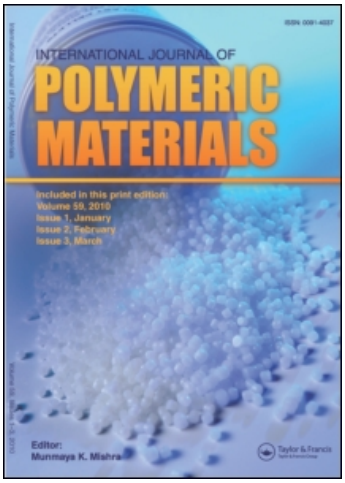

International Journal of Polymeric Materials
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713647664

Helix-Coil Transitions in Specific Sequence Biopolymers

Douglas Poland[a]
[a] Department of Chemistry, The Johns Hopkins University, Baltimore, Maryland, U.S.A.

To cite this Article Poland, Douglas(1975) 'Helix-Coil Transitions in Specific Sequence Biopolymers', International Journal of Polymeric Materials, 4: 1, 1 — 23
To link to this Article: DOI: 10.1080/00914037508072345
URL: http://dx.doi.org/10.1080/00914037508072345

Intern. J. Polymeric Mater., 1975, Vol. 4, pp. 1–23

Printed in Reading, England

Helix-Coil Transitions in Specific Sequence Biopolymers†

DOUGLAS POLAND

Department of Chemistry, The Johns Hopkins University, Baltimore Maryland 21218, U.S.A.

(*Received August 20, 1973*)

The theory of helix-coil transitions in specific sequence biological macromolecules is reviewed. The importance of the cooperative nature of the transition (resulting in the tendency for long helical sequences to exist in the transition region) is stressed in that heterogeneity of polymer composition (hence heterogeneity of helix stability) tends to more or less localize long sequences of helix producing a pronounced profile of helical probability. The appropriate thermodynamic parameters and the resulting nature of the transitions for DNA, collagen, and proteins are compared. Some new calculations are given for carboxypeptidase A that includes heterogeneity in both the helix-coil parameters σ and s.

While there has been a great deal of effort expended on the attempt to calculate the conformation of a protein from a knowledge of the amino acid sequence,[1] it is now generally recognized that this is an impossibly complex task (due to the problem of local minima in the free energy) without some preliminary knowledge of conformational preferences in the molecule. Of course, it would be cheating to use, even partially, knowledge supplied by X-ray studies. Furthermore, it would defeat the purpose of such an endeavor where the goal is to understand, in terms of the concepts of physical chemistry, why a particular protein has a particular conformation, and, would also defeat any hope of being able to determine the conformation of proteins which are not amenable to treatment by the X-ray technique.

There is a way to obtain preliminary knowledge about preferred conformations in a given protein. And, perhaps somewhat surprisingly, this knowledge does not come from the study of the conformations of small molecules and the science of intermolecular forces, but rather comes from the

† Presented at the Midland Macromolecular Meeting on "Order in Polymer Solutions", August 20–24, 1973.

study of the statistical mechanics of appropriate copolymers. How information about conformation is obtained from statistical mechanical analyses is the main subject of this review.

In Section 1 we briefly review the now familiar σ and s parameters of the Zimm-Bragg treatment of the helix-coil transition in homopolyamino acids with particular reference as to what physical significance one can (and cannot) assign to these parameters. Two techniques (with and without partition functions) for calculating helix probability profiles for heteropolymers are given in Section 2 with brief discussion of the application of these techniques to the double helix of DNA and the triple helix of collagen. We make a short detour in our review of techniques for treating specific sequence molecules in Section 3 to discuss a technique for treating oligomer-polymer binding that considerably relaxes many of the assumptions necessary in the standard models for treating helix-coil transitions. The specific problem of obtaining statistical mechanical parameters for use in proteins is discussed in Section 4 while the technique of extracting data from random copolymers is treated in Section 5. Brief mention is made of the importance of electrostatic interactions in Section 6. Finally, the techniques and parameters now available are used to calculate preliminary information about conformational preferences in carboxypeptidase A.

1 NATURE OF σ AND s

Though we will not review in detail here the many papers on the theory of helix-coil transitions in homopolymers,[2] there are a few results of the basic theory that are essential to the understanding of specific sequence molecules.

To start our very brief overview of important concepts, we first recall that the basic problem is to construct a model for the multiple equilibrium that arises when each residue or unit in a polymer can exist in either of two distinct conformations, helix or coil (h or c, respectively). When several residues in a row are in the h state, they lead to an ordered, helical sequence; residues in c states do not give rise to any ordered structure, but rather constitute a more or less random collection of conformations, the so-called random coil (note that the c state is not actually a single state but rather a collection of all conformations that do not give rise to helix). With the dichotomy h or c for each chain unit, there are 2^N possible microstates for a molecule of N units; for example, one such microstate is shown below.

$$\ldots c\,c\,h\,h\,h\,h\,c\,c\,c\,c\,c\,h\,h\,h\,c\,c\,c\ldots$$

To construct a statistical mechanical model for the multiple equilibrium between all 2^N microstates, one is faced with two problems; first, one must

formulate the free energy, G, for each microstate; second, one must sum over all the terms $\exp(-G/\boldsymbol{RT})$.

The first problem is approached by assuming that the free energy of a given microstate can be subdivided into a linear combination of appropriate free energies for helix and coil sequences. By taking the free energy of a unit relative to the coil state, one need only formulate a general expression for the free energy of a helical sequence of an arbitrary number (n) of units. This is really the heart of the standard treatment of helix-coil transitions. The form used is very simple; one assumes that as helical sequences become large one has

$$\begin{pmatrix}\text{free energy of a helical}\\ \text{sequence of } n \text{ units}\end{pmatrix} = \begin{pmatrix}\text{correction for}\\ \text{end effects}\end{pmatrix} + \begin{pmatrix}\text{number of helical units}\\ \text{in the sequence}\end{pmatrix} \times \begin{pmatrix}\text{free energy per unit}\\ \text{characteristic of } h\\ \text{states in a long sequence}\end{pmatrix}$$

or

$$G(n) \approx nG_{\text{unit}} + G_{\text{ends}} \tag{1}$$

What specifically is meant by large is very important and will be discussed shortly. The term for end effects simply reflects the fact that h states near the ends of a sequence of h states are in a different environment than h states surrounded on both sides by several h states.

The statistical weight, $\exp(-G(n)/\boldsymbol{RT})$, for a sequence of n h states then is given by

$$\exp(-G(n)/\boldsymbol{RT}) = \sigma s^{n} \tag{2}$$

where equation (2) defines the quantities σ and s in terms of the free energies introduced in Eq. (1).

To keep our discussion brief, we anticipate some results of the model:

$s > 1$	helix favored
$s = 1$	helix and coil equally favored
$s < 1$	coil favoured
$\sigma \ll 1$	long helical sequences favored

Experimentally, σ is found to be small for both nucleic acids (see Section 2) and polyamino acids (see Section 7) reflecting the fact that the end effects referred to in Eq. (1) represent an unfavorable free energy of units at the ends of helical sequences. The cooperative influence of σ is best illustrated by considering a sample equilibrium

$$\underset{\sigma s^3}{} \quad \underset{\sigma s^4}{}$$
$$..cchhhccccchhhhcc..$$

$$\underset{\text{maximize combinatorial entropy}}{\overset{\text{minimize end effects}}{\rightleftharpoons}} \quad \underset{\sigma s^7}{}..ccccchhhhhhhccccc.. \tag{3}$$

The reaction from left to right reduces the number of end effects (σ factors) and is more favored the smaller σ is while the reaction from right to left is favored by the fact that there are more ways (combinatorial entropy) to place many small sequences on the molecule than a few large ones: the equilibrium between small and long sequences thus is a balance between minimizing end effects (favoring long sequences) and maximizing the combinatorial entropy (favoring short sequences).

We now introduce a very important notion that is essential for an understanding of specific sequence molecules, namely that of the average helical sequence length

$$<L> = \text{average helical sequence length} \tag{4}$$

Actually the probability distribution of helical sequence lengths is quite broad. It can be shown[2] that almost all helical states exist in sequences having lengths in the range

$$<L> \pm <L>/2 \tag{5}$$

From Eq. (3) one easily sees that the smaller σ the larger $<L>$; a very useful result of the model is that at $s = 1$

$$<L> \approx \sigma^{-1/2} \tag{6}$$

Since σ is typically in the range of 10^{-4} for many systems, $<L>$ is of the order of one hundred h states. Thus specifically what one means by large in Eq. (1) is values of n big enough to encompass all the sequence lengths that are covered by Eq. (5); $<L>/10$ (≈ 10 for $\sigma = 10^{-4}$) is a generous lower limit above which one need not worry about the applicability of Eq. (1).

With the above considerations, we now make the approximation that we can use Eq. (2) for all values of n starting from unity even though we do not claim Eq. (2) is necessarily accurate for n values below about $n = 10$. And this is because h states in helical sequences of about 1 to 10 units make practically no contribution to the total population of h states; thus what statistical weights are assigned to these sequences is really almost immaterial (unless, of course, they are statistical weights that greatly favor these sequences, which Eq. (2) do not). The above argument has two very important results that may at first seem contradictory. The approximation of using Eq. (2) for

all sequence lengths makes the model mathematically equivalent to a nearest-neighbor correlation problem. In fact, the use of the mathematics of nearest-neighbor correlations does not imply that physically either σ or s reflect simply nearest-neighbor interactions; indeed, our above argument requires only that intermolecular potentials damp out to the extent that they can be truncated at the range of about 10 units distant from a given unit which is generous even for interactions between ionic sidechains. Equation (2) is all that is required to assign statistical weights to all 2^N possible microstates (combinations of helix and coil states). The problem of summing over all 2^N statistical weights is most elegantly accomplished by the matrix technique first applied to the problem by Zimm and Bragg.[3] We simply quote the result that the partition function is given by the matrix product

$$Z = \mathbf{e}\,\mathbf{W}^N\,\mathbf{e}^+ = (1\ 1)\begin{pmatrix} s & s \\ \sigma & 1 \end{pmatrix}^N \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{7}$$

Eq. (7) defining the matrix $\mathbf{W}$ and vectors $\mathbf{e}$ and $\mathbf{e}^+$. The matrix used here gives the statistical weight of a given unit in terms of the states of the following unit, the factor σ being assigned to a . . *ch* . . border for notational clarity in subsequent equations (a factor σ is assigned to each helical sequence and it is arbitrary whether it is assigned via the . . *ch* . . or . . *hc* . . border). How the results of the model quoted in this section are obtained from Eq. (7) is treated elsewhere.[2] The average fraction of *h* states is given by the standard relation

$$\theta = \frac{1}{N}\frac{\partial(\ln Z)}{\partial(\ln s)} = \theta(\sigma,s) \tag{8}$$

Since θ is a quantity that is experimentally measurable, Eq. (8) implies that this information can be translated via the statistical mechanical formulation of Z in Eq. (7) into information about σ and s. While both σ and s can be expected to be functions of temperature, the fitting of experimental data to the model of Eq. (2) is found to be sensitive only to the temperature variation of s which is usually found to be adequately represented by

$$s(T) = \exp\left(-\Delta H(T-T_{tr})/RTT_{tr}\right) \tag{9}$$

where ΔH is the average enthalpy difference between helix and coil units and T_{tr} is the temperature at which $s = 1$ (where $\theta = 1/2$ for long chains). The model thus contains three empirical parameters: σ, ΔH, and T_{tr}.

2 HELIX-COIL TRANSITIONS IN SPECIFIC SEQUENCE MOLECULES

In the last section, we reviewed some of the important concepts used to describe the transition between helix and coil in homopolymers. If the molecule

is composed of two or more different units, these units being in a specific sequence, then there may be regions in the molecule that have a greater tendency to form helix than others simply because a region is rich in a unit that is a strong helix former. In order to describe specific sequence heteropolymers, one can make a simple extension of Eq. (7). Suppose one has a chain containing three different kinds of units, *a*, *b*, and *c*, which might have the sequence

$$b\,a\,a\,c\,b\,c\,b\,b\,a \tag{10}$$

As with homopolymers, each unit is considered to be able to exist in *h* or *c* states, but now with *s* and σ factors appropriate to each particular kind of unit. One defines matrices appropriate to each kind of unit

$$\mathbf{W}_a = \begin{pmatrix} s_a & s_a \\ \sigma_a & 1 \end{pmatrix}, \qquad \mathbf{W}_b = \begin{pmatrix} s_b & s_b \\ \sigma_b & 1 \end{pmatrix}, \qquad \mathbf{W}_c = \begin{pmatrix} s_c & s_c \\ \sigma_c & 1 \end{pmatrix} \tag{11}$$

Then the partition function for the sequence illustrated in expression (10) is given by the specific matric product

$$Z = \mathbf{e}\,\mathbf{W}_b\mathbf{W}_a\mathbf{W}_a\mathbf{W}_c\mathbf{W}_b\mathbf{W}_c\mathbf{W}_b\mathbf{W}_b\mathbf{W}_a\,\mathbf{e}^+ \tag{12}$$

The probability that a particular unit, *e.g.* the fourth from the left, is in the helical state is given by

$$P_4 = \mathbf{e}\,\mathbf{W}_b\mathbf{W}_a\mathbf{W}_a\mathbf{W}_c'\mathbf{W}_b\mathbf{W}_c\mathbf{W}_b\mathbf{W}_b\mathbf{W}_a\,\mathbf{e}^+/Z \tag{13}$$

where

$$\mathbf{W}_c' = \frac{\partial\,\mathbf{W}_c}{\partial(\ln s_c)} \tag{14}$$

While the derivative required in Eq. (8) can be evaluated analytically (by expressing the matrix product of Eq. (7) in terms of the eigenvalues and eigenvectors of **W**) the expression in Eq. (13) must be evaluated by explicit matrix multiplication, an easy task for computer.

In analogy with Eq. (13), the probability that any given unit in the chain is in the helical state can be calculated thus giving a profile of the probability of helix as a function of the position (hence type) of a unit in the molecule; we will refer to this as the helix probability profile.

How does heterogeneity and specific sequence influence the helix-coil transition? Obviously this depends on how different the respective parameters s_a, s_b, etc., are. We will first discuss the results for parameters appropriate to nucleic acids. In DNA the double helix may be thought of as composed of a specific sequence of two kinds of units, A–T and G–C hydrogen bonded base pairs, the G–C pair being the stronger helix former contributing three hydrogen bonds instead of the two contributed by the A–T pair. Experiments on appropriate synthetic polynucleotides indicate that at the transition temperature for a heteropolymer containing equal quantities of A–T and G–C

base pairs $s_{A-T} \approx 1/2$ and $s_{G-C} \approx 2$ (making the average $s \approx 1$) with $\sigma \approx 10^{-4}$ (the same for A–T and G–C). In DNA there is an additional feature aside from Eq. (2), namely a long-range statistical weight required to describe the entropy of coil states when they exist in loops formed by disrupting the double helix in the interior. We will leave this complication out of our discussion here referring the reader to the literature;[2,4,5] the essential features of the effect of heterogeneity are not altered by this omission.

Using the above parameters characteristic of DNA, let us consider the following model calculation. We generate a specific sequence of say 1000 units long composed of two kinds of units, taking for simplicity $s_a = 1/2$, $s_b = 2$, $\sigma = 10^{-4}$, by picking the identity of each unit as *a* or *b* at random giving for a long chain 50% *a* units and 50% *b* units. The helix probability profile is then calculated numerically according to Eq. (13); since the average s $[= (s_a\, s_b)^{1/2}]$ is unity, we expect that the average fraction of helix in the whole molecule will be close to 1/2 and this is found to be so. Of great significance, for the nucleic acid-like parameters used, it is found that there is no correlation between whether a unit has a high helical probability and the fact that it is a strong helix former (here, unit *b*). Rather it is found[5] that there are long alternating regions in the molecule that are either approximately 100% helix or 100% coil; the average dimension of these regions is found to be given by Eq. (5) (a result for homopolymers) giving [from Eq. (6)] that the lengths of these regions fall in the range 100 ± 50 units. Thus, although the average fraction of helix is 50%, there are distinct helix and coil regions in the molecule, *i.e.* the helical regions are distinctly localized in the molecule. Since the average extent of the helical regions is approximately $\sigma^{-1/2} \approx 100$ and the sequence was generated by randomly assigning the designation *a* or *b* to each unit (each equally likely), a region containing 100 units obviously contains approximately equal numbers of strong and weak helix formers. What then causes the localization of distinct helical regions? If one looks at the average composition of the chain say in blocks of 10 units, then one begins to find a very good correlation between the location of helical regions and regions of the molecule where over a block of units there is a slight excess of strong helix formers.

From the above discussion, it is clear that the localization of helical regions has to be understood in statistical terms: first the concept from homopolymers of an average helical sequence length determined by σ [through the statistical mechanical description of the multiple equilibrium indicated in Eq. (3)] is a statistical mechanical one (there is no physical interaction extending 100 units to "hold" the sequence together); and second, the regions where the helical sequences are located in the molecule are determined by regions in the molecule that have a statistical excess (fluctuation) of strong helix formers averaged over many tens of units. Thus, two statistical concepts are required

to understand how distinct conformational preferences arise in a heteropolymer.

Collagen is a specific sequence molecule forming a triple-stranded helix that is receiving increased attention. Schwarz and Poland[6] have recently analyzed the helix-coil transition in the triple-stranded complex of the synthetic collagen analog $(Gly\text{-}Pro\text{-}Pro)_n$ for $n = 10, 15$, and 20. Natural collagen is known to have the general formula[7]

$$\ldots (Gly\text{-}X\text{-}Y)\ (Gly\text{-}X\text{-}Y) \ldots$$

where X can be proline or one of the amino acids and Y can be hydroxyproline or one of the amino acids. The helical unit per chain is a triplet of residues, Gly-X-Y; and since collagen is a triple helix, the total helical unit in the triple helix contains nine residues. Such a nonet of residues can contain 0, 3, or 6 imino acids (the sequence Gly-X-Y is the same at corresponding positions in each of the three chains in the triple helix). Thus the collagen triple helix can be described in first approximation by assigning each nonet of residues a factor s_0, s_3, or s_6 depending on the imino acid content. From the analysis of the melting behavior of $(Gly\text{-}Pro\text{-}Pro)_n$, we have evaluated s_6 as a function of temperature; using knowledge of s_6 together with experimental data on the melting temperatures of natural collagens as a function of imino acid composition, we have estimated s_0 and s_3. We find[6] at 25°C

$$s_0 = 0.48$$
$$s_3 = 2.17$$
$$s_6 = 10.80$$

and at 50°C

$$s_0 = 0.42$$
$$s_3 = 1.29$$
$$s_6 = 4.12$$

When compared with the nucleic acid parameters, $s_{A\text{-}T} \approx 1/2$ and $s_{G\text{-}C} \approx 2$, one would expect an even stronger influence of specific sequence for collagen due to the larger differences in the s's. For simple unwinding from the ends, model calculations[6] show that indeed helical regions are very sharply localized; more general model calculations allowing internal unwinding (via loops) are in progress. Note that the proline rich nonet (represented by s_6) is the strongest helix former; presumably, this is due to favorable hydrophobic interactions of the proline when in the triple helix.

Before ending this section, we want to indicate an alternate method of calculating helix probability profiles. Simha and Lacombe[8,9] have shown that average quantities for the (pseudo) nearest-neighbor helix-coil model can be calculated without using partition functions. Here, we just indicate the general method for infinite homopolymers referring the reader to the

original papers for details on treating specific sequence molecules. We begin by considering the two sequences of states

$$\begin{array}{ll} \text{(I)} & \ldots h\, c\, h \ldots \\ \text{(II)} & \ldots h\, h\, h \ldots \end{array}$$

From the basic nature of the partition function, one has the result that the ratios of the partition functions for the chains with the appropriate sequences indicated fixed is the ratio of the respective *a priori* probabilities of the given sequences. Thus

$$\frac{p(\ldots hch\ldots)}{p(\ldots hhh\ldots)} = \frac{p(h)\,P(h \mid c)\,P(c \mid h)}{p(h)\,P(h \mid h)\,P(h \mid h)}$$

$$= \frac{P(h \mid c)\;P(c \mid h)}{P(h \mid h)\;P(h \mid h)} = Z_{\mathrm{I}}/Z_{\mathrm{II}} = \sigma/s \tag{15}$$

where $p(\ldots hch\ldots)$ and $p(h)$, etc., refer to the *a priori* probabilities of the indicated sequence and $P\,(h/c)$ are conditional probabilities, *e.g.* that given h, c follows. The essence of the usefulness of Eq. (15) is that the partition functions Z_{I} and Z_{II} need not be evaluated, the ratio in the nearest-neighbor model being a simple combination of s and σ. In a similar fashion, consideration of the sequences

$$\begin{array}{ll} \text{(I)} & \ldots c\, h\, c \ldots \\ \text{(II)} & \ldots c\, c\, c \ldots \end{array}$$

yields

$$\frac{P(c \mid h)\;P(h \mid c)}{P(c \mid c)\;P(c \mid c)} = \sigma\, s \tag{16}$$

Eqs. (15) and (16) together with the relations

$$\begin{aligned} P(c \mid h) + P(c \mid c) &= 1 \\ P(h \mid c) + P(h \mid h) &= 1 \end{aligned} \tag{17}$$

give four equations in four unknowns and hence all the conditional probabilities can be determined. To calculate the *a priori* probability a unit is helix, one observes that for the sequences

$$\begin{array}{ll} \text{(I)} & \ldots c\, h \ldots \\ \text{(II)} & \ldots h\, h \ldots \end{array}$$

one obtains

$$\begin{aligned} p\,(h) &= p\,(\ldots c\, h \ldots) + p\,(\ldots h\, h \ldots) \\ &= p\,(c)\,P\,(c \mid h) + p\,(h)\,P\,(h \mid h) \end{aligned} \tag{18}$$

which together with the fact that the conditional probabilities have already been determined and the relation

$$p(c) + p(h) = 1 \tag{19}$$

yields $p(h)$.

The above technique has been applied to specific sequence heteropolymers[9] and can also be applied to the models for DNA and collagen requiring long-range correlations to treat interior loops with no approximations in the basic model.[10]

3 TECHNIQUE OF TREATING OLIGOMER-POLYMER BINDING

We make a short pause in our discussion of specific sequence molecules to discuss helix-coil transitions where helicity is a result of the binding of a small molecule to a large one. For this system one need not assume the specific properties of helix statistical weights given in Eqs. (2) and (8); thus, the validity of these particular functional forms can be tested.

The specific system we will discuss is the binding of oligomers of inosine, I_n (n in the range of 6 to 11), to polycytidylate.[11] Double helix is formed between I and C when an oligomer binds to the polymer. At low temperature the polyC will be saturated with I_n and hence be completely helical. If one heats this system, helix is lost via the dissociation of oligomers. Since this is an association equilibrium the melting temperature of the complex is dependent on the total oligomer concentration. Thus, one can also induce a helix-coil transition at constant temperature by changing the oligomer concentration, in short a titration curve.

To indicate how the helix-coil transition is treated using binding isotherms, consider the simple reaction

$$\mathrm{A} + \mathrm{B} \overset{K}{\rightleftharpoons} \mathrm{A{\cdot}B} \tag{20}$$

The equilibrium constant expression for the reaction of Eq. (20) is

$$K = \frac{[\mathrm{A{\cdot}B}]}{[\mathrm{A}]\,[\mathrm{B}]} \tag{21}$$

Eq. (21) can be solved for the fraction of A·B using the relation $f_{\mathrm{A}} + f_{\mathrm{A{\cdot}B}} = 1$

$$\text{Binding Isotherm} \qquad f_{\mathrm{A{\cdot}B}} = \frac{[\mathrm{B}]\,K}{1 + [\mathrm{B}]\,K} \tag{22}$$

Using the conservation relation for B

$$\text{Conservation of B} \qquad [\mathrm{B}] + f_{\mathrm{A{\cdot}B}}\,[\mathrm{A}]_0 = [\mathrm{B}]_0 \tag{23}$$

$[A]_0$ and $[B]_0$ being the total concentration of A and B in any form, respectively, then Eqs. (22) and (23) represent two equations in the two unknowns [B] and K if $f_{A\cdot B}$ (analog of fraction of helix) is measured experimentally.

In order to treat oligomer–polymer binding, one has a binding constant K plus another parameter that reflects the fact that an oligomer that binds to the polymer and interacts with the end of another oligomer that is already bound is in a different environment than an oligomer that binds in a free region of the polymer (this parameter is the analog of σ reflecting border effects). One thus has three unknown parameters: the binding constant; the oligomer–oligomer interaction parameters; and, the concentration of free oligomer. One also has analogs of Eqs. (22) and (23), requiring one more equation to determine all the unknowns. A relation for the slope of the adsorption isotherm as oligomer concentration is varied allows all the unknowns to be determined in terms of the two experimental quantities, the analogs of $f_{A\cdot B}$ and $df_{A\cdot B}/d[B]$ for the oligomer–polymer system. Of course, the actual forms of Eqs. (22) and (23) for oligomer–polymer binding are more complicated than that shown for the simple example of Eq. (20). However, even though an $(n + 1) \times (n + 1)$ matrix is required to formulate the partition function (analog of the quantity $1 + [B]\,K$), explicit relations for the oligomer–polymer binding constant and the oligomer–oligomer interaction constant are obtained[11] (*i.e.*, one can circumvent any need to extract eigenvalues of moderate sized matrices).

The constants can be determined at several temperatures simply by studying the adsorption isotherm at several temperatures. Then the enthalpy of binding is determined in the standard fashion

$$\Delta H = -\boldsymbol{R}\, d\,(\ln K)/d\,(1/T)$$

with no need to assume that ΔH is independent of temperature. In this fashion one can study K as a function of oligomer size and determine whether or not Eq. (2) indeed is valid. For the oligoinosinate-polycytidylate system, we find Eqs. (2) and (8) are, in fact, very good approximations.

4 APPLICATION OF HELIX-COIL THEORY TO PROTEINS

Obviously, the models we have been treating are not capable of describing the formation of globular structures that involve interactions between units far removed in the amino acid sequence. The goal of the approach outlined here is less ambitious, namely, simply to determine if there are regions in the primary sequence that have stronger helical forming propensities than others thus giving preliminary information about conformational preferences in the

chain. In order to do this, one needs σ and s factors for all of the naturally occurring amino acids. One obvious way to get this information is to study homopolymers of each of the amino acids using the technique of Section 1. However, except for polyproline which does not form the α helix, none of the homopolymers of the natural amino acids is soluble in water. There are two alternate approaches to get the required parameters, via sandwich compounds and random copolymers.

Sandwich compounds are simply a block of the desired amino acid bordered by blocks of a solubilizing agent such as blocks of D,L-lysine.[12,13,14] Unfortunately, few of the natural amino acids exhibit transitions in the range 0–100°C (either they are too stable or too unstable). Both alanine[12,13] and leucine[14] have been studied in this fashion; it is found, for example, that

$$\begin{array}{ll} 0°C & 80°C \\ s_{ala} = 1.08 & s_{ala} = 0.99 \\ s_{leu} = 1.28 & s_{leu} = 1.33 \end{array} \qquad (24)$$

Notice that helix in polyleucine becomes more stable as the temperature is increased (due presumably to hydrophobic interactions) while helix melts out (though the transition is very broad) in long chain of polyalanine at about 78°C.

The technique of studying random copolymers seems to be applicable to all the amino acids and so far has yielded parameters for glycine,[15] alanine,[16] leucine,[17] serine,[18] and phenylalanine.[19] The essence of the technique is to incorporate into synthetic water soluble polyamino acids [poly(hydroxypropyl-L-glutamine), PHPG, and poly(hydroxybutyl-L-glutamine), PHBG, are used] a certain small percentage (*e.g.*, 1–5%) of a natural amino acid as a copolymer unit. The introduction of the natural amino acid produces a perturbation on the helix-coil transition of the water soluble homopolymer and by suitable analysis yields σ and s for the natural amino acid. The analysis requires a theory for the helix-coil transition in a random copolymer, a subject that has occupied much attention in the literature which we now turn to. Once again, our search for information about conformation leads us to statistical mechanics.

5 HELIX-COIL TRANSITIONS IN RANDOM COPOLYMERS

We have already discussed how the helix-coil transition is treated for specific sequence molecules in Section 2. The additional problem presented by random copolymers is that a solution of such a copolymer represents a collection of many sequences, fixed once the synthesis is completed but randomly determined. To treat such a system, one must average over all the sequences

occurring, producing an average partition function from partition functions such as that given in Eq. (12).

The problem of calculating the proper average partition function for the nearest-neighbor model has been solved exactly.[20,21] Other approximate treatments[2] are available and require much less computer time (since one searches for the values of σ and s that give the best fit with experiment, a great many computations are required). It turns out that for a copolymer incorporating a small percentage of a natural amino acid the simplest possible approximation is very accurate; a hierarchy of approximations[22] that gives in highest order the exact solution and in lowest the approximation just mentioned is available and can be used to test in a very simple manner what degree of approximation is adequate for a given system.

The simple approximation mentioned above formulates the partition function as a matrix product, the appropriate matrix being (for a copolymer of two units)

$$\mathbf{W}=\begin{pmatrix}(p_a s_a + p_b s_b) & (p_a s_a + p_b s_b)\\ (p_a \sigma_a + p_b \sigma_b) & (p_a + p_b)\end{pmatrix} \tag{25}$$

where p_a and p_b are the *a priori* probabilities of occurrence of the two units ($p_a + p_b = 1$). The matrix of Eq. (25) actually allows each unit to be helix or coil and *a* or *b*; clearly this is incorrect since a unit cannot switch from *a* to *b* once the molecule is synthesized (while it can, of course, switch from helix to coil). Nonetheless, if σ is small (order of 10^{-4}) and s_a is not very different from s_b (for example, 0.7 and 1.0) then the matrix of Eq. (25) in Eqs. (7) and (8) is found to yield results essentially identical with the exact formulation.

The subject of helix-coil transitions in random copolymers is a very complex one involving particularly the proper average partition function to be used and how to evaluate it. While Eq. (25) is adequate for copolymers of amino acids, caution should be used in applying it to other systems without checking to see if less approximate treatments give the same results.[22]

6 ELECTROSTATIC INTERACTIONS

The protein myoglobin contains 153 residues and the three-dimensional structure determined by Kendrew shows the molecule to be about 80% helical. There are four main coil regions, the amino acid sequences of which are given below (the numbers representing position in the chain numbering from N terminus to C terminus)

I	43–50	-Phe-Asp-Arg-Phe-Lys-His-Leu-Lys-
II	78–85	-Lys-Lys-Gly-His-His-Glu-Ala-Glu-
III	119–124	-His-Pro-Gly-Asn-Phe-Gly-
IV	149–153	-Leu-Gly-Tyr-Glu-Gly

Coil region IV is at the end of the chain (ends have a natural tendency to be predominantly coil in homopolymers) and contains two glycine residues. As will be discussed in the next section, Gly, Pro, Asn, and Ser are very weak helix formers and the high concentration of these residues in region III makes this naturally a weak helix, strong coil region. However, there is no large concentration of weak helix formers in coil regions I and II (there being only one Gly in region II). These regions do contain large concentrations of basic, hence, positively charged, residues (Lys, Arg, His). Since the helical conformation puts sidechains closer together than they are in the more extended coil conformation, simple electrostatic repulsion would seem to be a possible explanation of why these regions are coil.

Preliminary calculations[23] that introduce electrostatic interactions between the specific sequence of positive and negative charges dictated by the primary sequence as well as heterogeneity in σ and s indicate that such electrostatic interactions as discussed above do indeed have a significant influence on the helix probability profile, influencing both strong (*e.g.* through attractive Lys-Glu interactions) and weak helix regions.

There are few studies on model systems that guide one as to how to incorporate electrostatic interactions in specific sequence molecules. Zimm and Rice[24] studied the helix-coil transition in polyglutamic acid induced by titration of the charged groups in a dioxane–water solvent. The electrostatic interactions were treated by summing over charge–charge pair interactions using a Debye-Hückel potential, truncating the extent of interaction at four residues distant. They obtained good agreement with experiment using no adjustable parameters (σ and s taken from polybenzylglutamate). Riem, *et al.*,[25] studied the titration of helical poly-(L-lysine) in 95% methanol using the same technique as Zimm and Rice. The polylysine system in methanol has the advantage that there is no helix-coil transition over the whole titration; thus, the whole behavior of the titration is governed by electrostatic interactions. The extent of the range of interaction was extended to seven residues distant. It was found that truncation of the extent of interaction at four residues, as in the treatment of Zimm and Rice, gave excellent agreement with experiment, while extending the range of interaction further gave progressively worse agreement. Thus, one can conclude that the treatment of electrostatic interactions by summing over pair interactions using a Debye-Hückel potential gives good agreement with experiment if the extent of interaction is truncated at four residues (the truncation is compensated by the fact that the Debye-Hückel potential underestimates the shielding at experimental electrolyte concentrations).

The whole topic of electrostatic interactions between charged sidechains deserves much more attention, both experimental and theoretical. These interactions are of major importance in determining conformation in specific sequence molecules.

A new technique of formulating the partition function for the titration of rigid macromolecules is available.[26]

7 HELIX PROBABILITY PROFILES FOR PROTEINS

In this section we will apply the concepts we have discussed for obtaining appropriate parameters and use them to calculate helix probability profiles. The example we will use is carboxypeptidase A whose tertiary structure has been determined from X-ray studies.[27] This protein has 307 residues and the crystal structure shows that the molecule is about 38% helix: specifically, residues 14–28, 72–88, 94–103, 112–122, 173–187, 215–231, 254–262, and 283–306 are helical. First, we will discuss previous work on calculating profiles and the parameters that we will use.

Helix probability profiles have been calculated for several proteins.[1,28,29] These calculations were based on grouping all the natural occurring amino acids into three categories with the following parameters:

$$\begin{aligned} s_1 &= 0.385 && \text{(helix breaker)} \\ s_2 &= 1.00 && \text{(helix indifferent)} \\ s_3 &= 1.05 && \text{(helix former)} \\ \sigma &= 5 \times 10^{-4} && \text{(for all)} \end{aligned} \tag{26}$$

Class 1 contains Pro, Ser, Gly, and Asn, class 2 contains Lys, Tyr, Asp, Thy, Arg, Cys, Phe, while class 3 contains all the rest. The helix probability profile for sperm whale myoglobin calculated by the author with the parameters of expression (26) is shown in Figure 1. The smooth curve shows the

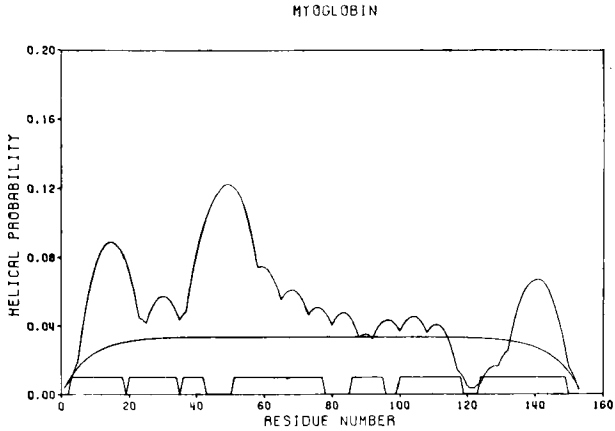

FIGURE 1 Helix probability profile for sperm whale myoglobin using the parameters of expression (26).

profile using an average value of s while the blocks at the bottom indicate the position of α helix in the amino acid sequence. The distinct cusps in the profile for the heteropolymer occur where amino acids of class 1 occur reflecting the low value of s_1 assigned to these residues.

Experimental data obtained to date by the techniques outlined in Section 4 has shown that the parameters assigned by expression (26) are not terribly accurate. In particular, expression (26) assigns an extremely low value of s to class 1 (containing four amino acids); in fact, all the structure (maxima and minima) in the curve shown in Figure 1 is determined by class 1. Helix-coil parameters are now known for the following amino acids[15–19] (given here for 60°C)

$$\begin{array}{llll} 1. & s_{\mathrm{Gly}} = 0.63 & \sigma_{\mathrm{Gly}} = 1.0 \times 10^{-5} & \\ 2. & s_{\mathrm{Ser}} = 0.74 & \sigma_{\mathrm{Ser}} = 7.5 \times 10^{-5} & \\ 3. & s_{\mathrm{PHPG}} = 0.96 & \sigma_{\mathrm{PHPG}} = 2.2 \times 10^{-4} & \\ 4. & s_{\mathrm{Ala}} = 1.01 & \sigma_{\mathrm{Ala}} = 8.0 \times 10^{-4} & (27) \\ 5. & s_{\mathrm{Phe}} = 1.00 & \sigma_{\mathrm{Phe}} = 1.8 \times 10^{-3} & \\ 6. & s_{\mathrm{Leu}} = 1.09 & \sigma_{\mathrm{Leu}} = 3.3 \times 10^{-3} & \end{array}$$

(PHPG is the synthetic homopolymer poly(hydroxypropylglutamine) mentioned in Section 4.) Lacking data on the rest of the amino acids, we must for the present group the remaining amino acids as to their resemblance to the residues listed in expression (27). We have done this as follows:

1. (like Gly) Pro
2. (like Ser) Asn
3. (like PHPG) Glu, Gln, Lys, Arg, Thr, Cys, Tyr, His, Trp, Met
4. Ala
5. Phe
6. (like Leu) Ileu, Val (28)

Obviously, this grouping is very crude and should be improved greatly in the near future by further theoretical and experimental studies. Notice that s_{Leu} determined from the sandwich compound technique given in expression (24) is significantly different from the value determined from the random copolymer technique given in expression (27). Presumably, this reflects the fact that more or less isolated Leu residues in the random copolymer have less stabilizing influence than when Leu is neighbored by other Leu residues in the sandwich compound. To reflect these facts, we introduce another class with s characteristic of Leu in the sandwich compound

$$\left(\begin{array}{l}\mathrm{Val}\\ \mathrm{Leu}\\ \mathrm{Ileu}\end{array}\right) \text{ neighboring } \left(\begin{array}{l}\mathrm{Val}\\ \mathrm{Leu}\\ \mathrm{Ileu}\end{array}\right), \quad s \sim 1.3 \qquad (29)$$

Using the parameters of expressions (28) and (29), one obtains the helix probability profile for sperm whale myoglobin shown in Figure 2, again

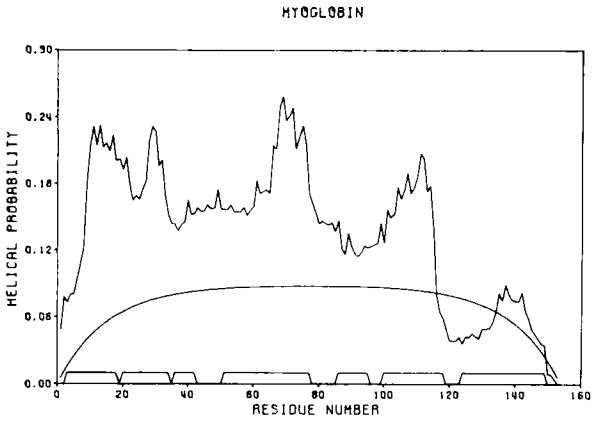

FIGURE 2 Helix probability profile for sperm whale myoglobin using the parameters of expressions (27) and (28).

showing the profile using an average σ and s for comparison. Clearly there are large differences in helix forming tendencies, these correlating moderately well with the regions found to be helical in the crystal structure. Preliminary calculations[23] introducing electrostatic interactions (see Section 6) between charged sidechains give even better correlation. We now examine in turn the influence of heterogeneous σ and s on helix probability profiles using the example of carboxypeptidase A mentioned at the beginning of this section.

All the remaining figures we will discuss are helix probability profiles for carboxypeptidase A, the blocks at the bottom of the graphs indicating the regions that are found to be helical in the crystal structure with the smooth curve showing the helix probability profile using an average value of σ and s for all residues; the parameters used are those of expressions (28) and (29). Figure 3 shows the helix probability profile when an average value of s is used for all residues but σ is heterogeneous; it is seen that heterogeneity of σ alone does not introduce major differences from the curve for σ and s both homogeneous, the effect being mainly to introduce a spiked fine structure. In Figure 4, σ is kept homogeneous while s is allowed to be heterogeneous. Heterogeneity of s is seen to introduce major minima and maxima in the profile. Finally, in Figure 5 both σ and s are allowed to be heterogeneous. Note that when σ alone was heterogeneous (Figure 3) not much effect was noted; however, when heterogeneity in σ is introduced on top of heterogeneity in s (Figure 4), the maxima and minima are much more pronounced. The reason for this

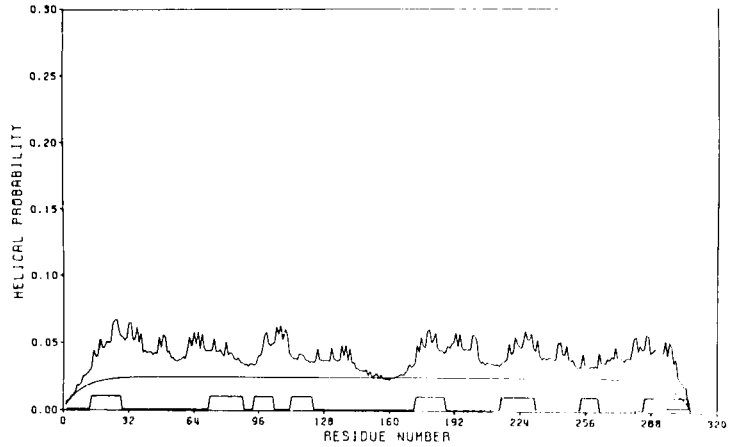

FIGURE 3 Helix probability profile for carboxypeptidase A using the parameters of expressions (27) and (28); s is kept homogeneous, σ is heterogeneous.

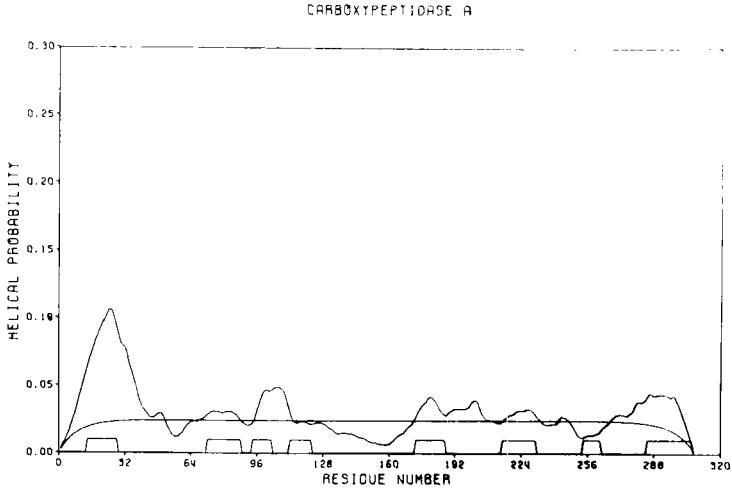

FIGURE 4 Helix probability profile for carboxypeptidase A using the parameters of expressions (27) and (28); σ is kept homogeneous, s is heterogeneous.

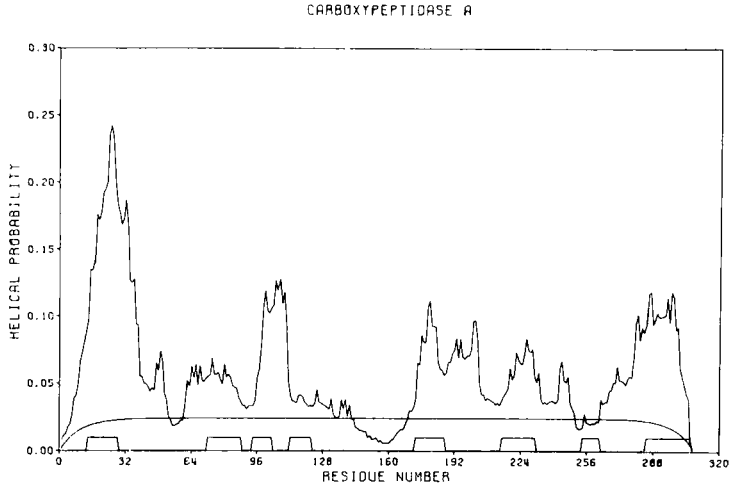

FIGURE 5 Helix probability profile for carboxypeptidase A using the parameters of expressions (27) and (28); both σ and s are heterogeneous.

is that there is a big spread in σ values in expression (27); specifically, $(\sigma^{-1/2})_{\text{Gly}} = 310$ and $(\sigma^{-1/2})_{\text{Leu}} = 18$. Recalling that the average helical sequence length is of the order of $\sigma^{-1/2}$ [Eq. (6)], heterogeneity in σ allows short regions of helix to exist with high probability.

While work has been done[30] on calculating σ and s from intermolecular potentials, there is of necessity a large empirical element in such calculations with regard as to the introduction of the influence of the solvent (water). This is illustrated by the fact that breaking an amide hydrogen bond in the gas phase requires about 5 Kcal/mol, while the ΔH per mole of residues of the helix-to-coil transition for polyamino acids[15–19] is found to be in the range of only a few tenths of a Kcal (this ΔH is, of course, the difference in enthalpy between the amide hydrogen bond and water hydrogen bonded to the amide group).

Clearly, there is much to be done in understanding conformational preferences in primary sequences. But even the beginnings indicated here give great promise that helical propensities (and perhaps also β-forming propensities) can be predicted with some accuracy, thus giving a firm starting ground for the fascinating problem of predicting and understanding protein tertiary structure from knowledge of the amino acid sequence.

Acknowledgement

This work was supported in part by an Alfred P. Sloan Research Fellowship.

References

1. H. A. Scheraga, *Conformation of Biological Molecules and Polymers*, The Jerusalem Symposia on Quantum Chemistry and Biochemistry, V, The Israel Academy of Science and Humanities, Jerusalem, p 51 (1973).
2. For a detailed review see: D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic Press, New York, 1970).
3. B. H. Zimm and J. K. Bragg, *J. Chem. Phys.* **31**, 526 (1959).
4. D. M. Crothers, *Biopolymers* **6**, 1391 (1968).
5. D. Poland and H. A. Scheraga, *Physiol. Chem. Phys.* **1**, 389 (1969).
6. M. Schwarz, Jr. and D. Poland, *Biopolymers*, **13**, 687 (1974).
7. W. Traub and K. A. Piez, *Advan. Protein Chem.* **25**, 243 (1971).
8. R. Simha and R. H. Lacombe, *J. Chem. Phys.* **55**, 2936 (1971).
9. R. H. Lacombe and R. Simha, *J. Chem. Phys.* **58**, 1043 (1973).
10. D. Poland, *Biopolymers*, **13**, 1859 (1974).
11. M. W. Springgate and D. Poland, *Biopolymers*, **12**, 2241 (1973).
12. W. B. Gratzer and P. Doty, *J. Amer. Chem. Soc.* **85**, 1193 (1963).
13. R. T. Ingwall, H. A. Scheraga, N. Lotan, A. Berger, and E. Katchalski, *Biopolymers* **6**, 331 (1968).
14. S. E. Ostroy, N. Lotan, R. T. Ingwall, and H. A. Scheraga, *Biopolymers* **9**, 749 (1970).
15. V. S. Ananthanarayanan, R. H. Andreatta, D. Poland, and H. A. Scheraga, *Macromolecules* **4**, 417 (1971).
16. K. E. B. Platzer, V. S. Ananthanarayanan, R. H. Andreatta, and H. A. Scheraga, *ibid.* **5**, 177 (1972).
17. J. F. Alter, G. T. Taylor, and H. A. Scheraga, *ibid.* **5**, 739 (1972).
18. L. J. Hughes, R. H. Andreatta, and H. A. Scheraga, *ibid.* **5**, 187 (1972).
19. H. E. Van Wart, G. T. Taylor, and H. A. Scheraga, *ibid.* **6**, 266 (1973).
20. G. W. Lehman and J. P. McTague, *J. Chem. Phys.* **49**, 3170 (1968).
21. S. S. Cohen and O. Penrose, *J. Chem. Phys.* **52**, 5018 (1970).
22. D. Poland and H. A. Scheraga, *Biopolymers* **7**, 887 (1969).
23. D. Poland, to be published.
24. B. H. Zimm and S. A. Rice, *Mol. Phys.* **3**, 391 (1960).
25. R. K. H. Liem, D. Poland, and H. A. Scheraga, *J. Amer. Chem. Soc.* **92**, 5717 (1970).
26. D. Poland, *J. Chem. Phys.* **58**, 3574 (1973).
27. F. A. Quiocho and W. N. Lipscomb, *Advan. Protein Chem.* **25**, 1 (1971).
28. N. Go, P. N. Lewis, M. Go, and H. A. Scheraga, *Macromolecules* **4**, 692 (1971).
29. P. N. Lewis and H. A. Scheraga, *Arch. Biochem. Biophys.* **144**, 576 (1971).
30. M. Go, N. Go, and H. A. Scheraga, *J. Chem. Phys.* **54**, 4489 (1971).

DISCUSSION

Prof. P. L. Luisi (*Swiss Federal Institute of Technology, Zürich*): It seems to me that the treatment by Zimm and Bragg (as all other statistico-mechanical treatments for macromolecules) can be applied only under equilibrium conditions among all the chain conformers. The best one can do with it is to get information on the average conformational properties of the chain when the polymer is in the random coil state (the helix being only one of the possible chain conformers in equilibrium). If this is so, one may have difficulties in understanding the meaning of studying in terms of s and σ processes such as the denaturation of biopolymers (often an irreversible process), or of

analyzing in those terms the rigid conformation of proteins in the native state. Could you clarify this point to me?

Prof. D. Poland: The theory of helix-coil transitions does apply to the reversible equilibrium between conformations, and when the theory is used to extract parameters from model systems or to compare theoretical with experimental melting curves, this reversibility must be demonstrated. With specific regard to proteins, while the native to denatured transition is rapidly reversible in many (but not all) proteins, our purpose is not to describe the single most probable conformation but rather the less ambitious task of determining whether or not there are regions in the random coil that retain high helical propensities in the absence of long range stabilizing influences present in the globular structure. Thus the calculations presented are only a beginning at understanding protein conformation; certainly more than σ and s are required for the final answer to understanding the low-temperature structure.

Prof. P. L. Luisi: In your analysis of the conformation of myoglobin and other proteins in terms of s and σ, long range interactions are necessarily absent. Does this reflect only the necessity of an early stage approximation, or does it reflect also the philosophy that only short range interactions are essential for the chain conformation (the long range interactions being a result of the near neighbor interactions, etc. . . .)?

Prof. D. Poland: Both. Conformational calculations indicate that side chain–backbone and near neighbor side chain–side chain interactions alone produce strong and varied conformational preferences for the different amino acids. A great deal of evidence seems to point to the fact that the local interactions point the direction strongly to the final conformation.

Prof. A. M. Jamieson (*Case Western Reserve University, Cleveland, Ohio*): I would like to clarify one point in my mind. Your results show that for a homopolymer of a 50% helix-former in the transition region, the macromolecule is in a state of high dynamic flux with helical segments diffusing back and forth rapidly along the chain, but for a random copolymer of equal amounts of a helix-former and coil-former, the macromolecule is in a dynamically rigid situation with regions of strong helix and strong coil, dictated by a small statistical excess of the appropriate species. Is this correct?

Prof. D. Poland: Remember that these are equilibrium or time averaged results. One cannot say that the helical regions in a specific sequence molecule

are dynamically rigid, only that they prefer to spend most of their time in a certain region of the molecule.

Prof. W. B. Rippon (*Case Western Reserve University, Cleveland, Ohio*): Your DNA data showed a line graph at the bottom with base distribution; this seemed to correlate with the calculated helicity. How many residues was the data in the line graph averaged over, or did it represent individual bases?

Prof. D. Poland: The composition was the average for blocks of 50 base pairs. And this is the interesting point: there is no correlation with individual bases but rather with the coarse grained composition.

Prof. W. B. Rippon: The complete sequence of collagen is essentially available now. In addition, there are at least two papers out where the generally accepted quarter staggering packing is rationalized on the basis of charged and hydrophobic interactions for side-to-side alignment. In addition, the position of residues is now correlated with electron microscope staining patterns. The sequence indicates regions lacking proline and it would be interesting to see if these are large enough to alter your conclusions.

Prof. D. Poland: Even in a random sequence there are lots of triplets that do not contain proline, this being reflected in our parameter s_0. More important is the fact that the beginning of the known sequences contains sequences that do obey the formula Gly-X-Y. Since these are at the ends (which are frayed anyway) they do not have much influence.

Prof. W. B. Rippon: Native collagens are crosslinked at the terminal ends—how would this affect your use of this data to determine s_0 and s_3 for proline?

Prof. D. Poland: Very little since the data we use from natural collagens is simply the melting temperature as a function of proline content; closing the molecule at one end does not change the melting temperature by more than a few degrees.

Dr. R. F. Boyer (*Dow Chemical Company, Midland, Michigan*): It is my understanding that a multi-strand system such as collagen will reform its triple strand with exact end register when cooled down in aqueous medium from the denaturing temperature. Such behavior would seem to imply the existence of long range order over the length of the polymer chains, rather than a series of random sequences of amino acids. Can you clarify this situation?

Prof. D. Poland: DNA finally rewinds in exact register, this long range ordering being due to local base pairing requirements. Even a randomly generated, though specific, model sequence will give the same result. Our calculations for collagen were based on model sequences with the amino acid sequence in each of the three chains identical. Thus there is a great regularity locally that tends to bring the chains into register; experimentally the chains are found to initially come together not in perfect register, a slow creep mechanism finally bringing them into register.

Prof. A. J. Hopfinger (*Case Western Reserve University, Cleveland, Ohio*): Can your model for the collagen triple helix distinguish between $(\text{Gly-X-Pro})_n$ and $(\text{Gly-Pro-X})_n$ sequences?

Prof. D. Poland: No, one cannot make this distinction without model conformational calculations. Thus we determine a mean s for (Gly-Pro-X) and (Gly-X-Pro); since the two sequences occur with almost exactly equal probability in known sequences this does not seem to present a problem.

Prof. A. J. Hopfinger: How were the thermal melt equations of the triple-helix tripeptides determined? What was the solvent?

Prof. D. Poland: The solvent for all our collagen calculations was water. The $(\text{Gly-Pro-Pro})_n$ melting curves were calculated according to a simple statistical mechanical model for mismatching of three chains with unwinding from the ends. The other collagen parameters were determined simply by fitting the known melting temperatures of natural collagens as a function of imino acid content with a ΔH and ΔS that were proportional to imino acid content.

Prof. A. J. Hopfinger: In the determination of σ and s for peptide residues, is the critical assumption that the σ and s determined from the block copolymers and/or homopolypeptides will be the same as those of a residue in a hetero-sequenced polypeptide chain?

Prof. D. Poland: Yes, although one has a check by seeing if the parameters determined from the copolymer and sandwich compound techniques agree.