BIOCHEMISTRY

 Volume 43, Number 8 March 2, 2004

Current Topics

Polymer Models of Protein Stability, Folding, and Interactions[†]

Huan-Xiang Zhou*

Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306

Received December 17, 2003

ABSTRACT: The unfolded state and flexible linkers in the folded structure play essential roles in protein stability and folding and protein−protein interactions. Intrinsic to these roles is the fact that unfolded proteins and flexible linkers sample many different conformations. Polymer models may capture this and complement experiments in elucidating the contributions of the unfolded state and flexible linkers. Here I review what can be predicted from these models and how well these predictions match experiments. For example, Gaussian chain models give quantitatively reasonable predictions of the effects of residual charge−charge interactions in the unfolded state and qualitatively reasonable results for the effects of spatial confinement and macromolecular crowding on protein stability. A wormlike chain model has met with success in quantifying the effects of flexible linkers in binding affinity enhancement and in regulatory switches. In future developments, more realistic models may emerge from molecular dynamics simulations, and these models will guide experiments to advance our understanding of the unfolded state and flexible linkers.

Great insight into the specific interactions that stabilize the folded structure of a protein can be gained from its structure determined by X-ray or NMR.[1] These interactions have been the focus of most discussions about the determinants of protein stability. However, appreciation of the role of the unfolded state has increased over the years. It is now not uncommon to see residual interactions in the unfolded state demonstrated or invoked (*1*, *2*). Of course, the unfolded chain is the starting point of protein folding. An elemental step in the folding process may be the formation of a contact between two residues that are far apart along the unfolded chain (*3*). Some proteins have been found to be unfolded under native conditions. Even in the folded state, linkers connecting individual domains may be quite flexible. Disordered proteins or disordered regions have recently attracted much attention (*4*−*6*).

It is difficult to completely characterize an unfolded protein (or a flexible linker) in experiments (*7*). By its very nature, an unfolded protein chain or a flexible linker samples many different conformations. Therefore, only distribution functions, not discrete conformations, can provide adequate descriptions. Simple models from polymer theory can capture the essence of the distribution functions without being bogged down by atomic details. Indeed, polymer models have yielded physical insight into and produced reasonable quantitative results for the roles of the unfolded state and flexible linkers in protein stability and folding and protein−protein interactions (*2*, *3*, *8*−*15*). The purpose of this paper is to present an overview of these advances.

[†] This work was supported in part by Grant GM58187 from the National Institutes of Health.

* To whom correspondence should be addressed: Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306. Phone: (850) 645-1336. Fax: (850) 644-7244. E-mail: zhou@sb.fsu.edu.

[1] Abbreviations: NMR, nuclear magnetic resonance; BPTI, bovine pancreatic trypsin inhibitor; FnFn3(10), tenth fibronectin type III domain of human fibronectin; drk-SH3, N-terminal SH3 domain of drk; DnaB-(24−136), residues 24−136 of the N-terminal domain of DnaB; RPA70AB, central region of the 70 kDa subunit of replication protein A; HEL, hen egg white lysozyme.

10.1021/bi036269n CCC: $27.50
Published on Web 02/07/2004

Chain Statistics of the Unfolded State

Regardless of the existence of residual interactions, the sampling of vastly different conformations remains an essential property of an unfolded protein. The simplest model for a flexible polymer chain has a Gaussian probability density for the displacement vector between any two residues:

$$p_G(\mathbf{r}) = (3/2\pi\langle r^2\rangle)^{3/2} \exp(-3r^2/2\langle r^2\rangle) \quad (1)$$

The mean square distance in this model only depends on the number of peptide bonds between the end residues; specifically, $\langle r^2\rangle = b^2 l$. The increase in $\langle r^2\rangle$ with l reflects the statistical behavior by which, relative to residues closer on the chain, residues far apart on the chain will have increasingly higher probabilities of sampling long distances. The Gaussian chain model gives a better description of the statistics of residues with a large chain separation.

The effective bond length b is a model parameter, which can be fixed by experimental data on the radius of gyration or hydrodynamic properties such as Stokes radius or intrinsic viscosity. For a Gaussian chain with N residues, the radius of gyration $R_g = (\langle R^2\rangle/6)^{1/2} = b(N/6)^{1/2}$, where $\langle R^2\rangle$ is the mean square end-to-end distance. According to the Zimm model (*16*), the Stokes radius and intrinsic viscosity of a Gaussian chain are

$$R_s = KbN^{1/2}; [\eta] = \Phi b^3 N^{3/2} \quad (2)$$

It should be noted that these results were derived for extremely long chains. Similar results had previously been obtained by Kirkwood and Riseman (*17*).

An important factor in the chain statistics of an unfolded protein is excluded-volume interactions, both local (within a residue and between neighboring residues) and nonlocal (between residues farther apart on the chain). Local excluded-volume interactions play a dominant role in restricting backbone torsional angles of a polypeptide chain to particular regions on the Ramachandran map (*18*). Gaussian statistics are retained for residues with large chain separations even in the presence of local excluded-volume interactions. However, the probability density may be distorted by nonlocal excluded-volume interactions. An approximate way to account for nonlocal excluded-volume interactions is to keep the Gaussian distribution of eq 1 but modify the mean square distance ($\langle r^2\rangle$) to equal $b'^2 n^{1+\epsilon}$ (*19*). The radius of gyration (R_g) then becomes equal to $b'[N^{1+\epsilon}/(2+\epsilon)(3+\epsilon)]^{1/2}$. The Stokes radius and intrinsic viscosity for extremely long chains are now given by

$$R_s = K'b'N^{1+\epsilon}/2; [\eta] = \Phi' b'^3 N^{3(1+\epsilon)}/2 \quad (3)$$

Decades ago, Tanford and co-workers (*20*, *21*) measured the Stokes radii and intrinsic viscosities of fully denatured proteins in GuHCl. These data are in qualitative agreement with eq 3. However, quantitative analysis of the data according to eq 3 is questionable, since this equation strictly applies to only extremely long chains and values for the K' and Φ' coefficients have large uncertainties. In recent years, Stokes radii and radii of gyration have been determined for a large number of denatured proteins and peptide fragments by modern techniques (*7*, *22*−*31*).

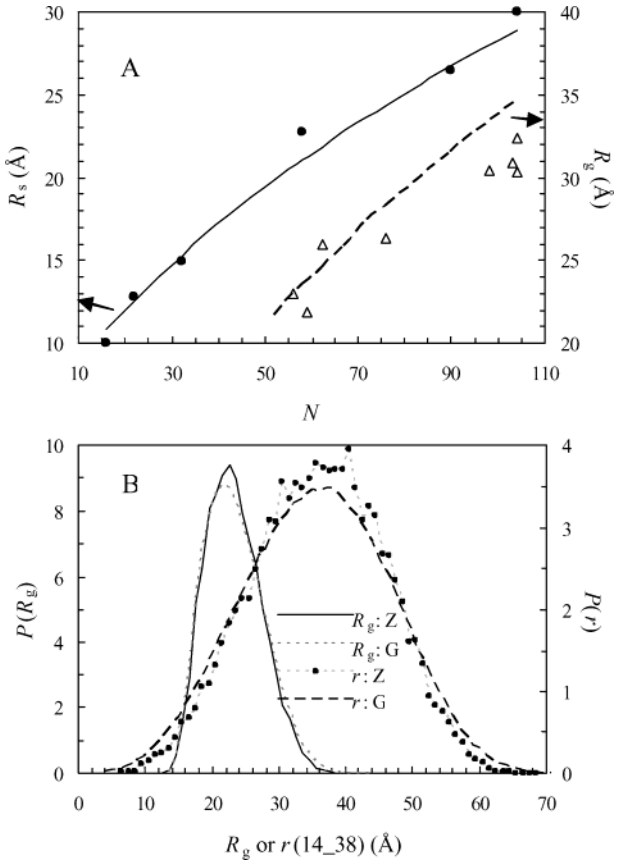

FIGURE 1: (A) Comparison of calculated (curves) and experimental (symbols) results for Stokes radius R_s and radius of gyration R_g. Calculated results for R_s are from ref *32*, and the chain conformations used for R_s give the displayed results for R_g. Experimental results for R_s are from refs *22*−*24*, and those for R_g are from refs *25*−*31*. (B) Comparison of distribution functions for the radius of gyration and the distance between residues 14 and 38 in BPTI obtained in two studies. Curves labeled Z are from ref *32*, and curves labeled G are from ref *33*.

Two recent studies explored the conformations sampled by unfolded proteins. In the first (*32*), chain conformations were generated using backbone torsion angles collected in a database of loop residues. As noted earlier, these torsion angles account for local excluded-volume interactions. Nonlocal excluded-volume interactions were introduced through hard-sphere repulsions between C_α atoms. Sixteen chains, with 16−104 residues, were studied. The hard-sphere diameters of C_α atoms were adjusted to reproduce experimental values of intrinsic viscosity. The resulting values for the Stokes radius and radius of gyration are in good agreement with experimental data (Figure 1A). The effective bond length, defined as $b = (\langle R^2\rangle/N)^{1/2}$, ranges from 5.7 Å when $N = 16$ to 8.8 Å when $N = 104$.

In the second study, Goldenberg (*33*) generated conformations of four protein chains (with 26−268 residues) by randomly sampling torsion angles. Excluded-volume interactions were included by a soft potential: $U_{ij} = (r_{ij}/r_{ij}^0 - 1)^2/4$ if interatomic distance r_{ij} is less than r_{ij}^0, the sum of adjusted van der Waals radii. Conformations thus generated were also able to reproduce experimental data on the Stokes radius and radius of gyration for denatured proteins. Despite the different approaches, chain statistics obtained in the two studies appear to show striking similarity. As Figure 1B shows, the distribution functions for the radius of gyration and the

distance between residues 14 and 38 calculated for the 58-residue BPTI by the two studies match well. Since these chain conformations share global properties (e.g., radius of gyration) with denatured proteins, however, one may question how well they represent the unfolded state under physiological conditions (*7*).

Wormlike Chain Model for Flexible Linkers

It was recognized in early studies of DNA hydrodynamics that the Gaussian chain model is inadequate for the statistics of residues close along the chain (*34*). At a short contour length, the direction of the chain has a persistent memory. In addition, DNA and polypeptide chains are practically inextensible so the total contour length is fixed for a given chain. These properties are captured by the continuous wormlike chain model of Kratky and Porod (*35*). The tangent vector at each point along the wormlike chain has a unit magnitude (to preserve the contour length) and shows correlation over a contour length of l_p (the persistence length). The mean square distance between two points spanning a contour length l_c is

$$\langle r^2 \rangle = 2l_p l_c - 2l_p{}^2[1 - \exp(-l_c/l_p)] \quad (4)$$

The contour length of a chain consisting of l peptide bonds is approximately $b_0 l$, where b_0 (3.8 Å) is the average distance between two adjacent C_α atoms. For long chains in which $l_c \gg l_p$, eq 4 reduces to $\langle r^2 \rangle = 2l_p l_c = 2b_0 l_p l$, a result consistent with the Gaussian chain model. On the other hand, when $l_c \ll l_p$, eq 4 reduces to $\langle r^2 \rangle = l_c{}^2$, a result expected for a rigid rod. No closed-form expression has been found for the probability density of the displacement vector, but a good approximation is (*36*)

$$p(\mathbf{r}) = (3/4\pi l_p l_c)^{3/2} \exp(-3r^2/4l_p l_c)[1 - w(r;l_c)] \quad (5a)$$

$$w(r;l_c) = 5l_p/4l_c - 2r^2/l_c{}^2 + 33r^4/80l_p l_c{}^3 + 79l_p{}^2/160l_c{}^2 + 329r^2 l_p/120l_c{}^3 - 6799r^4/1600l_c{}^4 + 3441r^6/2800l_p l_c{}^5 - 1089r^8/12800l_p{}^2 l_c{}^6 \quad (5b)$$

A convenient way to calculate $p(\mathbf{r})$ numerically is based on the fact that the wormlike chain is the continuum limit of the freely rotating chain (with the bond length $\rightarrow 0$ and the bond angle $\rightarrow \pi$). Interest in the wormlike chain model was revived by recent single-molecule experiments in which forces are applied to stretch DNAs or unfold proteins (discussed in the next section).

Many proteins have a modular design, with individually folded domains connected by flexible linkers. The flexibility may be reflected as missing atoms in X-ray structures (Figure 2). It was recently proposed that flexible linkers may be modeled as wormlike chains (*37*). Furthermore, it was suggested that loops connecting regular secondary structures (α-helices and β-strands) in protein structures represent conformations of a wormlike chain. Figure 3A−C displays the distribution functions of the end-to-end distances of loops with given chain lengths. They agree well with the predictions of a wormlike chain with a persistence length (l_p) of 3.04 Å. This procedure of obtaining the value of the persistence length of peptide linkers is similar in spirit to measurement of the persistence lengths of DNA and a disordered protein through direct visualization of the molecules by electron or atomic force microscopy (*38*−*40*).

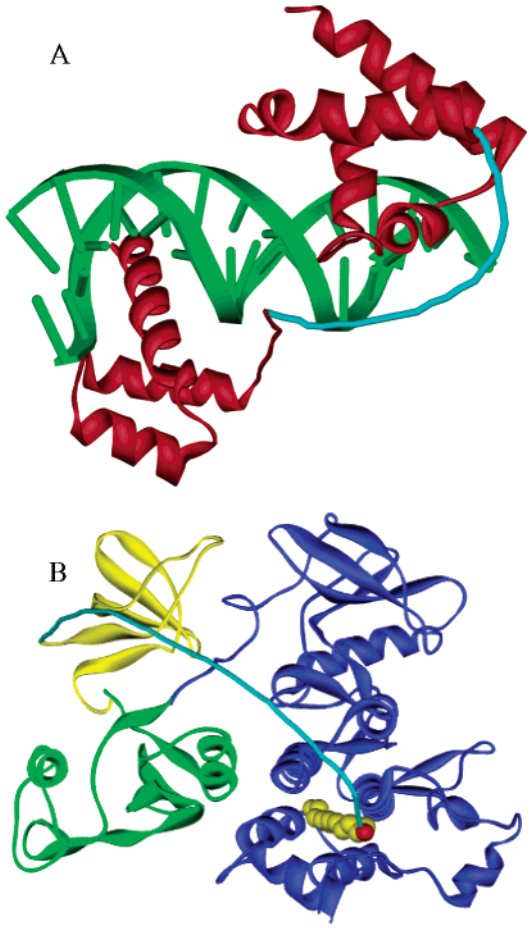

FIGURE 2: Flexible linkers (A) between the POU-specific domains and homeodomains of DNA-binding protein Oct-1 POU (*112*) and (B) between the SH3 domain (yellow) and a myristoyl group bound to the kinase domain (blue) of the tyrosine kinase c-Abl (*120*). The linkers are missing in the X-ray structures and are shown for illustration in cyan. C-Abl also has relatively rigid linkers between the SH3 domain and the SH2 domain (in green) and between the SH2 domain and the kinase domain.

Further insight into the wormlike chain model for protein loops and linkers can be obtained by examining the end-to-end distances of short peptides with conformations generated from backbone torsion angles of loop residues. Figure 3D−F shows that the distribution functions of these short peptides are also close to the wormlike chain model, with a somewhat larger persistence length of 4 Å. (Note that excluded-volume interactions between residues in such short peptides are insignificant.) As will be discussed next, the persistence length of 4 Å is in agreement with experimental results from mechanically unfolding a number of proteins. Two factors may contribute to the smaller l_p value of 3.04 Å found for protein loops. (i) The end residues of loops are also start or end residues of regular secondary structures. (ii) Excluded-volume interactions between a loop and the remaining folded structure may favor more compact conformations (see below). When the end residues are moved away from regular secondary structures, such loop fragments behave like a wormlike chain with a persistence length of 4 Å (Figure 3D−F). For flexible linkers connecting folded domains, the use of an l_p of 3.04 Å seems appropriate.

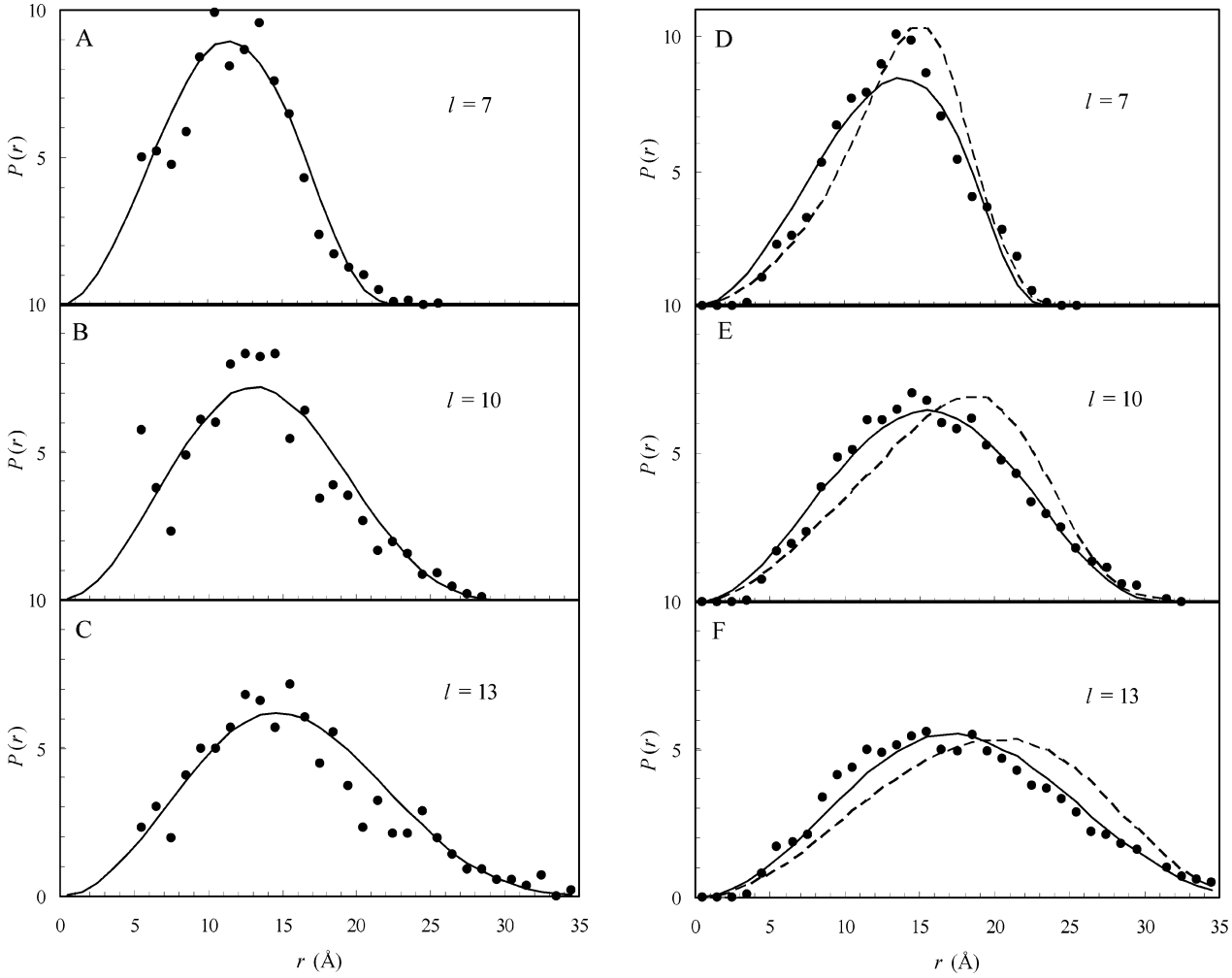

FIGURE 3: Distribution functions for the end-to-end distances of protein loops (*37*). The distribution function $P(r)$ for the end-to-end distance r is related to the probability density $p(\mathbf{r})$ of the end-to-end vector $\mathbf{r}$ via $P(r) = 4\pi r^2 p(\mathbf{r})$. Loops are defined as residues connecting regular secondary structures, and the end residues of a loop are also start and/or end residues of secondary structures. In panels A−C, symbols represent results calculated from protein loops with a given loop length l, and solid curves are predicted by the wormlike chain model with an l_p of 3.04 Å. In panels D−F, dashed curves are obtained from short peptide with backbone torsion angles of loop residues, solid curves are predicted by the wormlike chain model with an l_p' of 4 Å, and symbols are results calculated from loop fragments in loops with lengths ranging from 11 to 30 peptide bonds. Loop fragments do not include the first two and the last two residues of a loop.

Mechanical Unfolding of Proteins

The development of single-molecule techniques such as atomic force microscopy has provided new insight into the behavior of unfolded proteins. When a protein is unfolded by a mechanical force F, the chain can be extended by the force (Figure 4). When $F \rightarrow 0$, the extension z along the direction of the force is proportional to F. When $F \rightarrow \infty$, z approaches the contour length l_c. The force-extension curve contains information about the flexibility of the unfolded chain. For a wormlike chain under stretching, Marko and co-workers obtained a simple interpolation formula for the force-extension relation (*41*, *42*):

$$l_p F/k_B T = z/l_c + 1/4(1 - z/l_c)^2 - 1/4 \qquad (6)$$

where k_B is the Boltzmann constant and T the absolute temperature. Equation 6, initially used for analyzing DNA stretching data, has now been widely used to fit data for mechanically unfolding proteins (*40*, *43*−*54*). In most studies, the determined persistence lengths of unfolded proteins ranged from 3.3 to 4.2 Å (*43*−*46*, *49*−*54*). In a few studies (*40*, *47*, *48*), values of l_p in the range of 6.5−8 Å were reported. Chains with partial structures may have even longer apparent persistence lengths. It appears 4 Å has now been taken as a standard value for l_p in fitting single-protein mechanical unfolding data.

If a 90-residue unfolded protein is modeled as a wormlike chain with an l_p of 4 Å, then one obtains a value of 5.5 Å for the effective bond length defined as $b = (\langle r^2 \rangle / l)^{1/2}$, where $\langle r^2 \rangle$ is given by eq 4 and $l = 90$. This value of b is consistent with a chain without excluded-volume interactions and is significantly smaller than the value of ∼8.5 Å determined by fitting to hydrodynamic data (*32*). A possible explanation for this discrepancy is that, in a typical mechanical unfolding experiment, the fitting of the force-extension curve to eq 6 is dominated by the range of force above 50 pN (see Figure 4B). Under such large forces, the chain is extended so the chance for excluded-volume interactions is reduced. Under smaller forces, the chain is expected to coil up (Figure 4A) and thus encounter significant excluded-volume interactions. Rief et al. (*47*) found that the force-extension curve of titin Ig domains in the force range below 50 pN fitted to eq 6 better with an l_p of 8 Å, although the full curve fitted to eq

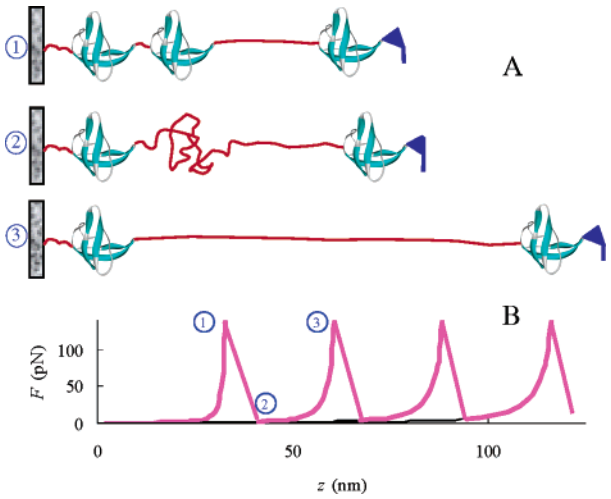

FIGURE 4: Single-protein unfolding by atomic force microscopy. (A) Illustration of the sequence of events in a stretching experiment. (1) One domain is unfolded and fully extended. (2) A second domain is just unfolded. (3) Both unfolded domains are fully extended. (B) Corresponding force-extension curve.

6 better with an l_p of 4 Å. For spectrin repeats, it took less than 50 pN to unfold each repeat, and fitting of the full force-extension curve to eq 6 gave an l_p of 8 Å, a relatively high value noted above.

Residual Charge−Charge Interactions in the Unfolded State

The pH dependence of the folding stability provides a probe of the unfolded state. Tanford (*55*) derived the following relation:

$$\Delta G(\mathrm{pH}) - \Delta G(\mathrm{pH}_0) = (k_B T \ln 10)\int_{\mathrm{pH}_0}^{\mathrm{pH}} Q_u \, d\mathrm{pH} - (k_B T \ln 10)\int_{\mathrm{pH}_0}^{\mathrm{pH}} Q_f \, d\mathrm{pH} \quad (7)$$

where ΔG(pH) is the unfolding free energy at a given pH and Q_f and Q_u are the net charges on the folded and unfolded protein, respectively. Since ΔG(pH) and Q_f can be determined experimentally, eq 7 can be used to calculate Q_u, which contains information about residual charge−charge interactions in the unfolded state.

For many years, it was assumed that such residual charge−charge interactions are negligible. Dill and co-workers (*56*) explicitly included charge−charge interactions in the unfolded state in developing a model for protein folding. The earliest experimental evidence for residual electrostatic effects was presented by Fersht and co-workers (*57*), who studied the effects of salt on the pH dependence of the folding stability of barnase. In a subsequent study (*58*), they determined the pK_a values of ionizable groups in the folded state, from which Q_f can be calculated. It was found that eq 7 could be satisfied only if Q_u was calculated by assuming pK_a values lower than model compound values. The pH dependence of the folding stability along with information for Q_f has now been determined for a number of other proteins (*59*−*66*). Elcock (*67*) made the first quantitative prediction about the contribution of residual electrostatic effects to the pH dependence of folding stability. He modeled the unfolded state by one structure partially unfolded from the folded structure (obtained by molecular dynamics simulations with expanded van der Waals radii) and was able to reproduce some of the experimental pH dependence of ΔG.

It has been recognized that the unfolded protein must sample different conformations so that distances between charges have a broad distribution rather than being fixed (*2*). At a particular distance r, the energy of interaction between two charges q_i and q_j in the unfolded state may be given by the Debye−Hückel potential $U_{ij}(r) = q_i q_j \exp(-\kappa r)/\epsilon r$, where κ is the ion screening parameter and ϵ is the dielectric constant of the solvent. If the probability density for the displacement vector between the charges is $p(\mathbf{r})$, then the average energy of interaction is

$$W_{ij} = \int d^3\mathbf{r} p(\mathbf{r}) U_{ij}(r) \quad (8a)$$

If the chain obeys Gaussian statistics, then eq 8a becomes

$$W_{ij} = q_i q_j (6/\pi)^{1/2}[1 - \pi^{1/2} x \exp(x^2)\,\mathrm{erfc}(x)]/\epsilon\langle r^2\rangle^{1/2} \quad (8b)$$

where $x = \kappa\langle r^2\rangle^{1/2}/6^{1/2}$. This interaction energy shifts the pK_a values of ionizable groups in the unfolded state and determines the net charge on the unfolded chain at a given pH.

The Gaussian chain model for treating residual charge−charge interactions has been used to rationalize the pH dependence of folding stability for a number of proteins (*2*, *68*, *69*). In all these applications, the root-mean-square distance between two charges was set to $\langle r^2\rangle^{1/2} = bl^{1/2} + s$, where l is the chain separation, b is the effective bond length fixed at 7.5 Å, and s is a shift distance fixed at 5 Å, purportedly to account for the fact the charges are located on side chains (as opposed to the backbone). By the chain statistics, residues close along the sequence are more likely to sample short distances, and thus, their electrostatic interactions are dominating. The quantitative account of the pH dependence of folding stability for the various proteins by the Gaussian chain model suggests that residual electrostatic effects may indeed be attributable to nonspecific interactions dominated by charges close along the sequence.

Kay and co-workers (*66*, *70*) recently determined the pK_a values of Asp, Glu, and His residues in both the folded and unfolded states of drk-SH3 and directly confirmed the validity of eq 7. The pK_a values provided unprecedented information for critically testing the Gaussian chain model. The model was found (*71*) to reproduce well the key experimental results: (i) pK_a shifts for the Asp and Glu residues from model compound values were small, and (ii) Glu2 and Asp8 had the largest pK_a shifts, 0.32 and 0.25 pH unit, respectively, in the downward direction. The calculations suggested that the two downward pK_a shifts were primarily due to interactions with sequentially neighboring positive charges on the N-terminus, Lys6, and His7. It was further demonstrated that repulsions among neighboring charges along the unfolded chain may contribute to folding stability, a mechanism initially advanced in a study of electrostatic contributions to the stability of a thermophilic cold shock protein (*72*).

Figure 5A displays the pH dependence of the unfolding free energy ΔG for FnFn3(10) predicted by the Gaussian chain model. Inclusion of the residual charge−charge interactions significantly improves the agreement with

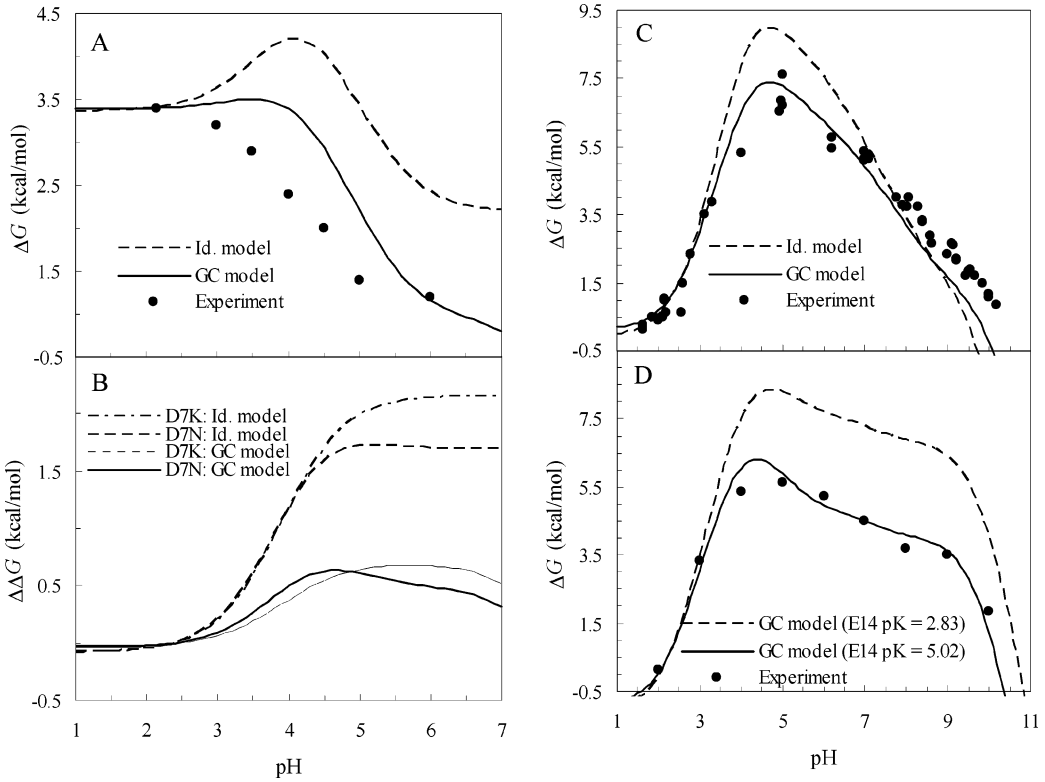

FIGURE 5: pH dependence of the unfolding free energy for (A) FnFn3(10), (B) its Asp7Lys and Asp7Asn mutants, (C) ribonuclease Sa, and (D) its 5K mutant. Symbols are experimental results (*64*, *73*), and curves are predictions by the Gaussian chain model (GC model) or an idealized model (Id. model) neglecting residual charge−charge interactions. In panel B, the reference is the unfolding free energy at pH 2.4.

experimental results of Koide et al. (*64*). These authors determined the pK_a values of all eight Asp and Glu residues of FnFn3(10) in the folded state, and found that five Asp or Glu residues have downshifted pK_a values but Asp7, Glu9, and Asp23 have elevated pK_a values. The latter three residues cluster on the surface of the folded structure, and their repulsions perhaps contribute to the elevated pK_a values. When residual charge−charge interactions in the unfolded state are neglected, ΔG from pH 2 to 4 is incorrectly predicted to increase. The increase is a result of the five downshifted pK_a values in the folded state (leading to a lower than expected Q_f from model compound pK_a values). The Gaussian chain model predicts downshifted pK_a values for all eight Asp and Glu residues, ranging from 0.04 pH unit for Glu9 to 0.17 pH unit for Asp80. The downshifts are due to sequentially neighboring positive charges (e.g., Asp7 next to Arg6 and Asp80 next to Arg78). They are the reason for the disappearance of the increase in ΔG with pH.

Experimental results of Koide et al. on the Asp7Lys and Asp7Asn mutants lend further support to the Gaussian chain model. With these mutations, the pK_a values of Glu9 and Asp23 are no longer significantly elevated in the folded state (perhaps by elimination of repulsions with Asp7). Now, except for Glu9 with a slight upshift in pK_a, all other Asp and Glu residues show a modest downshift in pK_a in the folded state. On the other hand, all Asp and Glu residues are predicted to have small downshifts in pK_a in the unfolded state. As a result, ΔG is predicted to have only a weak dependence on pH (Figure 5B). This prediction is in agreement with the experimental finding that the folding stability of each mutant is nearly identical at pH 2.4 and 7. When residual charge−charge interactions in the unfolded state are neglected, a significant increase in ΔG is predicted (Figure 5B).

The Gaussian chain model can be used to check experimental results for pH-dependent properties. Pace and co-workers (*65*, *73*) determined the pH dependence of ΔG for ribonuclease Sa and the pK_a values for all ionizable groups in the folded protein. The pH dependence of ΔG is reasonably well predicted by the Gaussian chain model (Figure 5C). Pace and co-workers made a mutant called 5K, in which five exposed Asp and Glu residues are replaced with lysines. The predicted pH dependence of ΔG for the 5K mutant now shows a significant discrepancy with the experimental results of Pace and co-workers (Figure 5D). One of the pK_a values, for Glu14, was found by these authors to shift from 5.02 in the wild-type protein to 2.83 in the 5K mutant (*65*). If the wild-type pK_a value for Glu14 is used in calculating Q_f, the predicted pH dependence of ΔG comes to good agreement with the experimental results.

Entropy-Based Strategies for Protein Stability

It was suggested earlier that unfavorable interactions in the unfolded state may contribute to folding stability. Since the unfolded state, with its ability to sample vastly different conformations, is favored by chain entropy over the folded state, a fruitful route to increased folding stability is restricting the chain entropy of the unfolded state. Theoretical bases for six different strategies have recently been summarized (*8*). These are (i) spatial confinement, (ii) macromolecular crowding, (iii) shortening of loop lengths, (iv) covalent linkage of dimeric proteins, (v) backbone cyclization, and (vi) catenation.

The theories developed for analyzing these entropy-based stabilization strategies are based on polymer models, specifically, the Gaussian chain model for the unfolded chain and the wormlike chain model for flexible loops and linkers. For example, the effect of spatial confinement on the unfolded state was found by calculating the fraction of conformations of a continuous Gaussian chain not crossing the boundary of a confining cage. Many of the more open conformations of the unfolded chain are eliminated by the confinement. Similarly, the effect of crowding on the unfolded state was found by calculating the fraction of conformations of the same chain not crossing the surface of any crowding macromolecule. Instead of being confined to a cage, the chain can extend to distant locations via interstitial spaces; thus, even very open conformations are possible under crowding. Significant stabilization is predicted for confinement, but only a marginal effect is predicted for crowding. These disparate predictions are consistent with experimental results. Eggers and Valentine (*74*) encapsulated α-lactalbumin in the pores of silica glass and found the melting temperature to increase by 32 °C. On the other hand, Minton and co-workers (*75*) recently studied the thermal unfolding of hen lysozyme in the presence of dextran, observing a mere 2.5 °C increase in the melting temperature even at 300 g of dextran/L. Shortle and Ackerman (*76*) observed persistent nonlocal structural order, even in the presence of 8 M urea, of a large fragment of staphylococcal nuclease confined to the pores of polyacrylamide gels. Thirumalai and co-workers (*77*) have attributed the persistent structure to the confinement effect. Confinement and crowding are likely physiologically important because proteins may be confined to cages such as the central cavity of chaperonins and usually exist in the crowded environment of the cytoplasm.

The fact that a protein loop is able to freely sample different end-to-end distances in the unfolded state but upon folding has a fixed end-to-end distance d determined by the folded structure gives rise to an entropy cost. For a loop modeled as a wormlike chain, the loop entropy is given by (*8*, *37*)

$$\Delta S_{loop}/k_B = \ln p(d) = -(3/2)\ln[(4\pi l_p/3)l_c] - 3d^2/4l_p l_c + \ln[1 - w(d;l_c)] \quad (9a)$$

When d is set to 0 and $p(d)$ is replaced by the probability density of a Gaussian chain, eq 9 is reduced to the classical result of Jacobson and Stockmayer (*78*)

$$\Delta S_{loop}/k_B = \ln p_G(d) = -(3/2)\ln[(2\pi b^2/3)l] \quad (9b)$$

For protein loops, it is important to recognize that the end-to-end vector spans a finite distance d in the folded state (as opposed to being confined to a tiny volume as considered by Jacobson and Stockmayer), and a wormlike chain may be a better model than a Gaussian chain. Equation 9a was found (*37*) to correlate well with the experimental results of Nagi and Regan (*79*) for the effect of loop length on the stability of the four-helix bundle protein Rop.

For a protein that folds only upon dimerization, a flexible linker connecting the subunits changes the folding process from bimolecular to unimolecular. The effective concentration, defined as the ratio of the equilibrium folding constants for the single-chain variant and the dimeric protein, measures the effect of the linker. The linker is unrestrained in the unfolded state but has its end-to-end distance restricted to a fixed value d in the folded state. It was found (*80*)

$$C_{eff} = K^u/K^b = p(d) \quad (10)$$

Equation 10 was shown to predict well the effects of linkers on folding stability for a number of proteins (*8*, *80*). A derivation of this equation will be presented in a later section.

A flexible linker can also be introduced to connect the N- and C-termini, resulting in a circular protein. The linker is now restrained in both the folded and unfolded states. The change in the unfolding free energy by the circularization was found to be given by (*81*)

$$\exp(\beta\Delta\Delta G_{cycl}) = p(d)/\int d^3\mathbf{r}\, p(\mathbf{r})p_G(\mathbf{r}) \quad (11)$$

where $\beta = (k_B T)^{-1}$. The linker and the unfolded linear protein were modeled as a wormlike chain and a Gaussian chain, with probability densities $p(\mathbf{r})$ and $p_G(\mathbf{r})$ for their respective end-to-end vectors. Predicted values of $\Delta\Delta G_{cycl}$ were found to be in quantitative agreement with the experimental results of Deechongkit and Kelly (*82*) on circularizing the PIN1 WW domain by linkers with two to seven residues. Another example is the circularization of DnaB-(24−136) via a nine-residue linker by Dixon and co-workers (*83*). These authors proposed an estimate of $\Delta\Delta G_{cycl}$ that is equivalent to assuming a uniform probability density for the linker end-to-end vector. The average distance (d) between the N- and C-termini of folded DnaB-(24−136) is 13.4 Å. Dixon and co-workers measured a stabilization of 1.9 kcal/mol for the circularization. Equation 11, without adjusting parameters, predicts a $\Delta\Delta G_{cycl}$ of 2.0 kcal/mol. For this protein, eq 11 finds a maximum stabilization of 2.1 kcal/mol by circularization with a five-residue linker. With a 20-residue linker, the stabilization is reduced to 1.7 kcal/mol.

Rate of Contact Formation

Attention is now turned to the kinetics of protein folding. Protein folding may be viewed as a process of accumulating native contacts. In simulations of protein folding, the fraction of native contacts formed is often used to monitor the folding process. The rate of forming a single contact between residues in an unfolded protein or an unstructured polypeptide has been measured in a large number of studies (*84*−*95*).

Contact formation in a polymer chain can be cast as crossing an entropy barrier. The distribution function $P(r)$ for the distance r between the contact-forming residues defines a potential of mean force $[-k_B T \ln P(r)]$. The bottom of the potential well corresponds to the maximum of the distribution function (see, for example, Figure 1B or 3). The value of $P(r)$ at a short distance, the contact distance a, is small, corresponding to a potential barrier. Assuming that the inter-residue motion is diffusive, Szabo and co-workers (*96*) calculated the diffusion-limited rate of contact formation as the inverse of the mean first passage time:

$$k = \int_a^\infty dr P(r)/\int_a^\infty dr[DP(r)]^{-1}[\int_r^\infty dr' P(r')]^2 \quad (12a)$$

where D is the inter-residue diffusion constant. If $a \ll \langle r^2\rangle^{1/2}$, then eq 12a is reduced to

$$k = 4\pi D a \times p(a) \quad (12b)$$

where the probability density $p(r) = P(r)/4\pi r^2$. For Gaussian chain statistics (eq 1), eq 12b becomes

$$k = 3(6/\pi)^{1/2} Da/\langle r^2 \rangle^{3/2} \quad (12c)$$

Using an alternative argument, Wang and Davidson (*97*) had derived an expression agreeing with eq 12c within a numerical factor.

Equation 12c has been used to rationalize experimental results for contact formation rates (*84*, *85*, *87*−*90*, *94*). For example, residues 18 and 80 of the 104-residue cytochrome *c* have a root-mean-square distance $\langle r^2 \rangle^{1/2}$ of 68.4 Å in the unfolded conformations generated in ref *32*. When $a = 4$ Å and $D = 4 \times 10^{-7}$ cm^2/s, a contact rate of 2×10^5 s^{-1} is predicted. The measured rate is in the range of $0.2-1.0 \times 10^5$ s^{-1} (with higher rate for lower denaturant concentrations) (*84*, *85*, *88*). The distribution of the distance between residues 18 and 80 in the unfolded conformations of cytochrome *c* deviates from Gaussian statistics, and can be fitted to $r^{2.3} \exp[-(r/73.4)^{3.4}]$. Calculation according to eq 12a with this fitted function gives a rate of 0.7×10^5 s^{-1} (when $a = 4$ Å and $D = 4 \times 10^{-7}$ cm^2/s). Portman (*98*) has discussed the validity of the assumption of inter-residue diffusion leading to eq 12a and, importantly, the apparent need to use a diffusion constant much smaller than the value expected for free diffusion of the contact-forming residues.

For shorter peptides, the persistence length becomes a hindering factor for end-to-end contact formation. This hindering effect may well be present in contact formation rates for chains with fewer than 10 peptide bonds (*89*, *94*). Lapidus et al. (*89*) found that their data for quenching of tryptophan triplet states by cysteine in Cys-(Ala-Gly-Trp)$_n$-Trp peptides with an n of 1−6 could be fitted to a wormlike chain model with a persistence length of ∼6 Å. The kinetics of this process have been studied by molecular dynamics simulations (*98*, *99*).

The folding of an entire protein proceeds through the accumulation of native contacts. Makarov et al. (*3*, *100*) outlined the following scenario. Diffusional motion of residues changes the conformation of the unfolded chain. A native pair of residues, which are separated by more than 12 residues on the chain and have a $C_\alpha - C_\alpha$ distance of less than 6 Å in the folded structure, will form a contact as soon as they are separated by less than 6 Å. This native contact may unravel or may be followed by another native contact. When all N_{nc} native contacts are formed, folding is complete. The rate for forming a native contact was determined with eq 12c, and the contact breaking rate was a universal constant k_r. The following result was derived for the overall folding rate:

$$k_f = N_{nc} k_r \exp[-\beta F(N_{nc})] \quad (13)$$

where $F(m)$ ($=F_0 + m\Delta F$) is the free energy of forming m native contacts. Equation 13 was found to account for the correlation of folding rates with contact order initially observed by Plaxco et al. (*101*). Ideas similar to those behind eq 13 had previously been proposed by Debe and Goddard (*102*).

Forming a native contact may involve more than just meeting a distance requirement. The two contact-forming residues may have to be in their native local conformations before a contact is formed. The residues can reach their native conformations by local fluctuations. The local conformational fluctuations may be modeled as stochastic transitions between nonnative and native states, just like in studying bimolecular reactions (*103*). An important question is how much the additional requirement of correct local conformations will slow the rate of contact formation. The answer has been found to depend on the time scale of the local conformational fluctuations relative to the inter-residue diffusional motion (*104*). If the fluctuations are slow, then the contact formation rate is scaled by the equilibrium fractions of the native states. For example, if each residue occupies the native conformation with a probability of 1/16, then the rate would be reduced by 256-fold. However, when local fluctuations become rapid, the rate of contact formation would be significantly higher than the equilibrium estimate. In particular, if the local fluctuations occur at a picosecond time scale relative to a nanosecond time scale of inter-residue diffusion, then the rate of contact formation was found to be reduced by just 7-fold. Thus, the rapid local fluctuations appear to play an important role in accelerating protein folding.

Gaussian Chain Model for the Folding Transition State

To model the transition state ensemble in protein folding, Wolynes and co-workers (*9*, *105*) introduced the following energy function

$$H = \sum_{i,j} \mathbf{r}_i \cdot \Gamma_{ij} \cdot \mathbf{r}_j + B \sum_i \mathbf{r}_i^2 + \sum_{nc} \epsilon_{ij} u(r_{ij}) \quad (14)$$

where the first term stipulates the Gaussian chain statistics, the second term constrains the radius of gyration, and the last term represents interactions between natively contacting residues. The free energy of the system was evaluated by a variational approach. The transition state ensemble was obtained by searching for saddle points on the free energy surface. For each saddle point, magnitudes of mean square fluctuations of residues indicate whether they are structured. For λ repressor, Wolynes and co-workers found that helices 1 and 4 are structured in the transition state but helices 2 and 3 are not. This result is consistent with the experimental observation of Oas and co-workers (*106*).

The entropy component of the free energy obtained from eq 14 arises from residues freezing into configurations that optimize favorable interactions. In other models for the transition state ensemble, free energy functions were constructed directly (*107*−*110*). The entropy components of these models consisted of the loop entropy for each pair of contact-forming residues (eq 9). The distributions of structures in the transition state ensembles identified by these models also largely agreed with experiments.

Bimolecular versus Intramolecular Binding

Most proteins function by interacting with other proteins (and DNA). The interactions occur via transient formation of protein−protein (protein−DNA) complexes. Over the course of evolution, interaction partners were connected by peptide linkers, whereby formerly individual proteins now reside within a single polypeptide chain. Linkers between individually folded domains may be relatively rigid or quite flexible (Figure 2). Even domains connected by relatively

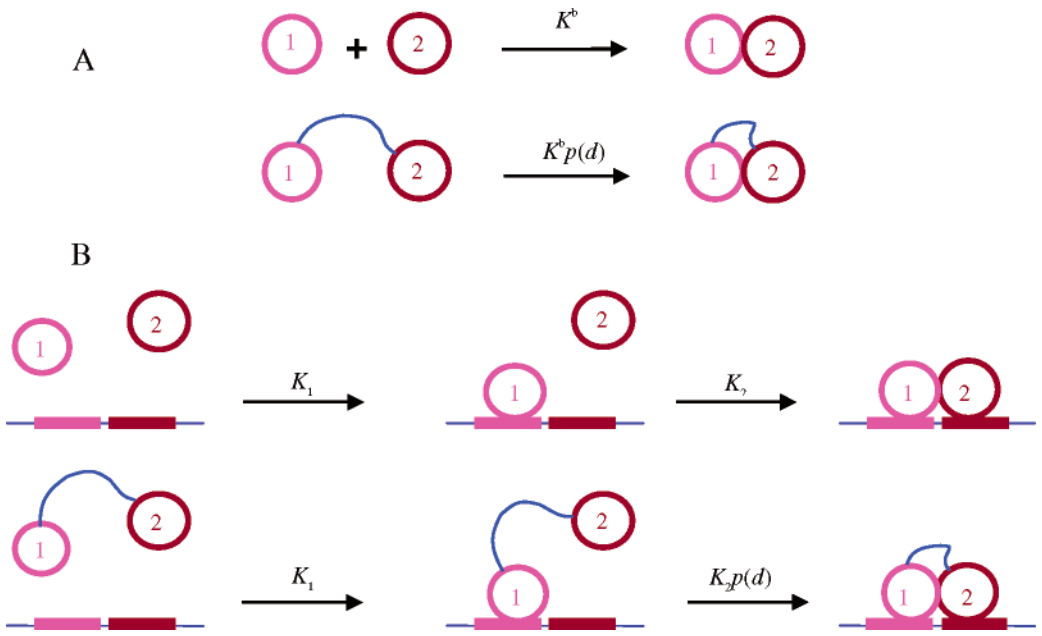

FIGURE 6: Models of the effect of a flexible linker on binding: (A) a linker between two binding partners and (B) a linker between two domains which bind to different sites of the same target.

rigid linkers may dissociate during protein function. In this case, it is difficult to isolate the role of the linker from those of the domains. The situation may be simpler for flexible linkers.

In the generic model illustrated in Figure 6A, a flexible linker between two binding domains changes the binding process from bimolecular to unimolecular. The role of the linker can be measured by the effective concentration, defined as the ratio of the unimolecular and bimolecular binding constants. In 1966, Wang and Davidson (*97*) related the cyclization of DNA with complementary single-stranded ends to the bimolecular association of the single-stranded ends after the original DNA is broken into two fragments. They derived the effective concentration for this case as the Jacobson−Stockmayer factor [$p_G(0) = (3/2\pi b^2 l)^{3/2}$]. For the binding between two protein domains connected by a flexible linker, the effective concentration was derived to be $p(d)$ (*10*, *80*; see eq 10). This result accounts for the fact that the end-to-end vector of the linker spans a finite distance d upon binding of the domains and a short peptide linker may be better modeled as a wormlike chain than a Gaussian chain (a similar argument was made above in discussing loop entropy; see eq 9).

To highlight the assumptions made in deriving eq 10 for the effective concentration, a brief derivation is given here. With a potential of mean force $U(\mathbf{r},\Omega_1,\Omega_2)$, where $\mathbf{r}$ is the displacement between the domains and Ω_1 and Ω_2 specify their orientations, the bimolecular equilibrium constant is (*111*)

$$K^b = \int_b d^3\mathbf{r}\, d^3\Omega_1\, d^3\Omega_2 \exp[-\beta U(\mathbf{r},\Omega_1,\Omega_2)]/(8\pi^2)^2 \quad (15a)$$

where the subscript b specifies the bound state. When a flexible linker, with a probability density $p(\mathbf{r})$ for the end-to-end vector, is introduced, the simultaneous action of the linker and the potential of mean force gives $p(\mathbf{r})\exp(-\beta U)$ as the distribution function. The unimolecular binding constant is

$$K^u = \int_b d^3\mathbf{r}\, d^3\Omega_1\, d^3\Omega_2\, p(\mathbf{r}) \exp(-\beta U)/ \int_u d^3\mathbf{r}\, d^3\Omega_1\, d^3\Omega_2\, p(\mathbf{r}) \exp(-\beta U) \quad (15b)$$

where the subscript u specifies the unbound state. In the bound state, the end-to-end distance of the linker has a fixed value of d. In the unbound state, if the potential of mean force has a range that is much shorter than the width of the probability density $p(\mathbf{r})$, than the Boltzmann factor $\exp(-\beta U)$ can be dropped. Equation 15b becomes

$$K^u = p(d)\int_b d^3\mathbf{r}\, d^3\Omega_1\, d^3\Omega_2 \exp(-\beta U)/(8\pi^2)^2 = p(d)K^b \quad (15c)$$

which is the same as eq 10.

The relation between bimolecular and unimolecular binding reactions can be extended to kinetics. Since the bound state is held together by numerous short-range interactions and dissociation entails breaking these interactions, whether a linker is present appears to be immaterial for the dissociation process (*15*, *80*). Thus, the dissociation rates are expected to be the same

$$k_d^u = k_d^b \quad (16a)$$

which in turn means for the association rates

$$k_a^u/k_a^b = K^u/K^b \quad (16b)$$

$$= p(d) \quad (16c)$$

For DNA cyclization, Wang and Davidson (*97*) have validated eq 16b experimentally.

Enhancement of Binding Affinity by a Flexible Linker

An extension of the model shown in Figure 6A is two linked domains not binding to each other but simultaneously

binding to a third target. The linker in this case is found to play the important role of enhancing binding affinity (*10*). As Figure 6B illustrates, the binding constant of the linked domains to the target is given by

$$K = K_1K_2p(d) \quad (17)$$

where K_1 and K_2 are the binding constants of the individual domains to the target. Relative to the binding constants of the individual domains, the binding constant of the linked domains is scaled by a factor $K_2p(d)$ or $K_1p(d)$. The value of $p(d)$ typically is in the millimolar range, whereas K_1 and K_2 are usually $\geq 10^6$ M^{-1}. Thus, an affinity enhancement of $\geq$1000-fold is expected.

DNA seems to be ideally suited for affinity enhancement by domain linking. Indeed, DNA-binding proteins typically have two or more domains for binding to adjacent sites. The linkers connecting the DNA-binding domains may be rigid or flexible. A dimerization domain may also serve the role of a linker. Figure 2A shows the POU-specific domain and homeodomain of Oct-1 POU bound to DNA (*112*). Klemm and Pabo (*113*) investigated the role of the flexible linker connecting the two domains by studying the DNA binding of the two domains without the linker, obtaining binding constants of 0.6×10^6 and 6.7×10^6 M^{-1}, respectively. The intact Oct-1 POU was found to have a binding constant of 1.4×10^{10} M^{-1}. From these results, one obtains a K/K_1K_2 of 3.6 mM. This value was compared with the prediction of eq 17 (*10*). The end-to-end distance of the linker, measured between the last residue (Glu75) of the POU-specific domain and the first residue (Arg102) of the homeodomain, is 27.6 Å (*112*). With a linker length of 26 peptide bonds, eq 17 with eq 5 predicts $K/K_1K_2 = p(d) = 5.8$ mM, agreeing well with the experimental result. van Leeuwen et al. (*114*) have further studied the effect of linker length on the DNA binding affinity of Oct-1 POU. When the linker length was mutated from 26 to 12, 18, 31, and 40 peptide bonds, they found the ratio of the binding constants before and after mutation to be 3.4, 1.1, 1.0, and 0.8, respectively. These correlate very well with the predictions of eq 17: 4.0, 1.4, 1.0, and 1.0, respectively.

The central region of the 70 kDa subunit of replication protein A (RPA70AB) has two domains that bind to single-stranded DNA in tandem (*115*). By cleaving the linker between the two domains and investigating the DNA binding affinities of the two individual domains, Chazin and co-workers (*11*) recently showed that the covalent linkage of the two domains contributes to an affinity enhancement of ~100-fold. This affinity enhancement was found to be well modeled by eq 17.

A potential important application of the linkage effect is in engineering high-affinity antibodies. Many antibodies can be raised against the same antigen, each targeting a different epitope. High affinity can be achieved by covalently linking antibodies targeting different epitopes. This strategy was successfully implemented by Neri et al. (*116*) by linking two single-chain Fv fragments, D1.3 and HyHEL-10, against hen egg white lysozyme (HEL). The HEL binding constants of the two Fv fragments were 10^6 and 10^8 M^{-1}, respectively. The linked Fv fragments were found to have an HEL binding constant of $>1.3 \times 10^{10}$ M^{-1}. This result was shown to be predicted well by eq 17 (*12*). The linker introduced by Neri et al. has 18 peptide bonds and spans a distance of 28.6 Å. According to eq 5, $p(d)$ in this case is 3.4 mM. When $K_1 = 10^6$ M^{-1} and $K_2 = 10^8$ M^{-1}, eq 17 predicts $K = 3.4 \times 10^{11}$ M^{-1}.

The association and dissociation rates of two domains linked to the same target have been found through the use of eq 16 (*12*). It was found that the affinity enhancement was achieved by reducing the dissociation rate without affecting the association rate. This prediction was validated by a recent experimental result (*117*, *118*).

Role of Flexible Linkers in Regulatory Switches

The equilibrium between the bound and unbound states may serve as a regulatory switch for many proteins. For example, dissociation of a binding partner may expose a DNA-binding surface or an enzyme active site. The advantage afforded by covalently linking the interaction partners is that the binding equilibrium is concentration-independent since the reaction is unimolecular. The equilibrium constant may be set to a fixed value, or it may by modulated by a triggering event such as phosphorylation or binding of a small ligand.

Recently, Dahlquist and co-workers (*13*) studied the interdomain interaction of the response regulator NarL. Binding of the receiver domain blocks the DNA-binding surface of the effector domain. The binding between the domains is weakened by phosphorylation of the receiver domain. The role of the linker was studied by comparing the unimolecualar binding of the domains in the intact protein with the bimolecular binding when the linker was cleaved. The linker has 13 peptide bonds and spans a distance of 19.2 Å. Equation 5 was used to determine $p(d) = 20$ mM. In the dephosphorylated and phosphorylated states, the two domains had bimolecular binding constants of 2.5×10^3 and 80 M^{-1}, respectively. Equation 15c would predict $K^u = 50$ and 1.6, meaning 2% of the dephosphorylated intact protein and 38% of the phosphorylated intact protein would be competent for DNA binding. These predictions agreed well with the DNA binding assay results reported by Dahlquist and co-workers: phosphorylated NarL bound DNA at the same level as the effector domain alone, whereas a 10-fold excess of nonphosphorylated NarL bound DNA at the same level as the effector domain alone.

Superti-Furga, Kuriyan, and co-workers (*119*, *120*) uncovered an important regulatory element of the c-Abl tyrosine kinase, in the form of the myristoylated N-terminus latching onto the kinase domain (Figure 2B). One potential mechanism for activating c-Abl is the dissociation of the myristoylated N-terminal latch. How often does this latch spontaneously come off? An answer to this question was obtained by incorporating experimental results to the wormlike chain model for linkers (*14*). The myristoylated N-terminus is connected to the SH3 domain of c-Abl by a flexible linker, which consists of 79 peptide bonds and spans an end-to-end distance of 56.8 Å (*120*). Equation 5 gives a $p(d)$ of 0.5 mM. Superti-Furga and co-workers (*119*) measured a dissociation constant $1/K^b$ of 2.3 μM for the myristoyl group. The binding constant K^u for the intact c-Abl is then predicted by eq 15c to be 217 M^{-1}, which means an equilibrium unbound fraction of just 0.5%. The small unbound fraction may be important for keeping the kinase activity of c-Abl at low levels.

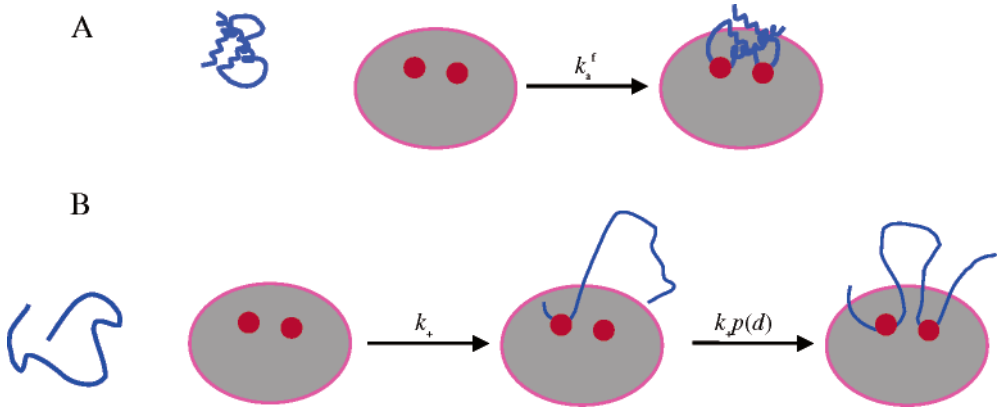

FIGURE 7: Binding of (A) a folded protein and (B) an unfolded protein to a target.

Unregulated kinase activity of c-Abl is a known cause for chronic myelogenous leukemia.

The unimolecular nature of covalently linked binding may also be important when a rapid binding rate is required, such as in deactivating ion channels in neural cells (*15*). For example, the *Shaker* potassium channel employs its N-terminal segment for inactivation (*121*).

Folding upon Binding

Many proteins are unfolded before they bind with a target (*4*−*6*). Wolynes and co-workers (*122*) studied the impact of being unfolded during binding and proposed that binding kinetics may be accelerated. The unfolded chain is more open and able to sample different conformations, and thus is able to interact with the binding surface on the target at a greater distance from the target than the folded protein. They constructed a free energy functional for the binding process by adding a binding term to their previous functional for protein folding (*107*). The problem was then solved in a Born−Oppenheimer fashion; the potential of mean force for a given protein−target distance was obtained first, which was then used in the Debye formula (*123*) to find the association rate.

An alternative explanation of the acceleration of the binding process by an unfolded protein is illustrated in Figure 7. The rate k_a^f of a rigid folded protein binding to a target is small because of the slow process of coming to the correct site with the correct orientation by random diffusion. k_a^f has been estimated to be on the order of 10^5 M^{-1} s^{-1} (*124*). For the unfolded chain, the first binding site can be reached without the burden of satisfying orientational constraints. The rate constant, k_+, is roughly what is expected for the binding of a point ligand to a small patch on the target surface. This is given by $4Da$ (*125*) (D is the diffusion constant and a the patch radius) and is on the order of 10^7 M^{-1} s^{-1}. The occupation of the second binding site becomes unimolecular, with a rate $p(d)k_+$ according to eq 16c. The overall binding rate is

$$k_a = k_+ \frac{K_+ p(d)}{K_+ p(d) + 1} \qquad (18)$$

where K_+ is the equilibrium constant for the first binding site. When $K_+p(d) \gg 1$, the association rate approaches k_+, which is 2 orders of magnitude higher than the rate for the folded protein.

Future Developments

I have given an overview of recent polymer theory-based models for the unfolded state and flexible linkers, and their applications in analyzing experimental data on protein stability, folding, and interactions. What is captured by these models is the fact that unfolded protein chains and flexible linkers sample many different conformations. What are missed are realistic details of the conformations actually sampled by an unfolded protein or a flexible linker connecting folded domains. Such details can be obtained from molecular dynamics simulations. These simulations may suggest more realistic polymer models and help fix model parameters. On the experimental side, the theoretical models may benefit from systematic and critical tests. The interplay of theoretical models, molecular dynamics simulations, and experiments will greatly advance our understanding of the unfolded state and flexible linkers.

REFERENCES

1. Baldwin, R. L. (2002) Making a network of hydrophobic clusters, *Science 295*, 1657−1658.
2. Zhou, H.-X. (2002) A Gaussian-chain model for treating residual charge−charge interactions in the unfolded state of proteins, *Proc. Natl. Acad. Sci. U.S.A. 99*, 3569−3574.
3. Makarov, D. E., Keller, C. A., Plaxco, K. W., and Metiu, H. (2002) How the folding rate constant of simple, single-domain proteins depends on the number of native contacts, *Proc. Natl. Acad. Sci. U.S.A. 99*, 3535−3539.
4. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function, *Biochemistry 41*, 6573−6582.
5. Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics, *Protein Sci. 11*, 739−756.
6. Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins, *Curr. Opin. Struct. Biol. 12*, 54−60.
7. Millet, I. S., Doniach, S., and Plaxco, K. W. (2002) Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins, *Adv. Protein Chem. 62*, 241−262. Other chapters in this volume also provide valuable insight to the characterizations of the unfolded state. In particular, R. L. Baldwin presented a perspective of unfolded proteins (pp 361−367) and A. K. Dunker, C. J. Brown, and Z. Obradovic presented a summary of disordered proteins (pp 25−49).
8. Zhou, H.-X. (2004) Loops, linkages, rings, catenanes, cages, and crowders: entropy-based strategies for stabilizing proteins, *Acc. Chem. Res. 37*, 123−130.
9. Portman, J. J., Takada, S., and Wolynes, P. G. (1998) Variational theory for site resolved protein folding free energy surfaces, *Phys. Rev. Lett. 81*, 5237−5240.
10. Zhou, H.-X. (2001) The affinity-enhancing roles of flexible linkers in two-domain DNA-binding proteins, *Biochemistry 40*, 15069−15073.

11. Arunkumar, A. I., Stauffer, M. E., Bochkareva, E., Bochkarev, A., and Chazin, W. J. (2003) Independent and coordinated functions of replication protein A tandem high affinity single-stranded DNA binding domains, *J. Biol. Chem. 278*, 41077−41082.
12. Zhou, H.-X. (2003) Quantitative account of the enhanced affinity of two linked scFvs specific for different epitopes on the same antigen, *J. Mol. Biol. 329*, 1−8.
13. Eldridge, A. M., Kang, H. S., Johnson, E., Gunsalus, R., and Dahlquist, F. W. (2002) Effect of phosphorylation on the interdomain interaction of the response regulator, NarL, *Biochemistry 41*, 15173−15180.
14. Zhou, H.-X. (2003) How often does the myristoylated N-terminal latch of c-Abl come off? *FEBS Lett. 552*, 160−162.
15. Zhou, H.-X. (2002) A model for the binding of the inactivation N-terminal to the ion pore of Shaker potassium channel: both electrostatic attraction and covalent linkage are required for rapid inactivation, *J. Phys. Chem. B 106*, 2393−2397.
16. Zimm, B. H. (1956) Dynamics of polymer molecules in dilute solution: viscoelasticity, flow birefringence and dielectric loss, *J. Chem. Phys. 24*, 269−278.
17. Kirkwood, J. G., and Riseman, J. (1948) The intrinsic viscosities and diffusion constants of flexible macromolecules in solution, *J. Chem. Phys. 16*, 565−573.
18. Ramachandran, G. N., Venkatachalam, C. M., and Krimm, S. (1966) Stereochemical criteria for polypeptide and protein chain conformations. 3. Helical and hydrogen-bonded polypeptide chains, *Biophys. J. 6*, 849−872.
19. Peterlin, A. (1955) Excluded volume effect on light scattering of the coiled linear macromolecule, *J. Chem. Phys. 23*, 2464−2465.
20. Tanford, C., Kawahara, K., and Lapanje, S. (1967) Proteins as random coils. I. Intrinsic viscosities and sedimentation coefficients in concentrated guanidine hydrochloride, *J. Am. Chem. Soc. 89*, 729−736.
21. Tanford, C. (1968) Protein denaturation, *Adv. Protein Chem. 23*, 121−282.
22. Damaschun, G., Damaschun, H., Gast, K., Gernat, C., and Zirwer, D. (1991) Acid denatured apo-cytochrome *c* is a random coil: evidence from small-angle X-ray scattering and dynamic light scattering, *Biochim. Biophys. Acta 1078*, 289−295.
23. Pan, H., Barany, G., and Woodward, C. (1997) Reduced BPTI is collapsed. A pulsed field gradient NMR study of unfolded and partially folded bovine pancreatic trypsin inhibitor, *Protein Sci. 6*, 1985−1992.
24. Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A., and Smith, L. J. (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques, *Biochemistry 38*, 16424−16431.
25. Kataoko, M., Hagihara, Y., Mihara, M., and Goto, Y. (1993) Molten globule of cytochrome *c* studied by small-angle X-ray scattering, *J. Mol. Biol. 229*, 591−596.
26. Smith, C. K., Bu, Z., Anderson, K. S., Sturtevant, J. M., Engelman, D. M., and Regan, L. (1996) Surface point mutations that significantly alter the structure and stability of a protein's denatured state, *Protein Sci. 5*, 2009−2019.
27. Segel, D. J., Fink, A. L., Hodgson, K. O., and Doniach, S. (1998) Protein denaturation: a small-angle X-ray scattering study of the ensemble of unfolded states of cytochrome *c*, *Biochemistry 37*, 12443−12451.
28. Plaxco, K. W., Millet, I. S., Segel, D. J., Doniach, S., and Baker, D. (1999) Chain collapse can occur concomitantly with the rate-limiting step in protein folding, *Nat. Struct. Biol. 6*, 554−556.
29. Kamatari, Y. O., Ohji, S., Konno, T., Seki, Y., Soda, K., Kataoka, M., and Akasaka, K. (1999) The compact and expanded denatured conformations of apomyoglobin in the methanol−water solvent, *Protein Sci. 8*, 873−882.
30. Choy, W. Y., Mulder, F. A., Crowhurst, K. A., Muhandiram, D. R., Millett, I. S., Doniach, S., Forman-Kay, J. D., and Kay, L. E. (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques, *J. Mol. Biol. 316*, 101−112.
31. Millet, I. S., Townsley, L. E., Chiti, F., Doniach, S., and Plaxco, K. W. (2002) Equilibrium collapse and the kinetic 'foldability' of proteins, *Biochemistry 41*, 321−325.
32. Zhou, H.-X. (2002) Dimensions of denatured proteins chains from hydrodynamic data, *J. Phys. Chem. B 106*, 5769−5775.
33. Goldenberg, D. (2003) Computational simulation of the statistical properties of unfolded proteins, *J. Mol. Biol. 326*, 1615−1633.
34. Hearst, J. E., and Stockmayer, W. H. (1962) Sedimentation constants of broken chains and wormlike coils, *J. Chem. Phys. 37*, 1425−1433.
35. Kratky, O., and Porod, G. (1949) Röntgenuntersushung gelöster Fagenmoleküle, *Rec. Trav. Chim. 68*, 1106−1123.
36. Gobush, W., Stockmayer, W. H., Yamakawa, H., and Magee, W. S. (1972) Statistical mechanics of wormlike chains. I. Asymptotic behavior, *J. Chem. Phys. 57*, 2839−2843.
37. Zhou, H.-X. (2001) Loops in proteins can be modeled as worm-like chains, *J. Phys. Chem. B 105*, 6763−6766.
38. Bednar, J., Furrer, P., Katritch, V., Stasiak, A., Dubochet, J., and Stasiak, A. (1995) Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA, *J. Mol. Biol. 254*, 579−594.
39. Rivetti, C., Guthold, M., and Bustamante, C. (1996) Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis, *J. Mol. Biol. 264*, 919−932.
40. Li, H., Oberhauser, A. F., Redick, S. D., Carrion-Vazquez, M., Erickson, H. P., and Fernandez, J. M. (2001) Multiple conformations of PEVK proteins detected by single-molecule techniques, *Proc. Natl. Acad. Sci. U.S.A. 98*, 10682−10686.
41. Marko, J. F., and Siggia, E. (1995) Stretching DNA, *Macromolecules 28*, 8759−8770.
42. Bustamante, C., Marko, J. F., Siggia, E., and Smith, S. (1994) Entropic elasticity of lambda-phage DNA, *Science 265*, 1599−1600.
43. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., and Gaub, H. E. (1997) Reversible unfolding of individual titin immunoglobulin domains by AFM, *Science 276*, 1109−1112.
44. Oberhauser, A. F., Marszalek, P. E., Erickson, H. P., and Fernandez, J. M. (1998) The molecular elasticity of the extracellular matrix protein tenascin, *Nature 393*, 181−185.
45. Rief, M., Gautel, M., Schemmel, A., and Gaub, H. E. (1998) The mechanical stability of immunoglobulin and fibronectin III domains in the muscle protein titin measured by atomic force microscopy, *Biophys. J. 75*, 3008−3014.
46. Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J., and Fernandez, J. M. (1999) Mechanical and chemical unfolding of a single protein: a comparison, *Proc. Natl. Acad. Sci. U.S.A. 96*, 3694−3697.
47. Rief, M., Pascual, J., Saraste, M., and Gaub, H. E. (1999) Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles, *J. Mol. Biol. 286*, 553−561.
48. Yang, G., Cecconi, C., Baase, W. A., Vetter, I. R., Breyer, W. A., Haack, J. A., Matthews, B. W., Dahlquist, F. W., and Bustamante, C. (2000) Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme, *Proc. Natl. Acad. Sci. U.S.A. 97*, 139−144.
49. Furuike, S., Ito, T., and Yamazaki, M. (2001) Mechanical unfolding of single filamin A (ABP-280) molecules detected by atomic force microscopy, *FEBS Lett. 498*, 72−75.
50. Best, R. B., Li, B., Steward, A., Daggett, V., and Clarke, J. (2001) Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation, *Biophys. J. 81*, 2344−2356.
51. Brockwell, D. J., Paci, E., Zinober, R. C., Beddard, G. S., Olmsted, P. D., Smith, D. A., Perham, R. N., and Radford, S. E. (2003) Pulling geometry defines the mechanical resistance of a β-sheet protein, *Nat. Struct. Biol. 10*, 731−737.
52. Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P. E., Oberhauser, A. F., and Fernandez, J. M. (2003) The mechanical stability of ubiquitin is linkage dependent, *Nat. Struct. Biol. 10*, 738−743.
53. Janovjak, H., Kessler, M., Oesterhelt, D., Gaub, H., and Muller, D. J. (2003) Unfolding pathways of native bacteriorhodopsin depend on temperature, *EMBO J. 22*, 5220−5529.
54. Hertadi, R., Gruswitz, F., Silver, L., Koide, A., Koide, S., Arakawa, H., and Ikai, A. (2003) Unfolding mechanics of multiple OspA substructures investigated with single molecule force spectroscopy, *J. Mol. Biol. 333*, 993−1002.
55. Tanford, C. (1970) Protein denaturation, *Adv. Protein Chem. 24*, 1−95.
56. Stigter, D., Alonso, D. O., and Dill, K. A. (1991) Protein stability: electrostatics and compact denatured states, *Proc. Natl. Acad. Sci. U.S.A. 88*, 4176−4180.
57. Oliveberg, M., Vuilleumier, S., and Fersht, A. R. (1994) Thermodynamic study of the acid denaturation of barnase and its

dependence on ionic strength: evidence for residual electrostatic interactions in the acid/thermally denatured state, *Biochemistry 33*, 8826−8832.
58. Oliveberg, M., Arcus, V. L., and Fersht, A. R. (1995) pK_a values of carboxyl groups in the native and denatured states of barnase: the pK_a values of the denatured state are on average 0.4 units lower than those of model compounds, *Biochemistry 34*, 9424−9433.
59. Pace, C. N., Laurents, D. V., and Thomson, J. A. (1990) pH dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1, *Biochemistry 29*, 2564−2572.
60. Swint-Kruse, L., and Robertson, A. D. (1995) Hydrogen bonds and the pH dependence of ovomucoid third domain stability, *Biochemistry 34*, 4724−4732.
61. Tan, Y.-J., Oliverberg, M., Davis, B., and Fersht, A. R. (1995) Perturbed pK_a-values in the denatured states of proteins, *J. Mol. Biol. 254*, 980−992.
62. Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions, *Biochemistry 38*, 4896−4903.
63. Whitten, S. T., and Garcia-Moreno, E. B. (2000) pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state, *Biochemistry 39*, 14292−14304.
64. Koide, A., Jordan, M. R., Horner, S. R., Batori, V., and Koide, S. (2001) Stabilization of a fibronectin type III domain by the removal of unfavorable electrostatic interactions on the protein surface, *Biochemistry 40*, 10326−10333.
65. Laurents, D. V., Huyghues-Despointes, B. M., Thurlkill, R. L., Daily, M. D., Schell, D., Briggs, J. M., Antosiewicz, J. M., Pace, C. N., and Scholtz, J. M. (2003) Charge−charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI=3.5) and a basic variant (pI=10.2), *J. Mol. Biol. 325*, 1077−1092.
66. Tollinger, M., Crowhurst, K. A., Kay, L. E., and Forman-Kay, J. D. (2003) Site-specific contributions to the pH dependence of protein stability, *Proc. Natl. Acad. Sci. U.S.A. 100*, 4545−4550.
67. Elcock, A. H. (1999) Realistic modeling of the denatured states of proteins allows accurate calculations of the pH dependence of protein stability, *J. Mol. Biol. 294*, 1051−1062.
68. Zhou, H.-X. (2002) Residual electrostatic effects in the unfolded state of the N-terminal domain of L9 can be attributed to non-specific non-local charge−charge interactions, *Biochemistry 41*, 6533−6538.
69. Zhou, H.-X. (2002) Residual charge interactions in unfolded staphylococcal nuclease can be explained by the Gaussian-chain model, *Biophys. J. 83*, 2981−2986.
70. Tollinger, M., Forman-Kay, J. D., and Kay, L. E. (2002) Measurement of side-chain carboxyl pK_a values of glutamate and aspartate residues in an unfolded protein by multinuclear NMR spectroscopy, *J. Am. Chem. Soc. 124*, 5714−5717.
71. Zhou, H.-X. (2003) Direct test of the Gaussian-chain model for treating residual charge−charge interactions in the unfolded state of proteins, *J. Am. Chem. Soc. 125*, 2060−2061.
72. Zhou, H.-X., and Dong, F. (2003) Electrostatic contributions to the stability of a thermophilic cold shock protein, *Biophys. J. 84*, 2216−2222.
73. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A., and Pace, C. N. (2001) The effect of net charge on the solubility, activity, and stability of ribonuclease Sa, *Protein Sci. 10*, 1206−1215.
74. Eggers, D. K., and Valentine, J. S. (2001) Molecular confinement influences protein structure and enhances thermal protein stability, *Protein Sci. 10*, 250−261.
75. Sasahara, K., McPhie, P., and Minton, A. P. (2003) Effect of dextran on protein stability and conformation attributed to macromolecular crowding, *J. Mol. Biol. 326*, 1227−1237.
76. Shortle, D., and Ackerman, M. (2001) Persistence of native-like topology in a denatured protein in 8 M urea, *Science 293*, 487−489.
77. Klimov, D. K., Newfield, D., and Thirumalai, D. (2002) Simulations of β-hairpin folding confined to spherical pores using distributed computing, *Proc. Natl. Acad. Sci. U.S.A. 99*, 8019−8024.
78. Jacobson, H., and Stockmayer, W. H. (1950) Intramolecular reaction in polycondensations. I. The theory of linear systems, *J. Chem. Phys. 18*, 1600−1606.
79. Nagi, A. D., and Regan, L. (1997) An inverse correlation between loop length and stability in a four-helix-bundle protein, *Folding Des. 2*, 67−75.
80. Zhou, H.-X. (2001) Single-chain versus dimeric protein folding: thermodynamic and kinetic consequences of covalent linkage, *J. Am. Chem. Soc. 123*, 6730−6731.
81. Zhou, H.-X. (2003) Effect of backbone cyclization on protein folding stability: chain entropies of both the unfolded and the folded states are restricted, *J. Mol. Biol. 332*, 257−264.
82. Deechongkit, S., and Kelly, J. W. (2002) The effect of backbone cyclization on the thermodynamics of β-sheet unfolding: stability optimization of the PIN WW domain, *J. Am. Chem. Soc. 124*, 4980−4986.
83. Williams, N. K., Prosselkov, P., Liepinsh, E., Line, I., Sharipo, A., Littler, D. R., Curmi, P. M. G., Otting, G., and Dixon, N. E. (2002) *In vivo* protein cyclization promoted by a circularly permuted *Synechocystis* sp. PCC6803 DnaB mini-intein, *J. Biol. Chem. 277*, 7790−7798.
84. Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. (1996) Diffusion-limited contact formation in unfolded cytochrome *c*: estimating the maximum rate of protein folding, *Proc. Natl. Acad. Sci. U.S.A. 93*, 11615−11617.
85. Hagen, S. J., Hofrichter, J., and Eaton, W. A. (1997) Rate of intrachain diffusion of unfolded cytochrome *c*, *J. Phys. Chem. B 101*, 2352−2365.
86. Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M., and Kiefhaber, T. (1999) The speed limit for protein folding measured by triplet−triplet energy transfer, *Proc. Natl. Acad. Sci. U.S.A. 96*, 9597−9601.
87. Lapidus, L. J., Eaton, W. A., and Hofrichter, J. (2000) Measuring the rate of intramolecular contact formation in polypeptides, *Proc. Natl. Acad. Sci. U.S.A. 97*, 7220−7225.
88. Hagen, S. J., Carswell, C. W., and Sjolander, E. M. (2001) Rate of intrachain contact formation in an unfolded protein: temperature and denaturant effects, *J. Mol. Biol. 305*, 1161−1171.
89. Lapidus, L. J., Steinbach, P. J., Eaton, W. A., Szabo, A., and Hofrichter, J. (2002) Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: testing a 1-dimenional diffusion model for peptide dynamics, *J. Phys. Chem. B 106*, 11628−11640.
90. Lapidus, L. J., Eaton, W. A., and Hofrichter, J. (2002) Measuring dynamic flexibility of the coil state of a helix-forming peptide, *J. Mol. Biol. 319*, 19−25.
91. Hudgins, R. R., Huang, F., Gramlich, G., and Nau, W. M. (2002) A fluorescence-based method for direct measurement of submicrosecond intramolecular contact formation in biopolymers: an exploratory study with polypeptides, *J. Am. Chem. Soc. 124*, 556−564.
92. Neuweiler, H., Schulz, A., Bohmer, M., Enderlein, J., and Sauer, M. (2003) Measurement of submicrosecond intramolecular contact formation in peptides at the single-molecule level, *J. Am. Chem. Soc. 125*, 5324−5330.
93. Buscaglia, M., Schuler, B., Lapidus, L. J., Eaton, W. A., and Hofrichter, J. (2003) Kinetics of intramolecular contact formation in a denatured protein, *J. Mol. Biol. 332*, 9−12.
94. Krieger, F., Fierz, B., Bieri, O., Drewello, M., and Kiefhaber, T. (2003) Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding, *J. Mol. Biol. 332*, 265−274.
95. Huang, F., and Nau, W. M. (2003) A conformational flexibility scale for amino acids in peptides, *Angew. Chem., Int. Ed. 42*, 2269−2272.
96. Szabo, A., Schulten, K., and Schulten, Z. (1980) First passage time approach to diffusion controlled reactions, *J. Chem. Phys. 72*, 4350−4357.
97. Wang, J. C., and Davidson, N. (1966) On the probability of ring closure of lambda DNA, *J. Mol. Biol. 19*, 469−482.
98. Portman, J. J. (2003) Non-Gaussian dynamics from a simulation of a short peptide: loop closure rates and effective diffusion coefficients, *J. Chem. Phys. 118*, 2381−2391.
99. Yeh, I.-C., and Hummer, G. (2002) Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent, *J. Am. Chem. Soc. 124*, 6563−6568.
100. Makarov, D. E., and Plaxco, K. W. (2003) The topomer search model: a simple, quantitative theory of two-state protein folding kinetics, *Protein Sci. 12*, 17−26.

101. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins, *J. Mol. Biol. 277*, 985−994.
102. Debe, D. A., and Goddard, W. A., III (1999) First principles prediction of protein folding rates, *J. Mol. Biol. 294*, 619−625.
103. Zhou, H.-X., and Szabo, A. (1996) Theory and simulation of the time-dependent rate coefficients of diffusion-influenced reactions, *Biophys. J. 71*, 2440−2457.
104. Zhou, H.-X. (2003) Theory for the rate of contact formation in a polymer chain with local conformational transitions, *J. Chem. Phys. 118*, 2010−2015.
105. Portman, J. J., Takada, S., and Wolynes, P. G. (2001) Microscopic theory of protein folding rates. I. Fine structure of the free energy profile and folding routes from a variational approach, *J. Chem. Phys. 114*, 5069−5081.
106. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. (1997) The energy landscape of a fast-folding protein mapped by Ala → Gly substitutions, *Nat. Struct. Biol. 4*, 305−310.
107. Shoemaker, B. A., Wang, J., and Wolynes, P. G. (1999) Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble, *J. Mol. Biol. 287*, 675−694.
108. Galzitskaya, O. V., and Finkelstein, A. V. (1999) A theoretical search for folding/unfolding nuclei in three-dimensional protein structures, *Proc. Natl. Acad. Sci. U.S.A. 96*, 11299−11304.
109. Alm, E., and Baker, D. (1999) Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures, *Proc. Natl. Acad. Sci. U.S.A. 96*, 11305−11310.
110. Guerois, R., and Serrano, L. (2000) The SH3-fold family: experimental evidence and prediction of variations in the folding pathways, *J. Mol. Biol. 304*, 967−982.
111. Zhou, H.-X. (2001) Disparate ionic-strength dependence of on and off rates in protein-protein association, *Biopolymers 59*, 427−433.
112. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules, *Cell 77*, 21−32.
113. Klemm, J. D., and Pabo, C. O. (1996) Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage, *Genes Dev. 10*, 27−36.
114. van Leeuwen, H. C., Strating, M. J., Rensen, M., de Laat, W., and van der Vliet, P. C. (1997) Linker length and composition influence the flexibility of Oct-1 DNA binding, *EMBO J. 16*, 2043−2053.
115. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA, *Nature 385*, 176−181.
116. Neri, D., Momo, M., Prospero, T., and Winter, G. (1995) High-affinity antigen binding by chelating recombinant antibodies (CRAbs), *J. Mol. Biol. 246*, 367−373.
117. Walker, D., Moore, G. R., James, R., and Kleanthous, C. (2003) Thermodynamic consequences of bipartite immunity protein binding to the ribosomal ribonuclease colicin E3, *Biochemistry 42*, 4161−4171.
118. Zhou, H.-X. (2003) Association and dissociation kinetics of colicin E3 and immunity protein 3: convergence of theory and experiment, *Protein Sci. 12*, 2379−2382.
119. Hantschel, O., Nagar, B., Guettler, S., Kretzschmar, J., Dorey, K., Kuriyan, J., and Superti-Furga, G. (2003) A myristoyl/phosphotyrosine switch regulates c-Abl, *Cell 112*, 845−857.
120. Nagar, B., Hantschel, O., Young, M. A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., and Kuriyan, J. (2003) Structural basis for the autoinhibition of c-Abl tyrosine kinase, *Cell 112*, 859−871.
121. Zagotta, W. N., Hoshi, T., and Aldrich, R. W. (1990) Restoration of inactivation in mutants of Shaker potassium channels by a peptide derived from ShB, *Science 250*, 568−571.
122. Shoemaker, B. A., Portman, J. J., and Wolynes, P. G. (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism, *Proc. Natl. Acad. Sci. U.S.A. 97*, 8868−8873.
123. Debye, P. (1942) Reaction rates in ionic solutions, *Trans. Electrochem. Soc. 82*, 265−272.
124. Zhou, H.-X. (1997) Enhancement of protein−protein association rate by interaction potential: Accuracy of prediction based on local Boltzmann factor, *Biophys. J. 73*, 2441−2445.
125. Hill, T. L. (1975) Effect of rotation on the diffusion-controlled rate of ligand-protein association, *Proc. Natl. Acad. Sci. U.S.A. 72*, 4918−4922.

BI036269N