Calculation of the Zimm–Bragg Cooperativity Parameter σ from a Simple Model of the Nucleation Process[1]

Robert R. Matheson, Jr.,[2] and Harold A. Scheraga*

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853.
Received November 9, 1982

ABSTRACT: A simple model for nucleation of structure by formation of hydrophobic pockets (Matheson, R. R., Jr.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 819) is applied to the problem of helix nucleation in homopoly(amino acids) in water. The nuclei for helix growth are considered to be four-residue pockets which are compatible with helical backbone geometry. The free energy of formation of such a pocket from a disordered structure depends on favorable hydrophobic interactions and an unfavorable entropy loss (and electrostatic repulsions when charged residues are involved); polar residues lack the contribution from hydrophobic interactions. A comparison of calculated and observed values for the Zimm–Bragg nucleation parameter σ indicates that the model is capable of semiquantitative accuracy. This implies that the dominant factors controlling the abilities of various types of residues to nucleate a helix are their conformational entropies and the capacity of their side chains to participate in hydrophobic bonding and possible electrostatic repulsions. In the model these factors tend to make the values of σ for charged or uncharged polar residues less than those for nonpolar ones. The results obtained provide additional support for the basic concepts of the model for nucleation of folding of polypeptides and proteins in dilute aqueous solution by formation of hydrophobic pockets.

I. Introduction

In an earlier article,[3] we proposed a simple model and computational method for predicting nucleation sites for the folding of proteins. It was based on the concept that the formation of a hydrophobic pocket constitutes the nucleation step in the folding process. The evidence cited in that earlier report was not sufficient to demonstrate the validity of the model unambiguously. Consequently, it is of interest to apply the model to the related problem of the helix–coil transition of homopoly(amino acids). Considerable experimental data are available for the parameters that characterize the helix–coil transition, so that a fairly rigorous test of the model can be made.

Detailed statistical-mechanical theories of the helix–coil transition are well-known.[4] The various theories all suppose that the internal partition function for an isolated chain molecule can be written in the form[5]

$$Z = \sum_{i,j}\prod(u_i v_j) \tag{1}$$

where u_i is the statistical weight of a sequence of i units in the coil (i.e., nonhelical) state and v_j is the corresponding quantity for a sequence of helical units of length j, with the coil and helical sequences alternating. The statistical weights u_i and v_j are assumed to depend only on the length of a sequence (i and j, respectively) and hence are of the form

$$u_i = a_i u^i \tag{2a}$$

and

$$v_j = b_j v^j \tag{2b}$$

where u and v are the corresponding statistical weights of a residue.

In the notation of Zimm and Bragg[6]

$$u_i = 1 \qquad \text{for all } i \tag{3a}$$

and

$$v_j = \sigma s^j \qquad \text{for all } j \tag{3b}$$

With these choices for statistical weights, the formulation of the helix–coil theory is essentially equivalent to the nearest-neighbor Ising model. The quantity s in eq 3b is to be interpreted as the equilibrium constant for extending an already very long sequence of helical units by one additional unit. The quantity σ is a "nucleation parameter" which enters in the statistical weight for a particular chain conformation exactly once for each helical sequence.

The particular case of a two-residue "helix" with statistical weight v_2 will be especially important in this work. We shall calculate this quantity according to the model[3] of nucleation by formation of a hydrophobic pocket and then use the relation

$$v_2 = \sigma s^2 \tag{4}$$

to compute σ and compare the results with experimental observations. In section II, the nucleation model will be reviewed, and its relevance to the statistical weight v_2 and the attendant estimation of σ will be discussed. In sections III and IV, the results for the naturally occurring amino acids will be presented and compared with available data. In section V, our calculation will be compared with other attempts to predict σ. Finally, in section VI, the conclusions that can be drawn, both about the model and about the physical processes involved in helix nucleation, will be explored.

II. Calculation of σ

The five basic assumptions of the model for nucleation of folding by formation of a hydrophobic pocket[3] are as follows: (1) the nucleation step consists of the formation of a pocket in the polypeptide chain; (2) the interactions among nonpolar groups are the dominant stabilizing forces in this pocket, and the pocket which gains the most stability from these nonpolar interactions (subject to favorable electrostatic interactions implied in assumption 3) can be identified as the nucleation site; (3) electrostatic interactions are the only ones other than hydrophobic bonds that need to be considered, and they contribute *only when charged groups are constrained to be near one another* (whether by the hydrophobic bonds or by the assumptions introduced into the model in this paper); (4) the stability and rate of formation of a nucleating pocket are independent of residues that are not part of the pocket; (5) all pockets can be considered as being formed from the "completely" unfolded polypeptide chain. Of course, this model implicitly assumes, and is applicable only to, dilute aqueous solutions. It is to be noted that, since the model is concerned only with the initial step of the folding process, it takes no cognizance of the final folded structure.

0024-9297/83/2216-1037$01.50/0

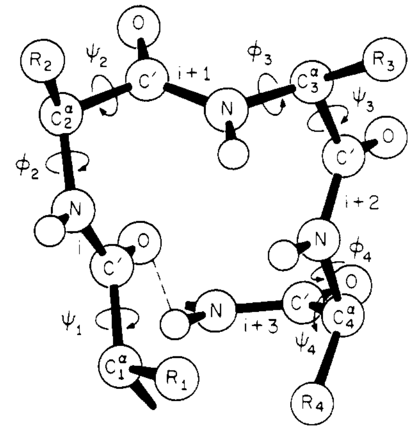

Figure 1. Schematic representation of the nucleating structure consisting of the three peptide units (i to $i + 2$) linking four α-carbon atoms (1–4). Even though ϕ and ψ of the two central residues (2 and 3) are restricted to the α-helical conformation, we postulate that they can vary sufficiently from the standard approximate α-helical values ($\phi = -57°$, $\psi = -47°$) to allow residues 1–4 to form the relevant hydrophobic bonds required by the model, viz., $R_2 \cdots R_3$ and $R_1 \cdots R_4$. No attempt is made to describe the optimal conformation of such a pocket quantitatively. Rather, we restrict ourselves to the plausible assertion that the total nonpolar interaction energy is of the same order as the sum of the mean values for the two discrete hydrophobic pairings shown here ($k = 2$). In a similar vein, our simple model assigns an approximate minimum value to the five possible neighbor–neighbor hydrophobic bonds ($l = 5$) which might potentially involve the R groups of the pocket, viz., $R_0 \cdots R_1$, $R_1 \cdots R_2$, $R_2 \cdots R_3$, $R_3 \cdots R_4$, $R_4 \cdots R_5$. *After nucleation*, the helical pocket can be stabilized, and grow, by formation of the hydrogen bond between the CO of peptide unit i and the NH of peptide unit $i + 3$; this requires that residue 4 acquire the helical conformation.

Thus, both the folding of proteins and synthetic poly(amino acids) can be analyzed with this model, even though the final state of the former is a compact globular structure and the latter is a rigid, elongated, α-helical structure.

In accordance with assumption 1, we suppose that a long polypeptide chain that is initially in a random (coil) conformation begins to fold to a helical conformation by creation of a pocket. This particular pocket is a nucleation site; i.e., a point in the chain at which a nucleus for the process of further folding exists. In accordance with the usual concept of a nucleus,[7] we expect this to be the intermediate state of folding beyond which the free energy of the folding chain becomes increasingly negative as folding progresses. For the helix–coil transition, we shall identify the structure of this nucleus from the following physical considerations. The dominant structural feature of the α-helix is the hydrogen bond between the CO group of the ith *peptide unit* and the NH group of the $(i + 3)$th peptide unit[8] (see Figure 1). It contributes a substantial amount of stability to the helix,[9,10] although other nonbonded interactions are also involved.[11,12] The incipient helix of only three consecutive peptide units (i to $i + 2$ of Figure 1) cannot form this hydrogen bond; thus, it is natural to identify this incipient helix as the nucleus for further folding, i.e., helix growth.

This incipient helix of three successive, helical, peptide units links four α-carbon atoms (see Figure 1). Thus, only two amino acid *residues* (2 and 3) are entirely in the helical conformation, since both sets of dihedral angles[13] ϕ and ψ for the two central residues of the nucleus are restricted to the helical region of conformational space; residues 1 and 4 are not in the helical conformation, but ψ_1 and ϕ_4 take on appropriate values to allow a hydrophobic bond to form between R_1 and R_4. Consequently, the statistical weight of this nucleus is v_2. On the other hand, in the language of our nucleation model it is a four-residue pocket. In this four-residue pocket, two hydrophobic bonds can form, one between residues R_1 and R_4 and the other between residues R_2 and R_3, the latter being one of intermediate strength (see below). The possibility of such a hydrophobic bond between two neighboring side chains of an α-helix is demonstrated in Table VIII of ref 12.

We digress for a moment to examine the logical consistency of identifying a four-residue pocket uniquely as a helix nucleus. In particular, we must determine whether this identification is compatible with assumption 2. Assumption 2 maintains that the first folding event (nucleation) involves the most stable hydrophobic pocket, which, in a protein, can be of any size (within limits).[3] If the most stable pocket in some polypeptide sequence were larger than four residues, then it would not be a helix because the nuclei in our model[3] are hairpin-like structures. Hence, two cases arise. In the first, viz., homopolymers of polar amino acids, hydrophobic bonds cannot form; thus, the smallest pockets (viz., four residues[3]) are the most stable because all larger ones involve the uncompensated loss of more conformational entropy. For polar residues, our postulates are thus compatible. In the second case, viz., homopolymers of the very nonpolar (type I[3]) residues, the question is moot; i.e., one cannot study the helix–coil transition for homopolymers of type I amino acids in water because such polymers are not soluble. Thus, the helix–coil transition for nonpolar polymers in water is a hypothetical process. We are, therefore, not constrained by assumption 2 and can maintain with consistency that, if we assume that no folded structures other than helices are formed (because they would then not be helices), then the nucleation of helices will proceed by formation of a four-residue pocket.

A related question is raised by this digression. Since homopolymers of type I residues are insoluble in water, the parameters for their hypothetical helix–coil transitions in water are obtained by the host–guest technique (see, e.g., ref 14), in which the properties of the homopolymer are deduced from experiments on suitable random copolymers. The significance of assumption 2, when applied to copolymers (e.g., proteins), rests on the fact that our model of nucleation includes the dominant influence of the amino acid sequence in nucleation. The question then arises as to how accurate are the *intraresidue* parameters deduced by the host–guest technique in view of this sequence dependence. In response, we cite the demonstrated dominance of short-range interactions in determining protein structure[15] and two cases[16,17] in which the parameters deduced by the host–guest technique were shown to be insensitive to changes in the host residue. We may conclude, then, that the host–guest technique provides reliable parameters for the guest residues that would be applicable to the hypothetical helix–coil transition in homopolymers of the guest residues.

In accordance with assumptions 2–4, we suppose that the relative stabilities of the four-residue pockets for various amino acid sequences are determined solely by the strengths of the hydrophobic bonds which offset the entropy loss incurred during nucleation and stabilize the pockets. Therefore, eq 1 of ref 3 is applicable. Because nonpolar residues contribute hydrophobic bonds, whereas polar ones essentially do not, it is more difficult to nucleate an α-helix among polar residues, other factors being equal. For the specific case of a four-residue pocket in a homopolymer of a nonpolar amino acid (for polar and charged amino acids, see below), there are $N = 3$ peptide bonds, with $k = 2$ hydrophobic bonds in the pocket (see Figure

1), and a possible $l = 5$ neighbor–neighbor hydrophobic bonds in the random coil. Hence, eq 1 of ref 3 can be written as

$$\Delta G_{\text{nuc}} = \Delta G^{\text{H}\phi,\text{max}} + \Delta G^{\text{H}\phi,\text{min}}\left[1 - \frac{5}{1 + \Omega \exp(\Delta G^{\text{H}\phi,\text{min}}/RT)}\right] + \left[3 - \frac{5}{1 + \Omega \exp(\Delta G^{\text{H}\phi,\text{min}}/RT)}\right] RT \ln \Omega \quad (5)$$

where $\Delta G^{\text{H}\phi,\text{max}}$ and $\Delta G^{\text{H}\phi,\text{min}}$ are the standard-state free energies of formation of maximum- and minimum-strength hydrophobic bonds,[12] respectively, between pairs of nonpolar side chains, R is the gas constant, T is the absolute temperature, and Ω is the effective number of isoenergetic positions for internal rotation about backbone bonds, expressed "per peptide bond". (See ref 3 for the concept of an "intermediate-strength" hydrophobic bond as one of average strength between those of maximum and minimum strength.) The statistical weight of the four-residue pocket is given by the expression

$$v_2 = \exp(-\Delta G_{\text{nuc}}/RT) \quad (6)$$

The equality asserted in eq 6 requires comment. Molecular models of a four-residue sequence demonstrate that, although contacts between side chains R_2 and R_3 are possible when the central two residues are constrained to the α-helical conformation, those contacts are mimimal for all but the largest side chains.[12] Slight deviations from strictly α-helical backbone geometry can markedly improve such nonpolar contacts (see Figure 1). Obviously, the value of ΔG_{nuc} computed according to eq 5 takes none of these details into account. Consequently, our calculation of σ from eq 4–6 is only approximate. For a homopolypeptide bearing charged side chains, the formation of a pocket is accompanied by a juxtaposition of groups with like charge. Thus, the expression for ΔG_{nuc} in eq 5 must be modified by appending a term, ΔG_{es} to account for this unfavorable repulsive interaction. Obviously, local details of solvation and side-chain conformation will determine the magnitude of ΔG_{es}. In the spirit of our simple scheme, we shall take the electrostatic contribution to be a constant for all charged residues and arbitrarily choose the magnitude of this interaction so that $\exp(-\Delta G_{\text{es}}/RT) \sim 0.01$.

We shall consider two approaches to estimate Ω and compare the results of both estimates. In the first, we take Ω as a constant for all residues; the reason for considering this very simple approach is discussed in section IV. In the second, we obtain Ω from conformational energy maps as

$$\Omega^{-1} = \sum_{\alpha} e^{-\epsilon_\alpha/RT} \Big/ \sum_{j} e^{-\epsilon_j/RT} \quad (7)$$

where ϵ denotes the relative energies of the various conformational energy (or free energy) minima, the sum over j extends over all such minima, and the sum over α recognizes minima which correspond to the conformation of interest (here, the α-helix). Thus, Ω of eq 7 is simply the inverse of the Boltzmann probability of occurrence of the α-helical state. Various calculations are available for the ϵ_j's. Below we consider three sets.[18,19] No matter how Ω is estimated, the quantity $R \ln \Omega$ represents the entropy change per residue which accompanies the restriction of the four-residue structure to the α-helical conformation.

Table I
Experimental Values of the Parameters σ and s in Aqueous Solution at 25 °C

residue	best value (×10⁴)	range (×10⁴)[a]	s	ref
Ala	8	10-6	1.06	20
Arg (charged)	0.1	<1[b]	1.03	21
Arg (uncharged)	NA[c]	NA	NA	
Asn	0.1	<1[b]	0.80	22
Asp (charged)	70	250-10	0.66	23
Asp (uncharged)	210	320-10	0.76	23
Cys/2	NA	NA	NA	
Cys	NA	NA	NA	
Gln	33	50-20	0.96	14
Glu (charged)	6	46-0.1	0.97	24
Glu (uncharged)	100	190-10	1.32	24
Gly	0.1	<1[b]	0.60	25
His (charged)	0.1[d]	<1[b,d]	0.68[d]	26
His (uncharged)	210[d]	224-196[d]	0.80[d]	26
Ile	55	84-34	1.12	27
Leu	33	60-12	1.14	28
Lys (charged)	1.0	<1[b]	0.94	29
Lys (uncharged)	23[e]	28-18	1.17	30
Met	54	63-45	1.17	31
Phe	18	26-10	1.07	32
Pro	NA	NA	NA	
Ser	0.1	<1[b]	0.77	33
Thr	0.1	<1[b]	0.83	34
Trp	77	133-51	1.08	35
Tyr	66	125-40	0.99	36
Val	0.1	<1[b]	0.96	37, 38

[a] The experimental uncertainty is indicated as reported in the appropriate reference, except[b] in one circumstance. [b] When no optimum value of σ could be identified, the usual report was that any $\sigma < 10^{-5}$ was acceptable. Critical examination of the data indicates that $\sigma \times 10^{-4}$ is usually more reasonable, and it is so indicated here. [c] NA = not available. [d] These are preliminary values. [e] There is a misprint in the Abstract of ref 30. The value of σ in the body of the text ($\sigma = (2.3 \pm 0.5) \times 10^{-3}$) is the one listed here.

III. Results

The essence of our model is contained in the assertion that only hydrophobic bonding, conformational entropy, and electrostatic repulsion (for charged residues) need to be considered in analyses of nucleation. Furthermore, for maximum simplicity, the initial estimate of Ω was made by assuming that $\Omega = 7$ is a constant for all residues and that, for type I residues[3] (Ala, Cys/2, Cys, Ile, Leu, Met, Phe, Pro, Trp, and Val), only hydrophobic interactions are important. For other residues, only the entropy term $[3-5/(1 + \Omega)]\ RT \ln \Omega$ (and ΔG_{es} for charged residues) contributes to ΔG_{nuc}. This simplest model (with $\Omega = 7$) implies that homopolymers of all uncharged residues except type I should have similar values for σ (likewise, that homopolymers of all charged residues should have similar values for σ) and that these must be smaller than any σ for a residue of type I. Within type I, the values of σ should follow the strength of hydrophobic bonding,[12] viz., Ile > Trp ≃ Phe > Cys ≃ Met > Pro > Val > Leu > Ala > Cys/2.

Comparison of these crude predictions with experimental data[14,20-38] (see Table I) reveals a qualitative agreement. Of the seven type I residues for which data are available, only Val has a very small value of σ. Of the other residues studied, Tyr, Lys (uncharged), Gln, Glu

Table II
Effective Values of Ω Estimated[a] for Various Computational Models[b]

residue	ϵ = ΔE(vacuum)[18]	ϵ = ΔG(vacuum)[18]	ϵ = ΔE(hydrated)[19]
Ala	14	17	8.9
Arg	6.1	9.6	3.9
Asn	15	29	7.9
Asp	19	46	12
Cys/2	14	18	5.4
Cys	NA[c]	NA	NA
Gln	6.5	6.9	3.2
Glu	7.6	7.2	3.2
Gly	21	15	10
His	12	10	5.6
Ile	15	12	8.3
Leu	7.9	10	3.9
Lys	7.9	7.7	3.4
Met	7.0	12	4.4
Phe	24	23	29
Pro	11	15	25
Ser	NA	NA	11
Thr	21	8.2	3.7
Trp	21	24	7.8
Tyr	37	39	38
Val	6.2	6.8	3.1

[a] Ω is estimated according to eq 7. [b] These values may be compared with the value[3] of $\Omega = 7$ in the first approximation. [c] NA indicates data are not available. For Ser, no α-helical minima with energy less than 5 kcal/mol above the global minimum were found.[18]

(uncharged), His (uncharged), and Asp (both charged and uncharged) have large values ($>10^{-4}$) of the nucleation parameter. We can improve the model by examining these apparent anomalies in detail. Since the assumption of a constant value for Ω is certainly a questionable one, it is an obvious potential source of anomalous results. Therefore, we consider an improved method for estimating Ω, viz., that of eq 7. In Table II, we report effective values of Ω obtained from three available sets of calculations. If other factors are constant, larger values of Ω will be associated with smaller values of σ. It appears that His, Arg, Lys, Gln, and Glu have smaller values of Ω than the other polar residues, seen most clearly in the column for ΔG(vacuum). This can account qualitatively for their large experimental values of σ (no data are available for uncharged Arg). However, this same consideration cannot account for the remaining cases of Val, Tyr, and Asp.

Val is reported to possess a very small value for σ (see Table I). Because it is distinctly hydrophobic, our model predicts a comparatively large value of σ. Moreover, the various calculations of Ω (Table II) all suggest that Val statistically favors the α-helical conformation more than any other residue. At first sight this should also tend to give larger values of σ, but more careful examination actually resolves the Val anomaly. The reason that calculations on the blocked single residue of Val indicate a relative preference for the α-helical conformation is that this permits favorable nonbonded interactions between the Val side chain and backbone.[18] Now the basic role of nonpolar side chains in our model is as a source of *interresidue* interactions in *both* the folded and unfolded forms. In the α-helical conformation for Val, the interresidue hydrophobic bonds compete against the *intraresidue* interactions as in other residues. However, in the case of the Val side chain, one sees the combination of quite weak[12] *interresidue interactions and quite strong*[18] *intraresidue* interactions whose mutual competition is not successfully approximated by our simple scheme. Apparently, the net change in hydrophobic bonding is destabilizing. (See ref 39 for a more detailed discussion of this question, involving the role of hydrophobic bonding in the folded and unfolded forms.)

The anomalously large value of σ for Tyr can be partially accounted for by acknowledging the potential for hydrophobic bonding in poly(L-tyrosine). The side chain of Tyr has a negative net free energy of transfer from water to a nonpolar environment.[40] Also, poly(L-tyrosine) is markedly insoluble in neutral aqueous solution.[41] Clearly, Tyr–Tyr hydrophobic bonding is possible and it should be reclassified as a type I residue.[42] We shall approximate it as Phe in order to preserve our simple parameterization scheme that is based on the results of Némethy and Scheraga.[12]

Asp also displays an anomalously large value of σ. However, there is no compelling reason to consider hydrophobic bonding for Asp. The values of Ω of Table II also indicate that no resolution of the Asp anomaly can be reasonably based on any unusual conformational entropy contributions including those from intraresidue hydrogen bonds (because such hydrogen bonds were already accounted for in the single-residue calculations used to obtain Ω). Even the negatively charged form of the Asp residue is reported to have a large value for σ, and our model treats any pockets of charged residues as very unlikely structures. We are left with no plausible explanation of the very large value for σ in the case of Asp.

Since our model of competition between hydrophobic bonding and conformational entropy can produce qualitatively acceptable results for all residues except Asp, let us turn to a consideration of quantitative agreement. In acknowledgment of the qualitative results just described, we shall henceforth treat Tyr as a type I residue (approximated as Phe) and Val as a type I residue analogous to Ala[3] (all interresidue hydrophobic bonds disrupted in the helical conformation).

In order to produce even qualitatively reasonable results for His, Lys, Gln, and Glu it was necessary to consider the variations of Ω among the residues. The effect of this variation is most readily seen in the polar residues since our model predicts that only Ω and electrostatic charge determine σ for such residues. The three (nonconstant) sets of effective values for Ω were summarized in Table II. Using these values and an arbitrary, but reasonable, electrostatic effect of $\exp(-\Delta G_{es}/RT) \sim 0.01$ (corresponding to $\Delta G_{es} \simeq 4.6RT$), we obtain the results given in Table III. The values for σ calculated with Ω determined from the calculations denoted ΔG(vacuum)[18] are the most nearly correct. Comparison with the experimental results in Table I indicates that the calculated values for σ are uniformly high by a factor of about 3. The relative values of σ are reproduced faily well, except for His (uncharged), Thr, and Asp. We conclude that, for these polar residues, one need consider only conformational entropy (quantitatively) and the state of charge (qualitatively) to calculate the values of σ with the proper order of magnitude and relative order.

Turning to the nonpolar residues of type I, we calculate values for σ by including the effects of hydrophobic bonding. Results for Ω held constant at 7 and for Ω given by the calculations with ΔG(vacuum) appear in Table IV. For the aliphatic residues, the latter results are generally too large by a factor of about 3–5. For the aromatic residues this is not true: the value of σ for Phe is in exact agreement with experiment, and the values for Trp and Tyr are far too small. Evidently, the approximation in which Trp and Tyr are taken[3] to be identical[12] to Phe is very bad. The relative order of the aliphatic residues

Table III
Calculated Values of $\sigma \times 10^4$ Using Various Estimates of Ω for Polar and Charged Residues

residue	constant[3]	ΔE(vacuum)[18]	ΔG(vacuum)[18]	ΔE(hydrated)[19]
Arg (charged)	0.92	0.15	0.31	6.4
Arg (uncharged)	98[a]	160[a]	33[a]	680[a]
Asn	150	11	1.1	100
Asp(charged)	2.2	0.069	0.0034	0.34
Asp (uncharged)	170	5.2	0.26	26
Gln	110	140	110	1300
Glu (charged)	1.0	0.79	0.95	13
Glu (uncharged)	56	42	51	689
Gly	270	6.1	19	78
His (charged)	2.1	0.32	0.60	4.5
His (uncharged)	150	23	44	328
Lys (charged)	1.1	0.74	0.80	11
Lys (uncharged)	72	47	52	730
Ser	160	b	b	34
Thr	140	3.2	83	1100

[a] No experimental value is available for the s appropriate for uncharged Arg. The values of σ reported for neutral Arg in this table have been computed by setting $s = 1.0$ so that $v_2 = \sigma$. [b] No value for Ω is available from the indicated computational models for Ser. See Table II.

Table IV
Calculated Values of $\sigma \times 10^4$ Using Two Estimates of Ω for Nonpolar Residues

residue	constant[3]	ΔG(vacuum)[18]
Ala	100	5.1
Cys/2	120[a]	4.7[a]
Cys	1900[a]	NA[b]
Ile	1900	320
Leu	480	150
Met	900	207
Phe	1400	21
Pro	1000[a]	80[a]
Trp	1400	20
Tyr	1700	4.6
Val	130	18

[a] No experimental value of s is available. The values of σ reported here have been computed by setting $s = 1.0$ so that $v_2 = \sigma$. [b] Data are not available; see Table II.

agrees with available experimental data, except for Val, which is still predicted to have a value for σ greater than that for Ala, in contrast to observation.

IV. Discussion

The results obtained from the simple nucleation model are in generally good agreement with experimental values of σ. There are two outstanding exceptions, viz., Asp and the nonpolar, aromatic residues Trp and Tyr. As we have discussed, the case of Asp appears to elude explanation within the constraints of our model. The two aromatic residues (and Phe) have been treated as if they were equivalent as far as the strength of their nonpolar interactions is concerned. The comparison with experiment that has just been performed reveals that this approximation is a poor one. Moreover, only Trp, Tyr, and Asp have observed values of σ larger than their computed values. A plausible source of this discrepancy is to be found in our averaging over minimum- and maximum-strength hydrophobic bonds, which may be more of an oversimplification for these residues than for the others. Other possible origins of the problems with these residues are also possible. Thus, we must omit Trp, Tyr, and Asp from the general discussion that follows.

The attractively simple calculations required for our nucleation model succeed in reproducing the relative values of σ within the rather large experimental uncertainties. They also succeed in giving the correct order of magnitude, 10^{-3}–10^{-4}, of the available nucleation parameters. Thus, it appears that the model of competition between hydrophobic bonding and conformational entropy (with electrostatic repulsions for charged residues) is a reasonable one. Moreover, the formal identification of the statistical weight v_2 with the equilibrium constant for creation of a four-residue pocket is justified.

The calculations of σ in which Ω is held constant (at 7) have been included in Tables III and IV. This approximation is the one that was employed in the initial formulation[3] of this nucleation model. It was used with the intention that such a simple assumption would suffice to yield accurate relative measures of nucleation tendencies, even though the absolute results were acknowledged to be unreliable. We now find that this assumption works only modestly well. In light of the modifications developed here, it is probably worthwhile to recompute the "best" nucleating regions found in our previous evaluation of protein folding,[3] using better values of Ω. Such a study is contemplated.

Finally, we summarize the status of the parameterization scheme of our model. Best results are obtained when Ω is evaluated for each residue by considering the Boltzmann probability of the α-helical state using free energies for the conformational statistical weights. Tyr should be treated as a type I residue. The Val–Val pair is not an effective source of net pocket stabilization, because of the relatively small size of the Val side chain. However, this behavior of Val seems to be specific to the α-helical geometry and can probably be ignored in treating pockets in proteins. Trp, Tyr, and Asp are poorly treated in the current model. An ad hoc fit to the helix–coil nucleation data can be obtained by arbitrarily adjusting Ω for these residues. However, the suitability of an ad hoc choice of Ω for nucleation processes in proteins is unknown, and more work is needed to clarify the proper treatment of Trp, Tyr, and both charged and uncharged forms of aspartic acid.

V. Relationship of This Model to Other Interpretations of σ

There are two principal interpretations of the physical significance of the parameter σ. Applequist[43] observed that σ may be described as the equilibrium constant for the creation of one additional, long, helical sequence in a chain by a process which maintains a constant number of backbone hydrogen bonds. The analogy to an interfacial free energy is obvious. This interpretation has been very useful and has served as the conceptual basis for at least three attempts to calculate σ. Brant[9] used conformational energy calculations to estimate the energetic contributions to σ for poly(L-alanine) in the absence of solvent. Gō et al.[10,11,44] did the same for polyglycine, poly(L-alanine), and poly(L-valine) in the presence and absence of water. Finkelstein et al.[45] used model building to estimate the statistical weights for residues near the end of a helical sequence and thereby estimate σ. The reader is referred to the original references for details. Our model represents a quite different approach. It is based on the Zimm–Bragg description of σ, viz., that σ is a factor which accounts for the extra difficulty associated with beginning as opposed to extending a helical sequence. It is certainly noteworthy that, by focusing so completely on only the approximate statistical weight for the incipient helix, we can obtain the generally good agreement between prediction and experiment summarized in Tables II–IV. We may therefore conclude that the process of nucleation by formation of a pocket (or by formation of a bend, since the two are identical for pockets of four residues) is a realistic description of the actual microscopic process.

VI. Conclusions

The results that are reported here lend credence to the validity of the model of nucleation by formation of a hydrophobic pocket. The modifications of the original model that were made in section III do not bring into question any of the fundamental assumptions of the model. This is an important test case for the model, which was originally introduced to explain nucleation in the far more complex process of the folding of globular proteins. The current results, however, do not speak to the validity of assumption 5 or of the manner in which the stability of pockets larger than four residues was computed.[3]

Our results, based on values of Ω computed with ΔG(vacuum), are generally too large by a factor of about 3–5. Reference to Table II shows that other estimates of Ω could lower this discrepancy, but by too much for the nonpolar residues. Our calculations of σ are most sensitive to the parameter Ω, and it seems likely that most of the uncertainty and inconsistency with observations can be traced to this source. Although various types of conformational energy calculations give a consistent picture of the overall conformation of a residue, they lead to appreciable variation in the quantity Ω.

As pointed out above, the apparent success of our model suggests that it may simulate the physical processes that accompany helix nucleation. That is, the incipient helix is organized so that nonpolar contacts between the side chains can be maintained at the same time that the backbone dihedral angles at least approximate those which characterize the α-helix. In the absence of hydrogen bonds, these nonpolar contacts are the source of the only significant favorable interactions in the incipient helix (except for Trp, Tyr, and Asp). In homopolymers where the side chains cannot provide favorable hydrophobic bonds, nucleation may be relatively unlikely. This is because considerable entropy must be sacrificed before the first stabilizing hydrogen bonds can be formed. In some cases, e.g., uncharged Glu, the entropy loss (as embodied in the value of Ω) is not as unfavorable as in other cases, thereby increasing the value of σ. On the other hand, homopolymers composed of residues that can form strong hydrophobic bonds (those of type I) can more readily nucleate helices because the nonpolar interactions provide free energy of stabilization that decreases the net free energy deficit involved in forming an incipient helix. The conformational entropy and hydrophobic bonds or electrostatic repulsions between the side chains dominate over other considerations in determining σ. This accounts for our reasonable success in describing the general trend of the available data.

Finally, for completeness, consider the helix–coil transition of homopoly(amino acids) in nonaqueous solvents. Clearly, there is no "hydrophobic bonding" in such a system. We do not anticipate that a change in solvent will dramatically influence the short-range interactions which contribute to nucleation.[11] A nonaqueous (presumably less polar) solvent, however, will (a) be a better solvent for nonpolar side chains, (b) not quench the unfavorable interactions between aligned peptide dipoles in incipient and short helical sequences, and (c) increase the strengths of interresidue hydrogen bonds. Effects a and b will serve to make nucleation more difficult since they will destabilize the incipient helix. Effect c will tend to stabilize the longer helical sequences once they are formed. Thus, we expect the helix–coil transition to be characterized by generally smaller values of σ for polypeptides in nonpolar solvents than for the same polymers in aqueous solution. (See ref 46, where it is shown that the value of σ for Ala is larger than its value in water, and that for Leu remains unchanged, within experimental error, between water and a nonaqueous solvent.) Extending this expectation to the creation of helical stretches in proteins, we anticipate that nucleation of helical stretches will take place predominantly in the early stages of folding when most residues are accessible to the solvent (water). At this stage, the driving force of nonpolar association will aid nucleation more than in later stages of folding when solvent has been excluded from most of the volume pervaded by the folding chain. The presence of water, however, will disfavor helix growth (by competition for backbone hydrogen bonds), and growth of the nuclei into helices of reasonable length will occur more easily in later stages of folding.

Acknowledgment. We are indebted to J. B. Denton, R. A. Fredrickson, C. A. McWherter, J. A. Nagy, and G. Némethy for helpful comments on the manuscript.

References and Notes

(1) This work was supported by research grants from the National Institute of Arthritis and Metabolic Diseases of the National Institutes of Health, U.S. Public Health Service (AM-08465), and from the National Science Foundation (PCM79-20279).

(2) NIH Predoctoral Trainee, 1974–1978. Present address: Central Research and Development Department, Experimental Station 356-105 E. I. du Pont de Nemours and Co., Wilmington, DE 19898.

(3) Matheson, R. R., Jr.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 819.

(4) See, for example: Poland, D.; Scheraga, H. A. "Theory of Helix-Coil Transitions in Biopolymers"; Academic Press: New York, 1970.

(5) Gō, N.; Gō, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *59*, 1030.

(6) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526.

(7) Mandelkern, L. "Crystallization of Polymers"; McGraw-Hill: New York, 1964; Chapter 8.

(8) Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 205.

(9) Brant, D. A. *Macromolecules* **1968**, *1*, 291.

(10) Gō, M.; Gō, N.; Scheraga, H. A. *J. Chem. Phys.* **1970**, *52*, 2060.

(11) Gō, M.; Gō, N.; Scheraga, H. A. *J. Chem. Phys.* **1971**, *54*, 4489.

(12) Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1962**, *66*, 1773.

(13) IUPAC-IUB Commission on Biochemical Nomenclature *Biochemistry* **1970**, *9*, 3471.

(14) Denton, J. B.; Powers, S. P.; Zweifel, B. O.; Scheraga, H. A. *Biopolymers* **1982**, *21*, 51.

(15) Scheraga, H. A. *Pure Appl. Chem.* **1973**, *36*, 1.

(16) von Dreele, P. H.; Lotan, N.; Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 408.

(17) Kidera, A.; Mochizuki, M.; Hasegawa, R.; Hayashi, T.; Sato, H.; Nakajima, A.; Fredrickson, R. A.; Powers, S. P.; Lee, S.; Scheraga, H. A. *Macromolecules* **1983**, *16*, 162.

(18) Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1.

(19) Hodes, Z. I.; Némethy, G.; Scheraga, H. A. *Biopolymers* **1979**, *18*, 1565.

(20) Platzer, K. E. B.; Ananthanarayanan, V. S.; Andreatta, R. H.; Scheraga, H. A. *Macromolecules* **1972,** *5,* 177.
(21) Konishi, Y.; van Nispen, J. W.; Davenport, G.; Scheraga, H. A. *Macromolecules* **1977,** *10,* 1264.
(22) Matheson, R. R., Jr.; Nemenoff, R. A.; Cardinaux, F.; Scheraga, H. A. *Biopolymers* **1977,** *16,* 1567.
(23) Kobayashi, Y.; Cardinaux, F.; Zweifel, B. O.; Scheraga, H. A. *Macromolecules* **1977,** *10,* 1271.
(24) Maxfield, F. R.; Alter, J. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1975,** *8,* 479.
(25) Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971,** *4,* 417.
(26) Sueki, M.; Konishi, Y.; Scheraga, H. A. *Macromolecules*, to be submitted.
(27) Fredrickson, R. A.; Chang, M. C.; Powers, S. P.; Scheraga, H. A. *Macromolecules* **1981,** *14,* 625.
(28) Alter, J. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1972,** *5,* 739.
(29) Dygert, M. K.; Taylor, G. T.; Cardinaux, F.; Scheraga, H. A. *Macromolecules* **1976,** *9,* 794.
(30) Barskaya, T. V.; Ptitsyn, O. B. *Biopolymers* **1971,** *10,* 2181.
(31) Hill, D. J. T.; Cardinaux, F.; Scheraga, H. A. *Biopolymers* **1977,** *16,* 2447.
(32) Van Wart, H. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1973,** *6,* 266.
(33) Van Nispen, J. W.; Hill, D. J.; Scheraga, H. A. *Biopolymers* **1977,** *16,* 1587.
(34) Hecht, M. H.; Zweifel, B. O.; Scheraga, H. A. *Macromolecules* **1978,** *11,* 545.
(35) Nagy, J. A.; Powers, S. P.; Zweifel, B. O.; Scheraga, H. A. *Macromolecules* **1980,** *13,* 1428.
(36) Scheule, R. K.; Cardinaux, F.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1976,** *9,* 23.
(37) Alter, J. E.; Andreatta, R. H.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1973,** *6,* 564.
(38) Chang, M. C.; Fredrickson, R. A.; Powers, S. P.; Scheraga, H. A. *Macromolecules* **1981,** *14,* 633.
(39) Gō, M.; Scheraga, H. A. *Biopolymers*, to be submitted.
(40) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1971,** *246,* 2211.
(41) Fasman, G. D.; Bodenheimer, E.; Lindblow, C. *Biochemistry* **1964,** *3,* 1665.
(42) This reclassification does not affect any of the results of our previous paper[3] because there were no Tyr–Tyr interactions in the proteins considered there. However, this reclassification will be taken into account in future calculations on proteins.
(43) Applequist, J. *J. Chem. Phys.* **1963,** *38,* 934.
(44) Gō, M.; Hesselink, F. T.; Gō, N.; Scheraga, H. A. *Macromolecules* **1974,** *7,* 459.
(45) Finkelstein, A. V.; Ptitsyn, O. B.; Kozitsyn, S. A. *Biopolymers* **1977,** *16,* 497.
(46) Sridhara, S.; Ananthanarayanan, V. S.; Fredrickson, R. A.; Zweifel, B. O.; Taylor, G. T.; Scheraga, H. A. *Biopolymers* **1981,** *20,* 1435.

Computed Conformational States of the 20 Naturally Occurring Amino Acid Residues and of the Prototype Residue α-Aminobutyric Acid[1]

Max Vásquez,[2] George Némethy, and Harold A. Scheraga*

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853.

Received December 1, 1982

ABSTRACT: Conformational energy calculations were carried out on the *N*-acetyl-*N'*-methylamides of the 20 naturally occurring amino acids and of α-aminobutyric acid (a useful prototype for conformational studies of backbone–side chain interactions in polypeptides). Use was made of an improved version, ECEPP/2, of the computer program "empirical conformational energy program for peptides", in which some of the parameters describing interatomic energies and residue geometry have been updated. As a result of these changes, better agreement was found between the computed side-chain conformational distributions and those observed in crystalline peptides and proteins and in solutions of oligopeptides, but otherwise there is little change in the conformational properties of the various residues that were reported earlier (Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1977,** *10,* 1). The Boltzmann probabilities of occurrence of various backbone and side-chain conformational states for all residues are tabulated for easy reference, and they are used to analyze trends in conformational preferences that depend on the nature of the residues.

I. Introduction

The conformational behavior of amino acid residues is of fundamental importance for the theoretical conformational analysis of oligo- and polypeptides and proteins. The properties of individual residues constitute information that can be used as the basis of the conformational studies of larger peptides, because short-range effects play an important role in the determination of the preferred conformations of polypeptides and proteins.[3-5] In this laboratory, conformational energy computations have been carried out with a set of empirical potentials that have been incorporated into a computer program named ECEPP (empirical conformational energy program for peptides).[6,7] Soon after the development of ECEPP in 1975, it was used to analyze the low-energy structures of the *N*-acetyl-*N'*-methylamides of the 20 naturally occurring amino acids,[8] denoted as "terminally blocked single residues".

The results of that study were in general agreement with experimental data and with structural information on peptides available at that time. Recently, however, a discrepancy has been found[9] between the computed distributions[8] and observed distribution in crystalline oligopeptides[9] and in proteins[10] for the dihedral angle χ^1 (describing rotation about the C^α–C^β side-chain bond) in unbranched side chains. The observed data indicate that the g^- conformational state is favored over the other two conformational states (g^+ and t), while the earlier computations indicated a preference for the t state. There was good agreement between observations and computations for the other side-chain dihedral angles and for χ^1 of side chains with β-branching. An analysis of the interatomic interactions for the computed conformations indicated that the discrepancy occurred because the repulsive interactions involving the backbone peptide N atom were too high.[9] The basis for the assignment[6] of the nonbonded interaction energy for this atom was reconsidered, and a modified set of parameters was introduced into a recent revision of ECEPP, named ECEPP/2.[11] In addition, the need arose for minor updating or modification of some other parameters in ECEPP, as a result of new experimental or theoretical

0024-9297/83/2216-1043$01.50/0