The Structure and Energetics of Helix Formation by Short Templated Peptides in Aqueous Solution. 3. Calculation of the Helical Propagation Constant *s* from the Template Stability Constants t/c for Ac-Hel$_1$-Ala$_n$-OH, $n = 1-6$

D. S. Kemp,* Sherri L. Oslick, and Thomas J. Allen

Contribution from the Department of Chemistry, Room 18-582, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received August 22, 1995[⊗]

Abstract: From t/c values measured at 25 °C for the series Ac-Hel$_1$-A$_n$-OH, $n = 1-6$ in D_2O the helical propagation constant (*s* value) for alanine in this templated peptide series is 1.03; the corresponding value in 10 mol % trifluoroethanol−D_2O is 1.47. An iterative least squares analysis for calculating *s* values from t/c ratios is developed, errors implicit in its application are discussed, and the *s* value is related to the degree of helical fraying.

Previous reports[1,2] analyzed the detailed energies of the reporting conformational template Ac-Hel$_1$ and showed that in water and in water−trifluoroethanol (TFE) mixtures the conjugate Ac-Hel$_1$ A$_6$-OH forms an α-helix. As depicted in the first figure of the accompanying paper (*J. Am. Chem. Soc.* **1996**; *118*, 4240−4248) this α-helix is induced only by the te conformational state of the template but not by the alternative cs or ts states, and the cs, ts, and te mol fractions are defined by the experimentally determined t/c ratio, as seen in Figure 1. (The c-t designators refer to the s-cis/s-trans orientations of the acetamido function, and the s-e designators refer to the staggered/eclipsed orientation of the C-8−C-9 single bond.)

In this paper we analyze the relationship between the experimentally measurable template state ratio t/c and the helical stability of the te state, expressed as the average *s* value or propagating tendency of a single alanine residue. A rigorous separation of helical propagation effects from initiation has hitherto been possible only for helices formed from oligopeptides.[3]

An s value is an equilibrium constant expressing the tendency of a single amino acid residue to join a pre-existing helical segment to which it is linked by a peptide bond.[4] As normally calculated, *s* values embody that part of the helical disposition of a particular amino acid residue that is intrinsic to its local structure and independent of its neighboring residues or of its position along a peptide sequence. Since helicity can depend on amino acid sequence, a calculation of helicity based solely on *s* values provides only a crude approximation to the actual helical stability of a given heteropeptide, a point that was stressed when *s* values for oligopeptides were first reported[3] and subsequently.[5] A better approximation could be achieved

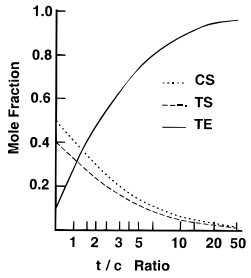

Figure 1. Mol fractions of template cs, ts, and te states as functions of the experimentally determined t/c ratio. $\chi_{cs} = 1/(1 + t/c)$; $\chi_{ts} = 0.79\chi_{cs}$; $\chi_{te} = 1 - \chi_{cs} - \chi_{ts}$.

by the addition of correction terms that reflect interactions between side-chain charge and helix dipole, between residue neighbors on the peptide sequence, between next nearest neighbors, etc. Although a protocol for experimental characterization of these interactions becomes intractable if most correction terms are energetically significant, a minimalist scenario can be envisaged in which position-dependent stabilization effects arise primarily for the charge-dipole interactions and neighboring interactions primarily involve van der Waals or polar contacts of the (i,i+3) and (i,i+4) types. Estimates of the energetics of several of these interactions have been reported.[6,7] The simplicity of the substate structure of a short helix initiated only at the N-terminus in the presence of a local reporter function may allow an exceptionally precise characterization of these effects, and this is the long-range aim of our studies of Ac-Hel$_1$-peptide conjugates.

We have chosen the series Ac-Hel$_1$-A$_n$-OH, $n = 1-6$, for the first study of the energetics of peptide propagation since it allows quantitation of the stabilizing effects of alanine residues

[⊗] Abstract published in *Advance ACS Abstracts,* April 15, 1996.

(1) Kemp, D. S.; Allen, T. J.; Oslick, S. L. *J. Am. Chem. Soc.* **1995**, *117*, 6641−6657. Kemp. D. S.; Allen, T. J.; Oslick, S. L.; Boyd, J. G. *J. Am. Chem. Soc.* **1996**, *118*, xxx.

(2) Kemp, D. S.; Boyd, J. G.; Muendel, C. C. *Nature* **1991**, *352*, 451−454.

(3) von Dreele, P. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 396−407. von Dreele, P. H.; Lotan, N.; Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 408−417. Wojcik, J.; Altmann, K.-H.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 121−134.

(4) Poland, D.; Scheraga, H. A. Statistical Mechanical Theory of Order-Disorder Transitions in Biological Macromolecules; Academic Press: New York, 1970.

(5) Wojcik, J.; Kidera, A.; Leed, A. R.; Nakajima, A.; Scheraga, H. A. *Macromolecules* **1990**, *23*, 3655−3662.

(6) For previous analysis of helical sequence dependences, see: Finkelstein, A. V.; Badretdinov, A. Y.; Ptitsyn, O. B. *Proteins, Struct., Funct., Genetics* **1991**, *10*, 287−299. Horovitz, A.; Matthews, J. M.; Fersht, A. R. *J. Mol. Biol.* **1992**, *227*, 560−568.

(7) Armstrong, K. M.; Fairman, R.; Baldwin, R. L. *J. Mol. Biol.* **1992**, *228*, 284−291. Schoemaker, K. R.; Fairman, R.; Schultz, D. A.; Robertson, A. D.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1990**, *30*, 1−11. Sancho, J.; Serrano, L.; Fersht, A. R. *Biochemistry* **1992**, *31*, 2253−2258. Bell, J. A.; Becktel, W. J.; Sauer, U.; Base, W. A.; Matthews, B. W. *Biochemistry* **1992**, *31*, 3590−3596. Horovitz, A.; Matthews, J. M.; Fersht, A. R. *J. Mol. Biol.* **1992**, *227*, 560−568.

S0002-7863(95)02911-8 CCC: $12.00

in isolation and in the structural context of an exceptionally short helix. This study is a prelude to a more extensive examination of larger templated alanine oligomers containing solubilizing residues. A preliminary report of the properties alanine-lysine conjugates has appeared.[8]

Helical Propagation Constants, *s* Values, from the Length Dependence of t/c Ratios Measured for Template-Conjugates of Amino Acid Oligomers

$$\mathrm{cs} \overset{K_1}{\rightleftharpoons} \mathrm{ts} \overset{K_2}{\rightleftharpoons} \mathrm{te}_0 \overset{K_3}{\rightleftharpoons} \mathrm{te}_1 \overset{K_4}{\rightleftharpoons} \mathrm{te}_2 \overset{K_5}{\rightleftharpoons} \quad (1)$$

$$\frac{[\mathrm{t}]}{[\mathrm{c}]} = \frac{[\mathrm{ts}] + \sum[\mathrm{te}_i]}{[\mathrm{cs}]} = K_1\,(1 + K_2\,(1 + K_3\,(1 + K_4\,(1 + K_5\,(1 + \ldots))))) \quad (2)$$

$$\frac{[\mathrm{t}]}{[\mathrm{c}]} = (K_1\,(1 + K_2)) + K_1K_2K_3\,(1 + K_4 + K_4K_5 + K_4K_5K_6 + \ldots) \quad (3)$$

$$\frac{[\mathrm{t}]}{[\mathrm{c}]} = A + B\,(1 + K_4 + K_4K_5 + K_4K_5K_6 + \ldots) \quad (4)$$

where $A = (K_1(1 + K_2))$ and $B = K_1K_2K_3$.

The mass action expressions pertinent to a templated peptide with a helical te state are shown in eqs 1−4. Definition of a mass action expression for an Ac-Hel$_1$-peptide conjugate starts with selection of a series of nonoverlapping states that blanket the entire manifold of conformations of the conjugate. A state is usually defined as a set of conformations that share conceptual or experimental features. Any system comprising n such states implicitly allows choice of $n!/(2!(n-2)!)$ fundamental equilibrium constants, each linking a pair of states. A mass action expression then contains $(n-1)$ of these fundamental equilibrium constants, so chosen that all n states are linked; it is defined by the numerical values of the $(n-1)K_i$ and by the branched or linear connectivities between the states.

Mass action expressions defined by linear connectivities can always be expressed as a nested series of parenthetic terms as in (2) or explicitly as a sum of products of equilibrium constants, as in (3) and (4). This linear choice for a helical peptide follows convention and is particularly convenient because of the relationship of the product sums of (3) and (4) to a simple geometric series. Provided that the listed states are inclusive and defined without overlap, expressions 1−4 are fully general, depending only on the assumption that equilibrium has been established.

A template with three conformations linked to a peptide with k conformations is expected to generate a conjugate with $3k$ conformations. For the peptide conjugates of Ac-Hel$_1$, the readily identifiable ^{1}H NMR signatures of the three cs, ts, and te template conformations allow definition of three distinguishable cs, ts, and te states, each containing a series of peptide conformations. However, within either the cs or the ts states, we group all peptide conformations together, reflecting an inability to detect nonrandom structure in these two peptide populations.[1]

The detectable structure of the te state of the template conjugate of a peptide containing n backbone amide residues allows it to be defined more explicitly. Nearly all computational models for short helices[4] are based on the assumption that interruptions or breaks in a helix sequence terminate it. Since initiation is an improbable event, the probable helical conformations of short peptides must contain a single sequence of continuously hydrogen bonded amide residues. Because the peptide-template conjugates we study are too short to permit independent initiation of helices within the peptide core, their helical sequences must begin at the peptide-template interface and extend outward from it. Evidence for this feature of the model is provided by the disruption of helicity seen for conjugates that contain lactate or sarcosine residues.[1]

The te state thus can be defined as a set of $n+1$ substates, each characterized by the number of helical amide−amide hydrogen bonds that it contains. The structure of the state sums of (3) and (4) reflects the conformational substructure of the te state. Each term on the right of (3) and (4) corresponds to one of the probable template-initiated helical conformations, and the magnitude of the product term

$$K_1K_2\prod_1^n K_{(i+2)}$$

reflects its relative abundance.

The subscript i in the te$_i$ substate label of (1) specifies the number of helical hydrogen bonds within the substate. The first substate te$_0$ is unique in lacking hydrogen bonds and must be present in any Ac-Hel$_1$ derivative, but from failure to detect te character in derivatives like Ac-Hel$_1$-OH, we conclude that the te$_0$ state is significantly less stable than the ts, which is the alternative contributor to the t state for unstructured derivatives of Ac-Hel$_1$. Like the cs and ts states, the te$_0$ substate of a peptide conjugate must contain a set of peptide conformations that lack specific stabilizing interactions with the template, and the conformational composition of these three sets of peptide manifolds is expected to be very similar.

The experimental constant $K_{1'}$ must rigorously equal ([ts] + [te$_0$])/[cs], or by the definitions of (1), $K_{1'} = K_1(1 + K_2)$. The fractional te character in the t state of a simple derivative like Ac-Hel$_1$-OH that lacks te substate structure is then $K_1K_2/(K_1(1 + K_2)) = K_2/(1 + K_2)$. Since te state character could not be detected in the t state under these conditions, this ratio must be small, and K_2 must accordingly be considerably smaller than 1. As a result, $K_{1'} = K_1(1 + K_2)$ is satisfactorily approximated by K_1, consistent with the usage of our previous papers, where $K_{1'}$ was shown to be invariant to local structural changes.[1]

The composite equilibrium constant $K_1K_2K_3$ describes the formation from the cs state of the te$_1$ state, containing a single hydrogen bond between the template and the first amino acid residue. The presence of substantial te character in t states of Ac-Hel$_1$-G-OH and Ac-Hel$_1$-NHMe implies that $K_1K_2K_3 \gg K_1K_2$, and the constant K_3 that describes the te$_0$ − te$_1$ equilibrium therefore must be significantly greater than 1. A large K_3 compensates for a small K_2. We have previously suggested that the hydrophobic environment of the hydrogen bond at the template-peptide junction may contribute to the anomalously large magnitude of K_3.[1] Additional terms must be added to the te state sum as the peptide chain is extended; each new term corresponds to a new, longer helical conformation. Thus the te state sum of Ac-Hel$_1$-A-A-OH is $(K_1K_2K_3 + K_1K_2K_3K_4) = K_1K_2K_3(1 + K_4)$.

A template-initiated helix with length sufficient to develop well-defined structure at the template-helix junction, at the C-terminus, and in the intervening core region has the property that extension by addition of new residues merely increases the length of the core region without changing the ends. For a

(8) Groebke, K.; Renold, P.; Tsang, K.-Y.; Allen, T. J.; McClure, K. F.; Kemp, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, in press.

sufficiently long helix the equilibrium constants K_i of (2)−(4) that describe the energetic effect of adding one additional amino acid residue to the pre-existing helix must therefore converge to a constant value characteristic of the core region, which we define as s in this context. If K_i for residues near the template junction and at the C-terminus are similar in magnitude to those in the core, (3) for this idealized case reduces to (5) which contains a geometric series in s. The additive constant A and multiplicative constant B are then simple functions of the template junction equilibrium constants. In the limiting case that effective energetic convergence of the helical propagation constants K_i occurs as early as the second helical amino acid residue, and A and B assume very simple forms: $A = K_1(1 + K_2) \approx K_1$, and $B = K_1K_2K_3$. The additive constant A thus reflects the intrinsic t-c disposition of the template in the absence of any helical initiation, and the multiplicative constant B reflects the intrinsic tendency of the template to initiate a helix in the linked peptide.

$$\frac{[t]}{[c]} = A + B\,(1 + s + s^2 + s^3 + ...) \tag{5}$$

$$\frac{[t]}{[c]} = A + B(0) \quad n = 0 \tag{6a}$$

$$\frac{[t]}{[c]} = A + B(1) \quad n = 1 \tag{6b}$$

$$\frac{[t]}{[c]} = A + B(1 + s) \quad n = 2 \tag{6c}$$

$$\frac{[t]}{[c]} = A + B(1 + s + s^2) \quad n = 3 \tag{6d}$$

$$\frac{[t]}{[c]} = A + B(1 + s + s^2 + s^3) \quad n = 4 \tag{6e}$$

$$\frac{[t]}{[c]} = A + B(1 + s + s^2 + s^3 + s^4) \quad n = 5 \tag{6f}$$

$$\frac{[t]}{[c]} = A + B(1 + s + s^2 + s^3 + s^4 + s^5) \quad n = 6 \tag{6g}$$

$$\text{T-h}_n\text{-c-c}_m\text{-OH} \xrightarrow{K_i} \text{T-h}_n\text{-h-c}_m\text{-OH} \tag{7}$$

K_i is independent of the length m.

Data suitable for determination of an s value consists of a set of t/c values obtained for a series of homopeptide conjugates Ac-Hel$_1$-X$_i$-OH, in which each successive member of the series contains an additional amino acid residue X. The series may begin with a peptide of length m and proceed to length n, but in specific cases a scan through the complete series $i = 1$ to n may be more appropriate. For such a series the most general and satisfactory protocol for fitting an average s value to a block of t/c data is based on an iterative linear least squares analysis, whose details can be examined by means of expressions 6a−6g. Provided the starting peptide length for the series is properly chosen and the likely condition (7) is met, A, B, and s are expected to be constants. In (7), "h" signifies an amino acid residue in a helical conformation and "c", a nonhelical residue. Effectively, (7) implies that a local helical equilibrium constant K_i is independent of the length m of the random coil region to which the corresponding helical region is linked. This condition is equivalent to the assertion that the same K_i describes the helical propagating tendency at a given site i of any peptide conjugate Ac-Hel$_1$-X$_n$-OH, regardless of its length n.

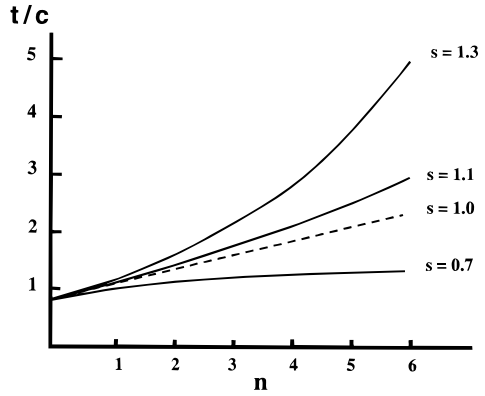

Figure 2. Calculated dependences of t/c on peptide length for hypothetical conjugates Ac-Hel$_1$-X$_n$-OH, for varying values of s. Curves are calculated from (5) with $A = 0.79$ and $B = 0.15$.

The iterative least squares analysis begins with selection of an initial value of s, which is then used to calculate values for each series on the right of (6a)−(6g). Each such value is then paired with the corresponding experimental value of t/c, and the resulting set of numerical pairs is subjected to a linear least squares fit for best values of A and B. Substitution of these along with s into the right hand expressions 6a−6g generates values for t/c$_{calc}$. From these and the corresponding values of t/c$_{exp}$ the error estimate, $\sum(t/c_{exp} - t/c_{calc})^2$ is calculated and used as a measure of the goodness of fit of the initial choice of s. The value of s is then varied, and the procedure is repeated to find the s value that minimizes the sum of the squares of the residuals. We define the value so obtained as $s_{effective}$. Intuitively, one might expect that $s_{effective}$ should be a reasonable approximation to the average of the K_i within the peptide region, even for cases in which a substantial variation of K_i is seen.

A graph of the t/c ratio as a function of the length of the peptide sequence permits a simple geometrical interpretation of the magnitude of $s_{effective}$. As shown in Figure 2, the constant A of eqs 4 and 6 is the zero intercept of the graph, and B is its initial slope. The value of $s_{effective}$ defines the curvature of the graph; the graph is linear for an $s_{effective} = 1$, concave upward for $s_{effective} > 1$, and concave downward for $s_{effective} < 1$. In the absence of significant random error in the data points, any graph of this type that is based on the general eqs 2−4 can be shown to generate a smooth curve, even if there is substantial variation in the magnitudes of the K_i values.

The propagation of error in the data points plays a major role in this analysis. Since the peptide length n assumes only integer values, t/c data appear discontinuously over the range of n. The experimental protocol therefore cannot be modified to focus data collection within the high t/c region, which bears a disproportionate weight in assigning curvature and which is generally subject to the greatest error. As described in the Experimental Section, for data sets in which substantial variation in t/c occurs, we have attempted to counterbalance the effect of errors at high t/c by assigning variances within the variance-covariance matrix[9] of the least squares analysis that are proportional to the magnitude of t/c.

The Systematic Residue Scan and the Problem of Implicit Errors in $s_{effective}$ as Assigned by the Least Squares Method

The above analysis ignores systematic variations in K_i that may occur with amino acid residues at the C-terminus or the

(9) Hamilton, W. C. Statistics in Physical Science; The Ronald Press Company: New York, 1964; pp 124−132.

template-peptide junction. Provided the peptide series is long enough and the experimental t/c values are sufficiently accurate, a simple analysis should be able to reveal these deviations. Inspection of (4) reveals that it can be rewritten as (8), which deletes a term in the te state sum on the right of (4) which is then subsumed into redefined template constants A' and B'. Factorization then shows that the original functional form of (4) is retained in (8). In effect, the initiation site has been redefined to include one hydrogen bonding site at the template-peptide junction, and the peptide helical region becomes one residue shorter.

$$\frac{[t]}{[c]} = A + B + BK_4(1 + K_5 + K_5K_6 + ...) = A' + B'(1 + K_5 + K_5K_6 + ...) \quad (8)$$

where $A' = A + B$ and $B = K_1K_2K_3K_4$.

If this redefinition is applied to all eqs 6a−6g for a given data set, a new block of equations is generated that contain all but the first of the original data points (that of (6b)). The null data point of (6a) is always kept, anchoring the series with the experimentally well-defined intrinsic t/c bias of the template. Least squares analysis of this new block results in changed values of additive and multiplicative constants A and B, with a corresponding change in the length of the geometric series. In this new analysis $s_{effective}$ is averaged over a more restricted region of the peptide sequence, closer to the C-terminus.

Generalizations through successive repetitions of the entire deletion and recalculation process results in a residue scan, which defines the sensitivity of $s_{effective}$ to peptide sequence. By excluding data points for residues near the peptide template junction, one can test for the sensitivity of $s_{effective}$ to energetic interactions peculiar to that junction. Restriction to a smaller data set will of course result in a reduction in precision for $s_{effective}$, and in practice the restricted data set should include at least three data points in addition to the null point.

Substantial variation might also occur among the K_i for larger i. In such a case three questions arise. Is the resulting $s_{effective}$ close to the $s_{average}$ of the K_i? Can one obtain an experimental measure of the heterogeneity of K_i values from the effect of a residue scan on $s_{effective}$? Does the form of the experimental data signal the heterogeneity in K_i? An overriding question is whether the experimental precision of a given data set blunts tests of these issues. We first address these questions by examining simulated data sets and then analyze the set of t/c ratios obtained for the series Ac-Hel$_1$-A$_n$-OH, $n = 1-6$ to test the practical limitations of this analysis.

Tests of the Reliability of $s_{effective}$ as a Measure of $s_{average}$

Only six homopeptide conjugates Ac-Hel$_1$-A$_n$-OH can be studied in water, since aggregation is observed for higher homologs of Ac-Hel$_1$-A$_6$-OH.[1] The small size of the t/c data set limits the scope of a residue scan and causes the calculated values of $s_{effective}$ to be unusually sensitive to errors in t/c values. Moreover the longest conjugate in the series is an atypical helix that contains only three alanine residues at the C-terminal region and three at the helix-template interface, with no helical core. Nevertheless, the Ac-Hel$_1$-A$_6$-OH helix with its eight amide residues corresponds to a nontemplated helical nonapeptide, and the average helix found within structures of the protein data base contains only 11 residues.[10] Even if s values for this series prove to be atypical of those expected for longer helices, they are likely to be representative of many short helices.

(10) Chothia, C. *Ann. Rev. Biochem.* **1984**, *53*, 537−572.

The iterative least squares procedure generates a series of $\sum(t/c_{experimental} - t/c_{calc})^2$ which show an approximate parabolic dependence on corresponding values for s. Both the minimal error value and the width of the parabola provide useful measures of the accuracy of $s_{effective}$, as can be seen from analyses of a model data set, constructed by substituting $A = 0.76$, $B = 0.15$, and $s = 1.035$ into eqs 6a−g. If the resulting t/c values are not rounded off, a least squares fit results in a value of $(1/(n-2))\sum(t/c_{model} - t/c_{calc})^2 \equiv (1/(n-2))\sum\delta^2$ that dramatically increases by six orders of magnitude when s is varied by ± 0.005 from the true value of 1.035. If error is introduced by rounding off t/c to three significant figures, the minimal $0.25\sum\delta^2 = 6.5 \times 10^{-6}$ corresponds to an s of 1.032 which is essentially the true value, but a mere 10% increase in $0.25\sum\delta^2$ results from a change in s of $\pm$ 0.02. Rounding t/c to two significant figures lowers the optimal s from the true value of 1.035 to 1.000, with $0.25\sum\delta^2 = 0.0010$; a change in s of $\pm$ 0.02 again causes a 10% increase in error. Since two significant figures provide a realistic estimate of the precision of experimental t/c measurements, s is characterizable within $\pm$5% if the data set consists of only seven t/c values.

A larger data set allows a more accurate characterization. If the same values of A, B, and s are used to simulate an 11-member t/c data set that is again rounded off to two significant figures, the least squares fit now yields an optimal s of 1.034 in excellent agreement with the true value, and $0.12\sum\delta^2 = 6.0 \times 10^{-4}$. In this case, a 10% increase in $(1/(n-2))\sum\delta^2$ corresponds to a change in s of $\pm$0.01.

Similar modeling was be used to explore the effects of large random errors, added to selected t/c values. Both $(1/(n-2))\sum\delta^2$ and the range of s that results in a 10% change in this error term are sensitive to the presence of random error, which is signaled by a significant decrease of both error parameters when individual data points are deleted from the analysis. A further modeling exercise tested the result of varying the K_i in eq 4. If the K_i are varied either systematically or randomly over the range of 0.6−1.5, $s_{effective}$ calculated for the full data set is always found to be an excellent approximation to $s_{average}$. Moreover residue scans reflect variations in K_i, provided they appear at sites other than at the C-terminus. A typical seven data point model was constructed using eq 4 by setting $A = 0.76$, $B = 0.15$, $K_4 = 0.8$, $K_5 = 1.0$, $K_6 = 1.2$, $K_7 = 1.4$, and $K_8 = 1.6$, with an $s_{average}$ of 1.20, and with calculated t/c values: 0.76, 0.91, 1.03, 1.15, 1.29, 1.50, 1.82. This model thus postulates a length dependent monotonic increase in helical stabilization. The full data set yields an optimal $s_{effective}$ of 1.24, in agreement with the average value, with $0.25\sum\delta^2 = 1.2 \times 10^{-3}$ and $s \pm 0.04$ for a 10% variance increase. Point deletions for $n = 1-5$ give similar errors and $s_{effective}$ values, but deletion of the t/c data point for $n = 6$ results in a dramatic decrease in optimal $s_{effective}$ to 1.11 ($s_{average} = 1.10$). This decrease is expected since the large value of K_8 is reflected solely in this point. Deletion of the data points for $n = 1$ and 2 yields an optimal $s_{effective}$ of 1.25, $0.50\sum\delta^2 = 9.4 \times 10^{-4}$, with $s \pm 0.025$ for a 10% variance increase, and deletion of the data points for $n = 1$, 2, and 3 yields an optimal $s_{effective}$ of 1.29, $\sum\delta^2 = 3.2 \times 10^{-4}$, with $s \pm 0.01$ for a 10% variance increase. Residue scans for this heterogeneous data set thus result in both an increase in $s_{effective}$ and a significant decrease in the error terms.

These results of all model experiments can be summarized. Individual site deletions appear to be best used as tests for unusually large errors in individual data points, signaled by a significant decrease in $(1/(n-2))\sum\delta^2$. Residue scans that result in a three- to five-fold decrease in $(1/(n-2))\sum\delta^2$ appear to be diagnostic for significant variations in K_i in the region of the

template junction. Unfortunately, the effects of large local deviations in K_i that characterize residues at the C-terminus cannot be distinguished from large random errors in the t/c data points for the longest peptide conjugates. The problem is implicit in the form of eq 4 which shows that only a few t/c depend on the K_i for which i is large. This serious ambiguity can only be addressed by examination of larger t/c data sets derived from study of longer helices.

Calculations of Values for $s_{effective}$ for Ac-Hel$_1$-A$_n$-OH, n = 1−6 in Water, 25 °C and 10 Mol % TFE−Water, 25 °C

The t/c ratios required for the least squares analysis of s values were obtained by integration of ^{1}H NMR resonances for each member of the series Ac-Hel$_1$-A$_n$-OH, n = 1−6. Spectra were taken in high purity D_2O as solvent, without water suppression and with protection from moisture during sample preparation. Small solvent deuterium isotope effects have been reported both for hydrophobic interactions[11] and for hydrogen bonding.[12] Although isotope effects may attend helix formation, a consistent picture of their magnitude and sign has not emerged. The t/c for Ac-Hel$_1$-A$_6$-OH is found to be 1.78 in D_2O and 1.75 in H_2O, but small weight can be placed on the latter number, which was obtained using water suppression which can alter the t/c value by a variety of mechanisms, including a state-elective NOE between water and template protons. Identical CD spectra of Ac-Hel$_1$-A$_6$-OH were observed at 25 °C in H_2O and in D_2O. Within the error limits of these techniques, there is no detectible isotope effect for helix formation for this conjugate.

The graphs of Figure 3a,b show the dependence of t/c on peptide sequence length for the oligomer series Ac-Hel$_1$-A$_n$-OH, n = 1−6, in solution in D_2O and 10 mol % TFE−D_2O at 25 °C. The span of t/c values seen in water alone is relatively modest and reflects the innate bias of the template Ac-Hel$_1$ against helix initiation. Addition of TFE results in a dramatic increase in the value of t/c. In the accompanying paper (*J. Am. Chem. Soc.* **1996**, *118*, 4240−4248) we note from study of NOE interactions in ^{1}H NMR spectra that TFE appears to stabilize existing structure without generating new conformations, and this is consistent with our recent analysis of the mechanism of helix stabilization by TFE.[13]

As shown in Table 1 a value of 1.03 ± 0.08 ($0.25\sum\delta^2$ = 0.009) is obtained for $s_{effective}$ in D_2O for the series Ac-Hel$_1$-A$_n$-OH, n = 1−6. This value is consistent with s_{Ala} = 1.07, reported by Scheraga and co-workers[3] for guest alanine in a host poly(hydroxybutyl)glutamine oligopeptide matrix, but very much smaller than estimates for s_{Ala} reported for a variety of small peptide systems.[14] Our $s_{effective}$ is also consistent with early reports of a lack of helicity for alanine block polymers.[15] The results of site deletions and a short residue scan are shown in Figure 4, which compares optimal $s_{effective}$ and the ranges of s that results in a 10% variation in $0.25\sum\delta^2$. Deletions at sites 1−5 do not change $s_{effective}$ significantly; however, deletion of the t/c value at site 6 decreases $s_{effective}$, and deletion of the t/c series 1, 2, and 3 modestly increases it, without a dramatic reduction in error. These results lie within the plausible error limits expected for the interpretation that K_i is constant throughout this series, although the data can be viewed as suggestive of an increase in K_i for n = 6. Elsewhere we have reported $s_{effective}$ for alanine of 1.01−1.03 for a 41-membered data set comprising the series Ac-Hel$_1$-Ala$_n$-NH$_2$, n = 1−6, together with several series Ac-Hel$_1$-Ala$_n$-K-Ala$_m$-NH$_2$.[8]

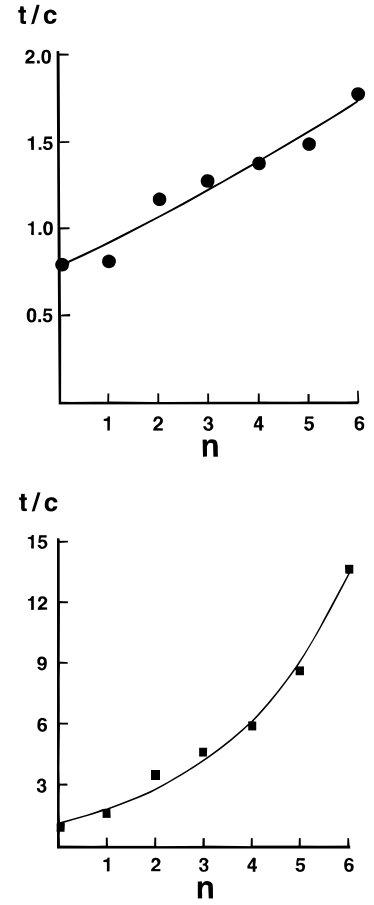

Figure 3. Length dependence of t/c for the series Ac-Hel$_1$-A$_n$-OH, n = 1−6, at 25 °C. a. In D_2O with ca. 1% TFA. b. In 10 mol % CF_3CD_2OD−D_2O with ca. 1% TFA.

Table 1. Least Squares Calculation of s Values

A.Ac-Hel$_1$-A$_n$-OH, n = 1−6, 25 °C, D_2O with ca. 1% TFA

1.Complete Scan, optimal s = 1.028; $\sum\delta_i^2$ = 0.0368

$0.79 = A + B\,(0)$
$0.80 = A + B\,(1)$
$1.17 = A + B\,(1 + s)$
$1.27 = A + B\,(1 + s + s^2)$
$1.37 = A + B\,(1 + s + s^2 + s^3)$
$1.47 = A + B\,(1 + s + s^2 + s^3 + s^4)$
$1.78 = A + B\,(1 + s + s^2 + s^3 + s^4 + s^5)$

B.Ac-Hel$_1$-A$_n$-OH, n = 1−6, in 10 mol % TFE D_2O, 25 °C, 1% TFA

1.Complete Scan, optimal s = 1.47; $\sum\delta_i^2$ = 1.8056

$0.79 = A$
$1.47 = A + B\,(1)$
$3.52 = A + B\,(1 + s)$
$4.63 = A + B\,(1 + s + s^2)$
$6.11 = A + B\,(1 + s + s^2 + s^3)$
$8.81 = A + B\,(1 + s + s^2 + s^3 + s^4)$
$14.64 = A + B\,(1 + s + s^2 + s^3 + s^4 + s^5)$

(11) Vdovenko, V. M.; Gurikov, Yu. V.; Legin, E. K. *Doklady Akad. Nauk, SSSR* **1967**, *172*, 126−129. Ben-Naim, A. *Israel J. Chem.* **1964**, *2*, 278−279. Schneider, H.; Kresheck, G. C.; Scheraga, H. A. *J. Phys. Chem.* **1965**, *69*, 4310−4324.

(12) Hermans, J., Jr.; Scheraga, H. A. *Biochem. Biophys. Acta* **1959**, *36*, 534−535. Scheraga, H. A. *Ann. NY Acad. Sci.* **1961**, 608−616. Calvin, M.; Hermans, J., Jr.; Scheraga, H. A. *J. Am. Chem. Soc.* **1960**, *81*, 5048−5050.

(13) Cammers-Goodwin, A.; Allen, T. J.; Oslick, S. L.; McClure, K. F.; Lee, J. H.; Kemp, D. S. *J. Am. Chem. Soc.* **1996**, in press.

(14) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. *Nature* **1990**, *344*, 268−270. Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5286−5290. O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646−651. Lyu, P. C.; Sherman, J. C.; Chen, A.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *68*, 5317−5320. Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. *Protein Science* **1994**, *3*, 843−852.

(15) Ingwall, R. T.; Scheraga, H. A.; Lotan, N.; Berger, A.; Katchalski, E. *Biopolymers* **1968**, *6*, 331−368.

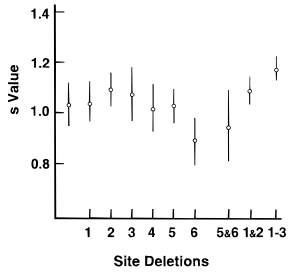

Figure 4. Effect of site deletions and a residue scan on the value of $s_{effective}$ for Ac-Hel$_1$-A$_n$-OH, $n = 1-6$, at 25 °C in D_2O. For the full data set the optimal value of $s_{effective}$ (open circles) and the corresponding error limits (the range of seffective that generates a 10% increase in $(1/(n-2))\sum\delta^2$) is shown on the left. The following six entries correspond to analyses of single site deletions, and the remaining three correspond to multiple site deletions (see text for discussion).

Table 2. te State Populations and Fractional Helicity in D_2O and 10 mol % TFE−D_2O for Ac-Hel$_1$-A$_6$-OH (Five Helical Sites)

a. General Analysis

te state	state function	helical mol fraction
Ac-Hel$_1$-c-c-c-c-c	1	$1/D$
Ac-Hel$_1$-h-c-c-c-c	s	s/D
Ac-Hel$_1$-h-h-c-c-c	s^2	s^2/D
Ac-Hel$_1$-h-h-h-c-c	s^3	s^3/D
Ac-Hel$_1$-h-h-h-h-c	s^4	s^4/D
Ac-Hel$_1$-h-h-h-h-h	s^5	s^5/D

$$\text{State Sum: } D = (1 + s + s^2 + ... + s^5) = \frac{(s^6 - 1)}{(s - 1)}$$

b. Specific Analysis

	helical mole fraction	
	$s = 1.03$ (D_2O)	$s = 1.47$ (10 mol % TFE)
Ac-Hel$_1$-c-c-c-c-c	0.155	0.05
Ac-Hel$_1$-h-c-c-c-c	0.16	0.08
Ac-Hel$_1$-hh-c-c-c	0.16	0.11
Ac-Hel$_1$-hh-h-c-c	0.17	0.16
Ac-Hel$_1$hh-h-h-c	0.17	0.24
Ac-Hel$_1$-hh-h-h-h	0.18	0.35

The value of $s_{effective}$ in 10 mol % TFE−D_2O for the series Ac-Hel$_1$-A$_n$-OH, $n = 1-6$, is 1.47 ± 0.06, ($0.25\sum\delta^2 = 0.45$). The results of single site deletions suggest that t/c values at sites 2 and 5 may have unusually large errors but are otherwise unexceptional. Deletion of the t/c values at sites 1 and 2 gives a value of $s_{effective}$ of 1.52 ± 0.05, ($0.50\sum\delta^2 = 0.49$) and a further deletion of site 3 increases $s_{effective}$ to 1.57 ± 0.03, ($\sum\delta^2 = 0.26$). As with those for the aqueous series, these results appear to lie within the plausible error limits of the interpretation that K_i is constant.

The structure of helices initiated by Ac-Hel$_1$ is shown schematically in Table 2. Provided $s_{effective}$ is a good approximation to each K_i, the fractional helicity at an amino acid site is fixed by the values of K_1 and $s_{effective}$. Within the te state, the fractional helicity at residue i is given by (9), and within the overall t state, the fractional helicity at residue i is given by (10).

$$\chi_i = \frac{(s^n - s^{i-1})}{(s^n - 1)}; \quad \chi_i \equiv \text{fractional helicity at residue } i \text{ within the (te) state} \quad (9)$$

$$f_{thi} = \left(1 - 0.79\,\frac{[c]}{[t]}\right)\chi_i; \quad f_{thi} \equiv \text{fractional helicity at residue } i \text{ within the (t) state} \quad (10)$$

The te-state fractional helicities calculated for Ac-Hel$_1$-A$_6$-OH for a range of values of $s_{effective}$ are shown in Figure 5. An s value of 1.03 results in a nearly linear decrease in fractional helicity with separation from the template junction. At residue four, the fractional helicity is less than 50%, and the helicity at residue six drops to less than 20%. Previously we have reported results of a glycine scan that are consistent with this degree of fraying.[2]

Values for $s_{effective}$ much larger than 1.0 show strong upward curvature in the graph, reflecting retention of structure at significantly greater separations from the template junction. In 10 mol % TFE−water, Ac-Hel$_1$-A$_6$-OH retains more than 50% helicity through the fifth hydrogen bonding site (at the φ and ψ angles of residue 4). Substantial fraying at the C-terminus (that is, at the φ and ψ angles of residue 5) is nevertheless calculated, even for s values of 2.0 or greater.

Spontaneous Helix Initiation within Templated Peptides and the Iniitiation Efficiency of Ac-Hel$_1$

Two fundamental questions remain to be addressed. What effect would spontaneous helix initiation have on the reporting properties of the Ac-Hel$_1$ template? The conventional theoretical models for peptide helicity assume that initiation is a random event, potentially occurring at any site within the peptide chain. By contrast, for a template-peptide conjugate, the initiation induced by the template at the N-terminus overwhelms spontaneous initiation at other sites. Helicities in these two cases are described by very different state sums, and the assumptions underlying the Zimm−Bragg and Lifson−Roeg formalisms[16] or any random initiation models render them inapplicable to helices formed by Ac-Hel$_1$ conjugates.

It might appear that the t/c reporting feature of Ac-Hel$_1$ is blind to spontaneously initiated helices, since to a first approximation these must populate the template cs, ts, and te states equivalently, resulting in no net change of t/c. A more careful analysis shows that an actual *decrease* in t/c can be anticipated if helices initiate spontaneously in the peptide region, but the magnitude of this decrease is difficult to estimate. The new spontaneously initiated helical states will populate the cs, ts, and te manifolds equally only if they share no substates with the template-initiated helical substates of the te manifold. Some of the substates corresponding to the longer and more stable helices will in fact appear in the te state sums of both the spontaneous and the template initiated manifolds. As a result, the te manifold will be stabilized to a lesser degree by spontaneous initiation than the cs and ts manifolds. Clearly the most useful applications of Ac-Hel$_1$ will involve peptide conjugates that are studied under conditions in which spontaneous helix initiation of the corresponding free peptide is insignificant. These include a majority of medium-sized peptide sequences in water at 25 °C.

The reporting features of Ac-Hel$_1$ have been achieved at a substantial price in initiating efficiency. If the template is viewed as an initiating species within the t state manifold, the constant B of (4) defines σ, which falls in the range of 0.12−0.18 in water and 0.6−0.8 in 10 mol % TFE−water. The low σ seen in water largely reflects the intrinsically unfavorable ts−te$_0$ equilibrium. However, the low σ also renders the measured t/c ratio relatively insensitive to change in helical stability. Values of t/c in excess of 15−20 are difficult to measure with experimental precision, and the insensitivity caused by a low σ expands the functional reporting range of this template. Though

(16) Zimm, B. H.; Bragg, J. *J. Chem. Phys.* **1959**, *31*, 526. Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963.

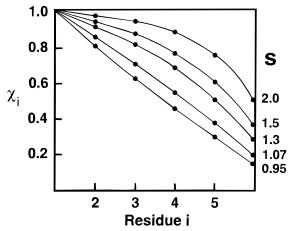

Figure 5. Fractional helicity χ_i at each site i of a heptapeptide (six helical residues) conjugated with an N-terminal initiation site and comprised of amino acids with constant s value. The five curves show the dependence of site fractional helicity on the value of *s*, as calculated from (10).

a relatively inefficient initiator, the template Ac-Hel$_1$ is nearly optimal as a reporter.

Structural variants of Ac-Hel$_1$ can be prepared that achieve different trade-offs between the effectiveness of initiating and reporting features. Elsewhere we have described an analog that lacks significant reporter features but which appears to be an exceptionally efficient helix initiator.[17]

Summary

An interative least squares analysis has been developed that assigns a value for the average helical propagation constant $s_{\text{effective}}$ to amino acid residues in a homopeptide conjugated with the reporting helical template Ac-Hel$_1$. The analysis requires a block of t/c ratios, assigned experimentally from ^{1}H NMR measurements of resonance ratios for a series of peptide conjugates of varying length. Residue scan and site deletion tests are described that are helpful in assessing the quality of the experimental data and the consistency of the $s_{\text{effective}}$ assignment. In D_2O at 25 °C in the presence of ca. 1% trifluoroacetic acid, the s value for alanine in the series Ac-Hel$_1$-A$_n$-OH, $n = 1-6$, is 1.03; in 10 mol % trifluoroethanol−D_2O under the same conditions, the s value is 1.47. The degree of fraying of an N-terminally initiated helix as been quantitatively related to $s_{\text{effective}}$, providing a complete characterization of helical structure for these simple systems.

Experimental Section

All Ac-Hel$_1$ derivatives were prepared, purified, characterized, and analyzed by ^{1}H NMR spectroscopy at 500 MHz as reported previously[1] or in the accompanying paper (*J. Am. Chem. Soc.* **1996**, *118*, 4240−4248).

CD Dilution. CD spectra were measured on an Aviv Model 62DS circular dichroism spectrometer at 25 °C in a 1 mm strain-free quartz cell. For ease in sample preparation a sample of Ac-Hel$_1$-A$_6$-OH was vacuum dried and dissolved in a 0.1 M perchlorate buffer, pH *ca*. 1 to form a stock solution. One aliquot of this was diluted into the original perchlorate buffer and another diluted into a 0.1 M D_2O-based perchlorate buffer, pD *ca*. 1 to create the samples for analysis. After dilution, the H_2O sample was in 100% H_2O, while the D_2O sample was in 90% D_2O. Concentrations of the sample solutions were determined by ninhydrin assay as described in the accompanying paper (*J. Am. Chem. Soc.* **1996**, *118*, 4240−4248).

Measurement of t/c Ratios for Ac-Hel$_1$-Ala$_n$-OH in D_2O and in 10 mol % TFE. The t/c ratios were measured from 500 MHz ^{1}H NMR spectra in D_2O as described previously,[1] with the following modifications. All spectra were carefully phased to pure absorption, and a linear base line correction routine (lvl/tlt) was applied to the region of integration. Integrations were measured 8 Hz outside the outermost resonance of each multiplet. t/c values reported are the average of five independently measured integration ratios. All t/c ratios were measured by careful integration of peaks in the range of 2.6−3.4 ppm, focusing on the 9b proton resonances. For these simple alanine derivatives, the two other previously described regions of the spectra were less suitable for integration due to either peak overlap or to interference by TFE. For spectra in pure water or low concentrations of TFE, the 9bc and 9bt proton resonances showed baseline resolution and could be integrated independently without corrections for overlapping resonances. Structural changes with higher TFE concentrations result in a shift of the 9bt resonances which can overlap with those of 13bc, 13bt, and 9at protons. In these cases this entire region was integrated. The unobscured 9bc resonance was then used as the measure of the c state abundance, and the appropriate corrections were made to the region of spectral overlap to determine the t state abundance, as described previously.[1]

Least Squares Analyses of t/c Data Blocks. To obtain contants *A* and *B* from a least squares fit[9] to equations of type (6a−g), programs of Mathematica 2.0 were employed to solve the matrix equation, $z =$ Inverse[(Transpose[*x*]. Inverse[*m*]·*x*)]·[Transpose[*x*]·Inverse[*m*]·*y*), where *y* is defined as the t/c data vector, $x = \{\{1,0\},\{1,1\},\{1,(s^2\text{-}1)/(s\text{-}1)\}, \{1,(s^3\text{-}1)/(s\text{-}1)\},\{1,(s^4\text{-}1)/(s\text{-}1)\},\{1,(s^5\text{-}1)/(s\text{-}1)\},\{1,(s^6\text{-}1)/(s\text{-}1)\}\}$, and *m* is the variance-covariance matrix, which in this study can only be approximated to within a scale factor. Since the t/c ratios are not known to be correlated, all covariances (off-diagonal terms) were set equal to zero. Relatively small variations are observed for t/c data measured in D_2O; for analysis of these, *m* is taken as the identity matrix (all variances proportional to 1). Large t/c variations were observed in 10 mol % TFE-D_2O, as noted in Table 1. For these the variances that appear along the diagonal of *m* were taken as proportional to the t/c ratio itself. Negligible changes in modeling outcome were noted when t/c was replaced by $(t/c)^{0.5}$.

Acknowledgment. Financial support from the National Institutes of Health, Grant 5 R37 GM13453, by the National Science Foundation Grant No. 9121702-CHE, and from Pfizer Research is gratefully acknowledged.

JA952911S

(17) Kemp, D. S.; Rothman, J. R. *Tetrahedron Lett.* **1995**, *36*, 4023−4026.