Construction and Analysis of *Lifson–Roig* Models for the Helical Conformations of α-Peptides

by **Daniel S. Kemp***

Department of Chemistry, Room 18-296, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Dedicated to Professor *Dieter Seebach* on the occasion of his 65th birthday

New forms of *Lifson–Roig* algorithms are introduced for modeling stabilities of helices formed by polypeptides derived from α-amino acids. The principles of constructing and generalizing these algorithms are developed, and their application to modeling of circular dichroism ellipticities and protection factors derived from H/D exchange of α-peptide backbone amide residues are critically reviewed. With the aim of comparing the properties of structured conformations formed by α- and β-peptides, the intrinsic limitations of *Lifson–Roig* algorithms and their underlying assumptions are analyzed. *Lifson–Roig* state sums that provide easy structural analysis are generated by new algorithms based on products of 8×8 and 16×16 matrices, and a simple protocol is introduced for generating new state sums that are tailored for specific purposes. The *N*- and *C*-capping of highly helical α-peptides by means of tailored helix-stabilizing templates is shown to result in helical conformational manifolds that approach those of helical β-peptides in conformational homogeneity.

1. Introduction. – In polar solvents, an unaggregated, helically-disposed polypeptide derived from α-amino acids must be modeled as an equilibrated manifold of fully helical, partially helical, and nonhelical conformations. As recently shown primarily by *Seebach* and co-workers [1], the conformations of helical analogs formed by β-amino acids are fully structured and highly stable. This contrast is particularly surprising in that in both cases the manifolds of unstructured conformations that serve as energetic references are believed to be strongly stabilized by entropic effects, and the stabilization of structured conformations is attributable to familiar, presumably well-understood enthalpic effects associated with H-bonding and hydrophobic interactions. This is a humbling result for those who profess a chemical understanding of the causes of the energetic properties of α-peptides. Clearly, these causes must be rethought.

As a modest step toward that goal, this report examines the underlying assumptions and generalizes the form of the *Lifson–Roig* (LR) bookkeeping models [2] currently used to assign stabilities from experimental data to helices formed by α-peptides. Why are bookkeeping models important for this class of molecules? The three cases shown in *Fig. 1* address this question by comparing basic features of conformational analysis that are employed under equilibrium conditions for simple organic molecules as well as for complex biopolymers.

Conformational modeling of a simple molecule studied under particular experimental conditions begins with a list generated by analogy or from prior experiment of the energetically accessible conformations. Spectroscopic data are then analyzed to yield mole fractions of the relevant conformational states. For the axial and equatorial methylcyclohexane of *Fig. 1*,*a*, the bookkeeping model is trivial, since only two states are present.

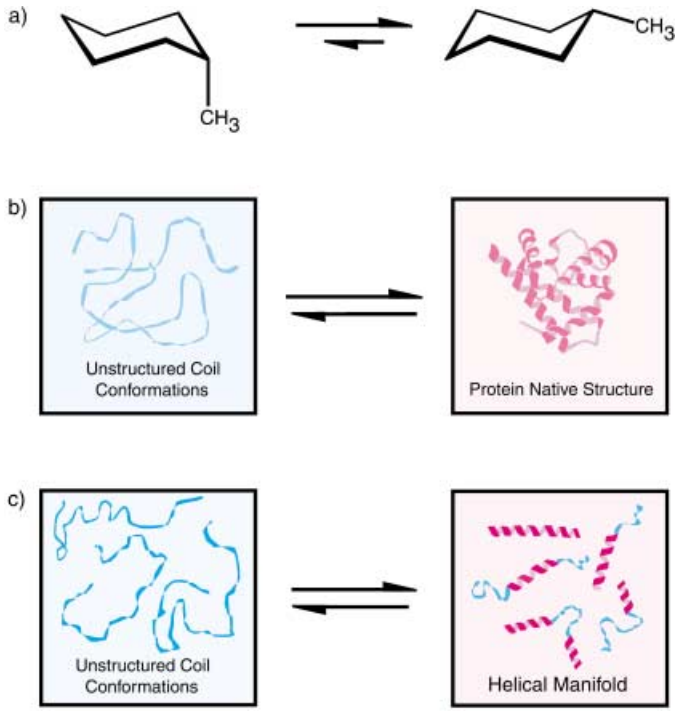

Fig. 1. *Relative energetics of conformations of molecules and their conformational manifolds. a*) A simple example of an equilibrium between two well-defined conformations of the same molecule. *b*) Equilibrium for the unfolding of a globular protein. Although a native protein (right) can often be modeled as a single conformation to a good approximation, the unstructured state (left) is a conformational manifold containing many disordered conformations. Owing to its conformational homogeneity, denaturation of a globular protein can often be modeled as a two-state process described by a simple equilibrium expression. *c*) The unfolding of a typical unaggregated α-peptide in a polar protic solvent. Both structured and unstructured states must be modeled as conformational manifolds.

A cyclohexane bearing a long straight- or branched-chain alkyl substituent is slightly less trivial. When spectroscopic data primarily reflect the axial/equatorial equilibrium of the cyclohexane moiety and are relatively insensitive to the presence of *gauche* and *anti* conformations of the alkyl function, then, for bookkeeping purposes, the conformations of the alkyl chains can be grouped as substates of axial and equatorial conformational manifolds. At least to a first approximation, the energetic distributions of the alkyl substates within these manifolds can be taken as identical, the alkyl chain can be treated as though it were a rigid substituent, and the axial/equatorial equilibrium of the alkylcyclohexane can be modeled conceptually as a two-state process, like that of methylcyclohexane. A bookkeeping model that accounts for conformational variability of the alkyl chain would clearly be more complex.

Reversible denaturation of a cooperatively-stabilized globular protein can occur completely over a small range of denaturant concentrations or temperatures, generating an exceptionally diverse manifold of unstructured conformations (*Fig. 1*,*b*) [3]. A two-state bookkeeping model is often applied to such denaturation equilibria;

the native protein structure is approximated as a single conformation, and the non-native conformations are grouped together as members of the unstructured manifold. The quantitative spectroscopic methods often used to characterize denaturation are primarily responsive to the unique structural features of the native state and relatively insensitive to subtle changes within the unstructured manifold; this spectroscopic feature contributes to the predictive success of a two-state model [4].

Models devised for random coil linear polymers in solution provide an alternative for describing the properties of the unstructured manifold [5]. The interaction of the polymer chain with itself and with solvent molecules are important variables in these models, which have been very successful in describing experimental properties of both homo- and heteropolymers in which the polarities of the monomer side chains are similar. They are much less successful, though, when applied to heteropolymers like biopeptides for which these side chains differ in charge and polarity [6].

These three examples contrast strongly with conformational equilibria between unstructured states and helices formed by unaggregated helical α-peptides in aqueous solution. As shown in *Fig. 1*, *c*, helical conformations of varying length but similar stability are formed, and the collective helical state, like the unstructured state, must be treated as a conformational manifold. Although assumption-dependent computational models can do so, no measurement allows experimental assignment of a mole fraction to each of these conformations. In the place of mole fractions, quantitative helicity assays for helical α-peptides rely upon averaged differences in the environments of helical and nonhelical α-C-atoms. The bookkeeping model is focused on fractional helicity, which is the fraction of potentially helical α-C-atoms within the peptide that is helical under particular experimental conditions. If mole fractions χ_i of all helical conformations of varying lengths i (on a residue basis) were defined for a peptide of overall length n, fractional helicity (FH) can be expressed as the sum of all terms $(i/n)\chi_i$. Any measured helical property clearly is an abundance-weighted statistical average over the conformations within the helical manifold. If this property depends on helix length i or on other context effects, the measured property lacks a simple structural interpretation.

Are the available physical methods for quantitating helicity sufficient to characterize these complex systems at a chemical level? There are grounds for cautious optimism if all available methods are applied to α-homopeptides of variable length. Such studies should resolve current paradoxes and achieve a deeper level of rigor than currently available. At the very least, one should be able to predict amino acid sequences for exceptionally stable helical α-peptides with properties that approach those expected for the two-state model of *Fig. 1*, *b*.

The complex character of the bookkeeping model required for quantitative modeling of a typical helical α-peptide in solution distinguishes the case of *Fig. 1*, *c* from conformational analysis familiar to the physical organic chemist. The *Lifson–Roig* (LR) algorithm was introduced more than 40 years ago to describe pH-, length-, and temperature-dependent helix/coil equilibria of high-molecular-weight polylysines and polyglutamates. With relatively minor changes, it has become the standard algorithm for modeling the properties of helical α-peptides in the size range of 10–50 residues. For a helical α-peptide of length n, the LR algorithm generates a state sum containing 2^n terms from which all other properties that can be correlated with

experiment are calculated. This complex state sum is rarely written or analyzed explicitly. As a result, for the nonspecialist, LR predictions often acquire a 'Delphic' status that impedes critical analysis. Clarification of the conformational differences between α- and β-peptides is likely to require a rigorous, incisive, quantitative comparison of the energetics of the conformations that make up their respective unordered and structured states. A first step is a demystification of the LR bookkeeping tool itself. Although the biophysicist and the structural biologist likewise address the conformational analysis of biopolymers, they usually do not share the preference of the physical organic chemist for a rigorous, molecularly-based understanding that is rooted, at least initially, in a thorough heuristic analysis of the simplest pertinent systems. This report attempts to frame the helicity problem from this perspective and emphasizes results and implications of modeling of simple helical α-homopeptides that are now experimentally accessible.

The helicity of α-peptides covers an extensive literature, primarily derived from the study of *charged* homo- and heteropeptide sequences [7]. The recent literature demonstrates impressive advances in predictive correlations of α-peptide helicity data with sequence and structure [8]. The narrow focus of this review is not intended to minimize the importance of either these results or recent refinements or alternatives to the classical L/R algorithms [9][10]. Its primary aim is a recasting of the usual exterior features of a L/R algorithm in forms that more readily reveal to the nonspecialist its underlying assumptions, its potential for generalization, and its intrinsic limitations. In the past, the conformational problems posed by biopolymers have been profitably addressed by an exceptionally eclectic group of bioscientists. The challenge posed by the striking energetic differences between conformations of α- and β-peptides may signal that the time is ripe for new contributions.

2. The Classical 4 × 4 *Lifson–Roig* Matrix Algorithm. – For an α-peptide with n residues that terminates in amide functions at each of its ends, a LR state sum combines, singles out, and weights 2^n of its conformations in the following way. Only conformations that result from changes in the n pairs of α-peptide backbone dihedral angles are considered. Side chain conformations are ignored. If the ϕ and/or ψ dihedral angles at each of the backbone α-C-atoms both lie within the helical limits of the *Ramachandran* diagram, the site conformation is assigned as h (for *helical*); if either ϕ or ψ lie(s) outside helical limits, the site conformation is assigned as c (for *coiled*). The conformation for the peptide sequence is assumed to be defined by the set of site assignments and is labeled by a sequence of h,c symbols, one for each of the n α-sites. Weighting rules are then used to translate each label into an energetic weight for that conformation, and the sum of weights for each of the 2^n conformations defines the state sum. For example, for the conformation hhhccc of a hexapeptide, ϕ and ψ of the first three α-C-atoms lie in the helical region, and those of the second three α-C-atoms lie in nonhelical regions. Since the backbone amide functions are planar, assigning ϕ and ψ helical values restricts the geometry of the first three residues substantially, but the second three may sample many different ϕ and ψ angles. Because the focus of the algorithm is helicity, these diverse c state conformations are grouped and assigned a collective *nonhelical* weight. The presence or absence of amide-forming end caps is a small but significant point. The two terminal residues of a simple α-peptide sequence of length n define its caps, and, for LR modeling, therefore, the peptide has at most $(n-2)$

helical α-C-atoms, and $2^{(n-2)}$ conformations and conformational labels. If the sequence is terminated by amide-forming caps like Ac and NH_2, then the corresponding numbers[1]) are n and 2^n, respectively.

For an α-homopeptide, *Eqn. 1* defines the conventional LR state sum for a homopeptide of length n, which is a product of two end vectors and a series of n identical matrices. Each matrix i in the ordered matrix product assigns the site weight of the amino acid residue at site i in the conformational label. For the label hhchcc, the fourth matrix of the product assigns a weight to the site conformation h that characterizes the fourth peptide residue. The state sum is a polynomial in two free variables, the so-called helix-initiation constant v and the helical propensity w. Although w is defined as dependent on both temperature and type of amino acid residue, v is usually taken to be independent of both.

$$\text{state sum} = \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} w & v & 0 & 0 \\ 0 & 0 & 1 & 1 \\ v & v & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}^n \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \tag{1}$$

$$(v+1)^n = \begin{bmatrix} 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} v & v \\ 1 & 1 \end{bmatrix}^n \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} v & v & 0 & 0 \\ 0 & 0 & 1 & 1 \\ v & v & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}^n \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \tag{2}$$

Why has a matrix product been used in this calculation, what are the helical weighting rules, and how does *Eqn. 1* implement these rules? These questions are best addressed by considering a directly related but conceptually simpler assignment process. A comparison of the vector-matrix product of *Eqn. 1* with the second vector-matrix expression in the series of equalities of *Eqn. 2* shows that the matrix of the former results from a single substitution of a new coefficient w for the v coefficient at the first row and column of the corresponding matrix of *Eqn. 2*.

The polynomials calculated by this simpler vector-matrix expression of *Eqn. 2* are familiar binomial series in 1 and v. For a capped α-pentapeptide, this series is $(1+v)^5 = 1+5v+10v^2+10v^3+5v^4+v^5$, a state sum that assigns weights to each of the $2^5 = 32$ conformations. In this sum, the single conformation labeled ccccc is weighted 1; the five conformations containing a single h and four c are each weighted 1^4v; the ten conformations containing two h and three c residues are each weighted 1^3v^2, and the single hhhhh conformation is weighted v^5, *etc.* This is the simplest of all models for the conformational energetics of a polymer that contains n flexible, identical, noninteracting subunits, and any of the three equivalent algorithms of *Eqn. 2* calculates its state sum. The model is based on the assumption that the energetic contribution of each conformation can be written as a product of n independent site conformational weights, 1 or v, at each of its α-C-atoms.

---

[1]) Each α-C of a LR peptide must be flanked by a pair of primary or secondary amide residues. In order for the two amino acid residues at the ends of an α-peptide of n residues to join a helix, the peptide must bear *N* and *C*-terminal caps that ensure participation of these residues in formation of helical H-bonds.

Although modeling calculations of the ϕ,ψ-dependent energies of α-peptide conformations show that the total area of energetically accessible space at normal temperatures is strongly dependent on the bulk of the side chain linked to the α-C-atom [11], an average value can be assigned to a typical α-amino acid side chain. From this, the parameter v can be estimated as a ratio of areas of accessible helical *versus* nonhelical ϕ,ψ space, in approximate agreement with $v = 0.05$, assigned from LR modeling [12]. Accepting this value, one can estimate the mole fraction of the ccccc conformation within the overall peptide manifold as $1/(1+v)^5 = 0.78$, that of each of the five conformations containing a single h residue as $v/(1+v)^5 = 0.04$, and that of the single hhhhh conformations as $v^5/(1+v)^5 = 2 \times 10^{-7}$. With this simple model, helical conformations are assigned very small weights, contrary to experiment[2]).

Clearly, a helix-stabilizing effect is missing from this model. Unlike the binomial expansion, an algorithm capable of distinguishing helical from nonhelical regions must identify clusters of adjacent h residues and weight them differently from isolated or paired h residues. In the simple LR model, this weighting is achieved by assigning different weights to h,c triads that appear within the sequence of the conformational label. Thus, the label chchhhcch contains an hhh triad that signals the presence of a short helical region, but the label hhchcchch does not.

To generate a list of triads that can be used in this weighting process, a pair of additional c-site weights are added at each end of the conformational label to signify *N*- and *C*-capping. This new sequence is used to construct an ordered series of overlapping h,c triads, as shown in *Eqn. 3* for an α-octapeptide conformation of the type cchhhchc. The underlined center h,c symbol of the ith triad in this sequence corresponds to the ith site h,c conformation that must be weighted by the ith matrix in the product of *Eqn. 1*. Because triad sequences overlap, the center symbol of any triad is identical with both the third symbol of its predecessor and the first symbol of its successor. Thus, the first symbol of any triad corresponds to the $(i-1)$th site conformation, and the third symbol to the $(i+1)$th site conformation. As a result, in order to weight h-site conformers in helical conformations more strongly than an isolated h-site conformer corresponding to the triad c$\underline{\text{h}}$c, or the paired h-site conformers of the triads h$\underline{\text{h}}$c and c$\underline{\text{h}}$h, the h-site conformer of the triad h$\underline{\text{h}}$h must be assigned a new weight w that is greater than the weight v used in the binomial model.

Conformation label + two end caps: c|cchhhchc|c (3)

Sequence of eight triads: c$\underline{\text{c}}$c c$\underline{\text{c}}$h c$\underline{\text{h}}$h h$\underline{\text{h}}$h h$\underline{\text{h}}$c h$\underline{\text{c}}$h c$\underline{\text{h}}$c h$\underline{\text{c}}$c

$$\begin{array}{c} \\ \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix} \end{array} \quad \begin{array}{c} \begin{matrix} \text{h} & \text{c} & \text{h} & \text{c} \end{matrix} \\ \begin{matrix} \text{hh} \\ \text{hc} \\ \text{ch} \\ \text{cc} \end{matrix} \begin{bmatrix} w & v & 0 & 0 \\ 0 & 0 & 1 & 1 \\ v & v & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \end{array} \qquad (4)$$

[2]) It might also be noted that by assigning the same value of v to all amino acid residues, this model ignores the substantial variations in *Ramachandran* h,c energetics predicted by molecular modeling.

Each of the eight triads of *Eqn. 3* corresponds to one of the eight nonzero coefficients of the 4 × 4 LR matrix of *Eqn. 1*. To clarify this bookkeeping feature, *Lifson* and *Roig* introduced the row and column identifiers of *Eqn. 4*. The two h,c symbols of each row identifier correspond to the first two of a triad, and the symbol of each column identifier corresponds to the third symbol of a triad. Thus, one can group the eight triads with the corresponding row/column identifiers (i,j) and their site weights w as follows: hhh: (1,1), w; hhc: (1,2), v; hch: (2,3), 1; hcc: (2,4), 1; chh: (3,1), v; chc: (3,2), v; cch: (4,3), 1; and ccc: (4,4), 1. In *Eqn. 5*, the triad sequence of *Eqn. 3* is repeated, with corresponding (i,j) terms and weights listed underneath each triad, defining two new series, each with a computational significance. The product of the eight (i,j) terms weights the conformation labeled cchhhchc as v^3w, equivalent to the state-sum term generated by a corresponding path traced through the matrix product.

ccc	cch	chh	hhh	hhc	hch	chc	hcc	
(4) (4,4)	(4,3)	(3,1)	(1,1)	(1,2)	(2,3)	(3,2)	(2,4)	(4)
1 1	1	v	w	v	v	1	1	1 = v^3w (5)

Important regularities appear in the above example. The overlapping feature of the triad sequences reflects the form $(i,j)(j,k)(k,l)$ of the matrix row/column identifiers, which, in turn, is defined by the conventions of matrix multiplication. In simple LR algorithms, a numerical weight of 1.0 is assigned to all *c*-site conformations, and, from *Eqn. 4*, these appear in even-numbered matrix rows, and, thus, for their (j,k) identifiers, j is even. The h-site conformations are assigned weights of either v or w, which appear in odd-numbered matrix rows. If one starts with the conformational label and constructs the product of (j,k) identifiers that corresponds to a particular conformation in a stepwise process, working from left to right, then, when a triad sequence ends in h, the k coefficient of the appropriate (j,k) pair must be odd; when the sequence ends in c, then k of the appropriate (j,k) must be even. Given the list of nonzero (i,j) coefficients and the conformational label, one can calculate the overall LR weighting without matrix multiplication, as shown by a second example below.

This process starts with selection of a nonzero coefficient from the first vector of *Eqn. 1*. For the α-pentapeptide sequence hhhch, the first residue to be weighted has an h-site conformation, so the odd-numbered coefficient (3) from this vector is selected, which must match the j for the first (j,k) coefficient. Since the first triad is chh, k for this coefficient is odd, and the choice is (3,1). The incomplete coefficient product is (3)(3,1)(1,k), but since the second triad is hhh, the incomplete coefficient must be (1,1). Continuing this analysis generates the full (j,k) sequence (3)(3,1)(1,1)(1,2)(2,3)(3,2)(2), which assigns a weight v^3w to the conformation hhhch. This weighting path can be traced through the explicit vector-matrix product of *Eqn. 6*; the nonzero matrix coefficients that contribute to the weighting are highlighted in red. The weighting of a particular h,c sequence by the algorithm of *Eqn. 1* can be regarded as the vector-matrix product that results when only the red nonzero coefficient is retained and all others are set equal to zero. The complete state sum is generated by the

algorithm of *Eqn. 6* and corresponds to all nonzero paths that can be traced through the matrix product.

$$\begin{bmatrix}0&0&\mathbf{1}&1\end{bmatrix}\cdot\begin{bmatrix}w&v&0&0\\0&0&1&1\\\mathbf{v}&v&0&0\\0&0&1&1\end{bmatrix}\cdot\begin{bmatrix}\mathbf{w}&v&0&0\\0&0&1&1\\v&v&0&0\\0&0&1&1\end{bmatrix}\cdot\begin{bmatrix}w&\mathbf{v}&0&0\\0&0&1&1\\v&v&0&0\\0&0&1&1\end{bmatrix}\cdot\begin{bmatrix}w&v&0&0\\0&0&\mathbf{1}&1\\v&v&0&0\\0&0&1&1\end{bmatrix}\cdot\begin{bmatrix}w&v&0&0\\0&0&1&1\\v&\mathbf{v}&0&0\\0&0&1&1\end{bmatrix}\cdot\begin{bmatrix}0\\\mathbf{1}\\0\\1\end{bmatrix} \quad (6)$$

The LR algorithm of *Eqn. 1* implements the following series of model assumptions. It retains the 2^n conformations of *Eqn. 2*, as well as the implicit assumptions that molecular conformations do *not* perturb each other and, thus, can be weighted independently, and that the weights of nonhelical regions depend only on the h,c dispositions of individual α-C-atoms. As with *Eqn. 2*, single or paired h terms spaced in the h,c sequence by intervening c terms are assigned v weights. Thus, in the LR model, the weight v^4 is assigned to each of the α-octapeptide h,c conformational labels chchchch, chhcchhc, and hhchchcc. As noted above, *Eqn. 1* weights clusters of three or more h residues differently, thus, defining the complete three-residue helical turn, with its single intramolecular H-bond, as the simplest helical conformation.

Three rules for weighting residues in helical clusters are used. First, within a helical cluster, the first and last residues are assigned v weights. Second, a w weight is assigned to each interior h residue in each cluster; thus the weights of the sequences chhhc, chhhhhc, hhhhhh, hhhhhhhhh, and chchchhhhh are, respectively, v^2w, v^2w^3, v^2w^4, v^2w^7, and v^4w^3. The helical propensity w reflects the enthalpic stabilization resulting from amide/amide H-bonding and hydrophobic effects that result from formation of a helical loop; w can be defined as the tendency of an amino acid *linked to a preexisting helix* to join and lengthen it by one residue. A propensity of $w = 1$ then corresponds to a residue that is indifferent to helix propagation. This second rule implements the simplest of the possible helix stability assumption, *i.e.*, the free energy of a helical region is a linear function of the length of its h cluster. Third, two regions containing consecutive h sequences separated by one or more c residues are treated as independent helices. Thus, the sequence hhhhhhc with one independent h sequence is assigned a weight v^2w^4, but the sequence hhhchhh is assigned the much smaller weight v^4w^2.

After his proposal of the original LR algorithm, *Lifson* questioned the severity of this third rule [13]. Its structural penalty implies that interruption of helical structures at a central position by a single nonhelical ϕ or ψ backbone angle is sufficient to destroy its propagating capacity at that site. An experimental measurement of the energetic cost of these broken sequences has proved difficult, and despite *Lifson*'s concern, the original broken sequence weighting is almost universally used today.

The two-state helix/nonhelix assumption of the model is embodied in the two nonzero coefficients that appear in rows or columns of both matrix and vectors. Numbered rows and columns can be assigned odd or even parities, and these are tied, respectively, to the h,c symbols that identify the conformation at each α-C-atom. The nonzero coefficients that appear in the odd-numbered first and third matrix rows assign to h identifiers either v or w weights, and the even-numbered nonzero coefficients in second and fourth rows assign to c identifiers unity weights. The row/column rules of

matrix multiplication ensure that, if nonzero coefficients are present in the first or third columns of a given matrix, its following matrix must assign v or w weights to the next residue in the peptide conformation. Correspondingly, nonzero coefficients in the second or fourth columns direct assignment of 1.0 weights by the following matrix.

In the next section, the triad sequences and their corresponding (j,k) matrix coefficients are used to tailor the nonzero coefficients of the matrices used in LR calculations to particular lists of conformation weights. The LR algorithm then becomes a plastic tool for defining and implementing new weighting schemes for the conformational manifolds of helical peptides.

The analysis of *Eqns. 5* and *6* shows that matrices and matrix multiplication that define the usual calculation of LR state sums are formalisms that ensure easy implementation of the multiplication of n (j,k) coefficients drawn from a list of eight. However, typical PC software is exceptionally efficient at carrying out matrix multiplication and is particularly suited for maximizing multiplication speed for matrices that contain high percentages of zero coefficients. Calculation of a state sum for an α-peptide containing 80 amino acids involves calculation and summation of more than 10^{24} terms, yet, a PC executes this numerical evaluation in less than seconds, and even a symbolic calculation of this state sum as a function of undefined w is complete in minutes. When first devised, the LR algorithm was primarily applied to linear oligopeptides containing hundreds of amino acids. In this 'pre-computer' age, complex matrix methods were used to render the LR algorithm tractable. At that time, the smallest matrix dimensions consistent with the modeling necessities were computational essentials. In the PC age, generalizations of the original 4×4 and 3×3 matrices of *Lifson* and *Roig* to matrices of much larger dimensions have many advantages.

3. The 8×8 *Lifson–Roig* Matrix: Calculations of Length- and w-Dependent Fractional Helicities for α-Homopeptides. – The algorithm of *Eqn. 1* yields the state sum ss(5) for an α-homopentapeptide, written explicitly in *Eqn. 7*. The following features can be identified. The state sum is a polynomial in variables v and w that contains a pure polynomial in v. This corresponds to the weight of conformations that include no helical regions. Because typical values for v lie close to 0.05, the first three terms of this polynomial provide a good approximation to this weight, and these terms are shared with the v polynomial of Eqn. 2.

$$\mathrm{ss}(5) = 1 + 5v + 10v^2 + 7v^3 + v^4 + v^2\,(3w + 2wv + 2w^2 + w^3) \tag{7}$$

The state sum terms in the remainder of the polynomial can be associated by inspection with h,c conformational labels. The last term, v^2w^3, weights the single completely-helical sequence hhhhh; the term $2v^2w^2$ corresponds to the two conformations chhhh and hhhhc, and the term $3v^2w$ represent cchhh, chhhc, and hhhcc. Two of these labels begin or end in cc sequences; they can be paired with labels hchhh and hhhch, which result from single h,c replacements that do not increase the length of the helical region. They correspond to the remaining w-weighted term $2v^3w$ in *Eqn. 7*. All state sums derived from *Eqn. 1* share these features, and all but one term in *Eqn. 7* can be translated unambiguously into one or more sequence labels. The exception belongs to a type that appears in all state sums for peptides containing seven or more residues. The term $3v^4w^2$ weights two distinct sequence types: (hhhhchh + hhchhhh) and

hhhchhh. These correspond to a four-residue helical sequence separated from a pair of h residues and to two three-residue helical sequences, respectively. From the weight alone, it is impossible to identify the ratio of sequence types that occur in the conformational manifold. This ambiguity results from the use of the same v symbol for two distinct types of site conformations, nonhelical single or paired h-weighted α-C-atoms and the pair of h-weighted α-C-atoms at the ends of helical sequences. This dual role for v is an intrinsic feature of LR algorithms based on 4×4 matrices, as can be easily demonstrated by assuming that the w,v assignments for coefficients of the odd-numbered rows of the LR matrix of *Eqn. 1* have not been made. We can assign them by examining short conformational labels for which assignments are obvious and follow the weighting rules listed in the above section. The appropriate (j,k) products for the tetrapeptide sequences hhcc, hhhc, and hhhh appear in *Eqns. 8–10*.

$$\text{c|hhcc|c} \qquad (3)(3,1)(1,2)(2,4)(4,4)(4) \tag{8}$$

$$\text{c|hhhc|c} \qquad (3)(3,1)(1,1)(1,2)((2,4)(4) \tag{9}$$

$$\text{c|hhhh|c} \qquad (3)(3,1)(1,1)(1,1)(1,2)(2) \tag{10}$$

The respective state weights from the helix-weighting rules for each of these conformations are: hhcc, v^2; hhhc, v^2w; and hhhh, v^2w^2. We can assign values consistent with these state weights to the coefficients (1,1), (1,2), and (3,2) by noting that all vector coefficients and those (j,k) coefficients with even i or j values possess a weight of 1.0. It follows from *Eqn. 8* that $(3,1)(1,2) = v^2$; from *Eqns. 9* and *10* $(3,1)(1,1)(1,2) = v^2w$ and $(3,1)(1,1)(1,1)(1,2) = v^2w^2$, implying that $(1,1) = w = v^2w^2/(v^2w)$ and that the product $(3,1)(1,2)$ must be assigned a value v^2 when it appears in a weighting sequence for either isolated hh pairs or for the first and last residues of a helical sequence. This is the reason for the ambiguous interpretation of the coefficient $3v^4w^2$ in *Eqn. 7*.

A more-general LR algorithm can be constructed that assigns different parameters to the weighting of isolated or paired h-site conformations or h terms that appear at the beginning or end of helical sequences. The triad labels associated with a 4×4 matrix are intrinsic to the problem, since the label hhc can correspond to a matrix weight for the second of a pair of adjacent h-site conformations that are not part of a helix, or it can correspond to the second residue of a helix. To differentiate these two distinct weighting roles, tetrad sequences constructed like the triad sequences from the conformational label are needed. The conformation cc | chhhchh | c thus yields such a tetrad sequence (*Eqn. 11*). A weighting of an h-site conformation that terminates a helical sequence corresponds to the tetrad hhhc, which is now distinguished from the tetrad chhc, which corresponds to a *nonhelical* pair of adjacent h-site conformations.

$$\text{cc}\underline{\text{c}}\text{h} \quad \text{cc}\underline{\text{h}}\text{h} \quad \text{ch}\underline{\text{h}}\text{h} \quad \text{hh}\underline{\text{h}}\text{c} \quad \text{hh}\underline{\text{c}}\text{h} \quad \text{hc}\underline{\text{h}}\text{h} \quad \text{ch}\underline{\text{h}}\text{c} \tag{11}$$

The 2×2 and 4×4 matrix products of *Eqn. 2* can be generalized to products of matrices with 2^k columns and rows. These new matrix products have the feature that the triad sequences that correspond to the weighting sequences of 4×4 matrices become $(k-1)$meric. To implement tetrad sequences, a LR algorithm based on 8×8 matrix products is, thus, required. The transformation of nonzero elements shown in *Eqn. 12* demonstrates how the 4×4 LR matrix can be converted into a fully equivalent 8×8 LR matrix. To obtain the same state function as in *Eqn. 1*, but using the 8×8 matrix of

Eqn. 12, the required vectors *v*1 and *v*2 are [0,0,0,0,0,0,1,1] and [0,1,0,1,0,1,0,1]. Although the 8×8 matrix quadruples the total number of coefficients found in the 4×4 matrix, it only doubles the number of *nonzero* coefficients. The computational time on a desktop PC, therefore, remains convenient.

$$\begin{bmatrix} w & v & 0 & 0 \\ 0 & 0 & 1 & 1 \\ v & v & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \mathrm{A} \\ \mathrm{B} \end{bmatrix} \qquad \begin{bmatrix} \mathrm{A} & 0 \\ 0 & \mathrm{B} \\ \mathrm{A} & 0 \\ 0 & \mathrm{B} \end{bmatrix} = \begin{bmatrix} w & v & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & v & v & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ w & v & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & v & v & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \tag{12}$$

$$\begin{array}{c} \\ \mathrm{hhh} \\ \mathrm{hhc} \\ \mathrm{hch} \\ \mathrm{hcc} \\ \mathrm{chh} \\ \mathrm{chc} \\ \mathrm{cch} \\ \mathrm{ccc} \end{array} \begin{array}{c} \begin{array}{cccccccc} \mathrm{h} & \mathrm{c} & \mathrm{h} & \mathrm{c} & \mathrm{h} & \mathrm{c} & \mathrm{h} & \mathrm{c} \end{array} \\ \begin{bmatrix} w & v & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & v & v & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ w & v & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & v & v & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \\ \mathbf{a} \end{array} \qquad \begin{array}{c} \begin{bmatrix} w & t & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & t & v & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ w & \frac{v^2}{t} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & t & v & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \\ \mathbf{b} \end{array} \tag{13}$$

New row/column identifiers can be added to the matrix of *Eqn. 12*, as shown in matrix **a** of *Eqn. 13*. The three h,c symbols of each row identifier correspond to the first three of a tetrad, and the third symbol of the *i*th tetrad in the sequence corresponds to the weight assigned by the *i*th matrix in the matrix product. Each column identifier corresponds to the fourth symbol of the tetrad. The presence of the first symbol distinguishes these labels from those applied to 4×4 matrices; for the upper part of the 8×8 matrix, these are h; for the bottom part, they are c. The tetrads hhhc and chhc to be distinguished correspond to the matrix coefficients (1,2) and (5,2). To make this distinction, they must be assigned different values in a LR algorithm. One must first check whether assigning different parameters to the two distinct roles of *v* consistently distinguishes between the nonhelical sequence chc and the helical sequences chhhc and chhhhc. The required weighting assignments are as follows: *w* is retained for tetrads hhhh (1,1) and chhh (5,1), and *v* is retained as the weight for tetrads cchc (7,3) and hchc (3,6), but *v* is changed to a new weight *t* for the sequences hhhc (1,2), hchh (3,5), and cchh (7,5). The former set of *v* corresponds to isolated h residues whose nonhelical character is well-defined, but the latter set of *t* correspond either to helical h residues or isolated, paired, nonhelical h residues. (This ambiguity must be clarified and corrected by the chhc (5,2) assignment, as shown below.)

One can check whether the v and t assignment consistently distinguishes between a nonhelical sequence chc and helical sequences chhhc and chhhhc. As seen in *Eqns. 14–18*, both cases test correctly.

cc|hccc|c cc|hc c|hcc hccc ccc|c

$$(7)\ (7,6)\ (6,4)\ (4,8)\ (8,8)\ (8) = v \tag{14}$$

cc|hchc|c cc|hc c|hch hchc chc|c

$$(7)\ (7,6)\ (6,3)\ (3,6)\ (6,4)\ (4) = v \tag{15}$$

cc|hhhc|c cc|hh c|hhh hhhc hhc|c

$$(7)\ (7,5)\ (5,1)\ (1,2)\ (2,4)\ (4) = t^2w \tag{16}$$

cc|hhhh|c cc|hh c|hhh hhhh hhh|c

$$(7)\ (7,5)\ (5,1)\ (1,1)\ (1,2)\ (2) = t^2w^2 \tag{17}$$

cc|hchhh|c cc|hc c|hch hchh chhh hhh|c

$$(7)\ (7,6)\ (6,3)\ (3,5)\ (5,1)\ (1,2)\ (2) = t^2wv \tag{18}$$

The coefficient $(5,2)$, which corresponds to the tetrad chhc, remains to be assigned, and the simplest conformation that embodies this sequence has the label hhcc, which generates the tetrad sequence and weighting product of *Eqn. 19*. The required weight is v^2, and assigning values of $(7)=(8)=(2,4)=(4,8)=1$, and $(7,5)=t$, implies that $(5,2)=v^2/t$. This analysis assigns all coefficients of an 8×8 LR matrix that embodies three fundamental parameters v, t, and w, and when the vector-matrix product of *Eqn. 20* that contains this matrix is applied to larger homopeptides, correct state sums are generated[3]).

cc|hhcc|c cc|hh c|hhc hhcc hcc|c

$$(7)\ (7,5)\ (5,2)\ (2,4)\ (4,8)\ (8) = v^2 \tag{19}$$

One feature of the v, w, and t variables remains to be defined. The new parameter t reflects the properties of the first and last residues of a helical sequence. Should it be defined like v as independent of temperature and amino acid side chain, or should it share the dependence on these variables that is assumed in LR models for w? Although the latter option has many attractive features, a major flaw is the introduction of two temperature- and residue-dependent parameters. A solution is provided if the variable t of *Eqn. 13* is replaced by wt, yielding the full state sum algorithm of *Eqn. 20*. The new variable t in wt is assigned the temperature and residue independence of v, and the temperature and residue dependence of w is applied to the first and last helix sites of

[3]) The assignment of distinct roles to the parameters v and t is conceptually convenient and allows heuristic modeling. At present, it is difficult to devise experiments that distinguish between these parameters, and when data are analyzed, v is usually set equal to t.

the helical conformation. The resulting new state sum $ss_{v,t}(8)$ for an eight-residue α-peptide is given in Eqn. 21.

$$ss_{v,t}(n) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} w & wt & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & wt & v & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ w & \frac{v^2}{wt} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & wt & v & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}^n \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \tag{20}$$

$$ss_{t,v}(8) = (1 + 8v + 28v^2 + 50v^3 + 45v^4 + 16v^5 + v^6) + \tag{21}$$

$$t^2 (6w^3 + 5w^4 + 4w^5 + 3w^6 + 2w^7 + w^8) +$$

$$t^2 (v (20w^3 + 12w^4 + 6w^5 + 2w^6) + v^2 (24w^3 + 9w^4 + 2w^5) + v^3 6w^3) +$$

$$t^4 (3w^6 + 2w^7)$$

Each term in $ss_{v,t}(8)$ identifies the properties of the conformations that it weights. The exponent of t divided by two defines the number of separate helical regions within a corresponding conformation; the exponent of w defines the overall length of its helix or helices, and the exponent of v defines the number of isolated or paired α-C-atom h weights. The sum of w and v exponents defines the total number of h weights within the conformation. The total number of each type of conformation can be obtained by summing numerical coefficients. *Eqn. 21*, thus, weights 149 nonhelical conformations, 21 conformations containing a single helical region and no other helical sites, and 5 conformations containing two helical conformations.

The construction of an 8×8 matrix from a 4×4 matrix can be generalized to form a series of LR matrices of size $2^k \times 2^k$, where k is any positive integer, and algorithms can be tailored using these large matrices to assign weights to conformations at the ith α-C-atom of the sequence that depend on the h,c weights at any of the $(k-1)$-preceding α-C-atoms. In the following section, 16×16 matrices are applied to the modeling of peptide protection factors. In the remainder of this section, the algorithm based on the 8×8 matrix **b** of *Eqn. 13* is used to explore the effects of variations of v, t, w, and n on α-peptide fractional helicity (FH) and the mole fraction of nonhelical conformations (χ_{Nonh}) within the LR manifold.

$$FH = \sum_{i=3}^{n} FH_i \qquad FH_i = \left(\frac{i}{n}\right) \cdot \chi_{Hel,i} \tag{22}$$

Fundamental features of LR algorithms are reflected in the model-predicted dependence of the mole fraction of nonhelical peptide conformations χ_{Nonh} and the fractional helicity FH on peptide length n and values for w, t, and v. The FH is conveniently

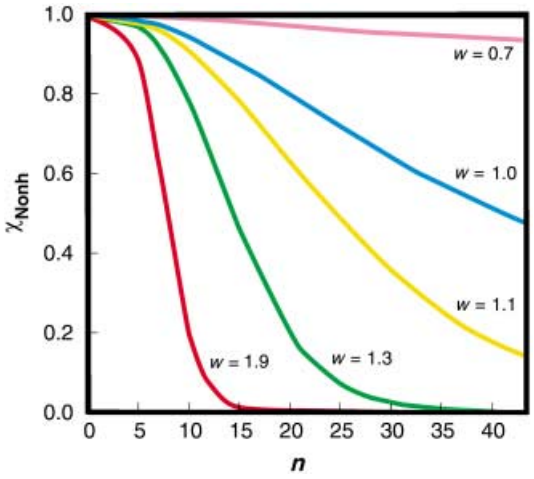

Fig. 2. Lifson–Roig *modeling of the dependence of the mole fraction of nonhelical conformations* χ_{Nonh} *as a function of peptide length* n *and helical propensity* w ($v = t = 0.048$). For $w < 1.1$, the nonhelical conformations dominate, even for long peptide sequences.

calculated as the sum over all helical lengths of FH_i, *i.e.*, the fractional helicity of helices of length i, in turn calculated from the mole fraction of these helices according to *Eqn. 22*.

In *Fig. 2*, the change in length dependence of χ_{Nonh} is shown as a function of the helical propensity w, which, for most α-amino acids, lies in the range of 0.5 to 1.6. Values below 1.0 correspond to helix breakers, and values significantly greater than 1.0 are strong helix formers. For very short α-peptides, χ_{Nonh} invariably approaches 1.0, implying that conformations containing helical regions are rare. For α-peptide lengths in the range of 10–15 residues, nonhelical conformations remain dominant, unless w assumes unusually large values. The α-peptide length for which $\chi_{Nonh} = 0.5$ is seen to vary significantly with w, and lengths greater than 30 residues are required for w values less than 1.1. This modeling prediction is consistent with the experimental properties of peptide sequences. Clearly, the conformations formed by β-peptides behave differently.

Fig. 3 provides the complementary picture of the effect of changes of length and helical propensity on FH. For any reasonable value of w, FH is predicted to be negligible for very short α-peptide sequences, and, for w-values of 1.1 or less, to increase modestly and nearly linearly with n. A cooperative dependence of FH on length is seen for larger w-values, and three distinctive length regions can be identified within each curve. In the first region, FH is close to zero, and χ_{Nonh} remains large, implying that most molecules lack helical regions. In this length range, the collective weight of t^2w^i for $i \leq n$ is substantially smaller than the random coil weight of the v polynomial of the state sum. The second region is characterized by a strong, nearly linear dependence of FH on n, reflecting a corresponding increase in the weight of t^2w^i terms; toward the upper limit of this region, χ_{Nonh} approaches zero. Nearly all molecules now contain helical regions, and the sum of all $\chi_{Hel,i}$ terms approaches 1.0. Inspection of the length graphs for $w = 1.5$ and 1.9 shows that FH for each lies significantly below 1.0, reflecting the weights of conformations that are partially helical. Although the weight

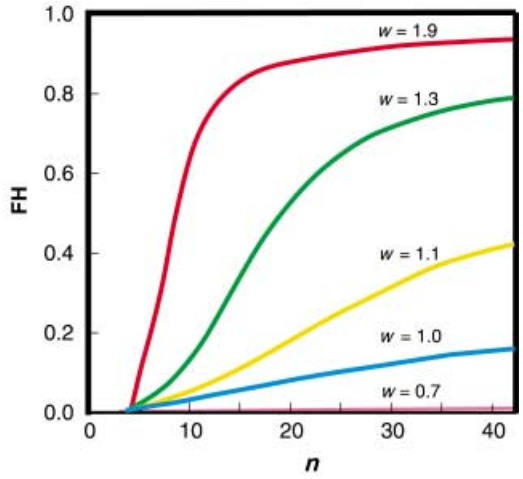

Fig. 3. Lifson–Roig *modeling of the fractional helicity as a function of peptide length* n *and the helical propensity* w ($v = t = 0.048$). For $w < 1.1$, FH remains modest, and little length-dependent cooperativity is evident.

t^2w^i of the completely helical state is the largest in the helical manifold, the presence of many partially helical conformations in this manifold lends it a strong entropic stabilization. For this reason, the slope in the third region of the graph is modest. As n is increased by one residue, the weight of the completely helical state is increased by w, but more than $(n-1)$ new partially helical states are added to the manifold. This length-dependent trade-off of enthalpic *versus* entropic stabilization of the helical manifold ensures that, as a state average, a helical α-peptide is partially structured and frayed at its ends.

As seen in *Fig. 4*, the LR model allows explicit calculation of the distribution of the lengths of helical regions within conformations of the manifold. In *Fig. 4*, the relative contribution to the overall FH of molecules containing helical regions of length i is plotted against this length for a series of w-values. If $w = 1.0$, $t^2w^i = t^2$, and there is no enthalpic distinction between long or short helical regions; the dependence of FH_i/FH on i is flat (not shown). For $w = 1.1$ or 0.9, a small enthalpic bias is evident, but entropy is seen to govern the distribution. As w increases, the most frequent length i for a helical region approaches the overall peptide length ($n = 19$), and the weight of very short helical regions is very small, but the effect of entropy-based conformational averaging is still evident, even for $w = 1.9$. At 25°, this w-value corresponds to a ΔG° of -1.6 kJ/mol for the addition of *one* new α-amino acid residue to a preexisting helix. With the assumptions that underlie the LR model, it is clear that a significantly larger ΔG° per residue is required to ensure dominance of the completely helical conformation.

The effects on FH of holding length and helix propensity constant and varying t or v are explored in *Fig. 5*. The primary effect of doubling t is evident as a marked reduction in the peptide length n at which FH assumes significant values. In effect, the curve has been shifted to the left, owing to an increase of the magnitude of all t^2w^i terms. The value of t has much less influence in *region 3* of the graph, since the weight of nonhelical

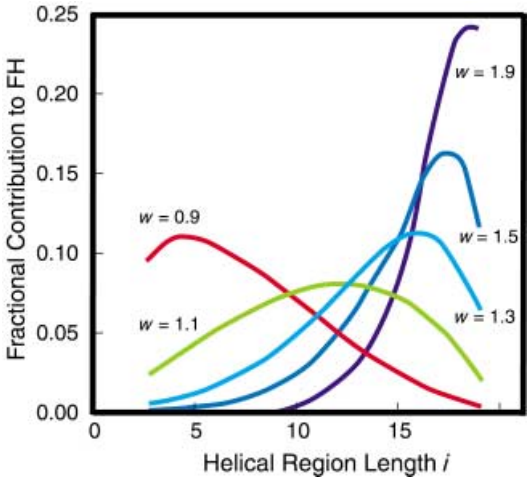

Fig. 4. Lifson–Roig *modeling of the relative contributions of long and short helical conformations to the helical manifold for a 19-residue peptide* ($v = t = 0.048$). The ordinate is the fraction of total FH that is contributed by helical conformations of lengths 3–19. The length homogeneity of the manifold increases with w, but, for $w < 2$, significant contributions from partial helices remain.

conformations now dominates the state sum. The state sum itself and its portion that weights the helical manifold are proportional to t^2, which largely cancels in their ratio. The effect of doubling v is smaller and opposes the effect of changes in t within *regions 1* and *2*. Its principal effect is to selectively stabilize the nonhelical and shorter, partially helical conformations, since v reflects the weight of nonhelical isolated and paired h-weighted α-C-atoms.

4. Modeling of CD-based Experimental Helicity Data Using *Lifson–Roig* Algorithms. – Three complementary experimental methods for quantitatively measuring helicity are particularly suited to study α-peptides in the small-to-medium-size range. A single parameter circular dichroism (CD) assay is currently the most convenient and widely used experimental characterization of α-peptide and protein helicity, since it mirrors the global conformational properties of a peptide backbone.

If a peptide demonstrates a characteristic CD signature, FH is often calculated as proportional to the experimental molar residue ellipticity at 222 nm ($[\theta]_{222,\mathrm{Exp}}$) [14]. Unfortunately, calibration problems currently remain unresolved and complicate quantitative interpretations [15] of CD data. Thus, calibrations based on two independent, NMR-based experimental methods are needed to fulfill the potential of $[\theta]_{222,\mathrm{Exp}}$ values as quantitative helicity monitors. These involve measurement of peptide NH protection-factors (PF) and so-called *t*/*c* values (derived from ^{1}H-NMR analysis of *Kemp*'s helix-inducing, N-terminal template 'AcHel'), which uniquely allow helical quantitation for short α-peptides [16] based on a LR algorithm tailored for modeling NMR data [17]. The analysis and LR modeling of PF data is described in detail in the next section.

As previously reported, new CD calibrations are required for alanine-rich peptides [15]. Polyalanines and alanine-rich peptides are the natural host contexts for exploring

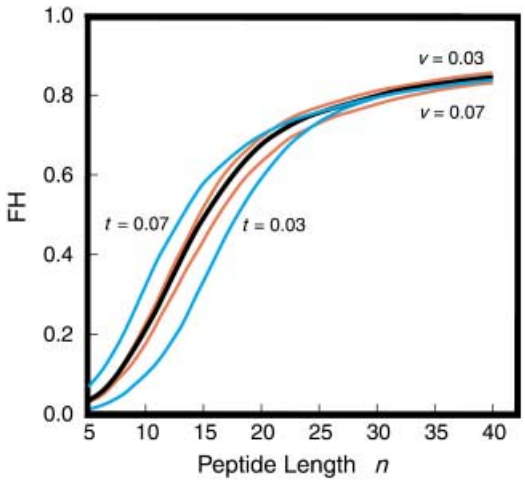

Fig. 5. Lifson–Roig *modeling of the dependence of FH on length* n *and the initiation parameters* t *and* v ($w = 1.3$). The black curve corresponds to $v = t = 0.05$. For green curves, v is fixed at 0.05 and t is varied to 0.03 and 0.07. For the red curves, $t = 0.05$ and $v = 0.03$ or 0.07. *Region 1* varies in approximate length from 10 for $v = 0.03$ to 5 for $v = 0.07$. The boundary between *regions 2* and *3* varies between lengths 20 and 25.

the relationship between helicity and amino acid composition and sequence, yet, recent evidence demonstrates that these peptides exhibit anomalously intense values of $[\theta]_{222,Exp}$ in H_2O at low temperatures, and the available evidence suggests that values for $[\theta]_{H222,n}$, the limiting ellipticity expected for a completely helical peptide of length n, is strongly temperature dependent [15].

The calibration value $[\theta]_{H222,n}$ needed to assign FH from experimental CD data is commonly modeled as proportional to $[\theta]_{H222,\infty}$, the residue ellipticity for a helical peptide of infinite length, as seen in *Eqn. 23*. The constant X in this equation is a length correction factor, and the logic of its role is clarified when $[\theta]_{H222,n}$ is converted to the molar ellipticity of *Eqn. 24*. Likely values for X fall in the range of 2–5, with a frequently cited value of 2.5 [18]. A recent estimate from the globular protein data base sets $[\theta]_{H222,\infty}$ equal to -37000 deg cm^2 dmol^{-1} [19], but, based on LR modeling of CD data, *ca.* 15% larger values are typical for alanine-rich peptides [20].

Rigorous experimental validation of *Eqn. 23* or the values of its coefficients has, hitherto, been difficult, since proper peptides for calibration have been unavailable. A definitive CD calibration of $[\theta]_{H222,n}$ requires length series of peptides of diverse types that have been established by *CD-independent* criteria to approach 100% helicity (FH = 1.0). Without such series, FH values calculated from CD data cannot be assigned error limits and lack the quantitative rigor required for defining and testing helicity algorithms. When the CD contributions of unstructured peptide residues are taken as insignificant at 222 nm, and when current values for $[\theta]_{H222,n}$ are used, LR algorithms can be applied to calculate expected values $[\theta]_{222,Calc}$ of experimental ellipticities $[\theta]_{222,Exp}$ from *Eqn. 25* in which FH_i is defined as in *Eqn. 22*[4]). For a large data set of $[\theta]_{222,Exp}$

[4]) For a helical peptide of length n, *Eqn. 25* is often approximated by $[\theta]_{calc} = FH \cdot [\theta]_{H,n}$. The error implicit in this approximation has recently been reported [25].

derived from a series of similar peptides, LR parameters w (and in some cases v) can be varied to yield a best fit between $[\theta]_{222,\mathrm{Exp}}$ and $[\theta]_{222,\mathrm{Calc}}$. When the CD contributions of unstructured (U) states are included in the calculation, then *Eqn. 26* yields $[\theta]_{222,\mathrm{Calc}}$.

$$[\theta]_{\mathrm{H222},n} = [\theta]_{\mathrm{H222},\infty} \cdot \left(1 - \frac{X}{n}\right) \tag{23}$$

$$[\theta]_{\mathrm{H222},n\ \mathrm{molar}} \equiv n \cdot [\theta]_{\mathrm{H222},n} \quad \text{and} \quad [\theta]_{\mathrm{H222},n\ \mathrm{molar}} = (n - X) \cdot [\theta]_{\mathrm{H222},\infty} \tag{24}$$

$$[\theta]_{222,\mathrm{Calc}} = \sum_{i=3}^{n} [\theta]_{\mathrm{H222},i} \cdot \mathrm{FH}_i \tag{25}$$

$$[\theta]_{222,\mathrm{Calc}} = [\theta]_{\mathrm{U222}} \cdot \chi_{\mathrm{Nonh}} + \frac{1}{n} \cdot \sum_{i=3}^{n} (i \cdot [\theta]_{\mathrm{H222},i} + (n - i) \cdot [\theta]_{\mathrm{U222}}) \chi_{\mathrm{h},i} \tag{26}$$

5. Modeling of Experimental Protection Factors Using *Lifson – Roig* Algorithms. – Protection Factors (PFs) can be measured for ^{15}N-labeled helical α-homopeptides of virtually any length and can define site helicities. For this reason, PFs provide the most-detailed and useful characterization of α-peptide helicity.

In D_2O, the rate of NH → ND exchange is dramatically reduced for a completely H-bonded conformation of a secondary amide [21]. The protection factor (PF) is defined at a fixed pH as the ratio of rate constants for exchange of a solvent-exposed model peptide NH relative to that of a partially helical NH of a candidate peptide. As developed by *Englander* and co-workers [22] and subsequently applied to the α-peptide helicity problem by *Baldwin* [23] and *Kallenbach* and co-workers [24], a series of ^{15}N-labeled site isomers of a helical peptide allow assignment of PF values for each H-bonded amide NH function of the peptide backbone. This uniquely-detailed, NMR-based multiparameter signature of site helicities complements and greatly extends the single-parameter calculation of CD-derived FH.

Derivation of an LR algorithm appropriate for PF modeling requires extension of the design principles developed in the preceding sections, and helicity modeling with these algorithms complements and adds incisive detail to the results shown in *Figs. 2–4*. The remainder of this section begins with algorithm construction, focusing initially on an important relationship between two distinct but closely related measures of site helicity. Finally, it presents results of LR modeling of these site helicity measures and generalizes the interpretation of experimental PF values.

The simplest measure of site helicity is the fractional population of α-C-atoms at site i that are part of helical conformations, $\chi_{\mathrm{H},\alpha\text{-}\mathrm{C}_i}$. Averaging $\chi_{\mathrm{H},\alpha\text{-}\mathrm{C}_i}$ over all α-C-atoms of an *N*- and *C*-capped helical α-peptide yields its FH. For the case of an α-pentapeptide, *Fig. 6,a*, illustrates how the value of $\chi_{\mathrm{H},\alpha\text{-}\mathrm{C}_i}$ can be calculated from the LR state sum. One selects from the state sum the helical weights that correspond to conformations for which an i-site h conformation is part of a helical region. For example, at site $i = 2$, reading down the column of identifiers, one finds that the appropriate conformations are A, B, C, D, D′, and E. Adding their weights and dividing by the state sum yields $\chi_{\mathrm{H},\alpha\text{-}\mathrm{C}_2}$.

$$\mathrm{PF}_i = \frac{1}{(1 - \chi_{\mathrm{H,NH}_i})} \qquad \chi_{\mathrm{H,NH}_i} = \frac{(\mathrm{PF}_i - 1)}{\mathrm{PF}_i} \tag{27}$$

Eqn. 27 relates PF_i to a different type of helical site weight, χ_{H,NH_i}, which is defined as the fractional population of NH protons of the residues that are intrahelically H-bonded to a backbone amide O-atom belonging to the residue at site $(i-4)$. The three intervening α-C-atoms at sites $(i-1)$, $(i-2)$, and $(i-3)$ must be α-helical for this H-bond to form, as seen in *Fig. 7* where the amide NH functions of the first three residues of the sequence lack PF_i values, although their α-C-atom must be helical for a PF_i to be defined at site 4. If the peptide is *C*-capped by a secondary amide, a $PF_{(n+1)}$ can, in principle, be measured for its NH. An *N*- and C-capped helical α-peptide of length n, thus, has n $\chi_{H,\alpha\text{-}C_i}$ values and a total of $(n-2)$ normal PF_i values including a C-terminal PF_{n+1}. A relationship between these is assigned at the close of this section.

The construction of a LR model for $\chi_{H,\alpha\text{-}C_i}$ is straightforward. Since helical weights must be selected from the state sum that correspond to conformations with helical regions that contain residue *i*, the *i*th matrix of a normal LR algorithm must be tailored to retain coefficients that assign w weights. From the 8×8 matrix of *Eqn. 20*, the coefficients $(1,1)=(5,1)=w$ and $(1,2)=(3,5)=(7,5)=tw$ are retained, but all other coefficients are set equal to zero. Symbolizing this matrix as 'mx' and the matrix of *Eqn. 20* as 'm', the value of $\chi_{H,\alpha\text{-}C_i}$ is obtained from *Eqn. 28*, where $v1=[0,0,0,0,0,0,1,1]$ and $v2=[0,1,0,1,0,1,0,1]$. One w weight assigned by mx would normally be reassigned as

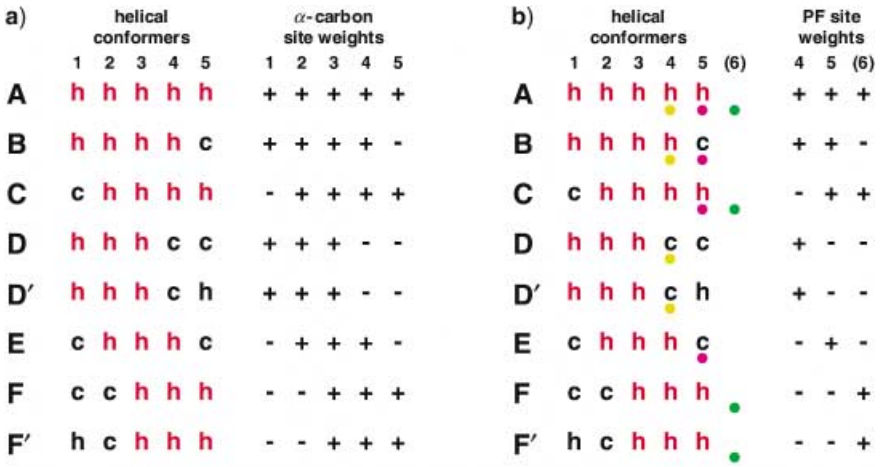

Fig. 6. a) *Site helicity calculated as* $\chi_{H,\alpha\text{-}C_i}$, *the helicity at each* α*-C-atom for an* α*-pentapeptide.* The seven helical conformations are abbreviated by capital letters and identified by their sequence labels, with helical regions shown in red. The $+/-$ labels in the state weights specify for each α-site *i* whether a particular conformation contributes to it. The contributions to site helicity at the first α-C-atom are found by reading down *column 1* and comprise conformations A, B, D, and D′, which corresponds to a *site 1* weight of $t^2(w^5+w^4+w^3(1+v))$; this is also the *site 5* weight. For *site 3*, the contributing conformations are A, B, C, (D + D′), E, (F + F′), which corresponds to a *site 3* weight of $t^2(w^5+2w^4+w^3(3+2v))$. b) *Site helicity calculated as* χ_{H,NH_i}, *the fractional population of NH protons at residue* i *that are intrahelically H-bonded to the amide O-atom at site* $(i-4)$. The colored dots beneath *sites 4*, *5*, and *6* symbolize the NH site of the H-bonds shown at the lower right of *Fig. 6*. The subsequences hhhh and hhhc identify conformations that contain a H-bond donor NH at the underlined site.Thus, *site 4* (yellow dots) contains conformations A, B, (D + D′); *site 5* (magenta dots) contains conformations A, B, C, E; *site 6* (green dots) contains conformations A, C, F + F′. Their respective site weights are: NH(4) and (6): $t^2(w^5+w^4+w^3(1+v))$; NH(5): $t^2(w^5+2w^4+w^3)$.

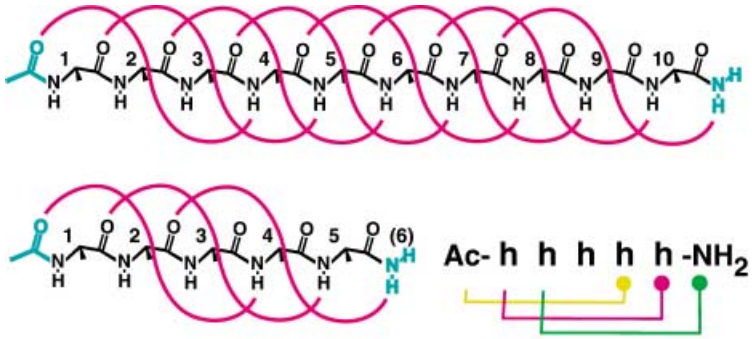

Fig. 7. *Schematic views of the α-helical H-bonding patterns of* N- *and* C-*capped deca- and pentapeptide sequences.* Each H-bond links atoms derived from five consecutive residues, but, owing to amide planarity, it constrains only the three pairs of dihedral ϕ/ψ angles for the central three residues. A shorthand notation for H-bonds is illustrated at the lower right for the capped pentapeptide.

v^2 by the (5,2) coefficient of the $(i+1)$th matrix termed 'mxx'. To eliminate this anomaly, mxx is defined as the matrix of *Eqn. 20* with the (5,2) coefficient set equal to zero.

$$\chi_{\mathrm{H},\alpha\text{-}\mathrm{C}_i} = \frac{v1 \cdot \mathrm{m}^{(i-1)} \cdot \mathrm{mx} \cdot \mathrm{mxx} \cdot \mathrm{m}^{(n-(i+1))} \cdot v2}{\mathrm{ss}} \tag{28}$$

Since the values of $\chi_{\mathrm{H,NH_i}}$ are defined by a helical conformation that extends over five consecutive peptide residues, they cannot be modeled by LR triads or tetrads. To ensure that the three residues that precede the *i*th residue are all helical, pentad sequences derived from the peptide conformational label must be used, as can be seen from inspection of the examples given in *Fig. 6,b* and *Fig. 7*. For a residue at site *i*, the appropriate first four terms of pentad sequences are hhhh and hhhc, which must be distinguished from chhh and chhc, which do not maintain helicity at site $(i-3)$. A 16×16 matrix termed 'nx' correlates the required pentads with row/column identifiers and weights: hhhhh (1,1) = *w*, hhhhc (1,2) = *tw*, hhhch (2,3) = 1.0, hhhcc (2,4) = 1; all other coefficients are set equal to zero. The ratio of matrix products required to calculate $\chi_{\mathrm{H,NH_i}}$ is given by *Eqn. 29*, in which the state sum ss can be calculated using 8×8 matrices in the usual way. The 16×16 matrix nx appears in the numerator of this equation, which also uses a matrix termed 'n', that implements a normal LR algorithm; it is constructed from the 8×8 matrix by generalizing the transformation of *Eqn. 12*, as shown in *Eqn. 30*.

$$\chi_{\mathrm{H,NH}_i} = \frac{v1 \cdot \mathrm{n}^{(i-1)} \cdot \mathrm{nx} \cdot \mathrm{n}^{(n-i)} \cdot v2}{\mathrm{ss}} \tag{29}$$

LR-modeled site dependences for $\chi_{\mathrm{H},\alpha\text{-}\mathrm{C_i}}$ and $\chi_{\mathrm{H,NH_i}}$ are compared in *Fig. 8* for a highly helical 19-residue α-homopeptide. Although $\chi_{\mathrm{H},\alpha\text{-}\mathrm{C_i}}$ is the conceptually more-incisive parameter, $\chi_{\mathrm{H,PF_i}}$ is easily calculated from experimental PF_i values, measured by well-defined and tested experimental protocols. Their large experimental ranges can define $\chi_{\mathrm{H,NH_i}}$ with precision; protocols for measuring $\chi_{\mathrm{H},\alpha\text{-}\mathrm{C_i}}$ are less precedented[5]).

[5]) Correlation of ^{13}C-NMR chemical shifts of α-C-atoms with residue helicity [26] suggests that, if site dependences are relatively unimportant, these parameters may provide an experimental assay for $\chi_{\mathrm{H},\alpha\text{-}\mathrm{C_i}}$.

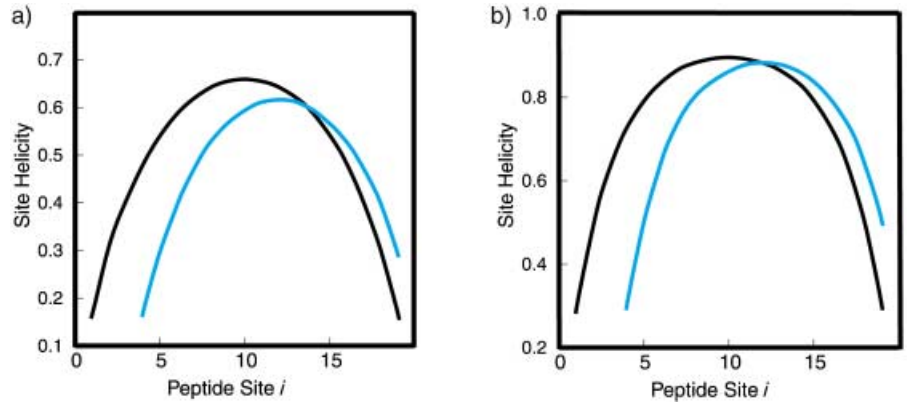

Fig. 8. Lifson–Roig *modeling of site helicities for a 19-residue peptide* ($v = t = 0.048$). a) $w = 1.3$. b) $w = 1.5$. The black curve corresponds to $\chi_{H,\alpha\text{-}C_i}$; the cyan curve corresponds to χ_{H,NH_i} values. As expected, the 19 $\chi_{H,\alpha\text{-}C_i}$ values are symmetrical about the peptide midpoint. For large w, as noted in the text, values of χ_{H,NH_i} at *sites 4–11* closely approximate the corresponding values of $\chi_{H,\alpha\text{-}C_i}$ at *sites 1–8*, and values of χ_{H,NH_i} at *sites 12–20* approximate $\chi_{H,\alpha\text{-}C_i}$ at *sites 11–19*.

$$\begin{bmatrix}
w & wt & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & wt & v & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w & \frac{v^2}{(wt)} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & wt & v & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
w & wt & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & wt & v & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w & \frac{v^2}{(wt)} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & wt & v & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix} \tag{30}$$

$$ss = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & p & 1 \end{bmatrix} \cdot \begin{bmatrix}
w & wt & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & wt & v & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
w & \frac{v^2}{(wt)} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & wt & \frac{v}{p} & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix} \cdot \begin{bmatrix}
w & wt & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & wt & v & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
w & \frac{v^2}{(wtp)} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & wt & v & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix} \cdot \begin{bmatrix}
w & wt & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & wt & v & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
w & \frac{v^2}{(wt)} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & wt & v & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}^{(n-3)} \cdot \begin{bmatrix}
w & wtc & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & wt & v & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
w & \frac{v^2}{(wt)} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & wt & v & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \tag{31}$$

As is evident from *Fig. 8*, for highly helical peptides, site-shifted values of χ_{H,NH_i} and $\chi_{H,\alpha\text{-}C_i}$ in the ascending and descending regions correlate closely. This correlation is rationalized through examination of the conformational weights that contribute to these two measures of site helicity. As shown above, *Eqn. 29*, which defines $\chi_{H,NH}$ at site i of the peptide sequence, is proportional to the summed weights of the list of conformations in the helical manifold that share h-site weights at the three sites $(i-1)$, $(i-2)$, and $(i-3)$. All of these conformations must appear in each of the corresponding lists that define α-C-atom site helicities at each of these sites. The list that defines χ_{H,NH_i} is, in fact, the logical intersection of the three lists that define $\chi_{H,\alpha\text{-}C_i}$ at each of the sites $(i-1)$, $(i-2)$, and $(i-3)$, and, as a result, the list of weighting terms that are summed in the numerator of *Eqn. 29* to calculate $\chi_{H,NH}$ are a corresponding logical intersection of the lists of weights for the $\chi_{H,\alpha\text{-}C_i}$ values[6]).

As expected for a logical intersection, the χ_{H,NH_i} list contains fewer conformations than any one of the associated three $\chi_{H,\alpha\text{-}C_i}$ lists, but it closely resembles the shortest of these. The missing conformations always contain unusually short helical regions. For large w values, their contributions to the state sum are small relative to that of average conformations, and, to a good approximation, the shortest of the $\chi_{H,\alpha\text{-}C_i}$ list equals the $\chi_{H,NH}$ list.

Inspection of the examples of *Fig. 6* demonstrates that, for i-values spanning the first half of the peptide sequence, each $\chi_{H,\alpha\text{-}C_i}$ list that corresponds to a larger i incorporates the conformations (or weights) of its predecessors; lists for central i-sites contain the largest numbers of conformations. Accordingly, for i-values spanning the second half of the peptide sequence, progressively more-complete $\chi_{H,\alpha\text{-}C_i}$ lists appear when the i-values are ranked in decreasing order. This pattern reflects the presence of substantial numbers of long, stabilized helical conformations that contribute to all site-helicities, except those close to the peptide termini. As a consequence, for $i < n/2$, χ_{H,NH_i} approximates $\chi_{H,\alpha\text{-}C_{(i-3)}}$; for $i > n/2$, χ_{H,NH_i} approximates $\chi_{H,\alpha\text{-}C_{(i-1)}}$. This i-mapping associates all but the two central values of $\chi_{H,\alpha\text{-}C_i}$ with experimental PF_i values. For large n and w, the pair of missing $\chi_{H,\alpha\text{-}C_i}$ that appear at the center of the sequence can be assigned with small error as equal to values for their neighbors.

Inspection of the LR-modeled site helicities of *Fig. 8* demonstrates the errors associated with this approximation, which, with the exception of $i=1$ or n, always slightly underestimates $\chi_{H,\alpha\text{-}C_i}$. For a medium-sized peptide and $w=1.3$, using χ_{H,NH_i} for the central values of $\chi_{H,\alpha\text{-}C_i}$ is seen to underestimate them by *ca.* 10%, but for the peptide with $w=1.5$, the error is clearly much smaller.

The most important experimental consequence of the modeling carried out in this section is the demonstration, evident from *Fig. 8*, that calculation of χ_{H,NH_i} from experimentally assigned values of PF_i provides an exceptionally powerful test of the degree of homogeneity of the conformational manifolds formed by α-peptide helices. The curvature seen in site dependent plots of χ_{H,NH_i} directly reflects the presence of

---

[6]) The pentapeptide of *Fig. 6* illustrates these assignments. The selections of h,c labels for assigning $\chi_{H,\alpha\text{-}C_i}$ conformation lists can be read from the $+/-$ selection lists of *Fig. 6,a*. For the first three sites, these are l(1) = [A, B, D, D′]; l(2) = [A, B, C, D, D′]; l(3) = [A, B, C, D, D′, F, F′]. Their logical intersection 1(1) ∩ 1(2) ∩ 1(3), which equals l(1), defines the conformational list for χ_{H,NH_i} at *site 4*, as confirmed by reading down the $+/-$ selection lists of *Fig. 6,b*. As noted in the legend for *Fig. 6*, for this pair of cases, $\chi_{H,\alpha\text{-}C_i} = \chi_{H,NHC_4}$.

partially helical conformations within the helical manifold. A manifold that is dominated by the completely helical conformation together with a few conformations with helical regions of nearly equal length yields site-dependent plots of χ_{H,NH_i} with a large central region in which χ_{H,NH_i} is nearly site-independent and approaches the limiting value of 1.0. The changes evident in *Fig. 8* that result when *w* is increased from 1.3 to 1.5 identify the range of *w* values required to approach this ideal form.

The next section addresses the effect of helix-stabilizing caps on the conformational homogeneity of a homopeptide characterized by *w* values in this range. The two independent factors of effective capping and consistently high helical propensities for the amino acid residues that make up an α-peptide sequence are jointly required to generate α-peptide helices that approach the stability and conformational homogeneity of helices formed by β-peptides.

6. Helix Stabilization by N- and C-caps. – This section describes results of LR modeling of helical manifolds formed by α-peptides that bear helix-stabilizing caps at each terminus. The helicities of peptides of moderate length are strongly enhanced by certain *N*- and *C*-capping functions. Citing evidence from the X-ray crystallographic data base for globular proteins, *Presta* and *Rose* noted sequence regularities within the nonhelical end regions of protein helices and proposed that helix start and stop signals rather than helical propensities may largely define the regions of helical structures within proteins [27]. Prior to this report and drawing upon proposals that aligned backbone amides create substantial local dipoles at the helix termini that are stabilized by neighboring charges [28], it was demonstrated that α-peptide helicity is enhanced when positive charges are placed near the peptide C-terminus or negative charges at the N-terminus [29]. However, charge is *not* an essential feature of a stabilizing cap [30]. Appropriately structured, uncharged α-peptide N-caps can give rise to helix stabilization comparable to that seen for negative charges; examples include the simple acetyl (Ac) function and the amino acid Gly, as well as amino acids bearing polar side chains, such as Asn, Ser, and Thr[7]). This report is primarily focused on changes of helicity that may result if appropriate amide-forming functions (helical templates) are linked at the ends of simple α-peptide sequences that alone are already substantially helical.

Along with peptide sequences of appropriate amino acid composition, caps of proven helix- enhancing efficacy are part of a tool kit for the molecular design of highly structured peptide sequences. It is unclear whether the upper limits of cap efficiency have been realized, in part because the range of mechanisms of cap stabilization are incompletely defined. It is also unclear whether the effectiveness of a cap is strongly

---

[7]) Ser, Thr, and Asn are also strong helix breakers when located *within* peptide sequences, and this effect has been attributed to intramolecular H-bonding between side-chain functions and backbone amides, which interferes with helix propagation. Depending on where they appear within a peptide sequence, these amino acids, thus, can either strongly stabilize or destabilize particular helical conformations within a manifold, and a first exploration of this point has recently been reported [31]. *Baldwin* and co-workers have proposed that two experimentally defined N- and C-capping parameters should be added to the helical propensity for each type of amino acid when applying LR modeling to heteropeptide sequences, and that existing data sets are sufficient to assign these 60 parameters [32]. An acute discussion of this proposal lies outside the scope of this report.

dependent on the local α-amino acid sequence at its attachment site. For selected cases, the large helix-enhancing effect of stabilizing caps is unequivocal. In H_2O, a solubilized, isolated $(Ala)_{12}$ sequence exhibits no helical CD features, but the same peptide, when appropriately capped, becomes exceptionally helical [33].

What is currently known about the mechanisms of helix stabilization by terminal caps? Abundant evidence confirms the effect of complementary charges that can stabilize terminal helix dipoles that are probably best attributed to the peripherial three-oriented amide residues that lack intrahelical H-bonds [34]. The mechanisms of helix stabilization by uncharged caps are less clear. Helix stabilization by N-terminal amino acids like Ser, Thr, and Asn may reflect formation of intramolecular H-bonds between polar side chains of these amino acids and one or more of the three amide NH functions that appear at this terminus. X-Ray structural evidence drawn from the protein data base has documented the formation of side-chain-to-backbone amide H-bonds by the above amino acid series [35].

Of the simple helix-stabilizing N-caps, acetyl and formyl rank among the best [36]. Why should such simple caps be strong helix stabilizers? The most plausible explanation appears to be a significant superiority in H-bond acceptor capacity of the O-atom of an acetamide relative to that of the corresponding O-atom of an α-aminoacyl amide, which may be attenuated by a competing and strongly stabilizing electrostatic interaction with its NH function. The strongest of the existing helix-stabilizing N-caps present both an oriented charge and a preorganized array of H-bonding acceptors that match the orientation of the NH H-bond donor sites that appear at the N-terminus of an α-helix [36].

A LR algorithm that has been tailored to model the effect of N- and C-caps on the helical stability of an α-peptide must introduce quantitative capping parameters p and c that are defined relative to a model, such as the same peptide that bears simple Ala residues at its N- and C-termini. The first such algorithm was introduced by *Doig et al.* and was based on a 4×4 matrix product [37]. In addition to applying the p-parameter to conformations with helical regions that extend to the peptide N-terminus, and the c-parameter to conformations with helical regions that extend to the C-terminus, this algorithm also includes p-weighted conformations like hhccc and hcchhhccc and c-weighted conformations like cchh and chhhhhcch. Solitary and paired h-sites at the N- or C-termini were also given capping weights. Mechanisms that justify this weighting are implausible, and subsequent publications have introduced algorithms that confine cap weighting to conformations with helical regions reaching either or both of the peptide termini [10].

The 8×8 matrix product of *Eqn. 20* is easily modified to accommodate capping weights that meet this restriction. The final w weights of all helical sequences that extend to the C-terminus are assigned by the (1,2) coefficient of the last matrix. All C-capping weights are, thus, correctly assigned if the value of the (1,2) coefficient of the final matrix in the product series is changed from *wt* to *wtc*, since only conformations with helical regions that extend to the peptide C-terminus are weighted by this coefficient of this matrix. Analogously, the p-weight for any conformations with helical regions extending to the peptide N-terminus is assigned by changing the value of the $v1$ vector at *column 7* from 1 to p. In the absence of other changes in the matrix product, peptide conformations with labels that begin with sequences hc or hhc are incorrectly

assigned *p* weights. The correct weights are assigned to these conformations only if the value of the coefficient (7,6) in the first matrix in the product series is changed from *v* to *v*/*p*, and the value of the coefficient (5,2) in the second matrix in the series from v^2/w to $v^2/(wp)$. *Eqn. 31* shows the corresponding algorithm, and *Eqn. 32* gives the state sum calculated for an *N*- and *C*-capped α-pentapeptide using this matrix product. Inspection shows that all conformations with helical sequences that extend to either or both of the caps are correctly assigned *p*- and *c*-weights, and these weights appear in no other state sum terms.

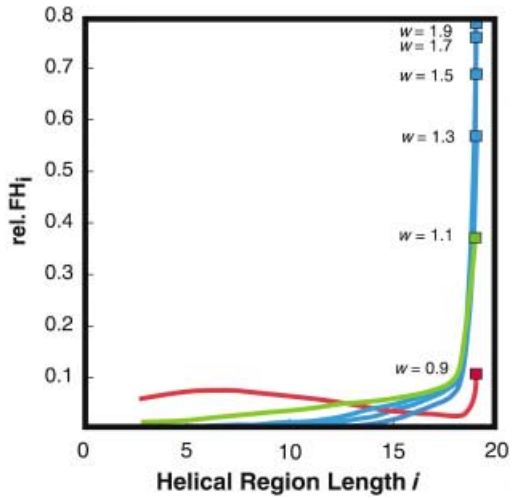

Fig. 9. Lifson–Roig *modeling of the relative contributions of long and short helical conformations to the helical manifold for a 19-residue peptide* ($v = t = 0.048$) *capped by N- and C-functions for which* $p = c = 10$, *equivalent to the stabilizing effect of an* N-*acetyl function*. The ordinate is the fraction of total FH that is contributed by helical conformations of lengths 3–19. With the exception of the strongly-destabilized peptide with $w = 0.9$, the introduction of caps dramatically increases the relative and absolute abundances of the completely helical conformations. These distributions should be compared with the corresponding examples of *Fig. 4*.

$$ss_{n,c}(5) = 1 + 5v + 10v^2 + 7v^3 + v^4 + t^2\,[w^3(1 + p + c) + w^3v(p + c) + w^4(p + c) + pcw^5] \quad (32)$$

In this state sum, both *p*- and *c*-weights are assigned to only one term, which corresponds to the completely helical conformation. This conformation is uniquely stabilized by both caps. Provided *p* and *c* are both large, such a double *pc* weighting provides the most effective mechanism for increasing the homogeneity of the conformational manifold formed by an α-peptide. The difference between cap-stabilized and simple helical α-peptides is best demonstrated by comparing *Fig. 4* with *Fig. 9*. For varying *w* in a 19-residue α-homopeptide, both graphs plot the modeled *relative* contributions to the FH by helical conformations of different length. For the uncapped examples of *Fig. 4*, the forms of the length distributions change dramatically as *w* is increased, but even for the largest value of *w*, no conformational length contributes more than 25% to the overall FH. All length distributions are broad, even

for very large values of *w*, reflecting a persistence of substantial length diversity within helical manifolds.

Very different length distributions are seen for the analogous capped examples of *Fig. 9*. For stabilized helices ($w > 1.0$), the *fully helical* conformation is the most abundant, contributing more than 50% to the FH for $w > 1.3$. All curves in *Fig. 9* display a similar shape, and the changes induced by an increase in *w* are primarily quantitative, not qualitative. For all *w*, the conformational homogeneity of these capped peptides has increased substantially from that of the uncapped peptides, and, for high *w*-values, the properties of the helical manifold can be approximated for many purposes by a single, completely helical conformation.

An alternative demonstration of the effect of strong helix-stabilizing caps on the structure of the helical manifold formed by an α-peptide is provided by the site helicities shown in *Fig. 10*, which include the results for the corresponding simple helical peptide shown in *Fig. 8,a*. For the 19-residue peptide and $w = 1.3$, addition of caps increases FH from 0.50 to 0.84. The maximal site helicity at the central residue is raised by 35%, and terminal site helicities are increased more than fourfold. The forms of the plot show striking differences that reflect the increased conformational homogeneity of the capped peptide. Its site helicities in the central region of the plot are nearly constant, implying that only relatively long helical sequences contribute to its manifold.

If *w* for the capped peptide is increased to 1.5, the overall FH increases to 0.93, and the site helicities lie above 0.9 in the region between *sites 3* and *17*, suggesting that addition of efficient helix-stabilizing *N*- and *C*-terminal caps to a highly helical α-peptide of medium length (comprised of helix-stabilizing residues with *w* values greater than 1.3) ensures its conformational homogeneity. To an excellent approximation, the conformational manifold of such a peptide can be replaced by a single,

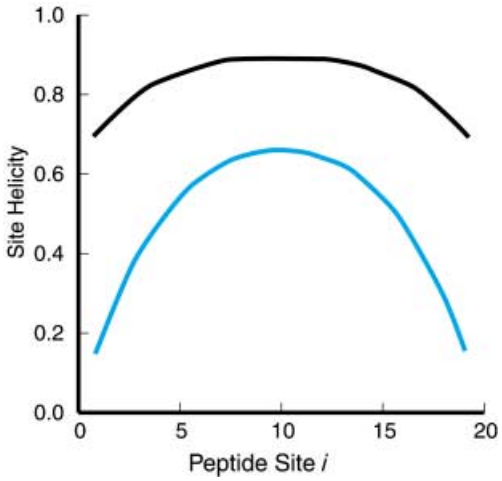

Fig. 10. Lifson–Roig *modeling of site helicities for a 19-residue peptide* ($v = t = 0.048$, $w = 1.3$) *with* (black curve) *and without* (cyan curve) N- *and* C-*capping residues for which* $p = c = 10$. By criteria of high central site helicity and central curve flatness, the addition of caps substantially reduces the conformational inhomogeneity of the helical manifold.

completely helical conformation. If the LR modeling assumptions are valid, α- and β-peptides should exhibit similar conformational integrity.

Although the capping values p and c of *Figs. 9* and *10* were chosen to generate dramatic graphs, they are not exceptional. Values approaching $p = 10$, the value used in this modeling exercise, have been assigned from experimental CD changes seen for Ac functions and simple, negatively charged and uncharged N-caps that provide a single H-bonding site proximate to the helix N-terminus. More complex N-caps that provide multiple, rigidly oriented H-bonding sites can lead to substantially more-efficient helix stabilization, with p-values that lie in the range of 30–100 [36]. Smaller effects have been demonstrated for C-caps, and almost all reported significant c-parameters result from C-terminal placement of positive charges [37][38], which usually gives rise to c-values in the range of 2–10. C-terminal helix-stabilizing caps that provide more than one optimally oriented H-bond donor are likely to increase c, but owing to the structural differences between the oriented amide functions at N- and C-termini, the design and synthesis of such templates pose challenging problems. They have yet to be realized.

Apart from insights they may offer concerning the energetics and intrinsic features of helix formation, maximally efficient caps are likely to find important practical applications. Selective stabilization of the completely helical peptide conformation depends on the availability of widely applicable N- and C-caps that exert large and nearly equal helix stabilizing effects on the linked peptide sequence. Developing such scaffolds is an important research goal.

7. Scope and Limitations of *Lifson–Roig* Algorithms: Alanine α-Homopeptides. – The *Lifson–Roig* [2], *Zimm–Bragg* [39], and related algorithms [40] were developed more than 40 years ago to model helix/coil equilibria of charged α-oligopeptides 20 to several hundred amino acids long. Recently, they have been nearly universally employed to model helicities of α-homo- and heteropeptides of lengths in the range of 10–50 residues. These algorithms are imperfect bookkeeping tools for the study of most α-heteropeptides. Unless LR algorithms are modified to include specific, independently quantified, context-dependent effects, they are 'sequence blind', and this is an intrinsic limitation of the matrix product used to construct them. Moreover, they incorporate a model for the α-peptide unstructured state that is almost certainly inappropriate for oligomers constructed from amino acids that bear charged as well as hydrophobic side chains [3][6]. Relevant to this point, a recent study has critically explored the scope and accuracy of the *Flory* hypothesis that underlies LR models for the unstructured state [41]. It is likely that a diverse class of algorithms for modeling and predicting α-peptide helicity from amino acid composition and sequence will be required to optimally address their heteropeptide applications, but there is one role for which tailored LR algorithms are almost certainly ideally suited. For polyalanines, a recent comparison of LR and similar models with a more-rigorous all-atom molecular-modeling simulation showed that, although quantitative differences were noted, fundamental helicity properties were in qualitative agreement [42]. Despite much effort, a rigorous understanding at a molecular level of the factors that govern α-peptide helicity is still lacking. Key missing elements of that understanding must be developed by focusing the full range of quantitative tools on the simplest and most versatile of the available helical α-peptide substrates. These are the spaced, solubilized polyalanines.

After briefly addressing the underlying important issues and the potential scope of polyalanine studies, this section closes by suggesting alternatives to the LR formalism that might better address the requirements of helicity modeling for α-heteropeptides.

Structurally, alanine is the simplest of the strongly helix-stabilizing proteinogenic α-amino acids, and α-polyalanines are the simplest hosts for exploring the helicity changes that result from the introduction of guest amino acids into their sequences [25]. LR Modeling of such changes is expected to permit assignment of helical propensities as well as context- and site-dependent effects on helicity that result from single or multiple guest substitutions. A preliminary to rigorous modeling of this type requires full characterization of the factors that define helicity for the α-polyalanine hosts. New findings imply that major aspects of that characterization are incomplete.

From studies of the properties of *N*-capped short α-polyalanine sequences and a large length series of spaced solubilized polyalanines, *Kemp* and co-workers have recently questioned the implicit LR assumption that the stability of an α-helical region is a linear function of its length [17][25]. Small values of w were required to model short helical conformations, and positive H-bonding cooperativity [43] that selectively stabilizes longer helical conformations was suggested as an explanation. Measurements of PF values for selected members of the spaced, solubilized α-polyalanine series have been initiated and provide independent support for the hypothesis of *a significant increase* of the value of w with the length of a helical region [44]. If this assumption can be quantitatively confirmed for α-polyalanines and validated as significant in magnitude for more general α-peptide sequences, this effect is expected to redefine our heuristic models for α-peptide helicity.

The interpretational complexities introduced by a hypothesis of a significant length dependence for helical propensities underlines the need to model helicity, at least initially, through studies of simple α-homopeptides that lack charged side chains. The existing LR models have the versatility to provide bookkeeping tools for these studies. As noted elsewhere [17][25], software for manipulation of symbolically-expressed polynomials allows each w^i-weighted term of a LR state sum, *e.g.*, that of *Eqn. 32*, to be replaced by terms that correspond to new *length-dependent* helical weights. In effect, the conventional LR state sum plays the role of a computational template that can be modified to yield more versatile expressions.

The complementarity between deficiencies of LR algorithms and compensating features of helical α-polyalanines deserves brief comment. The sequence-blind feature of LR algorithms is irrelevant, and the conformational manifold of these homopeptides has the simplifying feature that the properties of conformations with helical regions of a particular length are almost certainly similar[8]). Since the unstructured states of α-polyalanines lack the unpredictable hydrophobic interactions expected for many heteropeptides, their stabilities can probably be approximated satisfactorily by the v polynomial within a LR state sum.

The relative simplicity of the α-polyalanines also provides, at least potentially, an experimental approach for assigning helicity properties to individual conformations within a complex helical manifold. When a homopeptide sequence of length n is extended by one residue, close analogs of all conformations present in its manifold are

---

[8]) The N- and C-terminal sequence sites may be exceptions, as noted in recent reports [45].

retained, and $(n-1)$ new conformers are added to generate a new manifold. Only one of these, that of length $(n+1)$, lacks a representative in the manifold for the *n*-peptide. The presence of a strongly stabilizing *N*- or *C*-cap singles out and stabilizes a small subset of conformations within each manifold. In principle, the length-dependent properties of the individual conformers are assignable from length-dependent changes in the experimental properties of the manifold itself. *Kemp* and co-workers have recently applied this recursion tactic to assign length-dependent *w*-values for Ala residues in certain *N*-capped α-polyalanines containing 3–14 residues [17]. The full potential of the recursion tactic would be realized if the helicity properties of *N*- and *C*-capped α-peptides that approach conformational homogeneity could be used as independent calibrations of the properties assigned by recursion to their simple helical analogs that lack strong cap stabilization. An essential first step is a demonstration that helicities of cap-stabilized peptides result from independent contributions of cap *and* peptide.

One key example illustrates the potential of these studies. Currently, the reliability of calculations of FH from CD ellipticities is in doubt [15], owing to a lack of rigorous calibration standards. Measurement of ellipticities for the members of a *N*- and *C*-cap-stabilized homologous series of helical polyalanines may provide that calibration. A linear regression of $[\theta]_{H222,n\,molar}$ *versus* polyalanine length *n* according to *Eqn. 24* will yield the key parameters *X* and $[\theta]_{H222,\infty}$ of *Eqn. 23*. These are the missing parameters for rigorous LR calculations of $[\theta]_{222,Calc}$[9]). Extension of these results to other α-homo- and heteropeptides will allow experimental assessment of the errors implicit in the fundamental assumption of all CD-based FH calculations, assuming that the calibration constants $[\theta]_{H222,n}$ reflect only the backbone conformation of an α-peptide and are insensitive to its amino acid composition.

One can envisage alternative bookkeeping tools that lack the sequence-blind features of the LR algorithm and that are computationally more efficient when applied to modeling of helicities of heteropeptides by employing external and internal capping corrections, amino acid dependent helical propensities, and context dependencies. For α-peptides in the commonly encountered length range, the helical portion of the state sum can be modeled explicitly, without recourse to a matrix-based algorithm. Attractive features of alternatives to LR computational models would be a retention of the amino acid sequence in a computationally useful form, and a computational search capacity that could identify unexpected sequence-dependent context effects within an extensive data base comprised of helical sequences and their experimental helicity parameters. Replacing the *v* polynomials, which model the unstructured states of α-peptides within LR state sums, with more suitable approximations awaits energetic characterization, through experiment or theory, of the sequence- and amino acid dependent hydrophobic interactions that are believed to stabilize and reduce the conformational complexity of these states. Until such characterization is in hand, the existence of these stabilizing effects introduces an uncertainty into any algorithmic calculation of the helicity of heteropeptide sequences.

8. Conclusions. – Unlike helical β-peptides, helical α-peptides are conformationally inhomogeneous. A fully helical conformation formed by an α-peptide in H_2O is in rapid

---

[9]) Alternatively, it may invalidate *Eqns. 18* and *19*, requiring use of an empirically-derived set of calibration.

equilibrium with comparable 'concentrations' of partially helical conformers that, collectively, form a helical manifold. Conformational analysis of this complex system has primarily relied on *Lifson–Roig* (LR) state sums based on a series of model assumptions.

Conventional LR algorithms have been analyzed and modified in this report, then used to generate quantitative models for the properties expected for the conformational manifolds formed by helical α-peptides. The relative contributions of long and short helical conformers to the overall fractional helicity (FH) of an α-homopeptide are easily modeled as functions of the overall peptide length n and the helical propensity w of its amino acid residues. The site dependence of site helicities, which are directly related to experimental values of ^{1}H-NMR-derived protection factors (PF), provides an independent measure of the degree of conformational homogeneity within a helical manifold. In striking contrast to the conformationally homogeneous helices formed even by short β-peptides, as modeled by conventional LR algorithms, the helices formed by α-peptides approach conformational homogeneity only if the peptides are long, if all amino acids possess unusually high helical propensities, and/or if *N*- and *C*-capped by strongly helix-stabilizing functions.

A key LR model assumption is a linear length dependence of the stabilities of α-peptide helices. A recent report [17] frames the hypothesis that this assumption is incorrect: helical stabilities seem to be characterized by a *nonlinear*, large length dependence, with the implication that the conformational homogeneity of peptides containing more than 20 residues is grossly underestimated by conventional LR models. Under this hypothesis, shorter peptides retain the strong conformational inhomogeneity implied by these models.

A possible explanation for the striking differences exhibited by short helical β- and α-peptides is a relative reduction in the H-bonding capacity of the amide functions of α-peptides attributable to the strong electrostatic stabilization of these functions arising within α-peptide coiled-state conformations that allow opposing and proximate NH and CO dipoles of each peptide residue to interact. Such a stabilization in α-peptides should reduce the magnitudes of both helical propensities and *N*- and *C*-capping effects attributable to terminal α-amino acid functions.

Experimental Part

All modeling calculations were executed in *Mathematica 3.0*, *Wolfram Research*, *Inc.* Symbolic expression polynomials were rendered in easily interpretable form by use of the *Mathematica Expand* function, which was applied to calculated symbolic state sums to generate expressions similar to those of *Eqns. 21* and *32*. When applied to a polynomial, the *Mathematica* function coefficient [*p*,*x*,*i*] yields the coefficient for the term x^i within the polynomial, and the product $x^i \cdot$ coefficient [*p*,*x*,*i*] extracts the entire x^i from the polynomial. As a result, for a helical α-peptide, the mole fraction of conformations with helical regions of length i is given by $1/(n \cdot \mathrm{ss}) \cdot i \cdot x^i \cdot$ coefficient [*ss*,*x*,*i*]. Applications and extensions of this example allow generation of the modeling data cited in this report.

The following stepwise protocol was followed to tailor LR algorithms. The starting point is a set of full list of conformational labels for peptides of appropriate lengths. Each h,c conformational label in each list is then associated with a desired weighting scheme. (This weighting scheme may introduce general features into the algorithm, such as the distinction between *t*- and *v*-weights or the introduction of N- and C-terminal capping parameters. It also might introduce special weighting features that apply to a particular residue *k* in the peptide amino acid sequence. In this case, the weighting change must be applied only to the *k*-matrix and perhaps to its neighbors). Usually this list is constructed by making slight modifications in a standard LR weighting scheme.

Partial sequences, triads, tetrads, *etc.*, are then selected that allow the weighting distinctions to be made. The length of these partial sequences determines the minimal size of the LR matrices used in the algorithm. Thus, if tetrads are needed, then the 8×8 matrix algorithm is appropriate. Selected pairs of coefficients in the matrix are identified that implement the desired weighting, and a new algorithm is constructed. This algorithm is then tested for its capacity to apply the desired weights to each of the conformational labels for an appropriate range of peptides of lengths n. The process of associating state-sum terms with conformational labels is greatly facilitated by the use of matrices expressed in the w, v, t formalism of *Eqn. 20*.

REFERENCES

[1] D. Seebach, J. L. Matthews, *Chem. Commun*. **1997**, *21*, 2015; D. Seebach, S. Abele, K. Gademann, G. Guichard, T. Hintermann, B. Jaun, J. L. Matthews, J. V. Schreiber, L. Oberer, U. Hommel, H. Widmer, *Helv. Chim. Acta.* **1998**, *81*, 932.
[2] S. Lifson, A. Roig, *J. Chem. Phys.* **1961**, *34*, 1963.
[3] K. A. Dill, D. Shortle, *Annu. Rev. Biochem.* **1995**, *60*, 795.
[4] K. S. Thompson, C. R. Vinson, E. Freire, *Biochemistry* **1993**, *32*, 5491.
[5] P. J. Flory, 'Principles of Polymer Chemistry', Cornell University Press, New York, 1953; H. S. Chan, K. A. Dill, *Annu. Rev. Biophys. Chem.* **1991**, *20,* 447; R. W. Woody, *Adv. Biophys. Chem.* **1992**, *2*, 37.
[6] K. A. Dill, *Biochemistry* **1985**, *24*, 1501.
[7] A. Chakrabartty, R. L. Baldwin, *Adv. Protein Chem.* **1995**, *46*, 141.
[8] C. A. Rohl, R. L. Baldwin, *Methods Enzymol.* **1998**, *295*, 1; N. R. Kallenbach, E. J. Spek, *Methods Enzymol.* **1998**, *295*, 26.
[9] L. Serrano, *Adv. Protein Chem.* **2000**, *53,* 49; V. Villegas, A. R., Vigurera, F. X. Aviles, L. Serrano, *Fold. Des.* **1996**, *1*, 29.
[10] J. K. Sun, A. J. Doig, *Protein Sci.* **1998**, *7*, 2374; N. H. Andersen, H. Tong, *Protein Sci*. **1997**, *6*, 1920.
[11] Z. I. Hodes, G. Nemethy, H. A. Scheraga, *Biopolymers* **1979**, *18*, 1565; B. M. Petitt, M. Karplus, *Chem. Phys. Lett.* **1985**, *121*, 194.
[12] C. A. Rohl, J. M. Scholtz, E. J. York, J. M. Stewart, R. L. Baldwin, *Biochemistry* **1992**, *31*, 1263.
[13] M. Bixon, S. Lifson, *Biopolymers* **1967**, *3*, 509.
[14] R. W. Woody, *Methods Enzymol.* **1995**, *246*, 34.
[15] P. Wallimann, R. J. Kennedy, D. S. Kemp, *Angew. Chem., Int. Ed.* **1999**, *38*, 1290; P. Wallimann, R. J. Kennedy, W. Shalongo, D. S. Kemp, *J. Am. Chem. Soc.*, in press.
[16] D. S. Kemp, T. J. Allen, S. Oslick, J. G. Boyd, *J. Am. Chem. Soc.* **1996**, *118*, 4240.
[17] R. J. Kennedy, K.-W. Tsang, D. S. Kemp, *J. Am. Chem. Soc.* **2002**, *124*, 934.
[18] P. J. Gans, P. C. Lyu, M. C. Manning, R. W. Woody, N. R. Kallenbach, *Biopolymers* **1991**, *31*, 1605.
[19] N. A. Besley, J. D. Hirst, *J. Am. Chem. Soc.* **1999**, *121*, 9636.
[20] P. Luo, R. L. Baldwin, *Biochemistry* **1997**, *36*, 8413.
[21] C. L. Perrin, T. J. Dwyer, J. Rebek Jr., R. J. Duff, *J. Am. Chem. Soc.* **1990**, *112*, 3122.
[22] Y. Bai, J. S. Milne, L. Mayne, S. W. Englander, *Proteins: Struct., Funct., Genet.* **1993**, *17*, 75.
[23] C. A. Rohl, R. L. Baldwin, *Biochemistry* **1997**, *36,* 8435.
[24] H. X. Zhou, L. A. Hall, N. R. Kallenbach, *J. Am. Chem. Soc.* **1994**, *116*, 6482.
[25] J. S. Miller, R. J. Kennedy, D. S. Kemp, *J. Am. Chem. Soc.* **2002**, *124*, 945.
[26] W. Shalongo, D. Laxmichand, E. Stellwagen, *J. Am. Chem. Soc.* **1994**, *116*, 2500.
[27] L. G. Presta, G. D. Rose, *Science* **1988,** *240*, 1632; R. Aurora, G. D. Rose, *Protein Sci.* **1988**, *7*, 21.
[28] W. G. H. Hol, P. T. van Duijnen, H. J. C. Berendsen, *Nature* **1978**, *273*, 443.
[29] J. P. Schneider, W. F. DeGrado, *J. Am. Chem. Soc.* **1998**, *120*, 2764; G. Esposito, B. Danapal, P. Dumy, V. Varma, M. Mutter, G. Bodenhausen, *Biopolymers* **1997**, *41*, 27.
[30] B. Forood, E. J. Feliciano, K. P. Nambiar, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 838; B. Forood, H. K. Reddy, K. P. Nambiar, *J. Am. Chem. Soc*. **1994**, *116*, 6935; A. Chakrabartty, T. Kortemme, R. L. Baldwin, *Protein Sci.* **1994**, *3*, 843.
[31] M. Petukhov, K. Uegaki, N. Yumoto, S. Yoshikawa, L. Serrano, *Protein Sci.* **1999**, *8*, 2144.
[32] A. J. Doig, R. L. Baldwin, *Protein Sci.* **1995**, *4*, 1325.
[33] J. S. Miller, R. J. Kennedy, D. S. Kemp, *Biochemistry* **2001**, *40*, 305.
[34] B. Tidor, *Proteins: Struct., Funct., Genet.* **1994**, *19*, 310.
[35] M. Levitt, *Biochemistry* **1978**, *17*, 4277.

[36] W. Maison, E. Arce, P. Renold, R. J. Kennedy, D. S. Kemp, *J. Am. Chem. Soc.* **2001**, *123*, 10245; W. Maison, R. J. Kennedy, D. S. Kemp, *Tetrahedron Lett*. **2001**, *42*, 4975.
[37] A. J. Doig, A. Chakrabartty, T. M. Klingler, R. L. Baldwin, *Biochemistry* **1994**, *33*, 3396.
[38] S. Deechongkit, R. J. Kennedy, K.-Y. Tsang, P. Renold, D. S. Kemp, *Tetrahedron Lett.* **2000**, *41*, 9679.
[39] B. H. Zimm, J. K. Bragg, *J. Chem. Phys.* **1959**, *31*, 526.
[40] H. Qian, J. A. Schellman, *J. Phys. Chem.* **1992**, *96*, 3987; D. Poland, H. A. Scheraga, 'Theory of Helix-Coil Transitions', Academic Press, New York, 1970; H. A. Saroff, J. E. Kiefer, *Biopolymers* **1999**, *49*, 425.
[41] R. V. Pappu, R. Srinivasan, G. D. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12565.
[42] M. Takano, K. Nagayama, A.Suyama, *J. Chem. Phys.* **2002**, *116*, 2210.
[43] H. Guo, M. Karplus, *J. Phys. Chem.* **1994**, *98*, 7104.
[44] R. J. Kennedy, D. S. Kemp, unpublished results.
[45] S. Penel, E. Hughes, A. J. Doig, *J. Mol. Biol.* **1999**, *287*, 127; I. A. Sun, S. Penel, A. J. Doig, *Protein Sci.* **2000**, *9*, 730; D. A. E. Cochran, S. Penel, A. J. Doig, *Protein Sci*. **2001**, *10*, 463.

Received August 5, 2002