Entropic benefit of a cross-link in protein association

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · AUGUST 2002

Impact Factor: 2.63 · DOI: 10.1002/prot.10165 · Source: PubMed

3 AUTHORS, INCLUDING:

Richard Stephen Berry
University of Chicago
509 PUBLICATIONS **14,513** CITATIONS
SEE PROFILE

Tobin R Sosnick
University of Chicago
161 PUBLICATIONS **8,309** CITATIONS
SEE PROFILE

PROTEINS: Structure, Function, and Genetics 48:341–351 (2002)

Entropic Benefit of a Cross-Link in Protein Association

Muhammad H. Zaman,[1,3] R. Stephen Berry,[1] and Tobin R. Sosnick[2,3]*
[1]*Department of Chemistry, University of Chicago, Chicago, Illinois*
[2]*Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois*
[3]*Institute of Biophysical Dynamics, University of Chicago, Chicago, Illinois*

ABSTRACT **We introduce a method to estimate the loss of configurational entropy upon insertion of a cross-link to a dimeric system. First, a clear distinction is established between the loss of entropy upon tethering and binding, two quantities that are often considered to be equivalent. By comparing the probability distribution of the center-to-center distances for untethered and cross-linked versions, we are able to calculate the loss of translational entropy upon cross-linking. The distribution function for the untethered helices is calculated from the probability that a given helix is closer to its partner than to all other helices, the "Nearest Neighbor" method. This method requires no assumptions about the nature of the solvent, and hence resolves difficulties normally associated with calculations for systems in liquids. Analysis of the restriction of angular freedom upon tethering indicates that the loss of rotational entropy is negligible. The method is applied in the context of the folding of a ten turn helical coiled coil with the tether modeled as a Gaussian chain or a flexible amino acid chain. After correcting for loop closure entropy in the docked state, we estimate the introduction of a six-residue tether in the coiled coil results in an effective concentration of the chain to be about 4 or 100 mM, depending upon whether the helices are denatured or prefolded prior to their association. Thus, tethering results in significant stabilization for systems with millimolar or stronger dissociation constants. Proteins 2002;48:341–351.** © 2002 Wiley-Liss, Inc.

Key words: translational entropy; rotational entropy; probability distribution function; coiled coil; Gaussian random coil; tether; binding

INTRODUCTION

"Thus we are forced to the conclusion that there is no basis for estimating the standard free energy change for the binding of a molecule to a macromolecule from the corresponding energies of binding for molecules representing its component parts without a detailed knowledge of the properties for the system."[1]

Part of this dilemma posed by Jencks in his classic treatise on enzymology, is the difficulty of calculating from association constant of n component system (K_{ass}, units of M^{-n+1}), the related association constant for the lower order reaction where two of the components are tethered[2] (K_{ass}, units of M^{-n+2}). This issue can be cast in the specific situation of concentration-dependent bimolecular docking reaction $A + B \leftrightarrow A \cdot B$: Given the free energy of this reaction, ΔG^{o}_{bimol}, can one predict the concentration independent ΔG_{uni} for the corresponding unimolecular reaction where the components are cross-linked with a flexible tether $A \ldots B \leftrightarrow A \cdot B$?

We first note that the loss of entropy upon tethering is completely distinct from the loss of entropy upon binding (Fig. 1). Binding of two species of comparable sizes effectively reduces the independent translational and rotational freedom of the two bodies to that of a single body. Upon cross-linking, however, each of the two species retains a considerable amount of translational and rotational freedom, given that the tether is of moderate size and flexibility. The difference between the cross-linking and binding processes is clearly noted by considering the two-step process, $A + B \leftrightarrow A \ldots B \leftrightarrow A \cdot B$, where the tether is introduced, followed by the two tethered species binding to each other in a complex ($A \cdot B$). Hence, binding and tethering represent fundamentally different processes even though they both reflect a reduction of the dimensionality of the system.

Traditionally, the Sackur-Tetrode equation describing entropy in the gas phase has been used to estimate the entropy of binding, cross-linking, and association of atoms, molecules, and macromolecules.[3,4] However, this equation does not consider the molecular volume occupied by the solvent, and, hence, probably overestimates the entropy in the liquid phase. In the cell theory of liquids, another approach to the problem, the entire volume of the solvent

Abbreviations used: C_{eff}, effective concentration; C_{ref}, reference concentration; C_T, molar concentration; ΔG°_{bi}, standard state free energy of a bimolecular reaction; ΔG_{uni}, free energy of a unimolecular reaction; $\Delta S_{cross\text{-}linking}$, cross-linking entropy; e.u., entropic unit (1 e.u. = 1 cal mol^{-1} K^{-1}); K_{ass}, association constant; NN, nearest neighbor; $P(r)$, probability distribution of center-to-center distances; $P^{helix}_{dimer}(r)$, center-to-center distance distribution for dimeric system; $P^{helix}_{tether}(r)$, center-to-center distance distribution for tethered system; $P_{NN}(r)$, probability of the partner being the NN; r_{ave}, average inter-particle separation; ρ, number density; SSI, Streptomyces subtilisin inhibitor; V_{mol}, volume per particle; V_T, volume of the system.

Grant sponsor: NIH; Grant sponsor: Packard Foundation Interdisciplinary Science Program.

*Correspondence to: Tobin R. Sosnick, Dept. of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637. E-mail: trsosnic@midway.uchicago.edu

Received 23 Oct 2001; Accepted 18 March 2002

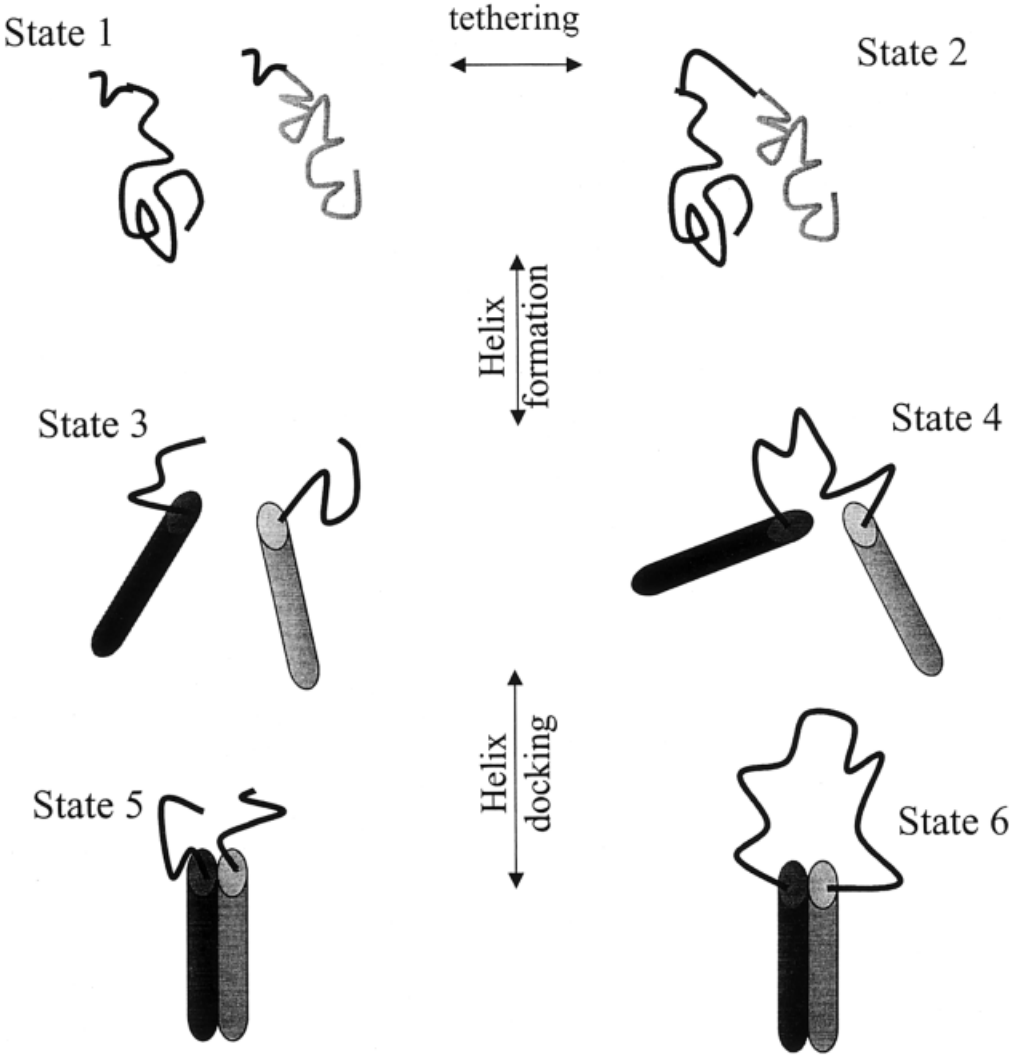

Fig. 1. Individual steps in folding, binding, and tethering. The left column represents the process of two untethered, denatured chains that form helices, and then bind to form a helical dimer. The two unstructured monomers are modeled as Gaussian random walks (State 1); the isolated (State 3) and bound (State 5) helices are modeled as thin rods. States 2, 4, and 6, are the corresponding states for a system where the two chains have been tethered. The introduction of the tether, a shift from the left to the right column, results in a decrease in the reaction order for the isolated chains (1 ↔ 2, and 3 ↔ 4), but not for pre-docked helices (5 ↔ 6). The loss of translational entropy upon introduction of the tether for the upper two transitions is calculated according to the NN method, while for the lower transition, the loss is calculated according to loop closure entropy.

is divided into fixed cages or cells.[5] Each cell contains a single given molecule, an unrealistic assumption that precludes the molecules from fully sampling the entire volume.

To compare the bimolecular to unimolecular system, we calculate the center-to-center probability distributions for the two species before and after their tethering. These distributions, $P^{helix}_{\mathrm{dimer}}(r)$ and $P^{helix}_{tether}(r)$, respectively, are used to calculate the entropy of each state according to

$$S = -R\int_0^\infty 4\pi r^2 P(r)\ln P(r)dr \qquad (1)$$

The distribution $P^{helix}_{\mathrm{dimer}}(r)$ inherently depends upon the concentration of reactants, typically chosen to be at 1 M standard state. The more concentrated the reactants, the less entropy is lost upon introduction of the tether. The distribution $P^{helix}_{tether}(r)$ depends upon the length and nature of the tether.

One of the major benefits of using probability distribution functions is that the presence of the solvent does not significantly affect the calculation of the benefit of cross-linking. Although the volume occupied by the solvent molecules may restrict the available volume of each helix, the distribution functions themselves are not significantly altered by the liquid. Additionally, any reduction should be

independent of the presence of the tether. Hence, the entropy of both states should be reduced by the same amount, which cancels out in the calculation of the overall change in entropy.

To calculate the untethered system's distribution, we introduce the "Nearest Neighbor" method, where $P^{helix}_{dimer}(r)$ is posited to reflect the probability that a partner can travel a given distance from a reference helix while still being its nearest neighbor (NN). The center-to-center and relative angle distribution functions of the tethered system, connected by Gaussian random coils or poly-L-alanine chains, are compared to their untethered counterparts to estimate the loss of translation and rotational entropy accompanying cross-linking. Here, we illustrate this methodology to the docking of two helices, as well as to the general association of three components where two of them are pre-tethered. The method is extended and compared to experimental results for a variety of proteins.

MATERIALS AND METHODS

The entropy calculations were performed using a program written in Mathematica® 4.1 developed by Wolfram Research Inc. A chain of n residues (unfolded peptide or the cross-link) was modeled either as a Gaussian random walk with $2n$ segments (length 1.5 Å = $\frac{1}{2}$(C_α-to-C_α distance)), or as a poly-alanine chain. For the latter, each alanine's conformation was specified by the occupation of three discrete regions in the Ramachandran Φ, Ψ plot (extended, α- and 3_{10} helical regions, which are the upper left, lower left, and upper right quadrants, respectively). The regions were approximated as ellipsoids according to Flory's method.[6] The Φ, Ψ values were randomly chosen within each region in proportion to the size of each ellipsoid to determine the overall conformation of the chain.

The excluded volume effects were investigated as outlined by Pappu et al.[7] Based upon a hard-sphere model, the sterically allowed Φ, Ψ angles for an alanine dipeptide were generated. A steric clash between two atoms exists when their contact distance is less than the hard-sphere contact distance. Equally distributed values of these sterically allowed Φ, Ψ angles were used to generate conformations of chains up to twelve alanine residues. Each chain's conformation was then screened for steric clashes between any two atoms of non-adjacent residues. The conformations without any clashes were tabulated to calculate the percentage of allowed conformations for each chain length.

RESULTS AND DISCUSSION

Loss of Translational and Rotational Entropy Upon Cross-Linking

The folding and binding of a pair of helices is modeled with six states, which are related by transitions representing either folding, cross-linking, or binding (Fig. 1). The horizontal arrows represent the cross-linking process whether the system is in the denatured ($1 \leftrightarrow 2$), helical ($3 \leftrightarrow 4$), or docked conformations ($5 \leftrightarrow 6$). The vertical arms represent the folding from a random coil to helical structure ($1 \leftrightarrow 3$ and $2 \leftrightarrow 4$), or the docking of two pre-folded helices ($3 \leftrightarrow 5$ and $4 \leftrightarrow 6$). The translational entropy depends upon the distribution of center-to-center distances whereas the rotational entropy depends upon the distribution of the relative angles between the two components. The reduction in translational entropy upon cross-linking of pre-folded helices is calculated by comparing the center-to-center probability distribution of the two helices before, $P^{helix}_{dimer}(r)$, and after tethering, $P^{helix}_{tether}(r)$, according to:

$$\Delta S_{trans} = S_{tether} - S_{dimeric} = -R\int_0^\infty (4\pi r^2\{(P^{helix}_{tether}(r)\ln P^{helix}_{tether}(r) - P^{helix}_{dimer}(r)\ln P^{helix}_{dimer}(r)\})dr \quad (2)$$

where R is the gas constant, and the distribution has normalization $\int_0^\infty 4\pi r^2 P^{helix}_{tether}(r)dr = 1$.

Although internal vibrational motions of the docked complex must be accounted for in a calculation of the energetics of a given binding process (ΔG°_{bimol} or ΔG_{uni})[8], these motions are not altered by introduction of an ideal tether. They should contribute equally in the bound state of both the tethered and untethered systems, and do not affect ΔS_{trans} as defined in Eq. 2. Thus, internal motions do not need to be considered in the present calculation of the change in entropy upon tethering.

The reduction in rotational entropy is obtained from a comparison of the rotational distribution function for the tethered species, $P_{tether}(\theta,\phi)$, with that of the untethered species, which is a uniform distribution $P_{untether}(\theta,\phi)$:

$$\Delta S_{rot} = -R\int [P_{tether}(\theta,\phi)\ln P_{tether}(\theta,\phi) - P_{uniform}(\theta,\phi)\ln P_{uniform}(\theta,\phi)]d(\theta,\phi) \quad (3)$$

where θ, and φ are the angles in spherical coordinates between the axes of the two helices with normalization $\int P(\theta,\phi)d(\theta,\phi) = 1$. The unfolded peptide and unstructured tethers are initially approximated as Gaussian random walks (or chains)[9,10] and the helices as thin, non-interacting rods. Both the tether and the unfolded polypeptide are assumed to be Gaussian random walks. The excluded volume effects of the chain are discussed in a later section.

Translational Entropy

The loss of translational entropy is calculated upon the introduction of the tether for each of the three horizontal transitions shown in Figure 1.

Docked helices

The process $5 \leftrightarrow 6$ represents the ligation of the tether and the formation of a closed loop. The ligation process is unimolecular (ignoring the covalent peptide bond formation). Hence, the concentration of reactants is irrelevant to the loss of translational and rotational entropy. According to Flory and Jacobson-Stockmayer Theory for Gaussian

chains, the entropic cost of loop closure of n segments is[6,9,10]

$$\Delta S_{5\text{-}6} = -3/2R \ln(n\pi/2) \quad (4)$$

Undocked helices

The process $3 \leftrightarrow 4$ represents the cross-linking of a system of two pre-folded, but undocked helices. Tethering of the helices results in a change from a bimolecular to a unimolecular system. The loss of entropy for this process is calculated from the change in the probability distribution of center-to-center distances that describes the configuration before and after the introduction of the tether, i.e., $P^{helix}_{\text{dimer}}(r)$ and $P^{helix}_{tether}(r)$, respectively.

The relative motion of the two cross-linked helices is restricted by the tether, which itself has an end-to-end probability distribution (Fig. 2). For an n residue tether, this distribution is approximated by a freely-jointed Gaussian random walk of $2n$ segments (each amino acid has two torsion angles):

$$P_{RC}(r) = (\beta/\sqrt{\pi})^3 e^{-\beta^2 r^2} \quad (5)$$

where $\beta = \sqrt{3/(2nl^2)}$, and l is the length of each step[6,11], taken to be 1.5 Å.

Rather than the end-to-end distribution of the tether, the relevant quantity for the calculation of the loss of entropy for the helices is their center-to-center distribution. This distribution can be calculated from the tether's end-to-end distribution by adding to the ends, an extra step of half the length of the helix chosen at random exit angles. We obtain $P^{helix}_{tether}(r)$ by simulating such a random walk (Fig. 2), and the translation entropy of the tethered helices from Eq. 1.

Next, the center-to-center distribution function for the unlinked system, $P^{helix}_{\text{dimer}}(r)$, is calculated. In principle, the two unlinked helices are uniformly distributed over the entire volume of the system, V_T. The application of this uniform distribution, however, greatly overestimates the accessible volume and results in an incorrect, system-size dependence on the entropy.

To circumvent this problem, we introduce the Nearest Neighbor (NN) Method. The relevant volume for a given partner helix is limited to the region around a reference helix where the partner is the nearest neighbor. As the partner diffuses a distance away from the reference helix, another helix is likely to become the closest helix. The distance the original partner can travel while remaining the nearest neighbor is dependent upon the number density, $\rho = N/V_T \equiv 1/V_{\text{mol}}$.

We posit that the probability of the partner being the closest, $P_{NN}(r)$, is proportional to the center-to-center distance distribution, $P^{helix}_{\text{dimer}}(r)$, between the partner and the reference helix, differing only by a scale factor required to maintain unit normalization. As the initial choice of both the reference helix and partner are arbitrary, and their mutual association is not fixed in time (i.e., a third helix can become the NN), issues related to distinguishability are circumvented.

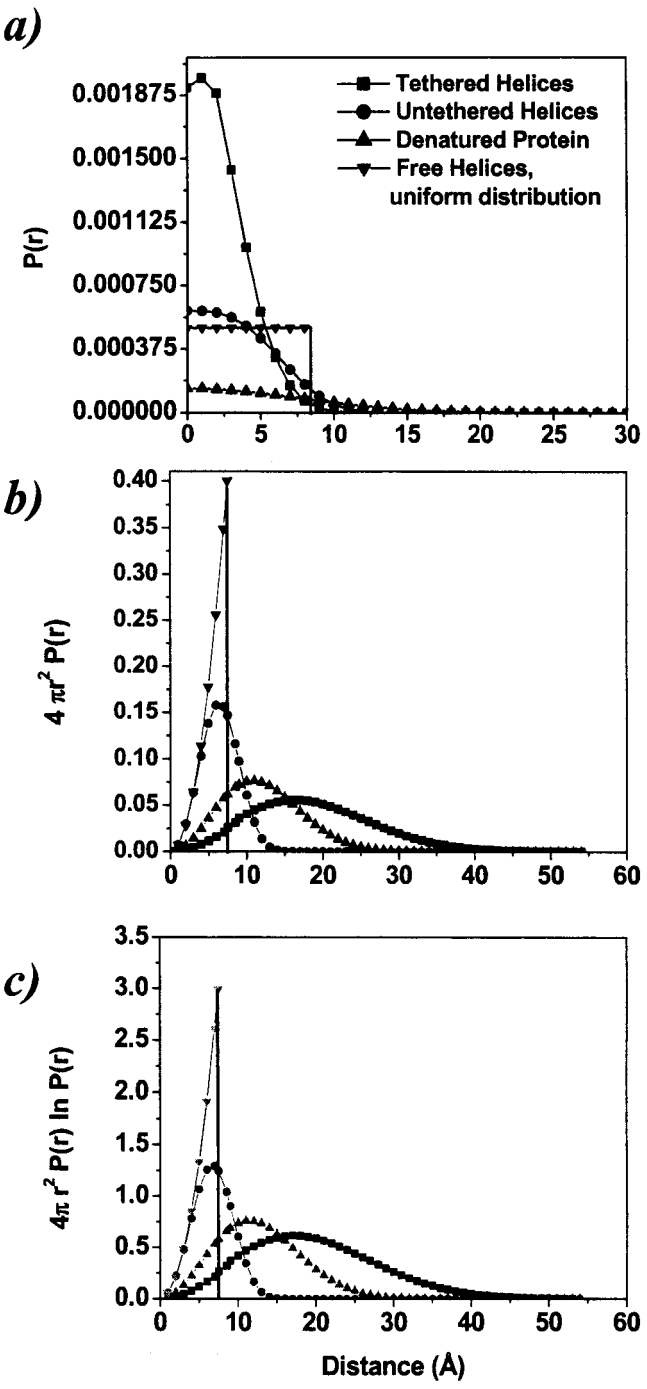

Fig. 2. Probability distribution functions. The center-to-center probability distribution functions for unfolded chains and cross-linked helices, either untethered, (calculated according to the NN Method), or connected by a six residue tether (modeled as a Gaussian random coil). **a:** $P(r)$ distribution. **b:** The unit normalized $4\pi^2 P(r)$ distribution. **c:** $4\pi r^2 P(r) \ln P(r)$. The area under the curve in *c* for each distribution (multiplied by $-R$) is the translational entropy. For comparison, the distribution is shown for a molecule uniformly distributed within a spherical volume, but having the same entropy as the NN distribution. Symbols are the same in all three panels.

The NN, and, hence, $P^{helix}_{\text{dimer}}(r)$ distribution can be calculated from the probability that no other helix is the NN. For the partner at a distance r from the reference helix, the

TABLE I. Change in Translational Entropy Upon Cross-Linking

Length of tether	$\Delta S_{trans}^{denatured}$ (e.u)[a]	$\Delta S_{trans}^{pre\text{-}folded}$ (e.u)[b]
6	3.89	−2.15
9	4.11	−0.838
12	4.32	0.104
15	4.51	0.84
18	4.69	1.45

[a]Tethering of denatured helices, Process 1 ↔ 2.
[b]Tethering of pre-folded helices, Process 3 ↔ 4.

probability that another particular helix is the NN is $v/V_T = \nu\rho/N$, where $\nu = 4\pi r^3/3$. The probability the original partner remains the NN is the probability that none of the other N helices are closer:

$$P_{NN}(r) = (1 - \nu\rho/N)^N \sim e^{-\nu\rho} \tag{6}$$

After normalization, we obtain $P_{\mathrm{dimer}}^{helix}(r) = \mathrm{e}^{-\nu\rho}/\rho$.

Having determined $P_{\mathrm{dimer}}^{helix}(r)$ in this manner, we can calculate the entropy of the untethered, dimeric helices according to Eq. 1. At 1 M standard state, $1/\rho = V_{\mathrm{mol}} = 1{,}661$ Å^3, and $\mathrm{S_{dimeric}} = -17.02$ e.u. Interestingly, this value of entropy is equivalent to that for a helix uniformly distributed within a volume of 1,839 Å^3. This volume is very close to V_{mol}, for which $\mathrm{S} = -16.82$ e.u. (Fig. 2).

The difference in entropy between the untethered system and the cross-linked helices (process 3 ↔ 4) provides the entropic cost of cross-linking in the helical, but undocked, state as per Eq. 2. For the introduction of a six-residue tether to a bimolecular system at 1 M standard state, the entropy is decreased by 2.15 e.u., or $\mathrm{T\Delta S_{trans}} = -0.64$ kcal mol^{-1} at T = 300 K. Values for other length tethers are listed in Table I.

Although the tethering the two undocked helices at 1 M standard state by six residues decreases the entropy of the system, a tether of twelve or more residues results in an increase in the entropy of the system. It may seem paradoxical that the tethering can result in an increase in entropy. The increase indicates the tethered but undocked helices sample more volume than do the isolated helices at the (high) standard state concentration of 1 M. The entropy of an untethered helix equates to it uniformly sampling a box of dimensions of only (12.2 Å)^3, whereas the mean center-to-center distance of two helices with a twelve-residue tether is over twofold larger. This increase for the tethered helices explains why a twelve-residue tether actually increases the entropy in the cross-linked state *when compared to the free state at 1 M standard state concentration* (see discussion below). Part of the effect comes from the fact that the tethered helix cannot be exchanged with another helix, however far away it moves, whereas the untethered one can when it moves further away than 12.2 Å. For a more realistic standard state concentration of 1 μM, the volume sampled by the untethered system is increased 10^6-fold, and the introduction of the tether does reduce the translational entropy of the helices.

The use of the distribution function and Eq. 1 to calculate the translational entropy correctly accounts for the concentration dependence, $\Delta\mathrm{S(C)} = \Delta\mathrm{S}(\mathrm{C_{ref}}) - \mathrm{R\ ln\ C_{new}/C_{ref}}$, where $\mathrm{C_{new}}$ and $\mathrm{C_{ref}}$ are the new and reference concentration, respectively. For example, an N-fold increase in solute concentration results in the distribution function being uniformly contracted N-fold along the r-axis (with a commensurate increase in height to maintain unit normalization): $P_{C_{new}}(r) = NP_{c_{ref}}(rN)$. The entropy of the system at higher concentration is

$$\begin{aligned} S(C) &= -R\int 4\pi r^2 P_C(r)\ln P_C(r)dr \\ &= -R\int_0^\infty [4\pi r^2 N P_{C_{ref}}(rN)\ln P_{C_{ref}}(rN) \\ &\quad + 4\pi r^2 N P_{C_{ref}}(rN)\ln N]dr \\ &= -R\int_0^\infty 4\pi r^2 P_{C_{ref}}(rN)\ln P_{C_{ref}}(rN)(Ndr) \\ &\quad -R\ln N\int_0^\infty 4\pi r^2 P_{C_{ref}}(rN)Ndr = S(C_{ref}) - R\ln N \end{aligned} \tag{7}$$

Hence, the method has the correct dependence on the solute concentration.

Denatured polypeptides

The analysis for tethering in the denatured state (Process 1 ↔ 2) is similar to that for the undocked helices. Again, the change in translational entropy is calculated from the change in the center-to-center distance distributions, $P_{\mathrm{dimer}}^{denatured}(r)$ and $P_{tether}^{denatured}(r)$. The $P_{\mathrm{dimer}}^{denatured}(r)$ distribution is identical to its helical counterpart, $P_{\mathrm{dimer}}^{helix}(r)$, as the NN Method did not assume any shape for the reactants.

The tether restricts the center-to-center distribution of the two denatured polypeptides. The tether's end-to-end distribution when convoluted with each polymer's end-to-center distribution is the desired center-to-center distance distribution between the two polypeptides. Alternatively, $P_{tether}^{denatured}(r)$ can be calculated by realizing that it is the end-to-end distribution for the portion of the chain connecting the two centers of the denatured (identical) helices. The number of residues in this portion is $N_{\mathrm{helix}} + N_{\mathrm{tether}}$.

The calculation of change in entropy upon the introduction of a cross-link in the unfolded state is carried out by using Eq. 2. We find $\Delta\mathrm{S_{trans}} = 3.9$ e.u. ($\mathrm{T\Delta S_{trans}} = 1.16$ kcal mol^{-1} at 300 K) for the introduction of a six-amino-acid-long tether in the denatured state, Process 1 ↔ 2. The increase in entropy indicates that the center of tethered polypeptide samples more configurational space than it does when it is untethered at a standard state concentration of 1 M.

TABLE II. Effect of Tether on the Association of Denatured and Pre-Folded Helices[†]

Length of tether	ΔS° (e.u.)	$\Delta G^\circ_{bi} - \Delta G_{uni} = -T\Delta S^\circ$	C_{eff} (M)[a]
6	−10.6 (−4.6)	3.2 (1.4)	0.0049 (0.10)
9	−12.0 (−7.1)	3.6 (2.1)	0.0024 (0.029)
12	−13.1 (−8.9)	3.9 (2.7)	0.0014 (0.012)
15	−14.0 (−10.3)	4.2 (3.1)	0.00092 (0.0056)
18	−14.7 (−11.5)	4.4 (3.4)	0.00063 (0.0032)

[†]Process $1 \leftrightarrow 5$ relative to Process $2 \leftrightarrow 6$, at 1 M standard state concentration. Values for Process $3 \leftrightarrow 5$ relative to Process $4 \leftrightarrow 6$ are given in parentheses. Stability is given in kcal M^{-1}, at T = 300 K.
[a]Calculated according to $C_{eff} = e^{\Delta S/R}$.

Size Dependence of Cross-Linking Entropy

The entropic cost of cross-linking depends upon the size of components. An increase from 33 to 66 residues, for example, results in the entropy of tethering increasing by 0.6 and 1.82 kcal M^{-1} (2 and 6.06 e.u., respectively) for denatured and pre-folded helices (Process $1 \leftrightarrow 2$, and $3 \leftrightarrow 4$), respectively, when they are connected by a six-residue helix. The increase in entropy for the denatured helices of 66 residue is because the center-to-center distance distribution is that of a 72-residue ($N_{helix} + N_{tether}$) random walk rather than the original 39 residue random walk. Likewise, the increase in entropy for the pre-folded helices is because of their increased length, which results in a more extended distribution function. The entropy values listed in Tables I and II, and mentioned elsewhere in this paper, are for tethers of different lengths. The helix, however, is kept at a constant 33 residues and length of 30 Å.

Rotational Entropy

The analysis of the loss of rotational entropy parallels that for translational entropy, except that the restriction in the relative angle, rather than the distance, between the two helices is the pertinent quantity. Of the three horizontal processes in Figure 1, the relative angle is likely to affect only the tethering of the two undocked helices (Process $3 \leftrightarrow 4$). The tether does not change the orientation between two docked helices (Process $5 \leftrightarrow 6$), nor does it significantly restrict the rotational freedom between the two denatured polypeptides (Process $1 \leftrightarrow 2$).

The angular distribution of the undocked helices, $P(\theta,\phi)$, relative to a uniform distribution, $P_{uniform}(\theta,\phi)$, is used to calculate loss of rotational entropy (Eq. 3). For arbitrarily shaped objects, a third angular degree of freedom is required. However, for the cylindrically symmetric helices with a freely jointed tether examined here, there is no restriction in this quantity.

A tether composed of amino acids has steric restraints due to restriction of each residue's Φ, Ψ dihedral angles. This restriction may result in a decrease in the angular freedom between the two helices. In order to investigate this effect, simulations are carried out with polyalanine tethers. The dihedral angles for each residue are chosen to reflect the restriction of the polypeptide backbone and side-chain moieties. The tether's conformation is coarsely specified for each residue by the occupation of three discrete regions in the Ramachandran Φ, Ψ plot.[6] The regions are approximated as ellipsoids according to Flory's method.[6] The Φ, Ψ values are randomly chosen within each basin according to their relative areas to determine the overall conformation of the chain.

To assess the loss of rotational entropy, the angle between the long axes of the helices is histogrammed for each configuration. This angle is assumed to be the same as the angle between the first and last segment of the tether (i.e., the helix is fixed in angle relative to the adjoining segment in the tether). The histogram of angles is coarsely divided into octants of a sphere [Fig. 3(a)]. For a six-residue tether, the angular distribution is nearly uniform, indicating that the helices are essentially freely jointed for this length of tether. Hence, the loss of rotational entropy upon tethering is quite small (<0.5 e.u.), and generally negligible as compared to the loss of translational entropy. Furthermore, the loss of rotational entropy upon tethering is considerably less than 45 e.u. ($T\Delta S$ = 13.5 kcal mol^{-1} at T = 300 K), the loss of rotational entropy for the binding of two 4 kD proteins.[4]

Entropic Benefit of Cross-Linking

Having calculated the entropic cost of cross-linking for each of the individual steps, and concluded that the loss of rotational entropy is insignificant, we can estimate the net benefit of cross-linking to the entire docking equilibrium of the unfolded (Process $1 \leftrightarrow 5$ relative to $2 \leftrightarrow 6$) or pre-folded helices (Process $3 \leftrightarrow 5$ relative to $4 \leftrightarrow 6$). Under the assumption that docking is intra-molecular, the introduction of a cross-link results in a loop closure penalty in the docked state. The tether either decreases or increases the entropy of free, undocked state, depending upon its length (Tables I, II). For example, closing a six-residue tether in the docked state (Process $5 \leftrightarrow 6$), results in a loop closure penalty equivalent to a reduction in entropy of 6.7 e.u. At 1 M standard state, the introduction of the tether increases the translational entropy by 3.9 e.u. for the denatured helices (Process $1 \leftrightarrow 2$), but decreases it by 2.15 e.u. for the pre-folded, undocked helices ($3 \leftrightarrow 4$).

Thus, the introduction of the tether in the denatured state (Process $1 \leftrightarrow 2$) favors the tethered state while in the docked state (Process $5 \leftrightarrow 6$), it favors the untethered state. Hence, for the entire reaction (Process $1 \leftrightarrow 5$ relative to $2 \leftrightarrow 6$), the tether's effect in both the undocked and docked states opposes the formation of the docked state by a total 10.6 e.u. When the undocked state is already helical, the tether's contribution favors the untethered state in both undocked and docked states. The net of these opposing factors still inhibits the formation of the docked state by a total of 4.6 e.u. Therefore, in either situation, the introduction of a tether results in a decrease in the amount of docked species.

However, the conclusion that tethering is entropically destabilizing explicitly depends upon the standard state concentration used in the calculation. The less concen-

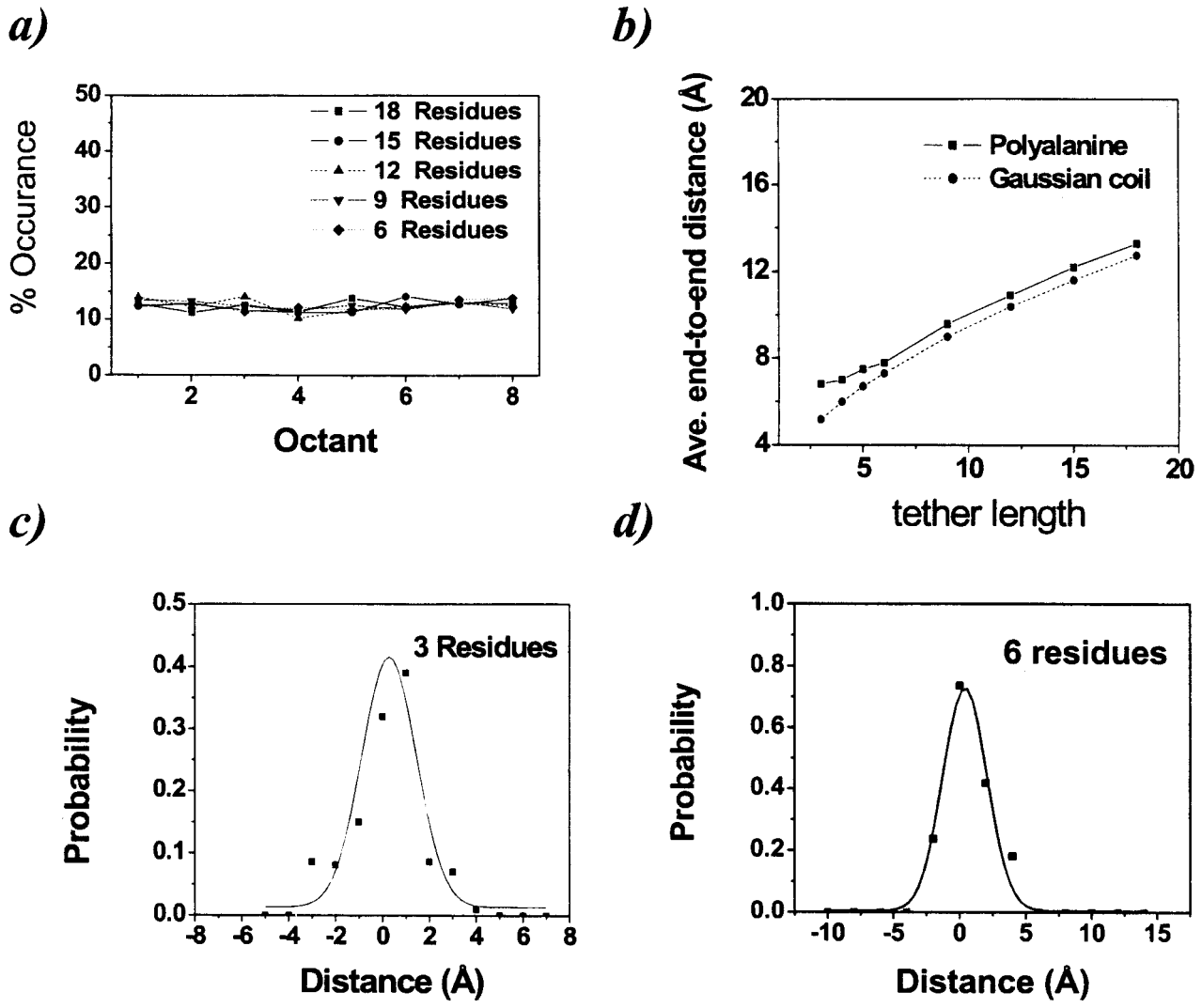

Fig. 3. Properties of a polyalanine tether. **a:** Rotational entropy. The angle between the two helices, approximated as the angle between the first segment and the last segment of the polyalanine tether, are shown. The uniform distribution indicates that the loss of rotational entropy upon tethering is minimal at these linker lengths. **b:** Comparison between end-to-end distances for Gaussian random coils and polyalanine chains. The distributions are significantly different for three-residue polyalanine chains, but are similar for six-residue (and longer) chains. **c:** Probability distribution of end-to-end distances for Gaussian random coils and polypeptides. The results for six or more residues are well approximated by a Gaussian distribution. The solid line represents a Gaussian fit whose width is within 10% of the corresponding Gaussian random coil distribution of the same length.

trated the reactants, the more translational entropy each helix has prior to binding. Thus, the loss of entropy upon introduction of the tether in the undocked state is greater when compared to reactants at lower concentrations. The loss of translational entropy generally is calculated relative to the free state at 1 M concentration, the concentration where free energies for bimolecular systems are calculated (i.e., $\Delta G^{\circ}_{bi} = -RT \ln K_{ass}$). As noted above, the use of 1 M standard state concentration explains the paradoxical result that a twelve-residue tether between two helices, whose average center-to-center distance is 25–30 Å, results in an *increase* in entropy relative to the free species, whose translational entropy is equivalent to the uniform exploration of a cube having a volume of only 1,661 $Å^3 \sim (12.2 Å)^3$. In essence, the helices explore more volume when tethered than they do at 1 M standard state concentration according to the NN method.

For macromolecules, the 1 M standard state concentration is unrealistically high. At a more realistic concentration of $C_{ref} = 1$ μM, the entropic penalty of adding the tether is increased by 26.7 e.u. (Eq. 7), and the introduction of the tether now is very entropically restrictive, as one expects. Further, at this lower concentration, the assumption that the tethered helices dock intramolecularly is more likely to be valid.

In terms of free energy for the denatured helices at 300 K connected by a six-residue tether, ΔG_{uni} is less than ΔG°_{bi} by 3.2 kcal mol^{-1} at 1 M standard state concentration, but is 5 kcal mol^{-1} greater at $C_{ref} = 1$ μM. Hence, the introduction of the tether does increase the population of

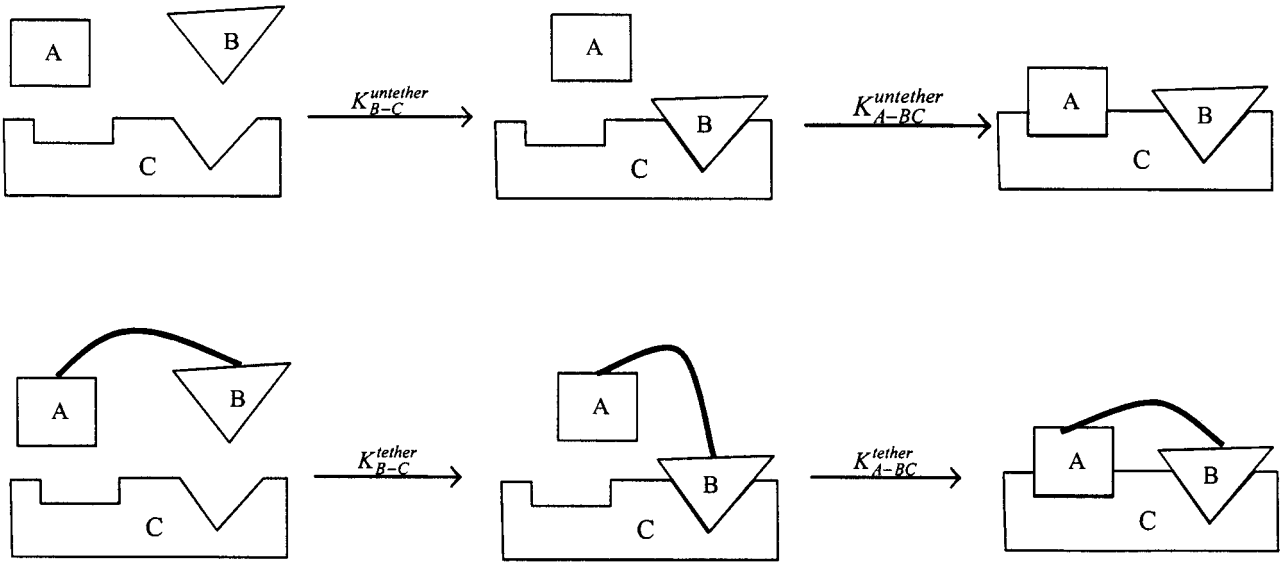

Fig. 4. Higher order reactions. Diagram depicted the sequential binding of three components, A, B, and C when A and B are untethered (upper) and pre-tethered (lower). For the untethered system, the association constant is given by $K_{A\text{-}B\text{-}C}^{untether} = K_{B\text{-}C}^{untether} K_{A\text{-}BC}^{untether} = [A \cdot B \cdot C]/[A][B][C]$ ($K_{B\text{-}C}^{untether}$, $K_{A\text{-}BC}^{untether}$ in units of M^{-1}, $K_{A\text{-}B\text{-}C}^{untether}$ in units of M^{-2}), while the association constant for the corresponding reaction where A and B are tethered, is given by: $K_{A\text{-}B\text{-}C}^{tether} = K_{B\text{-}C}^{tether} K_{A\text{-}BC}^{tether} = [A \cdot B \cdot C]/[A \ldots B][C]$ ($K_{B\text{-}C}^{tether}$, $K_{A\text{-}B\text{-}C}^{tether}$ in units of M^{-1}, $K_{A\text{-}BC}^{tether}$ is dimensionless).

docked species when concentration of the individual helices is below ~ 1 mM.

The tether's effect on stability, and its dependence on the choice of C_{ref}, also can be cast in terms of effective concentration of reactants, C_{eff}, upon the introduction of a tether. Once ΔS values are calculated at a given reference concentration, C_{ref}, the effective concentration upon cross-linking, C_{eff}, can be calculated according to:

$$C_{eff} = C_{ref} e^{\Delta S/R} \tag{8}$$

For six- and eighteen-residue tethers, denatured helices have an effective concentration of 4.1 and 0.57 mM, respectively (Table II). At a reference concentration equal to C_{eff}, the free energy of the bimolecular and tethered systems is equal. When the effective concentration induced by the tether exceeds the dissociation constant, the docked complex is stable, i.e., $K_{eq}^{complex} = [docked]/[undocked] = C_{ref}/K_{diss} > 1$.

Higher Order Reactions

The methodology presented above can be used as a framework for determining the binding constant for a multimeric association after two of the individual components are tethered.[2] The analysis is illustrated with the sequential binding of three individual components, *A, B*, and *C* (Fig. 4). If the tether does not interfere with either binding step (e.g., $K_{B\text{-}C}^{untether} = K_{B\text{-}C}^{tether}$), then the difference in the two processes is reduced to the difference in the second binding step where *A* binds to the complex $B \cdot C$. This situation directly corresponds to that for the docking of the pre-formed helix before and after the introduction of the tether (State 3 ↔ State 5 vs. State 4 ↔ State 6 in Fig. 1). Prior to *A* binding the complex, its effective concentration is altered by the presence of the tether, and consequently its translational entropy is changed. After binding, there is a loop closure penalty associated with the restriction of the ends of the tether. The loss in entropy, $\Delta S_{\mathrm{unbound}}$, upon introduction of the tether between *A* and the complex $B \cdot C$ is the same as that calculated using the NN method for the two helices (although excluded volume issues may become more important). The corresponding loss in entropy in the bound state, $\Delta S_{\mathrm{bound}}$, is the loop closure entropy penalty for the tether. This value depends upon the actual end-to-end distance of the tether in the bound state, which may not be zero, and will depend on the details of the system. The loop closure penalty for Gaussian random coil for a given *r* end-to-end distance (relative to a zero end-to-end distance) is[6,11]

$$\Delta S_{bound} = S(r) - S(0) = -R \ln\left[\frac{P_{RC(r)}}{P_{RC(0)}}\right] = R\beta^2 r^2 \tag{9}$$

where β is defined in Eq. 5.

The net effect of the introduction of the tether is $\Delta S_{\mathrm{tether}} = \Delta S_{\mathrm{bound}} - \Delta S_{\mathrm{unbound}}$. This entropy can be converted to C_{eff} according to Eq. 8. We obtain the association constant of the tethered reaction according to

$$K_{A\text{-}B\text{-}C}^{tether} = C_{eff} K_{A\text{-}B\text{-}C}^{untether} \tag{10}$$

An empirical value of $K_{A\text{-}B\text{-}C}^{tether}$ which equates to a C_{eff} that is higher than predicted by Eq. 10 generally implies that the tether provides an orientational benefit. Conversely, a lower value of C_{eff} implies that the tether may be directly interfering with binding or that it is insufficiently long or flexible, and is strained in the bound complex.

Gaussian Coil Approximation

The simulations of the poly-alanine tether used in the rotational entropy calculation can also be used to test the

validity of the Gaussian random coil approximation for real amino acid tethers. The average end-to-end distance distribution for poly-alanine tethers of various lengths is compared to that for an ideal Gaussian random coil in Figure 3 (*b–d*). A three-residue tether behaves significantly differently than the idealized coil, but a six-residue chain already is a reasonable approximation for a Gaussian coil, both in terms of the mean end-to-end distance and the distribution of end-to-end distances.

Excluded Volume Effects

In order to correct for the excluded volume of the helices and steric overlap, simulations are carried with finite size helices. The 30-Å-long, 10-Å-diameter helices overlap in about 1/3 of the configurations of a six-residue tether. For an eighteen-residue chain, the occurrence decreases to 5% [Fig. 5(a)]. The elimination of these disallowed configurations results in a more extended center-to-center probability distribution. Hence, the entropy of the tethered helices is higher, and the loss of translational entropy is, in fact, reduced. Numerically, this effect results in a decrease of only 0.43 e.u. (2.9% decrease from the original S_{tether}) and 0.12 e.u. (0.7% decrease) in cross-linking entropy for six- and eighteen-residues, respectively.

In order to investigate excluded volume effects due to the steric clashes of non-neighboring residues, we performed hard-sphere simulations similar to Pappu et al.[7] [Fig. 5(b)]. Polyalanine chains were constructed using dihedral angles randomly chosen from all the sterically allowed possibilities in a dipeptide (without restriction to a given region such as α-helical, as was done by Pappu et al.[7]). The number of overlapping chains increases nearly linearly from 7 to 16% going from six to twelve residues. Furthermore, the effect on the end-to-end probability distribution also is very small. As this distribution directly relates to the entropy, excluded volume effects of the tether with itself (and presumably with the helices that extend out in the opposite direction) has a minimal impact on the entropy of tethering.

Comparison to Other Methods

The NN method provides an explanation as to why the Cell Theory of Liquids provides a reasonable result, even though the basic premise, that the solute particles are localized to cells, is invalid. The distribution function for the untethered system, $P^{helix}_{dimer}(r)$, has nearly the equivalent entropy to that of a uniform distribution within volume equivalent to the average volume per solute molecule. Hence, the Cell Theory of Liquids provides a reasonable approximation even though the basis of the method is unrealistic.

Likewise, the NN method for calculating the entropy of a free helix (not the loss of entropy for the tethering process) agrees with the gas phase Sackur-Tetrode Eq. in that a solute particle has nearly the same effective volume per particle in both methods. We point out, however, that the Sackur-Tetrode Eq. for classical particles requires the ad hoc introduction of an extra factor of N^{-1} in the volume per particle to account for the particles' indistinguishability. No such correction is needed in the NN method.

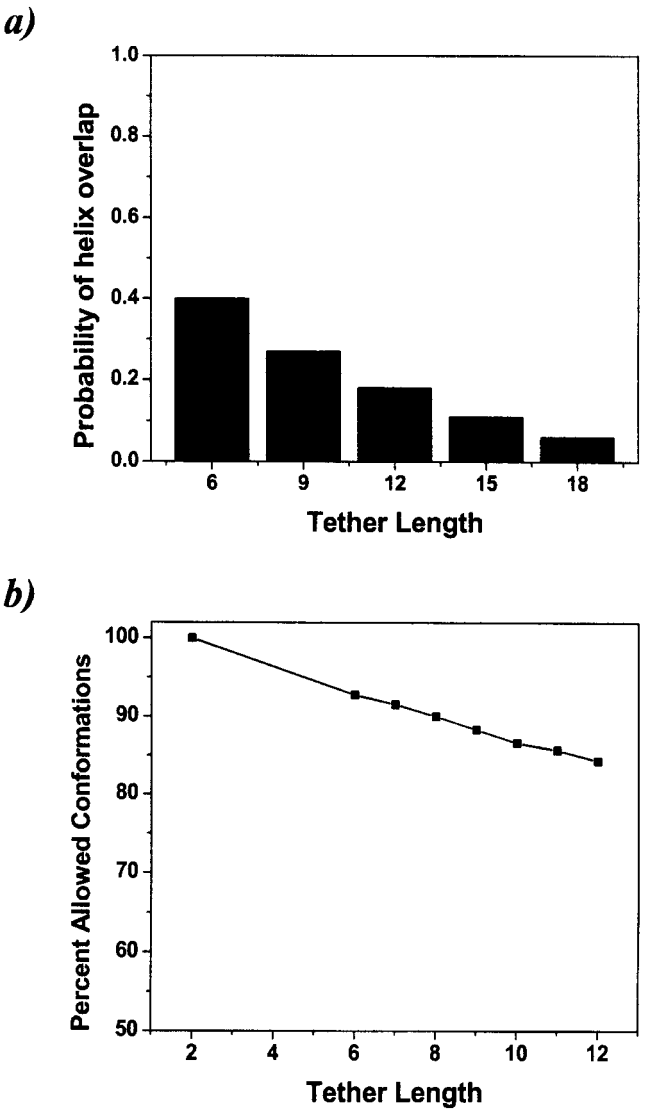

Fig. 5. Excluded volume effects. **a:** The fraction of configurations where the finite sized helices (30 Å long, 10 Å diameter) overlap for different length poly-alanine tethers. **b:** Number of allowed configurations for different length poly-alanine chains. The steric clashes of non-nearest neighbors were calculated using the hard-sphere model described in the text.

Amzel[3] as well as Mammen et al.[4] have presented similar corrections to the Sakur-Tetrode Eq. to account for restrictions due to the finite volume occupied by solvent molecules. The effective or free volume of the solute is the volume the center of the solute can sample without bumping into a solvent molecule. Using this method, Amzel[3] accurately estimates the entropy of liquid water. Amzel[3] also predicts the loss of entropy upon binding (Process $4 \leftrightarrow 6$) by comparing the free volume for the solute in solution to the solute's free volume when it is bound to another protein.

Mammen et al.[4] accurately estimated the standard entropy of condensation of monatomic gases using the concept of free volume. They also estimate the (maximum) loss of entropy for a binding reaction of a tethered system

TABLE III. Comparison Between Experiment and NN Method†

Protein (structure)	Experimental ΔS ($\Delta G^{\circ}_{bi} - \Delta G_{uni}$)	Predicted ΔS ($\Delta G^{\circ}_{bi} - \Delta G_{uni}$)
GCN4 coiled-coil[17]	~0	−10.66
(α-helical)	(~0)	(3.2)
Designed coiled-coil[19]	5 ± 8	−3.4
(α-helical)	(−1.5 ± 2.4)	(1.0)
SSI[20] (α/β)	−5 ± 4	−7.20
	(1.5 ± 1.2)	(2.2)
Arc repressor[21] (α/β)	−11.8 ± 0.5	−13.4
	(3.5 ± 0.2)	(4.0)

†Entropy is given in e.u., stability in kcal mol^{-1}, calculated according to $\Delta G = -T\Delta S$ at 300 K.

(Process 4 ↔ 6), as distinct from our calculation where entropic benefit of introducing the tether is calculated (Process 4 ↔ 6 relative to Process 3 ↔ 5). Regardless, our calculation is similar in that the loss of entropy is based upon a comparison of the volume accessible prior to docking to that available afterwards. However, certain details are different; for example, the present analysis is based upon probability distribution functions of the tether.

Another method used to calculate the translational entropy of binding or cross-linking in aqueous media is the empirical quantity "cratic entropy" (mixing entropy) given by:

$$\Delta S = -R \ln 55 \qquad (11)$$

The value of 55 reflects the ratio of the concentration of 1 M solute to the molarity of pure water. Cratic entropy was proposed by Gurney[12] and Kauzmann[13] to explain the entropy of mixing of ideal solutes in pure solutions. This correction is not derived from any princples of thermodynamics[14] or statistical mechanics.[15,16]

Comparison to Experiments

We have compared the NN method to the results for proteins where the entropy or the free energy between the monomeric and dimeric versions has been measured (Table III).

For variants of the dimeric GCN4-p1 coil coiled and analogs cross-linked with a disulfide bridged, Cys-Gly-Gly amino terminal tether, the difference between the dimeric and unimolecular stability was approximately zero.[17] For this system, however, the NN method predicts the dimeric stability is stronger by 3.2 kcal mol^{-1} for the difference in the stability of the two versions. The excess stabilization of the tethered species may reflect stabilizing interactions between the tether and the helices. The amount of denaturant sensitive surface buried in the native structure (the *m*-value) was increased by about 10% in the tethered version. Potentially, the helix was capped by the glycines in the tether, which stabilized additional helical structure.[18]

Yu and coworkers have performed detailed calorimetric measurements of the stability of homodimeric coiled coil and a version cross-linked with an internal disulfide bond.[19] They showed that the entire change in free energy upon cross-linking was due to the loss of entropy upon introduction of the tether. The heat capacity of these molecules was not affected by the cross-link and, therefore, the vibrational modes were not perturbed. Also, the $\Delta H_{folding}$ did not change upon introduction of the tether. The pre-docked conformation is essentially fully denatured in this system, thus the calculated loss of entropy is for the difference in ΔS between Processes 1 ↔ 2 and 5 ↔ 6. For a system with just an internal disulfide bond, the loop closure entropy in the folded complex is minimal (~1 e.u.). Therefore, the entire entropy change probably is only due to the tethering process in the denatured state. For such a system, according to the NN method, the loss of entropy is 3.4 e.u., which is in good agreement with the experimental results (5 ± 8 e.u.).

In a very similar study on Streptomyces subtilisin inhibitor (SSI), Tamura and Privalov[20] measured the loss of entropy upon cross-linking by calorimetric and magnetic resonance methods. For this homodimeric system, cross-linked by only a single internal disulfide bond, the NN method predicts the loss of entropy to be −7.20 e.u., which is close to the experimentally determined value of −(5 ± 4) e.u.

Robinson and Sauer[21] examined the effect of a 15-residue cross-link on stability and folding kinetics on dimeric are repressor. The cross-link connects the C-terminus of one Arc subunit to the N-terminus of the second subunit (Arc-L1-Arc). Comparison of the equilibrium stabilities of the linked and unlinked proteins yielded a C_{eff} = 2.7 ± 0.7 mM. We model the system by taking into account the length of the cross-link and the end-to-end distance between the C-terminus of one Arc subunit and the N-terminus of the second subunit. For the Arc-L1-Arc system, the NN method predicts C_{eff} = 1.2 mM, or a difference of 1.6 e.u. from the observed value.

CONCLUSION

A clear distinction exists between binding and cross-linking, two processes that are often considered to be equivalent. We have presented a method to calculate the entropic benefit of cross-linking. The major underpinning of the method is the realization that the probability distribution for the dimeric system represents the probability that a partner can travel a given distance while still being the closest molecule to a reference helix. Also, by comparing entropies based upon probability distributions, the method is independent of the nature of the solvent. We have also shown that the contribution of rotational entropy is relatively negligible for a reasonable length tether. We have also demonstrated the applicability of our method to a wide variety of protein systems. The methods outlined in this paper can be applied to higher order association processes and nucleic acid hairpins, although the exact results should be sequence dependent. For dimeric proteins, the introduction of a ten-residue tether results in an effective concentration of reactants in the millimolar range. When the concentration of the dimeric system is less than this concentration, the introduction of the tether will

increase the fraction of docked species and, generally, result in a net stabilization for systems with mM or stronger dissociation constants.

NOTE ADDED IN PROOF

H. Zhou (J Amer Chem Soc 2001;123:6730–6731)[22] calculated the effect of tethering using only the end-to-end distribution function of a worm-like chain in the tethered state: $C_{eff} = P(r)/N_{\mathrm{Avagodro}}$. Although this analysis does not consider the loss of entropy due to introduction of the tether, the predicted values are in good agreement with experiment.

ACKNOWLEDGMENTS

We sincerely thank Drs. S. Rice, E. Shakhnovich, M. Amzel, K. Sharp, K. Freed, T. Pan, N. Kallenbach, and A. Fernandez and members of our groups for many enlightening discussions. This work was supported by grants from the NIH (T.R.S.) and Packard Foundation Interdisciplinary Science Program (T.R.S., R.S.B.). M.H.Z. is a Burroughs-Wellcome Interdisciplinary Research Fellow.

REFERENCES

1. Jencks WP. Binding energy, specificity, and enzymic catalysis: the circe effect. Adv Enzymol 1975;43:219–410.
2. Jencks WP. On the attribution and additivity of binding energies. Proc Natl Acad Sci USA 1981;78:4046–4050.
3. Amzel LM. Loss of translational entropy in binding, folding, and catalysis. Proteins 1997;28:144–149.
4. Mammen M, Shakhnovich EI, Deutch JM, Whitesides GM. Estimating the entropic cost of self-assembly of multiparticle hydrogen-bonded aggregates based on the cyanuric acid, Melamine lattice. J Org Chem 1998;63:3821–3830.
5. Barker JA. Lattice theories of the liquid state. New York: MacMillan Press; 1963. 133 p.
6. Flory PJ. Statistical mechanics of chain molecules. Ithaca, NY: Cornell University Press; 1953, 464 p.
7. Pappu RV, Srinivasan R, Rose GD. The flory isolated-pair hypothesis is not valid for polypeptide chains: Implications for protein folding. Proc Natl Acad Sci USA 2000;97:12565–12570.
8. Brady GP, Sharp KA. Energetics of cyclic dipeptide crystal packing and solvation. Biophys J 1997;72:913–927.
9. Jacobson H, Stockmayer WH. Intramolecular reactions and polycondensation: I. The theory of linear systems. J Chem Phys 1950;18:1600–1606.
10. Jacobson H, Beckmann CO, Stockmayer WH. Intramolecular reaction in polycondensations. II. Ring-chain equilibrium in polydecamethylene adipate. J Chem Phys 1950;18:1607–1612.
11. Cantor CR, Schimmel PR. Biophysical chemistry: part III. New York: W. H. Freeman and Co.; 1980. p 863–864.
12. Gurney RW. Ionic processes in solution. New York: McGraw-Hill; 1953. 275 p.
13. Kauzmann W. Some factors in the interpretation of protein denaturation. Adv Protein Chem 1959;14:1–63.
14. Holtzer A. The "cratic correction" and related fallacies. Biopolymers 1995;35:595–602.
15. Janin J. For guldberg and waage, with love and cratic entropy. Proteins 1996;24:i–ii.
16. Gilson MK, Given JA, Bush BL, Mccammon JA. The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J 1997;72:1047–1069.
17. Moran LB, Schneider JP, Kentsis A, Reddy GA, Sosnick TR. Transition state heterogeneity in gcn4 coiled coil folding studied by using multisite mutations and crosslinking. Proc Natl Acad Sci USA 1999;96:10699–10704.
18. Krantz BA, Moran LB, Kentsis A, Sosnick TR. D/h amide kinetic isotope effects reveal when hydrogen bonds form during protein folding. Nature Struct Biol 2000;7:62–71.
19. Yu YB, Lavigne P, Kay CM, Hodges RS, Privalov PL. Contribution of translational and rotational entropy to the unfolding of a dimeric coiled-coil. J Phys Chem B 1999;103:2270–2278.
20. Tamura A, Privaloy PL. The entropy cost of protein association. J Mol Biol 1997;273:1048–1060.
21. Robinson CR, Sauer RT. Equilibrium stability and sub-millisecond refolding of a designed single-chain are repressor. Biochemistry 1996;35:13878–13884.
22. Zhou H. Single-chain versus dimeric protein-folding: thermodynamic and kinetic consequences of covalent linkage. J Am Chem Soc 2001;123:6730–6731.