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]

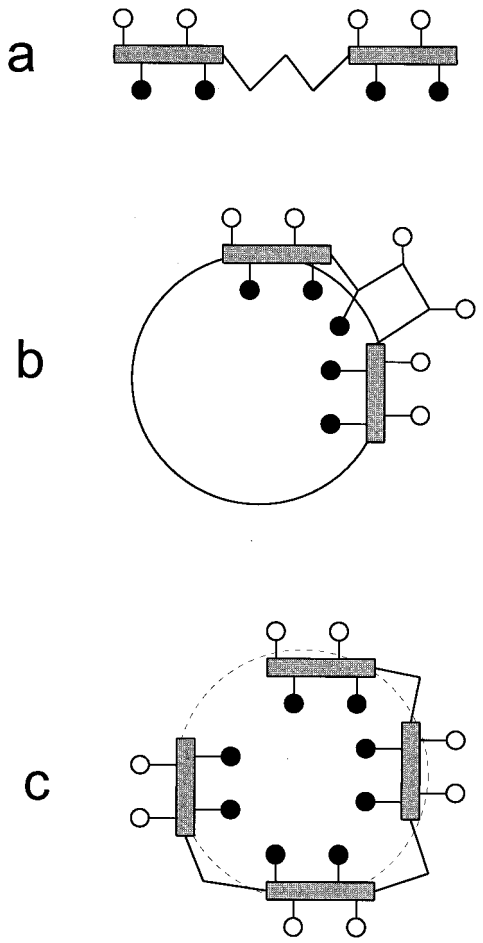

FIGURE 1 Three models of α-helix formation (the helices are shown as rectangles, solid circles are hydrophobic side chains): (a) ''peptide in aqueous solution'' (there are only specific interactions between residues within each α-helix); (b) ''peptide in complex with a micelle'' (there are specific intrahelical 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.

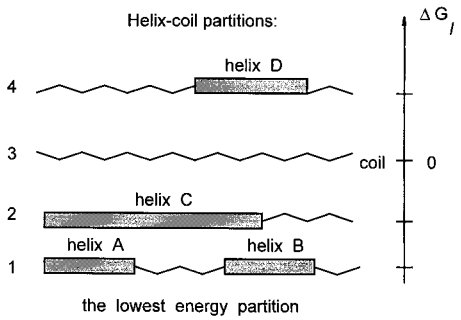

FIGURE 2 Helix–coil partitions as conformational states of the peptide chain. The coil in aqueous solution serves as a reference state with zero energy. The helices A, B, and C, shown as rectangles, with $\Delta G < 0$, are more stable than the coil. The helices compete with each other, and partition **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-

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.