Clusters in an Intrinsically Disordered Protein Create a Protein-Binding Site: The TolB-Binding Region of Colicin E9[†]

Kaeko Tozawa,[‡,§,||] Colin J. Macdonald,[‡,||] Christopher N. Penfold,[§] Richard James,[§] Colin Kleanthous,[⊥] Nigel J. Clayden,[‡] and Geoffrey R. Moore*[,‡]

School of Chemical Sciences and Pharmacy, University of East Anglia, Norwich NR4 7TJ, United Kingdom, School of Molecular Medical Sciences and Institute of Infection, Immunity, and Inflammation, Centre for Biomolecular Sciences, University of Nottingham, Nottingham NG7 2RD, United Kingdom, and Department of Biology (Area 10), University of York, York YO10 5YW, United Kingdom

Received February 25, 2005; Revised Manuscript Received June 17, 2005

ABSTRACT: The 61-kDa colicin E9 protein toxin enters the cytoplasm of susceptible cells by interacting with outer membrane and periplasmic helper proteins and kills them by hydrolyzing their DNA. The membrane translocation function is located in the N-terminal domain of the colicin, with a key signal sequence being a pentapeptide region that governs the interaction with the helper protein TolB (the TolB box). Previous NMR studies [Collins et al. (2002) *J. Mol. Biol. 318*, 787−904; MacDonald et al. (2004), *J. Biomol. NMR 30*, 81−96] have shown that the N-terminal 83 residues of colicin E9, which includes the TolB box, is intrinsically disordered and contains clusters of interacting side chains. To further define the properties of this region of colicin E9, we have investigated the effects on the dynamical and TolB-binding properties of three mutations of colicin E9 that inactivate it as a toxin. The mutations were contained in a fusion protein consisting of residues 1−61 of colicin E9 connected to the N terminus of the E9 DNase by an eight-residue linking sequence. The NMR data reveals that the mutations cause major alterations to the properties of some of the clusters, consistent with some form of association between them and other more distant parts of the amino acid sequence, particularly toward the N terminus of the protein. However, ^{15}N T_2 measurements indicates that residues 5−13 of the fusion protein bound to the 43-kDa TolB remain as flexible as they are in the free protein. The NMR data point to considerable dynamic ordering within the intrinsically disordered translocation domain of the colicin that is important for creating the TolB-binding site. Furthermore, amino acid sequence considerations suggest that the clusters of amino acids occur because of the size and polarities of the side chains forming them influenced by the propensities of the residues within the clusters and those immediately surrounding them in sequence space to form β turns.

One of the striking discoveries concerning protein structure in recent years has been the finding that many proteins are unfolded or only partly folded in their native states, folding into an ordered structure on binding a partner molecule (*1*−*4*). Coupling a protein-folding transition to an intermolecular interaction has been considered to be functionally advantageous because it may contribute to the specificity of the intermolecular recognition event (*5*), could enhance the rate of the intermolecular interaction (*6*, *7*), may allow a protein to bind to several different target molecules (*4*, *8*), and could provide for large intermolecular interfaces with relatively small proteins (*9*). A counterview is that coupled protein folding and binding events are not as frequent as recent reports suggest because a protein that is intrinsically disordered *in vitro* may not be so *in vivo*, as illustrated by *Salmonella typhimurium* FlgM. This has two unstructured domains in the dilute solutions required for *in vitro* NMR, and one becomes structured inside living *Escherichia coli* cells (*10*). Furthermore, where coupled folding and binding does occur, it is not always clear that the partner molecule recognizes the unfolded form of a protein, which folds once the molecules have made contact or binds to the folded form of the protein, which, being in equilibrium with the unfolded form, leads to protein folding through sequestration of the folded form in the complex. Neither of these caveats apply to the intrinsically disordered N-terminal domain of *E. coli* colicin E9 (*11*, *12*), an extracellular toxin secreted by producing strains of *E. coli* to kill competing bacteria (*13*, *14*). This domain retains disordered regions of considerable flexibility within the complex of the colicin and its partner protein TolB (*11*).

Cell killing by colicins normally occurs in three steps, which are carried out by separate domains: passage of the colicin across the outer membrane of the target cell following binding of the colicin to one or more outer membrane proteins, translocation of the colicin across the periplasm of

[†] This work was supported by the BBSRC through a project grant to R.J., the CEC through contract number QLRT-1999-0100, and the Joint Infrastructure Fund and the Wellcome Trust.

* To whom correspondence should be addressed. Telephone: +44-1603-592697. Fax: +44-1603-592003. E-mail: g.moore@uea.ac.uk.

[‡] University of East Anglia.

[§] University of Nottingham.

[||] These workers contributed equally to the study.

[⊥] University of York.

10.1021/bi0503596 CCC: $30.25
Published on Web 08/04/2005

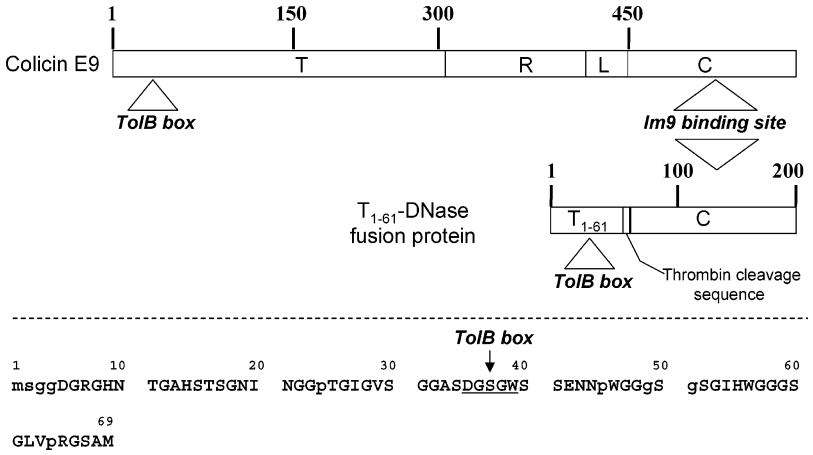

FIGURE 1: (Upper panel) Schematic diagram of intact colicin and the T_{1-61}-DNase fusion protein. T indicates the translocation domain; R indicates the receptor-binding domain; L indicates a short linking sequence between the R and C domains; and C indicates the cytotoxic DNase domain. T_{1-61} is the first 61 residues of the translocation domain. The TolB box (see the text) and the Im9-binding site on the colicin and fusion protein are indicated. (Lower panel) Primary sequence of the N-terminal 69 residues of the T_{1-61}-DNase fusion protein. Residues from the T_{1-61} region and the eight-residue thrombin cleavage sequence whose ^1H-^{15}N NH resonances have been assigned are indicated by uppercase lettering. The position of the TolB box is also indicated. The thrombin cleavage sequence was intended to enable proteolytic release of the T_{1-61} peptide. However, the fusion protein without cleavage was used for experiments reported here because of poor yields of stable intact T_{1-61} peptide following proteolysis.

the susceptible cell, and subsequent cell killing by enzymatic cleavage of nucleic acids in the cytoplasm or pore formation in the inner membrane (*13*−*16*). For colicin E9 (Figure 1), the killing activity is contained in the C-terminal domain, while the central section contains the receptor-binding domain and the N-terminal region is responsible for translocation of the cytotoxic domain into the target cell (*13*). A key step in the translocation mechanism for colicin E9 is its interaction with the 43-kDa TolB, which is located in the periplasm of the target cell. This interaction involves a pentapeptide sequence in the colicin from residues 35−39, Asp-Gly-Ser-Gly-Trp, which is known as the TolB box (*13*, *17*−*21*) and interacts with the β-propeller domain of TolB (*20*, *22*). The TolB box is contained within an 83-residue sequence that has been shown by NMR to be disordered in solution (*11*), consistent with crystallographic studies of the related colicin E3 in complex with its inhibitor protein Im3, which found that the N-terminal 83 residues were not visible in the electron-density map (*23*). NMR studies of the translocation domain of the intact 61-kDa colicin E9 were complicated by the presence of a glycine-rich region from residues 62 to 83, and to characterize the properties of the TolB box region by NMR, this glycine-rich region was deleted from a colicin E9 translocation domain construct. This construct also lacked residues 81−448 of the intact colicin so that the final T_{1-61}-DNase fusion protein[1] contains residues 1−61 of colicin E9 connected to the N terminus of the endonuclease domain of colicin E9 (E9 DNase) by an eight-residue thrombin cleavage sequence (Figure 1). The simplification of the NMR spectra of the N-terminal region containing the TolB box provided by this construct allowed the backbone dynamics of the TolB-binding epitope to be explored in detail (*12*). Here, we report ^1H-^{15}N NMR studies of the dynamical behavior of the Asp35Ala, Ser37Ala, and Trp39Ala variants of the T_{1-61}-DNase fusion protein and their interactions with TolB.

MATERIALS AND METHODS

Sample Preparation. Uniformly ^{15}N-labeled and $^{13}C/^{15}N$-labeled wild-type T_{1-61}-DNase fusion protein was obtained by growing *E. coli* ER2566 (F^- *λ*$^-$ *fhuA2 [lon] ompT lacZ::T7 gene1 gal sulA11* Δ*(mcrC-mrr)114::IS10R(mcr-73::*miniTn*10)2R(zgb-210::*Tn10*)1(Tet*^{S}) endA1 [dcm]) (New England Biolabs Inc.) cells containing plasmid pNP330 [encoding wild-type T_{1-61}-DNase fusion protein and the immunity protein for colicin E9 (Im9), with a histidine tag attached to Im9] in minimal medium containing $^{15}NH_4Cl$ (1 g/L) and $^{15}NH_4Cl$ (1 g/L) with $^{13}C_6$-glucose (4 g/L), respectively. Plasmids pNP332, pNP334, and pNP333, which encode the Asp35Ala, Ser37Ala, and Trp39Ala variants, respectively, were used for transformation of the host cells to express each mutant protein. Expression and purification of the fusion protein−Im9 complexes and isolation of the T_{1-61}-DNase fusion proteins from the His-tagged immunity protein were performed as previously described (*12*). Preparation of unlabeled TolB was carried out as described previously (*20*).

NMR Spectroscopy. All NMR samples contained 50 mM sodium phosphate buffer in 90% H_2O/10% D_2O at pH ∼6.3−6.8 and 0.1% sodium azide. Concentrations of the T_{1-61}-DNase fusion protein for NMR measurements were 50 μM−2.0 mM. All NMR spectra were acquired at 288 K with Varian Unity Inova 500 or 600 spectrometers equipped with triple-resonance pulsed field gradient probes, operating at ^1H frequencies of 499.865 and 599.162 MHz and ^{15}N frequencies of 50.66 and 60.72 MHz, respectively, using pulse sequences incorporated into the Varian (CA) "BioPack" suite of experiments. Resonance assignments were obtained from HNCO, HNCA, CBCA(CO)NH, HNCOCA, HNCACB, (HCA)CO(CA)NH, HCCONH, C(CO)NH, and HNN spectra

[1] Abbreviations: AABUF, average area buried upon folding; Im9, immunity protein for colicin E9; E9 DNase, endonuclease domain of colicin E9; HSQC, heteronuclear single-quantum coherence; DSS, 2,2-(dimethylsilyl)propanesulfonic acid; T_{1-61}-DNase fusion protein, residues 1−61 of colicin E9 connected to the N terminus of the E9 DNase by an eight-residue thrombin cleavage sequence; rmsd, root-mean-square deviation.

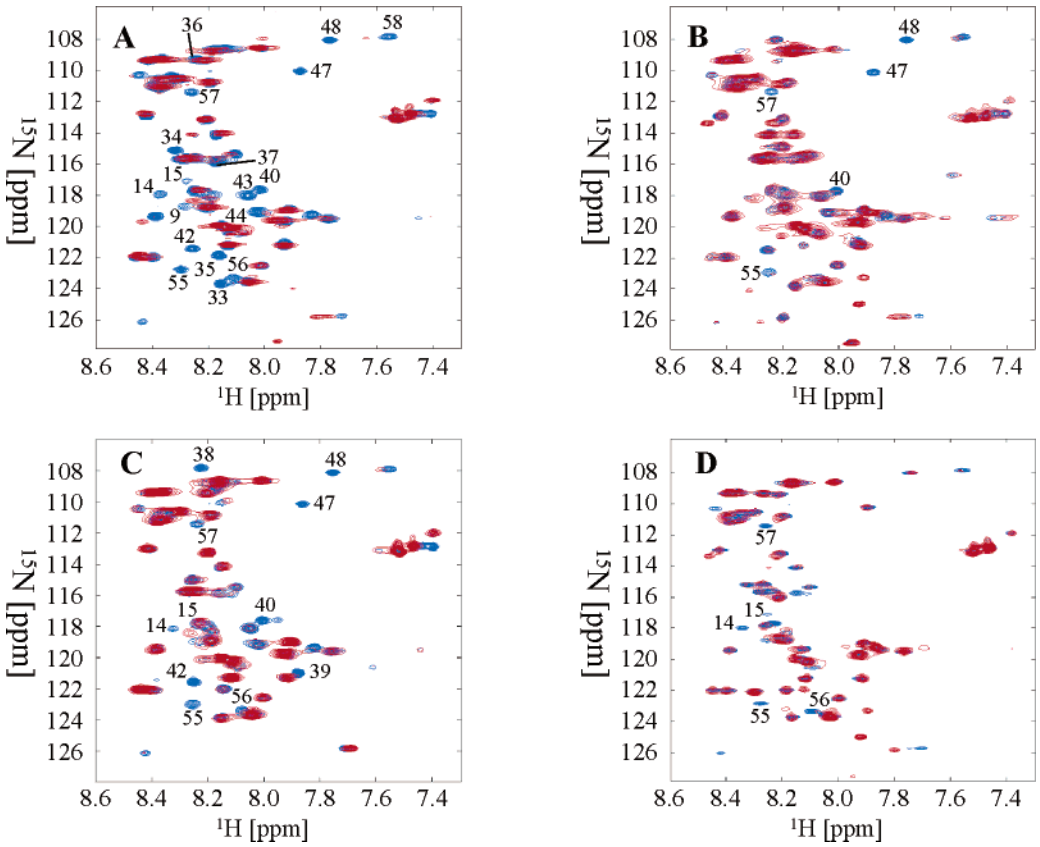

FIGURE 2: Backbone NH region of the 500 MHz ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled T_{1-61}-DNase fusion protein (A) and its Asp35Ala (B), Ser37Ala (C), and Trp39Ala (D) variants in the absence (blue) and presence (red) of unlabeled TolB. The samples of free fusion proteins were at a concentration of 1−1.6 mM in 90% H_2O, 10% $^{2}H_2O$, and 50 mM sodium phosphate buffer (pH 6.5), while the TolB-containing samples were at lower concentrations in 90% H_2O, 10% $^{2}H_2O$, and 50 mM sodium phosphate buffer (pH 6.8) because of relatively poor solubility of TolB. For the TolB interactions, the fusion protein concentrations were in the range of 0.067−0.2 mM, with a 1:2 molar ratio of fusion protein/TolB in all cases. The spectral overlay was carried out in NMRLAB (32).

(*24*, *25*). One-dimensional data were processed using Varian VNMR software and FELIX 95.0 (Biosym/MSI, CA), and multidimensional data were processed using NMRPipe (*26*). Prior to Fourier transformation, a cosine-bell window function was applied to each dimension for apodization. The indirect dimensions were first linear-predicted to double the number of data points and then zero-filled to round up the number of data points to the nearest power of 2. ^{1}H chemical shifts were referenced directly to external 2,2-(dimethylsilyl)-propanesulfonic acid (DSS), and the ^{13}C and ^{15}N chemical shifts were referenced indirectly to DSS (*27*). Spectra were analyzed with XEASY (28) and FELIX 95.0 (Biosym/MSI, CA).

Backbone NH ^{15}N relaxation times at 60.72 MHz were measured with standard procedures (*29*, *30*) using spectral widths of 8000 Hz (^{1}H) and 2200 Hz (^{15}N). Relaxation delays (τ) for the T_2 measurements were 10, 30, 50, 70, 90, 110, 130, 150 or 170, 210, and 250 ms with the experiments at 10, 50, and 150 or 170 ms repeated. The relaxation delays (τ) for the T_1 measurements were 10, 50, 80, 200, 500, 750, 1250, 2000, and 3500 ms, with the experiments at 10, 200, and 500 ms repeated. For all measurements, a pulse delay of 4 s was used. Heteronuclear NOE spectra were measured with the procedure described by Farrow et al. (*29*) as 1024 × 128 complex data points with 32 transients per point. Proton saturation was achieved with a pulse train of 120° pulses every 5 ms for 3 s. Steady-state NOE values were determined from spectra recorded in the presence and absence of proton saturation. For the spectra recorded with proton saturation a 2 s, relaxation delay was followed by the period of saturation, while those recorded without proton saturation used a relaxation delay of 5 s. For the determination of peak height uncertainties, three sets of the saturated/unsaturated experiments were run. Relaxation times and heteronuclear NOEs were calculated as described previously (*12*). Reduced spectral density analyses were carried out following the procedure of Mandel et al. (*31*) to determine the values of $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ as previously described (*12*).

RESULTS

Resonance Assignments for the T_{1-61} Regions of the Fusion Proteins. We previously reported assignments for 53 of the expected 58 backbone NH resonances for the first 61 residues of the wild-type sequence of the T_{1-61}-DNase fusion protein and the 7 backbone NH resonances of the linker sequence (*12*). In the present work, corresponding assignments were obtained for the Asp35Ala, Ser37Ala, and Trp39Ala variant proteins with a combination of triple-resonance experiments using ^{13}C/^{15}N-labeled samples and ^{1}H-^{1}H-^{15}N-NOESY−heteronuclear single-quantum coherence (HSQC) spectra. As previously indicated (*12*), we found the HNN experiment (*25*) to be invaluable for these assignments, particularly in distinguishing between glycine resonances. Sequence-specific assignments are indicated in Figure 2 for selected signals. Variation in the concentration of the T_{1-61}−DNase fusion

protein over the range 50 μM−2.0 mM did not change the chemical shifts of the ^{1}H-^{15}N HSQC resonances.

The chemical-shift differences for the T_{1-61} resonances of the wild-type sequence from their sequence-corrected random-coil values (*33*) indicated that this region lacked stable secondary structure (*12*) and has a preference for populating the β region of (φ,ψ) space, which is normal for random polypeptide chains (*34*). The effect of the mutations did not change these findings, although the mutations had a more far-reaching effect on resonance chemical shifts than just to groups close in sequence to the site of mutation, as illustrated by the sum of the normalized absolute ^{1}H and ^{15}N residue-specific chemical-shift differences between the mutants and the wild-type protein given by (Figure 3)

$$|(\Delta\delta H)/(\Delta\delta H_{max})| + |(\Delta\delta N)/(\Delta\delta N_{max})| \qquad (1)$$

where $\Delta\delta = \delta_{mutant} - \delta_{wild\ type}$.

As expected, the mutated residues and their sequentially nearest neighbors have the largest chemical-shift differences. Excluding these data, the maximum chemical-shift differences for assigned resonances were ±0.065 ppm (^{1}H) and ±0.25 ppm (^{15}N), and the average $\Delta\delta$ values for the Ser37Ala, Asp35Ala, and Trp39Ala variants, respectively, were 0.083, 0.115, and 0.116 ppm. Perturbations significantly greater than these averages were seen for regions far removed from the mutation site; specifically, resonances of residues 55 and/or 56 were affected by the mutations, and residues 13 and 14 for two of the variants were also affected (Figure 3). These chemical-shift changes for residues distant in sequence from the mutation sites indicate the presence of networks of nonlocal contacts involving sequentially distant residues that have been brought into spatial proximity to form structural elements. The presence of structural elements is supported by the sequence dependence of the {^{1}H}-^{15}N heteronuclear NOE and ^{15}N T_2 relaxation times (Figure 4), as discussed below.

Polypeptide Chain Dynamics. Backbone dynamics of the variant fusion proteins were investigated with ^{15}N T_1, ^{15}N T_2, and {^{1}H}-^{15}N heteronuclear NOE data recorded at 60.72 MHz (Figure 4). The data indicate flexibility throughout the proteins, as is most clearly indicated by the {^{1}H}-^{15}N NOEs, which are in the range from +0.4 to −0.8. These values are considerably lower than the +0.82 expected for NH groups of a rigid globular protein that is tumbling isotropically (*35*). Similar behavior has been observed for other unfolded proteins, including urea-unfolded apomyoglobin (*36*), acid-unfolded apomyoglobin (*37*), and urea-unfolded lysozyme (*38*), as well as the N-terminal regions of the T_{1-61}-DNase fusion protein with a wild-type sequence (*12*) and the intact colicin E9 (*11*). As with the intact colicin E9 and the T_{1-61}-DNase fusion protein with a wild-type sequence (*11*, *12*), the sequence variation in ^{15}N T_2 values for the fusion protein variants reveals a marked change coincident with the TolB box pentapeptide (residues 35−39) with the residues N-terminal to this having longer T_2 values than those on the C-terminal side. The sequence variation in {^{1}H}-^{15}N NOE values also shows a change at the TolB box.

In principle, the sequence variation in relaxation parameters might result from either intra- or intermolecular interactions. However, we exclude intermolecular interactions as the causative factor because, as noted above, the chemical shifts of the ^{1}H-^{15}N HSQC resonances were independent of the concentration and also because the ^{15}N R_2 values were concentration-independent (data not shown).

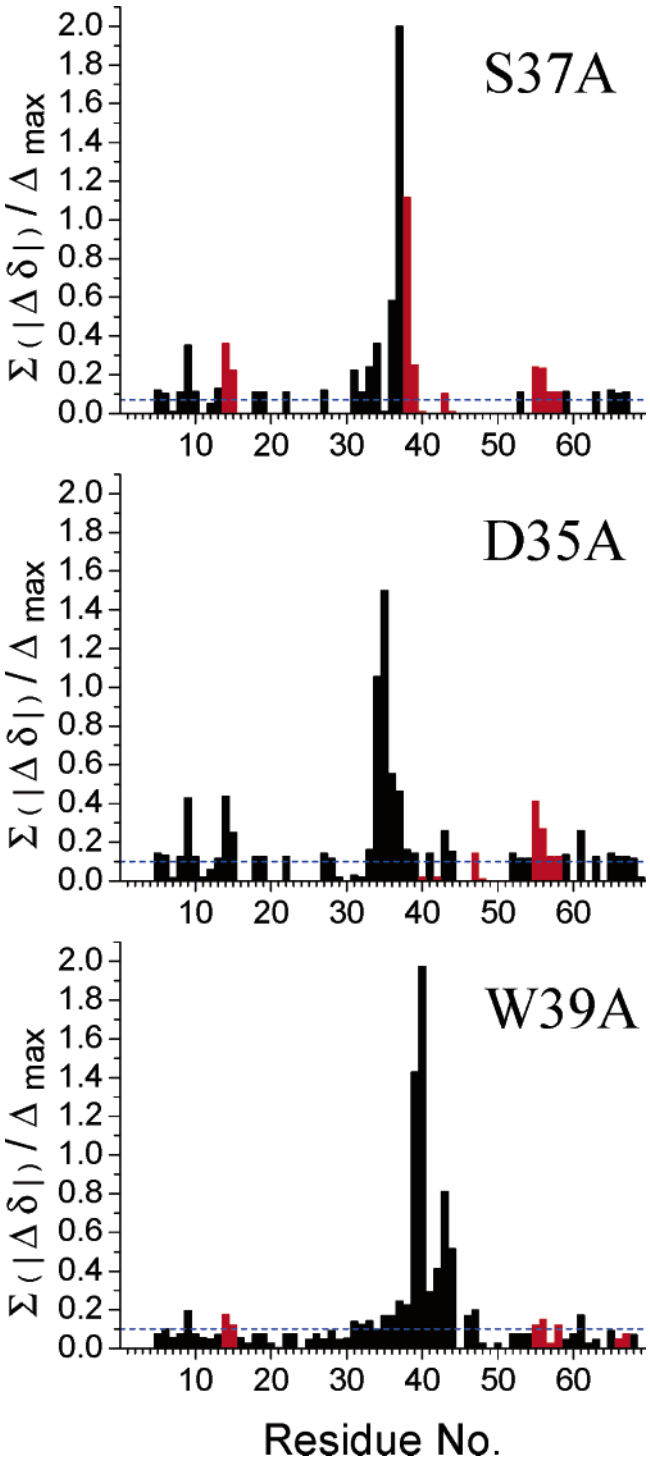

FIGURE 3: Chemical-shift differences between the three (Asp35Ala, Ser37Ala, and Trp39Ala) mutants and the T_{1-61}-DNase wild-type proteins. Differences are calculated as the sum of the ^{1}H and ^{15}N differences for each residue as a proportion of the maximum respective difference. Broken horizontal lines indicate the mean shift differences omitting residues exhibiting the largest ^{1}H or ^{15}N shift differences and their immediate sequential neighbors. Data points in red are for residues strongly affected by TolB binding judged by the appearance of ^{1}H-^{15}N HSQC spectra (Figure 2).

The relaxation data have been analyzed using the reduced spectral density method (*31*, *39*, *40*), which is the normal

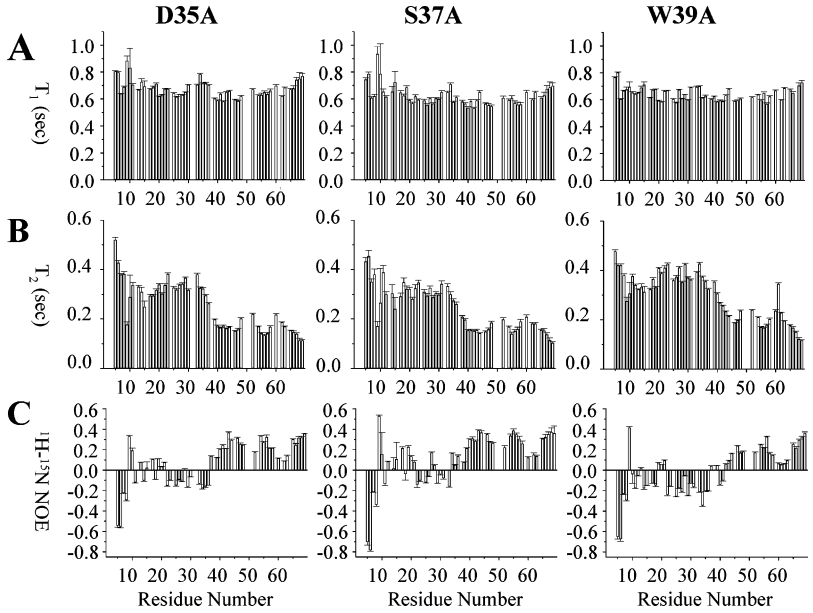

FIGURE 4: Backbone 60.72 MHz ^{15}N T_1 (A) and T_2 (B) relaxation times and {^{1}H}-^{15}N NOEs (C) of T_{1-61}-DNase fusion protein variants in 90% H_2O and 10% ^2H_2O (50 mM sodium phosphate buffer at pH 6.3) at 288 K.

approach for unfolded proteins because the Lipari−Szabo formalism (*41*) is not applicable to these (*39*, *40*, *42*). The sequence variation of the functions $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ are shown for the variant fusion proteins in Figure 5 along with the corresponding functions for the T_{1-61}-DNase fusion protein with a wild-type sequence (*12*). As noted previously for the wild-type sequence (*12*), the profiles of $J(0)$ for the mutants mirror the sequence variations of T_2 (Figure 4).

TolB Binding to T_{1-61}-DNase Fusion Proteins. The effect of adding TolB to samples of the fusion proteins can be seen in the ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled DNase fusion proteins with and without unlabeled TolB at the fusion protein/TolB molar ratio of 1:2 (Figure 2). The spectra are displayed at a relatively high threshold level so that the broad peaks of the DNase domain are not visible. The effect of TolB binding is to broaden some of the resonances of the T_{1-61} region so that they too are not visible at the threshold level of the spectra; Figure 6 provides a summary of the affected resonances to aid the Discussion later in the paper.

Resonances of fusion protein residues that are affected by TolB binding are also indicated in red in Figures 3, 7, and 8. Increasing TolB relative to the fusion protein up to a molar ratio of 1:4 did not cause any further changes in the fusion protein spectra, indicating that the spectra of the mixtures in Figure 2 are for the fully bound fusion proteins.

The broadening experienced by resonances of residues in the T_{1-61} region of the wild-type fusion protein on binding TolB is probably because of a reduction in their backbone motions. For the wild-type T_{1-61} DNase fusion protein, 21 resonances from the N-terminal 61 residues were affected by TolB, spanning from His 9 to Gly 58. Residues of the fusion protein with a wild-type sequence whose resonances were perturbed were also affected by TolB binding to intact colicin E9 (*11*), consistent with the fusion protein having the same TolB-binding properties as the intact colicin.

The fact that the presence of TolB affects the spectra of the three variants of the T_{1-61} fusion protein (Figure 2) indicates that TolB binds to all three proteins despite the corresponding mutations making colicin E9 nontoxic to cells (*18*). Thus, it appears that the mutations either lead to the formation of nonproductive complexes or reduces the binding affinity of the colicin for TolB sufficiently so that significant interaction between the two does not occur under physiological conditions. Note that the NMR experiment is able to detect weak interprotein complexes with K_d values as low as millimolar, while SPR experiments, in which binding of TolB to the Asp35Ala, Ser37Ala, and Trp39Ala mutant colicins could not be detected (*21*), are not sensitive to such weak interactions. The pattern of TolB-induced perturbations suggest that TolB interacts with the Ser37Ala variant in a similar way to how it interacts with the wild-type colicin but that it interacts differently with the Asp35Ala and Trp39Ala variants, most notably in that residues of the TolB box are not affected by TolB in these latter two variants (Figure 2). The only common feature for the wild-type protein and the three variants is that TolB binding perturbs Trp 56 and neighboring residues.

The effects of TolB on the backbone dynamics of the T_{1-61} region of the fusion protein were monitored by ^{15}N T_2 measurements. Figure 7 shows the sequence-dependent R_2 data for the T_{1-61}-DNase fusion protein−TolB complex with the data for the unbound fusion protein overlaid; the majority of the resonances affected by TolB binding were too broad for reliable R_2 rates to be obtained. For the complex, from Ser 34 toward the N terminus until residue 15, the effect of TolB binding was to increase R_2. This region exhibited some degree of restricted dynamics compared to the unbound T_{1-61}-DNase fusion protein but was still more flexible than the DNase domain of the fusion protein (*12*). Residues 5−13 of the bound fusion protein remain as flexible as they are in the free protein.

DISCUSSION

T_{1-61} Polypeptide Chain Dynamics. Dynamics of a polypeptide chain can be deduced from the backbone NH relaxation parameters T_1, T_2, and ^{1}H-^{15}N NOE through the use of the reduced spectral density functions $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$. The magnitudes of these functions are sensitive to motions

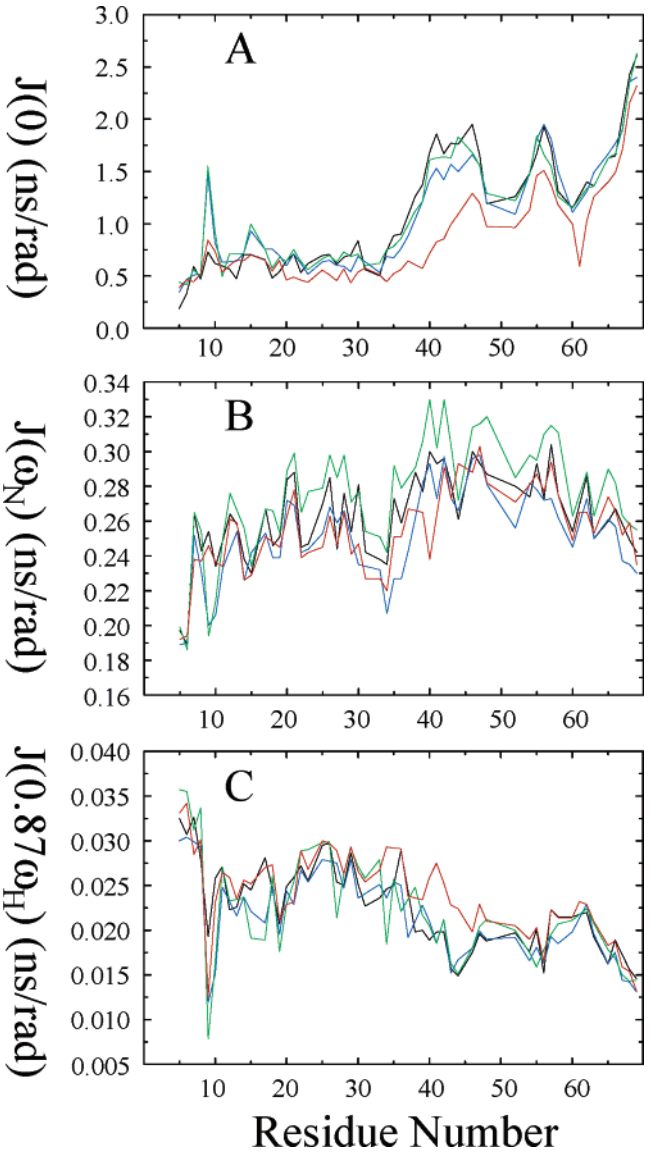

FIGURE 5: Sequence variations of the $J(0)$ (A), $J(\omega_N)$ (B), and $J(0.87\omega_H)$ (C) spectral density values of backbone NH groups calculated by the reduced spectral density mapping procedure (*30*) from the relaxation data in Figure 4 for the Asp35Ala (blue), Ser37Ala (green), and Trp39Ala (red) variants of the T_{1-61}-DNase fusion protein and, for comparison, the corresponding data for the T_{1-61}-DNase fusion protein with the wild-type sequence (*12*) shown in black.

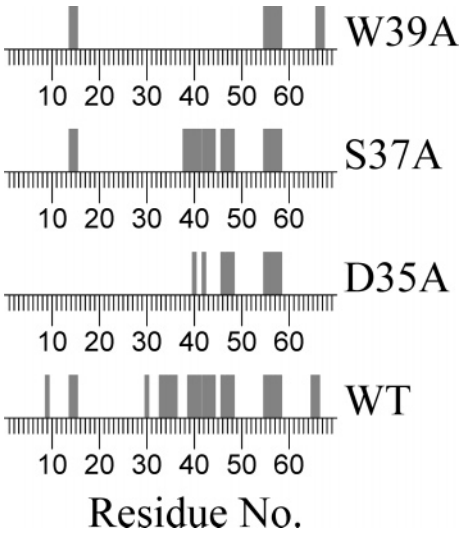

FIGURE 6: Summary of residues of the wild-type and variant T_{1-61}-DNase fusion proteins whose ^1H-^{15}N HSQC backbone NH resonances are affected by the presence of TolB. The data were obtained from the spectra presented in Figure 2.

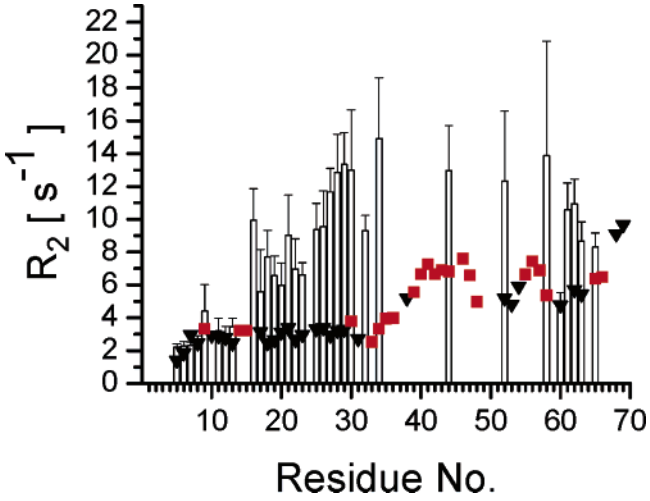

FIGURE 7: Histogram plot of R_2 values for T_{1-61}-DNase bound to TolB. The sample of the complex contained 55.5 μM T_{1-61} fusion protein with a wild-type sequence and 170.8 μM TolB in 50 mM phosphate buffer (pH 6.8). R_2 values for the free protein taken from Figure 4 are indicated as black triangles and red squares. Data points shown as red squares are for residues strongly affected by TolB binding judged by the appearance of ^1H-^{15}N HSQC spectra (Figure 2).

at the corresponding frequencies, with $J(0)$ reflecting slow internal motions on the millisecond−microsecond time scale as well as slow global rotational diffusion, $J(0.87\omega_H)$ reporting on the presence of internal motions on the picosecond time scale, and $J(\omega_N)$ lying between these extremes. As can be seen from Figure 5, $J(\omega_N)$ is rather uninformative about the dynamics of the T_{1-61} region but $J(0)$ and $J(0.87\omega_H)$ are more revealing. $J(0)$ is ~0.5 ns rad^{-1} at the N terminus of all of the proteins, highlighting the extreme flexibility of the T_{1-61} chain. Moving along the sequence of the wild-type fusion protein, the high degree of flexibility is maintained right up until residue 35. Then, over the next 10 residues, $J(0)$ increases to a maximum of 2.0 ns rad^{-1} at residue 45, showing that the local segmental motions are decreasing and the correlation time is increasing, as a consequence of some constraint on the local dynamics. Between residues 45 and 55, $J(0)$ shows a U-shaped dip toward increased local dynamics around residue 50, indicating that the local ordering is not fully maintained throughout this part of the sequence. However, because the minimum value for $J(0)$ of ~1.0 ns rad^{-1} is still greater than that seen for the initial 35 residues, some ordering must remain. Between residue 55 and 56, ordering is also present because $J(0)$ is ~2.0 ns rad^{-1} at residue 56, only decreasing to a minimum of ~1.2 ns rad^{-1} at residue 60. Beyond residue 60, $J(0)$ increases to the level of the globular E9 DNase domain. Mirroring the trends in $J(0)$, $J(0.87\omega_H)$ reveals faster motions in the first 35 residues of the T_{1-61} sequence than in the regions identified as being more structured, as indicated by a value of ~0.025 ns rad^{-1} for the first 35 residues compared with ~0.015 ns rad^{-1} for residues further along the sequence.

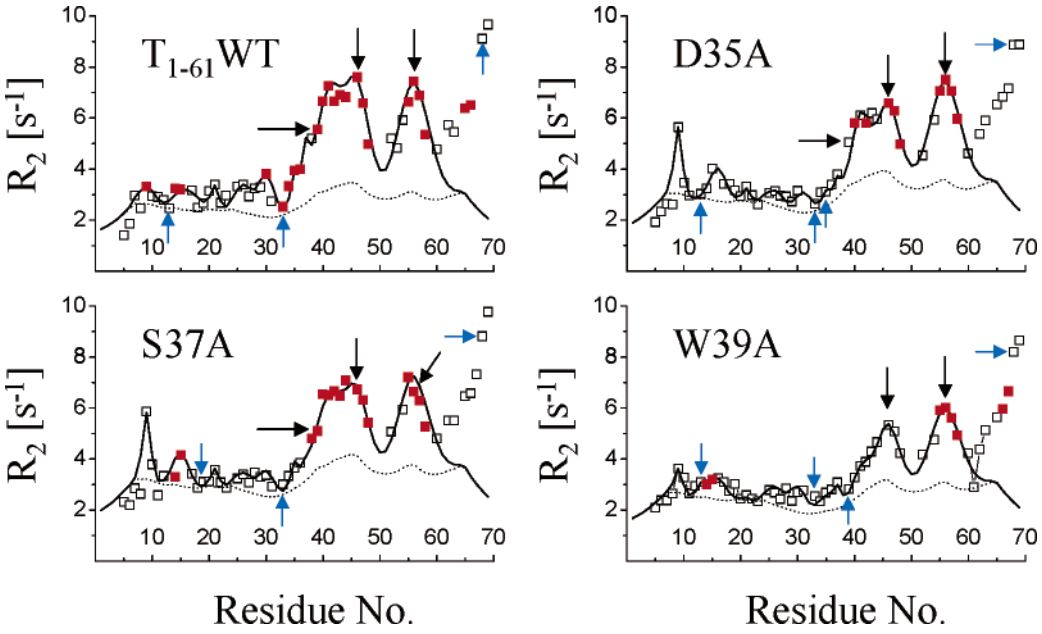

FIGURE 8: Plots of experimental R_2 values (squares) and the results of the cluster analysis (−) from eq 2 for the DNase fusion proteins. Broken lines represent the intrinsic relaxation of the polypeptide as described by the first term of eq 2. Data points in red are for residues strongly affected by TolB binding judged by the appearance of ^{1}H-^{15}N HSQC spectra (Figure 2). Black and blue arrows indicate the position of tryptophan and alanine residues, respectively.

The general trends in $J(0)$ seen for the Asp35Ala and Ser37Ala variants are similar to those seen for the wild-type sequence. Two marked differences are present though. First, residues 9, 15, and 16 in the variants have values for $J(0)$ in the range of 0.92−1.47 ns rad^{-1}, in contrast to the values of 0.5−0.6 ns rad^{-1} for the wild-type sequence and so does residue 13 for the Ser37Ala variant. These changes are consistent with an increase in the local structure accompanying the mutation. Second, $J(0)$ no longer increases at residue 35 but at residue 36 or 37. Hence, the local structure associated with residue 35 is destroyed in both variants. However, this does not affect the $J(0)$ maximum position at residue 45 nor the subsequent profile seen for $J(0)$. Thus, apart from some residues further toward the N terminus, the effects of the mutations are limited to the actual amino acid replaced and its near neighbors. In contrast, the mutation of Trp39 to alanine has a profound effect on $J(0)$, and instead of an increase in $J(0)$ around residue 35 as with the other fusion proteins, $J(0)$ only increases after residue 40. Also, the degree of structure within this local region is lower than in the other proteins as shown by the maximum value for $J(0)$ at residue 45 of 1.3 ns rad^{-1}. Furthermore, although the second region of the structure around 55−60 is preserved, the degree of structure within this region is also decreased, with a maximum $J(0)$ of only 1.51 ns rad^{-1}. In addition, the increase in $J(0)$ for residue 9 is much smaller than for the other mutants, with no evidence for an increase around 15 and 16.

Order in the Otherwise Disordered Polypeptide Chain. Further insight into the dynamics of the disordered regions of polypeptides can be gained through fitting of the ^{15}N T_2 relaxation data to models for polypeptide motion, as has been reported previously (*12*, *36*, *38*, *43*). These models generally assume that the effect on the motion of an amino acid because of neighboring residues decreases exponentially as the distance between them increases. Thus, the predicted sequence-dependent intrinsic relaxation rates would appear to form a central plateau that tails-off at the N and C termini. However, the overall glycine and alanine content of the regions of the proteins studied in the present paper was typically 40% with several regions of higher localized glycine content (up to 80%), and therefore, the model of Schwarzinger et al. (*36*) was used. This is a development of the dynamics model of Schwalbe and his colleagues (*43*) that we have used previously for the intact colicin E9 and wild-type T_{1-61}-DNase fusion protein (*12*). The Schwarzinger et al. (*36*) model takes into account the small size of glycine and alanine residues, thereby significantly modifying the predicted underlying polypeptide motional properties. The sequence-dependent R_2 relaxation rates are then given by eq 2

$$R_{2i} = k\cdot\sum_{j=1}^{N}\tau_j\cdot e^{(-|i-j|)/(\lambda_j)} + \sum_{\text{cluster}} R_{\text{cluster}}\cdot e^{(-(i-n_c)^2/(w_c)^2)} \qquad (2)$$

where the first term accounts for the intrinsic relaxation of the polypeptide and the second term describes additional interactions arising from local clusters of residues. R_{2i} is the relaxation rate of residue i; N is the total number of residues; and τ_j and λ_j are the intrinsic correlation time of residue j and the persistence length for segmental motion of the polypeptide, respectively. After Schwarzinger et al. (*36*), the intrinsic correlation time of a given residue was described using its radius of gyration, R_g, and taken to be proportional to R_g^3, and λ_j was assumed to be 2 for glycine and alanine and 7 for all other residues. The additional clusters were then defined by an intrinsic relaxation rate R_{cluster} centered at residue n_c with a half-width of w_c.

The results of applying this model to the wild-type T_{1-61}-DNase fusion protein and its Asp35Ala, Ser37Ala, and Trp39Ala variants are shown in Figure 8. No attempt was made to extend modeling beyond residue 69 because the properties of the highly structured DNase domain then begin

to dominate the R_2 relaxation rates (*12*). The intrinsic relaxation of the polypeptides determined by the first term of eq 2 are shown as broken lines in Figure 8, which shows that the experimental data follow the general trend, in particular reflecting the predicted peaks between residues 35 and 60. The polypeptides are therefore naturally predisposed to form localized regions of varying dynamics as a result of the distribution of glycine and alanine residues. This natural predisposition is then further enhanced by the interactions of various residues to form clusters with more restricted dynamics that are accurately modeled by inclusion of the second term of eq 2 (Figure 8).

The modeling reveals the presence in all three variants of similar clusters to those of the wild-type protein, having relaxation rates $R_{cluster}$ of $> \sim 1$ or $< \sim 1$ s^{-1}, respectively, for the major and minor clusters. Other than one extreme N-terminal cluster, the major clusters are centered in or close to the TolB box from Asp 35 to Trp 39 (*13*, *17*−*20*), at residues 35, 37, and 41 and at residues 46 and 56. These latter two clusters are centered on tryptophan residues and are the most rigid of all of the clusters having the highest $R_{cluster}$ rates and w_c values (Table 1 of the Supporting Information). Clusters at residues 35, 37, and 41, although centered on the smaller hydrophilic residues aspartate and serine, respectively, nevertheless surround the tryptophan at position 39 within the TolB box, which itself has a substantially increased value of R_2. The fact that there is no cluster centered at Trp 39 does not mean that this residue is not integral to the clustering between 35 and 46. Neither does it preclude the hydrophobic nature of this residue from being a trigger for cluster formation in this general region. However, the fact that the relaxation enhancement is not maximum compared to the neighboring residues indicates this is an example of cluster formation not solely dependent on amino acid hydrophobicity. This suggests that the nature of a cluster is more complex and arises from the interplay of multiple physicochemical properties of the interacting residues within the network. Including further minor clusters toward the N terminus of each protein considerably improves the overall fit of the simulation and leads to a root-mean-square deviation (rmsd) between the model and experimental data for the wild-type protein of 0.38. Similar rmsd values of 0.34, 0.4, and 0.24 were obtained for the Asp35Ala, Ser37Ala, and Trp39Ala mutants, respectively. These minor clusters are largely constructed of polar rather than hydrophobic residues.

Although the strongest clusters are centered at tryptophan residues, it nonetheless seems that the nature of the clusters observed in these systems cannot be accounted for solely by hydrophobic clustering. Consistent with this view, neither the Kyte and Doolittle approximation (*44*) nor hydrophobic cluster analysis using HCA_Draw (*45*) revealed the presence of any inherent hydrophobic clusters. However, the average area buried upon folding (AABUF), which is proportional to the hydrophobic contribution of a residue to the conformational free energy of a protein (*46*) and has been shown to correlate with sequence-dependent dynamic variations in urea-unfolded apomyoglobin (*36*) does appear to be associated with the formation of at least the major clusters in the T_{1-61}-DNase fusion protein, because the increase in R_2 mirrors an increase in AABUF (Figures 5 and 9). There is

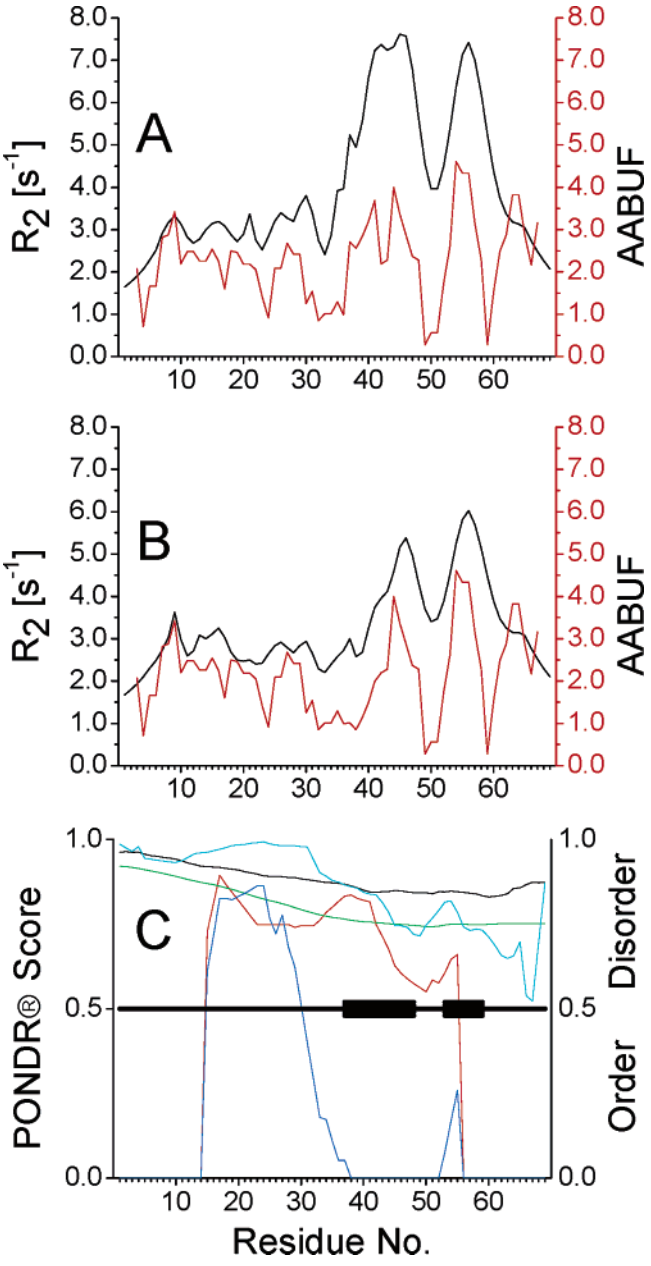

FIGURE 9: Plots of average area buried upon folding (*45*) in red for the T_{1-61}-DNase fusion protein (A) and its Trp39Ala variant (B) with their corresponding cluster analysis results from Figure 7 in black. AABUF values were calculated with the ExPASy tool ProtScale (http://us.expasy.org/tools/protscale.html) and normalized from 0 to 10. The outputs of various PONDR predictions (http://www.pondr.com) are shown in C with the VL-XT results as cyan, VSL1 results as black, VL3 results as green, XL1 results as red, and CaN results as blue. Also indicated by thick horizontal black bars are those residues exhibiting enhanced NMR R_2 rates.

no correlation of AABUF with $J(0.87\omega_H)$ or $J(\omega_N)$ but there is with $J(0)$. The effect of the Trp39Ala mutation on the AABUF and the corresponding correlation with the reduced clustering is also apparent (Figure 9), but similar plots of AABUF for the Asp35Ala and Ser37Ala mutants showed no significant differences between these mutants and the wild-type protein; however, there are clearly subtle variations in the clusters revealed by modeling the experimental R_2 data for these proteins (Figure 8).

When the R_2 data and clusters of the DNase fusion proteins are compared, the underlying importance of Trp 39 of the wild-type sequence is apparent from the severe disruption

to the clustering of the Trp39Ala variant, with $R_{cluster}$ rates being reduced to 26 and 37% of those of the wild-type sequence at residues 41 and 37 and to 56% at residue 46. Although replacement within the TolB box of the smaller polar residues, Asp35 or Ser37, with an alanine has less severe effects on the major clusters, they are nevertheless affected. Strikingly, the R_2 rates and cluster parameters $R_{cluster}$ and w_c are reduced, not only at the site of the mutation but also within the nearby clusters centered at residues 41, 46, and 56, suggesting some form of association between clusters. Most notable though is the appearance of clusters at His9 and Thr16. If clusters are stabilized by mutual association, then it would seem that to compensate for the loss of stability within the cluster containing the mutation new residues are co-opted into the cluster arrangement. Presumably, this enhances the stability of the clusters, possibly from polar interactions involving hydrogen-bonding networks because 4 of the 10 clusters are centered on serine or threonine residues, which can mask their polarity through hydrogen bonding. Also, although individual clusters have $R_{cluster}$ values between ~1 and ~4 s^{-1}, once these values are added to the underlying dynamics of the polypeptide, the result is to produce localized regions with similar maximum R_2 values (Figure 8), further indicating the occurrence of cooperative actions between clusters.

TolB Binding to the T_{1-61} Region of Colicin E9. From our previous NMR studies of intact colicin E9 (*11*), we proposed extending the TolB interaction region of colicin E9 from the pentapeptide TolB box sequence, Asp 35−Trp 39, to the 12 amino acids from Ala 33 to Asn 44. The effect of TolB on the ^1H-^{15}N HSQC spectrum of the T_{1-61}-DNase fusion protein with the wild-type TolB box is consistent with this, and with the improved resolution and increased number of resonance assignments for this protein compared to the intact colicin (*12*), it is clear that resonances of residues neighboring this region of the sequence are also affected by TolB, notably those of Ser 30, Gly 31, Gly 32, and Gly 48. Mutational analysis of the TolB box region of colicin E9 supports an extended TolB box, with S40A, E42A, and W46A mutations in colicin E9 inactivating biological activity and TolB binding (21). The perturbations to the ^1H-^{15}N HSQC spectra of the variant forms of the T_{1-61}-DNase fusion protein caused by the presence of TolB (Figure 2) show that, although the mutations render these proteins inactive as a toxin (*18*), TolB does bind to them, although not necessarily at the TolB box, at least for the Asp35Ala and Trp39Ala variants.

Interpreting the TolB-induced spectral perturbations to identify the TolB-binding site on colicin E9 is not possible because the perturbations may be a direct consequence of a group being in contact with bound TolB or a secondary effect of TolB-induced structural changes to the conformational ensemble of the T_{1-61} region of the fusion protein. However, although it is not possible to distinguish between these two with the NMR data of Figure 2 alone, in conjunction with the other NMR data presented herein and with previously reported mutagenesis and biophysical studies (*18*, *20*), they do allow a plausible model for the interaction of the T_{1-61} region of colicin E9 and TolB to be proposed.

The chemical shifts of the residues forming the minor clusters at the N-terminal end of the fusion protein are affected by TolB, but the backbone dynamics of this region reflected by ^{15}N T_2 measurements are unaffected by TolB (Figure 7). Therefore, in this case, it is most likely that structural changes to the T_{1-61} conformational ensemble are responsible for the perturbations seen in Figure 2 rather than this region forming part of the TolB-binding site. The cluster of residues centered on Trp 56 (Figure 7) is likely to be perturbed by TolB as a direct effect of its binding because resonances of these residues in the Asp35Ala and Trp39Ala variants are affected by TolB and the TolB box residues are not. However, colicin E9 lacking residues 54−164 of colicin E9 is able to bind TolB (*20*), and the Trp56Ala colicin E9 is toxic (*18*); therefore, this region is not essential for productive TolB−colicin E9 interaction. In summary then, the data suggests that the region including Ser 34−Asn 44 contains the main TolB-binding site and the TolB-induced perturbations for resonances further toward the N terminus are a result of conformational changes. Note that we are not able to precisely delineate the binding region on colicin E9 for TolB from the data available so that it may run some residues beyond 34−44. However, the NMR data (Figures 6 and 7) indicate that it does not extend beyond residues 33 and 48.

The issue of how the stretch of residues from Ser 34 to Asn 44 interacts with TolB is an important one, but the NMR data does not provide any direct evidence on this because the resonances of residues from Ser 34 to Asn 44 are not detected for the bound fusion protein. However, it is possible to make a reasonable suggestion. A run of 11 residues is too small to fold into a globular domain, but the conformational preferences of the amino acids suggest that this region will not be a random extended chain. The amino acids with the highest propensity for forming β turns are glycine, asparagine, proline, serine, and aspartate (*47*), which are 9 of the 11 residues in this sequence. Thus, we expect it to form one or more β turns, and predictions using BTPRED (*48*) suggest that it will consist of 2−4 β turns. Therefore, we propose that the TolB-binding sequence forms a series of turns and runs across the surface of TolB. There appears to be a precedent to this kind of structure with the $\alpha_2\beta_2$ protein methanol dehydrogenase (*49*, *50*). This consists of two domains, a 66-kDa eight-bladed β-propeller catalytic domain and a 8.5-kDa domain of unknown function that is essential for activity. Ghosh et al. (*50*) note that the small domain has an unusual structure by being layered across the surface of the β propeller in an extended fashion and without a hydrophobic core. The N-terminal 30 residues form a series of open turns with the remainder of the protein mostly in a single α helix. Ghosh et al. (*50*) suggest that the structure of the small domain is determined largely by its interaction with the β propeller. In summary, this is our proposal too for the T_{1-61} region of colicin E9 in its interaction with the six-bladed β-propeller domain of TolB.

The model that we have put forward for the interaction of TolB with the translocation domain of colicin E9 is a good example of the kind of system described by Fuxreiter et al. (*51*). These authors suggest that intrinsically disordered proteins that interact with structured partners often have preformed structural elements that act as the initial contact points for binding their partners and that the initial binding of the partners to these contact points reduces the dimensionality of the folding process. Although Fuxreiter et al. (*51*) focused on proteins whose preformed structural elements were likely to be helices, their scheme is not exclusive to

these and does encompass amino acid clusters of the kind that we have described for colicin E9.

Prediction of Disorder from the Colicin T_{1-61} Amino Acid Sequence. Analyses of amino acid sequence data to predict disorder is of growing importance as more genome sequences become available and predictive tools become more sophisticated. A leading set of predictors of natural disordered regions are the PONDR suite of programs (*2*, *52*), of which PONDR VL-XT is regarded as the most accurate predictor of disorder (*53*−*55*). This combines three neural network predictors (NNPs): VL1, which begins and ends 11 residues from the termini, and the XN and XC predictors, which begin at the respective N or C terminus and continue inward for 14 residues. Application of the PONDR approaches to the colicin E9 translocation region provides both a test of the PONDR predictors and additional insights into the nature of this region.

The results of using PONDR to predict disordered regions within the sequence of the wild-type T_{1-61} fusion protein are shown in Figure 9C. PONDR VL-XT predicts that the N-terminal 69 residues are completely disordered throughout. Therefore, because one continuous region of disorder is predicted, it is expected that the rate of false positive results (i.e., predicted disorder when the residue or region is ordered) would be less than 0.1%. The additional NNPs, VL3 and VSL1, concur with the general prediction of overall disorder with perhaps a slight tendency toward less disorder toward the C-terminal region of residues 1−69, particularly after approximately residue 30. The XL1 NNP does not provide predictions for the N- and C-terminal 14 residues and is optimized to predict regions of disorder containing greater than 39 residues (*53*). This algorithm predicts a single disordered region of 41 residues and would thus also be expected to be reasonably accurate. It is known that this NNP is less accurate than VL-XT at predicting disorder, but it is more accurate at predicting order; therefore, it is most interesting to note that the region C-terminal to residue 30 is now predicted to have an even greater tendency toward order than with the other PONDR NNPs, especially in the region between residues 44 and 55. NNPs trained on a broad data set of molecules containing regions that undergo disorder−order transitions upon complexation will naturally recognize those regions as disordered. However, CaN NNP, which is based on data sets of families of disordered proteins that fold on binding, has been shown (*56*) to identify binding sites within disordered sequences through apparent false negative predictions, i.e., predicting order when a residue is disordered. As can be seen from Figure 9, the CaN algorithm does indeed make false negative predictions for the wild-type T_{1-61} sequence, with an extreme minima centered on Trp 46. Significantly, this coincides with one of the strongest clusters identified from NMR relaxation rates and cluster analysis, as described above. The degree of correlation between the CaN prediction and the cluster analysis can be seen in Figure 9 by comparing those residues (*37*−*48*, *53*−*59*) that have significantly enhanced R_2 rates determined by the Schwarzinger et al. (*36*) model of polypeptide dynamics (i.e., $R_2 > 5.0 \pm 0.1$ s^{-1}) with the CaN extreme minima region between residues 38 and 52. Of the 14 extreme minima residues, 9 have been observed by NMR to be involved in binding to TolB (Figure 6).

In summary then, PONDR predictions of the essentially disordered nature of the wild-type T_{1-61} sequence are in agreement with experimental observations and, furthermore, indicate that some residues important for binding the partner protein TolB exhibit a propensity toward order consistent with a local disorder−order transition on binding, with the remainder of the T_{1-61} region staying disordered.

Possible Physiological Relevance of the Disordered Region of Colicin E9. As noted in the Introduction, there are a variety of possible functional reasons for intrinsic disorder of proteins, and often such disorder is associated with intermolecular interactions. In the case of colicin E9, the protein also has to traverse the outer membrane of a target cell and disorder might assist this because globular proteins do not generally pass across membranes unaided. It is a common feature of all structurally characterized colicins that their translocation domains contain some flexible regions, despite there being little primary and tertiary structural similarity between them. Thus, parts of the translocation regions of the TolA-dependent colicin N (15) and the TonB-dependent colicins B (*57*) and Ia (*58*) and not visible in X-ray diffraction electron-density maps. Furthermore, NMR studies show that the colicin N translocation region is intrinsically disordered and folds on binding TolA (*59*). The observation that parts of the TolB-bound N-terminal region of colicin E9 retains considerable flexibility (ref *11* and Figure 7) may be connected to the requirement for functional multiple protein complexes, which seems to be involved in the translocation of colicin A. This protein has a TolB interaction site between residues 1−20 and a TolA interaction site in the region from 34 to 107 (*19*) and forms the ternary complexes of colicin A−TolA−TolB and colicin A−TolA−TolR (*60*). Recently, various ternary complexes of colicin E9 have been detected in SPR experiments (*21*), and the ternary complex of TolB−colicin E9−OmpF has been isolated (*61*). The OmpF appears to interact with the TolB-bound colicin in flexible regions of the N-terminal sequence, which are unaffected by TolB binding (Figure 7). Therefore, the flexible translocation regions of colicins may be like fishing lines for their partner proteins, enhancing the rate at which productive intermolecular interactions occur through a fly-casting mechanism (*6*, *7*).

ACKNOWLEDGMENT

We thank Scott White (Birmingham) for pointing out the apparent similarity of our binding model for colicin E9 and TolB to the structure of methanol dehydrogenase.

SUPPORTING INFORMATION AVAILABLE

Cluster analysis of NMR relaxation data, n_c, w_c, and $R_{cluster}$ for WT T_{1-61}, D35A, S37A, and W39A. See http//www.uea.ac.uk/cap/grmresfoc/biochem.pdf. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: Re-assessing the protein structure−function paradigm, *J. Mol. Biol. 293*, 321−331.
2. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakouchova, L. M., and Obradović, Z. (2002) Intrinsic disorder and protein function, *Biochemistry 41*, 6573−6582.

3. Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics, *Protein Sci. 11*, 739−756.
4. Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins, *Curr. Opin. Struct. Biol. 12*, 54−60.
5. Spolar, R. S., and Record, M. T. (1994) Coupling of local folding to site-specific binding of proteins to DNA, *Science 263*, 777−784.
6. Pontius, B. W. (1993) Close encounters: Why unstructured, polymeric domains can increase rates of specific macromolecular association, *Trends Biochem. Sci. 18*, 181−186.
7. 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.
8. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996) Structural studies of p21Wal1/Cip1/Sdi1 in the free and Cdk2-bound state: Conformational disorder mediates binding diversity, *Proc. Natl. Acad. Sci. U.S.A. 93*, 11504−11509.
9. Gunasekaran, K., Tsai, C.-J., Kumar, S., Zanuy, D., and Nussinov, R. (2003) Extended disordered proteins: Targeting functions with less scaffold, *Trends Biochem. Sci. 28*, 81−85.
10. Dedmon, M. M., Patel, C. N., Young, G. B., and Pielak, G. J. (2000) FlgM gains structure in living cells, *Proc. Natl. Acad. Sci. U.S.A. 99*, 12681−12684.
11. Collins, E. S., Whittaker, S. B−M., Tozawa, K., MacDonald, C., Boetzel, R., Penfold, C. N., Reilly, A., Clayden, N. J., Osborne, M. J., Hemmings, A. M., Kleanthous, C., James, R., and Moore, G. R. (2002) Structural dynamics of the membrane translocation domain of colicin E9 and its interaction with TolB, *J. Mol. Biol. 318*, 787−804.
12. Macdonald, C. J., Tozawa, K., Collins, E. S., Penfold, C. N., James, R., Kleanthous, C., Clayden, N. J., and Moore, G. R. (2004) Characterisation of a mobile protein-binding epitope in the translocation domain of colicin E9, *J. Biomol. NMR* 30, 81−96.
13. James, R., Penfold, C. N., Moore, G. R., and Kleanthous, C. (2002) Killing of *E. coli* cells by E group nuclease colicins, *Biochimie 84*, 381−389.
14. Zakharov, S. D., and Cramer, W. A. (2002) Colicin crystal structures: Pathways and mechanisms for colicin insertion into membranes, *Biochim. Biophys. Acta 1565*, 333−346.
15. Vetter, I. R., Parker, M. W., Tucker, A. D., Lakey, J. H., Pattus, F., and Tsernoglou, D. (1998) Crystal structure of a colicin N fragment suggests a model for toxicity, *Structure 6*, 863−874.
16. Cao, Z., and Klebba, P. E. (2002) Mechanisms of colicin binding and transport through outer membrane porins, *Biochimie 84*, 399−412.
17. Pilsl, H., and Braun, V. (1995) Novel colicin 10: Assignment of four domains to TonB- and TolC-dependent uptake via the Tsx receptor and to pore formation, *Mol. Microbiol. 16*, 57−67.
18. Garinot-Schneider, C., Penfold, C. N., Moore, G. R., Kleanthous, C., and James, R. (1997) Identification of residues in the putative TolA box which are essential for the toxicity of the endonuclease toxin colicin E9, *Microbiology 143*, 2931−2938.
19. Bouveret, E., Rigal, A., Lazdunski, C., and Bénédetti, H. (1998) Distinct regions of the colicin A translocation domain are involved in the interaction with TolA and TolB proteins upon import into *Escherichia coli*, *Mol. Microbiol. 27*, 143−157.
20. Carr, S., Penfold, C. N., Bamford, V., James, R., and Hemmings, A. M. (2000) The structure of TolB, an essential component of the tol-dependent translocation system, and its protein−protein interaction with the translocation domain of colicin E9, *Structure 8*, 57−66.
21. Hands, S., Holland, L. E, Vankemmelbeke, Fraser, L., Macdonald, C. J., M., Moore, G. R., James, R., and Penfold, C. N. (2005) Interactions of tolB with the translocation domain of colicin E9 require an extended TolB box.
22. Abergel, C., Bouveret, E., Claverie, J.-M., Brown, K., Rigal, A., Lazdunski, C., and Bénédetti, H. (1999) Structure of the *Escherichia coli* TolB protein determined by MAD methods at 1.95 Å resolution, *Structure 7*, 1291−1300.
23. Soelaiman, S., Jakes, K., Wu, N., Li, C., and Shoham, M. (2001) Crystal structure of colicin E3: Implications for cell entry and ribosome inactivation, *Mol. Cell 8*, 1053−1062.
24. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR Spectroscopy*, Academic Press, San Diego, CA.
25. Panchal, S. C., Bhavesh, N. S., and Hosur, R. V. (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and ^{15}N sequential correlations in (^{13}C, ^{15}N) labeled proteins: Application to unfolded proteins, *J. Biomol. NMR 20*, 135−147.
26. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277−293.
27. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) ^{1}H, ^{13}C, and ^{15}N chemical shift referencing in biomolecular NMR, *J. Magn. Reson. B 101*, 63−71.
28. Bartels, C., Xia, T. H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules, *J Biomol. NMR 6*, 1−10.
29. Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., and Torchia, D. A. (1992) Pulse sequences for removal of the effects of cross-correlation between dipolar and chemical-shift anisotropy relaxation mechanism on the measurement of heteronuclear T_1 and T_2 values in proteins, *J. Magn. Reson. 97*, 359−375.
30. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. E., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation, *Biochemistry 33*, 5984−6003.
31. Mandel A., M., Akke, M., and Palmer, A. G., III (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: Correlations with structure and function in an active enzyme, *J. Mol. Biol. 246*, 144−163.
32. Günther, U. L., Ludwig, C., and Rüterjans, H. (2000) NMRLAB−Advanced NMR processing in MATLAB, *J. Magn. Reson. 145*, 201−208.
33. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil chemical shifts, *J. Am. Chem. Soc. 123*, 2970−2978.
34. Dyson, H. J., and Wright, P. E. (1991) Defining solution conformations of small linear peptides, *Annu. Rev. Biophys. Biophys. Chem. 20*, 519−538.
35. Kay, L. E. Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease, *Biochemistry 28*, 8972−8979.
36. Schwarzinger, S., Wright, P. E., and Dyson, H. J. (2002) Molecular hinges in protein folding: The urea-denatured state of apomyoglobin, *Biochemistry 41*, 12681−12686.
37. Yao, J., Chung, J., Eliezer, D., Wright, P. E., and Dyson, H. J. (2001) NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin: A model system for the initial steps of folding, *Biochemistry 40*, 3561−3571.
38. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein, *Science 295*, 1719−1722.
39. Peng, J. W., and Wagner, G. (1992) Mapping of the spectral densities of N−H bond motions in eglin c using heteronuclear relaxation experiments, *Biochemistry 31*, 8571−8586.
40. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) Spectral density function mapping using ^{15}N relaxation data exclusively, *J. Biomol. NMR 6*, 153−162.
41. Palmer, A. G., III (2004) NMR characterization of the dynamics of biomacromolecules, *Chem. Rev. 104*, 3623−3640.
42. Dyson, H. J., and Wright, P. E. (2004) Unfolded proteins and protein folding studied by NMR, *Chem. Rev. 104*, 3607−3622.
43. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea, *Biochemistry 36*, 8977−8991.
44. Kyte, J., and Doolittle, B. F. (1982) Hydrophobic cluster analysis: An efficient new way to compare and analyse amino acid sequences, *J. Mol. Biol. 157*, 105−132.
45. Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J.-P. (1987) A simple method for displaying the hydropathic character of a protein, *FEBS Lett. 224*, 149−155.
46. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehfus, M. H. (1985) Hydrophobicity of amino acid residues in globular proteins, *Science 229*, 834−838.

47. Creighton, T. E (1993) *Proteins: Structures and Molecular Properties*, 2nd ed., p 256, W. H. Freeman and Co., New York.
48. Shepherd, A. J., Gorse, D., and Thornton, J. M. (1999) Prediction of the location and type of β-turns in proteins using neural networks, *Protein Sci. 8*, 1045−1055.
49. Anthony, C., and Williams, P. (2003) The structure and mechanism of methanol dehydrogenase, *Biochim. Biophys. Acta 1647*, 18−25.
50. Ghosh, M., Anthony, C., Harlos, K., Goodwin, M. G., and Blake, C. C. F. (1995) The refined structure of the quinoprotein methanol dehydrogenase from *Methylobacterium extorquens* at 1.94 Å, *Structure 3*, 177−187.
51. Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins, *J. Mol. Biol. 338*, 1015−1026.
52. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., and Dunker, A. K. (1997) Identifying disordered regions in proteins from amino acid sequences, *Proc. IEEE International Conference on Neural Networks*, 90−95.
53. Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) Predicting protein disorder for N-, C-, and internal regions, *Genome Informatics 10*, 30−40.
54. Romero, P., Obradovic, Z., Li, X., Garner, E., Brown, C., and Dunker, A. K. (2001) Sequence complexity of disordered protein, *Proteins: Struct., Funct., Genet. 42*, 38−48.
55. Romero, P., Obradovic, Z., and Dunker, A. K. (1997) Sequence data analysis for long disordered regions prediction in the calcineurin family, *Genome Informatics 8*, 110−124.
56. Garner, E., Romero, P., Dunker, A. K, Brown, C., and Obradovic, Z. (1999) Predicting binding regions within disordered proteins, *Genome Informatics 10*, 41−45.
57. Hilsenbeck, J. L., Park, H., Chen, G., Youn, B., Postle, K., and Kang, C. (2003) Crystal structure of the cytotoxic bacterial protein colicin B at 2.5 Å resolution, *Mol. Microbiol. 51*, 711−720.
58. Wiener, M., Freymann, D., Ghosh, P., and Stroud, R. M. (1997) Crystal structure of colicin Ia, *Nature 387*, 461−464.
59. Anderluh, G., Hong, Q., Boetzel, R., MacDonald, C., Moore, G. R., Virden, R., and Lakey, J. H. (2003) Concerted folding and binding of a flexible colicin domain to its periplasmic receptor TolA, *J. Biol. Chem. 278*, 21860−21868.
60. Journet, L., Bouveret, E., Rigal, A., Lloubes, R., Lazdunski, C., and Bénédetti, H. (2001) Import of colicins across the outer membrane of *Escherichia coli* involves multiple protein interactions in the periplasm, *Mol. Microbiol. 42*, 331−344.
61. Housden, N. G., Moore, G. R., James, R., and Kleanthous, C. (2005) Cell entry mechanism of enzymatic bacterial colicins: Porin recruitment and the thermodynamics of receptor binding.

BI0503596