Molecular Thermodynamic Model To Predict the α-Helical Secondary Structure of Polypeptide Chains in Solution†

Yizu Zhu,‡ Chau-Chyun Chen,*,§ Jonathan A. King,‖ and Lawrence B. Evans‡

Departments of Chemical Engineering and Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and Aspen Technology, Inc., Cambridge, Massachusetts 02139

Received May 15, 1992; Revised Manuscript Received August 11, 1992

ABSTRACT: The native state of a protein molecule in aqueous solutions represents one of the lowest states of Gibbs energy [Anfinsen, C. B. (1973) *Science 181*, 223–230]. Much progress has been made about the rules of protein folding [King, J. (1989) *Chem. Eng. News 67*, 32–54] and the dominant forces in protein folding [Dill, K. A. (1990) *Biochemistry 29*, 7133–7155]. However, the quantitative contributions of different Gibbs energy terms to protein stability remains a controversial issue [Moult, J., & Unger, R. (1991) *Biochemistry 30*, 3816–3824]. A molecular thermodynamic model has been proposed for the Gibbs energy of folding a residue in aqueous homopolypeptides from a random-coiled state to either the α-helix state or the β-sheet state [Chen, C.-C., Zhu, Y., King, J. A., & Evans, L. B. (1992) *Biopolymers 32*, 1375–1392]. In this work, we present a generalization of the molecular thermodynamic model for the Gibbs energy of folding natural and synthetic heteropolypeptides from random-coiled conformations into α-helical conformations. The generalized model incorporates the intrinsic folding potential due to residue–solvent interactions, the cooperative folding effect due to residue–residue interactions, and the location and length of α-helices. The utility of the model was demonstrated by examining the stability of α-helical conformations of a number of natural polypeptides including C-peptide (residues 1–13) and S-peptide (residues 1–20) of RNase A (bovine pancreatic ribonuclease A), the Pα fragment in BPTI (bovine pancreatic trypsin inhibitor), and synthetic polypeptides (the copolymers of different amino acid residues) including alanine-based peptides (16 or 17 residues long) in water. The computed Gibbs energies correspond well with the experimental data on helicity. The results also accounted for the effects of amino acid substitution and temperature on the stability of α-helical conformations of the test polypeptides.

The local physical interactions determining the conformations of polypeptide chains in solution include hydrogen bonding, electrostatic interactions, van der Waals interactions, hydrophobic interactions, etc. (Dill, 1990; Kim & Baldwin, 1990). These interactions exist between residue–residue pairs, residue–solvent pairs, and solvent–solvent pairs.

A central problem in understanding how these interactions direct chain folding has been quantitative estimations of their contributions to the stability of the folded state. Among the important approaches have been the classical studies of Schellman (1955) and Zimm and Bragg (1959) based on statistical mechanics, the simulation work of McCammon et al. (1977) based on molecular dynamics, and the experimental work of Scheraga (1978) with host–guest copolymers. The useful prediction methods developed by Chou and Fasman (1974a,b; 1978) and many other groups applied statistical analysis to the ever-expanding protein conformation database and are not based on estimation of the physical interactions.

There are a number of recent works trying to quantify the various interactions that contribute to the protein stability. Ben-naim (1990) examined the solvent effects on the thermodynamics of protein folding and protein association with the concept of "conditional solvation free energy." Scholtz et al. (1991) applied the Zimm–Bragg theory (Zimm & Bragg, 1959) and the Lifson–Roig theory (Lifson & Roig, 1961) to describe the thermal unfolding transitions of short peptides in water. Gans et al. (1991) examined natural and synthetic peptides that contain intramolecular α-helical structure in aqueous solutions and suggested that a single set of Zimm–Bragg helix growth parameter s and the nucleation parameter σ can not account for all of the available data regarding the helical propensities for the common amino acids. Rooman et al. (1991) developed a method based on seven structure assignments to predict lowest energy protein backbone tertiary structures from the amino acid sequence. Potentials of mean force due to local interactions were derived from a database of known protein structures, based on statistical influences of single residues and residue pairs on the conformational states in their vicinity along the chain. Rashin (1992) carried out a computational study of the role of all ionizable groups of C-peptide in its helix–coil transition within the framework of continuum electrostatics.

It is now widely established that the native structure is thermodynamically stable, that hydrophobicity is the dominant force of folding, and that the principle opposing force is entropic (Dill, 1990). Several Gibbs energy expressions have been proposed for the folding of residues in polypeptide chains. Dill treated the Gibbs energy of folding as the result of the balance between the "solvophobic" interaction and the chain conformational entropy (Dill, 1985; Dill et al., 1989; Alonso & Dill, 1991; Fields et al., 1992). Moult and Unger (1991) considered the burying of the nonpolar surface (the hydrophobic effect) and the configurational entropy as the two main sources of the entropy of folding, and they suggested the enthalpy of folding to be of relatively minor importance.

† We acknowledge the support from the National Science Foundation under the Engineering Research Center Initiative to the Biotechnology Process Engineering Center (Cooperative Agreement CDR-8-03014) that made this research possible.

* To whom correspondence should be addressed.
‡ Department of Chemical Engineering, MIT.
§ Aspen Technology, Inc.
‖ Department of Biology, MIT.

0006-2960/92/0431-10591$03.00/0

Molecular thermodynamics takes advantages of a large body of experimental thermodynamic data developed by chemical engineers to quantify the intramolecular and intermolecular interactions governing the phase behavior and stability of both small molecules and large molecules (Prausnitz, 1979). This methodology has been successfully used to derive semiempirical Gibbs energy expressions and to estimate the associated thermodynamic parameters for mixtures of nonelectrolytes (Renon & Prausnitz, 1968; Fredenslund et al., 1975), electrolytes (Chen & Evans, 1986), polymers (Flory, 1941), micelles (Blankschtein et al., 1985), amino acids (Chen et al., 1989), antibiotics (Zhu et al., 1990), etc. These systems of small and large molecules share many of the important characteristics of aqueous protein systems. For example, the polypeptide chain structure exists in polymer systems. The short-range molecule–molecule interactions, such as dipole–dipole interactions, are common among nonelectrolyte systems. The long-range molecule–ion interactions and ion–ion interactions play dominant roles in electrolyte systems. The dipolar ion nature of proteins is shared by amino acids and zwitterionic surfactants. Moreover, the conformation issue is important to micellar systems.

In an earlier article (Chen et al., 1992), we successfully applied the molecular thermodynamic approach and developed a quantitative model for the Gibbs energy of folding for aqueous homopolypeptides. The resulting model has been shown to generate important results that are consistent with the critical observations reported in the literature for the folding of aqueous homopolypeptides. The model yields a hydrophobicity scale for the 20 amino acid side chains that compares favorably with established scales, including that of Nozaki and Tanford (1971) and that of Leodidis and Hatton (1990). In addition, the model generates qualitatively correct thermodynamic constants, and it accurately predicts thermodynamically favorable folding of a number of aqueous homopolypeptides from random-coiled states into α-helices. The model further facilitates estimation of the Zimm–Bragg helix growth parameter s and the nucleation parameter σ for amino acid residues. The calculated values of the two parameters fall into the ranges suggested by Zimm and Bragg.

In this paper, we generalized the molecular thermodynamic model as a model of the Gibbs energy of folding heteropolypeptides from random-coiled conformations into α-helical conformations. The generalized model incorporates the intrinsic folding potential due to residue–solvent interactions, the cooperative folding effect due to residue–residue interactions, and the location and length of α-helices. The utility of the model is demonstrated with investigations of α-helices of a number of natural and synthetic polypeptides. The contributions of side-chain interactions to helicity have been difficult to assess, both for homopolypeptide α-helices (Davidson & Fasman, 1967; Scheraga, 1978; Sueki et al., 1984; Wojcik et al., 1990) and for natural polypeptide segments (Kim & Baldwin, 1984; Marqusee et al., 1989; Padmanabhan et al., 1990). However, an extensive data set on the effect of amino acid sequence on α-helicity has been developed using the N-terminal amino acid sequence of ribonuclease A. The N-terminal 13 amino acids of ribonuclease A are helical in the crystal structure. Brown and Klee (1971) found that the S-peptide, the N-terminal 20 amino acid fragment cleaved by subtilisin from the native protein, was helical in solution at low temperature. Systematic studies on helicity in the S-peptide and its shorter 13 amino acid version, C-peptide, have been carried out by Baldwin and co-workers (Kim & Baldwin, 1984; Shoemaker et al., 1985; Marqusee & Baldwin, 1987). Kim and Baldwin (1984) showed that the helix formed by S-peptide is localized to residues 3–13 even in the isolated peptide. Five residues (14–18) with low helical propensity occur in a stretch next to the 3–13 helix. At 3 °C and pH 5, C-peptide was found to have ~25% helix content (Bierzynski et al., 1982).

THE MOLECULAR THERMODYNAMIC MODEL

The derivation of the molecular thermodynamic model for the Gibbs energy of folding for aqueous homopolypeptides has been presented by Chen et al. (1992). In this paper, we describe the treatments used to generalize the model for aqueous heteropolypeptides.

The Gibbs energy of folding residues in polypeptide chains is represented as the sum of two contributions: the local composition contribution and the configurational entropy contribution. The local composition contribution accounts for the local weak physical interactions that operate between residue and residue, residue and solvent, and solvent and solvent. The configurational entropy contribution accounts for the entropy loss of the aqueous polypeptide chains and solvent molecules, upon chain folding. Note that, in the earlier paper for aqueous homopolypeptides (Chen et al., 1992), the local composition contribution was treated as an enthalpy contribution.[1]

$$\Delta G = \Delta G^{lc} - T\Delta S^{config} \quad (1)$$

The Gibbs energy of folding the residues from a random-coiled state into an α-helix state is the Gibbs energy difference between the α-helix state α and the random-coiled state c.

$$\Delta G^{c\rightarrow\alpha} = \Delta G^{lc,c\rightarrow\alpha} - T\Delta S^{config,c\rightarrow\alpha} \quad (2)$$

In deriving the Gibbs energy expression, we chose to approximate the system as one single polypeptide chain with the amount of solvent molecules necessary to "solvate" all the amino acid residues on the random-coiled chain. In other words, the system consists of the polypeptide chain and the nearest neighbor solvent molecules that constitute the hydration shell. The water molecules that are not part of the hydration shell are excluded from the system. A typical coordination number of six is assumed in this treatment (Prausnitz et al., 1986). On the basis of our experience with molecular thermodynamic models, we believe the choice of the coordination number does not significantly alter the results.

The reference states were chosen to be pure liquid for water and a hypothetical amino acid residue aggregate state for residues in pure homopolypeptides. In this aggregate state for residues, the residues are surrounded by residues of the same amino acid. The reference state for water is characterized by the water-water physical interactions. The reference

[1] Notation: a, b, temperature coefficients of binary interaction parameters; c, conformational index; g, interaction energy; ΔG, Gibbs energy of folding per polypeptide chain; Δg, Gibbs energy of folding per residue; N, number of residues in a polypeptide chain; N_1, number of solvent molecules; N_2, number of polymer chain molecules of x segments; N_h, number of residues in a α-helical sequence, or helix length; N_s, helix starting residue; p, probability; R, gas constant; ΔS, entropy of folding per polypeptide chain; Δs, entropy of folding per residue; T, temperature (K); x, number of segments per polymeric chain; z, coordination number. Greek letters: τ, NRTL binary interaction energy parameter. Superscripts: α, α-helix; c, coil; FH, Flory–Huggins term; f, free water (or pure water state, unbonded to peptide chain); lc, local composition contribution; config, configurational entropy contribution; s, solvation state of water (bonded to peptide chain). Subscripts: i, j, k, any species or conformation; R, peptide unit, or "residue"; W, water.

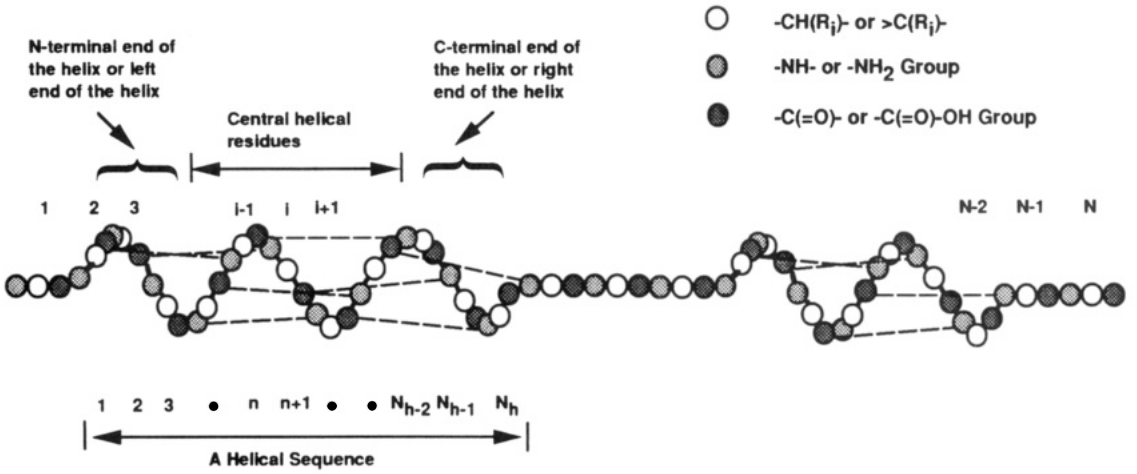

FIGURE 1: Definition of the α-helical conformation of a polypeptide chain.

state for residues is characterized by the residue-residue physical interactions.

The terminology used in the definition of an α-helical conformation of a polypeptide chain is defined in Figure 1. N is the total number of residues in the polypeptide, and i is the sequence index of the residue in the polypeptide chain. A helical sequence is defined as a series of consecutive residues in an α-helix. N_h is the number of helical residues in a specific helical sequence. If multiple helices do occur in a chain, there will be more than one N_h. The helical sequence index for a specific residue in the helical sequence is n. A central helical residue is meant to be a residue located at positions away from the left end (or N-terminal end) and the right end (or C-terminal end) of a helical sequence.

THE LOCAL COMPOSITION CONTRIBUTION

The nonrandom two-liquid (NRTL) theory of Renon and Prausnitz (1968) was adopted to represent the local composition contribution. This semiempirical theory provides a framework to account for the local compositions and the two-body interaction energies experienced between nearest-neighboring amino acid residues and solvent molecules in the system. It also allows us to conveniently address the sequence-dependent residue–residue interactions and the residue–solvent interactions.

For a given residue i in random-coiled conformations, we consider the interaction with the first neighboring species, including the i–1st and i+1st residues and the solvating water species. For a given residue i in helical conformations, as an approximation, we consider the interaction with the i–3rd, i–1st, i+1st, and i+3rd residues and the solvating water species, as shown in Figure 2. Therefore, the definitions for the interaction energies for a residue i in the random coil state and in the α-helix state are given as

$$g^{c}_{R(i)} = \frac{[g_{R(i-1),R(i)} + g_{R(i+1),R(i)} + 4g_{W,R(i)}]}{6} \tag{3}$$

$$g^{\alpha}_{R(i)} = \frac{[g_{R(i-3),R(i)} + g_{R(i-1),R(i)} + g_{R(i+1),R(i)} + g_{R(i+3),R(i)} + 2g_{W,R(i)}]}{6} \tag{4}$$

Here, $g_{R(j),R(i)}$'s represent the interaction energy between jth and ith residues along the peptide chain, $g_{W,R(i)}$ is the interaction energy between a water molecule and the ith residue, and the coordination number, z, is set to 6.

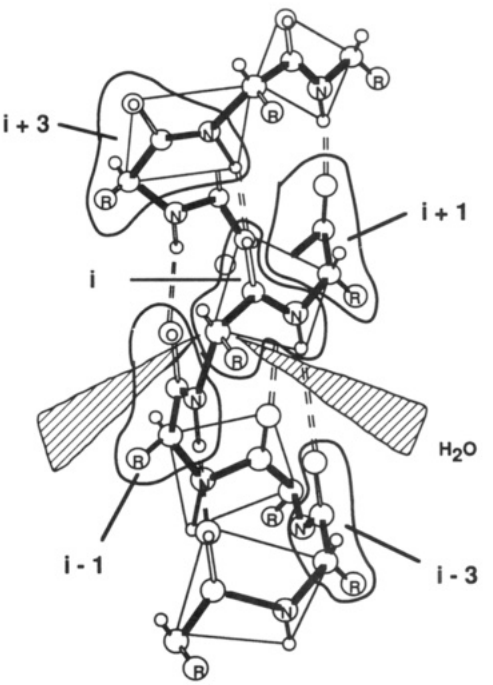

FIGURE 2: First neighbor interactions for a residue in the middle portion of an α-helical sequence.

It is necessary to take into consideration the end effect of the helix due to the deficit of side-chain–side-chain interactions. For instance, there always exist three residues at the N-terminus of a helix where their amino groups are not hydrogen bonded and three residues at the C-terminus of the helix in deficit of hydrogen bonding with their carboxyl groups. However, in the following discussion, only the derivation for central residues is given.

The interaction energies per water molecule in a solvation state (hydrogen bonded to amide groups or the side-chain groups of the polypeptide molecule) and a free state (free water beyond the hydration shell of the chain) are

$$g^{s}_{R(i),W} = \frac{[g_{R(i),W} + 5g_{W,W}]}{6} \tag{5}$$

$$g^{f}_{W} = g_{W,W} \tag{6}$$

where $g_{R(i),W}$ and $g_{W,W}$ are the interaction energies of contacts between amino acid residues i and a water molecule W and between two water molecules, respectively.

Given the local compositions and the interaction energies for the residues in the random-coil form and the α-helical form, and for the water molecules in the solvation state and the free water state, one can derive the expressions to determine the effective interaction energies for the system. Here, the interaction energy change caused by folding a central residue from its coil state to an α-helical state is

$$\Delta g_{R}^{c\rightarrow\alpha} = g_{R(i)}^{\alpha} - g_{R(i)}^{c} = \frac{[g_{R(i-3),R(i)} + g_{R(i+3),R(i)} - 2g_{W,R(i)}]}{6} = \frac{[\tau_{R(i-3),R(i)} + \tau_{R(i+3),R(i)} - 2\tau_{W,R(i)}]RT}{6} \quad (7)$$

The interaction energy change of a solvent molecule from its solvation state to its free state is

$$\Delta g_{R(i),W}^{s\rightarrow f} = g_{W}^{f} - g_{R(i),W}^{s} = \frac{g_{W,W} - g_{R(i),W}}{6} = \frac{-\tau_{R(i),W}RT}{6} \quad (8)$$

Here τ_{ij}'s are interaction energy parameters between residues or water i and j:

$$\tau_{ij} = \frac{g_{ij} - g_{jj}}{RT} \quad (9)$$

For each central residue, the formation of an α-helix releases two water molecules bonded to the residue. Therefore, the change of the system residue Gibbs energy for folding a central residue R(i) in the polypeptide chain from the random-coiled state into the α-helix state can be formulated as

$$\Delta g_{i}^{lc,c\rightarrow\alpha} = \Delta g_{R(i)}^{c\rightarrow\alpha} + 2\Delta g_{W,R(i)}^{c\rightarrow\alpha} \quad (10)$$

Equation 10 can be generalized for all residues:

$$\frac{\Delta g_{i}^{lc,c\rightarrow\alpha}}{RT} = -\frac{\Delta\tau_{i}}{z} \quad (11)$$

And here

$$\Delta\tau_{i} = 2[\tau_{W,R(i)} + \tau_{R(i),W}] - [\tau_{W,R(i)} + \tau_{R(i),W}] - \tau_{R(i+3),R(i)} \quad \text{(for N-terminal end residues)} \quad (12)$$

$$\Delta\tau_{i} = 2[\tau_{W,R(i)} + \tau_{R(i),W}] - [\tau_{R(i-3),R(i)} + \tau_{R(i+3),R(i)}] \quad \text{(for central residues)} \quad (13)$$

$$\Delta\tau_{i} = 2[\tau_{W,R(i)} + \tau_{R(i),W}] - [\tau_{W,R(i)} + \tau_{R(i),W}] - \tau_{R(i-3),R(i)} \quad \text{(for C-terminal end residues)} \quad (14)$$

Note that, for each residue at the two ends of the helical sequence, or residues 1, 2, and 3 and N_h–2, N_h–1, and N_h, the formation of an α-helix frees only one solvent molecule when folding into α-helical conformations.

In addition to the chain-end effect [represented by the term $-[\tau_{W,R(i)} + \tau_{R(i),W}]$], there are two basic contributions per residue: the contribution due to the intrinsic helix-forming potential of the residue as represented by the term $2[\tau_{W,R(i)} + \tau_{R(i),W}]$ and the contribution due to the cooperative effect of the neighboring residues to form α-helices (or the residue–residue interactions) as represented by the terms $\tau_{R(i-3),R(i)}$ and $\tau_{R(i+3),R(i)}$. This cooperative term cancels out in the Gibbs energy of folding formulation for aqueous homopolypeptides (Chen et al., 1992).

INTERACTION ENERGY PARAMETERS

The local composition contribution requires estimates for the binary interaction parameters, τ, to fully characterize the various weak two-body physical interactions that operate between nearest-neighboring residues and residues, residues and solvents, and solvents and solvents. These weak physical interactions may include hydrogen bonding, van der Waals interactions, "hydrophobic" interactions, local electrostatics, etc. Fully accounting for the various solvent–solvent, solvent–residue, and residue–residue interactions for aqueous polypeptide systems requires a 21 × 21 matrix of the binary interaction parameters. As discussed in our previous work, we estimated the binary interaction parameters with the UNIFAC (universal functional activity coefficient) group contribution method (Fredenslund et al., 1975).

The UNIFAC method was developed for the purpose of predicting activity coefficients for nonelectrolyte mixtures where little or no experimental data are available. In this method, each molecule is treated as the sum of the functional groups that constitute the molecule. The UNIFAC parameters for the contribution of each functional group were developed by regressing phase equilibrium data, including mutual solubility and vapor pressure measurements, of a wide range of nonelectrolyte systems composed of small molecules with various functional groups. We used the ASPEN PLUS software system (Aspen Technology, 1988) to identify the set of UNIFAC functional groups for each residue and to calculate the contribution of each functional group to the interaction parameters.

The estimation technique for the binary interaction parameters was carried out in two steps. First, infinite dilution activity coefficients were estimated for a given binary mixture with the UNIFAC method. Then, with the infinite dilution activity coefficients, the NRTL equation was solved for the NRTL binary interaction parameters, τ, with the NRTL nonrandomness factor set to 0.3. For more detailed information on the estimation technique for the interaction parameters, refer to our previous work (Chen et al., 1992).

The estimated 21 × 21 matrix of the binary interaction parameters at 25 °C are shown in Table I. In general, the estimated residue–water interactions appear to be stronger than the residue–residue interactions. Consequently, the model suggests the hydrophobic residue–water interactions play the primary role in polypeptide folding while the residue–residue interactions play a secondary role in the stabilization or destabilization of the folded structure.

The UNIFAC method is not applicable to mixtures with charged species or electrolytes because it does not accurately account for ion–ion interactions and ion–molecule interactions. In other words, the UNIFAC method does not adequately account for the local physical interactions between two charged residues and between solvent and charged residues. Therefore, the binary interaction parameters obtained from the UNIFAC method would underestimate the stabilization and destabilization effects of charged residues upon polypeptide chain folding.

Generally speaking, the magnitude of the binary interaction parameters decreases with temperature, since higher temperatures weaken the local physical interactions. As temperature is raised to a certain level, all mixtures essentially

Table I: Residue–Residue and Residue–Water Interaction Parameters (τ_{ij}) at 25 °C

	A	V	L	I	P	M	F	W	G	S	T
A	0.000	0.614	0.849	0.849	0.503	−1.231	−0.030	−0.100	−0.120	1.147	1.127
V	−0.470	0.000	0.429	0.429	0.342	0.261	0.698	0.889	−0.718	0.996	1.063
L	−0.594	−0.379	0.000	0.429	0.159	0.105	0.470	0.691	−0.808	0.953	0.996
I	−0.594	−0.379	−0.382	0.000	0.159	0.105	0.470	0.691	−0.808	0.953	0.996
P	−0.402	−0.300	−0.126	−0.126	0.000	0.359	0.788	0.909	−0.651	1.043	1.100
M	1.627	−0.066	0.268	0.268	−0.230	0.000	0.389	0.570	2.234	0.480	0.550
F	−0.082	−0.523	−0.261	−0.261	−0.671	−0.250	0.000	−0.463	−0.999	0.678	0.691
W	−0.051	−0.580	−0.308	−0.308	−0.721	−0.320	0.506	0.000	−1.134	−0.075	−0.073
G	0.176	1.271	1.593	1.593	1.070	−1.550	1.338	1.446	0.000	1.057	0.966
S	−0.433	0.379	0.798	0.798	0.209	0.168	0.818	1.389	−0.674	0.000	0.389
T	−0.620	−0.060	0.266	0.266	−0.173	−0.054	0.446	0.973	−0.718	−0.304	0.000
C	1.228	0.473	0.862	0.862	0.241	0.217	1.093	1.315	2.606	0.792	0.855
Y	6.695	4.283	3.421	3.421	4.984	4.803	1.751	1.425	8.070	−1.305	−1.291
N	−0.855	−0.131	0.271	0.271	−0.289	−0.325	0.339	0.946	−0.999	1.771	1.466
Q	−0.966	−0.419	−0.094	−0.094	−0.537	−0.463	0.056	0.631	−0.202	1.654	1.700
D	−0.912	0.219	0.775	0.775	−0.085	−0.825	0.429	0.896	−1.328	2.348	2.076
E	−1.160	−0.356	0.089	0.089	−0.574	−1.124	−0.007	0.466	−1.429	2.334	2.247
K	2.391	−0.496	−0.320	−0.320	0.986	1.117	0.409	0.537	2.677	−0.681	−0.667
R	−1.177	−0.600	−0.267	−0.267	−0.731	−0.667	−0.275	0.210	−0.137	3.217	3.176
H	−0.882	−0.188	0.202	0.202	−0.456	−0.263	0.111	0.122	−1.077	−0.637	−0.617
H_2O	2.220	3.988	4.920	4.920	3.834	5.548	7.704	7.825	1.409	2.881	3.277

	C	Y	N	Q	D	E	K	R	H	H_2O
A	−0.979	−3.032	1.778	1.667	2.224	2.351	−1.623	2.079	1.422	0.460
V	0.029	−2.321	1.868	1.855	1.275	1.496	0.584	2.331	0.953	0.976
L	−0.039	−1.962	1.865	1.845	1.060	1.197	0.547	2.355	0.745	1.271
I	−0.039	−1.962	1.865	1.845	1.060	1.197	0.547	2.355	0.745	1.271
P	0.078	−2.566	1.872	1.848	1.469	1.700	−0.822	2.318	1.087	0.932
M	−0.228	−2.536	1.258	1.264	1.482	1.808	−0.906	1.680	0.617	−1.486
F	−0.547	−1.275	1.442	1.439	1.328	1.429	−0.325	1.935	0.146	2.502
W	−0.634	−1.097	0.761	0.748	0.216	0.265	−0.113	1.211	−0.002	0.211
G	−1.727	−3.331	1.345	0.032	2.881	2.895	−1.754	−0.148	1.533	0.269
S	0.076	−0.071	−1.318	−1.248	−1.583	−1.573	0.839	−1.992	0.527	−1.288
T	−0.007	−0.160	−1.130	−1.278	−1.415	−1.539	0.527	−1.979	0.402	−1.114
C	0.000	0.154	1.583	1.630	−0.003	0.194	0.896	2.039	0.835	−0.674
Y	−0.745	0.000	−0.100	−0.171	0.096	0.240	−0.859	−2.026	−0.419	−0.530
N	−0.301	−0.654	0.000	0.366	−0.109	0.322	0.270	0.513	0.657	−2.137
Q	−0.372	−0.674	−0.298	0.000	0.048	−0.030	0.002	−0.443	0.678	−2.009
D	−0.567	0.483	−0.191	−0.254	0.000	0.399	1.425	−1.690	0.496	−1.063
E	−0.745	0.156	−0.527	−0.238	−0.300	0.000	1.657	−1.714	0.993	−0.939
K	−0.496	0.537	0.178	0.212	−0.939	−1.254	0.000	0.879	0.473	−1.459
R	−0.644	3.257	−0.436	0.493	2.546	2.559	−0.281	0.000	0.257	−1.315
H	−0.644	0.077	−0.260	−0.362	−0.409	−0.832	0.073	−0.070	0.000	−1.711
H_2O	3.113	5.487	3.954	4.276	3.220	3.689	4.380	0.006	3.029	0.000

become ideal solutions with τ's approaching zero. The temperature dependence of τ's is treated as follows:

$$\tau_{ij} = a_{ij} + \frac{b_{ij}}{T} \tag{15}$$

Here, a_{ij} and b_{ij} are regression constants determined from data within the temperature range under consideration. As an approximation, it is often assumed that b_{ij} is constant and a_{ij} equal to zero. In this work, we adopted this approximation.

THE CONFIGURATIONAL ENTROPY CONTRIBUTION

The configurational entropy of mixing a disoriented polymer and the solvents to form a random-coiled polymer solution is given by Flory (1941, 1942) and Huggins (1941, 1942):

$$\Delta S_{\text{mix}}^{\text{FH}} = -R\left[N_1 \ln\left(\frac{N_1}{N_1 + N_2 x}\right) + N_2 \ln\left(\frac{N_2 x}{N_1 + N_2 x}\right)\right] \tag{16}$$

Here, N_1 is the number of solvent molecules, N_2 is the number of polymer chain molecules of x segments long, and x is the number of the segments in a polymer chain.

When a residue folds from a random-coiled state to an α-helical state, the configurational entropy loss can be approximated as the difference between that of mixing solvents with a polymer chain of length n and that of mixing solvents with a polymer chain of length n–1.

$$\Delta s^{\text{config,c}\to\alpha} = (\Delta S^{\text{c,FH}}|_{n-1} - \Delta S^{\text{c,FH}}|_{n}) \tag{17}$$

THE GIBBS ENERGY OF FOLDING

The Gibbs energy of folding a residue i from a random-coiled state to an α-helical state can be summarized as follows:

$$\frac{\Delta g_i^{\text{c}\to\alpha}}{RT} = -\frac{\Delta \tau_i}{z} - \frac{\Delta s_i^{\text{config,c}\to\alpha}}{R} \tag{18}$$

Ignoring the chain-end effect and the cooperative effect, we can define an "intrinsic" Gibbs energy of folding for each residue i as follows:

$$\frac{\Delta g_i^{\text{c}\to\alpha^*}}{RT} = -\frac{\Delta \tau_i^*}{z} - \frac{\Delta s_i^{\text{config,c}\to\alpha}}{R} \tag{19}$$

where

$$\Delta \tau_i^* = 2[\tau_{\text{W,R}(i)} + \tau_{\text{R}(i),\text{W}}] \tag{20}$$

Table II shows the individual contributions of the 20 residues in the S-peptide of RNase to the Gibbs energy of folding and the elements of the Gibbs energy of folding for each residue

Table II: Individual Contributions of the 20 Residues in S-Peptide to the Gibbs Energy of Folding at 0 °C

i	residue	$\Delta g_i^{c\to\alpha}/RT$	intrinsic, $\Delta g_i^{c\to\alpha^*}/RT$	cooperative $\tau_{R(i-3),R(i)}/z$	cooperative $\tau_{R(i+3),R(i)}/z$	chain-end effect, $[\tau_{R(i),W} + \tau_{W,R(i)}]/z$
1	K	−0.027	−0.261		−0.295	0.529
2	E	0.727	−0.199		0.428	0.498
3	T	0.610	0.012		0.205	0.393
4	A	0.695	−0.175	0.435	0.435	
5	A	−0.401	−0.175	−0.211	−0.015	
6	A	−0.499	−0.175	−0.113	−0.211	
7	K	−0.608	−0.262	−0.295	−0.051	
8	F	−2.901	−2.905	−0.006	0.010	
9	E	0.077	−0.199	0.428	−0.151	
10	R	1.739	1.273	0.160	0.306	
11	Q	0.191	−0.024	0.262	−0.046	
12	H	0.596	0.320	0.181	0.096	
13	M	−0.765	−0.674	−0.121	0.031	
14	D	−0.233	0.016	0.009	−0.257	
15	S	0.104	0.220	−0.116	0.000	
16	S	0.516	0.220	0.087	0.209	
17	T	0.595	0.012	0.378	0.205	
18	S	0.509	0.219	0.000		0.290
19	A	0.233	−0.174	−0.079		0.486
20	A	0.199	−0.174	−0.113		0.486
total		1.358				

in the chain at 0 °C. Here we assumed the S-peptide has a 20-residue helical structure. $\Delta g_i^{c\to\alpha}/RT$ is the Gibbs energy of folding for each residue i; $\Delta g_i^{c\to\alpha^*}/RT$ is the intrinsic helix-forming tendency of each residue i; $\tau_{R(i-3),R(i)}/z$ and $\tau_{R(i+3),R(i)}/z$ are the cooperative contributions from the ith and the i–3rd residue interaction and the ith and i+3rd residue interaction, respectively; and $[\tau_{R(i),W} + \tau_{W,R(i)}]/z$ is the chain-end effect due to the deficit of side-chain contacts at the two ends of the helix. The model suggests that the Gibbs energy of folding a residue in a polypeptide chain is determined by both the intrinsic conformational preferences of the individual residues in a specific solvent and their cooperative interactions with neighboring residues in the chain.

The Gibbs energy of folding a polypeptide chain from a random-coiled conformation to an α-helical conformation j can be summarized as follows:

$$\frac{\Delta G_j^{c\to\alpha}}{RT} = \sum_i \frac{\Delta g_i^{c\to\alpha}}{RT} c_{ij} \qquad (21)$$

Here, c_{ij} is the conformational index of the residue i in the polypeptide chain in α-helical conformation j. It accounts for the location and length of the helix. The value of c_{ij} is 0 for residues in coil state and 1 for residues in helical state.

POLYPEPTIDE CONFORMATIONS IN AQUEOUS SOLUTIONS

Given the model for the Gibbs energy of folding for α-helical conformations, we developed a procedure to search through the Gibbs energy surface for the various conformations and to identify the most stable α-helical conformation of the polypeptide chain in water. We assumed there is only one helix in each α-helical conformation. This assumption is adequate for small peptides. Each α-helical conformation is defined by the location of the helix-starting residue, N_s, and the length of the α-helix, N_h. Note that N_s could be any residue of the chain except the last three residues at the C-terminal end and N_h could take any value from 4 to N since the minimum length of an α-helix is 4. The search was performed according to the following steps:

Step 1. Choose a helix-starting residue N_s. The possible starting residue could be any residue from residues 1 to N–4 of the polypeptide chain.

Table III: Calculated Thermodynamic Properties of Folding the S-Peptide at 0 °C with Helix-Starting Residue 1

chain length N_h	$\sum\Delta G_j^{c\to\alpha}/RT$	$\sum\Delta G_j^{lc,c\to\alpha}/RT$	$\sum\Delta S_j^{c\to\alpha}/R$
4 (min)	2.3158	−0.8766	−3.1923
5	2.3454	−1.6450	−3.9904
6	2.3563	−2.4322	−4.7885
7	2.2778	−3.3087	−5.5866
8	0.7174	−5.6673	−6.3847
9	0.7470	−6.4357	−7.1827
10	1.3617	−6.6191	−7.9808
11	0.1688	−8.6101	−8.7789
12	0.2585	−9.3185	−9.5770
13	0.7424	−9.6327	−10.3751
14	0.7008	−10.4724	−11.1732
15	0.9506	−11.0207	−11.9712
16	0.8412	−11.9282	−12.7693
17	0.9754	−12.5920	−13.5674
18	1.1952	−13.1703	−14.3655
19	1.3477	−13.7265	−15.1636
20	1.3586	−14.6030	−15.9617

Step 2. Choose a helical length N_h. Let the helix start from the smallest helical length of N_h = 4.

Step 3. Calculate $\Delta G_j^{c\to\alpha}$, the Gibbs energy of folding the polypeptide chain into a helical conformation j with a starting residue N_s and a helical length N_h.

Step 4. Go back to step 2 with the helix length increased by 1. Repeat the same calculations as in step 3. Compare the Gibbs energy of folding. The α-helical conformation with the lowest $\Delta G_j^{c\to\alpha}$ is considered as the most stable α-helical conformation with the helix-starting residue N_s.

Step 5. Go back to step 1 with the helix-starting residue moving along the polypeptide chain. Repeat steps 2–4.

Step 6. After calculating the $\Delta G_j^{c\to\alpha}$'s for all the potential α-helical conformations, compare the $\Delta G_j^{c\to\alpha}$'s for all the cases. The α-helical conformation with the lowest Gibbs energy of folding is the most stable α-helical conformation.

Table III shows the Gibbs energy of folding for the S-peptide and the individual contributions, $\Delta G^{lc,c\to\alpha}/RT$ and $\Delta S^{c\to\alpha}/R$ (both are dimensionless), versus the helical length N_h, with helix starting from residue number 1. The data indicate that the most stable helix (with N_s = 1) in water at 0 °C consists of residues 1–11 with the lowest Gibbs energy of folding, $\Delta G^{c\to\alpha}/RT$, at 0.1688.

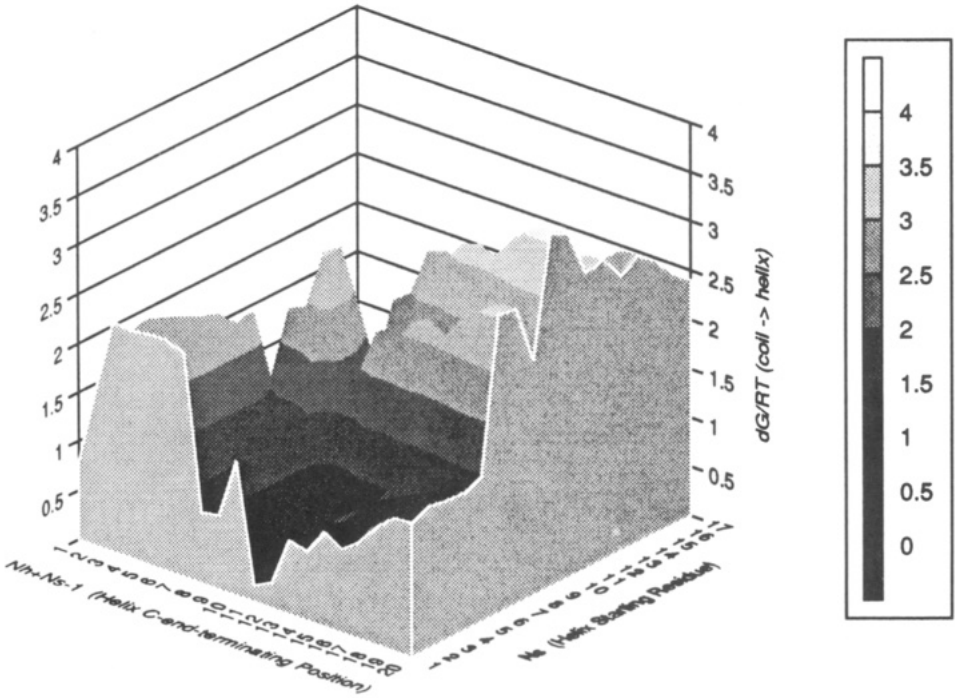

FIGURE 3: Gibbs energy surface of folding of S-peptide into an α-helical conformation.

Shown in Figure 3 is the three-dimensional contour map for the Gibbs energy of folding for the 20-residue long S-peptide as a function of the helix-starting residue, N_s, and the helix length, N_h. It can be seen that, when the helices start from residues 1, 2, 3, 4, and 5 with respective helix chain length of 11, 10, 9, 8, and 7, the conformations with these helical sequences have relatively lower Gibbs energies. Figure 3 also shows clearly that the helical conformation with the lowest Gibbs energy has a helix starting at residue 1 of the peptide, and the helix is 11 residues long. Moreover, in comparison to the thermal energy RT, the valleys and hills of the Gibbs energy surface are relatively shallow, and the marginal stability of the folded helices is suggested.

Figure 3 further suggests that a polypeptide chain should have multiple helical conformations. In other words, the folded form of the peptide chain is a mixture of α-helical conformations with different helical sequences such as 1–11, 2–11, 3–11, 4–11, and 5–11.

Note that the search procedure above assumes that there would only be one helix for each polypeptide chain. The search procedure needs to be modified when one considers multiple helices within a polypeptide chain. The consideration of multiple helices within a polypeptide chain is essential for long polypeptide chains.

HELIX POPULATION AND HELICITY

The conformational transition of a polypeptide chain can be treated as a chemical equilibrium process. The Gibbs energy of folding the random-coiled chain into a helical conformation j is related to the equilibrium constant between the two states:

$$\frac{\Delta G_j^{c\rightarrow\alpha}}{RT} = -\ln K_j = -\ln \frac{[\alpha\text{-helix}_j]}{[\text{coil}]} \tag{22}$$

Here, with activity coefficients assumed to be unity, [α-helix$_j$] and [coil] are the concentrations of the polypeptide molecules in helical conformation j and coil conformation, respectively. The helicity based on a "two-state" model can be defined as

$$[\%\ \text{helicity}] = \frac{100\dfrac{N_h}{N}e^{-(\Delta G_j^{c\rightarrow\alpha}/RT)}}{1 + e^{-(\Delta G_j^{c\rightarrow\alpha}/RT)}} \tag{23}$$

The lower the Gibbs energy of folding is, the higher the helical content will be. For example, the model predicts the most favorable helical conformation of S-peptide at 0 °C is the 1–11 helical sequence with $\Delta G^{c\rightarrow\alpha}/RT = 0.1688$. With the "two-state" model, the helical content of this specific conformation is 25%, comparable to the 30% helicity value obtained for C-peptide at 0 °C and pH 5 (Bierzynski et al., 1982) and for S-peptide under the conditions 0 °C, pH 5.4, and 1 M NaCl (Rico et al., 1983). Note that the definitions for the helicity are different. According to Bierzynski et al. (1982), their experimental helicity was defined as the ratio of the $[\theta]_{222}$ for the peptide to the $[\theta]_{222}$ for "complete helix formation". They treated the $[\theta]_{222}$ (from CD spectra) of "100% helicity" as an averaged value of $[\theta]_{222}$ when unfolded S-peptide recombines with folded S-peptide, if the change is attributed entirely to the 10 peptide groups of the helix formed by residues 3–12. Alternatively, the $[\theta]_{222}$ can be provided by the α-helix basis spectrum compiled from 15 proteins with an average helix length of around 10 residues. They assumed that the helix is completely melted out at high temperatures and used $[\theta]_{222}$ at 45 °C as a baseline for "0% helicity".

Considering the dispersity of the helix population, we can define the overall helicity of a polypeptide in solution. According to statistical thermodynamics, the probability of a given polypeptide in the α-helical conformation j, in analogy to a quantum state of the canonical ensemble (Prausnitz et al., 1986), with the Gibbs energy $\Delta G_j^{c\rightarrow\alpha}/RT$, is

$$p_j = \frac{e^{-(\Delta G_j^{c\rightarrow\alpha}/RT)}}{1 + \sum_k^{\text{all }\alpha} e^{-(\Delta G_k^{c\rightarrow\alpha}/RT)}} \tag{24}$$

The summation here is over all possible helical conformations

Table IV: Amino Acid Sequences of Some Natural and Synthetic Polypeptides

no.	polypeptides	sequence	note	ref
1	C-peptide	K-E-T-A-A-A-K-F-E-R-Q-H-M	1–13 of RNase A	*a*
2	S-peptide	K-E-T-A-A-A-K-F-E-R-Q-H-M-D-S-S-T-S-A-A	1–20 of RNase A	*a*
3	Pα	N-N-F-K-S-A-E-D-C-M-R-T-A-G-G-A	43–58 of BPTI	*b*
4	Pα5	S-A-E-D-C-M-R-T-A-G-G-A	47–58 of BPTI	*c*
5	3K(I)	A-A-A-A-K-A-A-A-A-K-A-A-A-A-K-A	alanine-based	*d*
6	3K(II)	A-K-A-A-A-A-K-A-A-A-A-K-A-A-A-A	alanine-based	*d*
7	4K	A-K-A-A-K-A-A-A-A-K-A-A-A-A-K-A	alanine-based	*d*
8	6K(I)	A-K-A-A-K-A-K-A-A-K-A-K-A-A-K-A	alanine-based	*d*
9	6K(II)	A-K-A-A-A-K-K-A-A-A-K-K-A-A-A-K-A	alanine-based	*d*
10	3E	A-E-A-A-A-A-E-A-A-A-A-E-A-A-A-A	alanine-based	*d*
11	(I+3)E,K	A-E-A-A-K-A-E-A-A-K-A-E-A-A-K-A	alanine-based	*d,f*
12	(I+4)E,K	A-E-A-A-A-K-E-A-A-A-K-E-A-A-A-K-A	alanine-based	*d,f*
13	C-RN16	K-E-T-A-A-A-K-F-L-R-A-H-A	C-peptide analogue	*e*
14	C-RN21	A-E-T-A-A-A-K-F-L-R-A-H-A	C-peptide analogue	*e*
15	C-RN23	A-A-T-A-A-A-K-F-L-R-A-H-A	C-peptide analogue	*e*
16	C-RN28	A-E-T-A-A-A-K-F-L-A-A-H-A	C-peptide analogue	*e*
17	C-RN54	A-A-T-A-A-A-K-F-L-A-A-H-A	C-peptide analogue	*e*
18	C-RN25	A-R-T-A-A-A-K-F-L-E-A-H-A	C-peptide analogue	*e*
19	C-RN26	A-D-T-A-A-A-K-F-L-R-A-H-A	C-peptide analogue	*e*
20	C-RN80	A-E-T-A-A-A-K-Y-L-R-A-H-A	C-peptide analogue	*e*
21	C-RN121	A-A-T-A-A-A-K-Y-L-R-A-H-A	C-peptide analogue	*e*
22	C-RN119	A-E-T-A-A-A-K-Y-L-A-A-H-A	C-peptide analogue	*e*
23	C-RN120	A-A-T-A-A-A-K-Y-L-A-A-H-A	C-peptide analogue	*e*
24	C-RN84	A-E-T-A-E-A-K-Y-L-R-A-H-A	C-peptide analogue	*e*
25	(I+4)K,E	A-K-A-A-A-E-K-A-A-A-E-K-A-A-A-E-A	de novo design	*f*
26	(I+3)K,E	A-K-A-A-E-A-K-A-A-E-A-K-A-A-E-A	de novo design	*f*
27	3ALA	Y-K-A-A-A-A-K-A-A-A-A-K-A-A-A-A-K	alanine-based	*g*
28	3LEU	Y-K-A-A-L-A-K-A-A-L-A-K-A-A-L-A-K	alanine-based	*g*
29	3PHE	Y-K-A-A-F-A-K-A-A-F-A-K-A-A-F-A-K	alanine-based	*g*
30	3ILE	Y-K-A-A-I-A-K-A-A-I-A-K-A-A-I-A-K	alanine-based	*g*
31	1VAL	Y-K-A-A-A-A-K-A-A-V-A-K-A-A-A-A-K	alanine-based	*g*
32	2VAL	Y-K-A-A-V-A-K-A-A-V-A-K-A-A-A-A-K	alanine-based	*g*
33	3VAL	Y-K-A-A-V-A-K-A-A-V-A-K-A-A-V-A-K	alanine-based	*g*

[a] Kim and Baldwin (1984). [b] Oas and Kim (1988). [c] Goodman and Kim (1989). [d] Marqusee et al. (1989). [e] Fairman et al. (1990). [f] Marqusee and Baldwin (1987). [g] Padmanabhan et al. (1990).

of the polypeptide chain. Then the α-helicity for all the possible helical conformations can be defined by

$$[\%\ \text{helicity}] = 100\sum_{k}^{\text{all}\ \alpha}\frac{N_{h,k}}{N}p_k \qquad (25)$$

With this population formulation the helicity for S-peptide at 0 °C is 50%. This value is much higher than the 25% helicity of the "two-state" model, and it differs from the reported experimental value of 30% helicity due to the different definitions of helicity.

PREDICTION OF POLYPEPTIDE CONFORMATIONS

Helix-forming natural and synthetic polypeptides provide excellent systems for testing the utility of the simple model. For example, according to Kim and Baldwin (1984), the C-peptide helix may function as an autonomous folding unit, and helix formation in isolated S-peptide (residues 1–20) is localized in a manner resembling that in the intact protein.

Table IV lists the amino acid sequences for 33 polypeptides studied in this paper. Table V gives the computed minimum Gibbs energy of folding of these natural and synthetic polypeptides and a comparison of the predicted helical conformations and helicities with experimental observations. It should be noted that the calculations were carried out at neutral pH. Although most of the ionizable groups on the side chain of the amino acid residues and the amino and carboxyl groups at the chain ends will be charged at this pH, a reasonable match between the model predictions and the experimental observations was obtained for a number of polypeptides without taking into account the charge effect.

The partial helix formation in both C- and S-peptide in ribonuclease A at 0 °C is a good test case for the model. C-peptide and S-peptide contain the first 13 and 20 residues of bovine pancreatic ribonuclease A (RNase A), respectively:

```
Lys-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Glu-Arg-Gln-His-Met-Asp-Ser-Ser-Thr-Ser-Ala-Ala
1                                  13                                  20
```

As shown in Table V (and Figure 3), the model suggests that the most favored helix in both C- and S-peptide proceeds from residues 1 to 11. In other words, by considering the chain-end effect, the residue–solvent interactions, and the residue–residue interactions, the model suggests that the residues 1–11 form the most stable α-helical sequence. Residues 12–20 are excluded from the helix sequence of S-peptide due to either the unfavorable residue–solvent interactions (i.e., the residues are inherently helix-breaking) or the unfavorable residue–residue interactions. The model predictions are in agreement with Kim and Baldwin's finding (1984) that the S-peptide helix terminates before residue Thr17, and it is not possible to tell whether Met13 is contained in the helix formed by the isolated S-peptide. Moreover, Kim and Baldwin did not find measurable helix shifts of Ala19 and Ala20. In native RNase A, residues 3–13 are helical (Wlodawer & Sjolin, 1983).

The molecular thermodynamic model appears to satisfactorily represent the contributions of the noncharged side chains to helical stability. However, the charged side-chain–helix dipole interaction may account for part of the helicity observed. In trying to assess the contributions to the helicity of their C-peptide analogues, Shoemaker et al. (1987) replaced His12 with Ala. The resulting destabilization indicated that these charged side chains were interacting with the helix macrodipole. Further experiments with substituents on the N-terminus supported this interpretation. Such charged side-chain–helix dipole interaction is not accounted for in the current model.

Table V: Comparison of the Model-Predicted and Observed Conformations of Polypeptides

no.	polypeptides	min $\Delta G_j^{c\to\alpha}/RT$	most stable α-helical conformation: model	most stable α-helical conformation: literature[a]	% helicity: two-state model	% helicity: complete model	% helicity: literature[a]
1	C-peptide	0.1688 (0 °C)	1–11	3–13	38.75%	59.90%	
2	S-peptide	0.1688 (0 °C)	1–11	3–13	25.18%	50.37%	
3	Pα	0.5356 (0 °C)	45–52		18.46%	46.80%	
4	Pα5	1.9964 (0 °C)	48–52		4.98%	36.65%	
5	3K(I)	0.6481 (1 °C)	1–16		34.34%	55.93%	72% helix at 1 °C, pH 7, and 0.01 M NaCl solution
6	3K(II)	0.6481 (1 °C)	1–16		34.34%	55.93%	67% helix at 1 °C, pH 7, and 0.01 M NaCl solution
7	4K	0.4656 (1 °C)	1–16		38.57%	56.93%	54% helix at 1 °C, pH 7, and 0.01 M NaCl solution
8	6K(I)	0.2925 (1 °C)	1–16		42.74%	57.95%	19% helix at 1 °C, pH 7, and 0.01 M NaCl solution
9	6K(II)	0.9566 (1 °C)	1–17		27.76%	53.08%	26% helix at 1 °C, pH 7, and 0.01 M NaCl solution
10	3E	1.1866 (1 °C)	1–16		23.39%	52.16%	71% helix at 1 °C, pH 7, and 0.01 M NaCl solution
11	(I+3)E,K	0.8201 (1 °C)	1–16		30.58%	54.19%	49% helix at 1 °C, pH 7, and 0.01 M NaCl solution
12	(I+4)E,K	1.4951 (1 °C)	1–17		18.32%	49.29%	80% helix at 1 °C, pH 7, and 0.01 M NaCl solution
13	C-RN16	−1.2660 (3 °C)	1–12		72.01%	67.74%	
14	C-RN21	−1.3613 (3 °C)	1–12		73.47%	68.08%	
15	C-RN23	−1.5635 (3 °C)	1–12		76.32%	69.19%	
16	C-RN28	−2.2111 (3 °C)	1–13		90.12%	73.34%	
17	C-RN54	−2.4133 (3 °C)	1–13		91.78%	74.43%	
18	C-RN25	−2.0180 (3 °C)	4–12		61.11%	69.72%	
19	C-RN26	−1.2545 (3 °C)	4–12		53.87%	67.39%	
20	C-RN80	1.5576 (3 °C)	1–9		12.05%	46.11%	
21	C-RN121	1.3555 (3 °C)	1–9		14.19%	47.35%	
22	C-RN119	1.0328 (3 °C)	1–13		26.25%	54.68%	
23	C-RN120	0.8307 (3 °C)	1–13		30.35%	55.91%	
24	C-RN84	0.2066 (3 °C)	1–13		44.85%	61.07%	
25	(I+4)K,E	1.4951 (1 °C)	1–17		18.32%	49.29%	70% helix at 1 °C, pH 7, and 0.01 M NaCl solution
26	(I+3)K,E	0.8201 (1 °C)	1–16		30.58%	54.19%	24% helix at 1 °C, pH 7, and 0.01 M NaCl solution
27	3ALA	0.7000 (0 °C)	2–17		31.23%	55.80%	78% helix at 0 °C, pH 7, and 1 M NaCl solution
28	3LEU	−2.6856 (0 °C)	2–17		88.11%	75.33%	80% helix at 0 °C, pH 7, and 1 M NaCl solution
29	3PHE	−6.5404 (0 °C)	2–17		93.98%	85.18%	23% helix at 0 °C, pH 7, and 1 M NaCl solution
30	3ILE	−2.6856 (0 °C)	2–17		88.11%	75.33%	41% helix at 0 °C, pH 7, and 1 M NaCl solution
31	1VAL	−0.2261 (0 °C)	2–17		52.36%	62.38%	49% helix at 0 °C, pH 7, and 1 M NaCl solution
32	2VAL	−1.1523 (0 °C)	2–17		71.52%	67.52%	34% helix at 0 °C, pH 7, and 1 M NaCl solution
33	3VAL	−1.6904 (0 °C)	2–17		79.46%	70.70%	17% helix at 0 °C, pH 7, and 1 M NaCl solution

[a] References listed in Table IV.

The model does not account for the contribution of ion pairing to helicity either. Oppositely charged residues located in the $i\pm3,4$ positions can form ion pairs on a face of the helix, stabilizing it in aqueous solution (Marqusee & Baldwin, 1987). Marqusee and Baldwin showed that removing the ion pairs, while leaving sufficient charged side chains to maintain solubility, did not prevent helix formation. The apolar side chains are making important contributions to helicity.

The Pα helix of BPTI (bovine pancreatic trypsin inhibitor) represents another example of local helix formation. Peptide Pα in BPTI is one of the two subdomains important for the first crucial intermediates in the folding of BPTI. It contains 16 residues corresponding to residues 43–58:

```
Asn-Asn-Phe-Lys-Ser-Ala-Glu-Asp-Cys-Met-Arg-Thr-Ala-Gly-Gly-Ala
43                                                          58
```

which include the C-terminal α-helix and a short β-strand in BPTI (Oas & Kim, 1988). Goodman and Kim (1989) reported that, at low temperature (0 °C), peptide Pα5, corresponding to residues 47–58, can form an α-helix in isolation.

The model suggests that Pα forms a most favored α-helix spanning residues 45–52 at 0 °C. The three-dimensional contour map of the Gibbs energy of folding of Pα peptide into α-helical conformation as a function of helical length and helical starting residue is given in Figure 4.

SITE-DIRECTED MUTATION

The model not only provides a methodology to predict preferred helical conformations and helical population but also facilitates the prediction of the changes in helical conformations and helicities associated with site-directed mutation. Such site-directed peptide mutation lays important foundations for studies in identifying active sites of enzymes and proteins.

Marqusee et al. (1989) reported that short, 16-residue, alanine-based polypeptides show stable monomolecular α-helix formation in water. The alanine-based polypeptides were solubilized by insertion of three or more residues of a single charge type, lysine(+) or glutamic acid(−). These 16-residue, alanine-based polypeptides have been designed such that the results cannot be explained by ion-pair formation or by charged group–helix dipole interactions. Therefore, these polypeptides are appropriate model systems to test our theory and the estimated interaction parameters. Table V shows that the model positively predicts significant helix formation for these polypeptides in aqueous solution at the experimental conditions. However, the model fails to correctly predict the trends associated with the helical content with site-directed mutation by the charged residues, due to the fact that the model does not properly account for the electrostatic interactions derived from charged side chains.

Another example is the study on the helix stability upon amino acid substitution for the 13-residue-long C-peptide analogues investigated experimentally by Fairman et al. (1990). As seen in Table V, the model indicates that, at 3 °C, peptide C-RN16 helix is less stable than C-RN21 after the substitution of Lys1 for Ala1 of the reference peptide C-RN16, consistent with the experimentally reported helix-

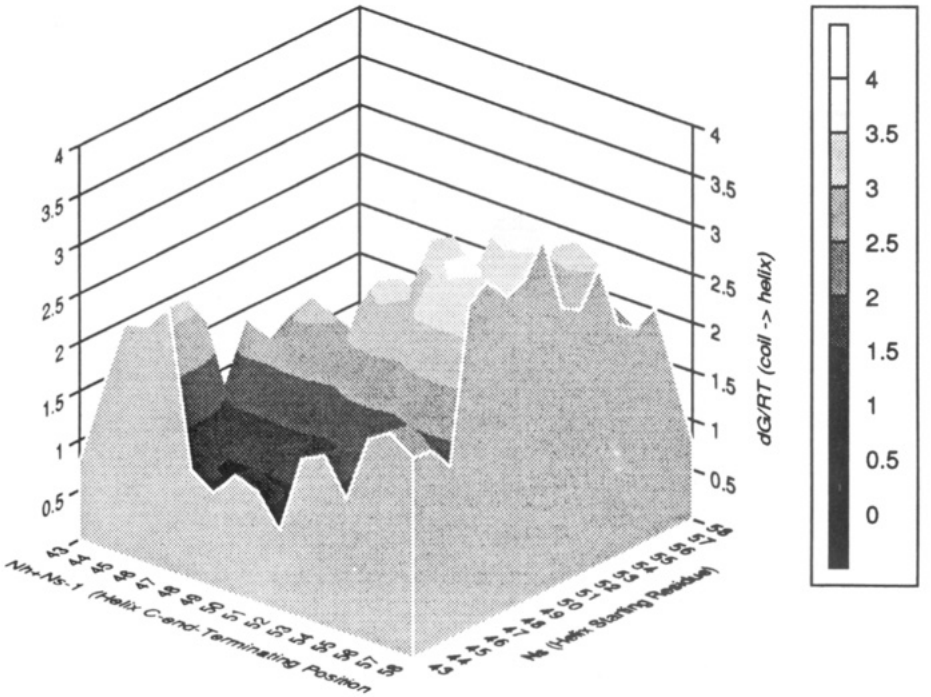

FIGURE 4: Gibbs energy surface of folding of Pα peptide into an α-helical conformation.

destabilizing replacement. The model-predicted helix stability increase for the replacement of Arg10 (peptide C-RN23) with Ala10 (C-RN54) in an Ala2 background also agrees with the Fairman et al. measurement. According to the model, the substitution of Ala2 and Arg10 in C-RN23 by Glu2 and Ala10 (peptide C-RN28) increases the α-helicity of the peptide, the replacement of Glu2 and Ala10 in C-RN119 by Ala2 and Arg10 (C-RN121) in a Tyr8 background lowers the helicity, and the mutation of Ala2 in C-RN120 by Glu2 (C-RN119), also in Tyr8 background, escalates the helix stability. Except for the mutation of C-RN120 to C-RN119, the predictions are in good agreement with the observations of Fairman et al. (1990).

Padmanabhan et al. (1990) synthesized a 17-residue alanine-based peptide 3ALA (see Table V) as a reference for further mutation studies. By substituting each of the five different apolar amino acids (Ala, Ile, Leu, Phe, Val) in turn for alanine in the 17-residue alanine-based peptide and determining the extent of α-helix formation, they measured and derived the relative helix-forming tendencies of the five apolar amino acids. Their rank order is Leu > Ala > Ile > Phe > Val. While our model, without adequate consideration of the steric hinderance effect, suggests these five apolar amino acids have the following rank order for the relative helix-forming tendencies: Phe > Leu = Ile > Val > Ala.

Recently, Merutka et al. (1991) and Lyu et al. (1992) investigated the quantitative role of salt bridges by altering the amino acid sequence of several short synthetic polypeptides. The present model is not capable of accurately predicting the effect of helix-stabilizing salt bridges because the model does not yet incorporate contributions from the electrostatic interactions involved in salt bridges.

TEMPERATURE EFFECT

Temperature has a profound influence on the conformational stability of polypeptides. The model predictions suggest that thermal unfolding of polypeptides does not follow a two-state mechanism; rather, the polypeptide chains unfold gradually as temperature rises. For example, as shown in Figure 5, the

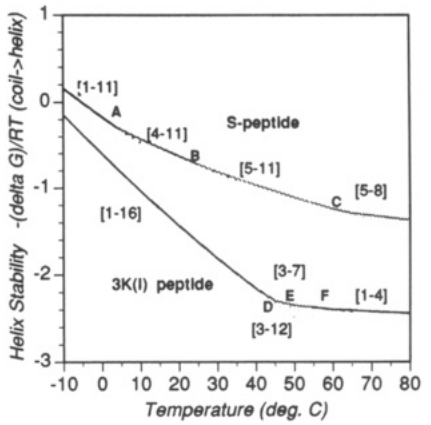

FIGURE 5: Predicted thermal stability of natural and synthetic polypeptides in water.

stable S-peptide helix 1–11 unfolds to a less stable helix 4–11 at a temperature near 5 °C (point A on the first curve), and 4–11 unfolds into a shorter, much less stable helix 5–11 (point B) at 30 °C, and this short helix unfolding occurs at 65 °C (point C of the curve). The 5–8 helical sequence unfolds further at elevated temperatures. This prediction appears to be consistent with the experimental results of Kim and Baldwin (1984). They reported that helix formation is strongly dependent on temperature, and S-peptide α-helix is essentially absent at 47 °C and is populated at 0 °C. This result is also in good agreement with the experimentally obtained thermal unfolding curve for C-peptide lactone (Bierzynski et al., 1982).

A similar thermal unfolding phenomenon is also predicted for the synthetic peptide 3K(I), a 16-residue alanine-based peptide. The unfolding conformational transitions take place at 45, 50, and 60 °C, corresponding to points D, E, and F on the lower curve of Figure 5, respectively. The model-predicted thermal unfolding curve is in excellent match with the experimentally observed thermal unfolding curve for 3K(I) peptide (Marqusee et al., 1989).

Moreover, Figure 5 illustrates that the model-predicted thermal unfolding of the alanine-based peptides is very broad,

spanning from 0 to 70 °C as Marqusee et al. (1989) reported experimentally.

CONCLUSIONS

Molecular thermodynamics offers a semiempirical approach to the development of Gibbs energy expressions for the study of the phase behavior of protein systems. The molecular thermodynamic model presented in this work effectively represents the Gibbs energy of folding for natural and synthetic polypeptides from random-coiled conformations into α-helical conformations. The model gives reasonable predictions regarding the formation of stable α-helix conformations, the location and length of each α-helical sequence, and the helical populations. Encouraging results were also obtained in relation to site-directed mutation studies and thermal unfolding studies.

Further investigations on the refinement of the molecular thermodynamic model to incorporate electrostatic interactions and the utility of the molecular thermodynamic model in representing polypeptide chain folding for larger protein molecules and protein systems with solutes are warranted.

ACKNOWLEDGMENT

We are grateful to Professor P. S. Kim of the Massachusetts Institute of Technology and Howard Hughes Medical Institute Research Laboratories for valuable discussions and comments on this work.

REFERENCES

Alonso, D. O. V., & Dill, K. A. (1991) *Biochemistry 30*, 5974–5985.
Anfinsen, C. B. (1973) *Science 181*, 223–230.
Aspen Technology (1988) *ASPEN PLUS User Guide*, Cambridge, MA.
Ben-naim, A. (1990) *Biopolymers 29*, 567–596.
Bierzynski, A., Kim, P. S., & Baldwin, R. L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2470–2474.
Blankschtein, D., Thurston, G. M., & Benedek, G. B. (1985) *J. Chem. Phys. 85*, 7268–7288.
Brown, J. E., & Klee, W. A. (1971) *Biochemistry 10*, 470–476.
Chen, C.-C., & Evans, L. B. (1986) *AIChE J. 32*, 444–454.
Chen, C.-C., Zhu, Y., & Evans, L. B. (1989) *Biotechnol. Prog. 5*, 111–118.
Chen, C.-C., Zhu, Y., King, J. A., & Evans, L. B. (1992) *Biopolymers 32*, 1375–1392.
Chou, P. Y., & Fasman, G. D. (1974a) *Biochemistry 13*, 211–222.
Chou, P. Y., & Fasman, G. D. (1974b) *Biochemistry 13*, 222–245.
Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.
Davidson, B., & Fasman, G. D. (1967) *Biochemistry 6*, 1616–1629.
Dill, K. A. (1985) *Biochemistry 24*, 1501–1509.
Dill, K. A. (1990) *Biochemistry 29*, 7133–7155.
Dill, K. A., Alonso, D. O. V., Hutchinson, K. A. (1989) *Biochemistry 28*, 5439–5449.
Fairman, R., Shoemaker, K. R., York, E. J., Stewart, J. M., & Baldwin, R. L. (1990) *Biophys. Chem. 37*, 107–119.
Fields, G. B., Alonso, D. O. V., Stigter, D., & Dill, K. A. (1992) *J. Phys. Chem. 96*, 3974–3981.
Flory, P. J. (1941) *J. Chem. Phys. 9*, 660–661.
Flory, P. J. (1942) *J. Chem. Phys. 10*, 51–61.
Fredenslund, A., Jones, R. L., & Prausnitz, J. M. (1975) *AIChE J. 21*, 1086–1099.
Gans, P. J., Lyu, P. C., Manning, M. C., Woody, R. W., & Kallenbach, N. R. (1991) *Biopolymers 31*, 1605–1614.
Goodman, E. M., & Kim, P. S. (1989) *Biochemistry 28*, 4343–4347.
Huggins, M. L. (1941) *J. Chem. Phys. 9*, 440.
Huggins, M. L. (1942) *J. Phys. Chem. 46*, 151–158.
Kim, P. S., & Baldwin, R. L. (1984) *Nature 307*, 329–334.
Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631–660.
King, J. (1989) *Chem. Eng. News 67*, 32–54.
Leodidis, E. B., & Hatton, T. A. (1990) *J. Phys. Chem. 94*, 6411–6420.
Lifson, S., & Roig, A. (1961) *J. Chem. Phys. 34*, 1963–1974.
Lyu, P. C., Gans, P. J., & Kallenbach, N. R. (1992) *J. Mol. Biol. 223*, 343–350.
Marqusee, S., & Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8898–8902.
Marqusee, S., Robbins, V. H., & Baldwin, R. L. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5286–5290.
McCammon, J. A., Gelin, B. R., & Karplus, M. (1977) *Nature 267*, 585–590.
Merutka, G., Shalongo, W., & Stellwagen, E. (1991) *Biochemistry 30*, 4245–4248.
Moult, J., & Unger, R. (1991) *Biochemistry 30*, 3816–3824.
Nozaki, Y., & Tanford, C. (1971) *J. Biol. Chem. 46*, 2211–2217.
Oas, T. G., & Kim, P. S. (1988) *Nature 336*, 42–48.
Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M., & Baldwin, R. L. (1990) *Nature 344*, 268–270.
Prausnitz, J. P. (1979) *Science 205*, 759–766.
Prausnitz, J. M., Lichtenthaler, R. N., & de Azevedo, E. G. (1986) *Molecular Thermodynamics of Fluid-Phase Equilibria*, p 294, Prentice-Hall, Inc., Englewood Cliffs, NJ.
Rashin, A. A. (1992) *Proteins: Struct., Funct., Genet. 13*, 120–131.
Renon, H., & Prausnitz, J. M. (1968) *AIChE J. 14*, 135–142.
Rico, M., Nieto, J. L., Santoro, J., Bermejo, F. J., Herranz, J., & Gallego, E. (1983) *FEBS Lett. 162*, 314–319.
Rooman, M. J., Kocher, J.-P. A., & Wodak, S. J. (1991) *J. Mol. Biol. 221*, 961–979.
Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg, Ser. Chim. 29*, 230–259.
Scheraga, H. A. (1978) *Pure Appl. Chem. 50*, 313–324.
Scholtz, J. M., Qian, H., York, E. J., Stewart, J. M., & Baldwin, R. L. (1991) *Biopolymers 31*, 1463–1470.
Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., Stewart, J. M., & Baldwin, R. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2349–2353.
Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987) *Nature 326*, 563–567.
Sueki, M., Lee, S., Powers, S. P., Denton, J. B., Konishi, Y., & Scheraga, H. A. (1984) *Macromolecules 17*, 148–155.
Wlodawer, L., & Sjölin, L. (1983) *Biochemistry 22*, 2720–2728.
Wojcik, J., Altmann, K.-H., & Scheraga, H. A. (1990) *Biopolymers 30*, 121–134.
Zhu, Y., Evans, L. B., & Chen, C.-C. (1990) *Biotechnol. Prog. 6*, 266–272.
Zimm, B. H., & Bragg, J. K. (1959) *J. Chem. Phys. 31*, 526–535.