Andrei L. Lomize
Henry I. Mosberg
College of Pharmacy,
University of Michigan,
Ann Arbor, Michigan 48109

Received 5 December 1996;
accepted 19 February 1997

Thermodynamic Model of Secondary Structure for α-Helical Peptides and Proteins

Abstract: *A thermodynamic model describing formation of α-helices by peptides and proteins in the absence of specific tertiary interactions has been developed. The model combines free energy terms defining α-helix stability in aqueous solution and terms describing immersion of every helix or fragment of coil into a micelle or a nonpolar droplet created by the rest of protein to calculate averaged or lowest energy partitioning of the peptide chain into helical and coil fragments. The α-helix energy in water was calculated with parameters derived from peptide substitution and protein engineering data and using estimates of nonpolar contact areas between side chains. The energy of nonspecific hydrophobic interactions was estimated considering each α-helix or fragment of coil as freely floating in the spherical micelle or droplet, and using water/cyclohexane (for micelles) or adjustable (for proteins) side-chain transfer energies. The model was verified for 96 and 36 peptides studied by ^{1}H-nmr spectroscopy in aqueous solution and in the presence of micelles, respectively ([set 1] and [set 2]) and for 30 mostly α-helical globular proteins ([set 3]). For peptides, the experimental helix locations were identified from the published medium-range nuclear Overhauser effects detected by ^{1}H-nmr spectroscopy. For sets 1, 2, and 3, respectively, 93, 100, and 97% of helices were identified with average errors in calculation of helix boundaries of 1.3, 2.0, and 4.1 residues per helix and an average percentage of correctly calculated helix–coil states of 93, 89, and 81%, respectively. Analysis of adjustable parameters of the model (the entropy and enthalpy of the helix–coil transition, the transfer energy of the helix backbone, and parameters of the bound coil), determined by minimization of the average helix boundary deviation for each set of peptides or proteins, demonstrates that, unlike micelles, the interior of the effective protein droplet has solubility characteristics different from that for cyclohexane, does not bind fragments of coil, and lacks interfacial area.* © *1997 John Wiley & Sons, Inc. Biopoly* ***42:*** *239–269, 1997*

Keywords: *α-helix stability; secondary structure prediction; micelles; protein folding*

INTRODUCTION

There are two types of theoretical approaches to the protein folding problem. Approaches originating from conformational analysis of peptides and polymer physics consider a protein molecule as a long polymer chain, the energy of which must be minimized by searching in the space of torsion angles,[1–3] or by using simplified lattice models.[4–6] An alternative way of looking at the problem is to represent a protein as a system of secondary structure elements,[7–12] as in every publication describing three-

Correspondence to: Henry I. Mosberg
Contract grant sponsor: National Institutes of Health
Contract grant numbers: DA03910 and DA00118

CCC 0006-3525/97/020239-31

dimensional structures of specific proteins. Only α-helices, β-sheets, or short covalently bridged cycles (as in conotoxins or in metallothioneins) can be stable enough to serve as nucleations initiating protein folding, and therefore they are present in 3D structures of all known proteins. Cooperative formation of backbone hydrogen bonds in α-helices and β-sheets provides their high intrinsic stability, and simultaneously, burial of the polar main chain, which gives an additional energy gain when the amphiphilic secondary structure elements aggregate with each other, creating the nonpolar protein core. A simultaneous or stepwise formation of the secondary structure frameworks by the hydrophobically collapsed peptide chain, which is usually supplemented by covalent cross-linking in small proteins, has been directly demonstrated in experimental studies of protein folding.[13–17] In terms of secondary structure, the protein folding process can be represented as a sequence of the following events: (1) formation of α-helices and β-sheets by the collapsed peptide chain, (2) assembly of the regular secondary structure elements into the protein core, and (3) joining of nonregular loops and the less stable ''peripheral'' helices and β-strands to the core and the association of independently formed domains. A theory of protein self-organization must reproduce all these events to calculate the protein 3D structure.

Formation of α-helices depends on various factors that can be studied separately by considering the following, increasingly complicated situations: (1) small linear peptides in aqueous solution, where stability of each helix depends only on interactions between its own residues (Figure 1a); (2) peptide–micelle complexes, where each helix is stabilized by a combination of the intrahelical and hydrophobic interactions with the micelle (Figure 1b); and (3) proteins, in which helices are stabilized by specific tertiary interactions along with intrahelical and nonspecific hydrophobic ones (Figure 1c; we denote as ''specific'' the interactions between atoms or groups that must be described by pairwise potentials, and as ''nonspecific'' the interactions of individual groups with a medium or averaged surrounding which can be described by transfer energies). The helix–coil transition is usually treated by Lifson–Roig and Zimm–Bragg theories.[18] However, even with essential modifications,[19–22] these theories and other models[23,24] deal only with intrahelical interactions (i.e., they describe formation of individual helices in water, Figure 1a), or can be modified for the specific case of dimeric coiled coils.[25] The goal of the present work is to develop a thermodynamic model of α-helix formation that would be applicable for micelle-bound peptides and for proteins (Figure 1b,c). The model, also for the first time, reproduces locations of the α-helices identified from medium-range NOEs in a representative set of peptides, instead of using average α-helicities derived from CD spectroscopy data, or qualitative comparisons with chemical shifts of $C^{\alpha}H$ protons, as in the previous theoretical studies of peptides in aqueous solution.[23,24]

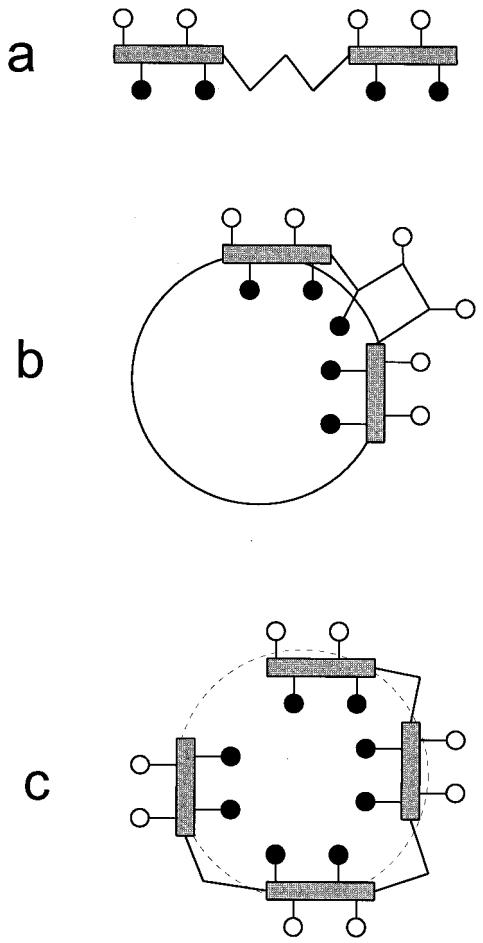

FIGURE 1 Three models of α-helix formation (the helices are shown as rectangles, solid circles are hydrophobic side chains): (a) ''peptide in aqueous solution'' (there are only specific interactions between residues within each α-helix); (b) ''peptide in complex with a micelle'' (there are specific intrahelical and nonspecific hydrophobic interactions of every α-helix with the micelle)—coil fragments may compete with helices for binding with the micelle; (c) ''droplet-like protein'' model (each helix and coil fragment floats in the liquid-like nonpolar spherical droplet created by the rest of protein).

The model can be briefly outlined as follows.

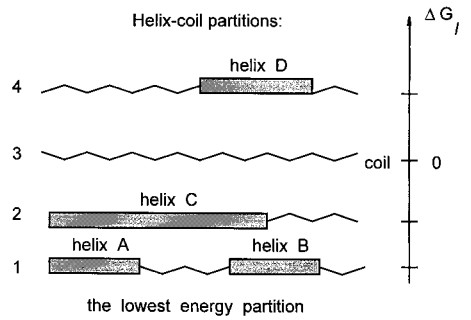

FIGURE 2 Helix–coil partitions as conformational states of the peptide chain. The coil in aqueous solution serves as a reference state with zero energy. The helices A, B, and C, shown as rectangles, with $\Delta G < 0$, are more stable than the coil. The helices compete with each other, and partition **1** consisting of two (A + B) helices can be of lower energy than partition **2** containing only helix C overlapped with A and B, even if helix C has lower energy than either of the individual A and B helices. Helix D ($\Delta G > 0$) is less stable than coil but may be detected spectroscopically. Partitions **1–4** are in equilibrium with each other and all may contribute to observed parameters of nmr and CD spectra.

Each partition of a peptide into helix and coil fragments (Figure 2) can be considered as a molecular conformational state defined by the variables N, k_1, m_1, . . . , k_i, m_i, . . . , k_N, and m_N, where N is the number of helices in the molecule, and where k_i and m_i ($i = 1, 2, \ldots, N$) represent the number of the first residue and the length, respectively, for each helix. Like coil or folded protein states, each helix–coil partition is an ensemble of conformers defined by torsion angles φ, ψ, χ, and, judging from molecular dynamics simulations,[26] interconversions of the partitions, i.e., lengthening, shortening, or breaking of helices, are slower than rotations of side chains and coil fluctuations.

For a peptide in aqueous solution (Figure 1a), the unfolding free energy, ΔG_l, of helix-coil partition l can be written as the sum of the helix–coil free energy differences, $\Delta G^{\alpha}(k_i, m_i)$, for all individual α-helices from the partition:

$$\Delta G_l = \sum_{i=1}^{N_l} \Delta G^{\alpha}(k_i, m_i) \qquad (1)$$

where N_l is the number of helices in partition l. The energies of individual helices in water, $\Delta G^{\alpha}(k_i, m_i)$, were calculated here with parameters derived from peptide substitution and protein engineering data, similar to that in the work of Munoz and Serrano,[23,24] but using a more physically justified approach for hydrophobic interactions between side chains in helices and a slightly different parametrization of some other interactions.

For a peptide in the micelle bound state (Figure 1b) ΔG_l, the free energy of its bound helix–coil partition l relative to a coil in aqueous solution, can be given by

$$\Delta G_l = (E_{\mathrm{el}} - T\Delta S_{\mathrm{imm}}) + \sum_i \Delta G^{\alpha}(k_i, m_i) + \sum_j \Delta G^{\mathrm{coil}}(k_j, m_j) \qquad (2)$$

where E_{el} is the peptide-micelle electrostatic interaction energy, ΔS_{imm} is the immobilization entropy of the peptide,[27] and the two sums in this equation are free energy changes for bound α-helical and coil fragments of m residues starting from residue k. Equation (2) can be simplified assuming, first, that the equilibrium is strongly shifted toward the bound peptide form, so that only bound helix–coil partitions need be considered, and second, that the total energy of electrostatic interactions of charged peptide groups with the micelle does not depend on the secondary structure of the peptide. Then the ($E_{\mathrm{el}} - T\Delta S_{\mathrm{imm}}$) term, which is of crucial importance for peptide–micelle binding, can be considered to be a constant for all bound helix–coil partitions and subtracted in calculations of their relative energies.

The energies of individual helices are additive [as in Eqs. (1) and (2)] when the helices do not interact with each other, i.e., for monomeric peptides lacking tertiary structure (Figure 1a,b), but the situation is more complicated in the presence of specific tertiary interactions. However, if the tertiary interactions are reduced, as in molten globules and in the intermediate and transition protein folding states,[14,28–31] the additivity approximation for helix energies can be applied. In a fluctuating compact state, each α-helix can be considered as floating in a dynamically averaged interior of a nonpolar spherical droplet created by the rest of the protein (Figure 1c) and stabilized independently of other helices by intrahelical and nonspecific hydrophobic interactions, similar to the micelle-bound peptides. Then, the energies of individual helices and coil fragments in a protein can also be simply summed:

$$\Delta G_l = \sum_i \Delta G^{\alpha}(k_i, m_i) + \sum_j \Delta G^{\mathrm{coil}}(k_j, m_j) + \Delta G' \qquad (3)$$

where the energies of the bound helices and coil

segments can be calculated similar to that for micelle-bound peptides, and the $\Delta G'$ term arises from loss of entropy by aggregating helices and is assumed to be a constant, independent of the helix–coil partition. Then, the relative energies of the helix–coil partitions can be approximated by the first two sums from this equation, which differ from unfolding free energies by the term $\Delta G'$.

All possible helix–coil partitions are in equilibrium with each other (Figure 2), including single helices, which are less stable than coil, but still detectable spectroscopically. This situation can be treated using Boltzmann averaging of the partitions[32] to calculate local α-helicities that can be compared with spectroscopically observed parameters. The number of the possible partitions grows rapidly with the chain length, which makes such calculations impossible for proteins. However, we show here that even the single *lowest* energy helix–coil partition (Figure 2) can satisfactorily reproduce experimentally observed locations of the helices, which are additionally stabilized by hydrophobic interactions with the micelles or with the rest of the protein. If the helix energies are additive, the search for the lowest free energy helix–coil partition (i.e., the global energy minimization with respect to the N, k_1, m_1, k_2, m_2, . . . , k_N, m_N variables) can be easily performed using the dynamic programming algorithm.[33]

METHODS

The computational procedure implemented here in the program FRAMEWORK consists of the following steps: (1) Calculation of α-helix and bound coil energies for each fragment of the molecule, depending on the chosen model [''peptide in aqueous solution,'' ''peptide in micelle,'' or ''droplet-like protein''; Eqs. (1)–(20)]. (2) Boltzmann averaging of helix–coil partitions to calculate the local α-helicities of every tripeptide fragment of the molecule [Eqs. (21) and (22)] or search for the lowest energy helix–coil partition [Eqs. (23) and (24). (3) Minimization of the average deviation of calculated and experimental boundaries of α-helices [Eq. (25)] with respect to several adjustable parameters of the model.

The average helix boundary deviation [Eq. (25)] was implemented, since the widely used percentage of correctly calculated secondary structure states (α, β, or nonregular) does not properly reflect success or failure of a prediction algorithm: a wrong prediction that sperm whale myoglobin, for example, is a single long helix would have a ''success'' rate as high as 89%, while the correct identification of all myoglobin helices with a small ($\sim$ 10%) error in the ends of each helix would produce the same success rate.

Free Energy of α-Helix in Aqueous Solution

The helix–coil free energy difference, $\Delta G^{\alpha}(k, m)$, for a fragment of peptide chain of m residues, starting from residue k, can be divided into the contribution of main-chain interactions (ΔG^{mch}), which is the free energy difference for the ''host'' polyAla peptide, the interactions of side chains with the helix backbone ($\Delta G^{\mathrm{sch}}_{\mathrm{int}}$) that describes free energy changes associated with replacement of the host Ala $C^{\beta}H_3$ group by other side chains,[34,35] the hydrogen-bonding and electrostatic interactions between polar side chains in water, $\Delta G^{\mathrm{sch}}_{\mathrm{hb}}$[36–38] and the hydrophobic interactions of side-chains $\Delta G^{\mathrm{sch}}_{\mathrm{pho}}$ (Refs. 39 and 40):

$$\Delta G^{\alpha}(k_i, m_i) = \Delta G^{\mathrm{mch}} + \Delta G^{\mathrm{sch}}_{\mathrm{int}} + \Delta G^{\mathrm{sch}}_{\mathrm{hb}} + \Delta G^{\mathrm{sch}}_{\mathrm{pho}} \quad (4)$$

Main-Chain Interactions. The helix–coil free energy difference for the host polyAla peptide is given by

$$\Delta G^{\mathrm{mch}}(k_i, m_i) = (m_i - 2)\Delta H - m_i T\Delta S \quad (5)$$

where ΔH is the enthalpy of the hydrogen-bonding interaction between two peptide groups in the α-helix, and ΔS is the conformational entropy change per residue during the helix–coil transition.[34] The ΔH and ΔS contributions measured by Hermans[41] and Scholtz et al.[42] are considered here as adjustable parameters of the model and must be determined independently by fit of calculated and experimentally identified positions of α-helices in peptides.

Side-Chain–Main-Chain Interactions. The energy of interaction between side-chains and the α-helix backbone $\Delta G^{\mathrm{sch}}_{\mathrm{int}}$ was calculated as the sum of corresponding published free energy differences $\Delta\Delta G^{\mathrm{sch}}_i$, measured by replacing the host Ala residue in model peptides and proteins:

$$\Delta G^{\mathrm{sch}}_{\mathrm{int}} = \sum_{i=k-1}^{k+m-1} \Delta\Delta G^{\mathrm{sch}}_i \quad (6)$$

where the replacement energies $\Delta\Delta G^{\mathrm{sch}}_i$ depend on the type of side chain i and its position within the α-helix or nearby: the energies can be different in the middle of the α-helix and near its termini, in positions denoted as N′-Ncap-N1-N2-N3-· · ·-C3-C2-C1-Ccap-C′. The corresponding α-helix propensities ($\Delta\Delta G^{\mathrm{sch}}$) measured for different peptides and proteins are not perfectly mutually consistent, and some of them reproduce the nmr-detected peptide helices more satisfactorily than others. Attempts to reproduce the peptide helices led to the parametrization and interpretation of the published α-helix propensity data described below.

Middle Helix, C-Turn, C-Cap, and N-Turn Positions. Because of the two-state behavior of proteins, the corresponding protein engineering scales were derived directly from thermodynamic measurements, while the corresponding energies for peptides have been obtained by using theories of the helix−coil transition. Remarkably, the averaging of two protein engineering scales measured in the middle helix positions (for α-helical dimers[43] and 44 site of T4 lysozyme[44]) gives a set of $\Delta\Delta G^{sch}$ values that is nearly identical (the correlation coefficient is 0.98) to the scale independently developed for 10 residues by Lyu et al.[45] using the model ''EXK'' peptide. The ''EXK'' peptide, which is stabilized by numerous ionic pairs and by the N-capping motif, also has a protein-like two state behavior, as can be seen from the similar $\Delta\Delta G^{sch}$ energies calculated using two-state and multistate models from CD data.[45] Thus, all these three middle-helix scales are consistent and can simply be averaged to reduce the experimental errors. The corresponding average $\Delta\Delta G^{sch}$ values used here (Table I) are close to the AGADIR scale[23,24] for all but Pro and Gly residues, and to the scale of Chakrabatty et al.[53] for all residues, except Val, Phe, Trp, Pro, and Gly.

In the helix C-turn (C2 position),[46] the experimental $\Delta\Delta G^{sch}$ energies are different: they are larger than in the middle of the α-helix by 0.3−0.5 kcal/mol for aromatic Trp, Phe, and Tyr residues and Cys, by ∼ 0.4 kcal/mol for β-branched Ile and Val side chains, by 0.1−0.2 kcal/mol for linear side chains containing a $C^{\gamma}H_2$ group (Leu, Met, Glu, Gln), and are unchanged for Gly and the short polar Ser and Asn side chains (Table I). These energetically unfavorable effects probably arise from shielding of unpaired carbonyls at the C-terminus of the α-helix by the γ substituents of the side chains and the larger accessibility of the nonpolar γ substituents themselves in the C-turn, compared to that in the middle of the α-helix. If the C2 side chain has a *trans* orientation ($\chi^1 \sim 180°$), its γ-methyl group or aromatic ring (of Phe, for example) reduces accessibility of the closest (C2) free C═O main chain oxygen by 26 or 36%, respectively, while the accessibilities of the nonpolar γ-methyl or aromatic ring themselves are increased by ∼ 11 Å^2 (the equivalent transfer energy is ∼ +0.2 kcal/mol) compared to that in the middle of α-helix. At the same time, the accessibilities of the C═O groups and side chains are not affected in the C-turn if the side chains have *gauche* orientations ($\chi^1 \sim -60°$). As discussed below, this solvation effect changes preferred conformations of side chains in the C-turn from *trans* to *gauche.*

The destabilization in C-turn positions is less for Lys and Arg compared with other residues with linear chains, and for His compared to other aromatic residues (Table I), probably because of small (∼ −0.2 kcal/mol) electrostatic attractions between the positively charged side chains and the helix dipole (as a result, the $\Delta\Delta G^{sch}$ energies of Lys and Arg in the middle of the helix and C-turn are identical, Table I). Repulsions of the Asp side chain with the helix dipole increases its $\Delta\Delta G^{sch}$ energy by ∼ +0.2 kcal/mol in C-turn positions compared to middle helix positions (Table I). The influence of electrostatic interactions is smaller for the Glu residue (∼ +0.1 kcal/mol), because its longer, flexible side chain can move away from the helix reducing the electrostatic repulsion. The electrostatic interactions of side chains at the C-terminus of the helix are weaker than at the N-terminus (−0.6 to −0.9 kcal/mol[54]) because the interactions depend on the spatial position of the charged groups relative to the helix dipole. The C^{α}-C^{β} bonds of side chains are tilted relative to the helix axis and directed toward the helix N-terminus. As a result, in the N-turn, the COO^- groups of Asp and Glu side chains are situated close to the helix dipole axis, near unsatisfied local dipoles of backbone NH groups, and may even form hydrogen bonds with them, while the positively charged side chains in the C-turn are far from the helix dipole axis. However, when His, Lys, or Arg residues occupy the C-cap position and their φ and ψ angles are in the left-handed helix area of the Ramachandran map (the structural motif of His[18] in barnase), the positively charged side chains are brought into the same position relative to the helix dipole as the negatively charged side chains in the N-turn: they are situated near the helix axis and can form hydrogen bonds with the main chain C═O groups, thus producing stronger electrostatic interactions: ∼ −0.6 kcal/mol.[55] Stabilization of α-helices by positively charged side chains, observed for model peptides,[56] may arise chiefly from this C-capping interaction. No special contributions for electrostatic interactions in the C-turn were used since they are already included in the C-turn $\Delta\Delta G^{sch}$ energies, and an average energy of electrostatic interactions for His, Lys, and Arg residues in the C-cap position was considered as an adjustable parameter, whose optimum value was found to be −0.4 kcal/mol. No other contributions were used for C-cap residues because experimental data here are contradictory: some studies[57] clearly demonstrate the significance of the C-capping interactions, especially for Asn residues, while others[48] show that these interactions are negligible.

In N-turn (N1-N3) positions, a small (−0.2 kcal/mol) correction of the middle helix scale was applied for the short polar Ser, Thr, and Asn residues and for Gly based on results of Serrano et al.[58] The $\Delta\Delta G^{sch}$ of Pro in N2 and N3 positions was reduced to 1 kcal/mol,[59,60] since the Pro side chain in the N turn of the α-helix causes steric hindrances with the preceding residue but does not produce an energetically unfavorable kink in the α-helix (this correlates with the much higher statistical occurrence of Pro in N-turn compared to middle helix positions).[61]

The pH dependence of all electrostatic contributions and pKs for charged side chains were taken into account as in the work of Munoz and Serrano.[24] Energies of electrostatic interactions of completely ionized side chains in N-turn positions with the helix dipole were considered as adjustable parameters, and their optimum values were −0.9 kcal/mol for Asp and Glu in the N1 and N2 positions (the ''capping box'' N3 residues were treated sepa-

Table I α-Helix ''Propensities'' ($\Delta\Delta G^{sch}$) and Transfer Energies (ΔG_{tr}^{sch}) of Side Chains

	$\Delta\Delta G^{sch}$ (kcal/mol)			ΔG_{tr}^{sch} (kcal/mol)		
Residue	Middle Helix[a]	C-turn Positions[b]	N-Cap Positions[c]	Cyclohexane Coil[d]	Cyclohexane α-Helix[d]	Protein α-Helix[e]
Leu	0.14	0.35	−0.17	−3.91	−3.01	−2.10
Ile	0.35	0.81	−0.10	−3.76	−2.63	−1.89
Val	0.46	0.88	0.01	−2.89	−1.97	−1.40
Phe	0.37	0.69	−0.17	−2.00	−1.99	−2.03
Trp	0.35	0.84	−0.41	−1.63	−1.58	−1.68
Met	0.20	0.31	−0.06	−1.79	−1.36	−1.47
Pro	3.40	3.40	−0.06	−2.23	−1.00	−0.49
Ala	0.0	0.0	0.0	−0.94	−0.63	−0.21
Cys	0.49	0.82	−0.08	−0.89	−0.57	−1.68
Tyr	0.42	0.82	−0.25	0.86	0.85	−0.63
Gly	0.86	0.91	−0.34	−0.29	−0.25	−0.14
Thr	0.54	0.79	−0.19	3.32	2.38	0.00
Ser	0.43	0.41	−0.34	3.99	3.06	2.10
His	0.55	0.78	−0.17	5.29	5.82	5.8
Lys	0.17	0.19	0.06	6.38	7.25	7.2
Gln	0.30	0.48	0.22	6.37	7.03	7.0
Glu	0.46	0.55	−0.17	7.53	8.23	8.2
Asn	0.63	0.66	−0.52	7.23	7.71	7.7
Asp	0.53	0.71	−0.51	9.39	9.81	9.8
Arg	0.14	0.14	0.0	15.75	16.49	16.5

[a] Average parameters for the host α-helix dimers,[43] EXK model peptide,[45] and T4 lysozyme[44] (the data for the 44 and 131 sites were averaged).

[b] Data of Horovitz et al.[46]

[c] Calculated using two-state helix−coil approximation from data of Chakrabatty et al.[47] and Doig and Baldwin[48] (peptide of 17 residues with uncharged N-terminus); pH dependencies of the energies were not taken into account: the data are for charged Glu and Asp and uncharged His and Cys residues (pH = 7). The N-capping energies for Asn, Ser, Thr, and Gly residues in the ''capping box'' combination (when Glu residue occupies N3 position in helix) for peptides were −0.58, −0.74, −0.59, and −0.25 kcal/mol, respectively, and the optimized N-capping energies for proteins for the Asp, Asn, Ser, and Thr residues were −0.8, −0.7, −0.5, and −0.4 kcal/mol, respectively, with any residue occupying the N3 position.

[d] Mole fraction based water−cyclohexane transfer energies[49] corrected for burial of side-chain analogues in α-helix and coil (as described below); for Ser and Thr forming hydrogen bonds with backbone of their own α-helix,[50] the transfer energies were corrected by −1.5 kcal/mol, the energy of a buried hydrogen bond in proteins.[51,52]

[e] Parameters obtained by minimization of helix boundary deviation with respect to side-chain transfer energies for a set of 30 proteins. For hydrophilic side chains (His and subsequent residues in the table), the energies could not be defined by this adjustment and correspond to the cyclohexane scale.

rately), and +0.5 kcal/mol for His, Lys, and Arg in the N1, N2, and N3 positions, close to the 0.6−0.9 kcal/mol estimated by mutagenesis.[54] $\Delta\Delta G^{sch}$ of Glu, Asp, and Gln residues in the capping box[62] N3 position (−0.40, −0.11, and −0.09 kcal/mol, respectively) were calculated using as the first approximation the all-or-none two-state model from CD data.[63]

N-Cap and N′ and C′ Positions. In contrast to the middle helix positions, the N-capping energies identified by protein engineering are highly variable: −2.2 to −0.4 kcal/mol for Asn, −2.1 to −0.9 kcal/mol for Ser, and −0.6 to 0.1 for Gly in the capping box motif.[58,64–66] This can be explained by ''context-dependent factors'': unlike residues in the α-helix, the replaced N-cap residues have different main chain φ and ψ angles,[64] and form extra hydrogen bonds with sequentially distant residues of the protein and with bound water molecules, as observed, for example, in substitution sites Thr[59]-Glu[62] of T4 lysozyme[64] and Ser[31]-Glu[34] of chymotrypsin inhibitor II.[66] The $\Delta\Delta G^{sch}$ energies for the N-cap position applied here (Table I) were calculated using the two-state model from CD data of Chakrabatty et al.[47] and Doig and Baldwin,[48] and for the special case of the capping box (Glu residue occupies N3 position), from data of Zhou et al.[63] for Gly, Asn, and Ser residues. These two-state peptide energies (−0.58, −0.74, and −0.25 kcal/mol, for Asn, Ser, and Gly, respectively, with the capping box motif, for example) are close to the lower limits estimated by protein engineering.

Table II The Replacement Energies ($\Delta\Delta G^{sch}$, kcal/mol) of Nonpolar Side Chains in the N′ Position[a]

N′ Position Residue	Intrinsic Contribution[b]	Energy of Interaction with Leu in N4 Position[c]	Energy of Interaction with Leu in N7 Position[c]
Val	−0.2	−0.4	0.0
Ile	−0.3	−0.4	−0.1
Leu	−0.4	−0.6	−0.2
Met	−0.6	−0.5	0.0
Phe[d]	0.0	−0.3	0.0

[a] The energies, relative to the reference Ala-containing peptide, were calculated using the all-or-none two state model (as described below) from θ^{222} ellipticities for a series of model peptides published by Munoz and Serrano.[72]

[b] The intrinsic helix-stabilizing contribution of bulky N′ aliphatic residues (Ile, Leu, Val, or Met) was observed even when the N4 residue is Ala, probably because of hydrophobic interactions between the N′ side chain and $C^{\beta}H_{2(3)}$ groups of N3 and N4 residues,[72] similar to helix-stabilizing interactions of nonpolar N-cap residues.[48]

[c] The same energies were applied when Ile or Met occupied the N4 and N7 positions.

[d] The same energies were used for Tyr and Trp residues.

A further helix-stabilizing contribution arises from hydrophobic interactions of flanking N′ and C′ residues with the α-helix.[67] The hydrophobic contact between side chains in the N′ and N4 positions, the ''hydrophobic-staple'' motif,[68] can be detected by nuclear Overhauser effects (NOEs) between the side chains in peptide α-helices.[63,69−71] The corresponding contributions to helix energy (Table II) were estimated using the two-state model from θ^{222} ellipticities published by Munoz and Serrano[72] for a series of model peptides, and it was assumed that the ''hydrophobic staple'' motif can exist only in combination with characteristic N-capping residues (Ser, Thr, Asn, Asp, or Gly). These energies are smaller by 0.1−0.4 kcal/mol than was estimated using the AGADIR program.[72]

A similar hydrophobic interaction at the C-terminus of the α-helix between the bulky C′ and C4 or C1 side chain when the C-cap residue is Gly,[73] the ''Schellman motif,'' is very common for proteins,[61,74] and has also been found in crystal structures of α-helical peptides[75] and detected by nmr spectroscopy for peptides in the presence of SDS micelles[76−79] or trifluoroethanol.[71,80] The Schellman motif is only marginally stable in water: no NOEs between the C′ and C1 or C4 side chains were observed in a model peptide, though analysis of CD spectra indicated that the interaction does contribute a little to α-helix stability.[80] The optimum adjustable energy for the Schellman motif (considered as a single parameter for all combinations of bulky Val, Phe, Tyr, Lys, Arg, Leu, Ile, Met, or Trp residues, but excluding contacts of two positively charged side chains) was identified as −0.3 kcal/mol.

Interactions Between Side Chains in the α-Helix. *H-Bonding and Electrostatic Interactions of Side Chains.* The helix-stabilizing interactions of polar side chains in water, arising from their hydrogen bonding and electrostatic attraction, have been investigated using several model peptides[36−38,81,82] and range up to −0.5 kcal/mol. Based on the published data, the following estimations of the interaction energies with completely ionized side chains were used in the present work: Glu^{i}–Lys, $Arg^{i\pm4}$: −0.5 kcal/mol; Asp^{i}–Lys, $Arg^{i\pm3}$, Asp^{i}–Arg^{i+4}, Gln^{i}–Asn, Asp, Glu^{i+4}; and Glu^{i}–Asn^{i+4}, and Lys^{i}–Asp^{i+4}: −0.4 kcal/mol; Asp^{i}–Arg, Lys^{i+3}: −0.3 kcal/mol. All these pairs are present in α-helical proteins. Two more hydrogen bonding side-chain pairs were taken into account with a tentative assigned energy of −0.5 kcal/mol: the Glu, Asp^{i}–Trp^{i+4} pair present as Glu^{136}-Trp^{140} in colicin (1col, the four-letter codes indicate names of Protein Data Bank files[83]), and Asp^{87}–Trp^{91} in interleukin 4 (1rcb; the Trp residue is in the last turn of each of the helices and has $\chi^1 \sim -60°$), and the Ser^{i}–Gln^{i+1} pair that can be present only at the C-terminus of the α-helix (as Ser^{245}–Glu^{246} of thermolysin, 2tmn), or immediately preceding a Pro-induced helix kink (as Ser^{215}–Gln^{216} in glucoamylase, 3gly, and Ser^{21}–Gln^{22} in cytochrome c′, 2ccy).

The total contribution of the interactions between the polar side chains was calculated assuming additivity of their pairwise $\Delta\Delta G_{ij}^{sch}$ energies:

$$\Delta G_{hb}^{sch} = \sum_{i=k}^{k+m-4} \sum_{j=i+3,4} \Delta\Delta G_{ij,hb}^{sch} \qquad (7)$$

This additivity approximation means that each side chain *i* interacts simultaneously with all surrounding side chains (in $i-4$ and $i+4$ positions, for example) rather than adopting any fixed orientation in solution. This situation can be expected for long flexible Lys, Arg, Glu, and Gln side chains, but not for aromatic side chains since they have preferred χ^1 orientations in the α-helix, as discussed below, and therefore do not interact simultaneously and equally well with other side chains in opposite $i+4$ and $i-4$ directions.

Hydrophobic Interactions Between ''Rotationally Frozen'' Side Chains. The energies of hydrophobic interactions between side chains in water can be estimated

Table III Energies of Hydrophobic Interactions (kcal/mol) Between Side Chains in Different Positions (*i*, *i* + 1, *i* + 3, and *i* + 4) and χ^1 Orientations (''Forward'' or ''Back'')[a] Estimated from Decrease of Nonpolar Accessible Surfaces of the Side Chains

			i Forward			
i + 4 Back	Phe, Tyr	Trp	His	Leu	''$\gamma - \delta$''[b]	''γ''[b]
Phe, Tyr	−0.74	−0.74	−0.34	−0.88	−0.44	−0.44
Trp	−1.08[c]	−0.81[c]	−0.88[c]	−0.93[c]	−0.64[c]	−0.64[c]
His	−1.15	−0.94	−0.27	−0.64	−0.50	−0.50
Leu	−0.70	−0.79	−0.65	−0.98	−0.48	−0.48
''$\gamma - \delta$''[b]	−0.77	−1.03	−0.47	−0.98	−0.48	−0.48
''γ''[b]	−0.61	−0.63	−0.31	−0.20	−0.06	−0.06
Trp^{i-4} Forward	−1.08	−0.88	−0.88	−0.81	−0.76	−0.43
i + 1 Back	Phe, Tyr	Trp	His	Leu	$\gamma - \delta$	γ
Phe, Tyr	−0.88	−0.41	−0.34	−0.45	−0.45	−0.16
Trp	−1.15	−0.27	−0.74	−0.37	−0.19	0
His	−0.54	−0.61[c]	−0.27	0	0	0
Leu	−0.20	−0.64	0	0	0	0
γ and $\gamma - \delta$	0	−0.33	0	0	0	0
i + 3 Forward	Phe, Tyr	Trp	His	Leu	$\gamma - \delta$	γ
Phe, Tyr	−0.14	−1.01	0	−0.25	0	0
Trp	−0.20	−1.15	−0.68[c]	−0.41	0	0
His	0	−0.27	0	−0.26	0	0
Leu and $\gamma - \delta$	−0.38	−0.56	−0.25	−0.34	0	0
			i Back			
i + 3 Back	Phe, Tyr	Trp	His	Leu	$\gamma - \delta$	γ
Phe, Tyr	−0.47	−0.88	−0.47	−0.41	0	0
Trp	−0.34	−0.61	−0.14	−0.21	0	0
His	−0.14	−0.61	−0.14	−0.30	0	0
Leu and $\gamma - \delta$	−0.10	−0.38	−0.25	−0.40	0	0

[a] Forward: $\chi^1 \sim 180°$ (−60° for Thr); back: $\chi^1 \sim -60°$ (60° for Thr). For Val side chain ($\chi^1 \sim 180°$), $C^{\gamma 1}H_3$ group is oriented forward, and $C^{\gamma 2}H_3$ group is oriented back. For Ile side chain ($\chi^1 \sim -60°$), $C^{\gamma 1}H_2 - C^{\delta 1}H_3$ ($\gamma - \delta$) group is oriented forward, and $C^{\gamma 2}H_3$ (γ) group is oriented back.

[b] The γ denotes $C^{\gamma}H_{2(3)}$ group of Val, Ile, Thr, Gln; or Glu side chains and $S^{\gamma}H$ group of Cys; and $\gamma - \delta$ denotes $C^{\gamma}H_2 - C^{\delta}H_{2(3)}$ fragment of Ile, Lys, and Arg residues (and the entire side chain of Met).

[c] The interaction energy may be overestimated because formation of the contact between the side chains decreases accessibility of polar NH groups in the aromatic ring of Trp or His, an effect neglected in calculations of only nonpolar areas.

based on the decrease of their nonpolar surface area when they are brought in contact, similar to surface calculations for interpretation of mutagenesis data.[84] The accessible surface calculations (as described below) indicate that hydrophobic contacts of side chains are formed mostly by their $C^{\gamma}H_{2(3)}$ and $C^{\delta}H_{2(3)}$ groups and aromatic rings, depending on χ^1 and χ^2 torsion angles. In the α-helix, the χ^1 and χ^2 conformers are constrained: only one or two χ^1 orientations (''forward'' $-\ \chi^1 \sim 180°$, or ''back'' $-\ \chi^1 \sim -60°$) are allowed, and usually there is a single energetically and statistically preferred χ^2 conformer.[44,85,86] Therefore, the energies of hydrophobic interactions of side chains *i* and *j* can be tabulated (Table III) based on the type of interacting nonpolar groups (-$C^{\gamma}H_{2(3)}$, -$C^{\gamma}H_2$-$C^{\delta}H_2$-, Leu, or aromatic side chains), their mutual χ^1 orientations, and the distance in the amino acid sequence (±1, ±3, or ±4). The side chains form contacts in the following situations (Table III): (1) the *i* and *i* + 4 residues are oriented toward each other (''forward'' and ''back,'' respectively, and also forward–forward if residue *i* is Trp); (2) the *i* and *i* + 3 side chains are oriented in the same (forward or back) direction; and (3) the *i* and *i* + 1 side chains are oriented toward each other, and at least one of them has an aromatic ring.

The interaction energy $e_{ij}(l)$, of side chains *i* and *j* in a fixed mutual orientation *l* is given by

$$e_{ij}(l) = C_{s,\mathrm{ali}}\Delta\sigma_{ij}^{\mathrm{ali}}(l) + C_{s,\mathrm{aro}}\Delta\sigma_{ij}^{\mathrm{aro}}(l) \qquad (8)$$

where $C_{s,\mathrm{ali}}$ and $C_{s,\mathrm{aro}}$ and $\Delta\sigma^{\mathrm{ali}}$ and $\Delta\sigma^{\mathrm{aro}}$ are solvation

constants and decrease of accessible surface for aliphatic and aromatic groups, respectively, and l is a variable defining mutual orientation of the side chains ($l = 1$: i forward, j forward; $l = 2$: i forward, j back, and so on). Based on the mole fraction water/cyclohexane transfer energies of hexane and benzene and their accessible surfaces, the $C_{s,\text{ali}}$ and $C_{s,\text{aro}}$ constants can be estimated as -0.020 and -0.025 kcal/mol/Å^2, respectively (with a smaller probe radius for aromatic groups as described below). For contacts between two aromatic side chains, Eq. (8) was modified to take into account that solubility of aromatic groups in aromatic solvents is higher than in aliphatic ones and therefore the solvation constant $C_{s,\text{aro}}$ for cyclohexane underestimates the energy of interaction of two aromatic side chains:

$$e_{ij}(l) = C'_{\text{aro}} C_{s,\text{aro}} \Delta\sigma_{ij}^{\text{aro}}(l) \qquad (9)$$

where the scaling coefficient C'_{aro} is an adjustable parameter. The corresponding interaction energies $e_{ij}(l)$ with the obtained value, $C'_{\text{aro}} = 2.7$, are given in Table III.

The energies of side-chain interactions in α-helices, which were calculated from the decrease of nonpolar accessible surface, are in agreement with experimental estimations from CD spectroscopy data, although some of them differ from the theoretical estimates of Creamer and Rose[87] and Munoz and Serrano.[23] The experimentally determined energies of $\text{Phe}_i-\text{Met}_{i+4}$ and $\text{Met}_i-\text{Phe}_{i+4}$ interactions are -0.75 and -0.54 kcal/mol,[22] or -0.65 and -0.20 kcal/mol,[88] respectively. These are close to the -0.77 and -0.44 kcal/mol for the corresponding Phe_i- ''γ-δ''$_{i+4}$ and ''γ-δ''$_i-\text{Phe}_{i+4}$ interactions in Table III. The energies of $\text{Tyr}_i-\text{Val}_{i+4}$, $\text{Tyr}_i-\text{Leu}_{i+4}$, $\text{Leu}_i-\text{Tyr}_{i+4}$, $\text{Leu}_i-\text{Tyr}_{i+3}$, $\text{Tyr}_i-\text{Leu}_{i+3}$, and $\text{Tyr}_i-\text{Val}_{i+3}$ contacts, identified as -0.49, -0.69 to -0.53, -0.84, -0.58, -0.17 to 0.12, and -0.24 kcal/mol, respectively,[21] differ from the corresponding energies in Table III by less than 0.2 kcal/mol.

An interesting specific case is the pair of bulky interacting aromatic side chains i and $i + 4$ with χ^1 angles $\sim 180°$ and $-60°$, respectively. Despite hindrances between these side chains, they form a helix stabilizing pair.[39] In proteins, the hindrances of the interacting aromatic rings are usually avoided by rotating the φ angle of the second $(i + 4)$ residue from $-55°$ to $-110°$, thus displacing its aromatic ring. Such a distortion of helix geometry is tolerated when the $(i + 4)$ residue is the last one in the α-helix. In 30 α-helical proteins shown in Table VI, pairs of interacting aromatic residues were found 18 times. In 14 of these occurrences, the pairs were present with the $i + 4$ residue in the last (C1) position and with a distorted φ_{i+4} angle $\sim -110°$, exactly as in the structural motif of the $\text{Phe}^8/\text{His}^{12}$ pair in ribonuclease A. The four exceptions are as follows: the $\text{Phe}^{61}/\text{Phe}^{65}$ pair of Met apo-repressor (Phe^{65} is in the C2 position, the next residue is Thr with $\chi^1 \sim +60°$), the $\text{Phe}^{83}/\text{His}^{87}$ pair of erythrocruorin (the His^{87} is in the C4 position and the helix C-turn is distorted by Pro^{89}), the $\text{Phe}^{17}/\text{Tyr}^{21}$ pair of the ''HMG-box'' domain (the helix has a kink induced by Pro^{23}), and the $\text{His}^{48}/\text{Tyr}^{52}$ pair of phospholipase A2 (both side chains are in the middle of the α-helix and form hydrogen bonds with the buried Asp^{99} residue). The φ_{i+4} angle is not distorted only in the case of phospholipase A2, in which the hindrances between the two aromatic rings are reduced by rotating the Tyr side chain by 23° from the equilibrium ($-60°$) position. Therefore, the helix-stabilizing aromatic pairs were taken into account only when residue $i + 4$ was the last one in the helix.

Dynamic Averaging of Hydrophobic Side-Chain Interactions. The pairwise interaction energies (Table III) were calculated for side chains with fixed χ^1 orientations. However, the total contribution of these interactions to stability of the helix is reduced because of flexibility of the side chains. This contribution was calculated for a statistical ensemble of N flexible side chains adopting only two allowed χ^1 orientations and N' ''rigid'' side chains with fixed orientation (such as Val or Ile, for example). Denoting the orientations forward and back as ''↑'' and ''↓,'' respectively, the states of the ensemble can be defined as follows:

State of System	Orientations of Flexible Side Chains 1 2 3 ··· i ··· N	Energy
1	↓ ↓ ↓ ··· ↓ ··· ↓	E_1
2	↑ ↓ ↓ ··· ↓ ··· ↓	E_2
···	···	···
2^N	↑ ↑ ↑ ··· ↑ ··· ↑	E_{2^N}

Then, the total energy $\Delta G_{\text{pho}}^{\text{sch}}$, averaged over the set of the states, is given by

$$\Delta G_{\text{pho}}^{\text{sch}} = E_0 + \frac{1}{Q} \sum_{n=1}^{2^N} E_n e^{-(E_n/kT)} \qquad (10)$$

where E_0 is the sum of interactions between ''rigid'' side chains (which are not included in the set of N flexible side chains),

$$Q = \sum_{n=1}^{2^N} e^{-(E_n/kT)} \qquad (11)$$

and the energy E_n of state n is given by

$$E_n = -\sum_{i=1}^{N} \sum_{j} e_{ij}(n) \qquad (12)$$

where $e_{ij}(n)$ are interaction energies [Eqs. (8) and (9), Table III] of side chain i with side chains j ($j = i \pm 1$,

3, 4), which belong to the same set of N flexible side chains or have fixed orientations.

The total energy ΔG^{sch}_{pho} can be approximately divided into contributions of individual side chains g^{sch}_j:

$$\Delta G^{sch}_{pho} = \sum_{i=1}^{N+N'} g^{sch}_i \qquad (13)$$

where the contributions for flexible side chains are given by

$$g^{sch}_i = \frac{1}{Q} \sum_{n=1}^{2^N} [e^{-(E_n/kT)} \sum_j 0.5 e_{ij}(n)] \qquad (14)$$

The calculations with Eqs. (10)–(14) were simplified by dividing the whole set of N flexible side chains in a helix into smaller subsets of interacting side chains and taking into account constraints on χ^1 conformers in α-helices evident from theoretical conformational analysis,[85] statistical analysis of protein structures,[44,86] and nmr spectroscopy of peptides[77,79,89,90]: there are only two allowed χ^1 orientations for linear side chains (Lys, Arg, Leu, Met, Glu, Gln, Cys), only one χ^1 conformer of β-branched Val and Ile side chains ($\chi^1 \sim 180°$ and $-60°$, respectively) in which one of their γ-substituent groups make contacts forward and another one simultaneously make contacts back, and a preferred conformer of Thr with $\chi^1 \sim -60°$. Moreover, determination of side-chain conformers in peptide α-helices from ^{1}H-nmr spectroscopy data[77,79,89,90] and qualitative analysis of published $i/i \pm 3, 4$ NOEs between side chains[71,91–93] show that aromatic side chains are always oriented forward ($\chi^1 \sim 180°$) in middle helix positions, but back ($\chi^1 \sim -60°$) in C-turn, exactly as reflected in statistics of side-chain conformers in middle and C-turn positions of protein α-helices.[86] A few examples of the ''reversed'' orientations of aromatic side chains in C-turn positions are Tyr[64] of a bacteriorhodopsin 34-65 fragment,[89] Phe[124] of a CheY 110-129 fragment,[92] and Tyr[16] of histocompatibility complex-derived peptide.[93] The *trans* ($\chi^1 \sim 180°$) orientation of aromatic side chains in α-helices is energetically more preferred in vacuo[85]; the change of χ^1 to $-60°$ in C-turn positions can be explained by accessibility effects discussed above. A similar C-turn effect can also be expected for residues with γ-aliphatic groups (Met, Lys, Arg, Leu, Glu, Gln, and Thr) and Cys. Therefore, in calculations with Eqs. (10)–(14), all these side chains in C-turn positions were directed back.

Micelle-Bound Peptides

Bound α-Helix. The energy $\Delta G^{\alpha}(k_i, m_i)$ of α-helices in the micelle bound state relative to coil in water can be calculated by considering formation of the bound helix as a two-stage process: (1) formation of α-helix in aqueous solution, and (2) transfer of the helix from water to the micelle. The transfer energy of the helix depends on its orientation at the micelle–water interface and the depth of its immersion into the micelle (Figure 3). The buried portion of the helix was approximated by the helix arc that can be specified by the phase angle φ, counted from the first (k') residue, and size ω (Figure 3b). A completely buried α-helix with $\omega = 360°$ has an undefined φ angle. The equilibrium (the lowest energy) position of an α-helix freely floating at the water–micelle interface was found by minimization of the total helix transfer energy with respect to the ω and φ variables. The energy of an α-helix in the micelle bound state can be given by

$$\begin{aligned}\Delta G^{\alpha}(k_i, m_i) &= \Delta G^{mch} + \Delta G^{sch}_{int} + \Delta G^{sch}_{hb} + \Delta G^{sch}_{pho} \\ &+ \min_{k',m',\varphi,\omega}\left\{(m'-2)\frac{\omega}{360}\Delta G^{mch}_{tr} + \sum_{\substack{j=k' \\ j\in A(\varphi,\omega)}}^{k'+m'-1} (\Delta G^{sch,\alpha}_{tr,j} - f^{sch} g_j)\right\}\end{aligned} \qquad (15)$$

where the first four terms, describing stability of the α-helix in water, were discussed previously; ΔG^{mch}_{tr} is the transfer energy of a helix backbone NH···CO group[94]; $(m' - 2)\ \omega/360$ is the number of buried NH···CO

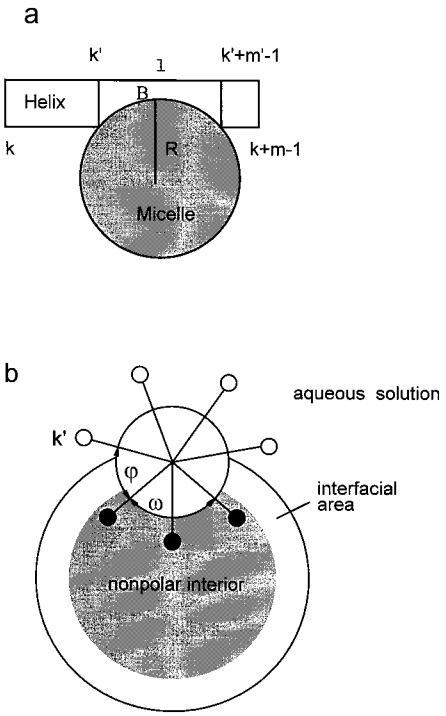

FIGURE 3 The α-helix bound to a spherical micelle of radius R. (a) The α-helix from residue k to ($k + m - 1$) has a buried segment from residue k' to $k' + m' - 1$. (b) The segment immersed into the micelle is approximated by the helix arc, which is defined by its size ω and phase angle φ relative to the 1-st (k') residue.

groups; $A(\varphi, \omega)$ is the set of side chains from the buried helix arc, and $\Delta G^{\mathrm{sch},\alpha}_{\mathrm{tr},j}$ is transfer energy of α-helix side chain j estimated from water/cyclohexane data (Table I). The part of the transfer energy that is already accounted for in the $\Delta G^{\mathrm{sch}}_{\mathrm{pho}}$ term, g_j from Eq. (13), is subtracted. The contribution g_j should be subtracted completely if there is no residual hydrophobic interactions between side chains within nonpolar media. The additional adjustable parameter f^{sch} has been incorporated in the model to account for the possibility of such interactions, which were found to be significant ($f^{\mathrm{sch}} < 1$) for the ''protein droplet.'' Summation from residue k' to $k' + m' - 1$ indicates that the immersed helical segment can be shorter than the whole helix (Figure 3a). The immersion term $\min_{k',m',\varphi,\omega}\{\cdots\}$ must be < 0, otherwise the α-helix does not gain energy from interaction with the micelle and is treated as not bound, with energy defined only by the first four terms in this equation. Since Eq. (15) includes optimization with respect to the variables k', m', φ, and ω, it automatically gives the buried segment and arc of the α-helix.

A different energy expression for an α-helix in the micelle-bound state could be obtained by considering the coil at the water/membrane (or micelle) interface as the reference state[95] and by using special ''interfacial'' transfer and α-helix propensity scales.[96,97] The coil in aqueous solution was chosen here as the reference state in order to use well-studied parameters of α-helix formation in water.

For the specific case of a completely immersed α-helix ($\omega = 360°$) with fixed length m', the last ''transfer'' term in Eq. (15) is related to averaged hydrophobicity, $\langle \Delta G^{\mathrm{sch}}_{\mathrm{tr}} \rangle$, used for hydrophobicity profiles,[98] with a window size of m' residues:

$$\langle \Delta G^{\mathrm{sch}}_{\mathrm{tr}} \rangle = \frac{1}{m'} \sum_{j=k'}^{k'+m'-1} \Delta G^{\mathrm{sch},\alpha}_{\mathrm{tr},j} \qquad (16)$$

and, for the case of $\omega = 180°$ and $m' = 19$, it is related to the ''strip-of-helix index,''[99] $\Delta G_{1/2}$, describing average hydrophobicity of a side of an amphiphilic helix, given by

$$\Delta G_{1/2} = 1/6 \min_{\varphi}\left\{ \sum_{j=k',\, j \in A(\varphi,180°)}^{k'+18} \Delta G^{\mathrm{sch},\alpha}_{\mathrm{tr},j} \right\} \qquad (17)$$

However, it must be emphasized that the stability of the α-helix depends on the *total* transfer energy of the side chains, as in Eq. (15), not on the *average* energy, as in Eqs. (16) and (17). The transfer energy of the helix arc in Eq. (15) is not related directly to hydrophobicity moment[100] since the moment depends on amino acid composition at the water-exposed surface of the helix: the Ser for Asp replacement there, for example, would change the hydrophobicity moment, since transfer energies of these residues are different, but does not change the transfer energy of the *buried* helix arc, affecting only the intrinsic energy of the helix [first four terms in Eq. (15)]. A combination of the intrinsic stability and transfer energy terms for the α-helix, recently investigated by protein engineering,[84] was implemented by Ptitsyn and Finkelstein in the ALB program,[101] but without numerical values of transfer energies and assuming that the buried arc of the α-helix must be 180°.

For geometrical reasons, a short helix cannot be immersed too deeply—the ω angle depends on the length l of the helix segment (k' to $k' + m' - 1$) embedded into the micelle (Figure 3a). The corresponding matching condition for a helix in a spherical micelle was introduced as the inequality $\omega < \omega_{\max}$, assuming that the maximum helix arc $\omega = \omega_{\max}$ at the given length l corresponds to the point B in Figure 3a. Then the geometrical relationship between the length l of the helix of radius r immersed into a micelle of radius R and the $\omega_{\max}$ is given by

$$l = 2\sqrt{R^2 - (R - d)^2} \qquad (18)$$

where

$$d = (R + r) - \left[r \cos\left(\frac{\omega_{\max}}{2}\right) + \sqrt{R^2 - r^2 \sin^2\left(\frac{\omega_{\max}}{2}\right)} \right]$$

The radius R of sodium dodecyl sulfate (SDS) and dodecylphosphocholine (DPC) micelles was chosen as 18.5 Å, though it can vary depending on the detergent, ionic strength, and other conditions,[102] and the radius r of α-helix was chosen as 3.5 Å. The latter is simply an intermediate value between 2.3 Å, the radius of an α-helix backbone,[103] and 4.7 Å, one half of the average interhelical distance in proteins used in geometrical analysis of helix packing.[104]

Bound Coil. The nmr spectroscopy studies of linear peptides forming nonregular structures in complexes with micelles[105–109] and the thermodynamics of their binding[95,110,111] demonstrate that the peptides are located at the water–micelle interface with water-exposed hydrophilic groups and small clusters of partially buried hydrophobic side chains detectable from nonhelical medium-range NOEs. The clusters are still rather flexible judging from the significant dispersion of conformations calculated using the distance geometry algorithm or molecular dynamics simulations with nmr-derived constraints. Based on this general picture of a micelle-bound coil, its possible states were approximated by different combinations of side chains completely buried in the micelle interface (''in'') or completely exposed to water (''out''; Figure 4). The ''in'' and ''out'' orientations of side chains are related to the ''bent'' or ''extended'' configuration of four consecutive C^{α} atoms, which corresponds to dihedral angle C^{α}_{i-1}-C^{α}_{i}-C^{α}_{i+1}-C^{α}_{i+2} close to 0° or 180°, respectively. In the bent (β-turn-like) configuration, the C^{α}-C^{β} bonds of residues i and $i + 1$ have approximately the same direction, and two adjacent i and $i + 1$ side chains can

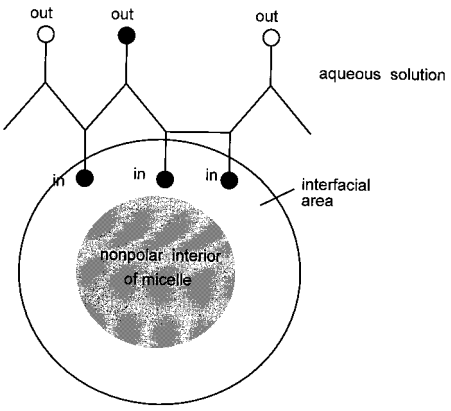

FIGURE 4 Micelle-bound fragment of coil. ''In'' and ''out'' denote side chains completely removed from and exposed to water, respectively. The solid circles are hydrophobic side chains.

be buried within the interface, while in the unfolded configuration, the side chains have opposite orientations, and only one of them can be buried. Immersion of 3 or more *sequential* coil side chains was forbidden in the model because two corresponding sequential bent configurations of C^{α}-C^{α} bonds would form a turn of helix. Energy of a coil fragment of m sequential hydrophobic residues, starting from residue k, is given by

$$\Delta G^{\text{coil}}(k, m) = \min_l \{0, \Delta G_l^{\text{coil}}\} \tag{19}$$

where ΔG_l^{coil} is the energy of coil fragment in in–out configuration l given by

$$\Delta G_l^{\text{coil}} = \sum_{j \in B_l} [f^{\text{coil}} \Delta G_{\text{tr},j}^{\text{sch,coil}} + (\Delta H^{\text{coil}} - T\Delta S^{\text{coil}})] \tag{20}$$

where B_l is the set of buried (in) side chains in configuration l, $\Delta G_j^{\text{sch,tr}}$ are transfer energies of side chains in the extended chain (as described below), f^{coil} is the average fraction of burial of side chains at the micelle–water interface (''the fraction of the hydrophobic free energy consumed at the interface,'' according to the definition of Jacobs and White[95]), the term $(\Delta H^{\text{coil}} - T\Delta S^{\text{coil}})$ includes all additional (mostly unfavorable) contributions to free energy of bound coil, such as the decrease of conformational freedom of a coil residue when it is embedded into the micelle, possible ''perturbation energy'' of the micelle, or partial dehydration of backbone peptide groups adjacent to the buried side chain. The f^{coil}, ΔH^{coil}, and ΔS^{coil} are adjustable parameters of the model. $\Delta G^{\text{coil}}(k, m) < 0$, otherwise the fragment of coil was considered as not bound, with zero energy.

The ''Droplet-Like Protein'' Model

The droplet-like protein model assumes that each α-helix or fragment of coil can be considered as floating in a nonpolar spherical droplet created by the rest of the protein (Figure 1c) to calculate the lowest energy helix–coil partition (Figure 2) using energy expressions as for peptide–micelle complexes, but with modified adjustable parameters. The geometrical matching condition (18) was applied to a spherical droplet of radius $R = (3V/4\pi)^{1/3}$, with the volume V calculated as a sum of volumes[9] of individual hydrophobic (with $\Delta G_{\text{tr}}^{\text{sch}} < 0$) residues of the protein. Calculations of the radius R from the total protein volume may cause discrepancies when applied to proteins of variable shape, dimers, and multidomain proteins. In dimers, like Trp repressor, the total volume of the globule may be twice that of the individual protein, while in multidomain proteins, in contrast, many α-helices may be formed by smaller domains. Fortunately, the results were found to be only weakly sensitive to the radius R.

Averaging of Helix–Coil Partitions

The parameter related to $(i, i + 3)$ NOE cross-peak intensities is occupancy $\langle P_i^{\alpha} \rangle$ of the corresponding helix turn from residue i to $i + 2$, which can be calculated by Boltzmann averaging of helix–coil partitions for the whole peptide molecule:

$$\langle P_i^{\alpha} \rangle = \frac{1}{Q} \sum_{l'} e^{-\Delta G_{l'}/kT} \tag{21}$$

where the index l' includes only partitions containing the helix turn from residue i to $i + 3$ as a part of any helix, and

$$Q = \sum_{l} e^{-(\Delta G_l/kT)} \tag{22}$$

where index l includes all possible helix–coil partitions, and ΔG_l is the unfolding free energy of partition l. Only partitions with one and two helices were taken into account.

The calculated helices were defined as continuous systems of helix turns with occupancies $\langle P_i^{\alpha} \rangle$ exceeding a detectability cutoff P_d, which is the same for all peptides considered. It was assumed that every helix turn with $\langle P_i^{\alpha} \rangle > P_d$ would be observed by the presence of $i/i + 3$, 4 NOEs, but that every helix turn with $\langle P_i^{\alpha} \rangle < P_d$ would not be detected, although, in fact, some marginally stable helices may or may not be detected, depending also on peptide concentration (usually, ≥ 1 mM), and on mixing time and signal/noise ratio in a particular NOESY experiment. The cutoff P_d is an adjustable parameter.

The Lowest Energy Helix–Coil Partition

The lowest energy $E_{\text{min}}^{\text{tot}}$ and corresponding helix–coil partition of a protein molecule of M residues was calculated in a recurrent manner, considering its fragments growing from the C- to N-terminus, $[M - 1, M]$, $[M$

$-2, M], \ldots, [M-n, M], \ldots, [1, M]$, and calculating corresponding lowest energies for each fragment $E_{\min}^{(1)}$, $E_{\min}^{(2)}, \ldots, E_{\min}^{(n)}, \ldots, E_{\min}^{(M-1)} = E_{\min}^{\text{tot}}$, where the lowest energy $E_{\min}^{(n)}$ of a protein fragment $[M-n, M]$, in the absence of bound coil, is given by

$$E_{\min}^{(n)} = \min_m \{\Delta G^{\alpha}(M-n, m) + E_{\min}^{(n-m)}\} \quad (23)$$

or in the presence of bound coil, by

$$E_{\min}^{(n)} = \min_{m,l} \{\Delta G^{\text{coil}}(M-n, l) + \Delta G^{\alpha}(M-n+l, m) + E_{\min}^{(n-m-l)}\} \quad (24)$$

where $n = 1, 2, \ldots, M-1$, $\Delta G^{\alpha}(k, m)$ is the energy of α-helix of m residues starting from residue k, and $\Delta G^{\text{coil}}(M-n, l)$ is the energy of bound coil of l residues preceding the α-helix.

Peptide Helices Detected by NMR

The model has been tested using all available ^{1}H-nmr spectroscopy studies of peptides in aqueous solution and in the presence of micelles for which complete signal assignment and listed medium-range NOEs are available, and which have no long-range NOEs suggestive of tertiary structure (Tables IV and V). A few peptides with significant signal overlap or experimental indications of dimerization or aggregation were excluded. Peptides studied in trifluoroethanol and methanol solutions were not used since these solvents may affect all parameters of the helix–coil transition. The experimentally detected helices were identified from the published data solely as uninterrupted sequences of at least two medium-range $d_{\alpha\beta}(i, i+3)$, $d_{\alpha N}(i, i+3)$, $d_{\alpha\beta}(i, i+4)$, or $d_{\alpha N}(i, i+4)$ NOEs. All peptides lacking the NOEs were considered to be nonhelical. Twelve peptides with ambigous systems of NOEs (at the end of Table IV) were analyzed separately. Intraresidue, sequential, and $(i, i+2)$ NOEs and chemical shifts of $C^{\alpha}H$ protons were not used to identify α-helices, since nonhelical conformations of the peptides stabilized by local interactions[112,182] can significantly contribute to these parameters. Sometimes, specific N- or C-capping NOEs help to identify helix boundaries more precisely. The identified locations of α-helices (Tables IV and V) often are tentative because of signal overlap and are usually close to or identical with those in the original publications.

Average Error in Calculation of Helix Boundaries and Estimation of Adjustable Parameters of the Model

The average deviation (in the number of residues) of calculated and experimental boundaries for α-helices in a set of M peptides was given by

$$\langle \Delta n_{\text{hel}} \rangle = \frac{1}{M} \sum_{m=1}^{M} \left[\frac{1}{N_m} \sum_{i=1}^{N_m} \min\{\min_j\{|n_i^{\text{exp}} - n_j^{clc}| + |k_i^{\text{exp}} - k_j^{\text{clc}}|\}, (n_i^{\text{exp}} - k_i^{\text{exp}})\} + \frac{1}{N'} \sum_{j'}^{N'} \min\{\min_i\{|n_{j'}^{clc} - n_i^{\text{exp}}| + |k_{j'}^{\text{clc}} - k_i^{\text{exp}}|\}, (n_{j'}^{\text{clc}} - k_{j'}^{\text{clc}})\} \right] \quad (25)$$

where N_m is the number of experimentally identified helices in molecule m; k and n are numbers of the first and last residues, respectively, in the experimental i and corresponding calculated j (with a minimum deviation from i) helices; and N' is the number of falsely calculated helices. The expression takes into account that calculated and experimental helices sometimes do not match each other. If the entire length, $(n_i^{\text{exp}} - k_i^{\text{exp}})$, of an experimental helix i is less than its deviation $|n_i^{\text{exp}} - n_j^{\text{clc}}| + |k_i^{\text{exp}} - k_j^{\text{clc}}|$ from any calculated helix j, then the deviation of the helix boundaries is equal simply to its length $(n_i^{\text{exp}} - k_i^{\text{exp}})$. The second term in Eq. (25) with j' indices serves to estimate the deviations for N' falsely calculated helices that did not match any experimental helix in the first sum with j indexes. The expression gives significant deviation when a long α-helix is broken into smaller ones in calculations.

Optimization of the $\langle \Delta n_{\text{hel}} \rangle$ deviation with respect to adjustable parameters of the model was done simply by grid scan with gradually decreasing step, since it had been found that optimization methods with finite difference approximation of derivatives do not work reliably due to the multiple minima problem and very shallow dependence of the deviation $\langle \Delta n_{\text{hel}} \rangle$ on some parameters.

Calculation of Nonpolar Contact Areas Between Side Chains in α-Helix

All changes in accessible surfaces of interacting nonpolar side chains were calculated in an $(\text{Ala})_{11}$ α-helix with two interacting i and $i+1$, $+3$, or $+4$ residues incorporated in the middle of the helix, with the exception of aromatic $i, i+4$ pairs, which were in the 6th and 10th positions. An ideal α-helix with the different pairs of ''probe'' side chains was minimized with CHARMm (250 iterations of the adopted-basis Newton-Raphson method) to allow the interacting side chains to form a geometrically optimum contact. Initial χ^1 angles of side chains for energy minimization were $-60°$ or $180°$. Contacts of the ''γ'' type (Table III) were calculated using the Val side chain, and contacts of the ''$\gamma - \delta$'' type, using the Lys side chain with initial angles $\chi^2 = \chi^3 = \chi^4 = 180°$ and taking into account changes of accessible surface only for the $C^{\gamma}H_2$ and $C^{\delta}H_2$ groups. It was assumed that linear side chains of the $\gamma - \delta$ type (Lys, Arg, Ile, Met) have the energetically and statistically prefered $\chi^2 \sim 180°$ conformer.[85,86] The Cys and Met residues were considered simply to be

Table IV Peptides Studied by ^{1}H-NMR Spectroscopy in Aqueous Solution

Peptide	Reference	*T* (K)	pH	Helix Positions[a] Experiment	Helix Positions[a] Calculation
Bovine pancreatic trypsin inhibitor 1–15, 13–21, 16–28, 24–32, 29–44, 41–51	112	271	4.6	None	None
BPTI 45–58	112	271	4.6	4–9	4–9
Annexin 1–32	78	293	6.4	6–12, 19–27	5–13, 18–28
Myoglobin 109–133	113	280	5.0	None[b]	None
Myoglobin 101–118	114	278	4.0	None[b]	None
Myoglobin 124–150	115	278	4.0	2–26	3–26
Myohemerythrin 1–18	115	278	5.1	10–15	11–15
Myohemerythrin 18–39	115	278	5.1	5–18	5–16
Myohemerythrin 40–63, 63–70, 69–87, 86–92, 109–118	115	278	5.1	None	None
Myohemerythrin 93–108	115	278	5.1	3–13	5–13
Plastocyanin 1–8, 11–20, 17–26, 26–37, 30–39, 36–47, 57–63, 61–70, 72–80, 83–93, 92–99	116	278	6.5	None	None
Thermolysin 233–248	91	295	2.5	4–14	**None**
Thermolysin 245–260, 258–276	91	278	2.5	None	None
Barnase 1–21	117	279	4.5	None	None
Adenylate cyclase 532–562	118	295	4.1	20–32	20–30
Adenylate cyclase 225–267	119	298	6.4	11–22	11–24
Adenylate cyclase 196–267[c]	120	298	4.8	22–25	22–28
C-peptide analogue 1–13	69	276	5.2	4–12	2–12
Viral coat protein 1–25	121	275	4.0	9–17	10–18
Viral coat protein 1–25	121	283	4.0	10–16	10–17
Model peptide	122	298	6.3	3–15	1–16
Transthyretin 71–93	123	283	4.0	5–10	3–12
Transthyretin 71–93	123	298	4.0	None	**5–12**
Model peptide	124	278	5.0	1–17	2–17
Platelet-adhesion peptide	125	303	6.5	14–25	14–23
Calcitonin related peptide	126	300	3.7	None	None
Angiogenin	127	283	2.1	None	None
S-peptide	127	283	2.1	None	None
Amyloid β-peptide	128	278	1.0	None	None
Calmodulin-binding peptide	129	288	6.5	None	None
Phospholipase A2 38–59	130	278	2.5	1–13, 15–18	**3–9**
Lysozyme 59–81	131	281	5.0	4–9, 10–20	3–11, 13–18
Salivary histatin	132	303	3.8	None	None
Model peptide	133	278	2.0	5–16	5–15
Model peptide	133	288	2.0	5–14	5–14
Erythrocruorin 1–19	134	277	3.0	3–13	**None**
Erythrocruorin 31–50, 61–80, 71–90	134	277	3.0	None	None
Model peptide	70	283	5.7	4–16	4–16
Model peptide	135	278	4.0	None	None
λ-Repressor 9–23	136	277	7.0	3–15	3–15
Titin 1–38	137	290	7.3	5–20, 25–32	5–23, 25–34
Laminin 641–660	138	288	3.5	None	None
Substance P	139	303	4.0	None	None
Merozoite antigen	140	275	4.9	11–34	12–34
Galanin 1–30	141	276	4.0	None	None

Table IV (*Continued from the previous page.*)

Peptide	Reference	*T* (K)	pH	Helix Positions[a]	
				Experiment	Calculation
Blood coagulation factor 4–11	142	288	5.2	4–11	4–11
Basolateral sorting signal	143	278	4.6	None	None
Che-Y 110–130	71	278	7.0	4–18	7–18
Che-Y 13–31	92	298	2.5	None	None
Che-Y 37–51, 90–106	92	278	7.0	None	None
Che-Y 63–78	92	278	7.0	None	**3–13**
Ferredoxin 62–79	92	278	7.0	None	None
Ferredoxin 91–109	92	278	7.0	5–14	4–14
Ferredoxin 120–139	92	278	7.0	5–14	4–17
Ras protein 63–77	92	278	7.0	None	None
Ras protein 84–106	92	278	7.0	7–13	7–**20**
Ras protein 122–140	92	278	7.0	7–14	11–14
Ras protein 149–167	92	278	7.0	7–11	5–**16**
Model peptide	68	278	6.1	4–11	4–14
Model Ala^{17} peptide	80	278	6.0	5–15	5–14
Model Met^{17} peptide	80	278	6.0	5–14	5–14
Model Phe^{17} peptide	80	278	6.0	5–15	5–14
Model Phe-Cys peptide	88	278	3.0	5–15	5–14
Model Cys-Phe peptide	88	278	3.0	5–14	5–14
Model Phe-Met peptide	88	278	3.0	5–15	5–14
Model Met-Phe peptide	88	278	3.0	5–14	5–14
NO synthase 725–747	144	278	7.0	6–20	7–20
Neuromodulin	145	298	5.0	None	None
Estrogen receptor 500–528	146	277	4.0	2–24	2–22
Protein G 1–20	147	278	3.0	None	None
Protein G 21–40	147	278	5.1	None	None
ComA protein 88–102	148	278	7.0	4–14	5–14
SH3 domain 13–29	149	278	5.4	None	None
Chymotrypsin inhibitor 2, 1–25[d]	150	298	4.6	16–23[e]	13–19
Barnase 1–36[d]	151	279	6.8	30–34[f]	None
Ras protein 13–30[d]	92	278	7.0	6–9[f]	None
Erythrocruorin 11–30[d]	134	277	3.0	13–16[f]	None
Erythrocruorin 80–100[d]	134	277	3.0	5–8[f]	None
Che-Y 113–130[d]	71	278	2.5	3–6, 8–11[g]	4–15
Ferredoxin 8–30[d]	92	278	7.0	8–11, 13–16[g]	6–15
Thermolysin 299–316[d]	91	295	2.5	8–12[h]	None
SH3 domain 56–12[d]	149	278	5.4	12–16[h]	None
SH3 domain 29–55[d]	149	278	3.0	3–6, 9–13[h]	None
Erythrocruorin 51–70[d]	134	277	3.0	11–16	None
Erythrocruorin 90–110[d]	134	277	3.0	3–15, 18–21	None

[a] Significant (≥4 residues) discrepancies of experimentally identified and calculated locations of helices are indicated in bold. The numbering starts with the first residue in each peptide, and the Boltzmann averaging of helix–coil partitions was applied for calculations.

[b] The 109–133 and 101–118 myoglobin peptides were considered as nonhelical, although each of them has one tripeptide segment with $i–i + 3$ NOEs, since the NOEs disappear in the longer 100–150 peptide.

[c] Only the N-terminal 196–232 segment of the long 196–267 peptide of adenylate cyclase was considered in calculations, since the C-terminal 225–267 fragment was studied independently (the previous line of table).

[d] The supplementary set of peptides with some uncertainty in interpretation of nmr data.

[e] The α-helical NOEs were not detected at lower temperature.

[f] Only one medium range NOE was observed for the peptide.

[g] Two single separate (i, $i + 3$) NOEs were observed.

[h] The (i, $i + 3$) and (i, $i + 2$) NOEs were observed between the protons of side chains, but there is no medium-range NOEs of backbone protons.

Table V Peptides Studied by ^{1}H-NMR Spectroscopy in the Presence of Micelles

Peptide	Reference	T (K)	pH	Detergent[d]	Helix Positions[a] Experiment	Helix Positions[a] Calculations
Annexin 1–32	78	293	6.4	DPC	5–15, 19–29	5–15, 18–29
ALDH signal peptide	152	313	5.1	DPC	5–18	2–19
Mellitin	153	303	3.5	DPC	2–26	2–24
Glucagon	154	310	6.0	DPC	11–14, 17–28	9–17, 19–28
Histocompatibility complex peptide	93	283	6.1	DPC	9–16	**2**–16
Glucagon-like peptide	155	310	6.0	DPC	5–15, 18–29	6–12, 15–28
Thiolase 1–21	156	293	5.2	DPC	4–14	2–14
δ-Hemolysin	157	300	7.0[b]	DPC	5–23	2–24
Rhodanese 1–23	158	293	3.3	DPC	2–20	2–21
Sufractant peptide 1–17	159	307	3.6	DPC	10–17	6–17
M13 coat protein	160	311	5.1	SDS	6–20, 24–45	4–18, 24–45
Calcitonin	161	310	3.7	SDS	6–22	6–21
T4 lysozyme 1–13	162	297	5.4	SDS	2–10	3–10
T4 lysozyme 59–81	131	303	4.0	SDS	4–9, 11–22	2–10, 12–22
Bombolitin III	163	313	5.0	SDS	4–16	2–17
Dynorphin A 1–17	164	310	3.2	SDS	4–9	3–13
L6L8 OmpA peptide	165	298	2.9	SDS	4–21	5–22
des-8 OmpA peptide	165	298	2.9	SDS	5–21	5–21
des-6-9 OmpA peptide	165	298	2.9	SDS	4–18	5–18
β-endorphin 12–26	166	323	5.0	SDS	4–11	2–11
Bacteriorhodopsin 1–35	167	308	3.5	SDS	8–32	8–30
Bacteriorhodopsin 34–65	76, 77	303	3.0	SDS	8–29	4–30
Growth hormone releasing factor 15–32	168	298	6.6	SDS	5–18	4–16
Uteroglobin 18–47	169	313	5.5	SDS	7–10, 15–27	4–10, 13–29
Eledoisin	170	298	5.4	SDS	4–11	4–11
Substance P	139	303	4.0	SDS	4–11	4–11
Met-enkephalin	106	303	4.1	SDS	None	None
Bradykinin	109	303	4.0	SDS	None	None
Retro-bombolitin I	171	313	5.0	SDS	2–16	2–16
Antihemophilic factor 2303–2324	172	298	5.5	SDS	3–6, 8–19	2–5, 8–20
Prion protein 109–122	173	298	3.7	SDS	5–14	4–13
Prion protein 109–141	173	298	3.7	SDS	5–26, 29–31	4–27, 29–31
Apolipoprotein C-I 7–24	174	310	4.8	SDS	2–17	2–17
Apolipoprotein C-I 35–53	175	310	4.8	SDS	2–19	2–18
PhoE signal peptide	176	298	1.6	SDS	2–18	5–18
Lipid-associating peptide	177	298	5.0	SDS	2–19	2–19
Presequence peptide p25[c]	178	301	3.8	DPC	4–11	4–12, 22–24
Presequence peptide p25[c]	179	298	4.6	DPG	4–13, 16–24	4–12, 22–24
Mastoparan X[c]	180	318	3.5	DPPC	3–14	3–14
KRR-lamb peptide[c]	181	298	5.0	POPC	13–26	4–8, 11–28

[a] Significant ($\geq$4 residues) discrepancies between experimentally identified and calculated locations of helices are indicated in bold. The numbering starts with the first residue in each peptide, and the Boltzmann averaging of helix–coil partitions was applied for calculations.

[b] Temperature and pH were not indicated in the publication.

[c] This peptide was not included in the main set used for optimization of adjustable parameters of the model.

[d] DPPC: L-α-dipalmitoylglycerophosphocholine; POPC: palmitoyloleoylphosphatidylcholine; DPG: dodecylphosphoglycol; DPC: dodecylphosphocholine; SDS: sodium dodecylsulfate.

of the aliphatic γ and $\gamma - \delta$ types, respectively. For His and Trp side chains, an optimum (i.e., forming maximum contact surface) χ^2 conformer was chosen for every pair of interacting side chains. For the Leu side chain, two preferred combinations of χ^1, χ^2 angles (180°, 60° and −60°, 180°)[85,86] were used. For aromatic pairs, the $i + 4$ residue was in the C2 position, and its initial φ angle was −110°, as discussed above.

The decrease of accessible surfaces of side chains i and j in mutual orientation l, $\Delta\sigma_{ij}(l)$ [Eqs. (8, 9)], for the energetically optimized α-helix, is given by

$$\Delta\sigma_{ij}(l) = [\sigma_i^o(l) - \sigma_i(l)) + (\sigma_j^o(l) - \sigma_j(l)), \quad (26)$$

where $\sigma_i(l)$ is the accessible surface of side chain i when it is in contact with side chain j, and $\sigma_i^o(l)$ is the accessible surface of the same side chain in the same conformation and position in the α-helix but lacking contact with side chain j. All calculations of accessible surfaces were performed using QUANTA.

For aromatic side chains, the $\Delta\sigma_{ij}(l)$ areas were calculated with a smaller probe radius of 0.1 Å to estimate contact surfaces rather than accessible ones.[183] This is because the water–cyclohexane transfer energy consists of two components: the ''truly hydrophobic energy'' reflected in the water-vapor distribution coefficients, and the transfer energy from vapor to cyclohexane arising from dispersion attraction of the solute with the cyclohexane. For aliphatic compounds, each of the water–vapor and vapor–cyclohexane components contribute approximately equally to the total water–cyclohexane transfer energy, but the transfer energy of nonpolar aromatic compounds (such as toluene, the analogue of the Phe side chain) from water to vapor is close to zero[49]; thus all of the favorable water–cyclohexane transfer energy arises from dispersion attraction to cyclohexane. The calculations with the probe radius of the water molecule (1.4 Å) reflects removal of a group from water. However, part of the aromatic atoms of side chain i removed from water do not directly contact atoms of side chain j, so their dispersion interactions are reduced. The smaller probe radius of 0.1 Å for aromatic groups better reflects direct dispersion contacts responsible for their transfer energies.

All-or-None Two-State Approximation

The replacement energies, $\Delta\Delta G^{sch}$ [Eq. (6)] for residues in N-capping and N′ positions (the ''hydrophobic staple'' motif) were estimated from CD data published for series of substituted peptides[47,63,72] as follows. First, the helix–coil free energy difference ΔG was calculated for each peptide of a series, taking into account that actual α-helices identified from the medium-range NOEs for some model peptides[63,72] are shorter than the whole peptide:

$$\Delta G = -RT \ln[\theta_{hel}/(\theta_{max} - \theta_{hel})] \quad (27)$$

Here θ_{hel}, the mean residue ellipticity appearing from the two-state equilibrium between helix and coil, is given by $\theta_{hel} = \theta^{222}(n/n_{hel})$, where θ^{222} is the measured mean residue ellipticity of the peptide at 222 nm, n is the number of residues in the peptide, n_{hel} is the number of residues in the α-helix identified by system of medium-range NOEs, and θ_{max}, the ellipticity of the helix of n_{hel} residues, is given by $\theta_{max} = \theta_\infty/(1 - 2.5/n_{hel})$, where $\theta_\infty = -40{,}000$. Then, the replacement $\Delta\Delta G^{sch}$ energies relative to the reference Ala-containing peptide of the series were calculated.

Transfer Energies of Side Chains

Side-chain transfer energies $\Delta G_{tr}^{sch,\alpha}$ were estimated from mole-fraction based water–cyclohexane transfer energies of their analogues[49] (isobutane for Leu, toluene for Phe and so on), ΔG_{tr}^{anal}, taking into account that the analogues lose part of their accessible surface when incorporated into an α-helix. For nonpolar side chains (aromatic, aliphatic and Cys), the corrected transfer energy is given by

$$\Delta G_{tr}^{sch,\alpha} = \Delta G_{tr}^{anal}(\sigma^\alpha/\sigma^{anal}) \quad (28)$$

where σ^{anal} is the accessible surface (with a probe radius of 1.4 Å) of a side-chain analogue, and σ^α is the accessible surface of the corresponding side chain in an α-helix averaged over a set of possible χ^1, χ^2 conformers (with $\chi^3 = \chi^4 = 180°$ for Lys, Arg, and Met):

$$\sigma^\alpha = \sum_i p_i \sigma_i^\alpha \quad (29)$$

where p_i is the occupancy of allowed side-chain conformer i in an α-helix, calculated from statistical data,[44] and σ_i^α is the accessible surface of the side chain of conformer i in the α-helix. For polar side chains, instead of Eq. (28), the transfer energy is given by

$$\Delta G_{tr}^{sch,\alpha} = \Delta G_{tr}^{anal} - C_s(\sigma^{anal} - \sigma^\alpha) \quad (30)$$

where $(\sigma^{anal} - \sigma^\alpha)$ is the buried nonpolar surface of the aliphatic (CH_2-CH_2-···) part of the polar side chain, and C_s is the solvation constant (−0.02 kcal/mol/Å^2). The transfer energies of side chains in a coil conformation were estimated similarly but using dynamically averaged accessibilities of side-chain atoms,[184] instead of the accessible surface σ^α calculated with Eq. (29).

RESULTS AND DISCUSSION

The calculations of secondary structure were done first for 96 and 36 peptides studied by ^{1}H-nmr spectroscopy in aqueous solution and in the presence of micelles, respectively (Tables IV and V), and

second, for 30 α-helical proteins (Table VI). In both cases, the average deviation of helix boundaries between experimental and calculated helices was minimized with respect to a few adjustable parameters. The deviations and parameters thus determined are presented in Table VII.

Peptides in Aqueous Solution

Agreement with NMR Spectroscopy Data. The thermodynamic model of α-helix formation works satisfactorily for the main set of 96 peptides in aqueous solution with straightforwardly interpreted nmr data (Table IV): the average error in the calculation of helix ends is 1.3 residues per helix, corresponding to an average of 93% correctly calculated helix–coil states (Table VII). Of 48 α-helices in the set, 3 were missed and, additionally, 2 helices were falsely predicted. The discrepancies can be explained by approximate parametrization of intrahelical interactions, the use of the same ''detectability cutoff'' for different nmr experiments, or by uncertainties in identification of α-helix ends from medium-range NOEs.

An additional set of 12 peptides (at the end of Table IV) reflects situations in which the interpretation of nmr data is less clear. Several peptides with only one $(i, i + 3)$ or $(i, i + 4)$ NOE (barnase 1-36, ras 13-30, and two erythrocruorin fragments) were calculated to be nonhelical. At the same time, two peptides, each with two interrupted $(i, i + 3)$ NOEs (Che-Y 113-130 and ferredoxin 8-30), were calculated as forming α-helices (Table IV). Three peptides (thermolysin 299-316 and SH3 domain fragments) have NOEs between side-chains of i and $i + 3$ residues, but no α-helical NOEs involving backbone $C^{\alpha}H_i$ protons. This can be expected for β-turns formed by residues i to $i + 3$, where the $C^{\alpha}H_i$ proton does not participate in any NOE contacts (the C^{α}-H_i bond goes in the direction opposite to that in the α-helix because of a positive ψ_i angle) but the side chains i and $i + 3$ are close to each other. No α-helices were calculated in this case (Table IV). Thus, all these a priori ambigous cases actually are consistent with the computational results. However, a more interesting situation occurs for several erythrocruorin fragments. In the absence of tertiary interactions and aggregation, all α-helices formed by short peptides should remain the same when the peptides are included into a longer fragment, as can actually be seen for peptides 109-133, 124-150, and 100-150 of myoglobin,[113,114,218] 225-267 and 196-267 of adenylate cyclase,[119,120] and 1-36, 34-65, 1-71, 190-233, and 163-231 of bacteriorhodopsin.[77,79,90,219] However, the situation for erythrocruorin peptides is different: the α-helical medium-range NOEs of residues 61–67 are observed in the 51-70 fragment of erythrocruorin, but disappear in the 61-80 fragment studied under the same conditions. Similarly, residues 90-100 of erythrocruorin appear as α-helical in the 90-110 peptide but not helical in the 80-100 peptide. In both cases, the α-helical medium-range NOEs are present in peptides better matching the long amphiphilic helices 53-71 and 94-111 of erythrocruorin. However, calculations show no detectable α-helices in the peptides (Table IV), contradicting the NMR data, perhaps because the ''fragment-specific'' α-helices are stabilized by some tertiary interactions within the peptides or by their dimerization.

Adjustable Parameters of the Model. Nearly all parameters of the model were taken from previously published experimental data or estimated from accessible surface calculations (for side-chain interactions). There were only eight adjustable parameters determined by minimization of the helix boundary deviation, calculated for the main set of 96 peptides (Table IV). The parameters are enthalpy and entropy of α-helix formation by the host polyAla peptide (-1.24 kcal/mol and 4.1 cal/mol/K, respectively), ''detectability cutoff'' (0.18), contribution of the Schellman motif (-0.3 kcal/mol), the energy of electrostatic interaction of completely ionized, positively charged residues in C-cap position with the helix dipole (-0.4 kcal/mol), the scaling coefficient C'_{aro} for aromatic interactions from Eq. (9) (2.7), and the energy of electrostatic interaction of negatively and positively charged side chains in N-turn positions with the helix dipole (-0.9 and $+0.5$ kcal/mol, respectively).

The per residue enthalpic ΔH and entropic $T\Delta S$ terms could be separately defined since nmr studies were done at a variety of temperatures, from 271 to 303 K. The enthalpy of interaction found for two peptide groups in the polyalanine α-helix ($\Delta H = -1.24$ kcal/mol) differs from that for polyglutamic acid (-1.12 kcal/mol) and polylysine (-0.885 kcal/mol)[41] but is close to the -1.3 kcal/mol found for a model alanine-based peptide.[42] The corresponding main-chain free energy contribution per residue ($\Delta G = \Delta H - T\Delta S$) is only -0.12 kcal/mol at 273 K, slightly less than the -0.26 kcal/mol estimated by Chakrabatty et al.[53]

The detectability cutoff that was obtained, $P_d = 0.18$, indicates that helix turns with occupancy $< 18\%$ usually cannot be detected by medium-range NOEs. The P_d cutoff can be compared with

estimations of helix occupancies from nmr and CD spectroscopy data. All peptides with high α-helix content (>50% from CD data) have clear patterns of α-helical medium-range NOEs[69,124,133] and were calculated here as α-helical with occupancies of individual turns, $\langle P_\alpha \rangle = 0.30-0.69$. However, α-helices formed by peptides of myohemerythrin and bovine pancreatic trypsin inhibitor, with relatively low calculated occupancies ($\langle P_\alpha \rangle = 0.17-0.28$) but with clear patterns of $i/i \pm 3$, 4 NOEs, are not visible in CD spectra at all[112,115]; thus no comparison can be made in these cases. At the same time, estimations of helix occupancies from chemical shifts are consistent with $P_d \sim 18\%$: for example, α-helicities of the 110-130 and 113-130 fragments of CheY, which form α-helices detected in calculations (Table IV), were estimated from chemical shifts as 30 and 21%, respectively,[130] while α-helicities of five thermolysin peptides found to be nonhelical in the calculations were estimated as $\leq 18\%$.[91]

Formation of α-Helices by Protein Fragments. Experimental studies of many peptides summarized in Table IV show that formation of short α-helices by protein fragments is a very common phenomenon. Calculations for 60 all-α, α/β, α + β, and all-β proteins, using the theoretical model described here, lead to the same conclusion: approximately 50% of protein α-helices could be identified by nmr spectroscopy of the properly chosen protein fragments at low (278 K) temperature and pH 7, though the peptide α-helices usually are shorter than in proteins and less stable than coil (the calculations were done for overlapped protein fragments of 40 residues, shifting the 40-residue ''frame'' by a 5-residue step through the amino acid sequence of the protein; the results are not presented here). Rarely, the calculated peptide α-helices are longer than in the native protein structure: 37-64 of myoglobin (5mbn), 45-71 of Trp repressor (3wrp), 10-37 of myohemerythrin (2mhr), and 9-39 of cytochrome b5 (3b5c), or are formed by nonhelical segments of proteins: 81-85 of histone 5 (1hst), 407-414 of glucoamylase (3gly), 209-218 of cytocrome c peroxidase (2cyp), 19-34 of aspartate receptor (2asr), 38-48 of ubiquitin (1ubq), 82-91 of immunoglobulin (2rhe), and 24-29 of interleukin-1 β (4ilb). The 10-37 helix formed by the 1-40 peptide of myohemerythrin was detected by nmr as helices 10-15 and 22-35 formed by the shorter 1-18 and 18-38 peptides.[115]

The marginal stability of α-helices reflects the important contribution of the intrinsic α-helix energy to stabilization of the entire protein structure. The intrinsic helix-forming potency is especially strong (a significant part of protein α-helices could be detected by nmr in peptide fragments) for proteins that have small hydrophobic cores or subdomains or undergo conformational transitions, such as colicin A (1col), citrate synthase (2cts), Trp repressor (3wrp), annexin V (2ran), Ca^{2+}-binding proteins (4cpv, 3cln, 4icb), and all rod-like 4-α-bundle proteins (2ccy, 256b, 2mhr, 1lpe, 1fha, and 2asr) with the notable exception of tobacco mosaic virus coat protein (2tmv). Probably, the helix-stabilizing amino acid substitutions can be accumulated during evolution at the surface of any protein, but the process depends on the selection pressure, which is high when other ''folding forces'' are insufficient to stabilize the protein 3D structure (if the hydrophobic core is small, for example), and also on the stereochemical limitations for the surface residues: they cannot be replaced to increase intrinsic helix stability if they participate in specific quarternary interactions, such as packing of each viral coat subunit with other subunits and RNA. The peptide fragments of extracellular disulfide-rich α-helical proteins, such as mating pheromone (1erl), hydrophobic seed protein (1hyp), phospholipase A2 (4p2p), and endothelin 1 (1edp), form almost no α-helices, probably because the proteins are sufficiently stabilized by their disulfide bonds, and therefore the intrahelical interactions are less important.

Peptides in the Presence of Micelles

The thermodynamic model of α-helix formation for peptides in the presence of micelles successfully reproduces locations of the helices identified from published nmr spectroscopy data as uninterrupted sequences of at least two of the $d_{\alpha\beta}(i, i+3)$, $d_{\alpha N}(i, i+3)$, $d_{\alpha\beta}(i, i+4)$, or $d_{\alpha N}(i, i+4)$ NOEs (Table V). There are no missed or falsely calculated helices for the 36 micelle-bound peptides, and the average error in calculation of the helix boundaries is ~ 2 residues per helix (Table VII).

It is usually assumed that formation of amphiphilic helices by peptides depends mostly on the sequence pattern of hydrophobic and hydrophilic residues. However, the intrinsic α-helix energy contribution was also crucially important, since neglect of the corresponding terms in Eq. (15) resulted in significantly poorer agreement with the nmr-detected locations of helices. Surprisingly, the absolute values of the intrinsic and transfer energy terms were consistent with each other and could be combined together without incorporation of any additional weight factors in Eq. (15); attempts to incor-

Table VI Comparison of Helices in Proteins and the Calculated Lowest Energy Helix–Coil Partitions[a]

Protein, PDB Code, and Reference	Actual (Top Line) and Calculated (Bottom Line) Locations of Helices[b]					
434 Repressor, 1r69[185]	1–13	17–24	29–36	45–52	56–61	
	2–13	17–24	31–38	44–52	56–61	
Endonuclease III, 1abk[186]	3–16	29–38	44–57	61–67	69–76	81–99
	3–15	29–40	44–54	61–67	69–79	85–98
	108–113	119–130	139–148	156–166	175–185	200–202
	105–114	**126**–130	137–145	156–167	**169**–182	199–**208**
Uteroglobin, 1utg[187]	4–15	18–25	32–45	50–65		
	2–15	20–28	32–48	54–66		
Citrate synthase, 2cts[188]	6–28	38–42	60–62	71–77	89–98	104–117
	6–**14**	38–**52**	60–**67**	71–78	88–97	104–117
	122–130	137–150	153–160	165–194		209–217
	120–131	138–148	150–**166**	168–182, 185–193		209–217
	222–235	243–254	258–270		275–292	298–311
	211–235	**251**–255	258–**260**, **262**–270		279–292	298–309
	328–340	345–364	374–384	391–415	427–436	Loop
	331–340	345–367	374–388	392–415	427–434	**200**–**207**
Trp repressor, 3wrp[189]	9–32	35–42	45–63	68–75	79–91	
	12–29	35–43	45–63	68–75	79–**82**, **88**–91	
	94–104					
	93–106					
Fis protein, 4fis[190]	1–14	24–44		48–55	59–69	
	3–16	24–32, 36–43		50–57	59–69	
λ-Repressor, 1lmb[191]	9–29	33–40	44–52	59–69	Loop	78–91
	9–29	33–40	**49–57**	59–71	73–76	78–91
Colicin, 1col[192]	8–30		32–46	56–67	76–88	91–101
	8–**22**, **24**–30		32–48	56–63	79–89	91–105
	110–126	131–143	147–164		169–187	190–198
	107–120	130–143	148–**158**, **160**–165		167–187	189–195
	Loop					
	70–**77**					
Homeodomain, 1hdd[193]	10–23	28–38	42–58			
	12–22	24–40	44–58			
Histone 5, 1hst[194]	10–20	29–39	46–60	β		
	10–24	31–39	46–60	**75**–**79**		
Seed protein, 1hyp[195]	12–18	25–34	37–51	56–67		
	11–22	25–35	37–51	53–67		
Parvalbumin, 4cpv[196]	3–5	8–17	26–33	40–50	60–69	79–89
	2–6	8–21	27–33	40–54	60–**77**	79–89
	99–107					
	99–107					
Calmodulin, 3cln[197]	5–19	29–39	45–55	65–92		102–112
	6–19	29–39	45–55	65–**76**, **82**–92		102–116
	118–128	138–146				
	118–128	138–145				
Calbindin, 4icb[198]	4–15	β	26–36	38–40	47–54	64–75
	4–14	**21**–**24**	26–33	36–41	47–54	64–74
Myoglobin, 5mbn[199]	4–19	21–35	37–42	44–48	52–57	59–78
	4–17	21–34	36–43	45–50	52–**64**	**68**–79
	83–97	101–118	125–149			
	86–**90**	103–116	125–149			

Table VI (*Continued from the previous page.*)

Protein, PDB Code, and Reference	Actual (Top Line) and Calculated (Bottom Line) Locations of Helices[b]					
Erythrocruorin, 1ecd[200]	3–17	20–30	32–37	46–50	53–72	77–90
	3–17	20–30	32–**42**	46–51	56–73	76–87
	94–111	115–135				
	94–111	117–133				
Leghemoglobin, 1lh1[201]	5–20	22–36	38–43	45–47	58–81	88–100
	15–22	**29**–37	Missed	43–50	58–79	87–**93**
	107–122	128–152				
	112–125	128–149				
Cytochrome *c*, 3c2c[202]	4–**13**, **15**–17		Loop	50–58	64–72	74–82
	3–18		28–36	52–58	64–70	74–77
	Loop	98–110				
	93–**97**	100–111				
Mating pheromone, 1erl[203]	2–**9**, **12**–18		23–35			
	2–19		21–33			
E3-binding domain, 1bb1[204]	14–23	27–30	41–47			
	12–23	27–30	41–50			
Met-aporepressor, 1cmb[205]	*β*	30–46		53–66	73–76	86–95
	19–24	28–**36**, **38**–49		54–65	73–80	83–95
Fragment B of protein A[206], 1fc2[c]	5–14	21–32	Loop			
	9–15	20–31	**40**–**42**			
HMG-box domain, 1hme[207]	13–28	34–47	50–73			
	11–**21**	34–45	55–76			
Annexin V, 2ran[208]	Loop	14–25	32–40	44–58	62–69	72–82
	8–**12**	18–28	32–39	44–58	62–70	73–83
	85–96	104–113	116–130		134–141	144–154
	90–98	102–114	116–125, 127–134		Missed	145–156
	166–181		188–197	200–214		218–225
	164–**171**, **173**–184		187–195	200–**211**, **213**–216		218–225
	228–242	244–257	263–275	278–289	293–300	304–313
	230–244	248–257	262–**270**	275–287	289–298	303–313
Interleukin 4, 1rcb[209]	5–19	23–26	*β*	41–59	70–94	
	7–17	23–27	**29**–**36**	41–59	70–**86**, **90**–94	
	Loop	109–128				
	96–99	109–**123**				
Interferon γ, 1rfb[d]	6–16	28–37	44–59	65–81	87–95	107–116
	3–18	29–37	**39**–61	63–81	89–96	**102**–117
Glucoamylase, 3gly[210]	1–20		Loop	53–69	75–89	Loop
	3–**10**, **17**–22		**46**–**52**	56–70	72–85	**98**–**100**
	126–144	147–167	*β*	185–205		210–226
	126–144	**161**–**172**	**174**–**181**	185–**195**, **200**–205		215–227
	β	245–254	271–285		289–292	Loop
	236–**241**	245–255	271–**275**, **280**–288		290–294	**296**–**302**
	317–337	344–353	367–390	*β*	*β*	415–428
	315–**331**	344–353	369–392	**397**–**404**	**408**–**412**	414–**422**
Cytochrome *c* peroxidase, 2cyp[211]	15–33	42–54	73–78	84–99		103–119
	16–33	42–52	71–77	85–89, 92–99		104–**110**
	150–158	165–176		182–184	201–209	*β*
	151–161	165–**172**	**174**–177	Missed	201–207	**209**–**215**

Table VI (*Continued from the previous page.*)

Protein, PDB Code, and Reference	Actual (Top Line) and Calculated (Bottom Line) Locations of Helices[b]					
	233–240	242–253	255–272	288–293		
	Missed	242–252	254–272	**283**–291		
Phospholipase A2, 4p2p[212]	2–12	18–22	40–57	90–108	120–123	Loop
	2–13	19–22	38–**45**	90–109	118–123	**58**–**66**
Endothelin 1, 1edp[213]	9–15	Loop				
	9–17	**3**–**7**				
Cytochrome b562,[e] 256b[214]	3–19		23–42		46–48	
	3–**10**, **12**–20		23–**36**, **38**–43		Missed	
	56–81		84–105			
	56–**62**, **65**–**72**		**74**–**95**			
Bacterial photoreaction center subunit M,[e] 1prc[215]	3–7	36–41	53–76	81–90	111–137	
	2–**13**	33–41	**48**–74	81–93	112–**122**, **124**–138	
	143–166	169–171	177–190	198–224	232–237	241–254
	140–**160**	165–173	180–**195**	197–**217**	Missed	242–256
	260–284	292–298	315–317			
	260–284	288–302	315–318			
Hen lysozyme,[e] 1lys[d]	5–15	25–36	β	75–76	80–83	
	5–12	28–38	**55**–**62**	75–78	80–84	
	89–101		109–115	120–123		
	86–**92**, **94**–98		107–116	119–**128**		
Bovine trypsin inhibitor,[e] 5pti[216]	3–6	β	β	48–55		
	Missed	12–26	**29**–**41**	48–55		

[a] The α, 3_{10}, and distorted helices, were identified using the Cabsch and Sander algorithm[217] included in QUANTA. The numbering of residues corresponds to SEQRES record in Protein Data Bank (PDB)[83] files, starting from the first residue in the record. This sometimes differs from the numbering in the list of coordinates.

[b] The 3_{10} and distorted helices are underlined. The calculated helix ends with deviations $\geq$5 residues are indicated in bold.

[c] The residues were numbered starting from the first residue present in the list of coordinates.

[d] The reference was specified as ''to be published'' in the corresponding PDB file.

[e] This protein was not included in the main set used for optimization of adjustable parameters of the model.

porate such factors produced poorer agreement of the helix locations. This shows that cyclohexane is a reasonable approximation for the interior of micelles, or that the results of calculations are not very sensitive to the chosen hydrophobicity scale.

Incorporation of the micelle bound coil in the model [Eqs. (19, 20)] was less significant than the intrinsic helix stability terms, but it, too, improved agreement with nmr data. The optimum adjustable parameters of coil (f^{coil}, ΔH^{coil}, and ΔS^{coil}, Table VII) show that the coil segments interact weakly with micelles: the corresponding contributions per coil residue [$f^{\mathrm{coil}}\Delta G_{\mathrm{tr}}^{\mathrm{sch,coil}} + (\Delta H^{\mathrm{coil}} - T\Delta S^{\mathrm{coil}})$ from Eq. (20)], are only −0.72, −0.64, and −0.22 kcal/mol for Leu, Ile, and Val residues, respectively, and are positive (i.e., they were considered as zero) for all other residues at $T = 300$ K using the cyclohexane-derived transfer energies (Table I). As a result, the coil segments do not compete significantly with helices for binding with micelles. This correlates with weak hydrophobic binding observed for peptides that adopt nonregular structure in the complexes with vesicles or micelles.[110] The binding of such peptides results mostly from electrostatic interactions with lipid bilayers.[220]

The enthalpy determined for the helix–coil transition in the micelle-bound state (ΔH in Table VII) is identical to that in water, but the entropy of helix–coil transition is slightly smaller, since flexibility of the coil in the micelle-bound state is reduced.[97] The transfer energy found for the helix C═O · · · H—N group ($\Delta G_{\mathrm{tr}}^{\mathrm{mch}}$ in Table VII) is zero, although

Table VII Agreement of Calculated and Experimentally Identified Locations of α-Helices and Values of Adjustable Parameters

	Peptides in Water, Averaging of Partitions	Peptides in Micelles, Averaging of Partitions	Peptides in Micelles, Lowest Energy Partition	Proteins, Droplet Model, Lowest Energy Partition[a]
Number of molecules	96	36	36	30
Number of helices	48	41	41	205
Agreement with experiment				
Missing helices	3	0	1	4
Falsely calculated helices	2	0	0	23
Merged helices	0	0	0	2
Broken helices	0	0	1	16
Average helix boundary deviation (residues per helix)	1.26	2.02	2.39	4.1
Average % correct states	93	89	87	81
Adjustable parameters				
ΔH, kcal/mol	−1.24	−1.24	−1.51	−1.28
ΔS, cal/mol/K	4.10	3.90	3.96	3.81
P_d	0.18	0.30	—	—
ΔG_{tr}^{mch}, kcal/mol	—	0	0	+0.62
f^{coil}	—	0.5	0.5	0
ΔH^{coil}, kcal/mol	—	+0.7[b]	+0.7[b]	—[c]
ΔS^{coil}, cal/mol/K	—	1.8[b]	1.8[b]	—[c]
f^{sch}	—	1	1	0.57

[a] Transfer energies of nonpolar and weakly polar (Tyr, Ser, Thr) side chains (Table I) and N-capping energies for Asp, Asn, Ser, and Thr residues (−0.8, −0.7, −0.5, and −0.4 kcal/mol, respectively) were considered as additional adjustable parameters.

[b] The relationship between the ΔH^{coil} and ΔS^{coil} parameters was poorly defined because of insufficient data: different combinations of these parameters provided almost identical average helix boundary deviation.

[c] Not defined because the optimum contribution of coil is zero ($f = 0$).

it was estimated as +0.55 kcal/mol for transfer from water to benzene, or +0.62 kcal/mol for transfer from water to CCl_4.[94] The energy is zero probably because the backbone of micelle-bound helices resides within the micelle interface and is still solvated, as could be expected from the distribution of water in lipid bilayers.[111] The found fraction of immersion for coil side chains at the micelle interface f^{coil} is 0.5 (Table VII), exactly as was shown by Jacobs and White.[95]

Of the set of 34 peptides used to find adjustable parameters of the model, 10 were studied with neutral DPC micelles and 24 with anionic SDS micelles. Although the parameters may depend on the type of micelles, all these data were considered together. In fact, the stronger electrostatic binding of positively charged side chains with anionic micelles may shift the helix–coil equilibrium toward helix formation. The weak helix-stabilizing action of anionic micelles has been observed for the PhoE peptide,[176] which forms a longer and more stable α-helix in SDS micelles than in DPC micelles (in SDS, the helix is lengthened by a Lys-Lys-Ser segment and some of its NH protons have slower deuterium exchange rates), and for the mitochondrial presequence p25 peptide (at the end of Table V), which forms an extra arginine-containing α-helix in anionic DPG micelles[179] that is not observed in neutral DPC micelles.[178]

Two last peptides in Table V (mastoparan X and KRR-Lamb signal peptide) were studied in complexes with vesicles,[180,181] and all the observed helices are correctly reproduced in calculations, with the exception of a short, false calculated Thr^4-Leu^5-Arg^6-Lys^7-Arg^8 helix in the KRR-Lamb peptide (Table V). This apparent contradiction arises due to different geometry of the micelles and the bilayer: this helix was calculated because the hydrophobic Leu^5 residue can be buried within the micelle while the polar C-turn of the helix can be exposed to water, but this is geometrically forbidden for the short α-helix embedded in the planar bilayer of vesicles parallel to its surface. For the same reason, a polar C-terminal fragment (Lys-Arg-Gln-Gln-NH_2)

of mellitin included in the long α-helix formed by the peptide in micelles[153] (Table V) becomes unfolded when the monomeric peptide binds to phosphatidylcholine vesicles parallel to the bilayer surface.[221]

Table V only shows locations of the α-helices in peptides. The model also provides quantitative estimations of α-helix stabilities (the ΔG^{α} energies), occupancies of their turns, and helix arcs immersed into a micelle (the φ and ω angles), which can be correlated with mean residue ellipticities, chemical shifts, deuterium exchange rates, peptide–micelle binding constants, or other physicochemical data. However, this would require a separate investigation and some modifications of the model: the calculation of peptide–micelle binding constants, for example, requires taking into account the unbound peptide form, the peptide immobilization entropy, and energy of electrostatic peptide–micelle interactions. Although the model reproduces α-helix locations, it may be too simplified for more quantitative comparisons, since the transfer energies of side chains from water to lipid bilayer are different from those to cyclohexane,[95,111,222,223] the portions of the α-helices, immersed into the micelles, may differ geometrically from their arcs, and some side chains may be partially exposed to and partially removed from water.

All the results in Table V have been obtained using Boltzmann averaging of helix–coil partitions. The lowest energy helix–coil partition approach does not reproduce many marginally stable α-helices in water, but works satisfactorily for peptides in the presence of micelles (Table VII). Therefore, this approach was applied further for proteins.

Proteins

The droplet-like protein model considers every α-helix as floating in a spherical nonpolar droplet created by the rest of protein, thus eliminating the protein three-dimensional structure from consideration, and takes into account only nonspecific hydrophobic and intrahelical interactions, neglecting the specific tertiary packing of the helices and loops. This is a crude simplification since even molten globules and intermediate folding states may form native-like three-dimensional ''folds'' with some tertiary contacts.

The capability of the model was investigated for 30 mostly α-helical peptides and proteins of ellipsoidal shape with known 3D structure (first 30 proteins in Table VI). The peptides in the set (endothelin, for example) are considered together with proteins, since they have fixed tertiary structure, in contrast to the previously considered peptides in aqueous solution and in micelles. The molecules are of very different size, from 17 to 437 residues, with single-domain, multidomain (2cts, 2ran), and dimeric (1utg, 3wrp, 1rfb) organization, including disulfide-rich (1hyp, 1erl, 4p2p, 1edp) and heme-containing (3c2c, globins) proteins, sometimes with some β structure (3gly, 2cyp, 4p2p). The set includes glucoamylase (3gly) with a significant polar water-filled interior. Some proteins from the set (three globins, three Ca^{2+}-binding proteins, and 434 and λ repressors) are structurally and evolutionarily related. Transmembrane, $\alpha + \beta$, and rod-like 4-α-bundle proteins were not used to adjust the parameters of the model, but were analyzed later (four proteins of those types are shown at the end of Table VI as examples).

Initially, the lowest energy helix–coil partitions for the proteins were calculated with all parameters obtained for peptide–micelle complexes, and using the radius of SDS micelles in Eq. (18). This might be expected to yield secondary structures of the SDS-denatured proteins. This ''pure micellar'' approximation (results not presented here) identifies most protein α-helices (the boundary deviation was 5.6 residues per helix) because the intrahelical interactions in proteins and in micelles are the same, and the protein α-helices are usually amphiphilic and thus are stabilized in micelles. The calculations show that some β-structural segments could also form amphiphilic α-helices in micelles. The deviation was reduced to 4.12 residues per helix by varying, at first, adjustable parameters of the model shown in Table VII, and then side-chain transfer energies. The transfer energies were adjusted only for 12 hydrophobic and weakly hydrophilic side chains (Table I) since more hydrophilic residues are very rarely found in inner arcs of α-helices. Consequently the deviation was found to be very weakly sensitive to changes in the transfer energies of polar side chains. Attempts to minimize the deviation by varying additional parameters not shown in Table VII, including α-helix propensities $\Delta\Delta G^{sch}$, were unsuccessful, with the exception of N-capping energies of Asn, Asp, Ser, and Thr residues (Table I), which had probably been slightly (less by 0.3 kcal/mol) underestimated due to the two-state interpretation of CD data for model peptides.

Adjustable parameters of the model (Table VII) may reflect the real situation in a typical hydrophobically collapsed state of α-helical proteins. The enthalpy and entropy of the helix–coil transition (ΔH and ΔS in Table VII) obtained for proteins are

rather close to those determined with the Boltzmann averaging model for free and micelle-bound peptides. However, the paramerters of bound coil and the transfer energy of the helix backbone are distinctly different from those for peptide–micelle complexes: the coil does not compete with α-helices for binding with the protein core ($f^{coil} = 0$), and burial of the α-helix main chain becomes energetically unfavorable: the value determined, $\Delta G_{tr}^{mch} = +0.62$ kcal/mol (Table VII), is exactly as was estimated for transfer of the C═O···H—N group from water to CCl_4.[94] These differences probably arise because the protein droplet has no water-saturated interface, which would allow incorporation of the coil side chains between the lipid or detergent head groups, or hydration of the helix backbone, as in micelles.

It was also found that the helix boundary deviation for proteins can be reduced by introducing a coefficient $f^{sch} = 0.57$ [Eq. (15)], which indicates that the side chains, buried in the droplet, still partially interact with each other as in the unbound α-helix. A possible interpretation of this result is that the buried side chains are partially surrounded by holes or loosely packed in the collapsed state with expanded volume.

Compared with the cyclohexane scale, the transfer energies determined for protein side chains (Table I) are redistributed in favor of sulfur-containing and aromatic residues, qualitatively more similar to the octanol scale, because the surrounding of the buried side chains in proteins is more polar than cyclohexane.

Comparison of the calculated and actual location of α-helices in amino acid sequences of the 30 proteins (Tables VI and VII) shows that the droplet-like model, neglecting specific tertiary interactions, can reproduce 97% of the 205 helices, although some of these were calculated as being broken into shorter helices. The success of the droplet-like protein model means that the helix locations are defined mostly by the short-range intrahelical and long-range nonspecific hydrophobic interactions that are taken into account in the model. At the same time, the discrepancies observed show that, in some cases, the specific tertiary interactions can significantly modify locations of the helices. Some discrepancies in Table VI must arise from approximations that are common for the treatment of micelle-bound peptides and proteins in the model, such as inaccuracies in intrahelix interactions, approximation of buried helix by an arc, and from the assumption that each side chain should be completely buried or exposed to water. However, all these factors should contribute only ~ 2 residues per helix to the helix boundary deviation judging from the results for peptides. Thus, an additional increase in the deviation for proteins (which totals ~ 4.1 residues per helix, Table VII) appears due to tertiary interactions.

Analysis of the discrepancies in Table VI shows several distinct situations. First, some β-structural fragments of proteins have been calculated as helices because β structure was not considered, even though it can be more stable than helices and must be accounted for. Second, a few α-helices are converted to loops by complicated combinations of tertiary interactions, including hydrophobic contacts of side chains and hydrogen bonds involving the extended main chain of the loops. These two situations underlie the 23 falsely calculated α-helices in Table VI. Third, disulfide bridges and other covalent bonds can partially destroy some α-helices (the 199-208 helix of endonuclease III participating in an FeS cluster, the C-terminus of helix 3-18 of cytochrome *c* covalently bound with heme, and the 3-7 helix of endothelin which forms two disulfide bonds), can stabilize other α-helices (C-terminus of 88-100 leghemoglobin helix interacting with heme, and the C-terminal part of the phospholipase 40-57 helix forming two SS bonds), or can rearrange helices (the calculated 2-19 helix of mating pheromone actually forms an α-hairpin from two helices connected by a disulfide bridge).

Another type of specific tertiary interaction affecting formation of α-helices in proteins is hydrogen bonding of buried polar residues. The appearance of a polar side chain within the nonpolar arc of a helix sometimes breaks it into shorter helices, thus avoiding the energetically costly transfer of the polar side chain from water to the presumably nonpolar interior of a protein droplet. In a real protein, the hydrophilic side chains can be embedded into a polar cavity within a generally nonpolar core. This situation is typical for some α-bundle proteins (e.g., glucoamylase, bacteriorhodopsin, and proteins with apoferritin-like ''folds'') consisting of several long α-hairpins that associate with each other, creating inner cavities filled by hydrophilic side chains and water. As a result, three α-helices of glucoamylase were calculated as being broken and one helix as being significantly shorter than observed (Table VI) because of the inner polar residues Glu^{10}, Asp^{162}, Arg^{194}, Glu^{198}, and His^{278}. The α-helix 30-46 of Met aporepressor (1cmb) was calculated as being broken because of the buried Arg^{43} side chain that participates in a network of hydrogen bonds in the protein. The Trp repressor (3wrp) has

a layer of 5 bound water molecules that are incorporated between α-helix 79-91 and the rest of protein and form hydrogen bonds with the backbone and side chains (Thr^{81} and Ser^{88}) of the helix. The inner water molecules stabilize the helix because they prevent desolvation of its buried backbone (as discussed above, the transfer of a CO···HN group from water to the protein interior costs 0.62 kcal/mol). The long α-helix connecting the two domains of calmodulin was also calculated as being broken (Table VI), as in the peptide-bound form of this regulatory protein, because the middle part of the helix contains 8 sequential polar residues, all exposed to water.

The last problem connected with tertiary organization arises for 4-α-bundles and coiled coils of essentially nonspherical, rod-like shape, in which long α-helices are packed by the ''knobs-into-holes'' mechanism requiring interhelical angles of $\sim -165°$ or $+20°$.[224] In this situation, judging from results obtained for cytochrome c', cytochrome b562, myohemerythrin, lipoprotein, ferritin, aspartate receptor, and leucine zipper (2ccy, 256b, 2mhr, 1lpe, 1 fha, 2asr, and 2zta PDB files, respectively), the droplet-like protein model often gives broken helices (for example, in cytochrome b562 shown in the end of Table VI). In the water-soluble 4-α-bundles and coiled coils, the close packing of side chains additionally stabilizes the helices, and thus makes some of them longer or allows sequentially adjacent helices to combine into a single helix. The droplet-like protein model gives more satisfactory results for transmembrane α-bundle proteins (the subunit M of photoreaction center is shown in Table VI as an example), which was unexpected since the hydrophobic matching conditions for α-helix in the protein droplet and lipid bilayer are different.

For ''$\alpha + \beta$'' and ''α/β'' proteins (hen lysozyme and bovine pancreatic trypsin inhibitor are shown in Table VI as examples), the model yields the same results as for predominantly α-helical proteins, i.e., most helices are determined correctly, but some β-structural segments are identified as forming amphiphilic α-helices, because formation of the β structure has been neglected in the calculations.

CONCLUSIONS

Predictions of protein secondary structure are traditionally based on the analysis of statistical correlations or sequence patterns observed in proteins. A physically more justified approach is the quantification of the energetics actually defining formation of α-helices and β-sheets to calculate stability of secondary structure, instead of its prediction. The energetics of α-helix formation have recently been quantified by protein engineering and by using model peptides. We show here that a simple model combining these intrinsic α-helix stability data, accessible surface estimations for side-chain interactions, the transfer energies of side chains, and a few adjustable parameters allows determination of nearly all α-helices formed by free and micelle-bound peptides and by α-helical proteins. Nonetheless, formation of some helices in proteins is significantly affected by specific tertiary interactions. Consequently a future algorithm of tertiary structure calculation must be able to modify the helices initially formed by intrahelical and nonspecific hydrophobic interactions.

This work was supported by grants DA03910 and DA00118 from the National Institutes of Health. We are grateful to Molecular Simulations, Inc., for making their molecular modeling software available to us.

REFERENCES

1. Levitt, M. & Warshel, A. (1975) *Nature* **253,** 694–698.
2. Nemethy, G. & Scheraga, H. A. (1990) *FASEB J.* **4,** 3189–3197.
3. Srinivasan, R. & Rose, G. D. (1995) *Proteins* **22,** 81–99.
4. Dill, K. A. (1985) *Biochemistry* **24,** 1501–1509.
5. Shakhnovich, E., Farztdinov, G., Gutin, A. M. & Karplus, M. (1991) *Phys. Rev. Lett.* **67,** 1665–1668.
6. Skolnick, J., Kolinski, A., Brooks, C. L., III, Godzik, A. & Rey, A. (1993) *Curr. Biol.* **3,** 414–423.
7. Ptitsyn, O. B. & Finkelstein, A. V. (1980) in *Protein Folding,* Jaenicke, R., Ed., Elsevier Press, pp. 101–115.
8. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339.
9. Chothia, C. (1984) *Ann. Rev. Biochem.* **53,** 537–572.
10. Cohen, F. E. & Kuntz, I. D. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation,* Fasman, G. D., Ed., Plenum Press, New York, pp. 647–705.
11. Chou, K. C., Nemethy, G. & Scheraga, H. A. (1990) *Acc. Chem. Res.* **23,** 134–141.
12. Karplus, M. & Weaver, D. L. (1994) *Protein Sci.* **3,** 650–668.
13. Ptitsyn, O. B. & Semisotnov, G. V. (1991) in *Conformations and Forces in Protein Folding,* Nall,

B. T. & Dill, K. A., Eds., AAAS, Washington, pp. 155–168.
14. Fersht, A. R. (1993) *FEBS Lett.* **325,** 5–16.
15. Baldwin, R. L. (1993) *Curr. Opin. Struct. Biol.* **3,** 84–91.
16. Matthews, C. R. (1993) *Ann. Rev. Biochem.* **62,** 653–683.
17. Dobson, C. M., Evans, P. A. & Radford, S. E. (1994) *Trends Biochem. Sci.* **19,** 31–37.
18. Qian, H. & Schellman, J. A. (1992) *J. Phys. Chem.* **96,** 3987–3994.
19. Finkelstein, A. V., Badretdinov, A. Y. & Ptitsyn, O. B. (1991) *Proteins* **10,** 287–299.
20. Doig, A. J., Chakrabatty, A., Klinger, T. M. & Baldwin, R. L. (1994) *Biochemistry* **33,** 3396–3403.
21. Shalongo, W. & Stellwagen, E. (1995) *Protein Sci.* **4,** 1161–1166.
22. Stapley, B. J., Rohl, C. A. & Doig, A. J. (1995) *Protein Sci.* **4,** 2383–2391.
23. Munoz, V. & Serrano, L. (1995) *J. Mol. Biol.* **245,** 275–296.
24. Munoz, V. & Serrano, L. (1995) *J. Mol. Biol.* **245,** 297–308.
25. Skolnick, J. & Holtzer, A. (1982) *Macromolecules* **15,** 303–314.
26. Soman, K. V., Karimi, A. & Case, D. A. (1993) *Biopolymers* **33,** 1567–1580.
27. Jahnig, F. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3691–3695.
28. Hughson, F. M., Barrik, D. & Baldwin, R. L. (1991) *Biochemistry* **30,** 4113–4118.
29. Milla, M. E., Brown, B. M., Waldburger, C. D. & Sauer, R. T. (1995) *Biochemistry* **34,** 13914–13919.
30. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995) *J. Mol. Biol.* **254,** 260–288.
31. Uchiyama, H., Perez-Prat, E. M., Watanabe, K., Kumagai, I. & Kuwajima, K. (1995) *Protein Eng.* **8,** 1153–1161.
32. Zhu, Y., Chen, C. C., King, J. A. & Evans, L. B. (1992) *Biochemistry* **31,** 10591–10601.
33. Bellman, R. E. (1957) *Dynamic Programming*, Princeton University Press, Princeton, NJ.
34. Scholtz, J. M. & Baldwin, R. L. (1992) *Ann. Rev. Biophys. Biomol. Struct.* **21,** 95–118.
35. Ptitsyn, O. B. (1992) *Curr. Opin. Struct. Biol.* **2,** 13–20.
36. Lyu, P. C., Gans, P. J. & Kallenbach, N. R. (1992) *J. Mol. Biol.* **223,** 343–350.
37. Stellwagen, E., Park, S.-H., Shalongo, W. & Jain, A. (1992) *Biopolymers* **32,** 1193–1200.
38. Scholtz, J. M., Qian, H., Robbins, V. H. & Baldwin, R. L. (1993) *Biochemistry* **32,** 9668–9676.
39. Armstrong, K. M., Fairman, R. & Baldwin, R. L. (1993) *J. Mol. Biol.* **230,** 284–291.
40. Padmanabhan, S. & Baldwin, R. L. (1994) *Protein Sci.* **3,** 1992–1997.
41. Hermans, J. (1966) *J. Phys. Chem.* **70,** 510–515.
42. Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M. & Bolen, D. W. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2854–2858.
43. O'Neil, K. T. & DeGrado, W. F. (1990) *Science* **250,** 646–651.
44. Blaber, M., Zhang, X. J., Lindstrom, J. D., Pepiot, S. D., Baase, W. A. & Matthews, B. W. (1994) *J. Mol. Biol.* **235,** 600–624.
45. Lyu, P. C., Liff, M. I., Marky, L. A. & Kallenbach, N. (1990) *Science* **250,** 669–673.
46. Horovitz, A., Matthews, J. M. & Fersht, A. R. (1992) *J. Mol. Biol.* **227,** 560–568.
47. Chakrabatty, A., Doig, A. J. & Baldwin, R. L. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11332–11336.
48. Doig, A. J. & Baldwin, R. L. (1995) *Protein Sci.* **4,** 1325–1336.
49. Radzica, A. & Wolfenden, R. (1988) *Biochemistry* **27,** 1664–1670.
50. Baker, E. N. & Hubbard, R. E. (1984) *Prog. Biophys. Mol. Biol.* **44,** 97–179.
51. Shirley, B. A., Stanssens, P., Hahn, U. & Pace, C. N. (1992) *Biochemistry* **31,** 725–732.
52. Chen, Y. W., Fersht, A. R. & Henrick, K. (1993) *J. Mol. Biol.* **234,** 1158–1170.
53. Chakrabatty, A., Kortemme, T. & Baldwin, R. L. (1994) *Protein Sci.* **3,** 843–852.
54. Nicholson, H., Anderson, D. E., Dao-Pin, S. & Matthews, B. W. (1991) *Biochemistry* **30,** 9816–9828.
55. Sancho, J., Serrano, L. & Fersht, A. R. (1992) *Biochemistry* **31,** 2253–2258.
56. Armstrong, K. M. & Baldwin, R. L. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11337–11340.
57. Zhou, H. X., Lyu, P. C., Wemmer, D. E. & Kallenbach, N. R. (1994) *J. Am. Chem. Soc.* **116,** 1139–1140.
58. Serrano, L., Sancho, J., Hirshberg, M. & Fersht, A. (1992) *J. Mol. Biol.* **227,** 544–559.
59. Yun, R. H., Anderson, A. & Hermans, J. (1991) *Proteins* **10,** 219–228.
60. Hurley, J. H., Mason, D. A. & Matthews, B. W. (1992) *Biopolymers* **32,** 1443–1446.
61. Richardson, J. S. & Richardson, D. C. (1988) *Science* **240,** 1648–1652.
62. Harper, E. T. & Rose, G. D. (1993) *Biochemistry* **32,** 7605–7609.
63. Zhou, H. X., Lyu, P., Wemmer, D. E. & Kallenbach, N. R. (1994) *Proteins* **18,** 1–7.
64. Bell, J. A., Becktel, W. J., Sauer, U., Baase, W. A. & Matthews, B. W. (1992) *Biochemistry* **31,** 3590–3596.
65. Zhukovsky, E. A., Mulkerrin, M. G. & Presta, L. G. (1994) *Biochemistry* **33,** 9856–9864.
66. elMastry, N. F. & Fersht, A. R. (1994) *Protein Eng.* **7,** 777–782.
67. Seale, J. W., Srinivasan, R. & Rose, G. D. (1994) *Protein Sci.* **3,** 1741–1745.

68. Munoz, V., Blanco, F. J. & Serrano, L. (1995) *Nature Struct. Biol.* **2,** 380–385.
69. Osterhout, J. J., Baldwin, R. L., York, E. J., Stewart, J. M., Dyson, H. J. & Wright, P. E. (1989) *Biochemistry* **28,** 7059–7064.
70. Lyu, P. C., Wemmer, D. E., Zhou, H. X., Pinker, R. J. & Kallenbach, N. R. (1993) *Biochemistry* **32,** 421–425.
71. Jimenez, M. A., Munoz, V., Rico, M. & Serrano, L. (1994) *J. Mol. Biol.* **242,** 487–496.
72. Munoz, V. & Serrano, L. (1995) *Biochemistry* **34,** 15301–15306.
73. Aurora, R., Srinivasan, R. & Rose, G. D. (1994) *Science* **264,** 1126–1130.
74. Schellman, C. (1980) in *Protein Folding,* Jaenicke, R., Ed., Elsevier/North-Holland, New York, pp. 53–61.
75. Karle, I. L., Flippen-Andersen, J. L., Uma, K. & Balaram, P. (1993) *Int. J. Peptide Protein Res.* **42,** 401–410.
76. Pervushin, K. V., Arseniev, A. S., Kozhich, A. T. & Ivanov, V. T. (1991) *J. Biomol. NMR* **1,** 313–322.
77. Lomize, A. L., Pervushin, K. V. & Arseniev, A. S. (1992) *J. Biomol. NMR* **2,** 361–372.
78. Macquaire, F., Baleux, F., Huynh-Dinh, T., Rouge, D., Neumann, J.-M. & Sanson, A. (1993) *Biochemistry* **32,** 7244–7254.
79. Pervushin, K. V., Orekhov, V. Y., Popov, A. I., Musina, L. Y. & Arseniev, A. S. (1994) *Eur. J. Biochem.* **219,** 571–583.
80. Viguera, A. R. & Serrano, L. (1995) *J. Mol. Biol.* **251,** 150–160.
81. Huyghues-Despointes, B. M., Scholtz, J. M. & Baldwin, R. L. (1993) *Protein Sci.* **2,** 80–85.
82. Huyghues-Despointes, B. M., Scholtz, J. M. & Baldwin, R. L. (1993) *Protein Sci.* **2,** 1604–1611.
83. Bernstein, F. C., Koetzle, T. F., Williams, J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
84. Lin, L., Pinker, R. J. & Kallenbach, N. R. (1993) *Biochemistry* **32,** 12638–12643.
85. Piela, L., Nemethy, G. & Scheraga, H. A. (1987) *Biopolymers* **26,** 1273–1286.
86. McGregor, M. J., Islam, S. A. & Sternberg, M. J. E. (1987) *J. Mol. Biol.* **198,** 295–310.
87. Creamer, T. P. & Rose, G. D. (1995) *Protein Sci.* **4,** 1305–1314.
88. Viguera, A. R. & Serrano, L. (1995) *Biochemistry* **34,** 8771–8779.
89. Maslennikov, I. V., Lomize, A. L. & Arseniev, A. S. (1991) *Bioorg. Chem.* (*Russia*n) **17,** 1456–1469.
90. Barsukov, I. L., Nolde, D. E., Lomize, A. L. & Arseniev, A. S. (1992) *Eur. J. Biochem.* **206,** 665–672.
91. Jimenez, M. A., Bruix, M., Gonzalez, C., Blanco, F. J., Nieto, J. L., Herranz, J. & Rico, M. (1993) *Eur. J. Biochem.* **211,** 569–581.
92. Munoz, V., Serrano, L., Jimenez, M. A. & Rico, M. (1995) *J. Mol. Biol.* **247,** 648–669.
93. Constantine, K. L., Mapelli, C., Meyers, C. A., Friedrichs, M. S., Krystek, S. & Mueller, L. (1993) *J. Biol. Chem.* **268,** 22830–22837.
94. Roseman, M. A. (1988) *J. Mol. Biol.* **201,** 621–623.
95. Jacobs, R. E. & White, S. H. (1989) *Biochemistry* **28,** 3421–3437.
96. Wimley, H. & White, S. H. (1995) *Biophys. J.* **68,** A300.
97. Deber, C. M. & Li, S.-C. (1995) *Biopolymers* (*Peptide Sci.*) **37,** 295–318.
98. Rose, G. D. (1978) *Nature* **272,** 586–590.
99. Stille, C. J., Thomas, L. J., Reyes, V. E. & Humpreys, R. E. (1987) *Mol. Immunol.* **24,** 1021–1027.
100. Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. (1982) *Nature* **299,** 371–374.
101. Ptitsyn, O. B. & Finkelstein, A. V. (1983) *Biopolymers* **22,** 15–25.
102. Israelachvili, J. N. (1992) *Intermolecular and Surface Forces,* Academic Press, New York, pp. 371–373.
103. Barlow, D. J. & Thornton, J. M. (1988) *J. Mol. Biol.* **201,** 601–619.
104. Chothia, C., Levitt, M. & Richardson, D. (1981) *J. Mol. Biol.* **145,** 215–250.
105. Hicks, R. P., Beard, D. J. & Young, J. K. (1992) *Biopolymers* **32,** 85–96.
106. Graham, W. H., Carter, E. S., II & Hicks, R. P. (1992) *Biopolymers* **32,** 1755–1764.
107. Brown, J. W. & Huestis, W. H. (1993) *J. Phys. Chem.* **97,** 2967–2973.
108. Ohno, Y., Segawa, M., Ohishi, H., Doi, M., Kitamura, K., Ishida, T., Inoue, M. & Iwashita, T. (1993) *Eur. J. Biochem.* **212,** 185–191.
109. Young, J. K. & Hicks, R. P. (1994) *Biopolymers* **34,** 611–623.
110. Deber, C. M. & Behnam, B. A. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 61–65.
111. White, S. H. & Wimley, W. C. (1994) *Curr. Op. Struct. Biol.* **4,** 79–86.
112. Kemmink, J. & Creighton, T. E. (1993) *J. Mol. Biol.* **234,** 861–878.
113. Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J. & Wright, P. E. (1993) *Biochemistry* **32,** 6337–6347.
114. Shin, H.-C., Merutka, G., Waltho, J. P., Wright, P. & Dyson, H. J. (1993) *Biochemistry* **32,** 6348–6355.
115. Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A. & Wright, P. E. (1992) *J. Mol. Biol.* **226,** 795–817.
116. Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H. C., Lerner, R. A. & Wright, P. E. (1992) *J. Mol. Biol.* **226,** 819–835.

117. Kippen, A. D., Arcus, V. L. & Fersht, A. R. (1994) *Biochemistry* **33,** 10013–10021.
118. Munier, H., Blanco, F. J., Precheur, B., Diesis, E., Nieto, J. L., Craescu, C. T. & Barzu, O. (1993) *J. Biol. Chem.* **268,** 1695–1701.
119. Craescu, C. T., Bouhss, A., Mispelter, J., Diesis, E., Popescu, A., Chiriac, M. & Barzu, O. (1995) *J. Biol. Chem.* **31,** 7088–7096.
120. Munier, H., Bouhss, A., Gilles, A.-M., Palibroda, N., Barzu, O., Mispelter, J. & Craescu, C. T. (1995) *Arch. Biochem. Biophys.* **320,** 224–235.
121. van der Graaf, M., van Mierlo, C. P. M. & Hemminga, M. A. (1991) *Biochemistry* **30,** 5722–5727.
122. Precheur, B., Munier, H., Mispelter, J., Barzu, O. & Craescu, C. T. (1992) *Biochemistry* **31,** 229–236.
123. Jarvis, J. A., Munro, S. L. & Craik, D. J. (1994) *Biochemistry* **33,** 33–41.
124. Merutka, G., Morikis, D., Bruschweiler, R. & Wright, P. E. (1993) *Biochemistry* **32,** 13089–13097.
125. Knott, H. M., Berndt, M. C., Kralicek, A. V., O'Donoghue, S. I. & King, G. F. (1992) *Biochemistry* **31,** 11152–11158.
126. Breeze, A. L., Harvey, T. S., Bazzo, R. & Campbell, I. D. (1991) *Biochemistry* **30,** 575–582.
127. Blanco, F. J., Jimenez, A., Rico, M., Santoro, J., Herranz, J. & Nieto, J. L. (1992) *Biochem. Biophys. Res. Commun.* **182,** 1491–1498.
128. Zagorski, M. G. & Barrow, C. (1992) *Biochemistry* **31,** 5621–5631.
129. Zhang, M. & Vogel, H. J. (1994) *Biochemistry* **33,** 1163–1171.
130. Jimenez, M. A., Carreno, C., Andreu, D., Blanco, F. J., Herranz, J., Rico, M. & Nieto, J. L. (1994) *Biopolymers* **34,** 647–661.
131. McLeish, M. J., Nielsen, K. J., Najbar, L. V., Wade, J. D., Lin, F., Doughty, M. B. & Craik, D. J. (1994) *Biochemistry* **33,** 11174–11183.
132. Raj, P. A., Soni, S. D. & Levine, M. J. (1994) *J. Biol. Chem.* **269,** 9610–9619.
133. Bradley, E. K., Thomason, J. F., Cohen, F. E., Kosen, P. A. & Kuntz, I. D. (1990) *J. Mol. Biol.* **215,** 607–622.
134. Czisch, M., Liebers, V., Bernstein, R., Chen, Z., Baur, X. & Holak, T. A. (1994) *Biochemistry* **33,** 9420–9427.
135. Xu, X., Cooper, L. G., DiMario, P. J. & Nelson, J. W. (1994) *Biopolymers* **35,** 93–102.
136. Marqusee, S. & Sauer, R. T. (1994) *Protein Sci.* **3,** 2217–2225.
137. Musco, G., Tziatzios, C., Schuck, P. & Pastore, A. (1995) *Biochemistry* **34,** 553–561.
138. Burke, C., Mayo, K. H., Skubitz, A. P. N. & Furcht, L. T. (1991) *J. Biol. Chem.* **266,** 19407–19412.
139. Young, J. K., Anklin, C. & Hicks, R. P. (1994) *Biopolymers* **34,** 1449–1462.
140. Mulhern, T. D., Howlett, G. J., Reid, G. E., Simpson, R. J., McColl, D. J., Anders, R. F. & Norton, R. S. (1995) *Biochemistry* **34,** 3479–3491.
141. Morris, M. B., Ralston, G. B., Biden, T. J., Browne, C. L., King, G. F. & Iismaa, T. P. (1995) *Biochemistry* **34,** 4538–4545.
142. Freedman, S. J., Furie, B. C., Furie, B. & Baleja, J. D. (1995) *J. Biol. Chem.* **270,** 7980–7987.
143. Aroeti, B., Kosen, P. A., Kuntz, I. D., Cohen, F. E. & Mostov, K. E. (1993) *J. Cell. Biol.* **123,** 1149–1160.
144. Zhang, M. & Vogel, H. J. (1994) *J. Biol. Chem.* **269,** 981–985.
145. Zhang, M., Vogel, H. J. & Zwiers, H. (1994) *Biochem. Cell. Biol.* **72,** 109–116.
146. Baumann, H. & Hard, T. (1995) *Eur. J. Biochem.* **230,** 879–885.
147. Blanco, F. J. & Serrano, L. (1995) *Eur. J. Biochem.* **230,** 634–649.
148. Munoz, V., Blanco, F. J. & Serrano, L. (1995) *Protein Sci.* **4,** 1577–1586.
149. Viguera, A. R., Jimenez, M. A., Rico, M. & Serrano, L. (1996) *J. Mol. Biol.* **255,** 507–521.
150. Itzhaki, L. S., Neira, J. L., Ruiz-Sanz, J., de Prat Gay, G. & Fersht, A. R. (1995) *J. Mol. Biol.* **254,** 289–304.
151. Sancho, J., Neira, J. L. & Fersht, A. R. (1992) *J. Mol. Biol.* **224,** 749–758.
152. Thornton, K., Wang, Y., Weiner, H. & Gorenstein, D. G. (1993) *J. Biol. Chem.* **268,** 19906–19914.
153. Inagaki, F., Shimada, I., Kawaguchi, K., Hirano, M., Terasawa, I., Ikura, T. & Go, N. (1989) *Biochemistry* **28,** 5985–5991.
154. Braun, W., Wider, G., Lee, K. H. & Wutrich, K. (1983) *J. Mol. Biol.* **169,** 921–948.
155. Thornton, K. & Gorenstein, D. G. (1994) *Biochemistry* **33,** 3532–3539.
156. Hammen, P. K., Gorenstein, D. G. & Weiner, H. (1994) *Biochemistry* **33,** 8610–8617.
157. Lee, K. H., Fitton, J. E. & Wutrich, K. (1987) *Biochim. Biophys. Acta* **911,** 144–153.
158. Hammen, P. K., Gorenstein, D. G. & Weiner, H. (1996) *Biochemistry* **35,** 3772–3781.
159. Johansson, J., Szyperski, T. & Wuthrich, K. (1995) *FEBS Lett.* **362,** 261–265.
160. van der Ven, F. J., van Os, J. W., Aelen, J. M., Wymenga, S. S., Remerowski, M. L., Konings, R. N. & Hilbers, C. W. (1993) *Biochemistry* **32,** 8322–8328.
161. Motta, A., Pastore, A., Goud, N. A. & Morelli, M. A. C. (1991) *Biochemistry* **30,** 10444–10450.
162. McLeish, M. J., Nielsen, K. J., Wade, J. D. & Craik, D. J. (1993) *FEBS Lett.* **315,** 323–328.
163. Bairaktari, E., Mierke, D. F., Mammi, S. & Peggion, E. (1990) *Biochemistry* **29,** 10090–10096.
164. Kallick, D. (1993) *J. Am. Chem. Soc.* **115,** 9317–9318.
165. Rizo, J., Blanco, F. J., Kobe, B., Bruch, M. D. &

Gierasch, L. M. (1993) *Biochemistry* **32,** 4881–4894.
166. Zetta, L., Consonni, R., De Marco, A., Longhi, R., Manera, E. & Vecchio, G. (1990) *Biopolymers* **30,** 899–909.
167. Pervushin, K. V. & Arseniev, A. S. (1992) *FEBS Lett.* **308,** 190–196.
168. Kloosterman, D. A., Scahill, T. A. & Friedman, A. R. (1993) *Peptide Res.* **6,** 211–218.
169. Improta, S., Pastore, A., Mammi, S. & Peggion, E. (1994) *Biopolymers* **34,** 773–782.
170. Wilson, J. C., Nielsen, K. J., McLeish, M. J. & Craik, D. J. (1994) *Biochemistry* **33,** 6802–6811.
171. Battistutta, R., Bisello, A., Mammi, S. & Peggion, E. (1994) *Biopolymers* **34,** 1535–1541.
172. Gilbert, G. E. & Baleja, J. D. (1995) *Biochemistry* **34,** 3022–3031.
173. Zhang, H., Kaneko, K., Nguyen, J. T., Livshits, T. L., Baldwin, M. A., Cohen, F. E., James, T. L. & Prusiner, S. B. (1995) *J. Mol. Biol.* **250,** 514–526.
174. Rozek, A., Buchko, G. W. & Cushley, R. J. (1995) *Biochemistry* **34,** 7401–7408.
175. Buchko, G. W., Rozek, A., Zhong, Q. & Cushley, R. J. (1995) *Peptide Res.* **8,** 86–94.
176. Chupin, V., Killian, J. A., Breg, J., de Jongh, H. H. J., Boelens, R. & de Kruijff, B. (1995) *Biochemistry* **34,** 11617–11624.
177. Buchko, G. W., Treleaven, W. D., Dunne, S. J., Tracey, A. S. & Cushley, R. J. (1996) *J. Biol. Chem.* **271,** 3039–3045.
178. Endo, T., Shimada, I., Roise, D. & Inagaki, F. (1989) *J. Biochem.* **106,** 396–400.
179. Chupin, V., Leenhouts, J. M., de Kroon, A. I. P. M. & de Kruijff, B. (1996) *Biochemistry* **35,** 3141–3146.
180. Wakamatsu, K., Okada, A., Miyazawa, T., Ohya, M. & Higashijima, T. (1992) *Biochemistry* **31,** 5654–5660.
181. Wang, Z., Jones, J. D., Rizo, J. & Gierasch, L. M. (1993) *Biochemistry* **32,** 13991–13999.
182. Wright, P. E., Dyson, H. J. & Lerner, R. A. (1988) *Biochemistry* **27,** 7167–7175.
183. Richards, F. M. (1977) *Ann. Rev. Biophys. Bioeng.* **6,** 151–176.
184. Zielenkiewicz, P. & Saenger, W. (1992) *Biophys. J.* **63,** 1483–1486.
185. Mondragon, A., Subbiah, S., Almo, S. C., Drottar, M. & Harrison, S. C. (1989) *J. Mol. Biol.* **205,** 189–200.
186. Kuo, C. F., McRee, D. E., Fisher, C. L., O'Handley, S., Cunningham, R. P. & Tainer, J. A. (1992) *Science* **258,** 434–440.
187. Morize, I., Surcouf, E., Vaney, M. C., Epelboin, Y., Buehner, M., Fridlandsky, F., Milgrom, E. & Mornon, J. P. (1987) *J. Mol. Biol.* **194,** 725–739.
188. Remington, S., Wiegand, G. & Huber, R. (1982) *J. Mol. Biol.* **158,** 111–152.
189. Lawson, C. L., Zhang, R. G., Schevitz, R. W., Otwinowski, Z., Joachimiak, A. & Sigler, P. B. (1988) *Proteins* **3,** 18–31.
190. Yuan, H. S., Finkel, S. E., Feng, J. A., Johnson, R. C. & Dickerson, R. E. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9558–9562.
191. Beamer, L. J. & Pabo, C. O. (1992) *J. Mol. Biol.* **227,** 177–196.
192. Parker, M. W., Postma, J. P. M., Pattus, F., Tucker, A. D. & Tsernoglou, D. (1992) *J. Mol. Biol.* **224,** 639–657.
193. Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. E. & Pabo, C. O. (1990) *Cell* **63,** 579–590.
194. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. (1993) *Nature* **362,** 219–223.
195. Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S. & Lehmann, M. S. (1992) *J. Mol. Biol.* **231,** 877–887.
196. Kumar, V. D., Lee, L. & Edwards, B. F. P. (1990) *Biochemistry* **29,** 1404–1412.
197. Babu, Y. S., Bugg, C. E. & Cook, W. J. (1988) *J. Mol. Biol.* **204,** 191–204.
198. Svensson, L. A., Thulin, E. & Forsen, S. (1992) *J. Mol. Biol.* **223,** 601–606.
199. Takano, T. (1977) *J. Mol. Biol.* **110,** 569–584.
200. Steigemann, W. & Weber, E. (1979) *J. Mol. Biol.* **127,** 309–338.
201. Arutyunyan, E. G., Kuranova, I. P., Vainshtein, B. K. & Steigemann, W. (1980) *Kristallografiya* (Russian) **25,** 80–100.
202. Salemme, F. R., Freer, S. T., Xuong, N. H., Alden, R. A. & Kraut, J. (1973) *J. Biol. Chem.* **248,** 3910–3921.
203. Weiss, M. S., Anderson, D. H., Raffioni, S., Bradshaw, R. A., Ortenzi, C., Luporini, P. & Eisenberg, D. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 10172–10176.
204. Robien, M. A., Clore, G. M., Omichinski, J. G., Perham, R. N., Appella, E., Sakaguchi, K. & Gronenborn, A. M. (1992) *Biochemistry* **31,** 3463–3471.
205. Rafferty, J. B., Somers, W. S., Saint-Girons, I. & Phillips, S. E. V. (1989) *Nature* **341,** 705–710.
206. Deisenhofer, J. (1981) *Biochemistry* **20,** 2361–2370.
207. Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D. & Thomas, J. O. (1993) *EMBO J.* **12,** 1311–1319.
208. Concha, N. O., Head, J. F., Kaetzel, M. A., Dedman, J. R. & Seaton, B. A. (1993) *Science* **261,** 1321–1324.
209. Wlodawer, A., Pavlovsky, A. & Gustchina, A. (1992) *FEBS Lett.* **309,** 59–64.
210. Aleshin, A. E., Hoffman, C., Firsov, L. M. & Honzatko, R. B. (1994) *J. Mol. Biol.* **238,** 575–591.
211. Finzel, B. C., Poulos, T. L. & Kraut, J. (1984) *J. Biol. Chem.* **259,** 13027–13036.
212. Finzel, B. C., Ohlendorf, D. H., Weber, P. C. & Sa-

lemme, F. R. (1991) *Acta Ctystallogr.* **B47,** 558–559.
213. Andersen, N. H., Chen, C. P., Marschner, T. M., Krystek, S. J. & Bassolino, D. A. (1992) *Biochemistry* **31,** 1280–1295.
214. Lederer, F., Glatigny, A., Bethge, P. H., Bellamy, H. D. & Matthew, F. S. (1981) *J. Mol. Biol.* **148,** 427–428.
215. Deisenhofer, J., Epp, O., Sinning, I. & Michel, H. (1995) *J. Mol. Biol.* **246,** 429–457.
216. Wlodawer, A., Walter, J., Huber, R. & Sjolin, L. (1984) *J. Mol. Biol.* **180,** 301–329.
217. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22,** 2577–2637.
218. Shin, H.-C., Merutka, G., Waltho, J. P., Tennant, L. L., Dyson, H. J. & Wright, P. E. (1993) *Biochemistry* **32,** 6356–6364.
219. Maslennikov, I. V., Arseniev, A. S., Tchikin, L. D., Kozhich, A. T. & Ivanov, V. T. (1993) *Bioorg. Chem.* (Russian) **19,** 5–20.
220. Tamm, L. K. (1994) in *Membrane Protein Structure. Experimental Approaches,* White, S. H., Ed., Oxford University Press, New York, pp. 283–313.
221. Okada, A., Wakamatsu, K., Miyazawa, T. & Higashijima, T. (1994) *Biochemistry* **33,** 9438–9446.
222. Wimley, W. C. & White, S. H. (1993) *Biochemistry* **32,** 6307–6312.
223. De Young, L. R. & Dill, K. A. (1990) *J. Phys. Chem.* **94,** 801–809.
224. Crick, F. H. C. (1953) *Acta Crystallogr.* **6,** 689–697.