Protein Science (1995), *4*:484–495. Cambridge University Press. Printed in the USA.

Stabilization of a strained protein loop conformation through protein engineering

ALEC HODEL,[1,2] ROGER A. KAUTZ,[1,3] AND ROBERT O. FOX[1,2]

[1] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511
[2] The Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06511

(RECEIVED August 25, 1994; ACCEPTED December 7, 1994)

Abstract

Staphylococcal nuclease is found in two folded conformations that differ in the isomerization of the Lys 116–Pro 117 peptide bond, resulting in two different conformations of the residue 112–117 loop. The *cis* form is favored over the *trans* with an occupancy of 90%. Previous mutagenesis studies have shown that when Lys 116 is replaced by glycine, a *trans* conformation is stabilized relative to the *cis* conformation by the release of steric strain in the *trans* form. However, when Lys 116 is replaced with alanine, the resulting variant protein is identical to the wild-type protein in its structure and in the dominance of the *cis* configuration. The results of these studies suggested that any nuclease variant with a non-glycine residue at position 116 should also favor the *cis* form because of steric requirements of the β-carbon at this position. In this report, we present a structural analysis of four nuclease variants with substitutions at position 116. Two variants, K116E and K116M, follow the "β-carbon" hypothesis by favoring the *cis* form. Furthermore, the crystal structure of K116E is nearly identical to that of the wild-type protein. Two additional variants, K116D and K116N, provide exceptions to this simple "β-carbon" rule in that the *trans* conformation is stabilized relative to the *cis* configuration by these substitutions. Crystallographic data indicate that this stabilization is effected through the addition of tertiary interactions between the side chain of position 116 with the surrounding protein and water structure. The detailed *trans* conformation of the K116D variant appears to be similar to the *trans* conformation observed in the K116G variant, suggesting that these two mutations stabilize the same conformation but through different mechanisms.

Keywords: NMR; proline isomerism; protein engineering; staphylococcal nuclease

In work toward rational design and redesign of protein molecules, determining the relationship between the amino acid sequence and the three-dimensional structure of protein loops and β-turns is essential. These elements of protein structure comprise most of the surface of proteins (Richardson, 1981) and are often critical in protein function such as catalysis and molecular recognition (e.g., antigen/immunoglobulin recognition; Foote & Winter, 1992). Because of their aperiodic structure, loops are the most difficult element of protein secondary structure to model. Several taxonomy schemes have been developed for the classification of β-turns (Wilmot & Thornton, 1990) and more general loop conformations (Ring et al., 1992). Efforts to predict the conformation of protein loops have included comparison to similar known loop structures (Chothia et al., 1989) and conformational searches utilizing Monte Carlo, simulated annealing, and exhaustive grid search techniques (Fine et al., 1986; Bruccoleri et al., 1988; Collura et al., 1993). Further work toward the understanding of the dominant interactions that determine loop conformations is important to the future of protein engineering and rational drug design.

We are working to define protein engineering principles by examining the sequence–structure relationships for a loop in staphylococcal nuclease, which adopts multiple conformations. The paradigm for this study has been to generate variants of nuclease with single amino acid substitutions and then to determine the effects of this mutation on the conformational equilibrium of the loop segment using X-ray crystallography and NMR spectroscopy. The results of such a study may then be analyzed with two different goals. The first goal is to determine whether there are simple, residue-based rules that govern the sequence dependence of the loop conformation. Chothia and Lesk (1987) found a few canonical loop conformations in immunoglobulin hypervariable loops, demonstrating the utility of this approach. In general however, the complex interactions between the loop endpoint geometry, the loop sequence, and the protein context

Reprint requests: Robert O. Fox, Department of Molecular Biophysics and Biochemistry, Bass Center, Room 432, Yale University, 266 Whitney Avenue, P.O. Box 208114, New Haven, Connecticut 06520-8114; fox@yalevms.bitnet.

[3] Present address: Department of Chemistry, Rm 18-020, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

suggest that the sequence-structure relationship may only be accurately predicted through atomic scale energy calculations. Thus, the second goal of our mutational study is to provide detailed experimental measurements of the mutational effects that will guide the development of detailed energy calculations suitable for systems of this scale. To this end, we have undertaken a highly rigorous computational study of the loop in staphylococcal nuclease in parallel with the experimental studies (Hodel et al., 1995).

Staphylococcal nuclease, a small protein (149 residues) lacking disulfide bridges, has been the subject of many early protein folding studies (reviewed in Tucker et al., 1979) and of recent renewed interest (Shortle & Meeker, 1986; Kuwajima et al., 1991; Fink et al., 1993; Green & Shortle, 1993). The crystal structures of nuclease (Hynes & Fox, 1991) and the nuclease-Ca^{2+}-3′,5′-diphosphothymidine (pdTp) complex (Cotton et al., 1979; Loll & Lattman, 1989) have been refined to high resolution. NMR studies of nuclease have revealed a slow exchange between two folded conformations observed as multiple resonances for each of the His 8, His 121, and His 124 $H^{\epsilon 1}$ protons (Fox et al., 1986). This conformational heterogeneity was shown to be due to a mixture of *cis* and *trans* isomers at the Lys 116–Pro 117 peptide bond in an equilibrium that favors the *cis* configuration with a fractional occupancy of 90% (Evans et al., 1987, 1989; Wang et al., 1990). Unfolded nuclease and a peptide analog of this segment favor the *trans* isomer of the Lys 116–Pro 117 peptide bond (Evans et al., 1987; Raleigh et al., 1992). Thus, the *cis* configuration found in the native molecule is a strained element.

We have previously proposed that the protein fold forces the ends of the loop segment to positions that allow only strained backbone conformations, the most favorable of which is the native conformation with a *cis* 116–117 peptide bond. There are several lines of evidence supporting this hypothesis. Mutations that alter the anchoring side chain of Asn 118 exhibit a greater fraction of molecules that adopt the *trans* isomer of the 116–117 peptide bond (Hodel et al., 1994). Mutations in the loop segment that reduce the conformational restrictions of this segment, such as K116G (Hodel et al., 1993) and P117G (Evans et al., 1987; Hynes et al., 1994), increase the fraction of the *trans* population while also increasing the stability of the protein. Mutations within the loop that do not change the intrinsic backbone conformation space of the protein, such as K116A (Hodel et al., 1993), have little effect on the stability or the *cis/trans* equilibrium. If the protein is destabilized through mutation outside of the loop (Alexandrescu et al., 1990) or through solvent conditions, such as high temperature or low pH, (Alexandrescu et al., 1989), the fraction of the *trans* configuration increases. It is our hypothesis that destabilizing the protein reduces the rigidity of the loop end anchorage, allowing less strained *trans* conformations to appear.

Our study of the nuclease variants K116A and K116G gave some insight into the conformational equilibrium of the residue 112–117 loop of nuclease A. The thermodynamic and structural similarity between nuclease A and its K116A variant suggested that only the β-carbon of the residue 116 side chain was important in the determination of the conformation of the residue 112–117 loop. The additional ϕ,ψ conformations available to a glycine residue allowed the variant K116G to adopt a *trans* configuration and relaxed the strain in the system. If any non-glycine residue was placed at position 116 of the K116G loop conformation, the side-chain β-carbon would be in close contact with the δ-carbon of the Pro 117 imino ring and the backbone carbonyl oxygen of Tyr 115. If one were to model the *trans* conformation of nuclease A after the conformation found in K116G, the steric strain induced by the β-carbon of position 116 could be responsible for driving the conformational equilibrium to favor the *cis* isomer of the 116–117 peptide bond. Such a model would lead to the hypothesis that any nuclease variant where position 116 was replaced with a non-glycine residue would favor the *cis* conformation.

In this paper we present an analysis of the nuclease variants K116E, K116D, K116N, and K116M. K116E and K116M follow the simple "β-carbon" hypothesis outlined above by exhibiting conformational behavior similar to that of wild-type nuclease. K116D and K116N provide exceptions to the above hypothesis as each contains a β-carbon at position 116, yet the *trans* conformation is significantly stabilized in each of these variants. We present X-ray crystallographic data that suggest that the *trans* conformations in the K116D and K116N variants are stabilized by specific tertiary interactions with the surrounding protein and water structure.

Results

^1H NMR spectroscopy

The histidine $H^{\epsilon 1}$ proton region of the NMR spectra of the K116E, K116D, K116N, and K116M nuclease variants are shown in Figure 1. As with nuclease A, two resonances appear for the protons of His 8, His 121, and His 124. In nuclease A, each of these resonances monitors different isomerization states of the Lys 116–Pro 117 peptide bond (Evans et al., 1987, 1989). The assignments of the variant histidine $H^{\epsilon 1}$ resonances were inferred from the spectrum of the wild-type protein (Alexandrescu et al., 1988; Kautz et al., 1990), assuming that the order of the histidine chemical shifts is conserved. These tentative assignments were confirmed through titration with the nucleotide inhibitor, pdTp, and through magnetization transfer. The addition of Ca^{2+} and pdTp to a solution of wild-type nuclease (Evans et al., 1989) and the variants K116A and K116G (Hodel et al., 1993) stabilized the *cis* conformation in all three proteins. Assuming that this effect is conserved in the variants studied here, solutions of each protein were titrated with Ca^{2+} and the inhibitor. As shown in Figure 2, the proposed *cis* resonances of the K116N variant increased, whereas the *trans* resonances decreased, confirming their assignments. Similar results were obtained for the K116D, K116E, and K116M variants (data not shown). Magnetization transfer experiments (Evans et al., 1989) were performed on selected histidine protons to confirm conformational exchange (data not shown).

Curve fitting of the histidine region of the NMR spectra (see Fig. 1) yielded the equilibrium constant and the free energy change for the *cis/trans* equilibrium of each protein (see Table 1). K116D favors the *trans* conformation with a population of 70%, yielding a $\Delta\Delta G_{cis \rightarrow trans}$ of −1.75 kcal/mol relative to nuclease A. K116N also favors the *trans* conformation more than nuclease A; however, it populates both the *cis* and *trans* conformations nearly equally with a $\Delta\Delta G_{cis \rightarrow trans}$ of −1.13 kcal/mol compared to the wild-type protein. The *cis/trans* equilibrium in K116E appears to be very similar to that of nuclease A. K116E favors the *cis* conformation at 84% occupancy with

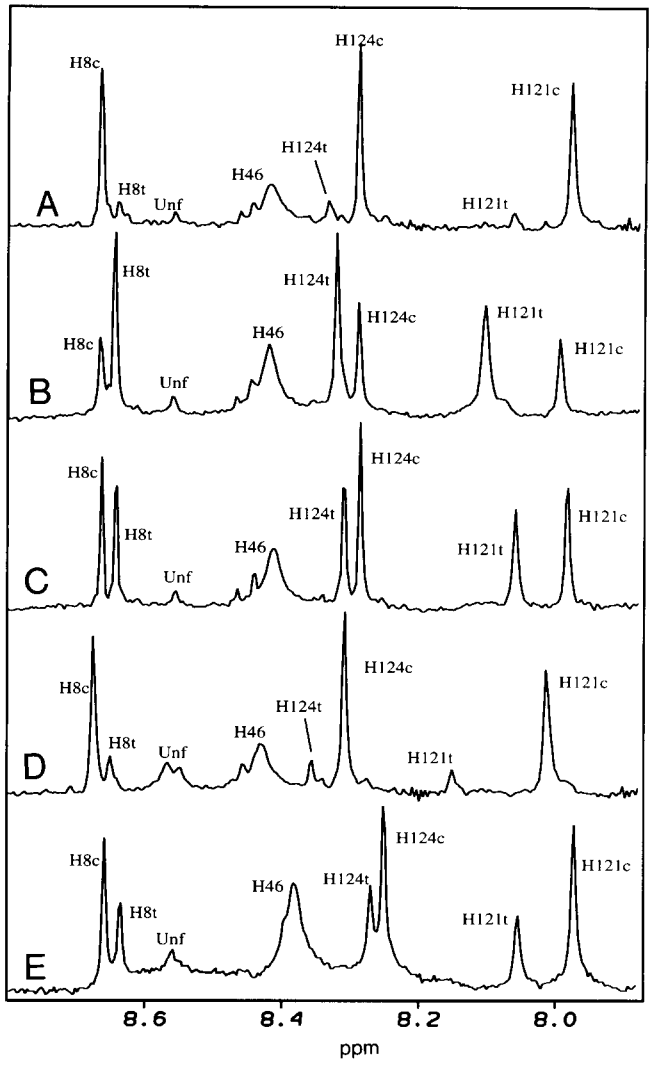

Fig. 1. Low-field region of 490 MHz ^{1}H NMR spectra showing resonances of the four histidine $H^{\epsilon 1}$ of **(A)** staphylococcal nuclease A, **(B)** the nuclease variant K116D, **(C)** K116N, **(D)** K116E, and **(E)** K116M. Backbone amide protons have been exchanged for deuterons, revealing the histidine $H^{\epsilon 1}$ resonances. These resonances are labeled by residue number suffixed with c, denoting the *cis* resonance, and t, denoting the *trans* resonance. Samples are in 200 mM acetic-d_3-acid-d buffered D_2O, pH* 5.3. All spectra were taken at 40 °C on the Yale-490 MHz in the Chemistry Instrumentation Center at Yale University.

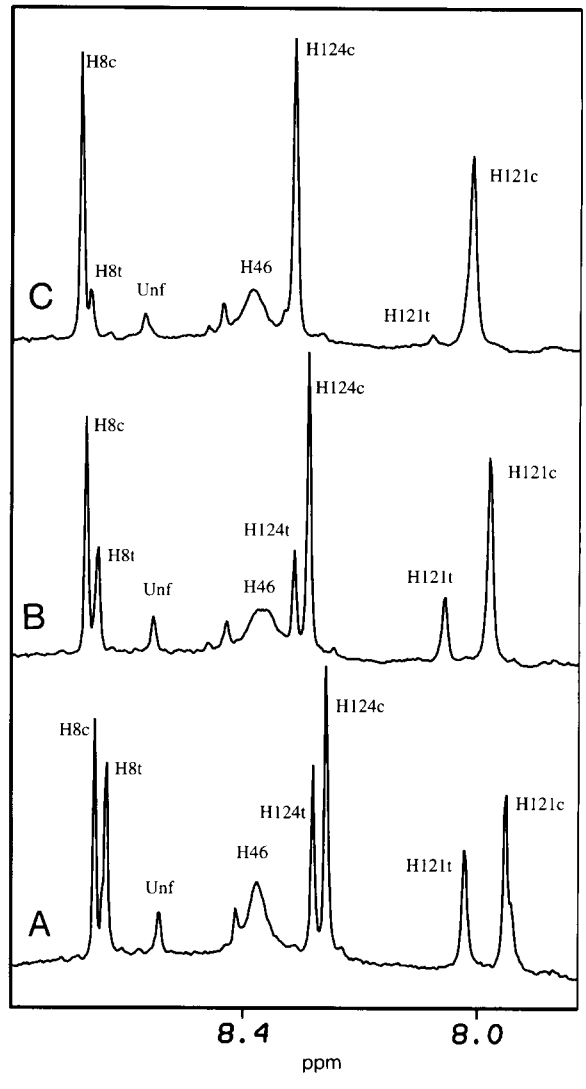

Fig. 2. Effect of ligand binding on the ^{1}H NMR spectrum of variant K116N. All spectra taken at pH* 5.3, 40 °C in 200 mM deuterated acetate buffered D_2O. **A:** No added ligands. **B:** 10 mM $CaCl_2$ and 1 mM 3′,5′-diphosphothymidine (pdTp) added. **C:** 10 mM $CaCl_2$ and 5 mM pdTp added. Intensity can be seen to shift from the *trans* form to the *cis* form as inhibitor is added.

a $\Delta\Delta G_{cis \rightarrow trans}$ from nuclease A of −0.20 kcal/mol. The *cis/trans* equilibrium in the K116M variant is modestly perturbed from that of nuclease A. K116M favors the *cis* conformation at 73% occupancy with a $\Delta\Delta G_{cis \rightarrow trans}$ from nuclease A of −0.81 kcal/mol.

Thermodynamic stability

Thermodynamic stabilities of K116D, K116N, K116M, and K116E all appear to be identical to that of nuclease A. As shown in Table 1, each of the three variants have the same unfolding midpoint concentration ($C_m \sim 0.7$ M) of guanidine hydrochloride (GuHCl) and the same free energy of unfolding ($\Delta G_D \sim$ 4.5 kcal/mol) as nuclease A within the precision of these measurements.

X-ray crystallography

Crystal structures of K116D, K116N, and K116E were similar to that of nuclease A except in the region of mutation (residues 112–117) and in a disordered loop (residues 44–51). Suitable crystals of K116M could not be prepared. Figure 3 shows the backbone traces of the three variants superimposed on the structure of nuclease A (Hynes & Fox, 1991). The RMS differences between the backbone coordinates of nuclease A and the three variants (not including residues 111–119 or 44–51) are 0.21 Å, 0.23 Å, and 0.18 Å for K116D, K116N, and K116E, respectively. These differences are on the order of the predicted uncertainty in the atomic coordinates as calculated from a Luzzati plot (Luz-

Table 1. *Thermodynamic measurements*

Protein	%*cis* (at 40 °C)	$\Delta G_{cis \rightarrow trans}$ (kcal/mol)	GuHCl C_m (M)	ΔG_D (kcal/mol)
Nuclease A	88 (1)	1.23 (0.07)	0.71 (0.02)	4.65 (0.1)
K116D	30 (1)	−0.52 (0.07)	0.71 (0.02)	4.54 (0.1)
K116N	54 (1)	0.10 (0.07)	0.68 (0.02)	4.41 (0.1)
K116E	84 (1)	1.03 (0.07)	0.72 (0.02)	4.65 (0.1)
K116M	73 (1)	0.62 (0.07)	0.70 (0.02)	4.43 (0.1)

zati, 1952; see Fig. 4). Thus, the mutations at position 116 have little effect on the global structure of the protein.

The conformation of residues 112–117 was clearly indicated in the SA-omit $F_o - F_c$ map of the K116E variant. As shown in Figure 5A, the density unambiguously defines the backbone conformation of each residue, including a *cis* peptide joining residues 116 and 117. The conformation defined by the electron density is very similar to the loop conformation found in nuclease A. Along with the *cis* peptide bond, the K116E structure retains the type VIa β-turn from residues 115–118 observed in the nuclease A structure. Figure 5B shows the superimposition of the nearly identical conformations of K116E and nuclease A.

As described in the Materials and methods, the electron density in the residue 112–117 regions of K116D and K116N was interpretable only after adding ordered waters and a bulk solvent correction to the models. Figure 6 shows the resulting electron density maps for the K116D variant contoured at 1.8σ and 1.5σ. The position of the side chain of Asp 116 was clearly defined by the electron density in a map contoured at 1.8σ. The position of this side chain requires the Asp 116–Pro 117 peptide bond to be in the *trans* isomer. The *trans* state of the 116–117 peptide was also suggested by a weak bump of electron density at the position of the backbone carbonyl oxygen of residue 116. The density for residues 115 and 114 was weak and patchy. However, the density for residues 112–113 clearly defined their backbone atomic positions. The side chains of residues 113–115 do not appear in the electron density. The electron density map yields some suggestion of the backbone conformation of residues 114 and 115 when it is contoured at 1.5σ (see Fig. 6B); however, the details of the backbone and side chain conformations of residues 114 and 115 are not unambiguously determined.

A model for the loop conformation was built into the density and refined to the data as shown in Figure 6. Along with the isomerization of the 116–117 peptide bond, the most striking difference between the K116D and nuclease A loop conformations is the rearranging of a hydrogen bond network including the loop residues and the surrounding ordered water molecules (see Fig. 7A,B). In nuclease A, Asn 118 and Tyr 115 form a hydrogen bond framework of local structure with two ordered water molecules. In the nuclease A structure, a hydrogen bond between the carbonyl oxygen of Tyr 115 and the amide nitrogen of Asn 118 completes the type VIa β-hairpin turn of residues 115–118. In K116D, the Asp 116 side chain takes the place of the Tyr 115 backbone atoms in completing hydrogen bonds with the amide of Asn 118 and the two water molecules. Thus, unlike nuclease A, K116D does not form a well-defined β-turn with residues 115–118, but instead completes a reverse turn by a hydrogen bond between the side chain of Asp 116 and the amide of Asn 118.

The conformation of residues 112–117 in K116D is very similar to that found in the variant K116G (Hodel et al., 1993), which also adopts a *trans* 116–117 peptide bond (see Fig. 7C). The main difference between the two conformations is the orientation of the 115–116 peptide bond, which is defined by the backbone dihedrals $\psi(115)$ and $\phi(116)$. Due to the steric clash between the carbonyl oxygen of Tyr 115 with the β-carbon of residue 116, the conformation of the 115–116 peptide group in K116G would be forbidden if any non-glycine amino acid were found at position 116. Similar to K116G, K116D adopts a $\psi(116)$ angle such that the β-carbon of the Asp 116 side chain is in close contact (3.2 Å) with the δ-carbon of the Pro 117 imino ring.

When the $F_o - F_c$ map is contoured at 1.5σ, the electron density near residue 116 shows two protuberances that are not explained by our initial model. It is possible that this density originates in well-ordered atoms of the *cis* conformation of K116D which, in solution, has an occupancy of 30%. We hypothesized that this *cis* conformation could be similar to the conformation found in wild-type nuclease A. A second model of the residue 112–118 loop was constructed for K116D with a *cis*

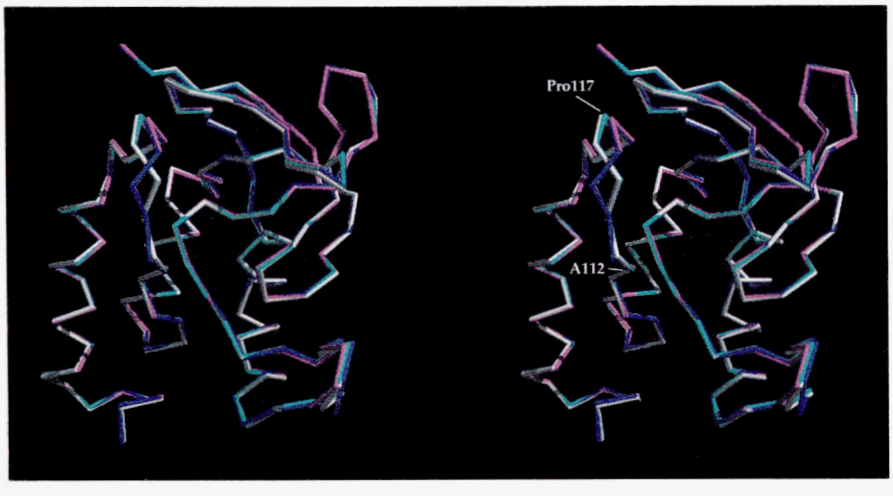

Fig. 3. Stereo representation of the backbone of nuclease A (white), K116E (green), K116D (blue), and K116N (magenta). The four structures were superimposed by minimizing the difference between the coordinates of the backbone atoms of all residues except the disordered loop of residues 44–51 and the loop containing the mutation (residues 111–119).

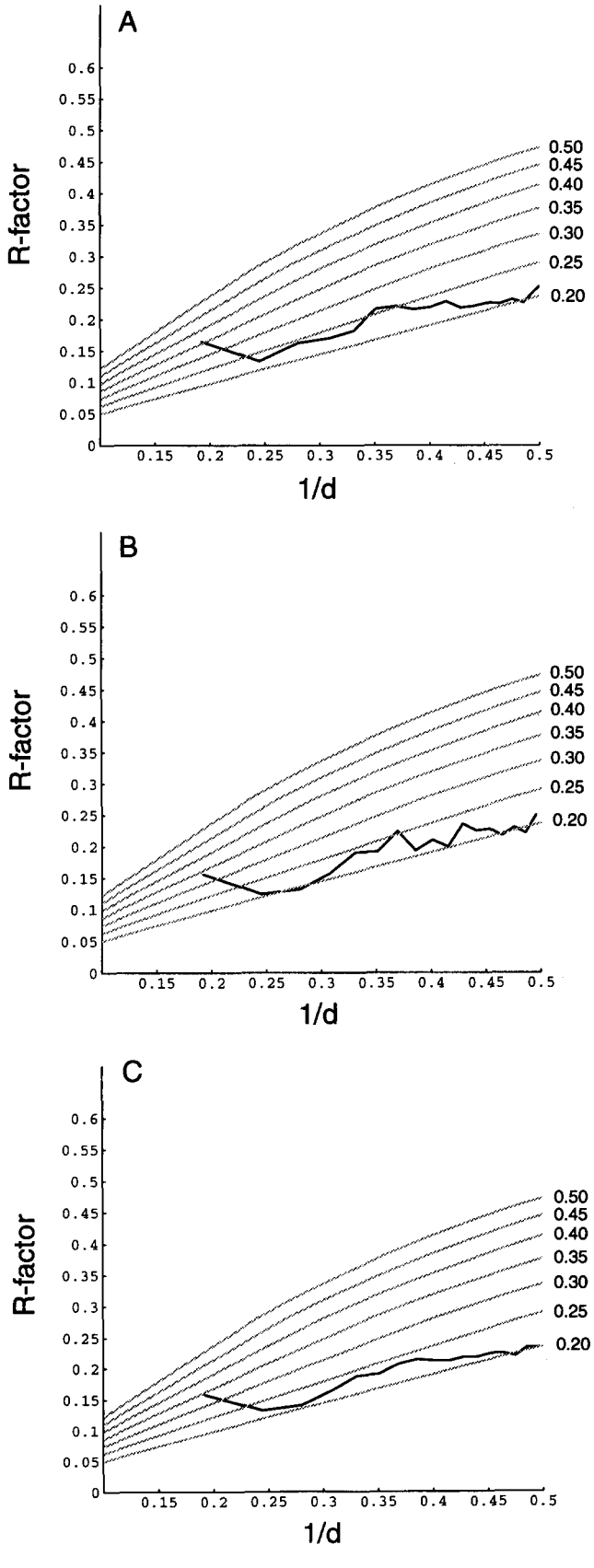

Fig. 4. Luzzati plots (Luzzati, 1952) of the mean atomic error for the (**A**) K116D, (**B**) K116N, and (**C**) K116E structures, where d is the Bragg spacing.

116–117 peptide bond and a conformation similar to that of the wild-type protein. When this second, *cis* model of the K116D loop is superimposed on the SA-omit electron density map, the unexplained density outcroppings are positioned consistently with the position of the 116 side chain and 115–116 backbone trace (see Fig. 6C).

The electron density from the $F_o - F_c$ SA-omit map in the loop region of K116N was noisy and ill-defined (see Fig. 8); yet, certain features of the conformation of residues 112–117 are strongly suggested. Similar to K116D, K116N appears to adopt a hydrogen bond between the Asn 116 side chain and the backbone amide of Asn 118. The density clearly defines the Asn 116 side chain in a position very similar to that observed in the Asp 116 side chain of K116D.

The detailed conformation of residues 113–115 is not clear in the electron density map. As shown in Figure 8A, the density is somewhat consistent with the conformation modeled into K116D, but a patch of density near the 115–116 peptide bond suggests an alternate position of this peptide group. In Figure 8B, a second possible conformation of residues 113–116 is shown where the carbonyl oxygen of Tyr 115 is directed into the previously mentioned patch of density. This change in the conformation brings the K116N loop conformation closer to that observed in K116G. The steric clash between the Tyr 115 carbonyl oxygen and the Asn 116 side chain suggests that this conformation is somewhat improbable. Both conformations shown in Figure 8 refine to the same *R*-factor, thus the data do not finely discriminate between possible conformations of the residue 113–115 segment.

Discussion

The K116E and K116M variants of nuclease are similar to nuclease A in their stability and *cis/trans* equilibrium. K116E also adopts a nearly identical structure to that of the wild-type protein. These characteristics are also shared by the variant K116A (Hodel et al., 1993), providing examples of four amino acids at position 116, namely alanine, glutamate, methionine, and lysine (the wild type), which result in proteins that are quite similar thermodynamically. The only similarity between these four amino acids is that they each contain a β-carbon but are not β-branched. The similarity between these four proteins lends support to the hypothesis that the β-carbon at residue 116 is the most important structural element at that position in the determination of the residue 112–117 loop structure. The K116M variant results in the largest perturbation of the *cis/trans* equilibrium from that observed for the wild-type protein of the three mutants considered above. Unfortunately, suitable crystals could not be prepared to investigate the structural basis of this change in the *cis/trans* equilibrium. We hypothesize that the change may be due to a more favorable hydrophobic interaction between the Met side chain and the remainder of the protein in the *trans* form of the protein.

The variants K116D and K116N provide exceptions to the β-carbon hypothesis. Both of these variants contain a β-carbon at position 116, yet $\Delta G_{cis \rightarrow trans}$ for these two proteins is significantly lower than that of the wild-type nuclease. The similar effect in K116D and K116N, and the absence of this effect in K116E, suggest that the specific geometry at position 116, rather than the charge, is important in the relative stabilization of the *trans* conformation. The crystal structures of these two variants

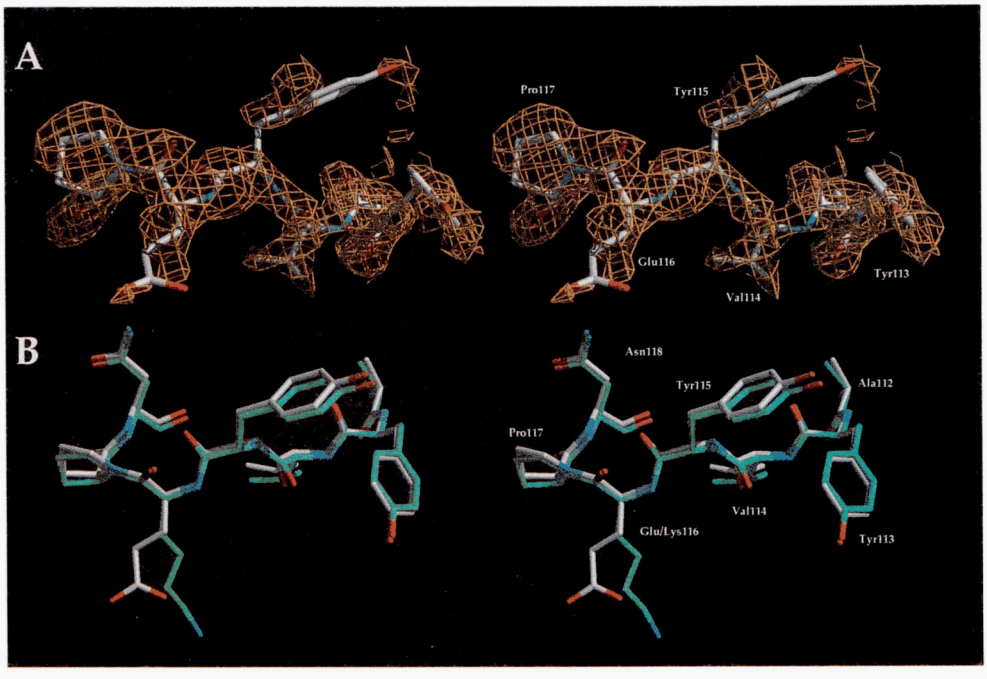

Fig. 5. Structure of the K116E variant. **A:** Stereo representation of the electron density in the region of residues 113–117 in the K116E variant. Density was calculated as an $F_o - F_c$ omit map using the model from nuclease A (Hynes & Fox, 1991) after refinement through the SA-omit procedure with residues 112–118 removed from the model. **B:** Structure of residues 112–118 of K116E (green) and nuclease A (white). Structures were superimposed as in Figure 3. Oxygen atoms are colored red; nitrogen is colored blue.

did not yield the detailed conformation of the residue 112–117 loop, but they did suggest an explanation for the deviation in the *cis/trans* equilibrium compared to that of nuclease A. Although both K116D and K116N are found in strained conformations (with the β-carbon of residue 116 in close contact with the δ-carbon of Pro 117), both variants exhibit a stabilizing short-range tertiary interaction involving hydrogen bonds between the Asp/Asn 116 side chain, Asn 118, and a network of water molecules. The geometry of this interaction must be compatible with a *trans* conformation, but not the *cis* conformation. Thus, the *trans* conformation is stabilized relative to the *cis* conformation. The tight geometry involved in this interaction is specific to the aspartate/asparagine size and shape, prohibiting the variant K116E from adopting a similar geometry.

We hypothesize that the *cis* conformation detected by NMR at 30% occupancy in K116D and 50% occupancy in K116N is similar to the conformation observed in wild-type nuclease. From here forward we shall refer to the backbone conformation observed or hypothesized in each variant by subscripting the name of the variant with either *cis* or *trans*. For example, our hypothesis above may be restated as $K116D_{cis}$ and $K116N_{cis}$ are similar to nuclease A_{cis}. This possibility was suggested by weak extra electron density in the K116D $F_o - F_c$ map (see Fig. 6C). More substantial evidence for this claim can be inferred from the similarity between nuclease A_{cis}, $K116A_{cis}$, and $K116E_{cis}$. It would be reasonable to assume that K116D would adopt this conformation with the single assumption that Asp 116 does not make any heretofore unobserved tertiary interactions in $K116D_{cis}$. Simple graphic modeling supports this assumption with the observation that the side chain of residue 116 should be completely solvated when the 116–117 peptide bond is *cis*.

If $K116D_{cis}$ and $K116N_{cis}$ are similar to the nuclease A_{cis}, then the *cis/trans* equilibria of these two variants is derived from the energetic differences between the conformations and hydrogen bonding networks shown in Figure 7. The small difference in $\Delta G_{cis \rightarrow trans}$ between K116D and K116N would then be attributed to a combination of the differences between solvation energy, hydrogen bonding energy, and reaction to the local electrostatic potential between aspartate and asparagine in the conformations shown in Figure 7. These variants would therefore provide a good test case for molecular modeling studies.

The similarity between the *trans* loop conformations found in K116D and K116N compared to that observed in K116G leads to the question of whether this is the *trans* conformation detected by NMR in wild-type nuclease A, K116A, K116M, and K116E. This is a plausible hypothesis, as one would expect K116E and K116D to have a similar backbone conformation space available. Thus, ignoring the interactions between the residue 116 side chain and the surrounding protein context, $\Delta G_{cis \rightarrow trans}$ should be similar in these two proteins, i.e., ~1 kcal/mol. For the uncertainty in this value, the effects of the specific side chain at position 116 can be observed in the difference between $\Delta G_{cis \rightarrow trans}$ of nuclease A (Lys at position 116), K116A, K116E, and K116M, which are 1.2 kcal/mol, 1.5 kcal/mol, 1.0 kcal/mol, and 0.6 kcal/mol, respectively. Thus, if $K116D_{trans}$ is similar to $K116E_{trans}$, the side-chain interactions of Asp 116 in K116D stabilize $K116D_{trans}$ with respect to $K116D_{cis}$ by approximately 1.5 kcal/mol.

The long-range goal of these efforts was to lay the foundation for the prediction of protein loop conformations from their primary sequence through two broad approaches, namely a knowledge-based approach and an energy-based approach. The

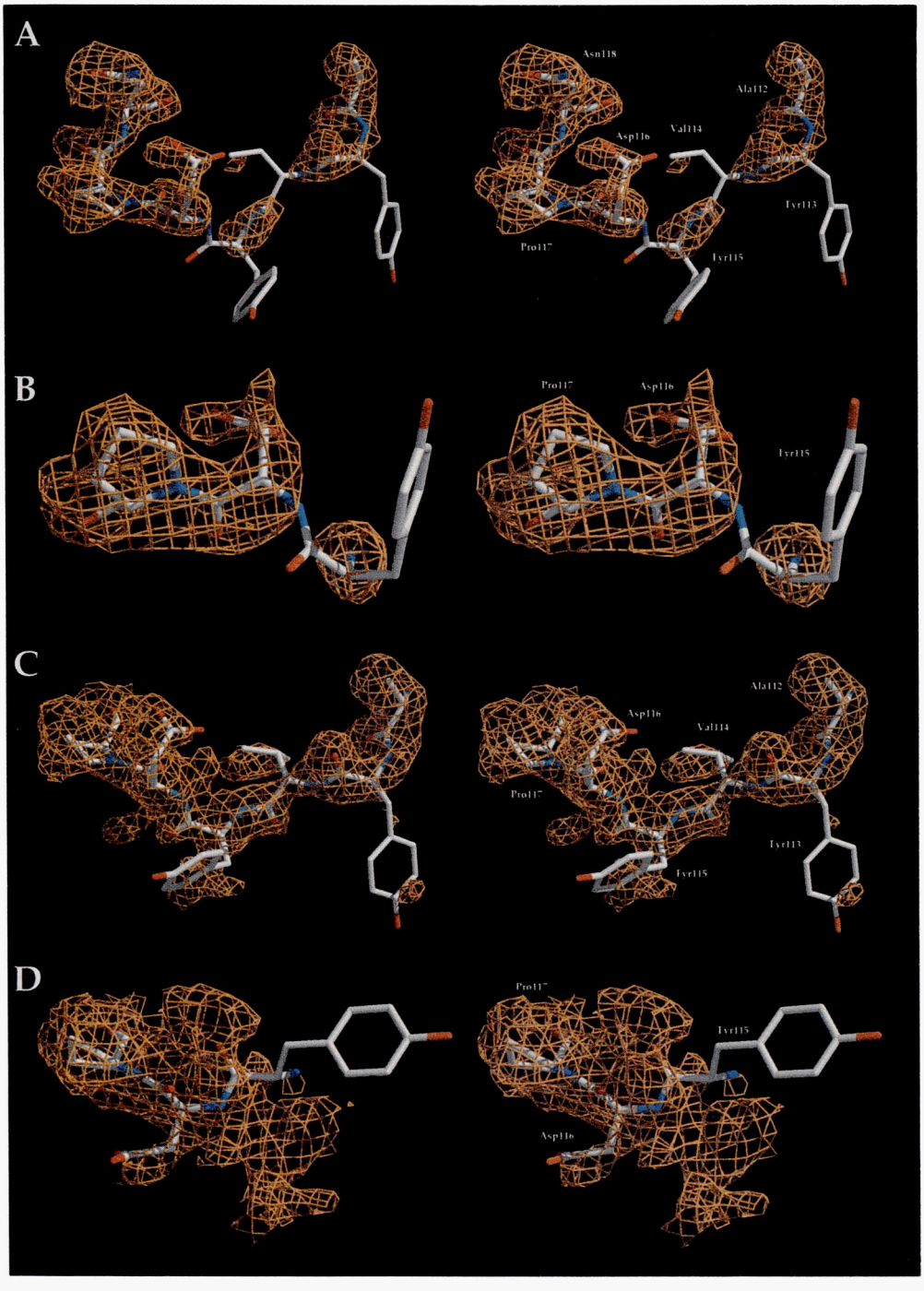

Fig. 6. Electron density surrounding residues 112–118 of the K116D variant. All density is calculated from an $F_o - F_c$ SA-omit map using the coordinates of nuclease A excluding residues 112–118. **A:** Density contoured at 1.8σ illustrates the essential features of residues 112–113 and 116–117. **B:** Map in (A) rotated to reveal the 116–117 peptide bond geometry. **C:** Density for the entire loop region becomes more clear in a map contoured at 1.5σ. Two protuberances near residue 116 also appear that are inconsistent with the *trans* K116D model. **D:** A model was constructed for the K116D residue 112–118 loop with a *cis* conformation identical to that of the wild-type nuclease A structure. The unexplained protuberances of the 1.5σ map are consistent with the placement of the 116 side chain and 115–116 backbone in this *cis* model of K116D.

first approach follows from the work of Chothia and Lesk (1987), who found a small number of canonical loop conformations that depend on only a few key residues. Toward this approach, the effects of site-directed mutagenesis on the conformation of the residue 112–117 loop was determined for several variants. From the work presented so far, it appears probable that the concept of canonical loop forms is applicable to this loop of staphylococcal nuclease. This conclusion is

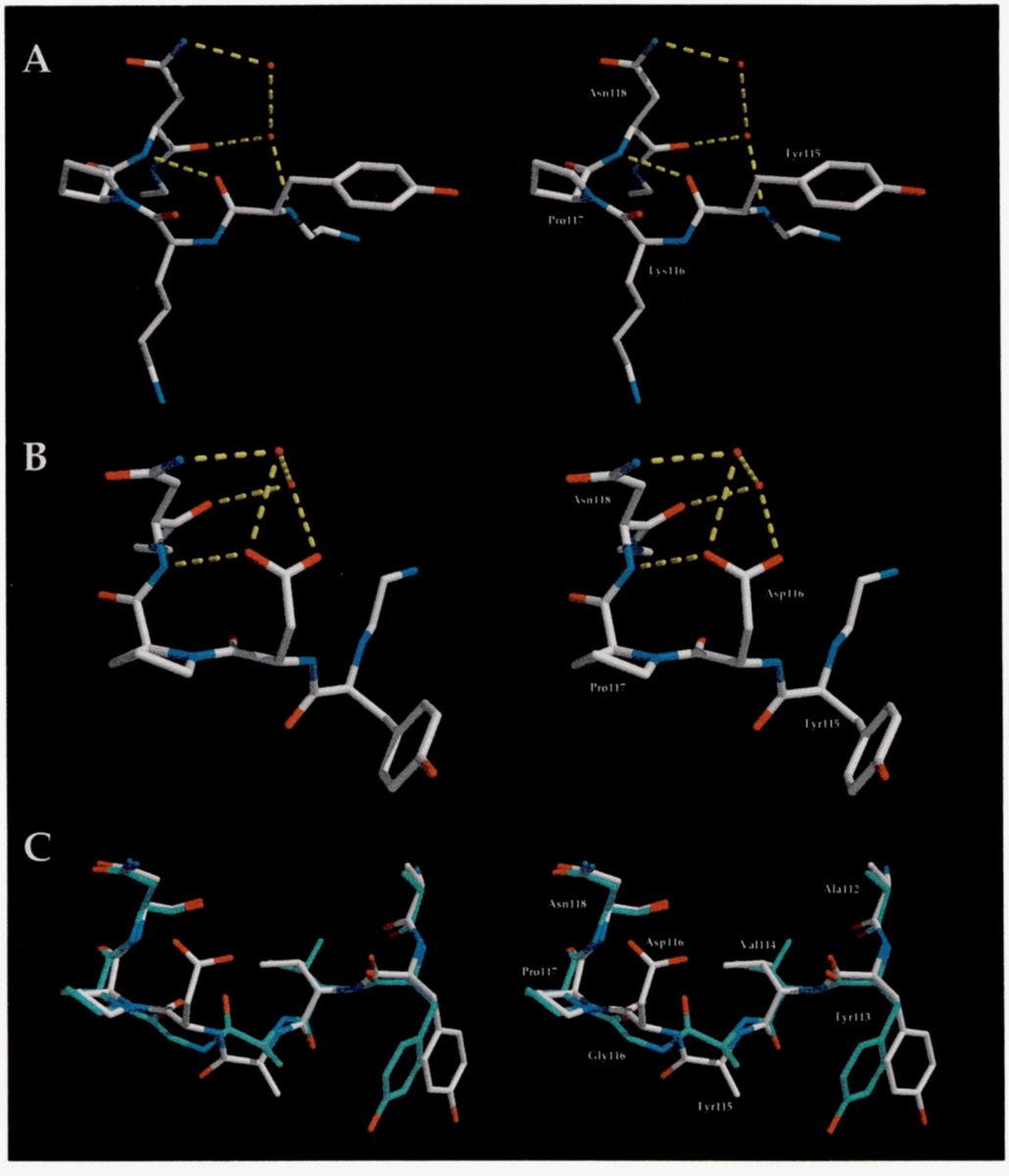

Fig. 7. A: Hydrogen bonding pattern between residues 115 and 118 of nuclease A (Hynes & Fox, 1991) and the surrounding ordered water. Residues 115–118 adopt a type IVa β-turn with a hydrogen bond between the carbonyl oxygen of Tyr 115 and the backbone amide of Asn 118. **B:** Hydrogen bonding pattern between residues 115 and 118 of K116D and the surrounding ordered water. The Asp 116 side chain replaces the backbone of Tyr 115 in many of the hydrogen bonds observed in the wild-type structure. **C:** Comparison of the residue 112–118 loop conformations from the K116D variant (white) and the K116G nuclease variant (Hodel et al., 1992) (green). Oxygen atoms are colored red; nitrogen is colored blue.

based on the fact that there are fewer unique loop conformations than there are variants, i.e., several variants of differing loop sequence share the same conformation. The sequence dependence of the residue 112–117 loop conformation demonstrates degeneracy in the amino acid to loop conformation code.

The analysis of the nuclease variants with substitutions at position 117 clearly define this residue as a key position in the determination of the loop conformation (Hynes et al., 1994). In their results, Hynes et al. describe three substitutions resulting in two new loop conformations. At position 117, a proline yields the wild-type *cis* conformation in a type VIa β-turn. Glycine and alanine yield a type I′ turn with a *trans* peptide bond and threonine yields a type I turn again with a *trans* peptide bond.

Six variants at position 116 have been generated yielding three observable loop conformations. The wild-type nuclease conformation with a *cis* 116–117 peptide bond is observed when position 116 is lysine, glutamate, or alanine. When aspartate or asparagine is placed at position 116, a conformation with a *trans* 116–117 peptide bond is observed. We believe that the formation of the hydrogen bond between the side chain of residue 116 and the backbone of residue 118 populates a conformation similar to the minor *trans* form of the wild-type protein to a suffi-

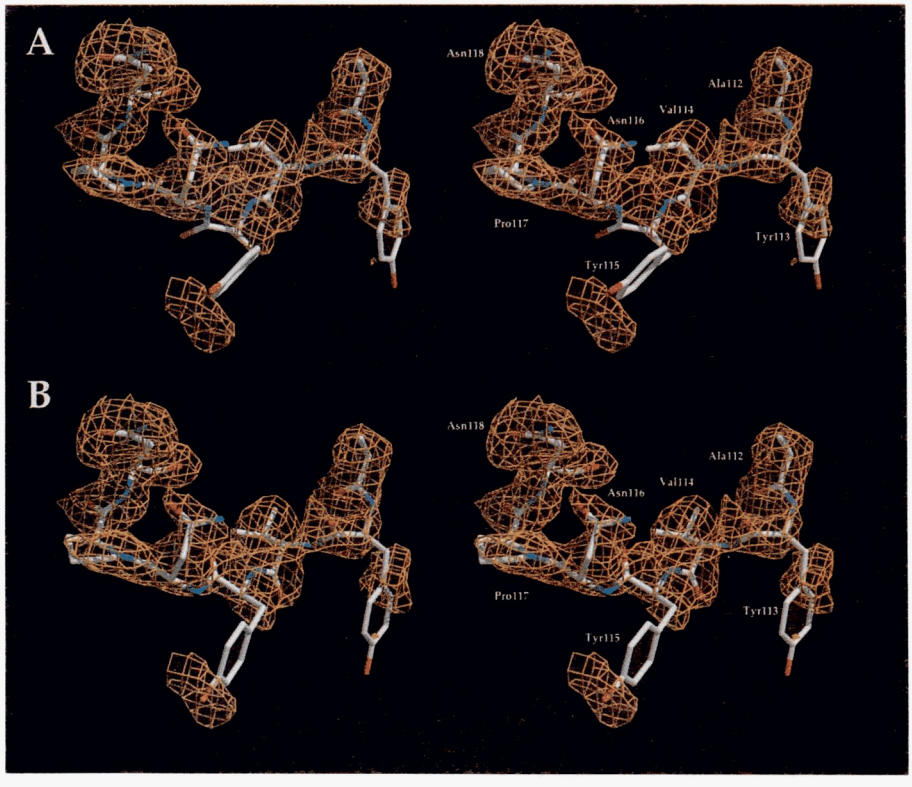

Fig. 8. Electron density surrounding residues 112–118 of the K116N variant. Shown is the $F_o - F_c$ SA-omit map calculated using the nuclease A structure excluding the atoms of residues 112–118. Map is contoured at 1.5σ. Density yields the conformation of residues 112–114 and 116–117, but conformation of residue 115 is not clear. Two models are shown with the map. **A:** A model similar to that of K116D was built and refined to the data. **B:** A model based on the conformation found in the K116G variant is also consistent with the electron density. Although the density in the SA-omit map favors the conformation of Tyr 115 in the K116G model (B), steric considerations suggest that the K116D conformation (A) is more likely to be adopted by K116N.

cient level that it can be observed crystallographically. The structure of this protein loop in nuclease K116G, although similar in its overall positioning on the protein surface, adopts a significantly different conformation at residue 116 (α_L), which is uniquely allowed for glycine residues (Hodel et al., 1993).

Although the K116D, K116N, and K116G variants all increase the population of the *trans* 116–117 peptide bond, they do so by different mechanisms and thus have distinct effects on protein stability. The K116G mutant examined in an earlier publication (Hodel et al., 1993) increases the *trans* population because the Gly residue can adopt a conformation (α_L) that is energetically unfavorable for side chains with a β-carbon. We hypothesized that the new loop conformation relieves the strain imposed by the remainder of the molecule without requiring the formation of a *cis* peptide bond, allowing more favorable intramolecular interactions, and thus increasing the stability of the K116G protein to denaturation. In the case of the two mutants examined here (K116D and K116N), the strain is not relieved by the mutation but is instead redistributed between the energetically less favorable peptide bond in the *cis* conformation and a displacement of the remainder of the protein structure from its lowest energy conformation when the peptide bond is in the *trans* conformation. The *trans* conformation has a faster thermal unfolding rate (Evans et al., 1989), and thus increasing the *trans* population (when the remainder of the protein is under strain) should decrease the stability of the protein. There are of course other contributing effects to the stability of each protein including electrostatic interactions.

In a theoretical study of this loop through rigorous free energy simulations, evidence was found suggesting that the conformation observed in K116D is very similar to that of the minor *trans* conformation of wild-type nuclease (Hodel et al., 1995). If this hypothesis is correct, the variants K116N, and K116D do not result in novel loop conformations but simply change the equilibrium between the two conformations populated by the wild-type protein. Again, this is in contrast to the substitutions at position 117 that result in novel conformations not populated by the wild-type protein. This analysis would not grant position 116 the status of a key position, i.e., one that determines the canonical loop form for the residue 112–117 loop. Rather, changes in position 116 result in only small energetic perturbations in the conformations populated by the wild-type protein.

We are engaged in a comprehensive exploration of the conformation space of this peptide segment through computational modeling. Our eventual goal in this exploration is to predict the

conformations of nuclease variants given the amino acid sequence of the residue 112–117 loop segment. A complete experimental description of these conformations along with their relative free energies will provide a much needed database for efforts dealing with the rational redesign of protein molecules through amino acid substitution.

Materials and methods

Preparation of the nuclease variants

The nuclease A gene was subcloned into M13mp18 to produce single- stranded template DNA. The K116D and K116E mutants were prepared by primer-directed mutagenesis using synthetic oligonucleotides (Zoller & Smith, 1983; Kautz et al., 1990). Plaques were screened by differential hybridization using 5′-^{32}P-labeled oligonucleotides and sequenced to verify the mutation. The K116N and K116M mutants were prepared by PCR mutagenesis using synthetic oligonucleotides (Higuchi, 1990) and were sequenced to verify the mutation. The mutant genes were subcloned into the plasmid pAS1, expressed in *Escherichia coli*, and protein prepared as described (Evans et al., 1989).

^{1}H-NMR determination of cis/trans equilibrium

Between 10 and 30 mg of lyophilized protein was resuspended in D_2O and adjusted to pH* 5.3. (pH* refers to glass electrode meter reading uncorrected for deuterium isotope effects [Bundi & Wüthrich, 1979].) The sample was then heated to 10 °C above the T_m for 5 min to facilitate exchange of labile protons, and any precipitate was removed by centrifugation. The deuterated protein was lyophilized again, and resuspended in 0.5 mL of 200 mM acetic-d_3-acid-d, pH* 5.3, with 1 mM trimethylsilyl propionate (Aldrich) as a chemical shift reference. Trace precipitate, if present, was removed by centrifugation before transferring to a 5-mm NMR tube. Spectra were taken at 40 °C on the Yale-490 MHz in the Chemistry Instrumentation Center at Yale University or a Bruker AM-500.

The assignments of the $H^{\epsilon 1}$ histidine peaks were performed by comparison to the wild-type protein spectrum. The resonances that corresponded to the *cis* and the *trans* conformations were differentiated by the addition of Ca^{2+} and the competitive inhibitor pdTp. Conformational exchange between the *cis* and *trans* peaks of individual histidine protons was confirmed through magnetization transfer experiments (Evans et al., 1989). Spectra were transformed and analyzed using the program Felix (Hare Research).

The fraction of molecules with a *cis* Lys 116–Pro 117 peptide bond was obtained from the relative areas of the resolved histidine resonances of spectra acquired at 40 °C. The nonlinear least-squares curve fitting of the Lorentzian lines was performed using the Leverberg–Marquardt algorithm (Press et al., 1986). Errors in the fitting parameters were estimated by applying normally distributed shifts to the experimental data and repeating the curve fitting to yield a distribution of fitting parameters. The ΔG values for the *cis/trans* equilibrium were obtained using $\Delta G = -RT \ln K$, where K is the equilibrium constant for the isomerization from *cis* to *trans*.

Thermodynamic stability determinations

The thermodynamic stability of nuclease and its variants was measured by GuHCl denaturation, monitoring the fluorescence of the single tryptophan at residue 140. Protein concentrations were 6 μM (0.1 mg/mL) in a buffer of 200 mM sodium acetate, pH 5.5. The excitation wavelength was 300 nm, and emission was monitored at 340 nm using an SLM model 8000 fluorimeter. Extrapolation to zero denaturant concentration provides an estimate of ΔG_D in the absence of denaturant (Pace, 1975).

Crystallization and X-ray intensity data collection

Nuclease K116D, K116N, and K116E were crystallized from low salt buffer (10.5 mM potassium phosphate, pH 8.15, 2 mg/mL protein) using 2-methyl-2,4-pentanediol as a precipitant (Arnone et al., 1969), yielding bipyramidal crystals in space group $P4_1$. A single crystal was used in each of the three data sets described in Table 2. X-ray diffraction intensities were measured using an area detector system and CuK_α radiation from a rotating anode generator. Redundant symmetry-related reflections were averaged and yielded reasonable R-merge statistics. The effective resolution of a data set is best defined by D2, the highest resolution shell with a mean intensity greater than or equal to $2\sigma(I)$. D2 varied from 1.9 to 2.0 Å resolution for the three data sets. Details of the data collection and reduction are given in Table 2.

Crystallographic refinement

Starting phases and coordinates for all three structures were derived from the structure of nuclease A (Hynes & Fox, 1991) with the solvent molecules removed. To avoid model bias, residues 112–118, a solvent exposed loop, were omitted from the nuclease A structure. This edited structure was then rigid-body refined to the low resolution diffraction data (15–3 Å), where the entire protein is treated as a rigid body. An overall temperature factor was then computed and applied to the model. Further rigid-body minimization and overall B-factor refinement were performed using data from 15 Å to 2 Å resolution. With residues 112–118 still omitted, the model was refined to the original data using alternate cycles of positional and individual restrained B-factor refinement. All refinement, including rigid-body minimization, positional, and B-factor, and simulated annealing refinements were carried out through the program X-PLOR (Brünger et al., 1987).

Nuclease crystals diffract anisotropically, with stronger amplitudes along the c^* axis relative to the a^* and b^* axes. To correct this anisotropy, the original data were scaled to the calculated data using a local scaling program (Hynes & Fox, 1991) adapting the methods of Matthews and Czerwinski (1975). In this procedure, a local scale factor is calculated as $\Sigma(F_{calc})/\Sigma(F_{obs})$ for a 5 × 5 × 5 (*hkl* units) box of reflections excluding the central reflection to prevent bias. The resulting scale factor is applied to the central F_o, which was omitted. Prior B-factor refinement of the model to the original data should preserve the overall temperature factor information in the anisotropic scaling of the data.

Each model was refined with the scaled data from 6 to D2 resolution through the simulated annealing omit technique (Hodel et al., 1992) where residues 112–118 were omitted from the structure. A $F_o - F_c$ electron density map was calculated in the residue 112–118 region from each annealed edited model. Electron density analysis and model building were performed through the graphics program FRODO (Jones, 1985).

The SA-omit $F_o - F_c$ map for the variant K116E demonstrated a good fit to the loop conformation of residues 112–118

Table 2. *Crystallographic and geometric parameters at the conclusion of refinement*

	K116E	K116D	K116N
a,b (Å)	47.4	48.2	47.2
c (Å)	63.4	63.1	63.0
Resolution range (Å)	6.0–1.9	6.0–1.9	6.0–2.0
Number of unique reflections	10,190	8,957	7,157
Percent unique reflections (to resolution)	99	88	92
D2[a] (Å)	1.9	1.9	2.0
$R_{symmetry}$[b] (%)	5.0	5.6	6.1
R-factor[c] (%)	18.8	18.9	18.0
Number of protein atoms	1,092	1,091	1,091
Number of water molecules	55	52	57
Number of total atoms	1,147	1,143	1,148
Mean B-factor	27.2	30.5	30.2
Main-chain atoms	23.2	26.4	26.1
Side-chain atoms	30.1	33.5	33.0
Backbone atoms residues 112–117	30.0	34.0	40.1
Side-chain atoms residues 112–117	38.7	44.0	55.2
Water	32.0	37.1	40.1
Deviations from ideal geometry			
Bond lengths (Å)	0.007	0.006	0.009
Bond angles (°)	1.62	1.54	1.72

[a] D2 represents the resolution shell where the $\langle I \rangle = 2\sigma$.
[b] Merging R-factor for symmetry-related reflections defined as $\sum|\langle I \rangle - I|/\sum\langle I \rangle$.
[c] Crystallographic R-factor defined as $\sum|F_{obs} - F_{calc}|/\sum F_{obs}$.

found in wild-type nuclease A. The Lys 116 side chain was replaced with glutamate and this new model was refined through alternating cycles of positional and B-factor refinement. Ordered water molecules were added to the model and the structure was refined to an R-factor of 18.7%.

The electron density maps calculated from the K116D and K116N data were weak and patchy in the residue 112–118 region, suggesting the need for further refinement of these models. With residues 112–118 still omitted, the annealed models were then subjected to multiple rounds of positional and B-factor refinement. Water molecules that were not in contact with the residue 112–118 region of the proteins were then added to the model structures. Further positional and B-factor refinement yielded models of K116D and K116N with R-factors of 20.2% and 18.5%, respectively. To obtain clear maps at the protein surface, we decided to calculate electron density maps using data from 40 Å to D2. A bulk solvent correction was added to the low resolution data utilizing a solvent mask filled with constant electron density. The B-factors of both models were then refined with the solvent mask present. $F_o - F_c$ omit electron density maps were then calculated for the residue 112–118 region from the complete modified data set.

The density of the $F_o - F_c$ omit map for the K116D variant suggested a radical change in the conformation of the residues 112–117 compared to that of nuclease A. A $F_o - F_c$ SA omit map contoured at 1.8σ clearly defines the positions of residues 112, 113, and 116–118, including a *trans* 116–117 peptide bond. When the map is contoured at 1.5σ, the path of the backbone trace for residues 114 and 115 becomes visible. A model for this conformation was built into the 1.8σ density, positioning residues 114 and 115 in unstrained conformations (with regard to the backbone ϕ and ψ angles) that were consistent with the 1.5σ density. The full model of K116D was then refined to an R-factor of 18.9% using data from 6 to 1.9 Å.

The density of the $F_o - F_c$ omit map for the K116N variant had similar features to that of K116D. The electron density demonstrated a good fit to the K116D loop model in residues 116–118, but not in residues 113–115. The density suggests small changes in the conformation of residues 113–114, and a different orientation of the 115–116 peptide group compared to the K116D model. A separate model of residues 112–117 was built into the density starting with the K116D loop model (see Fig. 8). Both the model with the K116D conformation and the rebuilt model were refined to the data (using data from 6.0 to 2.0 Å resolution) through rounds of positional and B-factor refinement. Both models yield the same R-factor of 18.0%.

Acknowledgments

This work was supported by an NIH grant to R.O.F. (AI23923), the Howard Hughes Medical Institute, and by shared instrumentation grants to Stanford University and Yale University (NIH RR 03475, NSF DMB8610557, and ACS RD259) for the NMR spectrometers. A.H. is a Howard Hughes Medical Institute Predoctoral Fellow.

References

Alexandrescu AT, Hinck AP, Markley JL. 1990. Coupling between local structure and global stability of a protein: Mutants of staphylococcal nuclease. *Biochemistry 29*:4516–4525.

Alexandrescu AT, Mills DA, Ulrich EL, Chinami M, Markley JL. 1988. NMR assignments of the four histidines of staphylococcal nuclease in native and denatured states. *Biochemistry 27*:2158–2165.

Alexandrescu AT, Ulrich EL, Markley JL. 1989. Hydrogen-1 NMR evidence for three interconverting forms of staphylococcal nuclease: Effects of mutations and solution conditions on their distribution. *Biochemistry 28*:204–211.

Arnone A, Bier CJ, Cotton FA, Hazen EE Jr, Richardson DC, Richardson JS. 1969. The extracellular nuclease of *Staphylococcus aureus*: Structure of the native enzyme and an enzyme-inhibitor complex at 4 Å resolution. *Proc Natl Acad Sci USA 64*:420–427.

Bruccoleri RE, Haber E, Novotny J. 1988. Structure of antibody hypervariable loops reproduced by a conformational search algorithm. *Nature 335*:564–568.

Brünger AT, Kuriyan J, Karplus M. 1987. Crystallographic *R* factor refinement by molecular dynamics. *Science 235*:458–460.

Bundi A, Wüthrich K. 1979. ^{1}H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-gly-gly-X-ala-OH. *Biopolymers 18*:285–297.

Chothia C, Lesk AM. 1987. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol 196*:901–917.

Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, Coleman PM, Spinelli S, Alzari PM, Poljak RJ. 1989. Conformations of immunoglobulin hypervariable regions. *Nature 342*:877–883.

Collura V, Higo J, Garnier J. 1993. Modeling of protein loops by simulated annealing. *Protein Sci 2*:1502–1510.

Cotton FA, Hazen EE Jr, Legg MJ. 1979. Staphylococcal nuclease: Proposed mechanism of action based on structure of enzyme-thymidine 3′,5′-bisphosphate-calcium ion complex at 1.5 Å resolution. *Proc Natl Acad Sci USA 76*:2551–2555.

Evans PA, Dobson CM, Kautz RA, Hatfull G, Fox RO. 1987. Proline isomerism in staphylococcal nuclease characterized by NMR and site-directed mutagenesis. *Nature 329*:2266–2268.

Evans PA, Kautz RA, Fox RO, Dobson CM. 1989. A magnetization-transfer nuclear magnetic resonance study of the folding of staphylococcal nuclease. *Biochemistry 28*:362–370.

Fine RM, Wang H, Shenkin PS, Yarmush DL, Levinthal C. 1986. Predicting antibody hypervariable loop conformations II. Minimization and molecular dynamics studies of MCPC603 from many randomly generated loop conformations. *Proteins Struct Funct Genet 1*:342–362.

Fink AL, Calciano LJ, Goto Y, Nishimura M, Swedberg SA. 1993. Characterization of the stable, acid-induced, molten globule-like state of staphylococcal nuclease. *Protein Sci 2*:1155–1160.

Foote J, Winter G. 1992. Antibody framework residues affecting the conformation of the hypervariable loops. *J Mol Biol 224*:484–499.

Fox RO, Evans PA, Dobson CM. 1986. Multiple conformations of a protein demonstrated by magnetization transfer NMR spectroscopy. *Nature 320*:192–194.

Green SM, Shortle D. 1993. Patterns of nonadditivity between pairs of stability mutations in staphylococcal nuclease. *Biochemistry 32*:10131–9.

Higuchi R. 1990. In: Innis MA, Gelford DH, Sninsky JJ, White TJ, eds. *PCR protocols: A guide to methods and applications*. San Diego, California: Academic Press. pp 177–183.

Hodel A, Kautz RA, Adelman DM, Fox RO. 1994. The importance of anchorage in determining a strained protein loop conformation. *Protein Sci 3*:549–556.

Hodel A, Kautz RA, Jacobs MD, Fox RO. 1993. Stress and strain in staphylococcal nuclease. *Protein Sci 2*:838–850.

Hodel A, Kim SH, Brünger AT. 1992. Model bias in macromolecular crystal structures. *Acta Crystallogr A 48*:851–858.

Hodel A, Rice LM, Simonson T, Fox RO, Brünger AT. 1995. Proline *cis-trans* isomerization in staphylococcal nuclease: Multi-substate free energy perturbation calculations. *Protein Sci 4*. Forthcoming.

Hynes TR, Fox RO. 1991. The crystal structure of staphylococcal nuclease refined at 1.7 Å resolution. *Proteins Struct Funct Genet 10*:92–105.

Hynes TR, Hodel A, Fox RO. 1994. Engineering alternative β-turn types in staphylococcal nuclease. *Biochemistry 33*:5021–5030.

Jones TA. 1985. Interactive computer graphics: FRODO. *Methods Enzymol 115*:252–270.

Kautz RA, Gill JA, Fox RO. 1990. Assignment of histidine resonances in the NMR spectrum of staphylococcal nuclease using site-directed mutagenesis. In: Craik C, ed. *Protein and pharmaceutical engineering*. New York: Wiley-Liss, Inc. pp 1–15

Kuwajima K, Okayama N, Yamamoto K, Ishihara T, Sugai S. 1991. The Pro 117 to glycine mutation of staphylococcal nuclease simplifies the unfolding-folding kinetics. *FEBS Lett 290*:135–138.

Loll PA, Lattman EE. 1989. The crystal structure of the ternary complex of staphylococcal nuclease, Ca^{2+}, and the inhibitor pdTp, refined at 1.65 Å. *Proteins Struct Funct Genet 5*:183–201.

Luzzati V. 1952. Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr 5*:802–810.

Matthews BW, Czerwinski EW. 1975. Local scaling: A method to reduce systematic errors in isomorphous replacement and anomalous scattering measurements. *Acta Crystallogr A31*:480–487.

Pace CN. 1975. The stability of globular proteins. *Crit Rev Biochem 3*:1–43.

Press WH, Flannery BP, Teukolsky SA, Vetterling WT. 1986. *Numerical recipes*. Cambridge, UK: Cambridge University Press.

Raleigh DP, Evans PA, Pitkeathly M, Dobson CM. 1992. A peptide model for proline isomerism in the unfolded state of staphylococcal nuclease. *J Mol Biol 228*:338–342.

Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem 34*:167–297.

Ring CS, Kneller DG, Langridge R, Cohen FE. 1992. Taxonomy and conformational analysis of loops in proteins. *J Mol Biol 224*:685–699.

Shortle DJ, Meeker AK. 1986. Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. *Proteins Struct Funct Genet 1*:81–89.

Tucker PW, Hazen EE, Cotton FA. 1979. Staphylococcal nuclease reviewed: A prototypic study in contemporary enzymology. IV. Nuclease as a model for protein folding. *Mol Cell Biochem 23*:131–141.

Wang J, Hinck AP, Loh SN, Markley JL. 1990. Two-dimensional NMR studies of staphylococcal nuclease: Evidence for conformational heterogeneity from hydrogen-1, carbon-13, and nitrogen-15 spin system assignments of the aromatic amino acids in the nuclease H124L-thymidine 3′,5′-bisphosphate-Ca^{2+} ternary complex. *Biochemistry 29*:4242–4253.

Wilmot CM, Thornton JM. 1990. β-Turns and their distortions: A proposed new nomenclature. *Protein Eng 3*:479–493.

Zoller ML, Smith M. 1983. Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. *Methods Enzymol 100*:468–500.