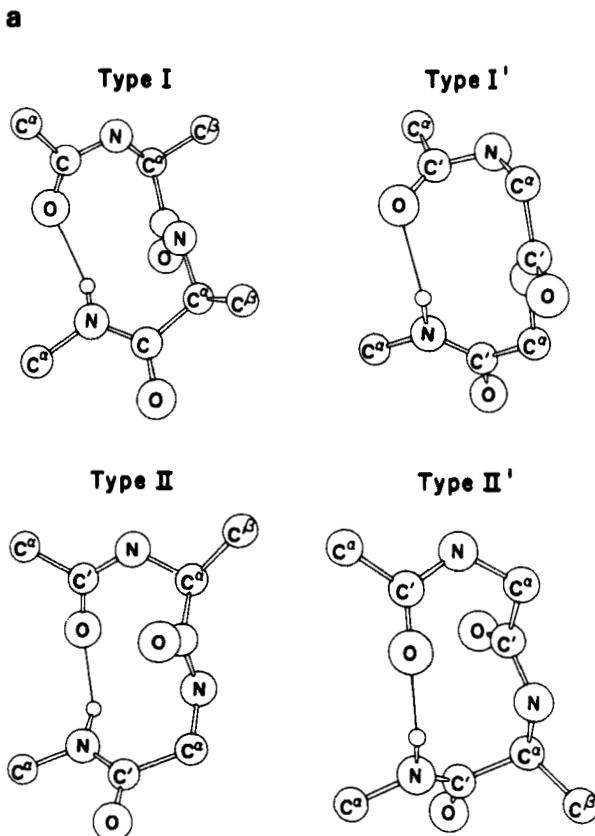


FIG. 2. Diagrams of various turns. (a), (b) Hydrogen-bonded turns, with dihedral angles as given in Table I. Side chains are shown in the *L* configuration, in those positions preferred by *L* residues. Positions favored for *D* residues are filled by glycine. Types III and III' turns are not illustrated, since they are only slightly different from types I and I' turns. (c) A stereoview of an open turn from ribonuclease A; this turn was located using the methods described in Section II,C,2.

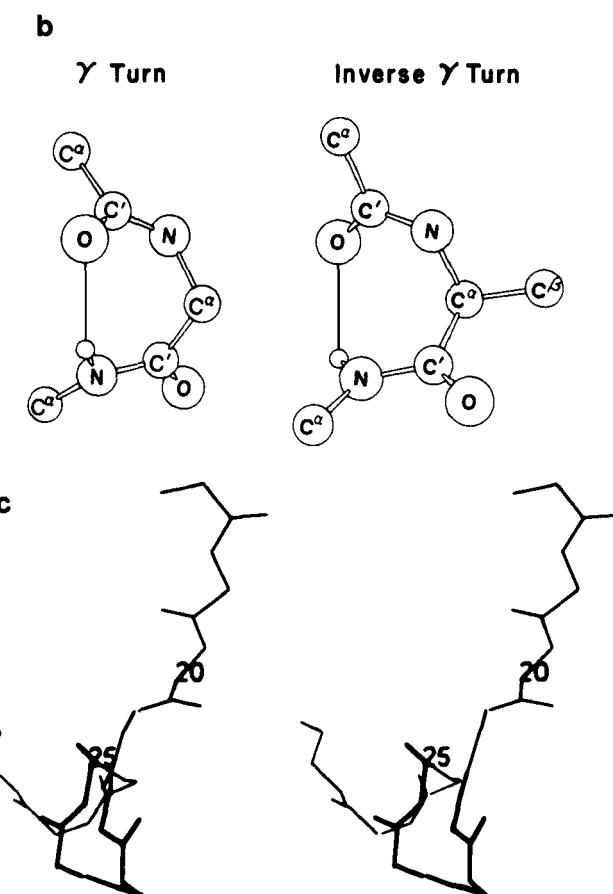


FIG. 2. (Continued)

proteins are homologous structures, the difference in molecular sizes leads to distinct geometric and energetic characteristics that may have a bearing on the roles of turns in these two classes of molecule. The following characteristics of peptides are of particular concern: (1) Nearly all atoms are exposed to solvent to some degree; hence, there is no sharp distinction between surface and interior groups. (2) Peptides adopt conformations that are necessarily determined by short-range interactions. (3) They are often conformationally mobile. (4) Their preferred conformations are strongly influenced by interactions with other molecules and by the nature of their environment. (5) They are small enough that their conformations can be analyzed in detail in solution as well as in the crystalline state.

TABLE I
Hydrogen-Bonded β and γ Turns: Dihedral Angles

Turn	$i + 1$		$i + 2$	
	ϕ	ψ	ϕ	ψ
β turns ^a				
Type I	-60	-30	-90	0
Type I'	60	30	90	0
Type II	-60	120	80	0
Type II'	60	-120	-80	0
Type III	-60	-30	-60	-30
Type III'	60	30	60	30
Type VIa (<i>cis</i>) ^b	-60	120	-90	0
Type VIb (<i>cis</i>) ^b	-120	120	-60	0
γ turns ^c				
Turn	70 to 85	-60 to -70		
Inverse turn	-70 to -85	60 to 70		

^a As originally defined by Venkatachalam (1968), except as noted.

^b Angles are taken from data presented in Richardson (1981).

^c As originally defined by Némethy and Printz (1972).

Two aspects of turns in peptides will be discussed. First, peptides serve as models for possible turns in proteins. Many of the sequence preferences in turns and the spectral parameters of turns have been established from model peptide studies. These topics are discussed in Section II,A. Second, turns are proposed to occur and to play functional roles in many naturally occurring peptides. Well-established examples of bioactive peptides with turn conformations are discussed in Section II,B.

A. Model Peptides

1. Methods for the Determination of Turns in Model Peptides

a. Overview. In this section a brief overview will be presented of methods available for determining the presence of turns in peptides. A previous review described methods used to determine turns in peptides and covered the literature through 1978 (Smith and Pease, 1980). The present review focuses on more recent developments. More detailed discussions of the various approaches and applications to turn-containing peptides can be found in more specialized reviews: for example, nuclear magnetic resonance (NMR) (Deslauriers and Smith, 1980;

Kessler, 1982); X-ray diffraction (Karle, 1981a); vibrational spectroscopy (Krimm, 1983); circular dichroism (CD) (Woody, 1974); and energy calculations (Némethy and Scheraga, 1977; Scheraga, 1981; Hagler, 1985). An example of a conformational study of a turn-containing model peptide will then be discussed to illustrate the application of some of the newer NMR methods.

Many methods exist that can yield information about the conformations of peptides. The most definitive is X-ray diffraction, and any peptide crystal of conformational interest ought to be studied using X-ray methods. Much of our current knowledge about details of turn conformations is based on observations of peptide crystal structures. This will be amply demonstrated in the next section where results from turn-containing model peptides are discussed.

However, the drawback of X-ray studies is that the molecule is viewed in conformation(s) determined by the interplay of intramolecular and crystal-lattice forces. Intermolecular forces are of primary importance in the formation of the crystal and can have profound effects on the conformation(s) adopted by the peptide (in contrast to the situation in protein crystals where the conformation of the protein is thought to be only slightly perturbed). Still, the observed conformation(s) can always be counted among those accessible to the peptide and in certain instances will be closely related to conformations in other environments of interest. For example, some peptide crystals have few strong peptide-peptide interactions in the crystal. Valinomycin exemplifies this situation; only weak nonpolar contacts exist between molecules, and the observed conformation can be deduced to arise predominantly from intramolecular forces (Smith *et al.*, 1975). In other cases, the peptides may interact with each other in the crystal via groups that are oriented appropriately in the low-energy solution conformation. Examples are the cyclic pentapeptides containing both a β and a γ turn; in crystals these molecules form linear arrays stabilized by hydrogen bonding between adjacent molecules using the C=O and NH of the peptide bond between residues $i + 1$ and $i + 2$ of the β turn (Karle, 1978, 1981b) (Fig. 3). These groups are unable to form intramolecular hydrogen bonds (in the absence of side-chain interactions) in a β turn. The solution and crystal conformations of these cyclic peptides are essentially the same (Karle, 1978; Pease and Watson, 1978; Pease, 1979; Karle, 1979, 1981b). The other type of crystal that offers an environment similar to solution is the solvent-rich crystal. An example is provided by [Phe⁴,Val⁶]antamanide, which has a water-filled channel between peptide molecules (Karle, 1977) (see Fig. 4).

One can speculate that the environment in some crystals may not be so

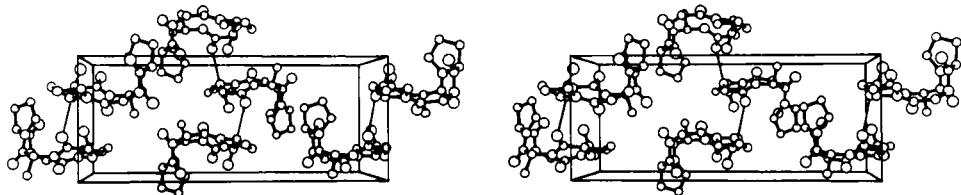


FIG. 3. Stereodrawing of the crystal packing in the cyclic pentapeptide cyclo-(Gly-Pro-D-Ala-Pro). Note the intermolecular hydrogen bonds (shown as thin lines) between the amide groups of the $i + 1$ to $i + 2$ peptide bond, which form infinite chains throughout the crystal lattice. Reprinted with permission from Karle (1978). Copyright 1978 by the American Chemical Society.

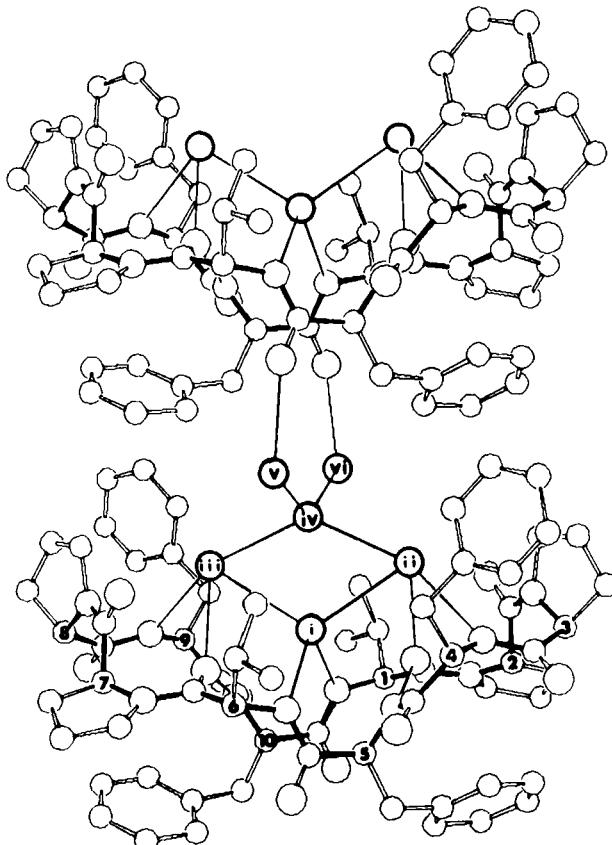


FIG. 4. Cross section of the crystal lattice of [Phe⁴,Val⁶]antamamide showing the water-filled channel (waters labeled with Roman numerals). Reprinted with permission from Karle (1977). Copyright 1977 by John Wiley & Sons., Inc.

very different from a receptor-binding site. Several examples of peptide crystals with amphiphilic packing environments have been reported (Karle, 1977; Karle *et al.*, 1983). One face of the peptide is involved in hydrophobic, side chain-to-side chain interactions, while the other interacts via hydrogen bonding. In Section II,B,2, examples of some naturally occurring peptides with amphiphilic bioactive conformations will be discussed. It has been suggested that amphiphilicity is a general characteristic of peptide hormones (Kaiser and Kézdy, 1984), and by analogy their receptors would be characterized by amphiphilic binding sites.

Solution conformational analysis of peptides is hindered by the interrelated obstacles of conformational heterogeneity and dynamics. Small linear peptides are usually flexible molecules and undergo conformational interconversions. The proportion of time the peptide spends in each of its conformational states and the rate of interconversions must be considered in choosing methods of conformational analysis and in interpreting data. Conformational interconversions not requiring peptide bond rotations may occur at millisecond or faster time scales in linear peptides; peptide bond rotations occur on the time scale of seconds (Jackman and Sternhell, 1969). If a spectroscopic method has a fast time scale relative to these rates, individual conformers will be "seen" (i.e., the observed spectrum will be a sum of the individual spectra). If the spectroscopic method has a slow time scale relative to these rates, observed spectral parameters will be weighted averages of interconverting conformers; the weighting functions can be complex (Jardetzky, 1980). Cyclic peptides are less flexible than their linear counterparts and are generally more tractable targets of solution conformational analysis.

Nuclear magnetic resonance is the most useful method of peptide conformational analysis in solution. NMR yields information about chemical environments of nuclei (chemical shifts), geometric relationships between nuclei (coupling constants), distances between nuclei [nuclear Overhauser enhancements (NOEs)], accessibility and hydrogen bonding of amide protons [exchange kinetics and sensitivity of resonance positions and linewidths to temperature, solvent, or paramagnetic probes (Kopple, 1983; Kopple and Zhu, 1983)], and dynamics of nuclei (relaxation times). This wealth of information is site specific, so that properties of individual residues can be studied. However, the interpretation of NMR data for peptides must be carried out with caution, since NMR is a slow method with a time scale of seconds to hundredths of seconds for the usual NMR measurements. Hence, conformational interconversions not requiring peptide bond rotations will lead to averaged NMR parameters. As Jardetzky (1980) has pointed out, the interpretation of these averaged NMR parameters in terms of one conformation is of little value.

Newly developed solid-state NMR methods have been applied to peptides in crystals (Pease *et al.*, 1981; Giersch *et al.*, 1981b, 1982a) and in other solid-like environments (notably, lipid bilayers) (Frey *et al.*, 1983). Conformational parameters used in solution (e.g., isotropic ^{13}C chemical shifts of prolyl residues) can now be determined for crystalline molecules. This methodology offers a dual advantage: The observed parameter (e.g., a ^{13}C chemical shift) can be correlated with a well-determined conformation in cases where the X-ray structure of the peptide crystal is available, and in instances where no structure has been determined by X ray, NMR data can yield a description of the conformation in the solid state.

Circular dichroism (CD) is exquisitely sensitive to conformation, since the arrangement of peptide bond chromophores with respect to each other and to asymmetric fields in the molecule will vary with different conformations. CD has a very fast time scale (10^{-15} second) (Cantor and Schimmel, 1980), but CD spectra in the peptide chromophore region are composed of broad overlapping bands arising from the ensemble of $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions of the molecule. Consequently, detailed interpretation of CD spectra for molecules larger than di- or tripeptides is difficult. Furthermore, several different conformational features yield similar CD spectra. Expected CD spectra have been calculated for β (Woody, 1974) and γ (Madison and Schellman, 1970) turns. Model peptide data (Bush *et al.*, 1978; Brahmachari *et al.*, 1979; Giersch *et al.*, 1981a; Madison and Kopple, 1980; Bandekar *et al.*, 1982) have given support to certain of these calculated spectral types and have raised doubts about others. At present it appears that γ turns give rise to a reliable CD band at long wavelength (~ 230 nm), and β turns yield any of three curve shapes that correlate with turn type (Fig. 5). It must be kept in mind that these turn spectra have relatively low ellipticities compared to α helix or β structure, so that decomposing a complex spectrum into various structural features including turns involves large errors.

Vibrational spectroscopy [infrared (IR) and Raman] has been used as a complementary approach to NMR and crystallography in determining turn conformations. The sensitivity of amide vibrations to hydrogen bonding has been of particular interest (Boussard *et al.*, 1979; Kopple *et al.*, 1975; Pease and Watson, 1978; Hseu and Chang, 1980; Aubry *et al.*, 1979; Rao *et al.*, 1980; Aubry and Marraud, 1983). Like CD, vibrational spectroscopy has a fast time scale (10^{-13} second), and one sees IR bands for virtually all species present. Normal coordinate analysis offers a way of analyzing all of the intrinsic vibrations for a particular conformation of a peptide. Bandekar and Krimm (1979, 1980) have focused on β turns and have reported expected bands for various turns. They com-

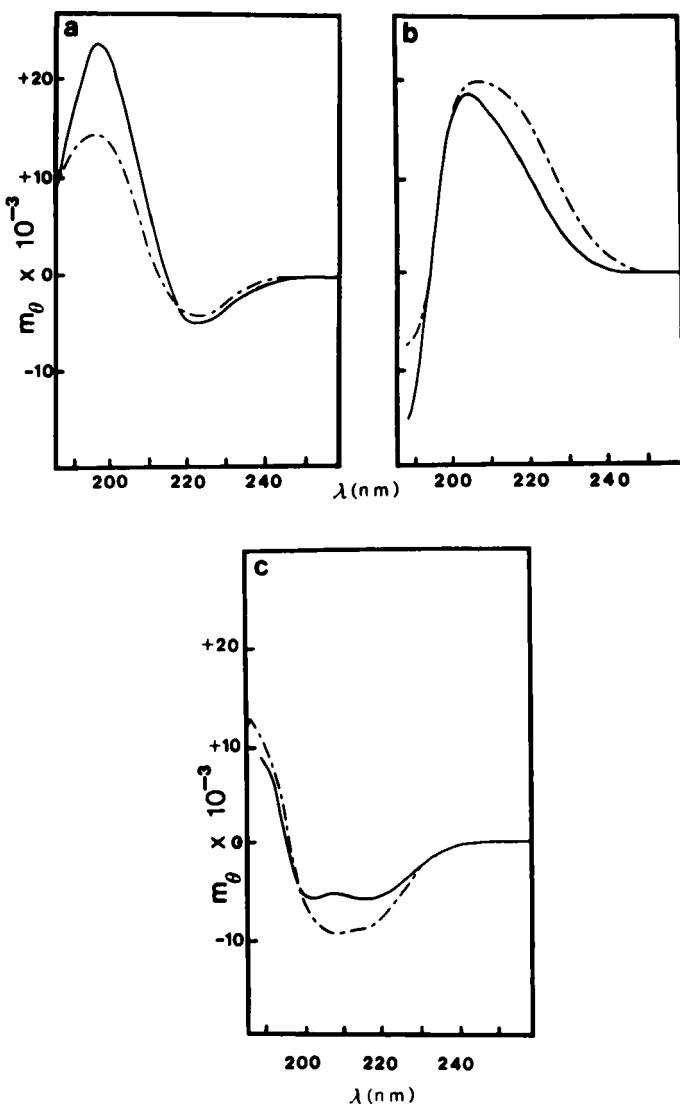


FIG. 5. Examples of CD curves of β turns. Here the model peptides are cyclic hexapeptides, which adopt conformations formed from two β turns. The solid lines are spectra in trifluoroethanol, and the dashed lines in water. (a) cyclo-(Ala-Pro-Gly)₂, type II β turns; (b) cyclo-(Ala-Pro-D-Ala)₂, type II β turns; (c) cyclo-(Gly-Pro-Ala)₂, type I β turns. Note that two different curve shapes are observed for type II turns; also, type II' turns yield spectra very similar to those for type I turns.

pared their predicted spectra with experimental data for cyclized β -turn models (i.e., a β turn cyclized via a covalent, nonpeptide linkage, see Section II,A,2,a) and found a consistent fit (Bandekar *et al.*, 1982). Also, J. W. Fox *et al.* (1981), Williams (1983), and Williams and Dunker (1981) have discussed the application of Raman methods to the determination of turn conformations. These approaches have not been extensively applied as yet, and it remains to be seen how useful they will be.

Energy calculations provide a means of exploring and characterizing the conformational space available to a peptide. Their value depends strictly on the validity of the functions used to compute the conformational energies and the effectiveness of the techniques used to search for structures. Furthermore, the nature of the medium around a peptide (solvent, lipid, receptor-binding site) will exert major influences on its preferred conformation(s). Until recently, conformational energy calculations were carried out *in vacuo* with highly simplified potential functions and rigid geometry (i.e., neglect of bond angle and length variations). Many peptides, including several naturally occurring and model peptides with putative turn-containing conformations, have been examined with these limitations. The results have helped to guide in the development of sequence-conformation correlations and in the interpretation of experimental results (for a discussion and references, see Smith and Pease, 1980).

In the past few years, progress in computer design has enabled refinements of methods of energy calculation. To date major advances that have occurred include the relaxation of rigid geometry assumptions through the use of valence force field or Urey-Bradley force fields, the treatment of molecules as fluctuating systems using molecular dynamics or Monte Carlo techniques, and the inclusion of solvent in calculations (for a review, see Hagler, 1985).

These approaches have been applied to turn-containing peptides. For example, lysine vasopressin, which, like many peptide hormones, is thought to be a flexible molecule in solution (Glickson, 1975), was studied using molecular dynamics methods (Hagler *et al.*, 1985). This nonapeptide with a six-residue cyclic moiety closed by a disulfide bridge was subjected to a valence force field, and its motions were followed in small steps over a period of several picoseconds. Figure 6 illustrates several "snapshots" along the motional trajectory of lysine vasopressin. The dynamic nature of the molecule is simulated and several hydrogen-bonded turns can be seen among the structures. This type of calculation yields a very different image of molecular conformational behavior than do calculations of static energy-minimized structures. If the molecule under

study, like vasopressin, is inherently flexible, calculation of a single static structure cannot adequately represent its conformations. Instead of calculating a continuous time course of the molecular behavior, as is done in molecular dynamics, Monte Carlo methods generate large numbers of configurations ($\sim 10^6$ or more) in a stochastic fashion, and these provide a description of the average structure and its fluctuations. The interaction between water and the cyclic hexapeptide, cyclo-(Ala-Pro-D-Phe)₂, was studied using Monte Carlo methods and potentials refined for aqueous solvation (Hagler *et al.*, 1980). As shown in Fig. 7, the interaction of solvent and peptide can be effectively studied using these approaches. The initial conditions were chosen to be a well-defined crystal structure of the cyclic hexapeptide in which 16 waters per unit cell had been crystallographically located (Brown and Teller, 1976). The results show a high degree of success in simulating the X-ray data, both in positions and degree of order of water sites.

b. Example of an NMR Conformational Analysis of a Turn-Containing Model Peptide. The model cyclic hexapeptide cyclo-(Pro-D-Tyr(Bzl)-Gly-Ile-Leu-Gln) was designed to mimic the local conformation of lysozyme residues 54–57 (Giersch *et al.*, 1985). This segment of lysozyme consists of an interior β turn (see Section III,B,2). This unusual turn is highly hydrophobic and virtually inaccessible to solvent water. This interior turn, like many such turns (Rose *et al.*, 1983), interacts with a buried water that is well located crystallographically. The model peptide was designed so that the hydrophobic interior turn could be studied apart from its protein environment. The Pro-D-Tyr sequence, which strongly favors a type II β turn, was used to encourage the adoption of a β turn by the remaining four residues, Gly-Ile-Leu-Gln. (The two β -turn conformation is a likely one for cyclic hexapeptides; see Section II,A,2,a.) Figure 8 illustrates how the protein served as a “model” for the hexapeptide.

¹H-NMR analysis of a peptide, such as the interior-turn model, begins with the assignment of all resonances to specific protons in the sequence. Recently developed two-dimensional NMR methods (Aue *et al.*, 1976; Nagayama *et al.*, 1979, 1980; Kumar *et al.*, 1980) have facilitated this potentially knotty problem. Two-dimensional correlated spectroscopy (COSY) gives information about the *J*-coupled connectivities (through three or fewer bonds) in a molecule. Normally, the data resulting from a COSY experiment are presented as a contour map in which the usual one-dimensional spectrum appears on the diagonal and cross-peaks appear at intersections of chemical shifts of coupled resonances. For example, Fig. 9 shows the COSY spectrum of the interior-turn model peptide

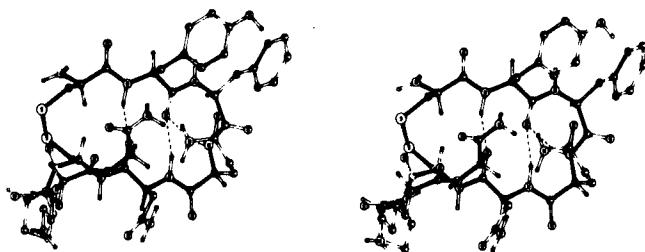
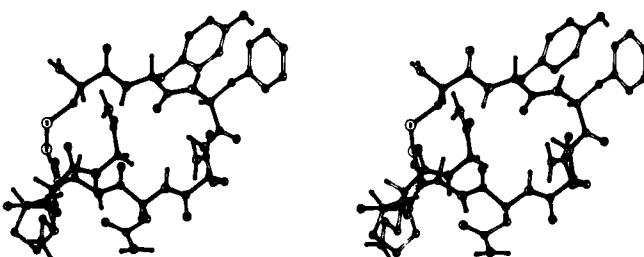
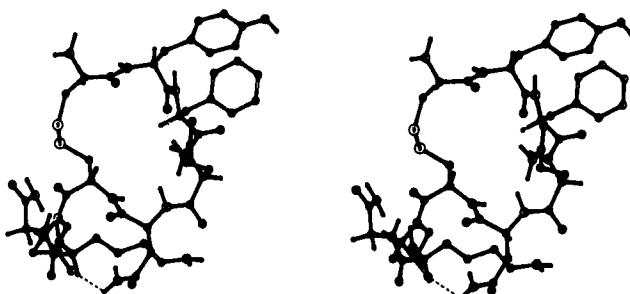
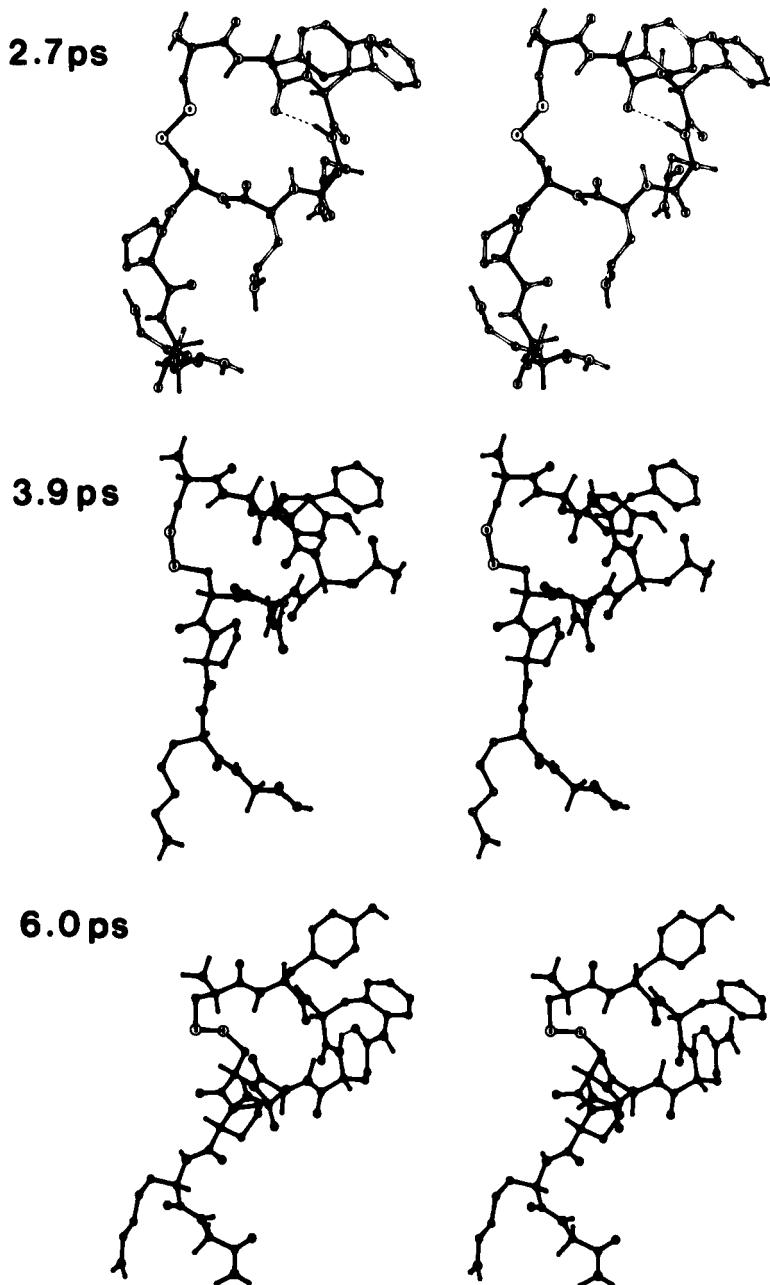
0.0 ps**0.2 ps****1.4 ps**

FIG. 6. Stereoviews of lysine vasopressin in the initial conformation and several subsequent conformations generated in a molecular dynamics simulation (time along the molecular dynamics trajectory given in picoseconds). Note the conformational changes demonstrated, including the making and breaking of intramolecular hydrogen bonds. Reprinted with permission from Hagler *et al.* (1985). Copyright 1985 by the American Association for the Advancement of Science.

FIG. 6. (*Continued*)

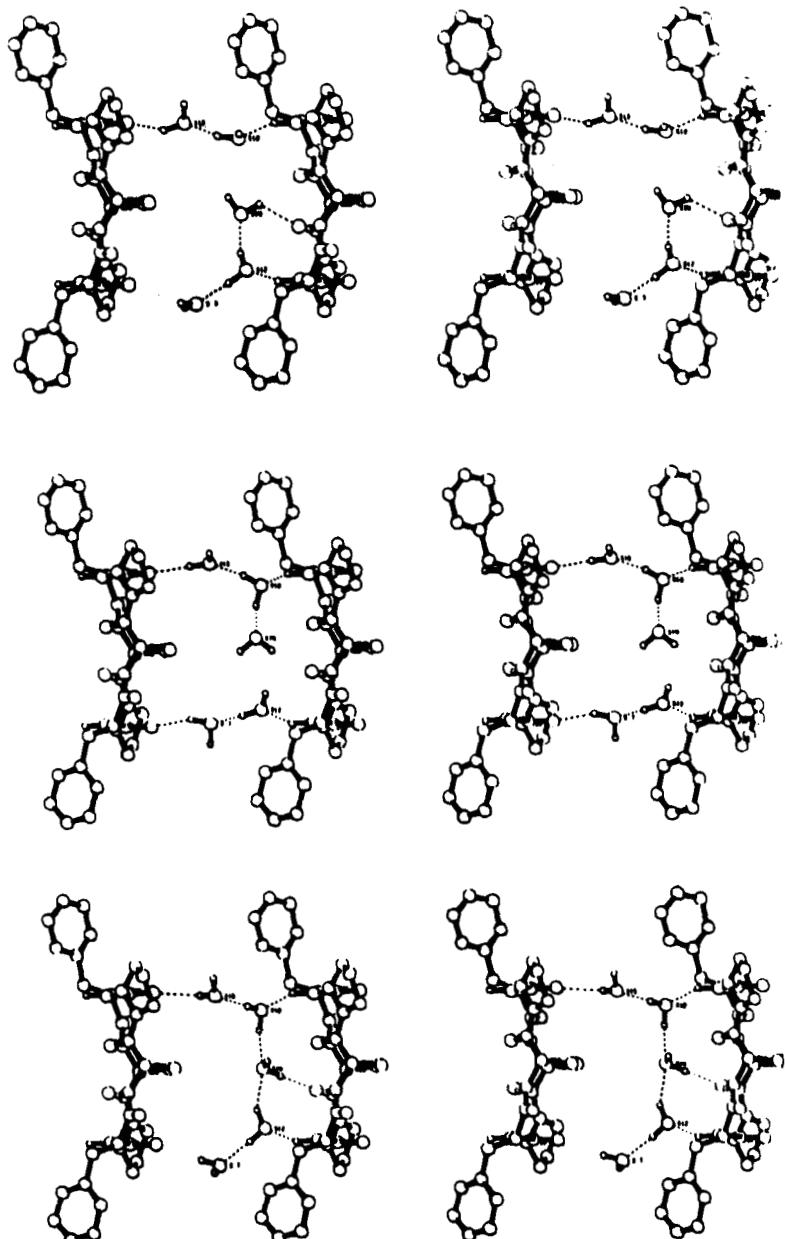


FIG. 7. Stereo "snapshots" of the arrangement of water molecules between molecules of cyclo-(Ala-Pro-D-Phe)₂ in the hydrated crystal. The configurations were generated using Monte Carlo methods and are not related by any temporal axis. Reprinted with permission from Hagler *et al.* (1980). Copyright 1980 by John Wiley & Sons, Inc.

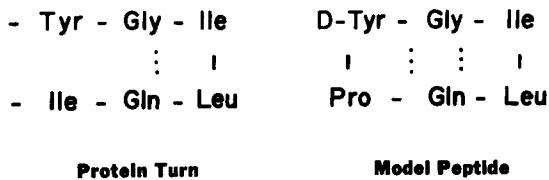


FIG. 8. Design of peptide model for an interior turn from lysozyme. The sequence around the turn in the protein is shown on the left. The cyclic hexapeptide incorporates that sequence in a backbone that is likely to encourage adoption of the same turn conformation in the peptide model (right) as in the protein. The critical feature of this cyclic peptide's structure is the Pro-D-Tyr dipeptide, which has a strong preference for a type II β turn.

in tetramethylene sulfone. The cross-peaks for tyrosine protons are mapped to illustrate the analysis. Using these data, the spectral assignments shown in Fig. 10 were obtained.

Figure 11 illustrates another two-dimensional NMR experiment, the NOE-correlated spectrum, where cross-peaks appear at the intersection of chemical shifts for two signals interacting via the NOE mechanism (dipolar coupling). In general, NOEs are observable in instances when two protons approach one another by less than 3 Å in the preferred conformation of the peptide (Noggle and Schirmer, 1971; Bothner-By, 1979). In this example, cross-peaks can clearly be seen between the Pro H $^{\alpha}$ and the Tyr NH resonance, between the Tyr NH and the Gly NH, and between the Leu NH and the Gln NH. In order to put the peptide solute into a motional regime where the dipolar interaction is efficient (Kopple, personal communication), the peptide was studied in tetramethylene sulfone, a solvent of high viscosity. Otherwise, NOE effects for peptides of this size are quite small (Bothner-By, 1979; Roques *et al.*, 1980).

Conformational analysis exploits the parameters described in Section II,A,1,a: chemical shifts, coupling constants, NH accessibility/hydrogen bonding, NOEs, and sometimes, relaxation times. Table II summarizes the critical data used here to develop a self-consistent picture of the hexapeptide's conformation. The observation of small temperature and solvent dependences for the Gly and Gln NH resonances indicate that these NH protons are inaccessible to solvent, consistent with the "design" conformation of the peptide (Fig. 8). Seeing one large (8 Hz) and one small (\sim 0 Hz) coupling constant for the Gly H $^{\alpha}$ -NH argues strongly for rigidity (viz., one strongly preferred conformation) since an ensemble of interconverting conformations would lead to averaged coupling constants in the range 5–7 Hz. Last, a particularly useful application of the observed NOEs, which is of general applicability to turns, can be illus-

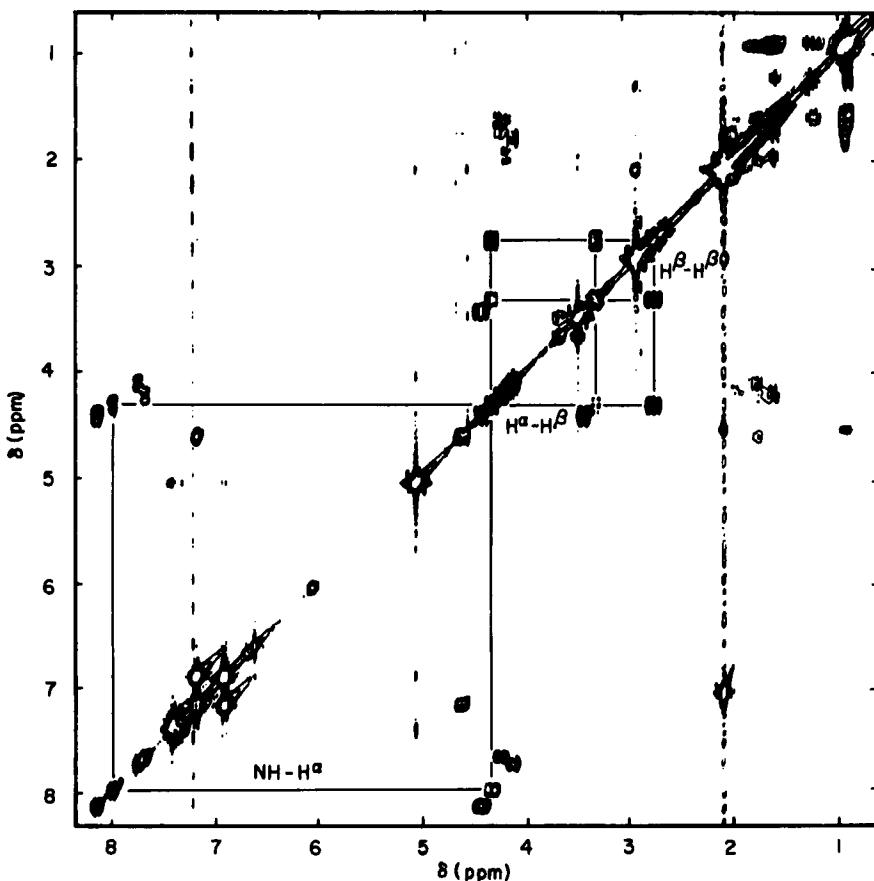


FIG. 9. Two-dimensional correlated proton NMR spectrum of the cyclic hexapeptide model in tetramethylene sulfone (250 MHz, peptide concentration 10 mg ml^{-1}). The spectrum is shown as a contour plot, with the usual one-dimensional spectrum appearing along the diagonal. Off-diagonal peaks occur at the intersections of chemical shifts for resonances that are J coupled to one another. Hence, J -coupled spin systems can be mapped out straightforwardly. In this case, the coupling of tyrosine resonances is labeled as an illustration.

trated here. In α helices, it has been noted that a consecutive sequence of $\text{NH}(i)$ to $\text{NH}(i + 1)$, $\text{NH}(i + 1)$ to $\text{NH}(i + 2)$, etc., NOEs occur (Wüthrich *et al.*, 1982). In the interior-turn hexapeptide, a pair of single $\text{NH}(i)$ to $\text{NH}(i + 1)$ NOE connectivities are seen, one at each end of the molecule. These, together with data indicating sequestering of the $i + 1$ NH , signal the presence of turns (Fig. 12). Note that it is the NH of residue $i + 3$ (Gly or Gln) in each turn that interacts with the NH of

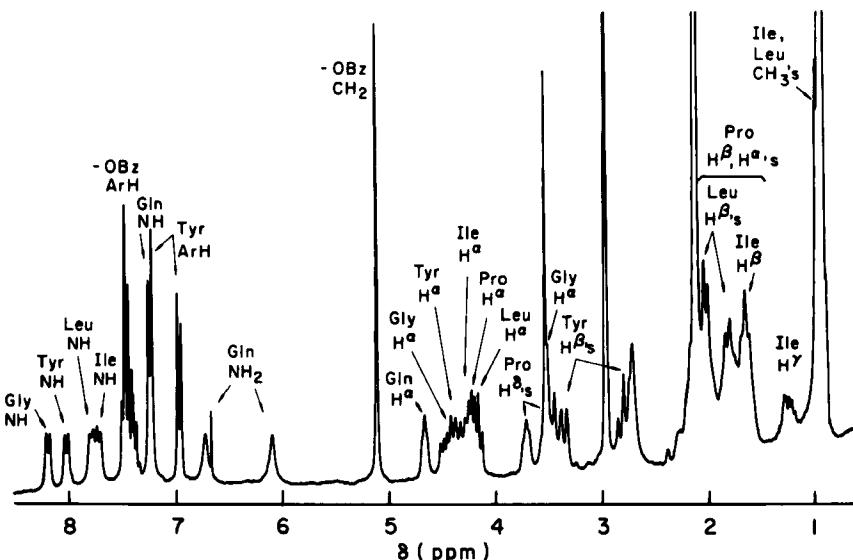


FIG. 10. One-dimensional proton NMR spectrum of the cyclic hexapeptide model in tetramethylene sulfone, with assignments as determined by experiments such as that shown in Fig. 9 (250 MHz, peptide concentration 10 mg ml⁻¹).

residue $i + 2$ (Tyr or Leu). Furthermore, if the $i + 1$ residue H^α interacts with the $i + 2$ residue NH as well, as is the case with Pro H^α and Tyr NH, a cogent case may be made for a type II β turn (see Fig. 12). This correlation has been pointed out previously (Khaled and Urry, 1976; Leach *et al.*, 1977), most recently in a cyclized β -turn model where various distances were calculated from predicted conformations and compared to observed NOEs (Narasinga Rao *et al.*, 1983). The distances for typical turns are shown in Fig. 13. Note that in the present peptide, the other end of the molecule also exhibits the NH(i)–NH($i + 1$) NOE, but not the NH to H^α connectivity. The presence of the NH(Leu)–NH(Gln) NOE along with the absence of an NH(Leu)– H^α (Ile) NOE leads to the conclusion that there is a β turn of type I comprising the sequence Gly–Ile–Leu–Gln, analogous to that occurring in the protein interior turn. While both the Gly and Gln NHs are sequestered, the NH resonance for Gln does not occur downfield in a region associated with strong hydrogen bonding (Wagner *et al.*, 1983). This observation is consistent with the finding that the hydrogen-bonding interaction in type I turns is less favorable than that in type II turns (see Section II,A,2,c and Giersch *et al.*, 1981a).

This discussion is intended to demonstrate salient aspects of an NMR

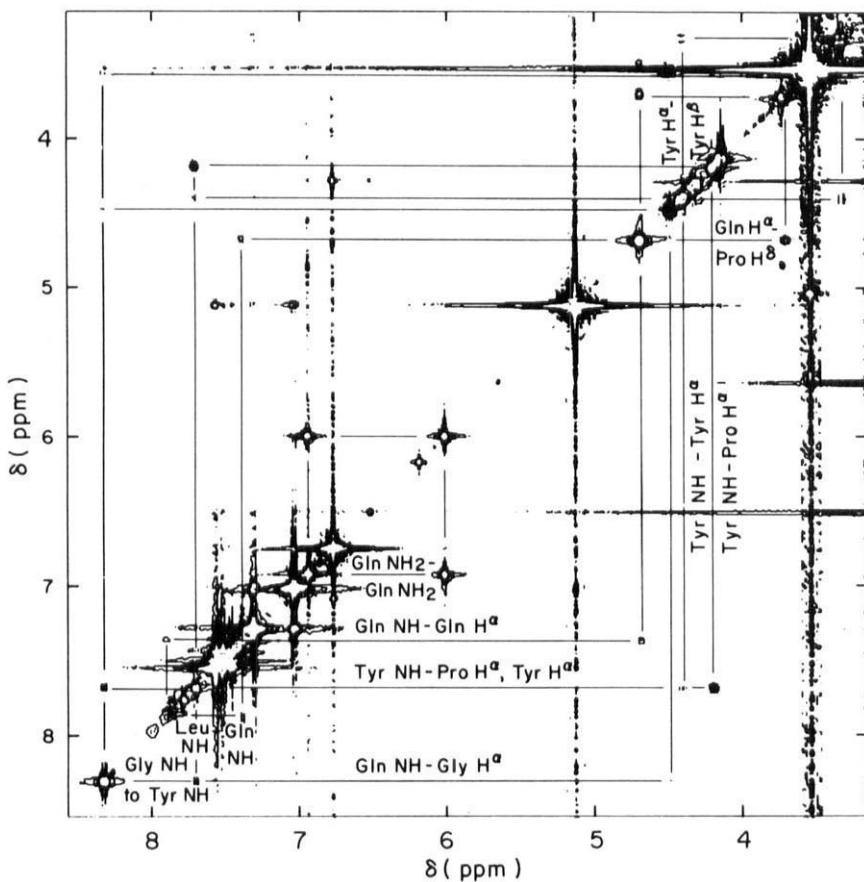


FIG. 11. Two-dimensional NOE spectrum of the cyclic hexapeptide model in tetramethylene sulfone (500 MHz, peptide concentration 10 mg ml^{-1}). Here, as in Fig. 9, the one-dimensional spectrum appears along the diagonal, but off-diagonal peaks occur at the intersection of chemical shifts of resonances for protons which are related by an NOE effect. Spatial relationships can therefore be mapped out by inspection of these data. Several interactions are labeled in the figure to illustrate the analysis. These interactions are discussed in the text.

conformational analysis of a turn-containing peptide in solution, using newly developed methods. It is clear that the use of NOEs can add considerably to the NMR analysis. The case described involves a rigid peptide, though, which minimizes complications from conformational averaging.

TABLE II
Evidence for Proposed Conformation of Interior Turn-Model Cyclic Hexapeptide: cyclo-(Pro-d-Tyr(BzL)-Gly-Ile-Leu-Gln) from ¹H-NMR Analysis

Accessibility of NHs	D-Tyr	Gly	Ile	Leu	Gln
$\Delta\delta/\Delta T^a$ (ppb/deg)	5.9	1.8	2.4	3.0	~0
δ^a (ppm)	8.56	8.27	7.91	7.54	7.26
$\Delta\delta/\Delta\text{solv}^b$ (ppm)	0.48	~0	0.75	0.15	~0
$J_{H\alpha^a-\text{NH}}(\text{Hz})$	8.5	8.0,~0	8.6	9.1	6.9
NOE's observed^c					
Gly NH: Gly H ^a (two), <u>Tyr NH</u>					
Tyr NH: Tyr H ^a , <u>Pro H^a</u>					
Gln NH: Gln H ^a , <u>Leu NH</u>					
Gln H ^a : Pro H ^b atoms					
Tyr H ^a : Tyr H ^b 's (two)					
Leu NH: Gln NH					

^a In CDCl₃ : (CD₃)₂SO, 98:2 (v/v), 8 mg ml⁻¹, 250 MHz.

^b Difference in chemical shift from CDCl₃ to (CD₃)₂SO.

^c In tetramethylene sulfone, 500 MHz, 23°C, 8 mg ml⁻¹. The underlined interaction gave rise to a larger peak in two-dimensional NOE experiments; interactions were not measured quantitatively. Mixing time was 150 msec.

2. What Has Been Learned about Turns from Studies of Model Peptides

a. Correlation of Sequence and Preferred Turn Type. Theoretical calculations of preferred β -turn conformations led to a fundamental distinction between turns predicted for *homochiral* and *heterochiral* sequences at positions $i + 1$ and $i + 2$ (Venkatachalam, 1968; Némethy and Printz, 1972; Chandrasekaran *et al.*, 1973; Maigret and Pullman, 1974). The former were predicted to favor type I (L-L) or I' (D-D) β turns and the latter type II (L-D) or II' (D-L) β turns. Glycine, lacking a side chain, could be accommodated in any position. The primary origin of these turn-sequence preferences is steric, arising from unfavorable nonbonded interactions between the side chains of residues in the $i + 1$ and $i + 2$ positions and either the C=O of residue $i + 1$ or the NH of residue $i + 2$. The favored turns place the side chain of the $i + 2$ residue on the opposite side of the turn from the $i + 1$ C=O (see Figs. 2 and 19).

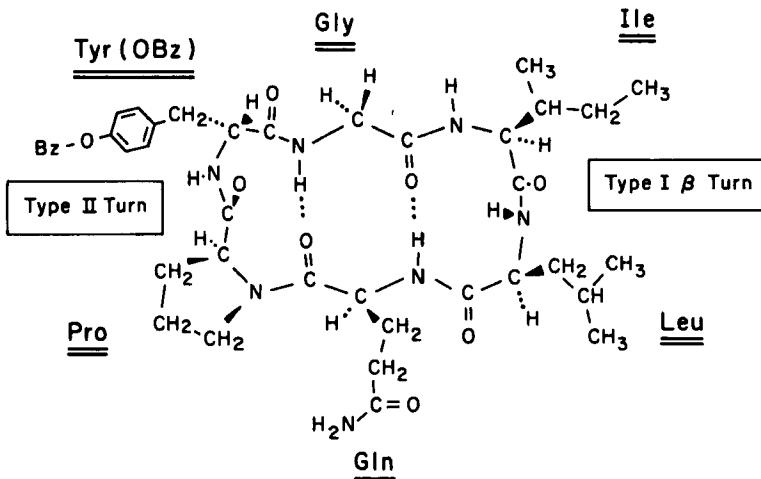


FIG. 12. Diagram of the conformation of cyclo-(Pro-d-Tyr(BzL)-Gly-Ile-Leu-Gln) showing close approaches of protons that yield NOE effects. Note particularly: Pro H^a and Tyr NH; Tyr NH and Gly NH; Leu NH and Gln NH.

Model peptide studies have provided considerable evidence in support of these predictions. A brief sampling of representative studies of linear and cyclic model peptides and cyclized turn models, where the peptide turn is constrained by a nonpeptide bridging unit, is presented in this section to illustrate the wealth of experimental evidence currently available. Fewer studies have been done on model peptides containing γ turns; examples of these will be described as well.

As pointed out previously (Section II,A,1,a), linear peptides are flexible molecules and as such are difficult targets of conformational analysis. However, their intrinsic flexibility makes them sensitive indicators of relative stabilities of various conformational states. The observation of one conformation for a linear peptide in solution or in crystals provides evidence for the stability of that conformation relative to other conformational states potentially available. Several linear peptides proposed to adopt β -turn conformations are listed in Table III, along with the environment (solvent, crystal) and methods used to analyze the conformation. This is a representative, not an exhaustive, compilation and emphasizes examples reported since 1980. Additional examples may be found in previous reviews (Smith and Pease, 1980; Toniolo, 1980).

Several observations can be made based on these data and those cited in previous reviews: (1) The majority of well-established turn conformations involve heterochiral sequences, and thus type II (or II') β turns,

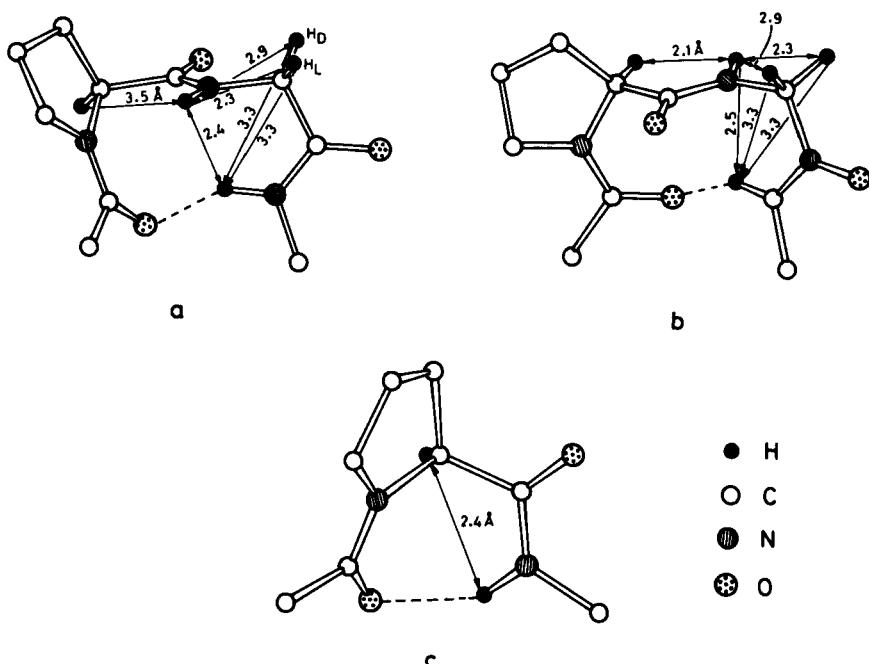


FIG. 18. Perspective drawings of hydrogen-bonded turn conformations. (a) Pro-Gly in a type I β turn; (b) Pro-Gly in a type II β turn; and (c) γ turn with Pro as residue $i+1$. Interproton distances that are short enough to yield significant NOE interactions are indicated. Note the close approach of the Pro H^a and Gly NH in the type II turn (b). Reprinted with permission from Narasinga Rao *et al.* (1983). Copyright 1983 by the American Chemical Society.

although some examples of type I turns are given. [In some apparently exceptional cases, for example, the type II turn in N-*i*-Bu-Pro-Ala-NH(*i*-Pr) (Aubry *et al.*, 1977), favorable interactions in the crystal seem to be important in determining the turn preference (M. Marraud, personal communication).] (2) Pro residues occur frequently, most often in position $i+1$. (3) Gly residues nearly always occur in positions favored for D residues.

Cyclic peptides and cyclized turn models have also yielded useful data pertinent to turn-sequence preferences. These conformationally restricted models have limited available conformational space and are more straightforward subjects for conformational analysis than linear peptides. The constraints imposed by cyclization can have an effect on the possible conformations of the sequence of residues, and the observed conformation may primarily reflect these stereochemical require-

TABLE III
Linear Peptides Adopting β -Turn Conformations^a

Peptide	Environment	Turn type	<i>i</i>	<i>i</i> + 1	<i>i</i> + 2	<i>i</i> + 3	N ··· O	Method	Reference
N-Ac-Pro-Gly-Phe-OH	Xstal	II		Pro	Gly	Phe	2.867	X ray NMR,CD	Brahmchari <i>et al.</i> (1981)
	TFE,MeOH	II		Pro	Gly	Phe		NMR,CD	
N-Ac-Pro-Gly-X-OH X = Gly, Ala, Leu or Ile	TFE, (CD ₃) ₂ SO	II		Pro	Gly	X			Brahmchari <i>et al.</i> (1982)
	² H ₂ O, (CD ₃) ₂ SO	I	Gly	Pro	X	Gly		NMR	
Gly-Pro-Asn-Gly -Leu-	I		Gly	Pro	Leu	Gly		NMR	Toma <i>et al.</i> (1980)
	(CD ₃) ₂ SO	I or II	Arg	Ala	Gly	Glu			
Gly-Pro-D-Leu-Gly	(CD ₃) ₂ SO	II	Gly	Pro	D-Leu	Gly		NMR	Prange <i>et al.</i> (1983)
N-iBu-Pro-Ala-NH(i-pr)	Xstal	II		Pro	Ala		3.05	X ray	Aubry <i>et al.</i> (1977)
	CCl ₄	I (and others)		Pro	Ala			IR	
N-iBu-Pro-D-Ala-NH(i-pr)	Xstal	II		Pro	D-Ala		3.10	X ray	Boussard <i>et al.</i> (1974)
	CCl ₄	II		Pro	D-Ala			IR	
N-Piv-D-Ala-Pro-NH(i-pr) ^b	Xstal	II'		D-Ala	Pro		2.925	X ray	Aubry <i>et al.</i> (1979)
	CCl ₄	II'		D-Ala	Pro			IR	
N-Piv-Pro-Gly-NH(i-pr)	Xstal	II		Pro	Gly		2.971	X ray	Aubry <i>et al.</i> (1980b)
	Xstal	I		Pro	Ser		2.875	X ray	
N-Piv-Pro-Ser-NHCH ₃	CCl ₄	I		Pro	Ser			IR	Aubry and Marraud (1983)
	Xstal	II		Pro	D-Ser		2.841	X ray	
N-Piv-Pro-D-Ser-NHCH ₃	CCl ₄	II		Pro	D-Ser			IR	Tanaka <i>et al.</i> (1979)
	Xstal	I		Pro	Gly			X ray	
Pro-Pro-Gly	Xstal	I		Pro				X ray	Aubry <i>et al.</i> (1983)
N-Piv-Pro-His-NHCH ₃	Xstal	I		Pro	His			X ray	Ayato <i>et al.</i> (1981)
	Xstal	II	Val	Pro	Gly	Gly	3.05	X ray	
t-Boc-Val-Pro-Gly-Gly-OBzI	MeOH	(II) ^d	Gly	Pro	Asn	Gly		CD'	Sato <i>et al.</i> (1981)

^a Abbreviations: Ac, acetyl; Xstal, crystal; TFE, trifluoroethanol; MeOH, methanol; t-Boc, tert-butyloxycarbonyl; Et, ethanol; i-Bu, isobutyrylcarbonyl; i-pr, isopropyl; Piv, tert-butylcarbonyl; Dnp, 2,4-dinitrophenyl; p-NA, p-nitroanilide.

^b Note that the homochiral analogs of this peptide adopt open (nonturn) conformations in the crystal (Aubry *et al.*, 1980a,c). Also, N-Piv-Gly-Pro-NHCH₃ was found to occur in an open conformation in the crystal (Aubry *et al.*, 1981a).

^c In the same study, several other tetrapeptide sequences were examined, and their tendencies to take up turn conformations determined by a CD method (see footnote d). It was found that heterochiral sequences favored turns more than sequences with Gly and one other residue in the turn, which favored turns more than homochiral sequences. For details, refer to Sato *et al.* (1981).

^d No definitive evidence was obtained for turn type; observation of an NOE between the Pro H^a and the Asn NH suggested that the turn is of type II.

^e A novel procedure based on the interaction of two chromophores bonded to the N and C termini of the tetrapeptides was used to deduce the tendency of these peptides to form turns. This method takes advantage of a large CD created by the proximity of the two chromophores.

^f The tendency to take up a turn form was qualitatively determined for a group of peptides with the X position occupied by various residues; the order, from greatest turn-forming tendency to least, is X = D-Ala > Asn > Gly > Ala > Gln.

ments. With cyclic peptides, the question posed in conformational analysis is not whether a turn will occur, but which turn will occur, since these molecules are forced to turn corners. In general, the preferred conformations of cyclic peptides and cyclized turn models have paralleled the results from linear peptides. Examples from some particularly informative conformational classes will be briefly discussed.

Cyclic hexapeptides have long been known to favor conformations that can be considered as two fused β turns (Schwyzer *et al.*, 1958) (Fig. 14). Varying the sequence of cyclic hexapeptides can therefore shed light on residue preferences for positions within turns, since several "frame-shifted" β turn-containing conformations are possible. In the absence of prolyl residues, conformational averaging is usually observed in these model peptides in solution, although a few exceptions have been reported (Blaha and Budesinsky, 1973; Kopple *et al.*, 1978). Cyclo-(X-Pro-Y)₂ peptides have been extensively studied in solution (e.g., Torchia *et al.*, 1972a,b; Kopple *et al.*, 1973, 1974, 1981, 1983; Kopple and Sarkar, 1979; Giersch *et al.*, 1981a; Kopple, 1981a; Kopple and Parameswaran, 1983) and favor C_2 symmetric conformations (on the NMR time scale). Their observed conformations reveal a strong preference for the prolyl residue in position $i + 1$ of a β turn. The restricted region of ϕ,ψ space available to a prolyl residue coincides with the ϕ,ψ angles required in this position for formation of an $i + 3 - i$ hydrogen bond in either a type I or II turn. As expected from steric considerations, type II turns occur when Y is a D residue; Pro-Gly sequences also occurred in type II turns. Type I turns occur when X is a Gly and Y is an L residue, except that presence of L-Val following the Pro residues leads to type II' β turns with Pro in position $i + 2$ of the turn, apparently because the bulky β -branched residue cannot readily be accommodated in the $i + 2$

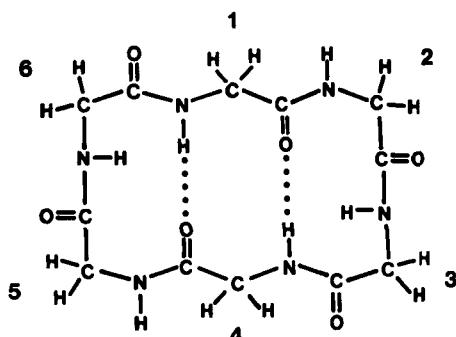


FIG. 14. Illustration of a generalized cyclic hexapeptide showing the conformation with two transannular hydrogen bonds that is essentially composed of two fused β turns.

position. Hence, the importance of positional preferences beyond $i + 1$ and $i + 2$ (viz., here Val prefers $i + 3$ to $i + 2$) is demonstrated. Effects of variation of residue in position $i + 3$ have been noted in the linear peptides *t*-Boc-Pro-Gly-X-OH (Table III) (Brahmachari *et al.*, 1982). In the cyclic hexapeptides a **D**-X-Pro-**L**-Y sequence also leads to type II' turns with Pro in position $i + 2$. Evidently, the presence of a **D** residue preceding the Pro serves as an inducement to form type II' turns. Although these conformational preferences are indicative of general sequence correlations within turns and are reinforced by their agreement with data from other peptide models, these findings may reflect factors specific to cyclic hexapeptides. For example, it has been pointed out that a strictly C_2 symmetric β -turn conformation in a cyclic hexapeptide requires an unfavorable carbonyl oxygen transannular approach (Madison, 1974; Snyder, 1984). Relaxation of C_2 symmetry may alleviate this O ··· O contact but may not be observable in NMR data. Several crystal structures have been determined for cyclic hexapeptides, and results from these are reviewed by Karle (1981a) and Giersch *et al.* (1981a). In most cases, the O ··· O contact has been minimized either by relaxation of C_2 symmetry or by opening up of the turns.

Cyclic pentapeptides are particularly suitable models for turns, since a strongly favored conformation consists of fused β and γ turns (Fig. 15).

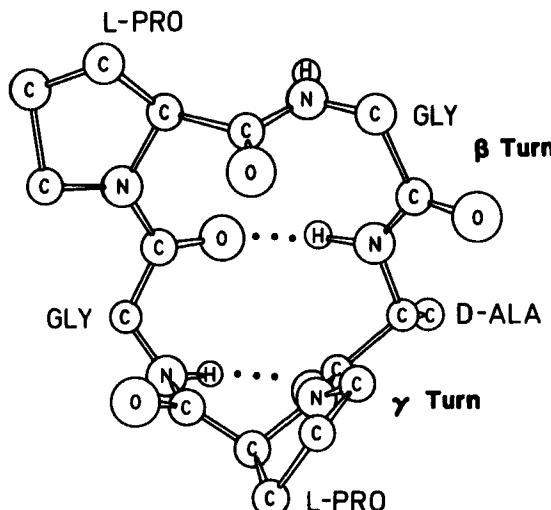


FIG. 15. Structure of the cyclic pentapeptide cyclo-(Gly-Pro-Gly-**D**-Ala-Pro) from X-ray diffraction (Karle, 1978). Its conformation consists of fused β and γ turns. Here the β turn comprises the sequence Gly-Pro-Gly-**D**-Ala, and is of type II. The γ turn is an inverse γ turn, with the sequence **D**-Ala-Pro-Gly.

In contrast to cyclic hexapeptides, they do not appear to suffer from unfavorable interactions in the two-turn conformation. Solution studies (e.g., Demel and Kessler, 1976; Pease and Watson, 1978; Bara *et al.*, 1978; Kondor and Kessler, 1979; Kessler and Kondor, 1979; Khaled *et al.*, 1976; Pease, 1979; Pease *et al.*, 1979; Kessler, 1982; Kessler *et al.*, 1982) and crystal-structure determinations (Karle, 1978, 1979, 1981b, personal communication; Einspahr *et al.*, 1980; Mauger *et al.*, 1982) provide additional support for the β -turn preferences discussed in this section. Table IV gives examples of solution and crystal structures of various cyclic pentapeptides that take up the β,γ -turn conformation; types II and II' β turns have been observed in crystals and types I, II, III, and II' in solution. An alternative solution conformation containing two γ turns has been proposed for several cyclic pentapeptides without prolyl residues (for a discussion, see Kessler, 1982). None of these peptides has been examined by X-ray crystallography.

Cyclized β -turn models also demonstrate the conformational preferences of various sequences for types I, II, or II' β turns. Two classes of these models have been described. In the first, a dipeptide is cyclized using ϵ -aminocaproic acid to yield a closed unit with three amide linkages and a $(\text{CH}_2)_5$ bridge across the dipeptide (Fig. 16). Extensive conformational analysis of these cyclized turn models has been carried out using rigid-geometry empirical-energy calculations, ^1H and ^{13}C NMR, CD, and vibrational spectroscopies (Deslauriers *et al.*, 1979; Némethy *et al.*, 1981; Deslauriers *et al.*, 1981; Maxfield *et al.*, 1981; Rae *et al.*, 1981;

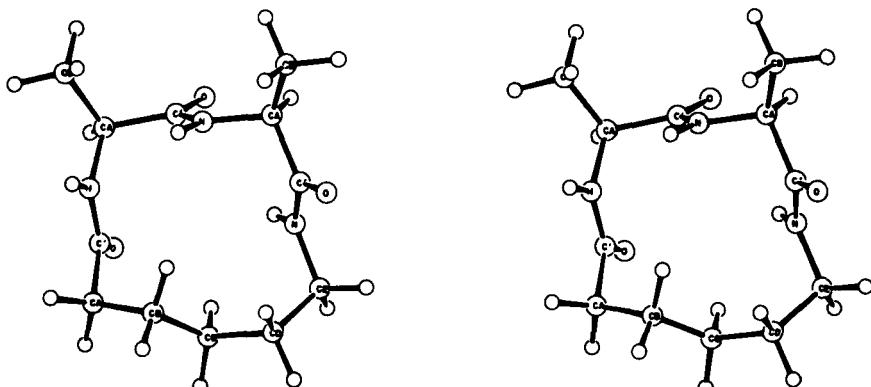


FIG. 16. Stereo drawing of the cyclized β turn cyclo-(l -Ala- l -Ala-Aca), where Aca is ϵ -aminocaproic acid. The conformation shown is the lowest energy conformer found for this peptide and corresponds to a type I β turn. Reprinted with permission from Bandekar *et al.* (1982). Copyright 1982 by Munksgaard International Publishers, LTD.

TABLE IV
Examples of Cyclic Pentapeptides That Adopt β,γ -Turn Conformations

Peptide	Environment	Type	β turn				γ turn ^a			Reference
			<i>i</i>	<i>i</i> + 1	<i>i</i> + 2	<i>i</i> + 3	<i>i</i>	<i>i</i> + 1	<i>i</i> + 2	
c-(Gly-Pro-Gly-d-Ala-Pro) ^b	Xstal, solution	II	Gly	Pro	Gly	d-Ala	d-Ala	Pro	Gly	Pease and Watson (1978); Karle (1978)
c-(d-Phe-Pro-Gly-d-Ala-Pro)	Xstal, solution	II	d-Phe	Pro	Gly	d-Ala	d-Ala	Pro	d-Phe	Pease (1979); Karle (1981b)
c-(Gly-Pro-Ser-d-Ala-Pro)	Xstal Solution	II'	Pro	Gly	Pro	Ser	d-Ala	None	Gly	Karle (1979)
c-(Gly-Pro-Ala-d-Phe-Pro)	Solution	I	Gly	Pro	Ser	d-Ala	d-Ala	Pro	Gly	Pease <i>et al.</i> (1979)
c-(Ala-Pro-Gly-Pro-d-Phe)	Solution	III	d-Phe	Ala	Pro	d-Phe	d-Phe	Pro	d-Phe	Pease and Giersch, unpublished
c-(Gly-Pro-Phe-Gly-Phe)	Solution	I	Gly	Pro	Phe	Gly	Gly	Phe	Gly	Demel and Kessler (1976); Kessler <i>et al.</i> (1982)
c-(Gly-Pro-d-Phe-Gly-Ala)	Solution	I/II'	Gly	Pro	d-Phe	Gly	Gly	Ala	Gly	Giersch <i>et al.</i> , unpublished
	Xstal	II	d-Phe	Gly	Ala	Gly	Gly	Pro	d-Phe	Karle, unpublished

^a All of the γ turns listed are inverse.

^b c, Cyclo.

' The NMR data for this peptide (both ¹H and ¹³C) suggest an equilibrium between two alternative β,γ -turn conformations. In particular, the proline spectral parameters that usually are diagnostic for a γ turn (upfield Pro C^B, downfield Pro H^a resonances) have values midway between nonturn and turn extremes.

Bandekar *et al.*, 1982). Dipeptide sequences studied include L-Ala-Gly, L-Ala-L-Ala, and L-Ala-D-Ala; preferred conformations were found to be as expected based on chirality: types II, I, and II β turns, respectively, although some indication of conformational averaging was reported in the Gly-containing peptide. Little definitive evidence of any intramolecular hydrogen bonding was found in any of the three peptides. One purpose in the design and synthesis of these molecules was to develop reference turn models for spectral parameters. Indeed, the CD and vibrational analyses (Section II,A,1) of the three peptides promise to be useful in studies of other turn-containing peptides.

The other class of cyclized β -turn model consists of Cys¹-Pro²-X³-Cys⁴ tetrapeptides with blocked termini and a disulfide bridge closing the cycle (Fig. 17). ¹H-NMR parameters have been determined for X = Aib (α -aminoisobutyric acid), L-Ala, D-Ala, Gly, L-Val, and L-Leu (Venkatachalapathi *et al.*, 1982; Narasinga Rao *et al.*, 1983), and an X-ray structure has been reported for X = Aib (Prasad *et al.*, 1981). In all cases, NMR data indicated the presence of β turns in these cyclized models, with hydrogen bonding from the Cys⁴ NH to the Cys¹ C=O. Observation of an NOE between the NH of the (*i* + 2)th residue and the H^a of the (*i* + 1)th residue argues for a type II turn in the X = D-Ala peptide. The Aib peptide was stabilized in the crystal by a second intramolecular hydrogen bond from the Cys⁴ NH to the Pro C=O; these two overlapping turns comprise the beginnings of a β_{10} helix. Aib has often been observed to promote β_{10} helical structures (Section II,A,2,d). Despite the clear-cut evidence for turns in these models, CD spectral parameters were not well correlated with turn type, perhaps because of interactions with the disulfide chromophore. These researchers also reported con-

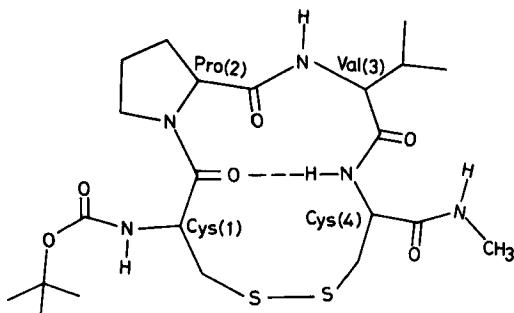


FIG. 17. Structure of a cyclized β turn, formed by ring closure of the sequence *t*-Boc-Cys-Pro-Val-Cys-NHCH₃. Reprinted with permission from Venkatachalapathi *et al.* (1982). Copyright 1982 by the American Chemical Society.

formational analysis of open analogs with the same dipeptide sequences as the disulfide-bridged cyclized turn models. Some indication of β -turn formation in solution was obtained for Piv-Pro-Aib-NHCH₃ (Piv = pivaloyl, i.e., *tert*-butylcarbonyl), Piv-Pro-D-Ala-NHCH₃, and Piv-Pro-Gly-NHCH₃ as previously described for closely related peptides (Boussard *et al.*, 1974; Aubry *et al.*, 1977, 1980b) (see Fig. 18). Observation of an NOE between Pro H^a and X NH again supported a type II β turn when X = D-Ala or Gly. When X = Val or Leu in the linear models, results were less clear-cut. A conformation stabilized by two γ turns was suggested.

In fact, relatively few studies have been specifically directed at γ turn-containing model peptides. In linear peptides, the γ turn or C₇ conformation has been observed predominantly in poor solvents, and often with Pro as the (*i* + 1)th residue (e.g., Higashijima *et al.*, 1977; Stimson *et al.*, 1977; Madison and Kopple, 1980). In the nomenclature proposed by Smith and Pease (1980) based on the original theoretical γ -turn dihedral angles (Némethy and Printz, 1972), the observed turns for L residues in position *i* + 1 are actually inverse turns (ϕ = -70, ψ = +70). Khaled *et al.* (1976) have proposed a γ -turn conformation for a protected tripeptide, *t*-Boc-Gly-Val-Gly-OMe, in CDCl₃. Similar results have been obtained by Giersch *et al.* (1982b) with *t*-Boc-X-Pro-Y-OBzl tripeptides in CDCl₃. No γ -turn conformations have been reported in crystals of linear peptides. In cyclic pentapeptides, several examples of γ turns have been observed both in solution and in crystals (Table IV). Particularly clear-cut examples contain proline residues in the *i* + 1 position; Pro ring ¹H- and ¹³C-NMR parameters display characteristic values when a γ turn is present (e.g., upfield-shifted ¹³C^β resonance, and low-field ¹H^a resonance). It is not clear whether the frequent finding of γ turns with Pro as the (*i* + 1)th residue derives from an intrinsic tendency of Pro to adopt this conformation or from the availability of diagnostic parameters that flag the presence of Pro-containing γ turns more clearly than those involving other residues. The narrow range allowed to Pro ϕ and ψ angles includes values associated with γ turns, suggesting that an intrinsic tendency may contribute to some extent. Structures of cyclic pentapeptides determined by X-ray diffraction have provided examples of Pro-containing γ turns; no cyclic pentapeptide structures have been reported with other (i.e., nonprolyl) residues in the (*i* + 1)th position. A γ turn with Ala in the (*i* + 1)th position was found in the crystal structure of the undecapeptide iodocyclosporin (Petcher *et al.*, 1976). However, many of the cyclic pentapeptides lacking proline which are proposed to adopt γ -turn conformations in solution have not been crystallized.

Reported examples of γ turns in linear peptides are stable only in poor

solvents, particularly chloroform. It appears that solvation of amide NH and C=O groups by hydrogen-bonding solvents competes effectively with the intramolecular interactions. This is in contrast to β turns which are observed in good hydrogen-bonding solvents. (This observation may reflect the stabilizing effect of a good solvent on a turn due to its non-hydrogen-bonded $i + 1$ to $i + 2$ peptide bond.) Note in particular that for the linear terminally blocked tripeptides described above the γ -turn conformation is one of very few available that provide an intramolecular hydrogen-bonding acceptor for the amide NH of residue i . In chloroform, intramolecular hydrogen bonding of the C=O has less influence on conformation than intramolecular hydrogen bonding of the NHs, since chloroform itself can act as a weak hydrogen bond donor. These observations underline the importance of examining alternative conformational possibilities and possible solvation effects in attempting to draw conclusions about conformational preferences and stabilities.

In sum, the many experimental data available from turn-forming model peptides may lead to some general conclusions. The set of peptides for which these generalizations apply may be a somewhat restricted one; those model peptides observed to take up turns nearly all contained proline, and many had glycines in their sequences. This finding surely reflects both the strong tendencies of proline and glycine to favor turns and the characteristics of the Pro- or Gly-containing peptides that made them amenable to physical studies (e.g., their tendency to crystallize, their reduced conformational freedom). With these conditions in mind, the following can be noted: (1) β turns are stable in a variety of environments, including strong solvents and crystals. Hence, intramolecular hydrogen bonding is not of primary importance to the stability of a β turn (Section II,A,2,b). (2) γ turns are stable only in the absence of competing hydrogen-bonding interactions with solvent or other peptides. Hydrogen bonding seems to play a major role in their stability. (3) Proline is strongly favored in position $i + 1$ of a β turn but can be accommodated in the $i + 2$ position following a Gly or D residue in position $i + 1$. [Note: An unusual type I Pro-Pro turn was found in the crystal structure of cyclo-(Gly-Pro-Pro)₂ (Czugler *et al.*, 1982).] Occasional cases of an L-X-Pro sequence with a *cis*-X-Pro bond in the $i + 1$ and $i + 2$ positions occur (see Section II,A,2,d). (4) Glycine can play the role of a D residue in peptides that are predominantly composed of L residues. It promotes formation of β turns favored by heterochiral sequences. (5) More examples of heterochiral β turns were observed than homochiral β turns, suggesting greater stability. (6) L residues in position $i + 1$ adopt inverse γ turns ($\phi = -70$, $\psi = 70$) as opposed to γ turns ($\phi = 70$, $\psi = -70$). (7) Although insufficient data are available to warrant definite conclusions

about propensities of specific residues (other than Pro and Gly) to occur in particular positions in turns, it may be tentatively suggested that bulky hydrophobic residues, especially β -branched ones, do not occur readily in position $i + 2$ of β turns, and that Ser, on the other hand, has a preference for position $i + 2$ of β turns (see Section II,A,2,c).

b. Importance of Hydrogen Bonding in β Turns. Examination of β turns in a set of 23 proteins in light of stereochemical requirements for hydrogen bonding revealed that formation of an $i + 3 \rightarrow i$ hydrogen bond is not possible in many of these turns (Némethy and Scheraga, 1980). These authors pointed out that the "standard" β -turn dihedral angles (Venkatachalam, 1968) do not lead to optimal hydrogen bonding, especially for type I turns. In several crystal structures of model peptides (Karle, 1981a) an $N_{i+3} \cdots O_i$ distance is found (3.2–3.5 Å) that is longer than found in more ideal hydrogen bonds. These observations together suggest that hydrogen bonding is not of major importance to the stability of β turns.

Model peptide studies in solution lend support to these suggestions. While shielding or inaccessibility of the NH of the $i + 3$ residue of β turns is observed in model peptides, the chemical shifts of these NHs (which are directly related to strength of hydrogen bonding, in the absence of magnetic anisotropic effects) show a range of values (Pease, 1979; Giersch *et al.*, 1981a). The lowest field resonances (~8 ppm) among these occur for the NH of the $i + 3$ residue in type II β turns. The NH of residue $i + 3$ in either type I or II' β turns resonates at a higher field chemical shift (~7.5 ppm), indicating weaker hydrogen bonding. NHs of the residue in the $i + 2$ position of γ turns in cyclic pentapeptides resonated at about the same chemical shift as the hydrogen-bonded NH in a type II β turn (Pease and Watson, 1978; Pease, 1979). In these studies and others (Kopple *et al.*, 1973, 1974; Kessler, 1982), resonances of exposed NHs in a hydrogen-bonding solvent (H_2O , DMSO) occur at lower field than those NHs supposedly involved in intramolecular hydrogen-bonding interactions. This observation suggests that even the best of the intramolecular hydrogen bonds are not of comparable strength to solvent-peptide hydrogen bonds. All of these data reinforce the idea that intramolecular hydrogen bonding is of varying importance in turns, and in most cases contributes little (less than solvent interactions could) to the stability of the turn.

c. Side Chain-Backbone Interactions in β Turns. Turns are intrinsically polar and contain amide groups that are not hydrogen bonded within the turn backbone elements. Thus, hydrogen-bonding partners for turn amide groups will include solvent, other peptide moieties, or side chains. Model peptide studies provide examples demonstrating the latter type

of interaction and raise the possibility of its general importance in turns in peptides and proteins.

Serine is a frequent participant in turns in proteins. Pointing out that it is often not possible to determine whether a side chain to backbone interaction contributes favorably to the stability of Ser-containing turns in proteins, Aubry and Marraud (1983) performed solution and solid-state conformational analyses on model peptides containing Ser. Piv-Pro-L-Ser-NHCH₃ and Piv-Pro-D-Ser-NHCH₃ were found both in solution and in crystals to adopt β -turn conformations, the former type I and the latter type II, in which the Ser O γ was involved in a hydrogen-bonding interaction with the Ser NH. These authors noted that this interaction can occur in Ser residues outside of turns; in fact, they suggest that it may account for a preferred rotamer distribution seen for serines in proteins. But, in turns, this O γ ··· NH hydrogen bond is favored when Ser is in the $i + 2$ position (where the Ser side chain is *axial*; see Section II,B,1 and Fig. 19).

A comparable investigation revealed that the basic His imidazole ring in position $i + 2$ of β turns also participated in an intramolecular interaction with the His amide NH in Piv-Pro-L-His-NHCH₃, both in crystals and in solution (Aubry *et al.*, 1983). Upon protonation of His, the turn conformation was apparently destabilized, and an open conformer was observed.

In proteins, the side chains of Asn residues are found to interact with backbone elements in turns (Richardson, 1981). Related studies with Asp- and Asn-containing model peptides suggest similar side chain-to-backbone interactions within β -turn conformations (M. Marraud, personal communication).

d. Participation of Unusual Amino Acids in Turns. Natural peptides from microbial organisms often incorporate unusual amino acids with unique stereochemical character: for example, N -methyl, α -methyl, or α,β -unsaturated amino acids (Umezawa *et al.*, 1978). Furthermore, peptide chemists seeking to develop analogs to bioactive peptides invoke unusual amino acids both to confer resistance to proteolysis and to favor particular conformations (Hruby, 1982). Several unusual amino acids have been incorporated into model peptides, and their propensities to occur in turns have been examined.

The presence of an additional substituent on the α -carbon of an amino acid, such as α -aminoisobutyric acid or Aib, leads to a marked reduction in its available conformational space. Aib residues have been found to occur frequently in turns in protected dipeptide amides, in tri-, and in tetrapeptides. For example, Piv-L-Pro-Aib-NHCH₃ adopts a type II β turn in crystals and in solution (Prasad *et al.*, 1982), Cbz-Aib-

L-Pro-NHCH₃ (Cbz = carbobenzoxy) takes up a type III β turn in crystals (Prasad *et al.*, 1979), Cbz–Aib–Aib–Ala–NHCH₃ and Cbz–Aib–Aib–Aib–Pro–OMe are proposed to exist in one and two β turn (all type III) -containing conformers, respectively, in solution (Iqbal *et al.*, 1981), and Cbz–(Aib–Pro)₂–OCH₃ takes up a type I β turn with Aib¹–Pro² at the corners both in solution and in crystals (Venkatachalapathi *et al.*, 1981). In longer peptides, Aib is found to promote 3_{10} helical conformations (i.e., linked type III β turns) (Rao *et al.*, 1980; Benedetti *et al.*, 1982; Jung *et al.*, 1983; and Balaram, 1983).

In theory, N-methylated residues can occur in positions $i + 1$ or $i + 2$ of β turns with little change in turn stability, and in fact this expectation has been used as support for the presence of turns in bioactive conformations of peptide hormones (see Section II,B,2,d). Some experimental evidence for the effect of N-methyl groups in β turns has been provided by Vitoux *et al.* (1981a,b). Conformations in solution and in crystals of protected dipeptide amides, Piv–L-Pro–N–Me–X–NHCH₃ (X = Gly, L-Ala, D-Ala), were compared with the homologous unmodified peptides. When X = Gly or D-Ala, the type II β -turn conformation preferred by the unmodified peptide was retained in the N-methylated version. By contrast, the homochiral peptide (X = L-Ala) displayed very different conformations when methylated or not. The unmodified peptide adopts a type II β -turn conformer in crystals (somewhat irregularly, possibly due to crystal packing interactions), and a type I β -turn conformer in solution. Presence of the $i + 2$ N-methylated residue caused a *cis*-Pro–N–Me–Ala peptide bond to be preferred, with retention of a β turn (type VI) which has a short ($N \cdots O = 2.79 \text{ \AA}$) linear hydrogen bond between the methylamide and the Piv C=O. Similar turns involving $i + 2$ N-methyl residues are found in the cyclic heptapeptide Ilamycin B1 (Iitaka *et al.*, 1974) and in the cyclic undecapeptide cyclosporin A in solution (H. Kessler, personal communication).

An interesting specific interaction between water and a β turn was observed in the N-methylated dipeptide series (Aubry *et al.*, 1981b). Piv–L-Pro–N–Me–D-Ala–NHCH₃ was crystallized in two forms: anhydrous and monohydrated. The conformation of the peptide in both crystal states was a folded one; in the anhydrous state, a type II β turn occurred, and in the monohydrated state, an open turn which included a water-bridged hydrogen bond between the amide NH and the Piv C=O (a 12-membered ring, Fig. 18) occurred.

α,β -unsaturated, or dehydroamino acids have been incorporated into analogs of bioactive peptides with retention of high potency (e.g., English and Stammer, 1978). One possible conformation of this constrained residue, based on both experimental and theoretical results, is $\phi = -70^\circ$, $\psi = \sim 0^\circ$ (Ajo *et al.*, 1980), which ought to be allowed in the $i + 2$

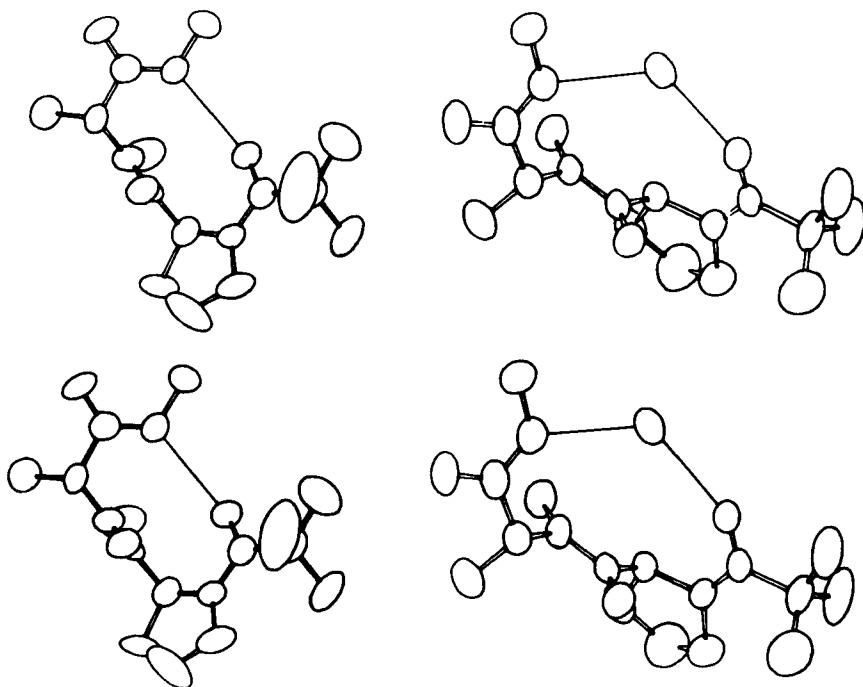


FIG. 18. Stereoscopic view of Piv-L-Pro-N-Me-d-Ala-NHCH₃ as it occurs in two crystal forms that differ by the presence of one molecule of water per peptide in that shown at the bottom. Note the β turn in the upper structure, which opens up to form a water-bridged enlarged cycle in the bottom structure. Reprinted with permission from Aubry *et al.* (1981b). Copyright 1981 by Munksgaard International Publishers, LTD.

position of β or γ turns. Solution conformational studies of linear (Bach *et al.*, 1983) and cyclic (Bach and Giersch 1985) model peptides suggest that either type of turn may incorporate a dehydro residues in the $i + 2$ position. Aubry *et al.* (1985) found that Piv-Pro- Δ Phe-NHCH₃ strongly favors a type II β turn in solution and in crystals (Δ Phe = α,β -dehydro-Phe).

The influence of an amide bond replacement, specifically CH₂—S, on turn conformations has been studied by synthesis of a cyclic "pseudo"-pentapeptide: cyclo-(Gly-Pro ψ [CH₂S]Gly-d-Phe-Pro) (Spatola *et al.*, 1983). The amide bond replacement occurs at the $i + 1 \rightarrow i + 2$ linkage of a β turn in the analogous cyclic pentapeptide; this bond is also the only amide linkage in the peptide not involved in an intramolecular hydrogen bond. NMR analysis demonstrated that the cyclic pseudopeptide in chloroform retains the β turn as well as a γ turn (around d-Phe-Pro-Gly) present in the corresponding pentapeptide.

*B. Naturally Occurring Peptides**1. Idea of a Bioactive Conformation: Roles of Turns*

In the past decade, a great many small peptides have been discovered which perform important biological functions, and many of these have been proposed to adopt turn-containing conformations (for a review, see Smith and Pease, 1980). The implication of these proposals has generally been that the biological activity of the peptide is mediated by the turn conformation.

Unfortunately, determining relationships between structure and activity in small peptides is an elusive goal. Most of the biologically active peptides whose activities are well defined and measurable are highly flexible molecules. Despite the availability of numerous analogs and data on their respective activities, the number of possible conformations of these peptides complicates efforts to relate structural parameters and activities. Furthermore, the conformation observed in a particular environment suitable for physical measurements (most commonly bulk solvents) may well not be the receptor-bound conformation. The situation is complicated further by the likelihood that conformational change on the part of the peptide effector molecule is a necessary aspect of its biological action (Williams, 1977). Nonetheless, a great deal of effort has been invested in studies of preferred solution conformations of biologically active peptides. There is, no doubt, a distribution of conformations in solution, and among the various conformers it is likely that the conformer that mediates biological activity might be found. But no physical method can yield a description of a single conformer among a population of conformers that interconvert rapidly relative to the time scale of the measurement. Moreover, there is no a priori strategy that would select the bioactive conformation from the population. Until a receptor can be purified and isolated and methods can be developed to examine the bound peptide effector molecule, only indirect approaches can be taken to determine the bioactive conformations of flexible peptides.

An approach that has proved useful in elucidating structure-activity relationships of bioactive peptides is the development of conformationally constrained analogs that retain the biological activity of the native molecule¹ (Veber, 1981; Hruby, 1982; Hruby and Mosberg, 1982; Hruby *et al.*, 1983; Schiller and DiMaio, 1983). The limited conformational space that is available to these constrained analogs must include

¹ It should be borne in mind that interpretation of bioactivity data for a peptide involves questions of rate of breakdown of the effector and its efficiency of transport to the target tissue. Receptor-binding assays may circumvent some of these complexities, but are not always available.

conformers that interact productively with the receptor. In many cases, the analogs that are so designed manifest antagonist activity (Hruby and Mosberg, 1982), although instances of agonist behavior have also been reported (Sawyer *et al.*, 1981). It has been suggested that a peptide effector molecule undergoes a conformational change after binding its receptor (Williams, 1977); this conformational change initiates the transduction of the binding message into a subsequent biological response (Hruby *et al.*, 1983). Although a constrained analog may be able to bind, it may be precluded from undergoing the conformational change necessary for eliciting a response.

Useful insights into the roles of turns in peptides may be gained by considering bioactive peptides for which conformationally constrained analogs are available. Several analogs have been designed based on the premise that stabilization of a turn conformation will lead to an active analog. Five examples will be discussed: somatostatin, melanocyte-stimulating hormone, bradykinin, luteinizing hormone-releasing hormone, and enkephalin. The strategies employed in the design derive in large part from correlations based on model peptide studies discussed above (see section II,A,2,a). For example, in four of the peptides discussed (somatostatin, melanocyte-stimulating hormone, luteinizing hormone-releasing hormone, and enkephalin), observation of enhanced activity in an analog with a D-residue substitution was used as the original indicator of a turn-containing bioactive conformation.

A small peptide that elicits a particular biological response must maintain a suitable orientation of binding groups for productive interaction with a receptor. A turn is a means of stabilizing a folded conformation in a small molecule via short-range interactions, which results in proper arrangement of groups essential for receptor binding. Small peptides are not subject to long-range interactions that stabilize protein conformations. The backbone conformation (i.e., the orientation of amide groups) may not be essential to the activity of many peptides. Instead, optimal placement of side chains seems to be the requirement for bioactivity. Somatostatin and enkephalin (both discussed below) offer compelling support for these ideas.

If the backbone conformation is not critical to activity, then favoring a type I, type II, or type II' β turn in a particular peptide would not alter activity provided that the favored orientations of the side chains were the same as those in the native molecule. Since turns are quasi-cyclic structures, one can consider the orientation of side chains with respect to the plane of the turn backbone as being either axial or equatorial. The various types of turn lead to axial (up or down) or equatorial dispositions of side chains, depending on the configuration of the corner residues.

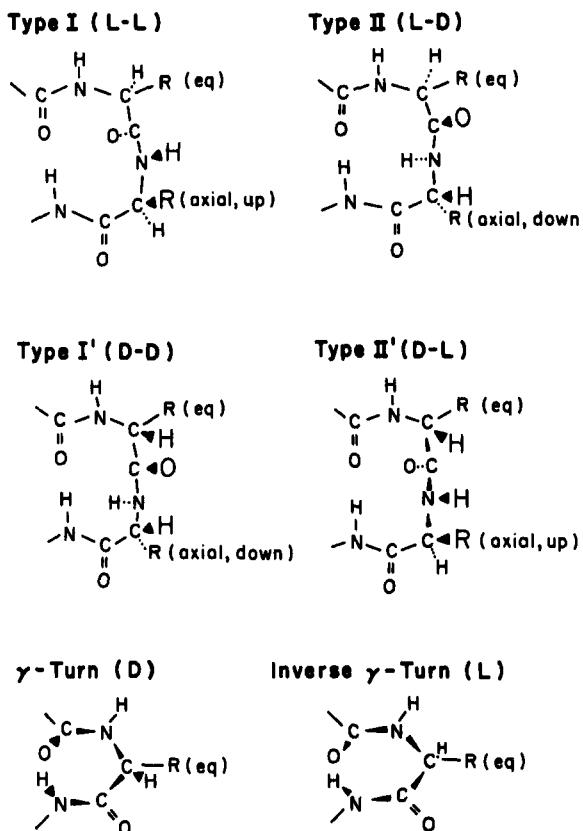


FIG. 19. Diagrammatic representation of the orientation of side chains in various hydrogen-bonded turns. The configurations of the residues shown are listed on the figure; they correspond to the favored configurations for the standard turn types. Note that the residue in position $i + 1$ of standard β turns always has its side chain oriented equatorially, and that in position $i + 2$ always has its side chain oriented axially (up or down).

For example, in a type I β turn, the side chains of L residues in position $i + 1$ and $i + 2$ of the turn are oriented equatorially and axially (up), respectively. Figure 19 illustrates the various possibilities for hydrogen-bonded β and γ turns. Note that all standard β turns orient the side chains of residues in position $i + 1$ equatorially and those of the residue in position $i + 2$ axially (up or down). This observation suggests that it may be worthwhile to think in terms of a general β -turn topography of binding groups and to expect complementary binding regions in receptors for bioactive molecules with β -turn conformations.

2. Naturally Occurring Peptides with Turn-Containing Bioactive Conformations

a. Somatostatin. Somatostatin is a tetradecapeptide with a cyclic structure formed by a cystine bridge, Ala¹-Gly²-Cys³-Lys⁴-Asn⁵-Phe⁶-Phe⁷-Trp⁸-Lys⁹-Thr¹⁰-Phe¹¹-Thr¹²-Ser¹³-Cys¹⁴. Synthesized in the hypothalamus and other endocrine glands, somatostatin inhibits the release of glucagon, insulin, gastrin, secretin, and growth hormone (Guillemin and Gerich, 1976).

The cyclic nature of this potent natural peptide effector makes it a promising target for correlating conformation and activity. Veber *et al.* (1979, 1981) have succeeded in designing a number of analogs of somatostatin that are smaller and less flexible than the parent molecule. Each of the analogs was designed based on a proposed bioactive conformation of somatostatin which includes a β turn at Phe⁷-Trp⁸-Lys⁹-Thr¹⁰ in the receptor-binding region. The original support for this model of the bioactive conformation came from the greatly enhanced activity (relative to somatostatin) of a D-Trp⁸ analog (Arison *et al.*, 1978). The "extra-turn" portion of the native molecule was replaced by progressively shorter bridges that retained the necessary stereochemical character to accommodate the essential β turn (Fig. 20). This work culminated in the development of a highly active cyclic hexapeptide analog of somatostatin in which the four residues from positions 7 to 10 are linked by the dipeptide segment, Phe-Pro (structure 3 in Fig. 20).

Some physical observables were found to correlate with biological activity of analogs. Most importantly, potency of somatostatin analogs followed in many cases the presence in ¹H-NMR spectra of a strongly upfield-shifted Lys H^γ resonance (Arison *et al.*, 1978; Veber, 1979). The perturbation of the Lys side-chain protons was attributed to a ring current shift due to proximity of the Trp ring. Interestingly, this upfield shift was not originally observed in the native peptide; it was first seen in the potent analog in which D-Trp was substituted for L-Trp at position 8. Subsequently, however, low-temperature data for methanol solutions of native somatostatin revealed the diagnostic high-field Lys H^γ signal (Arison *et al.*, 1981). This result and other data underline the fact that, despite its cyclic nature, native somatostatin is a flexible molecule with a distribution of conformational states. Observation of a strong correlation between activity and a spectral parameter in analogs suggests that the spectral parameter may be a characteristic of the bioactive conformation or at least of the requisite spatial arrangement of side chains. Furthermore, the magnitude of the ring current shift seen in the active somatostatin analogs suggests strongly that they favor one conformation

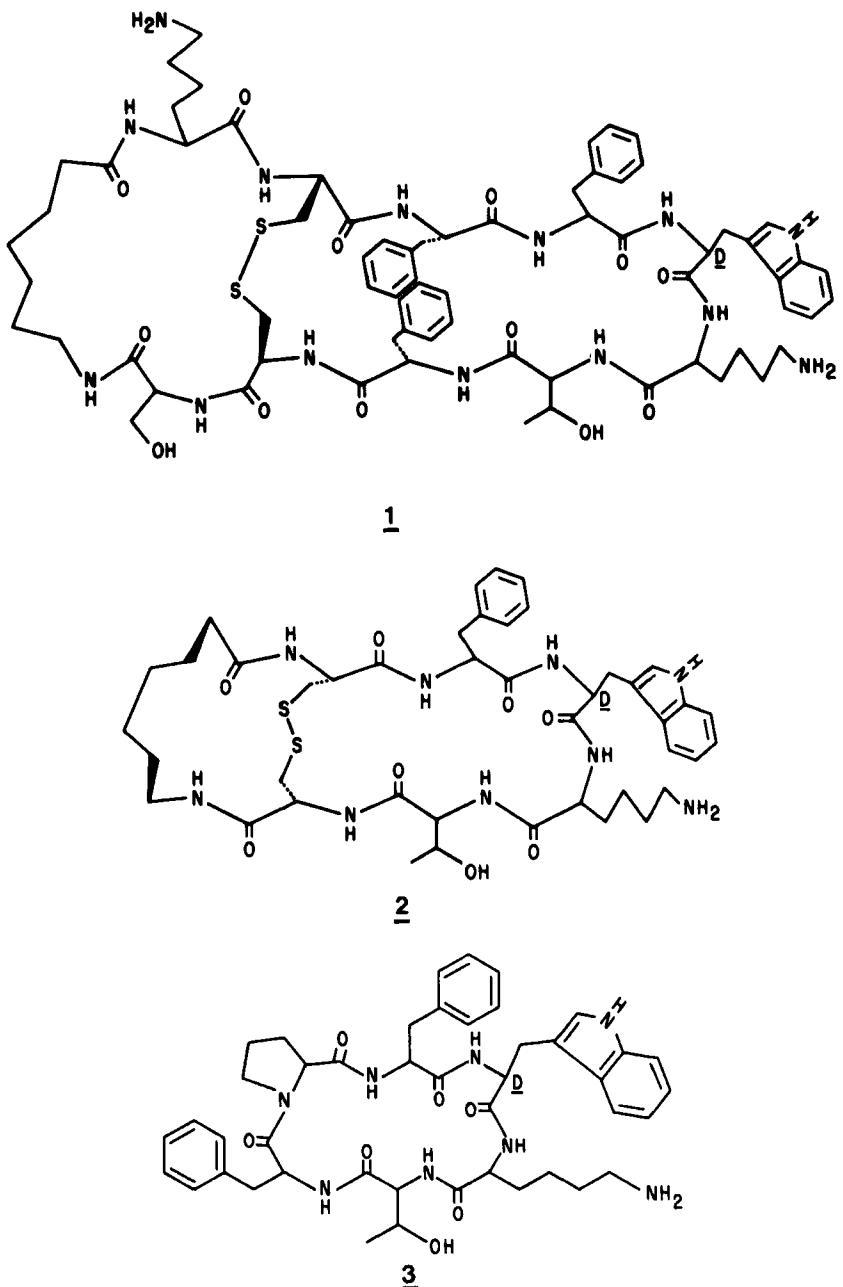


FIG. 20. Structures of three cyclic analogs of somatostatin designed by Veber *et al.* (1979, 1981) showing the progressive reduction of size of the analogs with retention of the tetrapeptide sequence encompassing the proposed β turn. Structures 1 and 2 reprinted with permission from Veber (1981). Copyright 1981 by Pierce Chemical Company.

with the Lys and Trp side chains spending most of the time in proximity. Stated otherwise, this sort of spectral parameter is a hallmark of existence of a highly preferred conformation in solution. In contrast, observation of the upfield shift *only* at low temperature in native somatostatin suggests that a distribution of conformations exists with the bioactive one present as a low-energy conformer (Arison *et al.*, 1981).

The solution conformation of the potent cyclic hexapeptide analog of somatostatin is proposed (Veber, 1981) to contain both the D-Trp-Lys type II' β turn and an unusual Phe-Pro type VI (Lewis *et al.*, 1973) (*cis*-Phe-Pro bond) β turn. NMR data (temperature dependences and exchange rates of NH proton resonances, coupling constants, and characteristic Pro ^{13}C signals) are consistent with this proposal (Veber, 1981; Veber *et al.*, 1981). In addition, an interesting stacking arrangement in which the Pro ring is sandwiched between the two Phe rings is suggested by several strong ring current shifts of Pro signals. One face of this peptide is very hydrophobic while the other is more hydrophilic, giving the peptide an overall amphiphilic character.

From the results on the potent hexapeptide and other constrained analogs, the likely mode of interaction of the parent hormone can be deduced (Veber, 1981). Considering the bioactive conformation of the peptide as the hand, it should be possible to develop a reasonable image of the glove (i.e., the receptor-active site). Furthermore, in cases of receptor heterogeneity, analogs with altered selectivity can shed light on binding interactions of distinct classes of receptors.

The potency of somatostatin analogs with D-Trp in position 8 relates to a point previously raised (Section II,B,1). The β turns favored for L-Trp-L-Lys and D-Trp-L-Lys sequences as the (*i* + 1)th and (*i* + 2)th residues would be types I and II', respectively. While these two turns differ in ϕ, ψ angles of the backbone, they show similar orientations of side-chain groups: equatorial Trp and axial (up) Lys (Fig. 19).

In sum, the somatostatin story provides a well-established example of a bioactive conformation that contains a turn, specifically a β turn. Further, the results from somatostatin and analogs [including a recently reported retroenantiomeric analog (Freidinger *et al.*, 1983)] illustrate the necessity of particular side-chain orientations for activity in a bioactive peptide and the relative unimportance of the backbone conformation.

b. α -Melanocyte-Stimulating Hormone. α -Melanocyte-stimulating hormone (α -MSH or α -melanotropin) exhibits strong effects on a variety of tissues of lower vertebrates, mammals, and, to some extent, man (Schwyzer and Eberle, 1977). These effects include activation of adenylate cyclase and tyrosinase and melanin deposition. It has been postulated that

this tridecapeptide, Ac-Ser¹-Tyr²-Ser³-Met⁴-Glu⁵-His⁶-Phe⁷-Arg⁸-Trp⁹-Gly¹⁰-Lys¹¹-Pro¹²-Val¹³-NH₂, has two "messages" (i.e., portions of its structure that interact with a receptor to elicit a particular response) within its sequence: Glu⁵ . . . Trp⁹ and Gly¹⁰ . . . Val¹³-NH₂ (Eberle and Schwyzer, 1976).

Sawyer *et al.* (1981, 1982, 1983) focused on the activity of the first message, i.e., melanin release, in a frog skin bioassay and found marked enhancement of potency in analogs with a d-Phe⁷ substitution. Based on theoretical considerations and model building, they concluded that a turn within the Glu⁵ . . . Trp⁹ region might be related to the activity of α -MSH.

From this reasoning, they designed what has turned out to be the most potent conformationally constrained analog of a bioactive peptide to date (Sawyer *et al.*, 1981, 1982, 1983). They substituted Met⁴ and Gly¹⁰ of α -MSH with cysteine residues and closed the disulfide bridge (Fig. 21). In the frog skin assay, [Cys⁴,Cys¹⁰]- α -MSH is a superagonist, 10,000 times more potent than native α -MSH. From this result, Sawyer *et al.* (1983) argued that the bridged analog constrains the peptide to a conformation favorable for hormone-receptor interaction. The magnitude

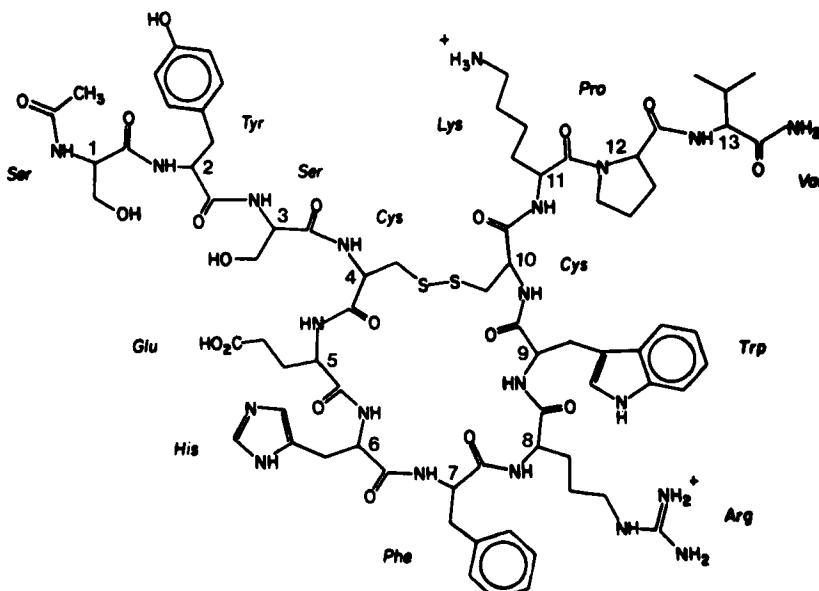


FIG. 21. Highly potent cyclized analog of α -melanocyte-stimulating hormone designed by Sawyer *et al.* (1981, 1982, 1983). Reprinted with permission from Sawyer *et al.* (1983). Copyright 1983 by Pierce Chemical Company.

of this enhancement, however, suggests that additional factors such as resistance to enzymatic degradation are playing a role. An alternative assay (mouse melanoma adenylate cyclase) found the constrained analog to be three times more potent than α -MSH, indicating similar (but not identical) requirements for productive receptor interaction in these distantly related vertebrates (Sawyer *et al.*, 1981). Model building led these researchers to conclude that a β -turn conformation was the likely bioactive conformation of their analog and to suggest that amphiphilicity was a potentially important attribute of the turn region (cf. somatostatin) (Sawyer *et al.*, 1983).

While the design strategy described yielded impressive results, there are as yet few data to substantiate the adoption by the analog of the intended conformation. The ring size of the disulfide-bridged analog is an unusual one (23 atoms), encompassing seven residues, and unlike the six-residue rings of oxytocin and vasopressin, is not readily correlated with a likely backbone conformation. It will be exciting to explore the conformational preferences of this superagonist.

c. *Bradykinin.* The linear nonapeptide hormone bradykinin, Arg¹-Pro²-Pro³-Gly⁴-Phe⁵-Ser⁶-Pro⁷-Phe⁸-Arg⁹, is one of the most potent endogenous vasodilators and acts both as a depressor and myotropic substance (Rocha é Silva *et al.*, 1949). Although it has been extensively examined in solution for evidence of a preferred conformation (Cann *et al.*, 1973; Ivanov *et al.*, 1975; London *et al.*, 1978, 1979; J. W. Fox *et al.*, 1981), no conclusive picture of either the distribution of conformational states or the bioactive conformation has emerged.

The approach of designing a conformationally constrained, cyclic analog has yielded fruitful results for bradykinin (Chipens *et al.*, 1981). The success of this approach argues that the bioactive conformation of this hormone contains a turn. The design of a cyclic analog was based on an analysis of the functions of various portions of the bradykinin sequence. Chipens (1983) has elaborated a general theory of peptide-receptor interaction that invokes a folded, quasi-cyclic conformation of the peptide effector (Fig. 22). The folding is stabilized by an interaction between the C-terminal residue (carboxyl) and a positive side chain of a residue near the N terminus. In bradykinin, they suggest that the side chain of Arg¹ performs the latter function, the Gly⁴ residue serves as a "hinge," and that the specificity sequence (i.e., "the message") is Phe⁵-Ser⁶-Pro⁷-Phe⁸. Therefore, the cyclic peptide analog, cyclo-(N^{ε1}-Lys¹-Pro²-Pro³-Gly⁴-Phe⁵-Gly⁶-Pro⁷-Phe⁸-Arg⁹), was synthesized. It was found to have potent depressor activity (of comparable magnitude but longer duration than bradykinin) and no myotropic activity. Circular dichroism spectra of the analog and the native molecule in aqueous solution were virtually identical.

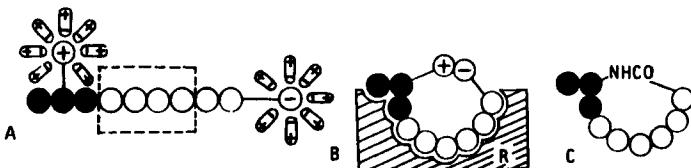


FIG. 22. Model proposed by Chipens (1983) for the interaction of peptide hormones with their receptors. A quasi-cyclic structure is proposed to exist for the peptide (B) when bound to the receptor. In (A), the black portion of the hormone is the so-called common fragment which plays a role in the adoption of the cyclic conformation. (C) Illustration of how a cyclic analog can potentiate the response of the native peptide, based on this model. Reprinted with permission from Chipens (1983). Copyright 1983 by Pierce Chemical Company.

These results argue strongly for the conclusion that the cyclic bradykinin molecule adopts a conformation close to the bioactive conformation of the native hormone when bound to the receptor responsible for the depressor activity. The cyclic structure of the analog requires at the minimum the presence of some open turns and possibly would contain a hydrogen-bonded turn. Proposals based on spectroscopic data for bradykinin have included γ turns around Pro⁷ (Cann *et al.*, 1973; Ivanov *et al.*, 1975; London *et al.*, 1978, 1979); this model could readily be accommodated in the cyclic analog.

d. Luteinizing Hormone-Releasing Hormone. Luteinizing hormone-releasing hormone (LH-RH), also known as gonadotrophin-releasing hormone (Gn-RH), a linear decapeptide of sequence *p*-Glu¹-His²-Trp³-Ser⁴-Tyr⁵-Gly⁶-Leu⁷-Arg⁸-Pro⁹-Gly¹⁰-NH₂, is synthesized in the hypothalamus and regulates levels of the pituitary hormones: luteinizing hormone (LH) and follicle-stimulating hormone (FSH) (Matsuo *et al.*, 1971; Burgus *et al.*, 1972). The levels of LH and FSH in turn regulate the critical reproductive processes of ovulation and spermatogenesis. The potential use of LH-RH and its analogs as contraceptive agents has led to a major effort in analog design; well over 1000 analogs have been synthesized and assayed (Vickery *et al.*, 1984). Of these, many possess potent biological activity, either as superagonists or antagonists.

Among the vast number of analogs synthesized, relatively few incorporate conformational constraints. In general, LH-RH has been refractory to conformationally based design of analogs. Furthermore, little experimental evidence has been obtained for any favored solution conformation(s) (Sprecher and Momany, 1979; Kopple, 1981a,b).

Nonetheless, several key findings combine to present a cogent case for a folded bioactive conformation of LH-RH, with a β turn of type II' at Tyr⁵-Gly⁶-Leu⁷-Arg⁸: (1) Substitution of Gly⁶ by a α residue leads to

retention of activity (Monahan *et al.*, 1973); (2) N-methylation of Leu⁷ is tolerated in active analogs (Ling and Vale, 1975); (3) constraining the Gly⁶-Leu⁷ link to a "bent" arrangement via incorporation of a lactam ring (Fig. 23) yields an active analog (Freidinger *et al.*, 1980).

Empirical energy calculations carried out by Momany (1976a,b, 1978) resulted in several models for the low-energy conformation of LH-RH. Among these, one with a Gly⁶-Leu⁷ β turn has gained favor based on the above activity data, although the other features of this conformation are not supported by any specific results.

To date, the most potent conformationally constrained analog of LH-RH is a cyclic antagonist synthesized by Rivier *et al.* (1981): cyclo-[Δ^3 -Pro¹-D-Phe(*p*-Cl)²-D-Trp³-Ser⁴-Tyr⁵-D-Trp⁶-NMe-Leu⁷-Arg⁸-Pro⁹- β -Ala¹⁰]. In cultured pituitary cells, this molecule reduces LH secretion 50% when it is present at a concentration 3.5 times that of LH-RH. This result reveals that the analog binds less strongly to receptors than LH-RH, but not markedly so (i.e., a factor of 3.5 in concentration corresponds to only 0.75 kcal mol⁻¹).

As for other peptides discussed in this section, the existence of this active cyclic analog of LH-RH argues for a folded bioactive conformation of LH-RH, containing turns. The activity of linear analogs with substitutions favoring a β turn at Gly⁶-Leu⁷ points more specifically to this site for a turn in the bioactive conformation of LH-RH.

Recently, a molecular dynamics study of LH-RH and its cyclic antagonist was carried out (Struthers *et al.*, 1984). The results of a molecular dynamics simulation of the conformational history of the two molecules over several picoseconds demonstrated drastic reduction in flexibility of the constrained analog relative to LH-RH. In fact, essentially one conformational family emerged for the cyclic antagonist, while the native hormone visited a variety of conformational states. The type of structure observed for the antagonist contained turns at Tyr-D-Trp-NMe-Leu-Arg and in the bridge region, at β -Ala- Δ^3 -Pro-D-Phe-D-Trp. These researchers then took advantage of the availability of the conformationally constrained active analog to search for conformations of the

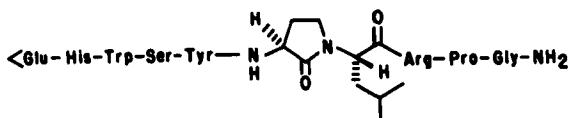


FIG. 23. LH-RH analog designed by Freidinger (1981) to limit the flexibility of the proposed turn region (around Gly⁶-Leu⁷) through the use of a lactam. Reprinted with permission from Freidinger (1981). Copyright 1981 by Pierce Chemical Company.

native hormone that resembled the analog and thus were likely candidates for the bioactive conformation. First, a low-energy conformation of LH-RH was superimposed on the cyclic antagonist, and the overlap was optimized by minimizing the root-mean-square (RMS) distances between atoms in common in the two molecules (residues 4–9, the putative message region, were chosen). This static fit of low-energy structures resulted in an RMS difference of 2.6 Å between atoms in common. A markedly improved fit was demonstrated when the native hormone was subjected to a force field augmented by a weighting function to favor homology to the constrained analog (i.e., "the template"). The RMS difference between native LH-RH and the constrained analog decreased to 0.38 Å (residues 4–9) for a modest energy cost of 4.9 kcal mol⁻¹. This "template-forcing" procedure acknowledges the capacity of the hormone for conformational adjustment and incorporates the strengths of molecular dynamics approaches into structure–activity correlations (Struthers *et al.*, 1984).

e. Enkephalin. The efforts to relate conformation and activity in enkephalins epitomize the difficulties of establishing structure/function correlations in flexible naturally occurring peptides. These endogenous opiate-like substances are linear pentapeptides; Met-enkephalin has the sequence Tyr¹–Gly²–Gly³–Phe⁴–Met⁵, and Leu-enkephalin has Leu in position 5 (Hughes *et al.*, 1975). Their small size and potent analgesic activity have encouraged analog design, and they have been the subjects of more experimental and theoretical conformational analyses in the past 7 years than any other bioactive peptides (Schiller, 1984). Moreover, the fact that rigid alkaloids like morphine compete effectively for the same receptors as the floppy enkephalins makes even more inviting the goal of defining the spatial arrangement of elements critical for activity (i.e., the "pharmacophore").

Nonetheless, a consistent model for the bioactive conformation of enkephalin remains elusive. Since 1980, several conformationally constrained analogs have been designed (DiMaio and Schiller, 1980; DiMaio *et al.*, 1982; Schiller *et al.*, 1981; Mosberg *et al.*, 1982, 1983a,b). The availability of these molecules may aid in the determination of a bioactive conformation, but as yet no conclusion has been reached.

The existence of multiple receptors with differing specificities further complicates the enkephalin story. Binding to two populations of these, the μ and δ receptors, can be assessed separately by choosing the appropriate assays (Lord *et al.*, 1977; Kosterlitz *et al.*, 1980). While they possess high affinity for both, native enkephalins bind more strongly to the δ receptors. It is generally accepted that the flexibility of the native effector molecule permits conformational change in response to interaction

with different receptors (Schiller and DiMaio, 1983; Schiller, 1984). One goal in the design of conformationally constrained analogs is to achieve enhanced selectivity of binding.

Some structure-activity relationships among enkephalin analogs have guided the design of conformationally constrained analogs: (1) Introduction of a D residue in position 2 is well tolerated in enkephalin (Pert *et al.*, 1976; Walker *et al.*, 1977); (2) analogs with N-alkyl carboxamide functions at the C terminus are active (Beddell *et al.*, 1977); (3) acylation of the N terminus destroys activity (i.e., the free amine is essential) (Ling and Guillemin, 1976).

A variety of β -turn conformations have been proposed for enkephalin in solution and receptor bound (for a review, see Schiller, 1984). They can be grouped into $1 \leftarrow 4$ and $2 \leftarrow 5$ categories, with the numerical designation corresponding to residues in the pentapeptide sequence that are hydrogen bonded. An X-ray structure determination of Leu-enkephalin crystals grown from ethanol/water found a $1 \leftarrow 4$ turn, with Gly-Gly in the corner positions and the geometry of a type I' β turn (Smith and Griffin, 1978). [Interestingly, an X-ray study of enkephalin crystals grown from water/dimethylformamide showed an extended conformation (Karle, 1983).] NMR studies in DMSO or water indicated that there was a conformational distribution favoring a $2 \leftarrow 5$ type of structure, with Gly-Phe in the turn corner positions (Roques *et al.*, 1976; Garbay-Jaureguiberry *et al.*, 1976; Jones *et al.*, 1976). Theoretical considerations based on overlap with the rigid opiates had originally suggested the same types of conformation (Bradbury *et al.*, 1976; Clarke *et al.*, 1978; Loew and Burt, 1978). Rigid-geometry empirical-energy calculations (Section II,A,1) also had yielded turn conformations of either the $1 \leftarrow 4$ or $2 \leftarrow 5$ type (Isogai *et al.*, 1977; Manavalan and Momany, 1981). Otherwise, little consensus exists among the various theoretically derived structures. An important observation was made in a study by Manavalan and Momany (1981) in which a range of low-energy conformers was found for enkephalins and analogs. These authors pointed out that despite the variations in backbone conformation, the orientations of side chains were not highly variant from structure to structure.

In spite of the apparent lack of accord on a bioactive conformation for enkephalins, a common theme in all these proposals is that the molecule is *folded* at the receptor. Furthermore, these neuroactive peptides illustrate a point previously discussed (Section II,B,1): Many different backbone conformations may be consistent with an active pharmacophore. It is the spatial disposition of side chains that appears to determine activity. Particularly strong support for this statement derives from the fact that the opiate alkaloids bind to the same receptor as the enkephalins. More-

over, it is likely that the search for a single bioactive enkephalin conformation is frustrated by (1) the sensitivity to environment of these flexible pentapeptides; (2) the likelihood that different active analogs have different backbone conformations while retaining the essential side-chain arrangement; and (3) the existence of multiple receptors that probably bind differently to the effector peptides.

Results obtained with conformationally constrained analogs are entirely consistent with the above. Schiller and co-workers (DiMaio and Schiller, 1980; DiMaio *et al.*, 1982) designed a series of potent cyclic enkephalin analogs by linking the C-terminal carboxyl group to an amino group on a side chain of varying length in position 2 (Fig. 24). The residue in position 2 was of the *D* configuration in active molecules. The analogs were found to be several times more potent than native enkephalin and to display μ -receptor selectivity. Comparison of cyclic and linear enkephalins revealed that they "share a common binding mode" (Schiller, 1984), and that the μ -receptor selectivity arises directly from cyclization (Schiller and DiMaio, 1983). The extreme rigidity of these analogs (ring sizes of 13 to 16 atoms) argues that they ought to adopt one conformation, both in solution and at the receptor, and thus determination of that conformation should shed light on the structure-activity relationship in enkephalins. Model building suggests that none of the proposed β -turn conformations can be accommodated in these cyclic molecules (DiMaio *et al.*, 1982). Ironically, preliminary conformational studies suggest that two of the cyclic enkephalins in the series (position 2 Lys or Orn, ring sizes of 15 and 16 atoms) have different conformations (Kessler *et al.*, 1983; Kessler *et al.*, 1985).

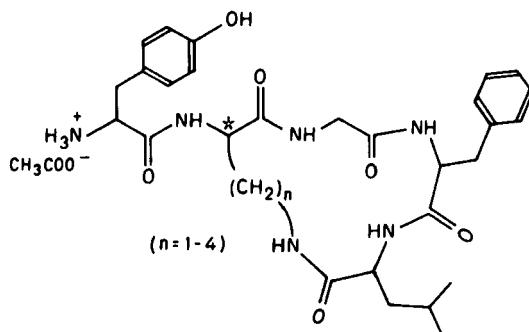


FIG. 24. Cyclized analogs of enkephalin designed by Schiller and co-workers (DiMaio and Schiller, 1980; DiMaio *et al.*, 1982). Reprinted with permission from Schiller and DiMaio (1983). Copyright 1983 by Pierce Chemical Company.

Another class of potent cyclic enkephalin analogs is formed by closing a disulfide bridge between a D-cysteine in position 2 and a D- or L-cysteine in position 5, creating a 14-membered ring (Schiller *et al.*, 1981). These analogs are nonselective with regard to μ and δ receptors, but replacement of the D-cysteine in position 2 by penicillamine (β,β -dimethylcysteine) resulted in δ selectivity which was enhanced further when the other cysteine was replaced by this restricted residue as well (Mosberg *et al.*, 1982, 1983a,b). Although all of these disulfide-bridged analogs appear from preliminary $^1\text{H-NMR}$ data (Mosberg and Schiller, 1984) to have similar conformations, with no intramolecular hydrogen bonding, they are of identical ring sizes and composition (in terms of number of residues and other covalent linkages) to the cyclized β -turn model studied by Venkatachalapathi *et al.* (1982). As discussed in Section II,A,2,a, this model peptide was found to adopt a β -turn conformation under a variety of conditions.

In sum, all available data seem consistent with a folded receptor-bound conformation for enkephalins. Nonetheless, in spite of all of the information gathered to date on enkephalins and analogs, a detailed picture of a bioactive conformation remains to be deduced.

III. TURNS IN PROTEINS

We now consider turns within proteins. A low-resolution view of any globular protein reveals a succession of more or less isodirectional chain segments interconnected by turns. This overall pattern, repeated in protein after protein, was already apparent to Kendrew more than 20 years ago from the 6 Å picture of myoglobin:

We could see at a glance that it contained the features we were looking for, namely a set of high-density rods of just the dimensions one would expect for a polypeptide chain. Closer examination showed that in fact it consisted of almost nothing but a complicated and intertwining set of these rods, sometimes going straight for a distance, then turning a corner and going off in a new direction. (Kendrew, 1961)

Turns enable the polypeptide chain to fold back upon itself. Their frequent occurrence is responsible for the globularity of globular proteins. In this section we discuss identification of turns in proteins of known structure, prediction of turns from the amino acid sequence, various roles that turns can play in the folding process, and possible functions of turns as recognition sites.

A. Identification of Turns from X-Ray Coordinates

Turns in proteins are curiously difficult to define in a satisfying way. In Section II it was seen that peptide turns are classified readily because they adhere closely to their ideal geometry. In contrast, protein turns are subject to wide structural variation. Nonetheless, rigorous definition of turn geometry is as important for proteins as it is for peptides. Prior definition is especially needed in predictive work in order to judge objectively the success of predictive methods; we return to this topic in Section III,C.

Turns in proteins are seen as changes in the overall chain direction. Although turns are conspicuous in physical models when situated at the ends of a helix or a strand of β sheet, they are harder to distinguish within less regular chain regions. It is not unusual to observe multiple reversals in chain direction within a span of 8 to 12 residues (Isogai *et al.*, 1980), as illustrated in Fig. 25. Indeed, it is often possible to subdivide the protein into segments and interconnecting turns in more than one way, and, in such cases, some degree of structural ambiguity seems inevitable. For this reason, a comprehensive geometric definition of turns in proteins is an unlikely goal.

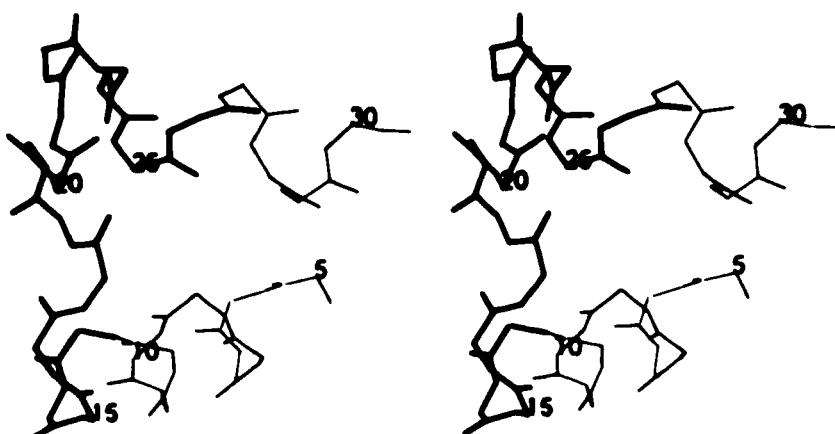


FIG. 25. Stereoview showing multiple changes in chain direction within a short chain segment, the backbone of lysozyme residues 13–26. Changes in the overall chain direction occurring at the termini of helices and strands of β sheet are classified readily due to the structural regularity of the intervening segment. Equivalent changes are more difficult to classify in less regular regions, often giving rise to instances of structural ambiguity. In this segment, turns can be defined either at positions 13–16, 17–20, 19–22, 23–26, or alternatively, only at positions 14–17 and 20–23. Neither definition is to be preferred based upon structural criteria alone.

The field has responded to this predicament by relaxing the search for uniqueness and accepting instead definitions that are objective and self-consistent. Definitions of turns in the literature can be classified as either reference or working. The idealized β turn, first recognized by Venkatachalam (1968), typifies a reference definition: four consecutive residues with a strictly defined stereochemistry are joined by a hydrogen bond between the C=O of the first turn residue (i) and the NH of the last residue ($i + 3$), as illustrated in Fig. 2.

In practice, turns seen in X-ray elucidated proteins often fail to satisfy such strict criteria. Both structural variation and measurement error can lead to nonideal geometries, and this complexity has given rise to a variety of working definitions. Strict turn definitions were soon relaxed in the work of Lewis *et al.* (1971), Kuntz (1972), and Crawford *et al.* (1973) who dropped the absolute requirement for hydrogen bonding within the turn. Kuntz (1972), in particular, considered turns with more than four residues. Later Lewis *et al.* (1973) and Chou and Fasman (1977) refined the classification of turns, introducing multiple subcategories that were derived by surveying proteins of known structure.

The most common working definitions associate turns with chain segments of 4 or 5 residues that fold sharply back upon themselves. This intuitive idea is implemented either by requiring small end-to-end distances or by demanding that the segment describe a high-angle turn, or both.

Representative working definitions, listed in Table V, identify turns from α -carbon coordinates alone, with four exceptions (viz., 3, 6, 14, 18). The most common strategy defines a chain site as a turn when the $C\alpha(i)$ – $C\alpha(i + 3)$ distance is less than 7 Å and the residues involved are not in a helix.

The histogram shown in Fig. 26 summarizes all nonhelical segments of four residues in length with small $C\alpha(i)$ – $C\alpha(i + 3)$ distances taken from 16 proteins of known structure. A clustering of these putative turn segments near their ideal reference values is apparent over interval 4.7–5.6 Å. By way of comparison, the corresponding $C\alpha(i)$ – $C\alpha(i + 3)$ distances for strict β turns (Venkatachalam, 1968) are 4.64 Å for type I, 4.96 Å for type II, and 5.39 Å for type III. Turns with larger $C\alpha(i)$ – $C\alpha(i + 3)$ separations are also evident in Fig. 26, with the next 2 Å interval from 5.6 Å to 7.6 Å containing 37% of the total segments found.

Distance-dependent strategies have the advantage of being easy to apply, and they serve to identify both β turns and more gradual turns as well. However, the chosen cutoff distance is always a compromise between overinclusion and overexclusion, as illustrated by the nonconverging tail at the high end of the distribution shown in Fig. 26. Ideally, a

TABLE V
Working Definitions Used to Identify Protein Turns from X-ray Data

Definition	Residues in turn	Reference
1. $\text{C}\alpha(i)-\text{C}\alpha(i+3)$ distance $< 7 \text{ \AA}$ and non-helical	4	Lewis <i>et al.</i> (1971)
2. $\text{C}\alpha(i)-\text{C}\alpha(i+3)$ distance $\leq 6.5 \text{ \AA}$ and chain angle $\geq 90^\circ$, or $\text{C}\alpha(i)-\text{C}\alpha(i+n-1)$ distance $\leq 6.5 \text{ \AA}$ and chain angle $\geq 120^\circ$	4	Kuntz (1972) ^a
3. $\text{C}\alpha(i)-\text{C}\alpha(i+3) < 5.7 \text{ \AA}$ and $\text{O}(1)-\text{N}(4)$ distance $< 3.2 \text{ \AA}$ and $\text{O}(1)-\text{N}(4)-\text{H}(4)$ angle $< 30^\circ$	5 - n	
4. Chain angle $> 120^\circ$	4	Crawford <i>et al.</i> (1973)
5. $\text{C}\alpha(i)-\text{C}\alpha(i+1)-\text{C}\alpha(i+2)-\text{C}\alpha(i+3)$ rotation angle approximately 0°	n	Schulz <i>et al.</i> (1974) ^b
6. Plot of (ϕ, ψ) angles versus residue number	4	Robson and Suzuki (1976) ^c
7. $\text{C}\alpha(i)-\text{C}\alpha(i+3)$ distance $< 7 \text{ \AA}$ and non-helical	4	Balasubramanian (1977)
8. $\text{C}\alpha(i-2, i-1, i, i+1)$ torsion angle between $\pm 90^\circ$	4	Chou and Fasman (1977)
9. $\text{C}\alpha(i)-\text{C}\alpha(i+3)$ distance $< 7 \text{ \AA}$ and non-helical	n	Levitt and Greer (1977)
10. Local minimum in radius of chain curvature	4	Nagano (1977)
11. $\text{C}\alpha(i)-\text{C}\alpha(i+3)$ distance $< 7 \text{ \AA}$ and nonhelical, or $\text{C}\alpha(i)-\text{C}\alpha(i+4)$ distance $< \text{C}\alpha(i)-\text{C}\alpha(i+3)$ distance $< 7 \text{ \AA}$	3 - n	Rose and Seltzer (1977) ^d
12. Profile of number of $\text{C}\alpha$ atoms within 8 \AA sphere centered at successive α -carbons	4	Kolaskar <i>et al.</i> (1980)
13. Profile of number of protein atoms within 8 \AA sphere centered at successive α -carbons	5	
14. Profile of number of protein atoms within 8 \AA sphere centered at successive α -carbons	n	Nishikawa and Ooi (1980)
15. Pattern of torsion angles on a circular plot	n	Rose and Roy (1980)
16. Distribution in the (curvature, torsion) plane	4 - 5	Srinivasan and Olson (1980)
17. Levitt and Greer criteria, method 8, this table	n	Rackowsky and Scheraga (1981) ^e
18. Profile of (ϕ, ψ) angles against residue number	n	Argos and Palau (1982)
19. Recognition of regular pattern in the space curve through successive α -carbons	4	Ramakrishnan and Soman (1982) ^f
20. Chain curvature $\geq 70^\circ$	n	Vishwanadhan and Sundaram (1982)
	2 - n	Louie and Somorjai (1983) ^g
		Kabsch and Sander (1983) ^d

(see footnotes on p. 53)

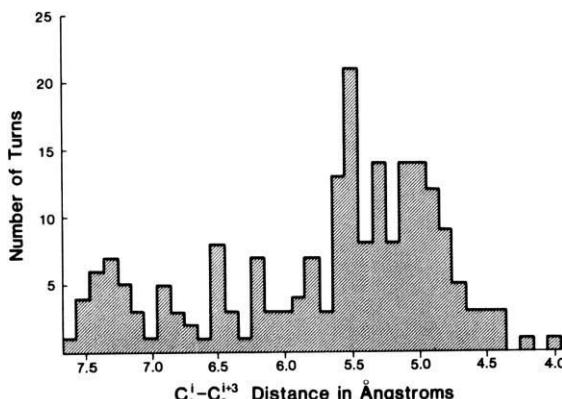


FIG. 26. Histogram of all distinct nonhelical segments with $C\alpha(i)$ – $C\alpha(i + 3)$ distances of 7.6 Å or less from 16 proteins of known structure. In all, 208 such segments were found. A clustering of distances around the values for ideal β turns is apparent. The $C\alpha(i)$ – $C\alpha(i + 3)$ distance for strict β turns is type I = 4.64 Å; type II = 4.96 Å; and type III = 5.39 Å. The proteins used were phospholipase A₂, concanavalin A, carboxypeptidase A, parvalbumin, cytochrome c, flavodoxin, high potential iron protein, lysozyme, myoglobin, papain, pancreatic trypsin inhibitor, ribonuclease A, rubredoxin, subtilisin, staphylococcal nuclease, and superoxide dismutase.

definition less sensitive to minor variations in the choice of parameters would be preferable.

Visual recognition of turns in physical models is less parameter sensitive than distance-dependent strategies. There are protein regions that exceed a $C\alpha(i)$ – $C\alpha(i + 3)$ cutoff distance of 7 Å but are recognized

^a The chain angle in a segment of n residues is the angle formed by the vector linking α -carbons i and $(i + 1)$ and the vector linking α -carbons $(i + n - 2)$ and $(i + n - 1)$.

^b Schulz defines the chain angle to be a change "in the overall direction of the polypeptide chain" (Schulz *et al.*, 1974).

^c The rotation angle is the conventional torsion angle as defined in the "Handbook of Biochemistry," 2nd edition, page A-26 (Sober, 1970).

^d The radius of chain curvature is the radius of a circle whose circumference passes through α -carbons $C\alpha(i - 2)$, $C\alpha(i)$, and $C\alpha(i + 2)$.

^e In differential geometry, a space curve can be characterized by its curvature and torsion at each point. The curvature is a measure of the rate at which the curve is turning away from its tangent line at a point. The torsion is a measure of the rate at which the curve is twisting out of the closest plane containing the tangent line (i.e., the osculating plane) at a point.

^f θ and Δ are equivalent to the dihedral and bend angles needed to construct a bent-wire backbone model (Rubin and Richardson, 1972).

^g A unique space curve is specified by the best fitting helix through the protein's α -carbons. A turn in this space curve is then defined to occur when three or more consecutive changes in axial angles, each greater than 40°, sum to approximately 180°. The axial angle is the angle between the long axes of the best fit helix at two consecutive α -carbons.

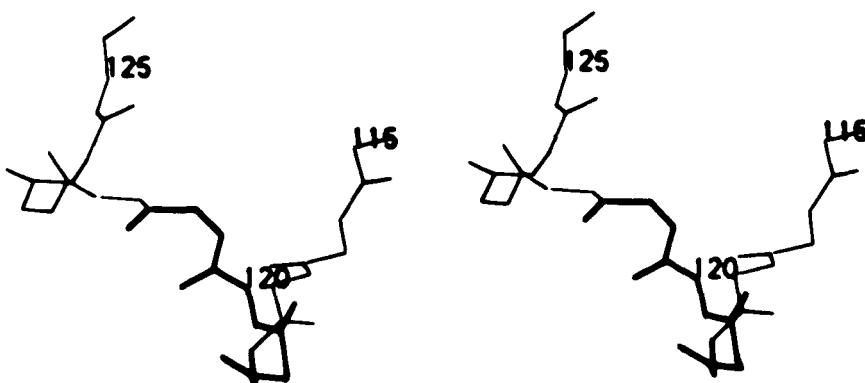


FIG. 27. Stereoview of a turn backbone in flavodoxin, residues 118–121, that is perceived as a change in overall chain direction, although the $C\alpha(i)-C\alpha(i+3)$ distance of 7.2 Å exceeds the 7 Å threshold.

visually as turns (e.g., Fig. 27). The discrepancy between the two approaches arises because a turn is signaled by both a small end-to-end distance and a change in overall chain direction. The former characteristic is a local property of the molecule, while the latter is a global property, requiring assessment of the chain leading into and away from the turn site. Visual processing is effective because it takes both characteristics into account, and a corresponding quantitative algorithm should also consider both aspects.

Several of the strategies in Table V combine both local and global chain characteristics by use of profiles that plot structural parameters of interest against the residue number. In Table V, methods 6, 10, 12, 13, 14, and 18 are all of this type. In method 10, for example, turn sites are signaled by local minima in a chain curvature profile and are not critically dependent upon the specific numeric values of such points (see Fig. 28). The approach of Louie and Somorjai (1983), method 19, also takes global chain properties into account.

The structural ambiguity previously mentioned and illustrated in Fig. 25 is analogous to cases of syntactic ambiguity seen in natural languages. It is an intrinsic property of proteins that there will be instances of multiple, self-consistent ways to partition the polypeptide chain into structural units. In such cases, no spark of geometric ingenuity can provide an escape from this inherent characteristic of the molecular syntax.

It is important to bear in mind that the goal of the preceding analysis is to identify physically meaningful structural features within the folded

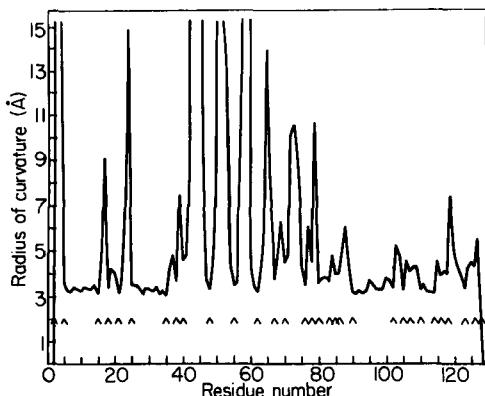


FIG. 28. A radius of curvature curve for lysozyme. The radius of curvature at α -carbon $C\alpha(i)$ is given by the radius of the circle whose circumference passes through points $C\alpha(i - 2)$, $C\alpha(i)$, and $C\alpha(i + 2)$. The radius of curvature curve is the radius of curvature at each point in the sequence plotted against the sequence number. In the resultant plot, strands of β sheet are associated with large values (e.g., the peaks around residues 47, 53, 58) while helices are associated with repeating small values (e.g., 7–14). Turns are local minima in the radius of curvature curve unless within a helix; these sites are annotated by arrows in the illustration. Reproduced from Rose and Seltzer (1977) with permission of Academic Press (London) Ltd.

protein. Experimental details about transient folding intermediates are only now beginning to emerge (Kim and Baldwin, 1982). As such information accumulates, algorithms to identify structural features will inevitably be qualified by experimental determination of these same structures.

B. Location of Turns in Proteins

Turns are usually situated at the protein surface (Kuntz, 1972). This topographical tendency is a consequence of the hydrogen-bonding requirements of backbone polar groups. In particular, turn stereochemistry precludes intraturn hydrogen bonding with NH and $C=O$ groups of middle residues (e.g., $i + 1$ and $i + 2$ of β turns or $i + 1$ of γ turns). Unlike the intrasegment hydrogen bonds seen in helices, turns must seek extrasegment partners for these middle residues. Energetically, the easiest way to meet this restriction is to dispose turns to the surface of the protein and there to solvate polar groups as necessary.

Residues found in turns are predominantly polar (Lewis *et al.*, 1971; Kuntz, 1972; Rose, 1978). Were this not the case, a destabilizing conflict

could arise between the requirements of polar main chain groups seeking hydrogen bonds and apolar side chains seeking solvent shielding. In sum, turns are highly polar secondary structures owing to the combined contributions of both backbone and side-chain groups.

1. Inside/Outside Regions in Proteins

Turns define nodes in a cagelike envelope about the protein, to be called the *turn shell*. The existence of such a shell arises from the ubiquitous nature of turns in proteins, their role as structural corners, and their tendency to lie at the solvent interface. As an illustration, Fig. 29 shows the lysozyme molecule within its turn shell. The exterior turn sites can be seen to define a tight envelope about the molecule.

The turn shell establishes a way to partition proteins into an inside and an outside without regard to atomic positions other than those in exterior turns. Since helices and strands of β sheet are punctuated at their termini by turns, the polypeptide chain appears to ricochet around within the shell boundary.

The partitioning of proteins into surface and interior regions has been quantified by many investigators (Lee and Richards, 1971; Shrake and

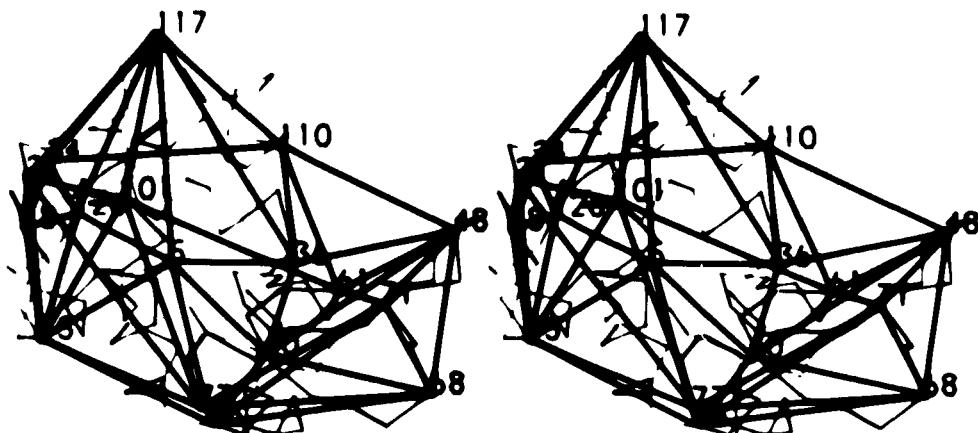


FIG. 29. Stereoview of the turn shell for lysozyme. The shell is formed by interconnecting all exterior turns. Turn sites were found using Rose and Seltzer (1977), with minima in the radius of curvature occurring at residues 5, 15, 18, 21, 24, 36, 40, 48, 61, 68, 71, 77, 86, 101, 110, 117, and 126. These sites are annotated in the figure. The α -carbon backbone is also displayed. It can be seen that the turn shell forms a tight envelope about the molecule.

Rupley, 1973; Finney, 1975; Kuntz and Crippen, 1979; Rose and Roy, 1980; Connolly, 1983). In the approach of Lee and Richards (1971), a detailed determination of the static molecular surface is calculated from X-ray coordinates and van der Waals' radii by rolling a water-sized probe uniformly around the molecule. The surface described by the probe when tangent to the protein is, by definition, the *molecular surface*, and the corresponding surface described by the center of the probe is the *accessible surface*. In either case, all protein atoms that remain inaccessible to the probe comprise the buried interior. This process, depicted in Fig. 30, has been widely applied during the past decade (Richards, 1977).

Molecular surface, calculated in this way, is developed from atomic positions, without regard to the secondary structure. Thus, the method can be used to confirm observations that the turns are at the surface of the protein.

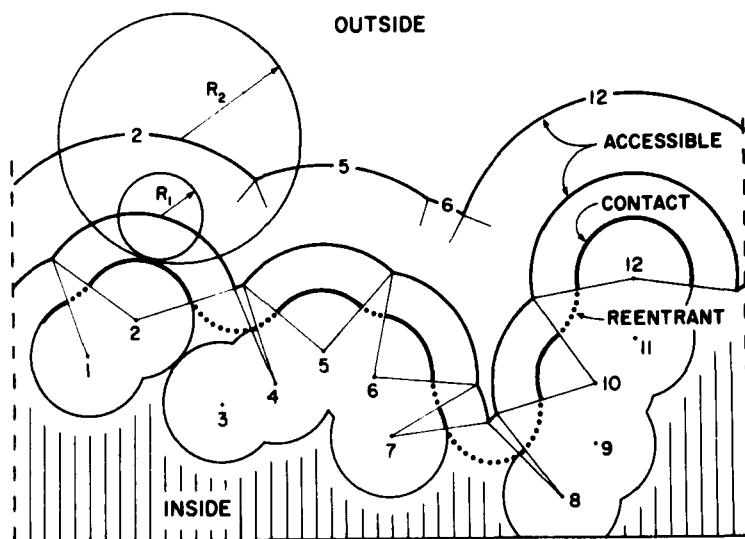


FIG. 30. A schematic representation of possible molecular surface definitions. The numbered atom centers represent a hypothetical protein, shown within its van der Waals envelope. Spheres with radii R_1 and R_2 represent probes of differing radii. The surface described by a probe where it touches a protein is the molecular surface; this surface can be partitioned into two components, viz., contact and reentrant surface, as shown in the figure. The corresponding surface described by the center of the probe is called the accessible surface. Courtesy of Frederic M. Richards (1977) and reproduced with permission of Annual Reviews, Inc.

Kauzmann distinguished "shape properties" from "short-range properties" of proteins (Kauzmann, 1959). Shape properties depend on the overall envelope of the molecule and include hydrodynamic, scattering, diffusion, and surface properties, among others. In large part, the shape properties of a protein are circumscribed by the positions of its exterior turns.

2. *Interior Turns*

Most turns do reside at the molecular surface. However, a few exceptions have been observed, situated within the hydrophobic interior. In a recent survey of refined protein structures that include solvent (Rose *et al.*, 1983), nine interior β turns were found within six proteins.

As mentioned, turns are intrinsically polar, irrespective of side-chain polarity. The finding that these polar structures are occasionally buried is rationalized by the unexpected observation in each case of one or more buried water molecules in a hydrogen-bonded complex with main chain atoms of the turn residues. These buried water molecules appear to function as structural prosthetic groups, supplementing intramolecular hydrogen bonding by providing necessary site-specific polarity (Rose *et al.*, 1983). In the case of the serine proteases, these prosthetic waters appear to be as conserved as the turns themselves (Greer, 1981; Edsall and McKenzie, 1983).

C. *Prediction of Turns*

There is a history of researchers who have attempted to predict protein secondary structure in general and turns in particular (Nagano, 1973; Chou and Fasman, 1974; Lim, 1974; Burgess *et al.*, 1974; Argos *et al.*, 1976; Garnier *et al.*, 1978) by recognizing correlations between sequence and conformational preference. Such work can be subdivided into two approaches, empirical and physicochemical.

1. *Empirical Approaches*

Empirical approaches are fundamentally statistical in character and require a data base of known structures. The typical procedure employs an empirical likelihood that relates every amino acid to each secondary structure under study. For example, to derive the likelihood that glycine will be in a turn, the ratio of occurrences of glycine in turns to total glycines is calculated from the data base and appropriately normalized, yielding the probability that an arbitrarily chosen glycine will be found in this conformation. Additional rules are then imposed to assign a joint probability to a sequence of residues.

These empirical probabilities are placed on familiar statistical ground by noting that occurrences of a residue type within a known sequence are binomially distributed (Feller, 1950); i.e., given a sequence of n residues, the number of ways k letters g (lysine) can be distributed is equivalent to the number of ways n trials can result in k "successes" and $n - k$ "failures." Many investigators have developed their counting statistics in a related fashion.

When reckoning empirical probabilities, the tally of frequencies of occurrence depends critically upon the a priori identification of helices, sheets, and turns within the data base of X-ray elucidated structures. Recognition of these secondary structures is not always straightforward, particularly in the case of turns, as discussed in Section III,A. This point is illustrated in a paper by Levitt (1978), who compared residue preferences for turns as determined by various groups, including himself (Lewis *et al.*, 1971; Finkelstein and Ptitsyn, 1971; Crawford *et al.*, 1973; Chou and Fasman, 1974, 1977; Robson and Suzuki, 1976; Levitt, 1978); he reported wide variation in the findings of these investigators. However, excluding an early study with a three-protein data base, just two of these groups used the same criteria to identify turns, and these two are in excellent agreement.

Empirical prediction methods have been widely applied and recently reviewed (Némethy and Scheraga, 1977; Chou and Fasman, 1978; Smith and Pease, 1980; Palau *et al.*, 1982). There can now be little doubt that the local sequence is an important determinant of the observed secondary structure (Schulz, 1974). However, at this writing, it appears unlikely to be the only important determinant (Schulz, 1974; Matthews, 1975; Palau *et al.*, 1982). For example, Kabsch and Sander (1984) searched proteins of known structure and identified several cases in which the same pentapeptide sequence adopts differing conformations. Whether larger data bases and more sensitive discrimination will lead in due course to significant improvement in predictive power remains an open question.

2. Turns and Hydrophobicity

Following Kauzmann's review early in this series (Kauzmann, 1959), it has been noted that residues with hydrophobic side chains tend to be situated within the protein interior. The complementary tendency appears to be a factor in determining the location of turns because turn sequences are rich in nonhydrophobic residues and are typically situated at the solvent-accessible surface of the molecule (Kuntz, 1972; Rose, 1978).

The hydrophobic effect has been interpreted in more than one way in

recent years (Hildebrand, 1979; Hansch and Leo, 1979; Ben-Naim, 1980; Tanford, 1980; Némethy *et al.*, 1981; Hvistendahl, 1983), resulting in conflicting measures of residue hydrophobicity. For example, tryptophan and tyrosine are highly hydrophobic residues as determined by Nozaki and Tanford (1971), but quite hydrophilic as determined by Wolfenden *et al.* (1981). For this reason, conclusions based upon any given scale of hydrophobicity may not survive when an alternative scale is used.

Before assessing the influence of hydrophobicity upon turn conformation, it is appropriate to review quantitative measures of hydrophobicity for the amino acids and their residues. Two general classes of hydrophobicity scales have been proposed: (1) those derived from solution measurements, and (2) those calculated empirically from X-ray elucidated coordinates. Solution scales are based upon comparisons of solubilities in aqueous and nonaqueous solvents; from these data partition coefficients are calculated. Empirical scales are based upon the extent to which each residue type is found buried within the protein's interior. In the discussion that follows, it is important to bear in mind that "buriedness" is a property of the three-dimensional structure and, unlike solubility, may be only nominally defined. This distinction will be drawn in greater detail later in the section. The two classes of hydrophobicity scales are now discussed, and an attempt is made to rationalize apparent discrepancies.

a. *Scales of Hydrophobicity: Solution Measurements.* The protein and its aqueous surrounding can be viewed as distinct solvent phases. A partition coefficient between these phases can be defined, and a residue's tendency to occur at the surface or in the interior can be described by this quantity. Ideally, the distribution of the residues between these two phases should be governed by the partitioning tendencies of their side chains, since backbone atoms remain the same from residue to residue. In practice, significant sources of nonideality arise because of the following:

1. The residues are covalently linked and cannot be treated independently. For example, a chain site with a polar residue adjacent in sequence to a nonpolar residue may not be able to satisfy partitioning tendencies of both residues simultaneously.
2. The interior of the protein is an anisotropic phase, unlike a bulk solvent.
3. The packing between segments of secondary structure is cooperative and longer range forces may predominate (Richardson, 1981; Rossmann and Argos, 1981).

4. The interior volume available to the residues varies with the molecular weight of the protein (Janin, 1979).

These complexities notwithstanding, trends in partitioning remain conspicuous; chain segments comprising turns do favor the aqueous phase for the most part (Kuntz, 1972; Rose, 1978).

The quantitative basis for treating these problems was established by McMeekin *et al.* (1935) and Cohn and Edsall (1943) who derived the free energy required to transfer amino acid solutes from an aqueous to an organic phase, $\Delta G_{\text{water} \rightarrow \text{org}}^\circ$, using the relative solubilities in each phase. The free energy of transfer, $\Delta G_{p \rightarrow p'}^\circ$, between phases p and p' is related to the partition coefficient between these phases, $K_{p \rightarrow p'}$, by the equation $\Delta G_{p \rightarrow p'}^\circ = RT \ln K_{p \rightarrow p'}$. Cohn and Edsall also provided the solubilities for many amino acids in both water and ethanol.

These studies were extended by Nozaki and Tanford (1971). Tanford (1962) noted that the free energy of transfer is not very sensitive to the choice of organic solvent and concluded that the principal factor being measured in this process is the energetic cost of removing solutes from water. Viewing the protein interior simply as another organic phase, Nozaki and Tanford (1971) developed a single scale of hydrophobicities, ΔG_t , for 11 of the amino acid residues.

Recently, Wolfenden and co-workers (1981) determined the free energy of completely stripping water away from the residue side chains, ΔG_h , by measuring the partitioning of side-chain analogs between aqueous solution and the dilute vapor phase. Wolfenden reasoned that ΔG_h is a better measure of the change that occurs in side-chain environment upon folding, because in this process interior groups are displaced from a hydrated state to one that is solvent shielded.

Surprisingly, there is no correlation between the free energy of transfer, ΔG_t , as measured by Nozaki and Tanford (1971), and the hydration potential, ΔG_h , as measured by Wolfenden *et al.* (1981); indeed, several residues that are hydrophobic in one scale are hydrophilic in the other. The correlation coefficient is -0.04 for the 11 residues common to both studies.

b. Scales of Hydrophobicity: Empirical Calculations. The other class of hydrophobicity scales, those calculated from X-ray elucidated proteins, is based on the implicit assumption that the best model for the protein's heterogeneous interior ought to be the protein itself. Many investigators have devised empirical scales by analyzing proteins of known structure (Jones, 1975; Chothia, 1976; Levitt, 1976; Manavalan and Ponnuswamy, 1978; Wertz and Scheraga, 1978; Krigbaum and Komoriya, 1979; Janin, 1979; Robson and Osguthorpe, 1979; Meirovitch *et al.*, 1980; Nishikawa

and Ooi, 1980). These studies assess the extent to which a given residue type partitions between the inside and outside of the protein by measuring the fraction of the residue population that is buried (i.e., excluded from solvent access), within proteins of known structure. An empirical partition coefficient, K_{emp} , from inside to outside is then defined as: $K_{\text{emp}} = \text{number of residues buried} \div \text{number of residues not buried}$.

A novel empirical approach was introduced by Sweet and Eisenberg (1983). They calculated the hydrophobicity of a residue by an iterative procedure that requires a scale of initial values of residue hydrophobicity and the observed frequencies of amino acid replacement within related proteins. Upon averaging these initial hydrophobicities over residue replacements from related structures, a new value is evolved that is largely independent of the initial values. Moreover, three different scales of initial values were found to converge to the same final scale, indicating the stability of the method.

c. *Comparisons between Scales of Hydrophobicity.* Various hydrophobicity scales are listed in Table VI. Table entries are characterized as a solution measurement, an empirical calculation from X-ray coordinates, or some combination of the two.

Pairwise comparisons between the hydrophobicity scales of each group are shown in Table VII. The table lists the absolute values of the correlation coefficients between residues in common to each pair of scales. Despite broad differences in methods and assumptions, overall accord between measured and empirically determined scales exists in certain cases. The comparison splits into two groups centered around either the scale of Nozaki and Tanford (1971) or that of Wolfenden *et al.* (1981). The former scale correlates well with the empirical scales of Wertz and Scheraga (1978), Meirovitch *et al.* (1980), and Robson and Osguthorpe (1979), and poorly with those of Chothia (1976) and Janin (1979). Conversely, the scale of Wolfenden *et al.* (1981) correlates well with the empirical scales of Chothia (1976) and Janin (1979) and poorly with those of Wertz and Scheraga (1978), Meirovitch *et al.* (1980), and Robson and Osguthorpe (1979). The scale of Nozaki and Tanford (1971) agrees well with the free energies of transfer between water and octanol as measured by Yunger and Cramer (1981) for the subset of amino acids in common to both studies. Scales due to Jones (1975), Levitt (1976), Manavalan and Ponnuswamy (1978), and Kyte and Doolittle (1982) are hybrids of both experimental and empirical measures, and comparisons vary.

While both measured and empirical scales are formally similar in that each is derived from a partition coefficient, there is a fundamental difference between them. The partition coefficient for a solution measure-

TABLE VI
Scales of Hydrophobicity for the Amino Acids, Their Residues, and Their Analogs

Type of scale ^a	Scale parameter	Number of residues in scale	Reference
Measured	ΔG (transfer H ₂ O to organic solvent)	11	Nozaki and Tanford (1971)
Measured/empirical	Extended scale of Nozaki and Tanford (1971) using Zimmerman <i>et al.</i> (1968)	20	Jones (1975)
Empirical	Fraction of residues at least 95% buried ^b	20	Chothia (1976)
Measured/empirical	Combination of Nozaki and Tanford (1971) and Chothia (1976)	20	Levitt (1976)
Measured/empirical	Average of scale values from Nozaki and Tanford (1971) for residues with an 8 Å sphere	20	Manavalan and Ponnuswamy (1978)
Empirical	Fraction of residues that are buried	20	Wertz and Scheraga (1978)
Empirical	$\Delta G_{\text{stabilization}}$ from pairwise contacts	20	Krigbaum and Komoriya (1979)
Empirical	$\Delta G_{\text{transfer}}$ from buried interior to solvent-accessible surface ^b	20	Janin (1979)
Empirical	Information theoretic measure of distribution between interior and surface	20	Robson and Osguthorpe (1979)
Empirical	Reduced distance from center of mass and average side-chain orientation angle	20	Meirovitch <i>et al.</i> (1980)
Empirical	Average contact number of α-carbons within an 8 Å sphere	20	Nishikawa and Ooi (1980)
Measured	$\Delta G_{\text{transfer}}$ H ₂ O to vapor	19	Wolfenden <i>et al.</i> (1981)
Measured	$\Delta G_{\text{transfer}}$ H ₂ O to octanol	16	Yunger and Cramer (1981)
Measured/empirical	Modified scale of Wolfenden <i>et al.</i> (1981)	20	Kyte and Doolittle (1982)
Empirical	Average of hydrophobicity values for residue replacements from related structures	20	Sweet and Eisenberg (1983)

^a Each scale is classified as a solution measurement, an empirical calculation, or a combination of the two.

^b Residue accessibility is measured using the method of Lee and Richards (1971).

ment, $K_{p \rightarrow p'}$, is dictated by the chemical properties of the solute : solvent system. Mathematically, $K_{p \rightarrow p'}$ is a continuous variable that can range from all solute in the aqueous phase to all solute in the nonaqueous phase. In contrast, the partition coefficient for an empirical calculation,

TABLE VII
Pairwise Correlation Coefficients between Hydrophobicity Scales^{a,b}

References	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Jones (1975) (1)	.99	.53	.66	.68	.83	.70	.63	.41	.35	.47	.34	.47	.40	.75	
Nozaki and Tanford (1971) (2)		.81	.82	.75	.91	.99	.66	.20	.28	.24	.04	.89	.63	.78	
Krigbaum and Komoriya (1979) (3)			.89	.87	.78	.74	.86	.79	.81	.79	.63	.15	.04	.78	
Wertz and Scheraga (1978) (4)				.93	.88	.72	.85	.71	.74	.71	.49	.35	.15	.80	
Meirovitch <i>et al.</i> (1980) (5)					.87	.75	.89	.79	.75	.78	.56	.29	.34	.82	
Robson and Osguthorpe (1979) (6)						.74	.83	.65	.64	.64	.48	.40	.21	.84	
Levitt (1976) (7)							.67	.62	.77	.67	.65	.05	.14	.75	
Manavalan and Ponnuswamy (1978) (8)								.86	.76	.88	.71	.25	.23	.82	
Chothia (1976) (9)									.91	.96	.89	.06	.00	.67	
Janin (1979) (10)										.87	.83	.22	.15	.58	
Kyte and Doolittle (1982) (11)										.89	.12	.07	.75		
Wolfenden <i>et al.</i> (1981) (12)											.30	.10	.57		
Yunger and Cramer (1981) (13)												.43	.22		
Nishikawa and Ooi (1980) (14)													.34		
Sweet and Eisenberg (1983) (15)															

^a The absolute value of the correlation coefficient is shown for residues in common to each pair of scales. Additional information about each scale is summarized in Table VI.

^b Horizontal scale of numbers in parentheses corresponds to those for references.

K_{emp} , is a discrete variable that is reckoned by partitioning the residue population into either of two states: (1) x , the fraction of residues classified as buried, and (2) $1 - x$, all other residues which, by subtraction, are classified as accessible. Unlike a solubility, K_{emp} is a nominal measure because it is not the protein but the investigator who decides the criterion by which a given residue is classified as buried or accessible.

i. *Comparison between ΔG_t and ΔG_h .* To better understand the differences between ΔG_h , ΔG_t , and the widely used empirical scales of Chothia (1976) and Janin (1979), the measurements are now examined in greater detail. First, the hydration potential is compared with the free energy of transfer from water to organic solvent. Wolfenden *et al.* (1981) chose the dilute vapor phase as a reference phase for transfer experiments because it is isotropic with negligible enthalpy change upon addition of solute. In contrast, the organic solvents employed by Nozaki and Tanford (1971) were hydrogen bond acceptors; these include both the aliphatic alcohols and dioxane.

Solution to vapor transfer energies reflect the aggregate energy cost of eliminating intermolecular interactions in general and hydrogen bonds in particular. The side chains with imidazole, indole, and phenol rings, together with the alcohols, are all amphipathic, containing both hydrophobic and polar groups. The free energy involved in taking these side chains out of water includes a favorable solvent entropy contribution together with an unfavorable enthalpic contribution upon breaking solute : solvent hydrogen bonds. The observed net hydrophilic values of ΔG_h for these residues indicate that the enthalpic component is dominant in the dilute vapor phase (at pH 7.0 and 25°C). However, part of this enthalpic cost is expected to be repaid in a hydrogen-bonding solvent such as ethanol.

If hydrogen bonding were a major source of discrepancy between ΔG_h and ΔG_t , then the largest differences between the two scales would be for the side chains of His, Trp, Tyr, Ser, and Thr. This is observed to be so. In the case of Trp and Tyr, the differences are particularly striking; both residues have highly hydrophobic ΔG_t values but strongly hydrophilic ΔG_h values.

Tanford finds the decrease in hydrophobicity due to an aromatic —OH group to be very small, on the order of 0.2 kcal when taken as the difference in transfer free energy between Phe and Tyr (Nozaki and Tanford, 1971). Wolfenden finds corresponding differences to be much larger (Wolfenden *et al.*, 1981): the difference in hydration potential between Phe and Tyr is 5.12 kcal, the approximate magnitude of a hydrogen bond. Such discrepancies suggest that ΔG_h potentials measure

hydrogen bond breaking, while ΔG_t potentials primarily reflect differences in hydrogen bond strength between the two solvents.

ii. Comparison between ΔG_h and Empirical Scales. A strong correlation is observed between the hydrophobicity scales of Wolfenden *et al.* (1981) and the empirical scales of Chothia (1976) and Janin (1979), as shown in Table VII. Ostensibly, the problem of totally removing residue solutes from water, as measured by ΔG_h , closely resembles the problem of creating a cavity within the protein in which these solutes can be almost totally buried. However, this comparison neglects the hydrogen-bonding capacity of groups inside the protein. Hence, for reasons discussed above, the strong correlation is unexpected.

Chothia (1976) classified residues as buried when 95% or more of their total surface area is occluded from solvent access; remaining residues are then classified as accessible. Similarly, Janin (1979) classified residues with less than 20 \AA^2 of accessible surface as buried, with remaining residues classified as accessible.

These are stringent criteria. Consider, for example, the phenolic —OH group in Tyr: in a linear tripeptide, this group accounts for 53/218 of that residue's accessible surface. If a Tyr residue were to be almost completely buried, with only its —OH group accessible, then either criterion would nevertheless classify the residue as accessible. Moreover, this situation would be indistinguishable from one in which the entire Tyr ring were solvent exposed. The use of stringent criteria for classifying a residue as buried results in systematic overestimation of the extent to which the residues are actually exposed. This tendency will be most pronounced in the case of the amphipathic side chains where the most favorable state for the side chain is one of partial burial.

In short, the hydration potential favors the aqueous phase in the case of the amphipathic residues because there is no compensation for broken hydrogen bonds upon transfer to the vapor phase. The empirical measures of Chothia and of Janin favor the solvent-accessible state for these same residues because their criteria for burial are stringent. When these scales are compared, the reasons leading to the observed correlation are unrelated. It remains an open question whether a continuous empirical measure of accessibility would exhibit closer agreement with ΔG_t .

3. Prediction of Turns: Hydrophobicity Profiles

In contrast to empirical methods, physicochemical approaches to prediction seek a correlation between observed turn sites and measurable physical parameters such as residue polarity, volume, or hydrophobicity. No data base of known structures is required. Instead, a table measuring

parameters of interest is needed for each of the 20 residues (or for as many as possible).

In a folding protein, the most hydrophobic chain segments are expected to have the greatest tendency to partition into the solvent-shielded interior. This effect is not governed solely by the partitioning tendency of individual side chains, since the residues are not independent of each other. Rather, it is the aggregate hydrophobicity over a stretch of covalently connected residues that will determine these partitioning tendencies.

It has been hypothesized that turns correspond to loci where the average hydrophobicity along the linear chain is at a local minimum (Rose, 1978). According to this hypothesis, complementary chain sites at local maxima in hydrophobicity interact with each other giving rise to the buried interior of the molecule (Rose and Roy, 1980). In this interaction, intervening sites of lesser hydrophobicity are disposed to the solvent-accessible surface, where they are often expressed as chain turns.

The quantitative relationship between turns and hydrophobicity has been explored recently by means of *hydrophobicity profiles* that graph the average hydrophobicity per residue against the sequence number (Rose, 1978; Both and Sleigh, 1980; Rose and Roy, 1980; Hopp and Woods, 1981; Kyte and Doolittle, 1982). The plotted curve reveals the local minima and maxima in hydrophobicity along the linear polypeptide chain. Typical profiles are shown in Fig. 31.

Both aspects of the above hypothesis have been tested utilizing hydrophobicity profiles. Local minima were shown to be in good agreement with turn sites, although some of the less-pronounced minima correspond instead to other surface regions (Rose, 1978; Kyte and Doolittle, 1982), particularly to solvent-accessible parts of helices. Similarly, local maxima were observed to coincide with chain segments that are buried within the tertiary fold (Rose and Roy, 1980; Kyte and Doolittle, 1982).

Some local minima in hydrophobicity occur within stretches of the polypeptide chain containing hydrophobic residues, consistent with Richards' observation that "some of the grease is on the surface" (Richards, 1977). Given the observed organization of protein molecules into turns and the chain segments between them, shielding of hydrophobic surface from solvent water appears to be achieved by burying sites of hydrophobic maxima within and between interacting segments (Chothia, 1976), despite the unavoidable solvation of some remaining sites of lesser hydrophobicity.

The hydrophobicity profile for an amino acid sequence is simple to construct, with only two significant sources of variation: (1) the choice of a hydrophobicity scale, and (2) the degree of averaging. The issue of

hydrophobicity scales was examined at length in Section III,C,2 and summarized in Table VI. A table of hydrophobicities for each of the 20 residue types is established using the chosen scale. Averaging is then achieved by evaluating the mean hydrophobicity within a moving window that is stepped along the sequence from each residue to the next. For good resolution, the size of the averaging window must be larger than a single turn (i.e., greater than 4 residues) but smaller than a complex segment, such as a turn-helix (i.e., less than approximately 12 residues). While smaller windows yield profiles with greater detail, most of the minor fluctuations in hydrophobicity appear to have little relation to overall segmentation of the molecule.

Rose (1978) used a hydrophobicity profile to predict turns, while Rose and Roy (1980) used the method to predict interior and exterior regions in the molecule. Both and Sleigh (1980) compared the hydrophobicity of hemagglutinins from different viral strains in order to identify antigenic sites. Hopp and Woods (1981) also emphasized use of the approach to predict antigenic determinants, referring to their running average as a hydrophilicity value. Kyte and Doolittle (1982) predicted molecular segmentation into interior and exterior regions for both soluble and membrane-bound proteins. Their hydropathy index was developed after consideration of several scales.

When interpreting a hydrophobicity profile, it is necessary to differentiate the major peaks, which correspond to structural features being predicted, from the minor fluctuations, which are to be ignored. An acceptable way to do this is by visual inspection. Rose (1978) went further and smoothed the discrete values in average hydrophobicity by fitting them to a quadratic polynomial, in imitation of the effect achieved using a French curve. Since Hopp and Woods (1981) were primarily interested in the few largest peaks, their results are not perturbed by minor fluctuations in other parts of the graph. Kyte and Doolittle (1982) used larger windows, reducing the inclination toward additional smoothing.

Some hydrophobicity scales lack values for all 20 residue types. Although the scale of Nozaki and Tanford (1971) covers only 11 of the residues, all of the hydrophobic amino acids are included except isoleucine. Both and Sleigh (1980) used this scale directly. Rose (1978) and Rose and Roy (1980) assigned isoleucine a ΔG_t value equal to leucine and values of zero to all hydrophilic residues. While this strategy does not affect the positions of local minima in hydrophobicity, it does prevent these minima from assuming quantitative values, as pointed out by other workers (Hopp and Woods, 1981; Kyte and Doolittle, 1982). Both Hopp

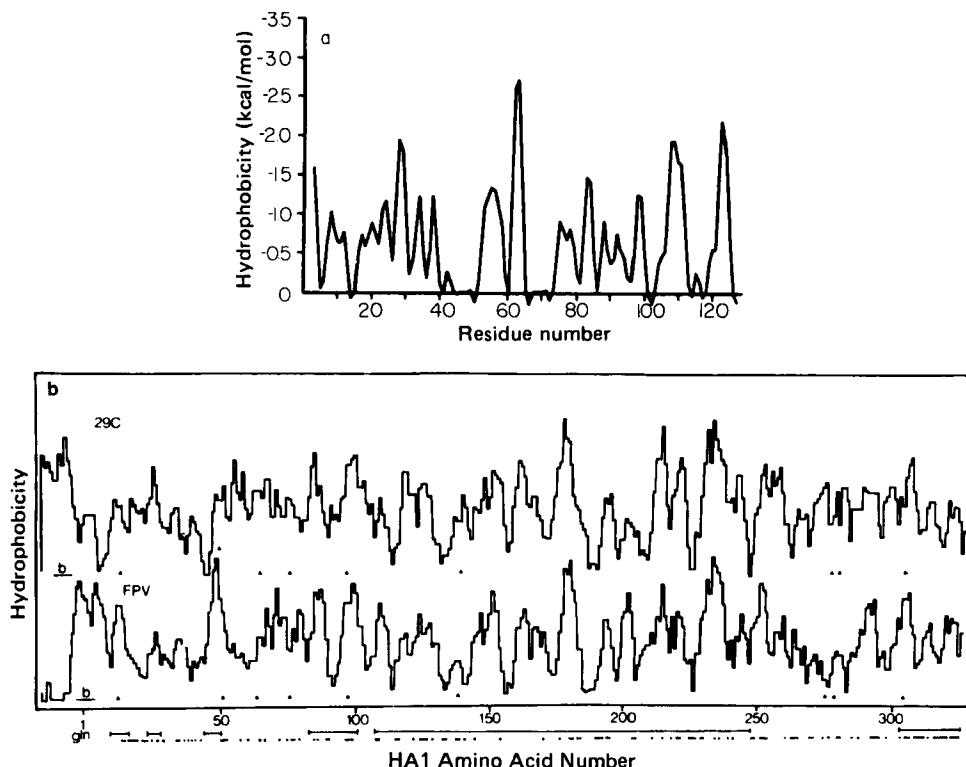


FIG. 31. Hydrophobicity profiles taken from the literature. (a) A smoothed hydrophobicity profile for lysozyme using a 5-residue moving average and the scale of Nozaki and Tanford (1971). Local maxima in hydrophobicity predict chain regions that are buried within the protein, while local minima predict exterior regions, turns in particular. Reproduced Rose (1978) with permission of Macmillan Journals Ltd. (b) Hydrophobicity profiles for hemagglutinin, Hong Kong influenza strain 29C, and avian influenza strain FPV, using a 5-residue moving average and the scale of Nozaki and Tanford (1971). Regions of similar hydrophobicity profile are indicated by solid lines below the figures, with homologous amino acids indicated by a dot. The solid triangles mark Cys residues; b is a base line for each curve. Courtesy of G. W. Both (Both and Sleigh, 1980) and reproduced with permission of IRL Press Ltd. (London). (c) A hydrophilicity profile for sperm whale myoglobin using a 6-residue moving average and the scale of Levitt (1976). The largest local maximum in hydrophilicity corresponds to a predicted antigenic determinant. Solid bars mark known antigenic determinants; (—), profile obtained using the Levitt scale, (---), profile obtained by raising the Levitt values for aspartic and glutamic acid to 3.0; (· · ·), profile obtained by assigning proline a scale value of 0. Courtesy of Thomas P. Hopp (Hopp and Woods, 1981) and reproduced with permission. (d) A hydropathy profile (i.e., both hydrophobicity and hydrophilicity) for dogfish lactate dehydrogenase using a 9-residue moving average and the scale of Wolfenden *et al.* (1981) with modification. Local maxima in hydrophobicity predict buried interior regions of the protein while local minima in hydrophobicity predict exterior regions. Solid bars above the graph mark known interior regions; solid bars below mark known exterior regions. Courtesy of Russell F. Doolittle (Kyte and Doolittle, 1982) and reproduced with permission of Academic Press (London) Ltd.

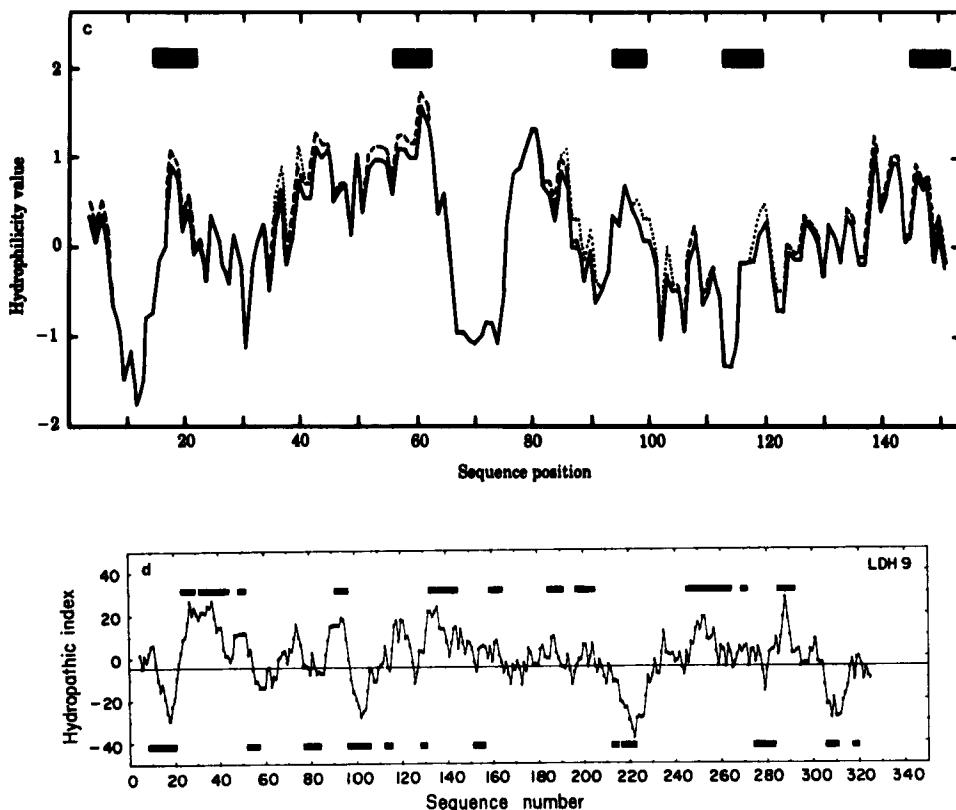


FIG. 31. (Continued)

and Woods (1981) and Kyte and Doolittle (1982) employed complete scales, resulting in quantitative values for all peaks. Hopp and Woods (1981) used the Levitt (1976) scale, with adjustment of the values of acidic residues and proline to maximize predictive success. Kyte and Doolittle (1982) supplemented and modified the Wolfenden scale based on chemical argument.

A simple Fortran program to calculate a hydrophobicity profile is included in the Appendix. Since a variety of hydrophobicity scales and window sizes have been used, these are left as program parameters. Several scales are furnished for convenience. The program is written in Vax-11 Fortran but should be readily adaptable to other machines. Similar programs (but not written in Fortran) have been published by Kyte and Doolittle (1982) and by Hopp and Woods (1983).

D. Larger Loops

Examination of proteins of known structure reveals the frequent occurrence of loops. Loops are contiguous chain segments in which the peptide backbone follows a loop-shaped path through space, with only a small end-to-end distance separating the segment termini. Richardson (personal communication) and Leszczynski *et al.* (unpublished results) studied such structures and found them to be both solvent accessible and compact.

Loops present themselves as overgrown turns, having few if any backbone : backbone hydrogen bonds. The observed compactness is due to loop side chains packed within the loop itself, as illustrated in Fig. 32. Upon analyzing 67 X-ray elucidated proteins, it was found that 57 have one or more compact surface loops. These moieties frequently play key functional roles in their respective proteins.

E. Turns and Protein Folding

Many protein-folding mechanisms have been proposed in recent years. Some of these do not allude explicitly to turns, but those that do ascribe either of two roles to these sites. At one extreme, turns are

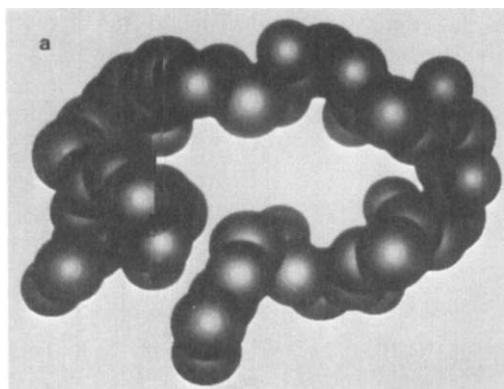


FIG. 32. Compact loops from cytochrome *c*, residues 40–54, and subtilisin, residues 74–82. Loops are contiguous chain segments in which the peptide backbone follows a loop-shaped course through space, with only a small end-to-end distance separating segment termini. The observed compactness, comparable to that of β sheet, results from loop side chains packed within the loop's own interior, as seen here. In the illustration, backbone atoms are black and side-chain atoms are gray. Actual CPK values (Sober, 1970) were used for the atomic radii, with no compensation for hydrogen atoms. (a) Cytochrome *c*, residues 40–54, backbone atoms only. (b) Cytochrome *c*, residues 40–54, backbone and side-chain atoms. (c) Subtilisin, residues 74–82, backbone atoms only. (d) Subtilisin, residues 74–82, backbone and side-chain atoms.

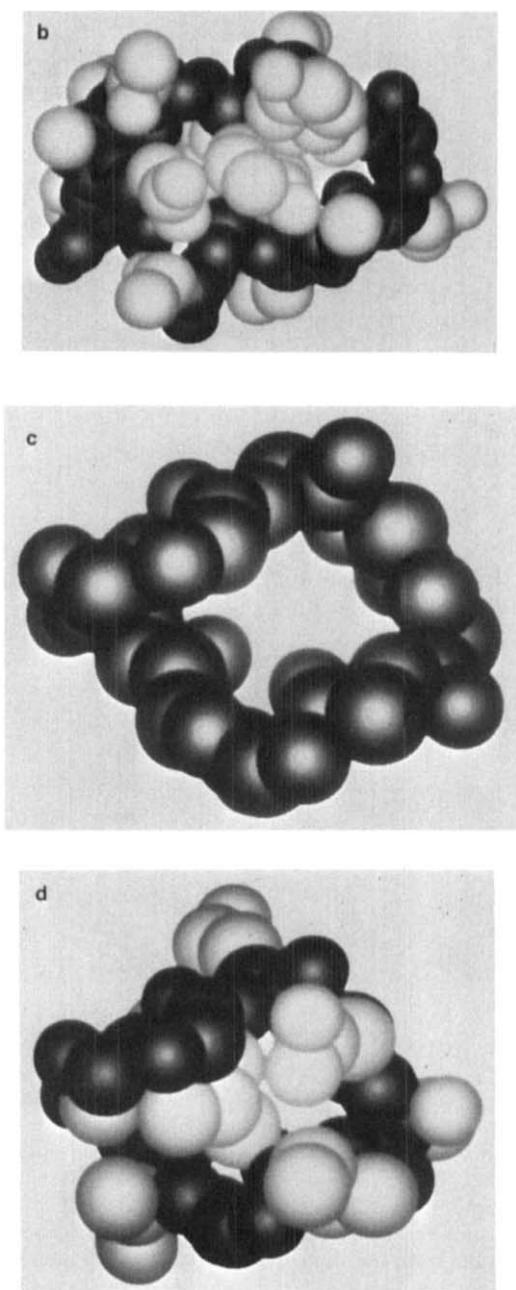


FIG. 32. (*Continued*)

portrayed as nucleating sites that are formed early in the folding process and direct further folding events (Lewis *et al.*, 1971; Zimmerman and Scheraga, 1977b). At the other extreme, turns are depicted as essentially passive structures that arise in consequence of formative interactions in other parts of the chain (Karplus and Weaver, 1976; Levitt and Chothia, 1976; Rose *et al.*, 1976; Kanehisa and Tsong, 1978; Matheson and Scheraga, 1978; Lesk and Rose, 1981; Bierzynski *et al.*, 1982). In fact, such extremes may not represent truly distinct alternatives. For example, during helix formation, conformational arrangement of both helical residues and the nonhelical residues that bracket them may be corequisite events.

It is plausible that turn conformation is closely determined by the local sequence, more so than in the case of either helix or sheet. Helices and strands of β sheet tend to run back and forth across protein molecules, linked by intervening turns (Levitt and Chothia, 1976; Rose and Seltzer, 1977). In a folding protein, a chain segment destined to be helical will be in rapid equilibrium between multiple conformational states. The linear sequence is known to be an important determinant of this conformational equilibrium (Chou and Fasman, 1978), but it is probable that segments distant in sequence yet close in space will codetermine the position of equilibrium. Thus, a linear sequence that favors a helix might eventually be stabilized in some other geometry by its spatial neighbors. However, regardless of the ultimate conformation adopted by folding segments, the interconnections formed between them will be turns. For this reason, turns may be less sensitive to the conformational constraints of spatial neighbors, and their positions better predicted from the local sequence alone.

Turns implicitly demarcate the chain segments between them. Partitioning of the chain into segments and turns can be predicted from the sequence and found in the native structure, as discussed in earlier sections. The success of such predictions suggests a folding mechanism wherein the protein's native pattern of segmentation emerges spontaneously during early stages in folding and largely persists. Later folding steps would then freeze the angles between segments as well as the individual segment geometry.

A folded protein is a dynamic structure. Particular mobility is expected at turn sites in that turns tend to lie at the molecular surface where their motion is less restricted by packing interactions. Evidence for the comparative mobility of these sites is seen in structural fluctuations taken from both X-ray temperature factors and molecular dynamics simulations (Northrup *et al.*, 1981), as shown in Fig. 33. Pronounced peaks in simulated values that are also peaks in experimental values (at residues 25, 44, 72, and 87) occur at residues within turns.

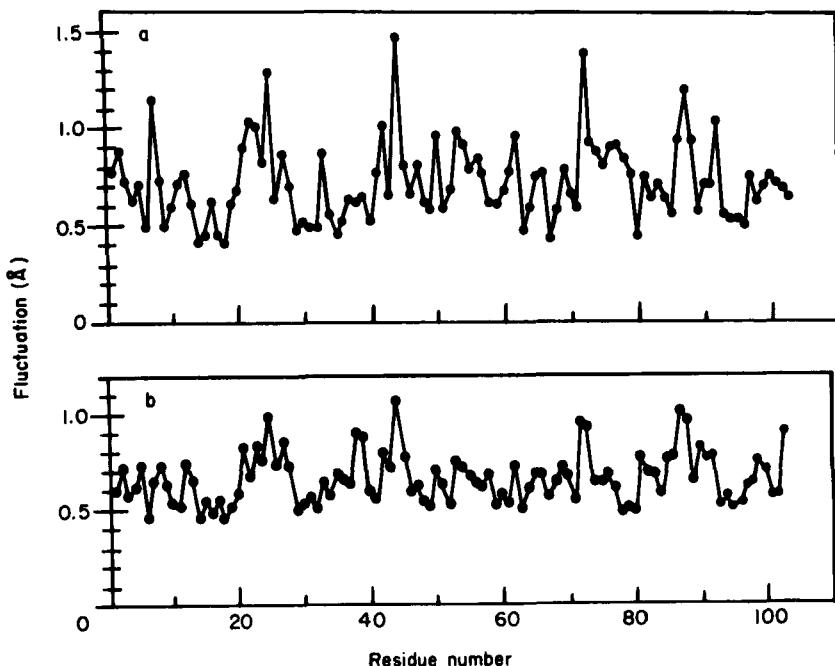


FIG. 33. Structural fluctuations in ferrocytochrome *c*, comparison of simulated and experimental values. The root-mean-square fluctuation of atoms about their mean positions is averaged for each residue. Residue averages are then plotted against the residue number for (a) a molecular dynamics simulation and (b) temperature factors from X-ray results. Courtesy of Martin Karplus (Northrup *et al.*, 1981) and reproduced with permission of Academic Press (London) Ltd.

Protons in the external turns of trypsin can be classified as labile in the neutron diffraction–hydrogen exchange experiments of Kossiakoff (1982), as illustrated in Fig. 34.

F. Possible Roles of Turns as Recognition Sites in Proteins

1. Introductory Comments: Reprise on Surface Localization

The surface localization of turns in proteins and the predominance in turns of amino acids bearing potentially reactive functional groups in their side chains (e.g., Asn, Ser, Pro, Thr, and Lys) have led to the suggestion that turns function as recognition sites for complex immunological, metabolic, genomic, and endocrinologic regulatory mechanisms.

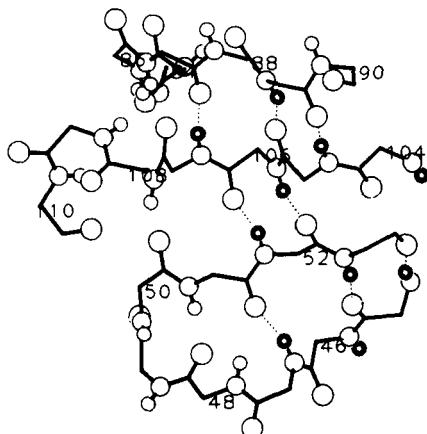


FIG. 34. Turn sites appear highly mobile when probed by the neutron diffraction-hydrogen exchange method of Kossiakoff (1982). In this experiment, crystals of trypsin are soaked in D₂O for 1 year, resulting in exchange of most of the amide hydrogens, including those in external turns. A case in point is the β turn at residues 48–51 between strands of β sheet. In the figure, filled circles denote unexchanged protons; open circles denote exchanged protons. This turn is seen to be fully exchanged, although the adjacent residues within the sheet are protected. Courtesy of A. A. Kossiakoff.

Although sites of antigenic recognition, phosphorylation, glycosylation, hydroxylation, and intron/exon splicing are frequently within turns, it is unclear whether this results from conformational recognition or propitious surface localization [Smith and Pease (Giersch), 1980]. Each of these processes and the possible role of turns will be discussed.

2. Immunological Recognition of Antigenic Sites in Proteins

Antigenic sites in proteins are distributed over the molecular surface. Perturbation of the immune network by a foreign protein results in a polyclonal antibody production. Antibodies are directed against antigenic sites on the protein surface. Although certain turns are coincident with antigenic sites in proteins, it is uncertain whether this coincidence relates to specific recognition of turn conformations or to frequency of occurrence of turns among other surface structures monitored by immune surveillance. However, by examining the location of antigenic sites in relation to the position of turns in various well-characterized protein antigens, several definite conclusions can be made.

Atassi and Smith (1978) described two types of antigen sites: continuous and discontinuous. A continuous site consists of a sequence of resi-

dues, defined by primary structure, while a discontinuous site results when residues distant in sequence are juxtaposed in space. Although this distinction was experimentally based, a continuous site may actually represent a limited case of a discontinuous site, resulting from the fact that the primary structure constrains these functional residues to a limited region of the protein surface (D. W. Fanning, J. A. Smith, and G. D. Rose, unpublished results). Table VIII lists protein antigenic sites containing residues in turns.

The antigenic sites of six proteins of known structure have been mapped in some detail. Sperm whale myoglobin has continuous (Atassi, 1975) and discontinuous (Lando *et al.*, 1982; Berzofsky *et al.*, 1982) antigenic sites (Table VIII). All five continuous sites and two discontinuous sites coincide with turns and are illustrated in Fig. 35. Staphylococcal nuclease has two continuous antigenic sites which each contain juxtaposed, overlapped turns (Smith, 1977) (Table VIII). An examination of Table VIII reveals that most of the residues in continuous antigenic sites in sperm whale myoglobin and staphylococcal nuclease are congruent with turns, and immunological recognition of turn sequences and conformations may be important. Furthermore, Leach (1983) has demonstrated in beef myoglobin the association of a turn located between the E and F helices with an antigenic region. Discontinuous antigenic sites in hen egg-white lysozyme also have residues located in one or more turns (Atassi, 1978) (Table VIII). Continuous antigenic sites have been localized to eight regions in hen egg-white lysosyme (Arnon, 1977). Lactate dehydrogenase C₄ has at least eight continuous antigen sites (T. E. Wheat and E. Goldberg, unpublished results). However, the recognition of turns in continuous antigenic sites in hen egg-white lysozyme and LDH-C₄ is not clearly demonstrated because of the nonspecific localization of antigenic sites within protein fragments. Cytochrome c has been mapped partially (Jemmerson and Margoliash, 1979), although an association of its immunodominant regions and turns is largely uncertain.

A dominant role for turn recognition in discontinuous sites is even less likely, since few residues identified in sperm whale myoglobin and hen egg-white lysozyme are contributed by turns (Table VIII). The coincidence of discontinuous as well as continuous antigenic sites and turns is probably a consequence of the ubiquity of the turn shell (see Section III,B,1).

Interestingly, 11 of the 12 continuous and discontinuous antigenic sites in sperm whale myoglobin, hen egg-white lysozyme, and staphylococcal nuclease contain one or more lysyl residues (Leach, 1983). This distribution within turns may facilitate surface exposure of these basic residues and enhance antigenicity. Furthermore, an association between

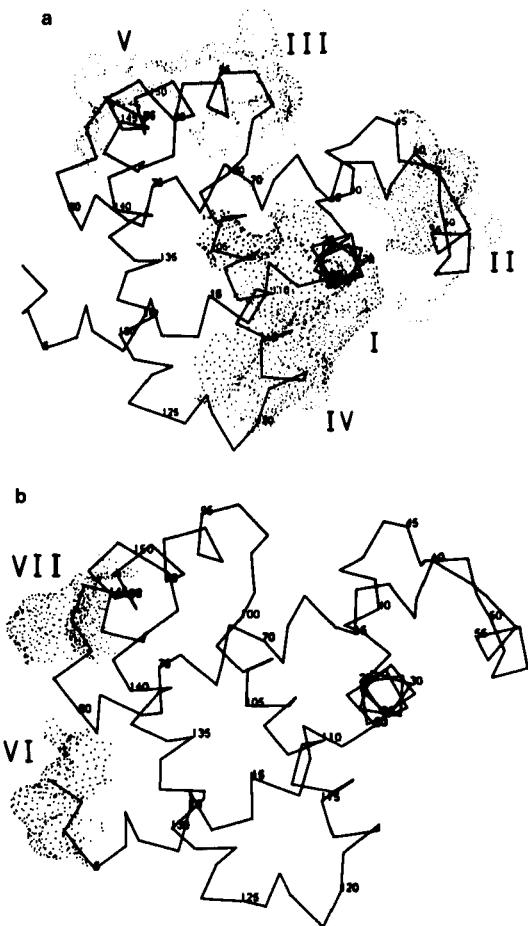


FIG. 35. Antigenic sites of sperm whale myoglobin. The (a) continuous (Atassi, 1975) and (b) discontinuous sites (Berzofsky *et al.*, 1982) are illustrated. The continuous line is the α -carbon backbone, labeled every 5 amino acid residues. The stippled area is the solvent-accessible surface about all atoms, although these are not shown, in the antigenic sites. The continuous sites are labeled I–V: I (residues 16–21); II (residues 56–62); III (residues 94–99); IV (residues 113–119); and V (residues 145–151). The discontinuous sites are labeled VI and VII: VI (residues 4 and 79) and VII (residues 83, 144, and 145).

antigenic sites and protein regions containing two or three juxtaposed, overlapped turns is seen in several of the examples in Table VIII. Such juxtaposed, overlapped turns may form compact loops, accentuate the relief of the surface topography, and enhance antibody–antigen interaction. However, not all such regions will be antigenic; three turns in hen

TABLE VIII
Protein Antigenic Sites Containing Residues Located in Turns

Protein	Type of antigenic site	Sequence and location of antigenic site ^a	Turn(s) ^b	Reference
Sperm whale myoglobin	Continuous ^c	Lys ¹⁶ -Val-Gly-Ala-Asp-Val ²¹	Glu ¹⁸ -Ala-Asp-Val ²¹	Atassi (1975)
	Continuous	Lys ⁵⁶ -Ala-Ser-Glu-Asp-Leu-Lys ⁶²	Region of residue 59	Atassi (1975)
	Continuous	Ala ⁹⁴ -Thr-Lys-His-Lys-Ile ⁹⁹	Thr ⁹⁵ -Lys-His-Lys ⁹⁸	Atassi (1975)
	Continuous	His ¹¹³ -Val-Leu-His-Ser-Arg-His ¹¹⁹	Region of residue 118	Atassi (1975)
	Continuous	Lys ¹⁴⁵ -Tyr-Lys-Glu-Leu-Gly-Tyr ¹⁵¹	Region of residue 149	Atassi (1975)
	Discontinuous ^c	Glu ⁴ . . . Lys ⁷⁹	Leu ⁷⁶ -Lys-Lys-Lys ⁷⁹ Lys ⁷⁸ -Lys-Gly-His ⁸¹	Berzofsky <i>et al.</i> (1982)
Staphylococcal nuclease	Discontinuous	Glu ⁸³ . . . Ala ¹⁴⁴ . . . Lys ¹⁴⁵	Region of residue 83	Berzofsky <i>et al.</i> (1982)
	Continuous	Lys ²⁴ -Leu-Met-Tyr-Lys-Gly-Gln-Pro ³¹	Region of residue 144 Met ²⁶ -Tyr-Lys-Gly ²⁹ Tyr ²⁷ -Lys-Gly-Gln ³⁰	Smith (1977)
	Continuous	His ⁴⁶ -Pro-Lys-Lys-Gly-Val-Glu ⁵²	Pro ⁴⁷ -Lys-Lys-Gly ⁵⁰ Lys ⁴⁹ -Gly-Val-Glu ⁵² Glu ⁵² -Lys-Tyr-Gly ⁵⁵	Smith (1977)

Hen egg-white lysozyme	Discontinuous	(Cys) ⁶ —Gly ⁷ ··· Ala ¹⁰ —Lys ¹³ —Arg ¹⁴ and Gly ¹²⁶ ··· Arg ¹²⁸	Region of residue 5 Region of residue 15 Ile ¹²⁴ —Arg—Gly—Cys ¹²⁷ Asn ⁵⁹ —Ser—Arg—Trp ⁶² Ser ⁶⁰ —Arg—Trp—Trp ⁶³ Region of residue 67 Region of residue 86	Atassi (1978)
	Discontinuous	Trp ⁶² and Leu ⁶⁵ and Asp ⁸⁷ ··· Thr ⁸⁹ and Asn ⁹³ ··· Lys ⁹⁶ —Lys ⁹⁷	Leu ¹⁷ —Asp—Asn—Tyr ²⁰ Asn ¹⁹ —Tyr—Arg—Gly ²² Tyr ²⁰ —Arg—Gly—Tyr ²³ Cys ¹¹⁵ —Lys—Gly—Thr ¹¹⁸	Atassi (1978)
	Discontinuous	Tyr ²⁰ —Arg ²¹ ··· Tyr ²³ and Lys ³³ —Phe ³⁴ and Asn ¹¹³ —Arg ¹¹⁴ ··· Lys ¹¹⁶	Atassi (1978)	

^a The notation (· · ·) indicates that the intervening residues are not involved in the delineated antigenic site. () indicates that the involvement of this residue in the antigenic site is uncertain. The superscripted numbers refer to the residue number.

^b Explicit four-residue turns are β turns identified from X-ray coordinates by Chou and Fasman (1977). The turns designated by region of a residue are located using the algorithm of Rose and Seltzer (1977).

^c The term continuous was originally introduced by Atassi (1975) and rigorously defined by Atassi and Smith (1978). The term discontinuous was originally introduced by Atassi (1978) and rigorously defined by Atassi and Smith (1978).

egg-white lysozyme, i.e., (103–106), (104–107), and (105–108), are overlapping adjacent turns without detectable antigenicity (Atassi, 1978).

Arnon *et al.* (1974) have examined the relationship between antigenicity and conformation in a lysozyme fragment formed by a disulfide bond between residues 64 and 80. This fragment contains no complete native antigenic sites. Arnon *et al.* (1974) have correlated changes in the binding capacity of synthetic loop analogs, having various residues replaced by alanine, to anti-loop antibodies with the relative probability of occurrence of β turns. By substituting one amino acid residue at a time, they were able to establish a relationship between differences in antigenicity and differences in the theoretically predicted probability for β -turn formation. A different experimental approach using immunological measurements of the degree of cross-reaction between lysozymes from different species in which only amino acid substitutions occurred (i.e., no additional inversions, additions, or deletions) found a correlation with the probability of turn formation (Fainaru *et al.*, 1974). Since no direct experimental measurements of the changes in turn conformation were made in these two studies and since the antigenicity of any antigenic site depends on its sequence as well as its conformation, these correlations must be viewed with caution.

3. Immunological Recognition of Viruses

Viruses are enveloped by lipo- and glycoproteins, and viral surfaces are a composite of the exposed surfaces of these molecules. Such molecular surfaces mediate the viruses' interaction with host-cell receptors and subsequent fusion with intracytoplasmic membranes after cellular penetration. Active immunization with isolated viral surface proteins elicits the development of cross-reactive and neutralizing antiviral antibodies of therapeutic value. In contrast, immunization with carrier-linked synthetic peptides mimicking exposed regions of a viral protein may or may not elicit cross-reactive antibodies and rarely induces protective immunity (Table IX) (M. Sela, personal communication). Because the X-ray crystallographic details of most of these viral proteins are unknown, the relationship between viral protein turns and immunogenic sites is still uncertain. However, the data in Table IX are included as a starting point for those seeking such a correlation.

Anderer (1963a,b) was the first to demonstrate that a peptide could inhibit the reaction between a virus and specific antiviral antibodies. From an enzymatic digestion of the tobacco mosaic virus (TMV) coat protein, a hexapeptide, Thr–Ser–Gly–Pro–Ala–Thr (the underlined residues frequently occur in turns), was isolated that could partially in-

hibit the reaction between "native" TMV and anti-TMV antibodies. Another synthetic peptide corresponding to the residues 89–108 (Glu–Leu–Thr–Ile–Pro–Ile–Phe–Ala–Thr–Asn–Ser–Asp–Cys–Glu–Leu–Ile–Val–Lys–Ala–Met) of the coat protein of coliphage MS-2 induces antibody-mediated bacteriophage neutralization (Langbeheim *et al.*, 1976). Immunization of guinea pigs and rabbits with two different synthetic peptides corresponding to regions in the carboxyl-terminal half of the foot-and-mouth disease virus (FMDV) VP1 polypeptide leads to the induction of serotype-specific neutralizing antibodies (Bittle *et al.*, 1982; Pfaff *et al.*, 1982). These anti-FMDV peptide antibodies are directed against region (141–160) (Val–Pro–Asn–Leu–Arg–Gly–Asp–Leu–Glu–Val–Leu–Ala–Gln–Lys–Val–Ala–Arg–Thr–Leu–Pro) and region (200–213) (Arg–His–Lys–Glu–Lys–Ile–Val–Ala–Pro–Val–Lys–Glu–Thr–Leu). Although there is no definite experimental evidence that turns occur in the synthetic peptides or the corresponding regions in the viral coat protein, except influenza (Muller *et al.*, 1982), these immunogenic peptides frequently contain prolyl residues and clusters of other turn-associated amino acids.

However, there is increasing evidence that immune recognition of other viruses focuses, in part, on immunogenic surface regions frequently coincident with turns. Analysis of the amino acid and DNA sequences predicts conformations with a preponderance of β structure alternating with surface-exposed β turns for the hepatitis B surface antigen polypeptide (P25) (Dreesman *et al.*, 1982), Rauscher murine leukemia virus polypeptide (p10) (Henderson *et al.*, 1981), adenovirus "spike" protein (Gingeras *et al.*, 1983), and herpes simplex virus type 1 glycoprotein D (Watson *et al.*, 1982). Some of these turns are sites of posttranslational modification, which may or may not preclude immune recognition of these regions.

Protein sequences derived from DNA sequence analysis and hydrophobicity/hydrophilicity algorithms were used to predict the locations of various protein immunogenic sites in hepatitis B virus (Lerner *et al.*, 1981; Hopp and Woods, 1981; Hopp, 1981). In each case the immunogenic surface regions were hydrophilic and contained prolyl residues. Dreesman *et al.* (1982) chose a different hydrophilic surface region (117–137) containing three predicted turns [i.e., (118–121), (129–132), and (134–137)] and synthesized a corresponding peptide with a disulfide bond between residues 124 and 137. Since cross-reactive antibodies specific for hepatitis B surface antigen developed in mice immunized with the synthetic peptide, the antigenicity of this synthetic molecule mimics that of the intact viral protein. Later, Gerin *et al.* (1983) examined the same protein region. At present, it is unclear whether the antivi-

TABLE IX
Immune Response to Synthetic Viral Peptides

Virus	Peptide	Cross reactive	Protective immunity	Reference
MS-2 coat protein	89-108(P ₂) 109-129(P ₃)	+	+	Langbeheim <i>et al.</i> (1976)
Moloney leukemia virus/R protein	C terminus (15 residues)	+	Not tested	Sutcliffe <i>et al.</i> (1980)
SV40 T antigen	N terminus (8 residues)	+	Not tested	Walter <i>et al.</i> (1980)
	C terminus (11 residues)	+	Not tested	
t antigen	C terminus (6 residues)	+	Not tested	Harvey <i>et al.</i> (1982)
Polyoma-virus middle T antigen	C terminus (6 residues)	+	Not tested	Walter <i>et al.</i> (1981)
Rous sarcoma virus pp60 ^{rr}	415-424	+	Not tested	Wong and Goldberg (1981)
FMDV-VP1	498-512	+	Not tested	Rohrscheider (quoted by Walter and Doolittle, 1983)
	1-41	-	-	Bittle <i>et al.</i> (1982)
	9-24	-	-	Bittle <i>et al.</i> (1982)
	17-32	-	-	Bittle <i>et al.</i> (1982)
	25-41	-	-	Bittle <i>et al.</i> (1982)
	141-160	+	+	Bittle <i>et al.</i> (1982)
	144-159	+	+	Pfaff <i>et al.</i> (1982)
	151-160	+	Partial	Bittle <i>et al.</i> (1982)
	200-213	+	+	Bittle <i>et al.</i> (1982)
Influenza virus HA1	20 peptides	+ (18 peptides)	-	Green <i>et al.</i> (1982)
	91-108	+	Partial	Muller <i>et al.</i> (1982)
	123-151	-	-	Jackson <i>et al.</i> (1982)
	138-164	+	Partial	Shapira <i>et al.</i> (1985)
	139-146(Gly ¹⁴⁴)	-	-	Shapira <i>et al.</i> (1985)
	139-146(Asp ¹⁴⁴)	-	-	Shapira <i>et al.</i> (1985)
	147-164	+	-	Shapira <i>et al.</i> (1985)
HA2	1-11 (fusion peptide)	+	-	Atassi and Webster (1983)
Poliovirus P63	18-30	-	Not tested	Baron and Baltimore (1982a)
	450-461	+	Not tested	Baron and Baltimore (1982a)
VP _g	1-22	+	Not tested	Baron and Baltimore (1982b)
	9-22	+	Not tested	Baron and Baltimore (1982b)

TABLE IX (Continued)

Virus	Peptide	Cross reactive	Protective immunity	Reference
VP1	11-17	-	-	Emini <i>et al.</i> (1983)
	70-75	-	-	Emini <i>et al.</i> (1983)
	70-80	-	-	Emini <i>et al.</i> (1983)
	93-103	+	+	Emini <i>et al.</i> (1983)
	97-103	-	-	Emini <i>et al.</i> (1983)
Adenovirus (human) type	Several peptides	+	Not tested	Branton (quoted by Walter and Doolittle, 1983)
	N terminus (2 peptides)	+	Not tested	Green <i>et al.</i> (1983)
	C terminus (3 peptides)	+	Not tested	Green <i>et al.</i> (1983)
Hepatitis B surface antigen	2-16	+	-	Lerner <i>et al.</i> (1981)
	12-16	-	-	Lerner <i>et al.</i> (1981)
	22-35	+	-	Lerner <i>et al.</i> (1981)
	31-35	-	-	Lerner <i>et al.</i> (1981)
	38-52	-	-	Lerner <i>et al.</i> (1981)
	47-52	-	-	Lerner <i>et al.</i> (1981)
	48-81	+	-	Gerin <i>et al.</i> (1983)
	95-109	+	-	Lerner <i>et al.</i> (1981)
	104-109	-	-	Lerner <i>et al.</i> (1981)
	110-137	+	Partial	Gerin <i>et al.</i> (1983)
	117-137	+	-	Dressman <i>et al.</i> (1982)
	122-137	+	-	Dressman <i>et al.</i> (1982)
	122-137	-	-	Bhatnager <i>et al.</i> (1982)
	125-137	+	-	Gerin <i>et al.</i> (1983)
	128-134	-	-	Bhatnager <i>et al.</i> (1982)
	134-146	+	-	Vyas <i>et al.</i> (1980)
	135-155	+	-	Neurath <i>et al.</i> (1982)
	138-149	+	-	Prince <i>et al.</i> (1982)
	139-147	+	-	Bhatnager <i>et al.</i> (1982)
	139-158	+	-	Bhatnager <i>et al.</i> (1982)
	140-158	+	-	Bhatnager <i>et al.</i> (1982)
	145-158	-	-	Bhatnager <i>et al.</i> (1982)
	150-158	-	-	Bhatnager <i>et al.</i> (1982)
	212-226	-	-	Lerner <i>et al.</i> (1981)
	221-226	-	-	Lerner <i>et al.</i> (1981)

ral antibodies elicited by these synthetic immunogens are specific for continuous or discontinuous antigenic sites. However, the frequency of occurrence of turns with or without prolyl residues and the propensity of these turns for surface distribution herald their importance as immunogenic sites for use in future antiviral immunotherapy. However, the elicitation of protective immunity by peptide immunogens is rare, and the factors necessary to enhance the production of neutralizing antibodies are largely unclear.

4. Recognition of Phosphorylation Sites

Protein kinase-dependent phosphorylation of proteins is one of the mechanisms leading to posttranslational modification (Uy and Wold, 1977) of protein structure (Huijing and Lee, 1973). Intracellular protein kinases are known to transfer a phosphate from adenosine triphosphate (ATP) to specific seryl, threonyl, and tyrosyl residues located in phosphorylation sites in membrane-bound sugar-transport proteins, membrane receptors, histones and acidic nuclear proteins, ribosomal proteins, and mitochondrial proteins. Aberrant membrane protein phosphorylation is associated with diseases, including muscular dystrophies (Roses and Appel, 1975; Roses *et al.*, 1976), hereditary stomatocytosis and spherocytosis (Beutler *et al.*, 1976; Greenquist and Shohet, 1976; Wyatt *et al.*, 1977), and sickle cell anemia (Beutler *et al.*, 1976; Hosey and Tao, 1976). However, the nature of the resulting membrane alterations is unknown.

Williams (1976) suggested that regions including basic amino acids adjacent to the phosphorylated hydroxyamino acid residues may function as the recognition site for the phosphorylase kinase. His analysis of the sequences juxtaposed in 25 phosphorylation sites in 11 proteins found that phosphorylated seryl or threonyl residues are frequently separated from an adjacent basic residue by one intervening residue. Although emphasizing that primary structure is the critical feature leading to recognition of a phosphorylation site, Williams (1976) acknowledged that phosphorylation might be regulated by the secondary or tertiary structure of a site. Small *et al.* (1978) later demonstrated that 24 of 30 seryl and threonyl phosphorylation sites in 14 proteins, including many of the sites examined by Williams (1976), are coincident with predicted turn conformations. No correlation between tyrosyl phosphorylation and turns has been made. One limitation of this study is that the turn-containing regions are defined by an empirical predictive method and are not confirmed by X-ray diffraction analysis. Among these phosphorylated residues predicted to be associated with β turns, no strong

positional preference is observed. Other phosphorylation sites are not related to turn-containing regions.

An understanding of turns and phosphorylation sites must await availability of structural data for phosphorylated proteins. Furthermore, it is unclear whether any class of protein kinase depends on turn recognition for initiation of phosphorylation.

5. Enzymatic Recognition of Glycosylation Sites in Proteins

Two different modes of glycosyl linkage between protein and carbohydrate are identified: (1) the carbohydrate is linked N-glycosidically to an asparagine residue of the protein; (2) the carbohydrate is linked O-glycosidically to a serine, threonine, hydroxylysine, or hydroxyproline residue of the protein (Neuberger *et al.*, 1972). In glycoproteins both types of glycosidic linkage may exist, and it has been suggested that glycosyltransferases recognize turn regions in either case.

a. *N*-Glycosylation. Glycoproteins containing carbohydrate linked N-glycosidically to an Asn residue invariably possess the sequence, —Asn—Xxx—[^{Thr}_{Ser}]— (Parodi and Leloir, 1979). When the primary structures of multiple glycoproteins are examined, this tripeptide sequence is found to be a necessary but not a sufficient requirement for glycosylation via the dolichol pathway (Marshall, 1974). Aubert *et al.*, (1976), realizing that asparagine, serine, and threonine occur frequently within turns, analyzed the sequences around the glycosylated residues in various glycoproteins by the empirical methods of Chou and Fasman (1978) in order to determine if turns are required within N-glycosylation sites. They found that in 19 of the 28 sites in 14 glycoproteins analyzed, the Asn residue is located within predicted turns. It has been suggested that the location of Asn within turns may provide a hallmark recognized by oligosaccharide transferase, which transfers the oligosaccharide dolichyldiphosphate to acceptor proteins. Such proposals should be interpreted with caution, however, because asparagines are strong turn formers in the predictive method used (Chou and Fasman, 1978).

The kinetics of N-glycosylation of various synthetic hexapeptides, Tyr—Asn—Gly—Xxx—Ser—Val, where Xxx is Thr, Ser, CysH, Val, or Thr(OCH₃), revealed that the best substrates have Xxx equal to serine or threonine. The authors propose that a hydrogen bond interaction between the hydroxyl oxygen of a threonine or serine residue (hydrogen bond acceptors) and the β -amide hydrogen of an asparagyl residue (hydrogen bond donor) is integral to glycosyl transfer (Bause and Legler, 1981). Space-filling models demonstrate that such hydrogen bond inter-

actions are facilitated in turns in which the asparagyl residue occupies the i or $i + 1$ position of a β turn (Bause *et al.*, 1982). Further studies with proline-containing peptide analogs substantiate this model (Bause, 1983).

Aubert *et al.* (1981, 1982) and Ronin and Aubert (1982) have proposed that a low concentration of dimethylsulfoxide ($\sim 10\%$) induces conformational features in synthetic Asn-Xxx-[$^{Thr}_{Ser}$]-containing peptides which enhance N-glycosylation of these substrates. In particular, upon addition of dimethylsulfoxide to an aqueous solution of a model heptapeptide substrate, a CD spectral change is observed which the authors attribute to adoption of turn conformation, and increased rates of N-glycosylation are noted.

Pless and Lennarz (1977) found that RNase A and α -lactalbumin, both containing the sequence —Asn-Xxx-[$^{Thr}_{Ser}$]—, are converted from non-acceptors to acceptors after reduction and alkylation under denaturing conditions. Additional fragmentation of α -lactalbumin by cyanogen bromide cleavage and tryptic or chymotryptic digestion produces polypeptides ranging in size from 7 to 123 residues, but differing little in their rates of *in vitro* glycosylation (Struck *et al.*, 1978). Seven other proteins that are not normally glycosylated, but contain one (or more) —Asn-Xxx-[$^{Thr}_{Ser}$]— sequence(s), were evaluated as acceptors for carbohydrates (Kronquist and Lennarz, 1978). After reduction and alkylation, two proteins, rabbit muscle triosephosphate isomerase and ovine prolactin, became acceptors of oligosaccharide chains. In the native triosephosphate isomerase molecule, the glycosylation site (195–197) occurs within an α -helical conformation (Levitt and Greer, 1977), and this structural feature may protect the Asn¹⁹⁵–Val¹⁹⁶–Ser¹⁹⁷ sequence from glycosylation. After cyanogen bromide cleavage, peptides from catalase and concanavalin A also become carbohydrate acceptors. Two of these glycosylation sites in concanavalin A, regions (118–120) and (162–164), occur in regions containing three juxtaposed and overlapped turns located between two β sheets (Levitt and Greer, 1977). Interestingly, the presence of these turns, even in a highly surface-exposed position characteristic of repeating turns, is insufficient to evoke glycosylation prior to CNBr cleavage.

However, the conformation of N-glycosylation sites is not invariably a turn. Wilson *et al.* (1983) have demonstrated by X-ray analysis of the Asn-linked oligosaccharide sites on the Hong Kong X31 1968 hemagglutinin that for the six sites analyzed, two were coincident with α -helical regions, two were coincident with β -sheet regions, and two were coincident with turns. The Asn residues in the influenza hemagglutinin glycosylation sites coincident with turns were located at the $i + 2$ position in

contrast to the $i + 1$ position found for the Fc portion of immunoglobulin G (Deisenhofer, 1981).

In summary, N-glycosylation is linked to the recognition of a unique receptor sequence. The evidence that turn conformational recognition is sometimes required is suggestive but still inconclusive. Several questions remain to be answered about N-glycosylation. Why are only some surface-accessible Asn-Xxx-[^{Thr}_{Ser}] sequences glycosylated, and is this related to turn conformation? Is there a relationship between the position (i.e., $i + 1$ or $i + 2$) of the Asn in glycosylation sites with turn conformations? Is there a functional difference between Ser and Thr-bearing receptor sequences? Is there a relationship between glycosylation site conformation and the type of carbohydrate linked (i.e., high-mannose versus complex carbohydrate)?

b. *O*-Glycosylation. The recognition of turn conformations by glycosyltransferases has been implicated in O-glycosylation. In what is now a familiar approach, Aubert *et al.* (1976) used the empirical methods of Chou and Fasman (1974) to predict possible turn conformations in nine glycosylation sites in five glycoproteins. They observed (1) one or more prolyl residues, and (2) one or more glycyl residues within or adjacent to their predicted turns in 89% of the glycosylation sites examined. In two-thirds of these glycosylation sites the glycosylated residue occupies a central position (i.e., the $i + 1$ or $i + 2$) within a predicted β turn.

Young *et al.* (1979) studied enzymatic O-glycosylation of synthetic peptides containing sequences from myelin basic protein (90–102). The minimum sequence glycosylated was Thr⁹⁸–Pro–Pro–Pro¹⁰¹, and upon N-terminal elongation of this peptide the level of glycosylation was not diminished. Unfortunately, the conformations of these peptides have not been examined, so the degree of conformational specificity remains ambiguous.

Experimental evidence is required to substantiate this appealing suggestion, but based on previous experience with N-glycosylation (vide supra), a simple relationship between O-glycosylation and turn conformation seems unlikely.

6. *Recognition of Connective Tissue Protein Hydroxylation Sites*

Proline hydroxylation in connective tissue protein (e.g., collagen and elastin) depends in part on the recognition of (Pro–Gly)-containing sequences, which occupy turn conformations. The preferred sequence is Xxx–Pro–Gly–Yyy, which favors a type II β turn (Venkatachalam, 1968). Brahmachari and Ananthanarayanan (1978, 1979) proposed that proline hydroxylase (EC 1.14.11.2) selectively hydroxylates prolyl resi-

dues in the $i + 1$ position of β turn-containing sequences, Xxx-Pro-Gly-Yyy. Although this proposal may be valid for nascent procollagen, at least two collagen-like molecules are exceptions: (Val-Pro-Gly-Gly)_n, which exists in a β -turn conformation (Urry and Ohnishi, 1974), is not effectively hydroxylated (Bhatnagar *et al.*, 1978), while (Pro-Pro-Gly)_n, which does not exist in a β -turn conformation (Zimmerman and Scheraga, 1977a), is optimally hydroxylated (Bhatnagar *et al.*, 1978). Although the identity of the preceding and the residue following the Pro-Gly site are posited to be important in the regulation of hydroxylation (Bhatnagar *et al.*, 1978; Brahmachari and Ananthanarayanan, 1978), the mechanisms underlying this regulation are poorly understood.

Hydroxyproline formation in collagen leads to enhanced stability of the collagen triple helix (Rosenbloom *et al.*, 1973; Berg and Prockop, 1973) due to hydrogen bonding between the hydroxyproline's hydroxyl group and the peptide backbone (Ramachandran *et al.*, 1973). Such conformational changes may affect the rate and/or site of subsequent hydroxylation.

The prolyl residues of elastin are generally located in Xxx-Pro-Gly-Yyy sequences (Urry and Long, 1976). Two of five polypeptide models of elastin developed by Urry and co-workers (Urry and Ohnishi, 1974; Urry *et al.*, 1975; Rapaka and Urry, 1978; Rapaka *et al.*, 1978), (Val-Pro-Gly-Val-Gly)_n and (Ala-Pro-Gly-Gly)_n, were found by Bhatnagar *et al.* (1978) to inhibit the hydroxylation of collagen by chick embryo proline hydroxylase. However, because there is no concomitant enhancement of the endogenous hydroxylation of either of these peptides, it is unlikely that hydroxylation sites in collagen and elastin have the same conformational features.

7. Leader/Signal Sequences

Secreted proteins are initially synthesized as larger precursors, which contain an additional amino-terminal sequence of 15–30 amino acids, called a leader (signal) sequence (for a recent review, see Silhavy *et al.*, 1983). This hydrophobic extension facilitates translocation of the peptide across the microsomal membrane and is subsequently cleaved endopeptidolytically by a membrane-bound signal peptidase (Jackson and Blobel, 1980; Date and Wickner, 1981; Mollay *et al.*, 1982; Perlman and Halvorson, 1983).

Leader (signal) sequences have some common structural features. All have a hydrophobic core region ranging in length from 9 to 20 amino acid residues. A hydrophilic region containing one or two basic residues precedes the hydrophobic core; proline or glycine residues often termi-

nate the core region. The cleavage site is carboxyl terminal to the hydrophobic core and may be separated from it by up to five amino acid residues (Inouye and Halegoua, 1980). Turns have been predicted both within the hydrophobic core (Inouye and Halegoua, 1980) and near the cleavage site (Perlman and Halvorson, 1983). Most leader (signal) sequences contain one or two helix-breaking and turn-promoting glycyl or prolyl residues in the hydrophobic region, and Inouye and Halegoua (1980) proposed that turns promoted by the glycyl or prolyl residues are essential for the function of the leader (signal) sequence.

Austen (1979) applied secondary structure-predictive methods to 21 signal sequences. Turns were predicted to coincide with the sites of signal peptidase action in 9 of 21 precursor proteins studied: prepro-parathyroid hormone, prepromellitin, prealbumin, pre- β -casein, pre-opiocortin, prelipoprotein, and three monoclonal immunoglobulin precursors. There are no predicted turns located at or adjacent to the cleavage sites in 10 other precursor proteins, although they were predicted to have α helix or β structure in these sites.

Secondary structures derived from CD spectra have been determined for three signal sequences: human preproparathyroid hormone (Rosenblatt *et al.*, 1980), M13 coat protein (Shinnar and Kaiser, 1985), and *Escherichia coli* λ -receptor protein (Briggs and Giersch, 1984). Turns were not detected in any of these isolated signal sequences.

APPENDIX: A FORTRAN PROGRAM TO CALCULATE A HYDROPHOBICITY PROFILE

```

c ASN ILE LEU GLY ARG GLU ALA LYS CYS THR c
c ASN GLU VAL ASN GLY CYS PRO ARG ILE TYR c
c ASN PRO VAL CYS GLY THR ASP GLY VAL THR c
c TYR SER ASN GLU CYS LEU LEU CYS MET GLU c
c ASN LYS GLU ARG GLN THR PRO VAL LEU ILE c
c GLN LYS SER GLY PRO CYS c
c
c The program asks for the name of the input file. c
c A corresponding output file is then created by c
c replacing the extension of the input file c
c with the extension .OUT c
c
c The only two program parameters are- c
c      a) The size of the averaging window, and c
c      b) The choice of a hydrophobicity scale. c
c The scale of choice is set at the beginning of the c
c program. Upon execution, the program requests the c
c window size. c
c
c For convenience, three scales are included with c
c the program. These are: c
c 1) Nozaki and Tanford (1971), modified by Rose (1978) c
c 2) Levitt (1976), modified by Hopp and Woods (1981) c
c 3) Wolfenden et al.,modified by Kyte and Doolittle (1982)c
c
c Output is normalized to fall within the range c
c -100 to +100, regardless of the scale used. c
c Output values are stored ten per output record. c
c Hydrophobic values are more (+) than hydrophilic c
c values. c
c
c Sample output from the sequential input file for c
c BPTI, shown above, would appear as follows (using c
c the Nozaki-Tanford scale and a window size of 7): c
c
c      56 residues. Window Size= 7 Scale=NT bpti. c
c 1. ASN ILE LEU GLY ARG GLU ALA LYS CYS THR c
c 11. ASN GLU VAL ASN GLY CYS PRO ARG ILE TYR c
c 21. ASN PRO VAL CYS GLY THR ASP GLY VAL THR c
c 31. TYR SER ASN GLU CYS LEU LEU CYS MET GLU c
c 41. ASN LYS GLU ARG GLN THR PRO VAL LEU ILE c
c 51. GLN LYS SER GLY PRO CYS c
c
c 1. *** *** *** -66 -66 -67 -78 -78 -78 -66 c
c 11. -70 -70 -70 -73 -73 -58 -52 -52 -52 -53 c
c 21. -39 -39 -51 -70 -70 -57 -66 -61 -61 -65 c
c 31. -65 -51 -48 -37 -42 -31 -31 -31 -45 -60 c
c 41. -75 -89 -97 -97 -84 -69 -54 -54 -54 -57 c
c 51. -57 -70 -71 *** *** ***

```

```

c      An optional code segment that creates output      c
c      suitable for plotting on the line printer is      c
c      included. Output is directed to a file with      c
c      extension .PLT      c
c
cccccccccccccccccccccccccccccccccccccccccccccccccccc
c
character inname*20, pname*20, buff*80, period*1, ans*1, label*4
character*3 aacids(1000), names(2,22), resnam, blanks
character*1 plot(21,100), tabs(4,100), digits(10), jdig
character*2 grpnam .
integer half, values(1000), indx(22), resnum, NT, HW, KD, group
real vals(1000), scales(3,20), scale(20), min, max
data period/'.', blanks/'  /
data aacids/1000*'  /
data NT/1/, HW/2/, KD/3/
data (names(1,j),j=1,22) /'ALA','ARG','ASN','ASP','CYS','GLN','GLU',
* 'GLY','HIS','ILU','LEU','LYS','MET','PHE','PRO','SER',
* 'THR','TRP','TYR','VAL','TRY','ILE'/
data (names(2,j),j=1,22) /'ala','arg','asn','asp','cys','gln','glu',
* 'gly','his','ilu','leu','lys','met','phe','pro','ser',
* 'thr','trp','tyr','val','try','ile'/
data indx/1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
* 18,19,20,18,10/
data (scales(1,j),j=1,20)/500,0,0,0,0,1650,0,0,0,500,1800,1800,0,
* 1300,2500,0,0,400,3400,2300,1500/
data (scales(2,j),j=1,20)/.5,-3.,-.2,-3.0,1.,-.2,-3.0,0.,.5,1.8,
* 1.8,-3.,1.3,2.5,0.,-.3,.4,3.4,2.3,1.5/
data (scales(3,j),j=1,20)/1.8,-4.5,-3.5,-3.5,2.5,-3.5,-3.5,-4.,
* -3.2,4.5,3.8,-3.9,1.9,2.8,-1.6,-.8,-.7,-.9,-1.3,4.2/
data tabs/400*'  /
data digits/'0','1','2','3','4','5','6','7','8','9'/

c
c
c      aacids(j) holds the 3 letter code for the jth amino acid in
c      the sequence.
c      names holds the three letter codes of the amino acids in
c      alphabetical order, both upper and lower case.
c      indx is the alphabetic index number (i.e. ala=1, val=20).
c      scales(1,1:20) holds the Nozaki-Tanford scale.
c      scales(2,1:20) holds the Hopp-Woods scale.
c      scales(3,1:20) holds the Kyte-Doolittle scale.
c      scale(1:20) holds normalized values for the scale of choice.
c      plot and tabs hold characters for plotted output to line printer.
c
c
c      Select the hydrophobicity scale here. Scales due to three
c      groups are included with the program. These include
c          a) Nozaki and Tanford (1971)

```

```
c      b) Hopp and Woods (1981)
c      c) Kyte and Doolittle (1982)
c A particular scale is chosen by setting the 'group' variable
c to either NT, HW, or KD. By default, the Nozaki-Tanford scale will
c be selected. Additional scales can be included by expanding the
c dimension of the array named 'scales' and using a data statement to
c initialize any new values.
c
c      group=NT
c
c
c      numres=0
c      Numres is the total number of residues in the sequence.
c
c      Transform the scale of choice into a scale ranging from [-100,+100].
c      First, determine minimum and maximum values.
c      min=scales(group,1)
c      max=min
c      do 1 i=2,20
c          if(scales(group,i).lt.min) min=scales(group,i)
c          if(scales(group,i).gt.max) max=scales(group,i)
1      continue
c      delta=max-min
c      factor=200./delta
c      offset=-min*factor-100.
c      Then transform the original values by this linear transformation.
c      do 2 i=1,20
c          scale(i)=scales(group,i)*factor+offset
2      continue
c
c
c      Read protein file name.
5      write(6,6)
6      format('$Enter protein file name:    ')
      read(5,7) inname
7      format(a20)
      open(unit=43,file=inname,type='old')
c      Input file now opened.
c
c      Determine input file type - Sequence or Brookhaven
4      write(6,9)
9      format('$Input type? (b=Brk,s=Seq)  ')
      read(5,3) ans
3      format(a1)
      if(ans(:1).eq.'s'.or.ans(:1).eq.'S') then
c
c      sequence-only file to be read here
      do 10 i=1,1000,10
          read(43,15,end=100) (aacids(j),j=i,i+9)
```

```
15      format(10(A3,x))
      if(aacids(i+9).ne.blanks) then
          numres=i+9
      else
          Line with less than ten residues encountered, so
          find out number of residues on last line.
          do 20 j=i+9,i,-1
              if(aacids(j).eq.blanks) then
                  goto 20
              else
                  numres=j
                  goto 100
              endif
20          continue
      endif
10      continue

c
c      else
c          Process Brookhaven file here.
        if(ans(:1).ne.'b'.and.ans(:1).ne.'B') goto 4
40      read(43,50,end=100) buff
50      format(a80)
        if(buff(:4).eq.'ATOM') then
            read(buff,60) resnam,resnum
60      format(17x,a3,2x,i4)
        if(numres.ne.resnum) then
            numres=resnum
            aacids(numres)=resnam
        endif
        endif
        goto 40
    endif
c
100     close(43,disp='keep')
c
c
c
c      Isolate stem of file name.
        jname=index(inname,period)
        jname=jname-1
c      jname is number of characters in file name stem
        pname=inname(:jname)//'.OUT'
        open(unit=46,name=pname,type='new',carriagecontrol='list')
c
c      Determine window size.
65      write(6,70)
70      format('$Window size= ')
        read(5,*) intrvl
```

```

c      Check to make sure the interval is an odd number.
      k1=intrvl/2.+0.5
      k2=intrvl/2.
      if(k1.eq.k2) then
          write(6,75)
          format(x,'Window size should be an odd number - try again... ')
75
      goto 65
      endif
c
c      Write the sequence on the output file.
      if(group.eq.1) grpnam='NT'
      if(group.eq.2) grpnam='HW'
      if(group.eq.3) grpnam='KD'
      write(6,110) numres,intrvl,grpnam,inname(:jname)
      write(46,110) numres,intrvl,grpnam,inname(:jname)
110     format(x,i4,' residues.',2x,'Window Size=',i3,2x,'Scale=',a2,2x,
           •           a,'.')
      do 120 i=1,numres,10
          jtwa=i+9
          if(jtwa.gt.numres) jtwa=numres
          write(46,125) i,(aacids(j),j=i,jtwa)
          format(x,i4,'.',2x,10(a3,2x))
120     continue
          write(46,*)
c
c      Ready to calculate the hydrophobicity profile.
      middle=intrvl/2.+1
      istart=middle
      iend=numres-middle+1
c
      do 200 i=1,numres
c
c      First, translate the amino acid 3 letter code into an index number
      resnam=aacids(i)
      do 210 j=1,2
      do 210 jj=1,22
          if(names(j,jj).eq.resnam) then
              idx=indx(jj)
              goto 220
          endif
210     continue
c      Error if fell through loop - couldn't find the 3 letter code.
      write(6,215) resnam
215     format(x,'Illegal 3 letter code - ',a3)
      goto 200
c
c      Then, look up the hydrophobicity value for this residue.
      score=scale(idx)
      values(i)=score

```

```
200      continue
c
c      Individual hydrophobicity values in array VALUES.
c      Take running average over interval.
if(intrvl.gt.0) then
    do 240 i=istart,iend
        score=0.
        do 250 j=1,intrvl
            score=score+values(i-middle+j)
250      continue
        vals(i)=score/intrvl
240      continue
        do 260 i=istart,iend
            values(i)=nint(vals(i))
260      continue
endif
c
c      Average hydrophobicity now in array VALUES(middle:numres-middle+1)
c      Print them on output file.
do 270 i=1,numres,10
    jtwa=i+9
    if(jtwa.gt.numres) jtwa=numres
    write(buff,275) i,(values(j),j=i,jtwa)
275      format(x,i4,'.',x,10(i4,x))
c
    if(i.lt.istart) then
        ktwa=istart-1
        if(ktwa.gt.jtwa) ktwa=jtwa
        do 276 j=i,ktwa
            jpt=8+(j-i)*5
            buff(jpt:jpt+3)=' ***'
276      continue
    endif
c
    if(jtwa.gt.iend) then
        kfwa=iend+1
        if(i.gt.kfwa) kfwa=i
        do 277 j=kfwa,jtwa
            jpt=8+(j-i)*5
            buff(jpt:jpt+3)=' ***'
277      continue
    endif
c
    write(46,280) buff(:58)
280      format(a)
270      continue
c
c
    close(43,disp='keep')
c
c
```

```

c      Make a hydrophobicity profile suitable for plotting on
c      the line printer.
c      write(6,300)
300      format('$Do you want a line printer plot of the profile? ')
      read(5,3) ans
      if(ans.eq.'n'.or.ans.eq.'N') then
          stop
c
c      else
          pname=inname(:jname)//'.plt'
          open(unit=46,name=pname,type='new',carriagecontrol='list')
          write(46,110) numres,intrvl,grpnam,inname(:jname)
c
          do 302 i=1,numres,100
              jtwa=i+99
              if(jtwa.gt.numres) jtwa=numres
c
c      form label for abscissa
          do 310 j=i+4,jtwa,5
              write(label,305) j
              format(i4)
              do 310 k=1,4
                  tabs(k,j-i+1)=label(k:k)
305          continue
c
c      then output the label
          do 315 j=1,4
              write(46,320) (tabs(j,k),k=1,jtwa-i+1)
              format(7x,100a)
320          continue
315          write(46,*)
c
c      form the plot
          do 325 j=i,jtwa
              jpt=j-i+1
              jscore=nint(values(j)/10.)+11
              k=j/10.
              jdig=digits(j-k*10+1)
              do 326 k=1,21
                  plot(k,jpt)=' '
326          if(j.lt.istart) goto 325
                  if(j.gt.iend) goto 325
                  plot(jscore,jpt)=jdig
325          continue
c
c      output the plot
          do 330 j=1,21
              jval=(11-j)
              write(46,335) jval,(plot(22-j,k),k=1,jtwa-i+1)
              format(x,i4,':',x,100a)
335

```

```

330      continue
c
c      then output label again
write(46,*)
do 340 j=1,4
      write(46,320) (tabs(j,k),k=1,jtwa-i+1)
340      continue

302      continue
c
c      endif
c
end

```

ACKNOWLEDGMENTS

We thank the following for critical reading of the manuscript: Buzz Baldwin, Marty Briggs, Arnie Hagler, Isabella Karle, Michel Marraud, and Dan Veber. We also thank Rick Lee and Bob Wood for useful discussion. We are especially grateful to John Edsall for his help at every stage in the preparation of this review. JAS wishes to acknowledge Howard M. Goodman for encouragement and support. LMG is a Fellow of the A.P. Sloan Foundation, 1984–1986, and GDR is recipient of an NIH Research Career Development Award. Financial support from NIH (LMG, GDR) and from Hoechst AG (West Germany) (JAS) is acknowledged.

REFERENCES

- Ajo, D., Granozzi, G., Tondello, E., and Del Pra, A. (1980). *Biopolymers* **19**, 469–475.
- Anderer, F. A. (1963a). *Z. Naturforsch. B: Anorg. Chem. Org. Chem. Biochem. Biophys. Biol.* **18**, 1010–1014.
- Anderer, F. A. (1963b). *Biochim. Biophys. Acta* **71**, 246–248.
- Argos, P., and Palau, J. (1982). *Int. J. Pept. Protein Res.* **19**, 380–393.
- Argos, P., Schwartz, J., and Schwartz, J. (1976). *Biochim. Biophys. Acta* **439**, 261–273.
- Arison, B. H., Hirschmann, R., and Veber, D. F. (1978). *Bioorg. Khim.* **7**, 447–451.
- Arison, B. H., Hirschmann, R., Paleveda, W. J., Brady, S. F., and Veber, D. F. (1981). *Biochem. Biophys. Res. Commun.* **100**, 1148–1153.
- Arnon, R. (1977). In "Immunochemistry of Enzymes and Their Antibodies" (M. R. J. Salton, ed.), p. 1. Wiley, New York.
- Arnon, R., Teicher, E., and Scheraga, H. A. (1974). *J. Mol. Biol.* **90**, 403–407.
- Atassi, M. Z. (1975). *Immunochemistry* **12**, 423–438.
- Atassi, M. Z. (1975). *Immunochemistry* **15**, 909–936.
- Atassi, M. Z., and Smith, J. A. (1978). *Immunochemistry* **15**, 609–610.
- Atassi, M. Z., and Webster, R. G. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80**, 840–843.
- Aubert, J. P., Biserte, G., and Loucheux-Lefebvre, M. H. (1976). *Arch. Biochem. Biophys.* **175**, 410–418.
- Aubert, J. P., Helbecque, N., and Loucheux-Lefebvre, M. H. (1981). *Arch. Biochem. Biophys.* **208**, 20–29.

- Aubert, J. P., Chiroutre, M., Kerckaert, J. P., Helbecque, N., and Loucheux-Lefebvre, M. H. (1982). *Biochem. Biophys. Res. Commun.* **104**, 1550–1559.
- Aubry, A., and Marraud, M. (1983). *Biopolymers* **22**, 341–345.
- Aubry, A., Protas, J., Boussard, G., and Marraud, M. (1977). *Acta Crystallogr. Sect. B* **33**, 2399–2406.
- Aubry, A., Protas, J., Boussard, G., and Marraud, M. (1979). *Acta Crystallogr. Sect. B* **35**, 694–699.
- Aubry, A., Protas, J., Boussard, G., and Marraud, M. (1980a). *Acta Crystallogr. Sect. B* **36**, 321–326.
- Aubry, A., Protas, J., Boussard, G., and Marraud, M. (1980b). *Acta Crystallogr. Sect. B* **36**, 2822–2824.
- Aubry, A., Protas, J., Boussard, G., and Marraud, M. (1980c). *Acta Crystallogr. Sect. B* **36**, 2825–2827.
- Aubry, A., Boussard, G., and Marraud, M. (1981a). *Acta Crystallogr. Sect. B* **37**, 1474–1477.
- Aubry, A., Vitoux, B., Boussard, G., and Marraud, M. (1981b). *Int. J. Pept. Protein Res.* **18**, 195–202.
- Aubry, A., Boussard, G., and Marraud, M. (1983). In "Peptides. Structure and Function" (V. J. Hruby and D. H. Rich, eds.), pp. 817–820. Pierce Chem. Co., Rockford, Illinois.
- Aubry, A., Boussard, G., and Marraud, M. (1985). *CR Hebd. Séances Acad. Sci. Ser. C*. In press.
- Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976). *J. Chem. Phys.* **64**, 2229–2246.
- Austen, B. M. (1979). *FEBS Lett.* **103**, 308–313.
- Ayato, H., Tanaka, I., and Ashida, T. (1981). *J. Am. Chem. Soc.* **103**, 6869–6873.
- Bach, A. C., II, and Giersch, L. M. (1985). *J. Am. Chem. Soc.* In press.
- Bach, A. C., II, Dunn, B. S., Giersch, L. M., Shimohigashi, Y., and Stammer, C. H. (1983). In "Peptides. Structure and Function" (V. J. Hruby and D. H. Rich, eds.), pp. 299–302. Pierce Chem. Co., Rockford, Illinois.
- Balaram, P. (1983). In "Peptides. Structure and Function" (V. J. Hruby and D. H. Rich, eds.), pp. 477–486. Pierce Chem. Co., Rockford, Illinois.
- Balasubramanian, R. (1977). *Nature (London)* **266**, 856–857.
- Bandekar, J., and Krimm, S. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76**, 774–777.
- Bandekar, J., and Krimm, S. (1980). *Biopolymers* **19**, 31–36.
- Bandekar, J., Evans, D. J., Krimm, S., Leach, S. J., McQuie, J. R., Minasian, E., Nemethy, G., Pottle, M. S., Scheraga, H. A., Stimson, E. R., and Woody, R. W. (1982). *Int. J. Pept. Protein Res.* **19**, 187–205.
- Bara, Y. A., Friedrich, A., Kessler, H., and Molter, M. (1978). *Chem. Ber.* **111**, 1045–1057.
- Baron, H. M., and Baltimore, D. (1982a). *J. Virol.* **43**, 969–978.
- Baron, H. M., and Baltimore, D. (1982b). *Cell* **28**, 395–404.
- Bause, E. (1983). *Biochem. J.* **209**, 331–336.
- Bause, E., and Legler, G. (1981). *Biochem. J.* **195**, 639–644.
- Bause, E., Hettkamp, H., and Legler, G. (1982). *Biochem. J.* **203**, 761–768.
- Beddell, C. R., Clark, R. B., Hardy, G. W., Lowe, L. A., Ubataba, F. B., Vane, J. R., Wilkinson, S., Chang, K. J., Cuatrecasas, P., and Miller, R. J. (1977). *Proc. R. Soc. London, Ser. B* **198**, 249–265.
- Benedetti, E., Bavoso, A., DiBlasio, B., Pavone, V., Pedone, C., Crisma, M., Bonora, G. M., and Toniolo, C. (1982). *J. Am. Chem. Soc.* **104**, 2437–2444.
- Ben-Naim, A. (1980). In "Hydrophobic Interactions." Plenum, New York.
- Berg, R. A., and Prockop, D. J. (1973). *Biochem. Biophys. Res. Commun.* **52**, 115–120.

- Berzofsky, J. A., Buckenmeyer, G. K., Hicks, G., Gurd, F. R. N., Feldmann, R. J., and Minna, J. (1982). *J. Biol. Chem.* **257**, 3189–3198.
- Beutler, E., Guinto, E., and Johnson, C. (1976). *Blood* **48**, 887–898.
- Bhatnagar, R. S., Rapaka, R. S., and Urry, D. W. (1978). *FEBS Lett.* **95**, 61–64.
- Bhatnager, P. K., Papas, E., Blum, H. E., Milich, D. R., Nitecki, D., Karel, M. J., and Vyas, G. N. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4400–4404.
- Bierzynski, A., Kim, P. S., and Baldwin, R. L. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2470–2474.
- Bittle, J. L., Houghten, R. A., Alexander, H., Shinnick, T. M., Sutcliffe, J. G., Lerner, R. A., Rowlands, D. J., and Brown, F. (1982). *Nature (London)* **298**, 30–33.
- Blaha, K., and Buděšínský, M. (1973). *Int. J. Pept. Protein Res.* **5**, 399–403.
- Both, G. W., and Sleigh, M. J. (1980). *Nucleic Acids Res.* **8**, 2561–2575.
- Bothner-By, A. A. (1979). In "Biological Applications of Magnetic Resonance" (R. G. Shulman, ed.), pp. 177–219. Academic Press, New York.
- Boussard, G., Marraud, M., and Néel, J. (1974). *J. Chem. Phys.* **71**, 1081–1091.
- Boussard, G., Marraud, M., and Aubry, A. (1979). *Biopolymers* **18**, 1297–1331.
- Bradbury, A. F., Smyth, D. G., and Snell, C. R. (1976). *Nature (London)* **260**, 165–166.
- Brahmachari, S. K., and Ananthanarayanan, V. S. (1978). *Curr. Sci.* **47**, 107–109.
- Brahmachari, S. K., and Ananthanarayanan, V. S. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5119–5123.
- Brahmachari, S. K., Ananthanarayanan, V. S., Brahms, S., Brahms, J., Rapaka, and Bhatnagar, R. S. (1979). *Biochem. Biophys. Res. Commun.* **86**, 605–612.
- Brahmachari, S. K., Bhat, T. N., Sudhakar, V., Vijayan, M., Rapaka, R. S., Bhatnagar, R. S., and Ananthanarayanan, V. S. (1981). *J. Am. Chem. Soc.* **103**, 1703–1708.
- Brahmachari, S. K., Bhatnagar, R. S., and Ananthanarayanan, V. S. (1982). *Biopolymers* **21**, 1107–1125.
- Briggs, M. S., and Giersch, L. M. (1984). *Biochemistry* **23**, 3111–3114.
- Brown, J. N., and Teller, R. G. (1976). *J. Am. Chem. Soc.* **98**, 7565–7569.
- Burgess, A. W., Ponnuswamy, P. K., and Scheraga, H. A. (1974). *Isr. J. Chem.* **12**, 239–286.
- Burgus, R., Butcher, M., Amos, M., Ling, N., Monohan, M., Rivier, J., Fellows, R., Blackwell, R., Vale, W., and Guillemin, R. (1972). *Proc. Natl. Acad. Sci. U.S.A.* **69**, 278–282.
- Bush, C. A., Sarkar, S. K., and Kopple, K. D. (1978). *Biochemistry* **17**, 4951–4954.
- Cann, J. R., Stewart, J. M., and Matseuda, G. R. (1973). *Biochemistry* **12**, 3780–3788.
- Cantor, C. R., and Schimmel, P. R. (1980). "Biophysical Chemistry," Parts I, II. Freeman, San Francisco.
- Chandrasekaran, R., Lakshminarayanan, A. V., Pandya, U. V., and Ramachandran, G. N. (1973). *Biochim. Biophys. Acta* **303**, 14–27.
- Chipens, G. (1983). In "Peptides. Structure and Function" (V. J. Hruby and D. H. Rich, eds.), pp. 865–868. Pierce Chem. Co., Rockford, Illinois.
- Chipens, G. I., Mutulis, F. K., Kataev, B. S., Klusha, V. E., Misina, I. P., and Myshlyakov, N. V. (1981). *Int. J. Protein Res.* **18**, 302–311.
- Chothia, C. (1976). *J. Mol. Biol.* **105**, 1–14.
- Chou, P. Y., and Fasman, G. D. (1974). *Biochemistry* **13**, 222–245.
- Chou, P. Y., and Fasman, G. D. (1977). *J. Mol. Biol.* **115**, 135–175.
- Chou, P. Y., and Fasman, G. D. (1978). *Annu. Rev. Biochem.* **47**, 251–276.
- Clarke, F. H., Jaggi, H., and Lovell, R. A. (1978). *J. Med. Chem.* **21**, 600–606.
- Cohn, E. J., and Edsall, J. T. (1943). "Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions." Van Nostrand-Reinhold, Princeton, New Jersey.

- Connolly, M. L. (1983). *Science* **221**, 709–713.
- Crawford, J. L., Lipscomb, W. N., and Schellman, C. G. (1973). *Proc. Natl. Acad. Sci. U.S.A.* **70**, 538–542.
- Czugler, M., Sasvári, K., and Hollósi, M. (1982). *J. Am. Chem. Soc.* **104**, 4465–4469.
- Date, T., and Wickner, W. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6106–6110.
- Deisenhofer, J. (1981). *Biochemistry* **20**, 2361–2370.
- Demel, D., and Kessler, H. (1976). *Tetrahedron Lett.*, pp. 2801–2804.
- Deslauriers, R., and Smith, I. C. P. (1980). In "Biological Magnetic Resonance," (L. Berliner and J. Reuben, eds.), Vol. 2, pp. 243–344. Plenum, New York.
- Deslauriers, R., Leach, S. J., Maxfield, F. R., Minasian, E., McQuie, J. R., Meinwald, Y. C., Nemethy, G., Pottle, M. S., Rae, I. D., Scheraga, H. A., Stimson, E. R., and Van Nispen, J. W. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2512–2514.
- Deslauriers, R., Evans, D. J., Leach, S. J., Meinwald, Y. C., Minasian, E., Nemethy, G., Rae, I. D., Scheraga, H. A., Somorjai, R. L., Stimson, E. R., Van Nispen, J. W., and Woody, R. W. (1981). *Macromolecules* **14**, 985–996.
- DiMaio, J., and Schiller, P. W. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7162–7166.
- DiMaio, J., Nguyen, T. M.-D., Lemieux, C., and Schiller, P. W. (1982). *J. Med. Chem.* **25**, 1432–1438.
- Dreesman, G. R., Sanchez, Y., Ionescu-Matiu, I., Sparrow, J. T., Six, H. R., Peterson, D. L., Hollinger, F. B., and Melnick, J. L. (1982). *Nature (London)* **295**, 158–160.
- Eberle, A., and Schwizer, R. (1976). *Clin. Endocrinol.* **5**, Suppl., 41s–48s.
- Edsall, J. T., and McKenzie, H. A. (1983). *Adv. Biophys.* **16**, 53–183.
- Einspahr, H., Cook, W. J., and Bugg, C. E. (1980). *Am. Crystallogr. Assoc. Abstr.* **7**, No. 2, 14.
- Emini, E. A., Jameson, B. A., and Wimmer, E. (1983). *Nature (London)* **304**, 699–703.
- English, J. L., and Stammer, C. H. (1978). *Biochem. Biophys. Res. Commun.* **87**, 780–782.
- Fainaru, M., Wilson, A. C., and Arnon, R. (1974). *J. Mol. Biol.* **84**, 635–642.
- Feller, W. (1950). "An Introduction to Probability Theory, and its Applications," Vol. 1. Wiley, New York.
- Finkelstein, A. V., and Ptitsyn, O. B. (1971). *J. Mol. Biol.* **62**, 613–624.
- Finney, J. L. (1975). *J. Mol. Biol.* **96**, 721–732.
- Fox, J. A., Tu, A. T., Hruby, V. J., and Mosberg, H. I. (1981). *Arch. Biochem. Biophys.* **211**, 628–631.
- Fox, J. W., Vavrek, R. J., and Stewart, J. M. (1981). In "Peptides. Synthesis, Structure, Function" (D. H. Rich and E. Gross, eds.), pp. 323–326. Pierce Chem. Co., Rockford, Illinois.
- Freidinger, R. M. (1981). In "Peptides. Synthesis, Structure and Function" (D. H. Rich and E. Gross, eds.), pp. 673–683. Pierce Chem. Co., Rockford, Illinois.
- Freidinger, R. M., Veber, D. F., Perlow, D. S., Brooks, J. R., and Saperstein, R. (1980). *Science* **210**, 656–658.
- Freidinger, R. M., Colton, C. D., Perlow, D. S., Whitten, W. L., Paleveda, W. J., Veber, D. F., Arison, B. H., and Saperstein, R. (1983). In "Peptides. Structure and Function" (V. J. Hruby and D. H. Rich, eds.), pp. 349–352. Pierce Chem. Co., Rockford, Illinois.
- Frey, M. H., Hexem, J. G., Leo, G. C., Tsang, P., Opella, S. J., Rockwell, A., and Giersch, L. M. (1983). In "Peptides. Structure and Function" (V. J. Hruby and D. H. Rich, eds.), pp. 763–771. Pierce Chem. Co., Rockford, Illinois.
- Garbay-Jaureguiberry, C., Roques, B. P., Oberlin, R., Anteunis, M., and Lala, A. K. (1976). *Biochem. Biophys. Res. Commun.* **71**, 558–565.
- Garnier, J., Osguthorpe, D. J., and Robson, B. (1978). *J. Mol. Biol.* **120**, 97–120.

- Gerin, J. L., Alexander, H., Shin, J. W.-K., Purcell, R. H., Dapolito, G., Engle, R., Green, N., Sutcliffe, J. G., Shinnick, T. M., and Lerner, R. A. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2365–2369.
- Gierasch, L. M., Deber, C. M., Madison, V., Niu, C., and Blout, E. R. (1981a). *Biochemistry* **20**, 4730–4738.
- Gierasch, L. M., Opella, S. J., and Frey, M. H. (1981b). In "Peptides. Synthesis, Structure, and Function" (D. H. Rich and E. Gross, eds.), pp. 267–276. Pierce Chem. Co., Rockford, Illinois.
- Gierasch, L. M., Frey, M. H., Hexem, J. G., and Opella, S. J. (1982a). In "NMR Spectroscopy: New Methods and Applications" (G. C. Levy, ed.), pp. 233–247. Amer. Chem. So., Washington, D.C.
- Gierasch, L. M., Lacy, J. E., Thompson, K. F., Rockwell, A. L., and Watnick, P. (1982b). *Biophys. J.* **37**, 275–284.
- Gierasch, L. M., Thompson, K. F., Rockwell, A. L., and Briggs, M. S. (1985). *Biopolymers*. In Press.
- Gingeras, T. R., Sciaky, D., Gelinas, R. E., Bing-Dong, J., Yen, C. E., Kelly, M. M., Bullock, P. A., Parsons, B. L., O'Neill, K. E., and Roberts, R. J. (1983). *J. Biol. Chem.* **257**, 13475–13491.
- Glickson, J. D. (1975). In "Peptides: Chemistry, Structure and Biology" (R. Walter and J. Meienhofer, eds.), pp. 787–802. Ann Arbor Science Publishers, Ann Arbor, Michigan.
- Green, N., Alexander, H., Olson, A., Alexander, S., Shinnick, T. M., Sutcliffe, J. G., and Lerner, R. A. (1982). *Cell* **28**, 477–487.
- Green, M., Brackmann, K. L., Lucher, L. A., Symington, J. S., and Kramer, T. A. (1983). *J. Virol.* **48**, 604–615.
- Greenquist, A. C., and Shohet, S. B. (1976). *Blood* **48**, 877–886.
- Greer, J. (1981). *J. Mol. Biol.* **153**, 1027–1042.
- Guillemin, R., and Gerich, J. E. (1976). *Annu. Rev. Med.* **27**, 397.
- Hagler, A. T. (1985). In "The Peptides," (V. J. Hruby and J. Meienhofer, eds.), Vol. 7. Academic Press, New York. In press.
- Hagler, A. T., Moult, J., and Osguthorpe, D. J. (1980). *Biopolymers* **19**, 395–418.
- Hagler, A. T., Osguthorpe, D. J., Dauber, P., and Hempel, J. C. (1985). *Science* **227**, 1309–1315.
- Hansch, C., and Leo, A. J. (1979). "Substitute Constants for Correlation Analysis in Chemistry and Biology." Wiley, New York.
- Harvey, R., Faulkes, R., Gillett, P., Lindsay, N., Paucha, E., Bradbury, A., and Smith, A. E. (1982). *EMBO J.* **1**, 473–477.
- Henderson, L. E., Copeland, T. D., Sowder, R. C., Smythers, G. W., and Oroszlan, S. (1981). *J. Biol. Chem.* **256**, 8400–8406.
- Higashijima, T., Tasumi, M., and Miyazawa, T. (1977). *Biopolymers* **16**, 1259–1270.
- Hildebrand, J. H. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76**, 194.
- Hopp, T. P. (1981). *Mol. Immunol.* **18**, 869–872.
- Hopp, T. P., and Woods, K. R. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3824–3828.
- Hopp, T. P., and Woods, K. R. (1983). *Mol. Immunol.* **20**, 483–489.
- Hosey, M., and Tao, M. (1976). *Nature (London)* **263**, 424–425.
- Hruby, V. J. (1982). *Life Sci.* **31**, 189–199.
- Hruby, V. J., and Mosberg, H. I. (1982). *Peptides* **3**, 329–336.
- Hruby, V. J., Mosberg, H. I., Sawyer, T. K., Knittel, J. J., Rockway, T. W., Ormberg, J., and Darman, P. (1983). *Biopolymers* **22**, 517–530.
- Hseu, Th., and Chang, H. (1980). *Biochim. Biophys. Acta* **624**, 340–345.

- Hughes, J., Smith, T. W., Kosterlitz, H. W., Fothergill, L., Morgan, B. A., and Morris, H. R. (1975). *Nature (London)* **258**, 577–579.
- Huijing, F., and Lee, E. Y. C., eds. (1973). "Protein Phosphorylation in Control Mechanisms." Academic Press, New York.
- Hvidt, A. (1983). *Annu. Rev. Biophys. Bioeng.* **12**, 1–20.
- Iitaka, Y., Nakamura, H., Takada, K., and Takita, T. (1974). *Acta Crystallogr. Sect. B* **30**, 2817–2825.
- Inouye, M., and Halegoua, S. (1980). *CRC Crit. Rev. Biochem.* **7**, 339–371.
- Iqbal, M., Nagaraj, R., and Balaram, P. (1981). *Int. J. Pept. Protein Res.* **18**, 208–213.
- Isogai, Y., Nemethy, G., and Scheraga, H. A. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74**, 414–418.
- Isogai, Y., Nemethy, G., Rackovsky, S., Leach, S. J., and Scheraga, H. A. (1980). *Biopolymers* **16**, 1183–1210.
- Ivanov, V. T., Filatova, M. P., Reissman, Z., Reutova, T. O., Efremov, E. S., Pashkov, V. S., Galaktionov, S. G., Grigoryan, G. L., and Ovchinnikov, Yu. A. (1975). In "Peptides: Chemistry, Structure, and Biology" (R. Walter and J. Meienhofer, eds.), pp. 151–157. Ann Arbor Science Publishers, Ann Arbor, Michigan.
- Jackman, L. M., and Sternhell, S. (1969). "Applications of Nuclear Magnetic Resonance Spectroscopy in Organic Chemistry." Pergamon, Oxford.
- Jackson, D. C., Murray, J. M., White, D. O., Fagan, C. N., and Tregear, G. W. (1982). *Virology* **120**, 273–276.
- Jackson, R. C., and Blobel, G. (1980). *Ann. N.Y. Acad. Sci.* **343**, 391–404.
- Janin, J. (1979). *Nature (London)* **277**, 491–492.
- Jardetzky, O. (1980). *Biochim. Biophys. Acta* **621**, 227–232.
- Jemmerson, R., and Margoliash, E. (1979). *J. Biol. Chem.* **254**, 12706–12714.
- Jones, C. R., Gibbons, W. A., and Garsky, V. (1976). *Nature (London)* **262**, 779–782.
- Jones, D. D. (1975). *J. Theor. Biol.* **50**, 167–183.
- Jung, G., Bosch, R., Katz, E., Schmitt, H., Voges, K.-P., and Winter, W. (1983). *Biopolymers* **22**, 241–246.
- Kabsch, W., and Sander, C. (1983). *Biopolymers* **22**, 2577–2637.
- Kabsch, W., and Sander, C. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1075–1078.
- Kaiser, E. T., and Kezdy, F. J. (1984). *Science* **223**, 249–255.
- Kanehisa, M. I., and Tsong, T. Y. (1978). *J. Mol. Biol.* **124**, 177–194.
- Karle, I. L. (1977). In "Peptides. Structure and Biological Function" (E. Gross and J. Meienhofer, eds.), pp. 681–690. Wiley, New York.
- Karle, I. L. (1978). *J. Am. Chem. Soc.* **100**, 1286–1289.
- Karle, I. L. (1979). *J. Am. Chem. Soc.* **101**, 181–184.
- Karle, I. L. (1981a). In "The Peptides" (E. Gross and J. Meienhofer, eds.), Vol. 4, pp. 1–54. Academic Press, New York.
- Karle, I. L. (1981b). In "Perspectives in Peptide Chemistry" (A. Eberle, R. Geiger, and T. Wieland, eds.), pp. 261–271. Karger, Basel.
- Karle, I. L., Karle, J., Mastropaoletti, D., Camerman, A., and Camerman, N. (1983). In "Peptides. Structure and Function" (V. J. Hruby and D. H. Rich, eds.), pp. 291–294. Pierce Chem. Co., Rockford, Illinois.
- Karplus, M., and Weaver, D. L. (1976). *Nature (London)* **260**, 404–406.
- Kauzmann, W. (1959). *Adv. Protein Chem.* **14**, 1–64.
- Kendrew, J. C. (1961). *Sci. Am.* **205**, 96–110.
- Kessler, H. (1982). *Angew. Chem. Int. Ed. Engl.* **21**, 512–523.
- Kessler, H., and Kondor, P. (1979). *Chem. Ber.* **112**, 3538–3551.
- Kessler, H., Hehlein, W., and Schuck, R. (1982). *J. Am. Chem. Soc.* **104**, 4534–4540.
- Kessler, H., Holzemann, G., and Geiger, R. (1983). In "Peptides. Structure and Function"

- (V. J. Hruby and D. H. Rich, eds.), pp. 295–298. Pierce Chem. Co., Rockford, Illinois.
- Kessler, H., Holzemann, G., and Zechel, C. (1985). *Int. J. Pept. Protein Res.* In Press.
- Khaled, M. A., and Urry, D. W. (1976). *Biochem. Biophys. Res. Commun.* **70**, 485–491.
- Khaled, M. A., Urry, D. W., and Okamoto, K. (1976). *Biochem. Biophys. Res. Commun.* **72**, 162–169.
- Khaled, M. A., Venkatachalam, C. M., Sugano, H., and Urry, D. W. (1981). *Int. J. Pept. Protein Res.* **17**, 23–33.
- Kim, P. S., and Baldwin, R. L. (1982). *Annu. Rev. Biochem.* **51**, 459–489.
- Knittel, J. J., Sawyer, T. K., Hruby, V. J., and Hadley, M. E. (1983). *J. Med. Chem.* **26**, 125–129.
- Kolaskar, A. S., Ramabrahman, V., and Soman, K. V. (1980). *Int. J. Pept. Protein Res.* **16**, 1–11.
- Kondor, P., and Kessler, H. (1979). In "Peptides. Structure and Biological Function" (E. Gross and J. Meienhofer, eds.), pp. 181–184. Pierce Chem. Co., Rockford, Illinois.
- Kopple, K. D. (1981a). *Biopolymers* **20**, 1913–1920.
- Kopple, K. D. (1981b). In "Peptides: Synthesis, Structure and Function" (D. H. Rich and E. Gross, eds.), pp. 295–298. Pierce Chem. Co., Rockford, Illinois.
- Kopple, K. D. (1983). *Int. J. Pept. Protein Res.* **21**, 43–48.
- Kopple, K. D., and Parameswaran, K. N. (1983). *Int. J. Pept. Protein Res.* **21**, 269–280.
- Kopple, K. D., and Sarkar, S. K. (1979). In "Peptides. Structure and Biological Function" (E. Gross and J. Meienhofer, eds.), pp. 201–204. Pierce Chem. Co., Rockford, Illinois.
- Kopple, K. D., and Zhu, P. P. (1983). *J. Am. Chem. Soc.* **105**, 7742–7746.
- Kopple, K. D., Go, A., Schamper, T. J., and Wilcox, C. S. (1973). *J. Am. Chem. Soc.* **95**, 6090–6096.
- Kopple, K. D., Schamper, T. J., and Go, A. (1974). *J. Am. Chem. Soc.* **96**, 2597–2605.
- Kopple, K. D., Go, A., and Pilipauskas, D. R. (1975). *J. Am. Chem. Soc.* **97**, 6830–6838.
- Kopple, K. D., Go, A., and Schamper, T. J. (1978). *J. Am. Chem. Soc.* **100**, 4289–4295.
- Kopple, K. D., Sarkar, S. K., and Giacometti, G. (1981). *Biopolymers* **20**, 1291–1303.
- Kopple, K. D., Zhu, P. P., and Go, A. (1983). *Biopolymers* **22**, 153–156.
- Kossiakoff, A. A. (1982). *Nature (London)* **296**, 713–721.
- Kosterlitz, H. W., Lord, J. A. H., Paterson, S. J., and Saterfield, A. A. (1980). *Br. J. Pharmacol.* **68**, 333–342.
- Krigbaum, W. R., and Komoriya, A. (1979). *Biochim. Biophys. Acta* **576**, 204–228.
- Krimm, S. (1983). *Biopolymers* **22**, 217–225.
- Krimm, S., and Bandekar, J. (1980). *Biopolymers* **19**, 1–29.
- Kronquist, K. E., and Lennarz, W. J. (1978). *J. Supramol. Struct.* **8**, 51–65.
- Kumar, A., Ernst, R. R., and Wüthrich, K. (1980). *Biochem. Biophys. Res. Commun.* **95**, 1–6.
- Kuntz, I. D. (1972). *J. Am. Chem. Soc.* **94**, 4009–4012.
- Kuntz, I. D., and Crippen, G. M. (1979). *Int. J. Pept. Protein Res.* **13**, 223–228.
- Kyte, J., and Doolittle, R. F. (1982). *J. Mol. Biol.* **157**, 105–132.
- Lando, G., Berzofsky, J. A., and Reichlin, M. (1982). *J. Immunol.* **129**, 206–211.
- Langbeheim, H., Arnon, R., and Sela, M. (1976). *Proc. Natl. Acad. Sci. U.S.A.* **73**, 4636–4640.
- Leach, S. J. (1983). *Biopolymers* **22**, 425–440.
- Leach, S. J., Némethy, G., and Scheraga, H. A. (1977). *Biochem. Biophys. Res. Commun.* **75**, 207–215.
- Lee, B. K., and Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.
- Lerner, R. A., Green, N., Alexander, H., Liu, F. T., Sutcliffe, J. G., and Shinnick, T. M. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3403–3407.

- Lesk, A. M., and Rose, G. D. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4304–4308.
- Levitt, M. (1976). *J. Mol. Biol.* **104**, 59–107.
- Levitt, M. (1978). *Biochemistry* **17**, 4277–4284.
- Levitt, M., and Clothia, C. (1976). *Nature (London)* **261**, 552–558.
- Levitt, M., and Greer, J. (1977). *J. Mol. Biol.* **114**, 181–239.
- Lewis, P. N., Momany, F. A., and Scheraga, H. A. (1971). *Proc. Natl. Acad. Sci. U.S.A.* **68**, 2293–2297.
- Lewis, P. N., Momany, F. A., and Scheraga, H. A. (1973). *Biochim. Biophys. Acta* **303**, 211–229.
- Lim, V. I. (1974). *J. Mol. Biol.* **88**, 857–894.
- Ling, N., and Guillemin, R. (1976). *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3308–3310.
- Ling, N., and Vale, W. (1975). *Biochem. Biophys. Res. Commun.* **63**, 801–805.
- Loew, G., and Burt, S. K. (1978). *Proc. Natl. Acad. Sci. U.S.A.* **75**, 7–11.
- London, R. E., Stewart, J. M., Cann, J. R., and Matwiyoff, N. A. (1978). *Biochemistry* **17**, 2277–2283.
- London, R. E., Stewart, J. M., Williams, R., Cann, J. R., and Matwiyoff, N. A. (1979). *J. Am. Chem. Soc.* **101**, 2455–2462.
- Lord, J. A. H., Waterfield, A. A., Hughes, J., and Kosterlitz, W. (1977). *Nature (London)* **267**, 495–499.
- Louie, A. H., and Somerjai, R. L. (1983). *J. Mol. Biol.* **168**, 143–162.
- McMeekin, T. L., Cohn, E. J., and Weare, J. H. (1935). *J. Am. Chem. Soc.* **57**, 626–633.
- Madison, V. (1974). In "Peptides, Polypeptides and Proteins" (E. R. Blout, F. A. Bovey, M. Goodman, and N. Lotan, eds.), pp. 89–98. Wiley, New York.
- Madison, V., and Kopple, K. D. (1980). *J. Am. Chem. Soc.* **102**, 4855–4863.
- Madison, V., and Schellman, J. (1970). *Biopolymers* **9**, 569–588.
- Maignret, B., and Pullman, B. (1974). *Theor. Chim. Acta* **35**, 113–128.
- Manavalan, P., and Momany, F. A. (1981). *Int. J. Pept. Protein Res.* **18**, 256–275.
- Manavalan, P., and Ponnuswamy, P. K. (1978). *Nature (London)* **275**, 673–674.
- Marshall, R. D. (1974). *Biochem. Soc. Symp.* **40**, 17–26.
- Matheson, R. R., and Scheraga, H. A. (1978). *Macromolecules* **11**, 818–829.
- Matsuo, H., Baba, Y., Nair, R. M. G., Arimura, A., and Schally, A. V. (1971). *Biochem. Biophys. Res. Commun.* **43**, 1374–1439.
- Matthews, B. W. (1975). *Biochim. Biophys. Acta* **405**, 442–451.
- Mauger, A. B., Stuart, O. A., Hightet, R. J., and Silverton, J. V. (1982). *J. Am. Chem. Soc.* **104**, 174–180.
- Maxfield, F. R., Bandekar, J., Evans, D. J., Leach, S. J., Némethy, G., and Scheraga, H. A. (1981). *Macromolecules* **14**, 997–1003.
- Mayer, R., and Lancelot, G. (1981). *J. Am. Chem. Soc.* **103**, 4738–4742.
- Meirovitch, H., Rackovsky, S., and Scheraga, H. A. (1980). *Macromolecules* **13**, 1398–1405.
- Mollay, C., Vilas, U., and Kreil, G. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2260–2263.
- Momany, F. A. (1976a). *J. Am. Chem. Soc.* **98**, 2990–2995.
- Momany, F. A. (1976b). *J. Am. Chem. Soc.* **98**, 2996–3000.
- Momany, F. A. (1978). *J. Med. Chem.* **21**, 63–68.
- Monahan, M. W., Amoss, M. S., Anderson, H. A., and Vale, W. (1973). *Biochemistry* **12**, 4616–4620.
- Mosberg, H. I., and Schiller, P. W. (1984). *Int. J. Pept. Protein Res.* **23**, 462–466.
- Mosberg, H. I., Hurst, R., Hruby, V. J., Galligan, J. J., Burks, T. K., Gee, K., and Yamamura, H. I. (1982). *Biochem. Biophys. Res. Commun.* **106**, 506–512.
- Mosberg, H. I., Hurst, R., Hruby, V. J., Galligan, J. J., Burks, T. F., Gee, K., and Yamamura, H. I. (1983a). In "Peptides. Structure and Function" (V. J. Hruby and D. H. Rich, eds.), pp. 279–282. Pierce Chem. Co., Rockford, Illinois.

- Mosberg, H. I., Hurst, R., Hruby, V. J., Gee, K., Yamamura, H. I., Galligan, J. J., and Burks, T. F. (1983b). *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5871–5874.
- Muller, G. M., Shapira, M., and Arnon, R. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79**, 569–573.
- Nagano, K. (1973). *J. Mol. Biol.* **75**, 401–420.
- Nagano, K. (1977). *J. Mol. Biol.* **109**, 251–274.
- Nagayama, K., Wüthrich, K., and Ernst, R. R. (1979). *Biochem. Biophys. Res. Commun.* **90**, 305–311.
- Nagayama, K., Kumar, A., Wüthrich, K., and Ernst, R. R. (1980). *J. Magn. Reson.* **40**, 321–334.
- Narasinga Rao, B. N., Kumar, A., Balaram, H., Ravi, A., and Balaram, P. (1983). *J. Am. Chem. Soc.* **105**, 7423–7428.
- Némethy, G., and Printz, M. P. (1972). *Macromolecules* **5**, 755–758.
- Némethy, G., and Scheraga, H. A. (1977). *Q. Rev. Biophys.* **10**, 239–352.
- Némethy, G., and Scheraga, H. A. (1980). *Biochem. Biophys. Res. Commun.* **95**, 320–327.
- Némethy, G., McQuie, J. R., Pottle, M. S., and Scheraga, H. A. (1981). *Macromolecules* **14**, 975–985.
- Neuberger, A., Gottschalk, A., Marshall, R. D., and Spiro, R. G. (1972). In "Glycoproteins: Their Composition, Structure, and Function" (A. Gottschalk, ed.), Vol. 5 (Part A), pp. 450–490. Elsevier, Amsterdam.
- Neurath, A. R., Kent, S. B. H., and Strick, N. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7871–7875.
- Nishikawa, K., and Ooi, T. (1980). *Int. J. Pept. Protein Res.* **16**, 19–32.
- Noggle, J. H., and Schirmer, R. E. (1971). "The Nuclear Overhauser Effect. Chemical Applications." Academic Press, New York.
- Northrup, S. H., Pear, M. R., Morgan, J. D., McCammon, J. A., and Karplus, M. (1981). *J. Mol. Biol.* **153**, 1087–1109.
- Nozaki, Y., and Tanford, C. (1971). *J. Biol. Chem.* **246**, 2211–2217.
- Palau, J., Argos, P., and Puidgdomenech, P. (1982). *Int. J. Pept. Protein Res.* **19**, 394–401.
- Parodi, A. J., and Leloir, L. F. (1979). *Biochim. Biophys. Acta* **559**, 1–37.
- Pease, L. G. (1979). In "Peptides. Structure and Biological Function" (E. Gross and J. Meienhofer, eds.), pp. 197–200. Pierce Chem. Co., Rockford, Illinois.
- Pease, L. G., and Watson, C. (1978). *J. Am. Chem. Soc.* **100**, 1279–1286.
- Pease, L. G., Niu, C. H., and Zimmerman, G. (1979). *J. Am. Chem. Soc.* **101**, 184–191.
- Pease, L. G., Frey, M. H., and Opella, S. J. (1981). *J. Am. Chem. Soc.* **103**, 467–468.
- Perlman, D., and Halvorson, H. D. (1983). *J. Mol. Biol.* **167**, 391–404.
- Pert, C. B., Pert, A., Chang, J. K., and Fong, B. T. W. (1976). *Nature (London)* **194**, 330–332.
- Petcher, T. J., Weber, H.-P., and Rüegger, A. (1976). *Helv. Chim. Acta* **59**, 1480–1488.
- Pfaff, E., Mussgay, M., Bohm, H. O., Schultz, G. E., and Schaller, H. (1982). *EMBO J.* **1**, 869–874.
- Pless, D. D., and Lennarz, W. J. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74**, 134–138.
- Prange, T., Sakarellos, C., Toma, F., Pascard, C., and Fermandjian, S. (1983). *J. Am. Chem. Soc.* **105**, 6306–6309.
- Prasad, B. V. V., Shamala, N., Nagaraj, R., Chandrasekaran, R., and Balaram, P. (1979). *Biopolymers* **18**, 1635–1646.
- Prasad, B. V. V., Ravi, A., and Balaram, P. (1981). *Biochem. Biophys. Res. Commun.* **103**, 1138–1144.
- Prasad, B. V. V., Balaram, H., and Balaram, P. (1982). *Biopolymers* **21**, 1261–1273.
- Prince, A. M., Ikram, H., and Hopp, T. P. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79**, 579–582.
- Rackowsky, S., and Scheraga, H. A. (1981). *Macromolecules* **14**, 1259–1269.

- Rae, I. D., Leach, S. J., Minasian, E., Smith, J. A., Zimmerman, S. S., Weigold, J. A., Hodes, Z. I., Némethy, G., Woody, R. W., and Scheraga, H. A. (1981). *Int. J. Pept. Protein Res.* **17**, 575–592.
- Ramachandran, G. N., Bansal, M., and Bhatnagar, R. S. (1973). *Biochim. Biophys. Acta* **322**, 166–171.
- Ramakrishnan, C., and Soman, K. V. (1982). *Int. J. Pept. Protein Res.* **20**, 218–237.
- Rao, Ch. P., Nagaraj, R., Rao, C. N. R., and Balaram, P. (1980). *Biochemistry* **19**, 425–431.
- Rapaka, R. S., and Urry, D. W. (1978). *Int. J. Pept. Protein Res.* **11**, 97–108.
- Rapaka, R. S., Okamoto, K., and Urry, D. W. (1978). *Int. J. Pept. Protein Res.* **11**, 109–127.
- Richards, F. M. (1977). *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.
- Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 167–339.
- Rivier, J., Rivier, C., Perrin, M., and Vale, W. (1981). In "LHRH Peptides as Male and Female Contraceptives" (G. I. Zatuchni, J. D. Shelton, and J. J. Sciarra, eds.), pp. 13–23. Harper, New York.
- Robson, B., and Osguthorpe, D. J. (1979). *J. Mol. Biol.* **132**, 19–51.
- Robson, B., and Suzuki, E. (1976). *J. Mol. Biol.* **107**, 327–356.
- Rocha é Silva, M., Beraldo, W. T., and Rosenfeld, G. (1949). *Am. J. Physiol.* **156**, 261–273.
- Ronin, C., and Aubert, J. P. (1982). *Biochem. Biophys. Res. Commun.* **105**, 909–915.
- Roques, B. P., Garbay-Jaureguiberry, C., Oberlin, R., Anteunis, M., and Lala, A. K. (1976). *Nature (London)* **262**, 778–779.
- Roques, B. P., Rao, R., and Marion, D. (1980). *Biochimie* **62**, 753–773.
- Rose, G. D. (1978). *Nature (London)* **272**, 586–590.
- Rose, G. D., and Roy, S. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4643–4647.
- Rose, G. D., and Seltzer, J. P. (1977). *J. Mol. Biol.* **113**, 153–164.
- Rose, G. D., Winters, R. H., and Wetlaufer, D. B. (1976). *FEBS Lett.* **63**, 10–16.
- Rose, G. D., Young, W. B., and Giersch, L. M. (1983). *Nature (London)* **304**, 654–657.
- Rosenblatt, M., Beaudette, N. V., and Fasman, G. D. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3983–3987.
- Rosenbloom, J., Harsh, M., and Jimenez, S. A. (1973). *Arch. Biochem. Biophys.* **158**, 478–484.
- Roses, A. D., and Appel, S. H. (1975). *J. Membr. Biol.* **20**, 51–58.
- Roses, A. D., Roses, M. J., Miller, S. E., Hull, K. L., Jr., and Appel, S. H. (1976). *N. Engl. J. Med.* **294**, 193–198.
- Rossmann, M. G., and Argos, P. (1981). *Annu. Rev. Biochem.* **50**, 497–532.
- Rubin, B., and Richardson, J. S. (1972). *Biopolymers* **11**, 2381–2385.
- Sato, K., Kawai, M., and Nagai, U. (1981). *Biopolymers* **20**, 1921–1927.
- Sawyer, T. K., Darman, P. S., Hruby, V. J., and Hadley, M. E. (1981). In "Peptides: Synthesis, Structure and Function" (D. H. Rich and E. Gross, eds.), pp. 387–390. Pierce Chem. Co., Rockford, Illinois.
- Sawyer, T. K., Hruby, V. J., Darman, P. S., and Hadley, M. E. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1751–1755.
- Sawyer, T. K., Cody, W. L., Knittle, J. J., Hruby, V. J., and Hadley, M. E., Hirsch, M. D., and O'Donohue, T. L. (1983). In "Peptides. Structure and Function" (V. J. Hruby and D. H. Rich, eds.), pp. 323–331. Pierce Chem. Co., Rockford, Illinois.
- Scheraga, H. A. (1981). *Biopolymers* **20**, 1877–1899.
- Schiller, P. W. (1984). In "The Peptides" (S. Udenfriend and J. Meienhofer, eds.), Vol. 6, pp. 219–268. Academic Press, New York.
- Schiller, P. W., and DiMaio, J. (1983). In "Peptides. Structure and Function" (V. J. Hruby and D. H. Rich, eds.), pp. 269–278. Pierce Chem. Co., Rockford, Illinois.

- Schiller, P. W., Eggiman, B., DiMaio, J., Lemieux, C., and Nguyen, T. M.-D. (1981). *Biochem. Biophys. Res. Commun.* **101**, 337-343.
- Schultz, G. E., Barry, C. D., Friedman, J., Chou, P. Y., Fasman, G. D., Finkelstein, A. V., Lim, V. I., Ptitsyn, O. B., Kabat, E. A., Wu, T. T., Levitt, M., Robson, B., and Nagano, K. (1974). *Nature (London)* **250**, 140-142.
- Schwyzer, R., and Eberle, A. (1977). *Front. Horm. Res.* **4**, 18-25.
- Schwyzer, R., Sieber, P., and Gorup, B. (1958). *Chimia* **12**, 90-91.
- Shapira, M., Gibson, M., Muller, G., and Arnon, R. (1985). *Proc. Natl. Acad. Sci. U.S.A.* In Press.
- Shinnar, A., and Kaiser, E. T. (1984). *J. Am. Chem. Soc.* **106**, 5006-5007.
- Shrake, A., and Rupley, J. A. (1973). *J. Mol. Biol.* **79**, 351-372.
- Silhavy, T. J., Benson, S. A., and Emr, S. D. (1983). *Microbiol. Rev.* **47**, 313-344.
- Small, D., Chou, P. Y., and Fasman, G. D. (1978). *Biochem. Biophys. Res. Commun.* **79**, 341-346.
- Smith, G. D., and Griffin, J. F. (1978). *Science* **199**, 1214-1216.
- Smith, G. D., Duax, W. L., Langs, D. A., DeTitta, G. T., Edmonds, J. W., Rohrer, D. C., and Weeks, C. M. (1975). *J. Am. Chem. Soc.* **97**, 7242-7247.
- Smith, J. A. (1977). Ph.D. Dissertation, University of Melbourne.
- Smith, J. A., and Pease, L. G. (1980). *CRC Crit. Rev. Biochem.* **8**, 315-399.
- Snyder, J. (1984). *J. Am. Chem. Soc.* **106**, 2393-2400.
- Sober, H. A., ed. (1970). "Handbook of Biochemistry," 2nd Ed., Chemical Rubber Co., Cleveland, Ohio.
- Spatola, A. F., Giersch, L. M., and Rockwell, A. L. (1983). *Biopolymers* **22**, 147-151.
- Sprecher, R. F., and Momany, F. A. (1979). *Biochem. Biophys. Res. Commun.* **87**, 72-77.
- Srinivasan, A. R., and Olson, W. K. (1980). *Int. J. Pept. Protein Res.* **16**, 111-123.
- Stimson, E. R., Zimmerman, S. S., and Scheraga, H. A. (1977). *Macromolecules* **10**, 1049-1060.
- Struck, D. K., Lennarz, W. J., and Brew, K. (1978). *J. Mol. Chem.* **253**, 5786-5794.
- Struthers, R. S., Hagler, A. T., and Rivier, J. (1984). *ACS Symp. Ser.* (J. Vida and M. Gordon, eds.), pp. 239-267. Am. Chem. Soc., Washington, D.C.
- Sweet, R. M., and Eisenberg, D. (1983). *J. Mol. Biol.* **171**, 479-488.
- Sutcliffe, J. G., Shinnick, T. M., Green, N., Liu, F.-T., Niman, H. L., and Lerner, R. A. (1980). *Nature (London)* **287**, 801-805.
- Tanaka, I., Ashida, T., Shimoniski, Y., and Kakirdo, N. (1979). *Acta Crystallogr. Sect. B* **35**, 110-114.
- Tanford, C. (1962). *J. Am. Chem. Soc.* **84**, 4240-4247.
- Tanford, C. (1980). "The Hydrophobic Effect," 2nd ed. Wiley, New York.
- Toma, F., Lam-Thanh, H., Piriou, F., Heindl, M. C., Lintner, K., and Fermandjian, S. (1980). *Biopolymers* **19**, 781-804.
- Toniolo, C. (1980). *CRC Crit. Rev. Biochem.* **9**, 1-44.
- Torchia, D. A., di Corato, A., Wong, S. C. K., Deber, C. M., and Blout, E. R. (1972a). *J. Am. Chem. Soc.* **94**, 609-615.
- Torchia, D. A., Wong, S. C. K., Deber, C. M., and Blout, E. R. (1972b). *J. Am. Chem. Soc.* **94**, 616-620.
- Umezawa, H., Takita, T., and Shiba, T., eds. (1978). "Bioactive Peptides Produced by Microorganisms." Halsted, New York.
- Urry, D. W., and Long, M. M. (1976). *CRC Crit. Rev. Biochem.* **4**, 1-45.
- Urry, D. W., and Ohnishi, T. (1974). *Biopolymers* **13**, 1223-1242.
- Urry, D. W., Mitchell, L. W., Ohnishi, T., and Long, M. M. (1975). *J. Mol. Biol.* **96**, 101-117.

- Uy, R., and Wold, F. (1977). *Science* **198**, 890–896.
- Veber, D. (1979). In "Peptides. Structure and Biological Function" (E. Gross and J. Meienhofer, eds.), pp. 409–419. Pierce Chem. Co., Rockford, Illinois.
- Veber, D. F. (1981). In "Peptides. Synthesis, Structure and Function" (D. H. Rich and E. Gross, eds.), pp. 685–694. Pierce Chem. Co., Rockford, Illinois.
- Veber, D. F., Holly, F. W., Nutt, R. F., Bergstrand, S. J., Brady, S. F., Hirschmann, R., Glitzer, M. S., and Saperstein, R. (1979). *Nature (London)* **280**, 512–514.
- Veber, D. F., Freidinger, R. M., Perlow, D. S., Paleveda, W. J., Jr., Holly, F. W., Strachan, R. G., Nutt, R. F., Arison, B. H., Homnick, C., Randall, W. C., Glitzer, M. S., Saperstein, R., and Hirschmann, R. (1981). *Nature (London)* **292**, 55–58.
- Venkatachalam, C. M. (1968). *Biopolymers* **6**, 1425–1436.
- Venkatachalapathi, Y. V., Nair, C. M. K., Vijayan, M., and Balaram, P. (1981). *Biopolymers* **20**, 1123–1136.
- Venkatachalapathi, Y. V., Prasad, B. V. V., and Balaram, P. (1982). *Biochemistry* **21**, 5502–5509.
- Vickery, B. H., Nestor, J. J., Jr., and Hafez, E. S. E., eds. (1984). "LHRH and Its Analogs: Contraceptive and Therapeutic Applications," MTP Press, Lancaster, England.
- Viswanadhan, V. N., and Sundaram, K. (1982). *Int. J. Pept. Protein Res.* **20**, 371–379.
- Vitoux, B., Aubry, A., Cung, M. T., Boussard, G., and Marraud, M. (1981a). *Int. J. Pept. Protein Res.* **17**, 469–479.
- Vitoux, B., Boussard, G., Cung, M. T., and Marraud, M. (1981b). In "Peptides. Synthesis, Structure and Function" (D. H. Rich and E. Gross, eds.), pp. 299–302. Pierce Chem. Co., Rockford, Illinois.
- Vyas, G. N., Milich, D. R., Gottfriend, T., Nitecki, D., and Karels, M. (1980). *Prog. Immunol., Proc. Int. Congr. Immunol., 4th, 1980*. Abstr. No. 9.6.16.
- Wagner, G., Pardi, A., and Wüthrich, K. (1983). *J. Am. Chem. Soc.* **105**, 5948–5949.
- Walker, J. M., Berston, G. G., Sandman, C. A., Coy, D. H., Schally, A. V., and Kastin, A. J. (1977). *Science* **196**, 85–87.
- Walter, G., and Doolittle, R. F. (1983). In "Genetic Engineering: Principles and Methods" (J. K. Setlow and A. Hollaender, eds.), Vol. 5, pp. 61–91. Plenum, New York.
- Walter, G., Scheidman, K.-H., Carbone, A., Laudano, A. P., and Doolittle, R. F. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5197–5200.
- Walter, G., Hutchinson, M. A., Hunter, T., and Eckhart, W. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4882–4886.
- Watson, R. J., Weiss, J. H., Salstrom, J. S., and Enquist, L. W. (1982). *Science* **218**, 381–383.
- Wertz, D. H., and Scheraga, H. A. (1978). *Macromolecules* **11**, 9–15.
- Williams, R. E. (1976). *Science* **192**, 473–474.
- Williams, R. J. P. (1977). *Angew. Chem. Int. Ed. Engl.* **16**, 766–777.
- Williams, R. W. (1983). *J. Mol. Biol.* **166**, 581–603.
- Williams, R. W., and Dunker, A. K. (1981). *J. Mol. Biol.* **152**.
- Wilson, I. A., Ladner, R. C., Skehel, J. J., and Wiley, D. C. (1983). *Biochem. Soc. Trans.* **11**, 145–147.
- Wolfenden, R., Andersson, L., Cullis, P. M., and Southgate, C. C. B. (1981). *Biochemistry* **20**, 849–855.
- Wong, T. W., and Goldberg, A. R. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7412–7416.
- Woody, R. W. (1974). In "Peptides, Polypeptides, and Proteins" (E. R. Blout, F. A. Bovey, M. Goodman, and N. Lotan, eds.), pp. 338–350. Wiley, New York.
- Wüthrich, K., Wider, G., Wagner, G., and Braun, W. (1982). *J. Mol. Biol.* **155**, 311–319.
- Wyatt, J. L., Greenquist, A. C., and Shohet, S. B. (1977). *Biochem. Biophys. Res. Commun.* **79**, 1279–1285.

- Young, J. D., Tsuchiya, D., Sandlin, D. E., and Holroyde, M. J. (1979). *Biochemistry* **18**, 4444–4448.
- Yunger, L. M., and Cramer, R. D. III (1981). *Mol. Pharmacol.* **20**, 602–608.
- Zimmerman, S. S., and Scheraga, H. A. (1977a). *Biopolymers* **16**, 811–843.
- Zimmerman, S. S., and Scheraga, H. A. (1977b). *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4126–4129.
- Zimmerman, S. S., Eliezer, N., and Simha, R. (1968). *J. Theor. Biol.* **21**, 170–201.