Model for the Catalytic Domain of the Proofreading ϵ Subunit of *Escherichia coli* DNA Polymerase III Based on NMR Structural Data[†]

Eugene F. DeRose,[‡] Dawei Li,[‡] Thomas Darden,[‡] Scott Harvey,[§] Fred W. Perrino,[§] Roel M. Schaaper,[‡] and Robert E. London*[,‡]

Laboratory of Structural Biology and Laboratory of Molecular Genetics, NIEHS, Box 12233, Research Triangle Park, North Carolina 27709, and Department of Biochemistry, Wake Forest University Medical Center, Winston−Salem, North Carolina 27157

Received July 9, 2001; Revised Manuscript Received October 12, 2001

ABSTRACT: The DNA polymerase III holoenzyme (HE) is the primary replicative polymerase of *Escherichia coli*. The ϵ subunit of the HE complex provides the 3′-exonucleolytic proofreading activity for this enzyme complex. ϵ consists of two domains: an N-terminal domain containing the proofreading exonuclease activity (residues 1−186) and a C-terminal domain required for binding to the polymerase (α) subunit (residues 187−243). Multidimensional NMR studies of ^{2}H-, ^{13}C-, and ^{15}N-labeled N-terminal domains (ϵ186) were performed to assign the backbone resonances and measure H^N-H^N nuclear Overhauser effects (NOEs). NMR studies were also performed on triple-labeled [U-^{2}H,^{13}C,^{15}N]ϵ186 containing Val, Leu, and Ile residues with protonated methyl groups, which allowed for the assignment of H^N-CH_3 and CH_3-CH_3 NOEs. Analysis of the $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO shifts, using chemical shift indexing and the TALOS program, allowed for the identification of regions of the secondary structure. H^N-H^N NOEs provided information on the assembly of the extended strands into a β-sheet structure and confirmed the assignment of the α helices. Measurement of H^N-CH_3 and CH_3-CH_3 NOEs confirmed the β-sheet structure and assisted in the positioning of the α helices. The resulting preliminary characterization of the three-dimensional structure of the protein indicated that significant structural homology exists with the active site of the Klenow proofreading exonuclease domain, despite the extremely limited sequence homology. On the basis of this analogy, molecular modeling studies of ϵ186 were performed using as templates the crystal structures of the exonuclease domains of the Klenow fragment and the T4 DNA polymerase and the recently determined structure of the *E. coli* Exonuclease I. A multiple sequence alignment was constructed, with the initial alignment taken from the previously published hidden Markov model and NMR constraints. Because several of the published structures included complexed ssDNA, we were also able to incorporate an A−C−G trinucleotide into the ϵ186 structure. Nearly all of the residues which have been identified as mutators are located in the portion of the molecule which binds the DNA, with most of these playing either a catalytic or structural role.

The mechanisms by which organisms faithfully replicate their DNA are of considerable interest. The best characterized replisome is the DNA polymerase III holoenzyme (HE) of the bacterium *Escherichia coli*, which has been studied in significant detail both genetically and biochemically. HE consists of 17 subunits, the products of 10 different genes (*1*). It contains two core polymerases, responsible for simultaneously replicating the leading and lagging strands. Each core consists of three tightly bound subunits: the 130 kD α subunit (the polymerase), the 27.5 kD ϵ subunit (a proofreading 3′ → 5′ exonuclease), and the small (8 kD) θ subunit, whose function is as yet undetermined. These three subunits are bound together in the linear order $\alpha-\epsilon-\theta$. In addition to the two cores, HE contains two sliding clamps (dimers of the β subunit) that tether the two polymerases to the DNA to enable high processivity and the multisubunit DnaX complex. The DnaX complex dimerizes the two cores and also performs the loading and unloading of the sliding clamp, which is of particular importance in the discontinuously produced lagging strand.

The fidelity of the replication process is determined primarily by (i) the polymerase, which preferentially selects correct nucleotides over incorrect nucleotides during the base insertion step, and (ii) the ϵ subunit, which can remove polymerase misinsertion errors through its 3′ → 5′ exonuclease activity. The estimated contributions to the overall fidelity are 10^4-10^5 for the base insertion step (misinsertion rate $10^{-4}-10^{-5}$) and 10^2-10^3 for the proofreading step, producing a final misincorporation rate of about 10^{-7} (*2*). Mechanistically, the precise interaction between the polymerase and the exonuclease is of particular interest, because these two activities are located at distinct sites, requiring the shuttling of the 3′ terminal base from one site to the other, which may be a rate-limiting factor in fidelity.

Our studies have focused on the functional and structural aspects of the ϵ subunit, the product of the *dnaQ* gene. Genetic studies have revealed ϵ to have both a fidelity role, as evidenced by the strong mutator phenotype of *E. coli*

[†] This work was supported in part by NIH Grant CA75350 (F.W.P.).
* To whom correspondence should be addressed.
[‡] NIEHS.
[§] Wake Forest University Medical Center.

10.1021/bi0114170 CCC: $22.00
Published on Web 12/04/2001

strains carrying *dnaQ* mutations (*3*), and a structural role within HE. The structural role is evidenced by the conditionally lethal phenotype of strains lacking the ϵ proofreading subunit (*3, 4*). This phenotype likely results from a destabilization of the α subunit in the absence of ϵ. Our studies have revealed ϵ to consist of two domains: an N-terminal domain (186 amino acids) that contains both the catalytic activity and the binding site for interaction with the θ subunit and a C-terminal domain (57 amino acids) that contains the site for binding to the polymerase (*5, 6*). Furthermore, a genetic analysis of *dnaQ* mutator mutants has provided insight into the important catalytic and structural residues of ϵ (*7*). The presumed active site contains essential aspartic and glutamic acid residues contributed by three separate sequence motifs, Exo I−III. Interestingly, while Exo motifs I and II have readily recognizable counterparts in other proofreading exonucleases, the Exo III motif is significantly dissimilar (*7, 8*), placing ϵ in a separate subgroup within the superfamily of exonucleases and suggesting an evolutionarily distinct lineage (*9*).

Crystallography has been used to analyze the structure of the proofreading exonuclease of a number of DNA polymerases, including *E. coli* DNA polymerase I and the polymerase of bacteriophages T4 and RB69 (*10*−*14*). These structures have confirmed the essential role of the conserved Exo I−III motifs as constituents of the active site, participating in catalysis and metal binding. As indicated, the Exo III motif as present in *E. coli* (and related organisms) appears distinct, and a similarity of structure should not be assumed a priori. The *E. coli* proofreading system provides a potentially distinct system because of at least two additional features. First, the *E. coli* proofreading exonuclease is contained on a separate subunit, rather than being on the same polypeptide as the polymerase. Second, the system contains the tightly bound θ subunit, which is not found in other organisms either as a free subunit or as a corresponding domain of other replication proteins.

Here, we present the results of our structural analysis of the 186-residue N-terminal catalytic domain of ϵ. The full-length protein, which recently has been shown to consist of two domains (a catalytic N-terminal domain, ϵ186, and a C-terminal α-subunit-binding domain (*5, 6*)), was found to be completely unsuitable for NMR analysis. Although ϵ186 proved to be much better behaved than the full-length protein, the limited stability of the isolated catalytic domain at high concentrations presented significant problems, requiring the use of conditions which effectively increased the rotational correlation time of the protein. This ultimately required the use of methods developed to deal with larger proteins. Similar problems have recently been reported for ϵ186 by Hamdan et al. (*15*). The assignment of the backbone resonances of ϵ186 was accomplished by using triple-labeled [U-^{2}H,^{13}C,^{15}N]ϵ186. To obtain additional structural constraints, we utilized the selective protonation strategy developed by Kay and co-workers (*16*−*18*) in which the δ1-methyl group of isoleucine, the γ-methyl groups of valine, and the δ-methyl groups of leucine remain selectively protonated in a deuterated background. After the assignment of these methyl resonances, long-range methyl−methyl and methyl−amide nuclear Overhauser effects (NOEs) were obtained, which provided a limited resolution structure of the exonuclease catalytic domain.

Molecular modeling approaches have typically proven valuable when strong homology to proteins of known structure is available or when homology is lower but other information exists which can be used to complement the information available (*19*). In the present case, the homology of ϵ to proteins of known structure is quite low. A pairwise analysis of the fractional identity relative to the aligned positions in ϵ186 obtained using the program ClustalX (*20*) gives 20% identity for the Klenow exonuclease domain (*11*), 17% for the T4 exonuclease domain (*13*), and 19% for *E. coli* Exonuclease I (*21*). These low values preclude the development of a useful homology model in the absence of additional experimental data. In the present study, we have utilized the initial NMR structural data in combination with a multiple sequence alignment with these proteins to develop a model for ϵ186. Despite the significant amino acid divergence, our model shows ϵ to be structurally similar to other, evolutionarily distant proofreading exonucleases. Further, the derived structure provides useful insights into the basis for previously identified *dnaQ* mutator mutants (*7*).

EXPERIMENTAL PROCEDURES

ϵ186 Protein. A bacterial expression system was used with four different isotopic-labeling strategies to produce [U-^{15}N]ϵ186, [U-^{13}C,^{15}N]ϵ186, [U-^{2}H,^{13}C,^{15}N]ϵ186 and a selectively protonated ϵ186 for the NMR studies. For the production of the selectively protonated ϵ186, the approach described by Goto et al. (*18*) in which *E. coli* are grown in a 99% D_2O medium containing $^{15}NH_4Cl$, [U-^{2}H,U-^{13}C]D-glucose, [3,3-$^2H_2$, U-$^{13}C_4$]α-ketobutyrate, and [3-^{2}H,U-$^{13}C_5$]α-ketoisovalerate (Cambridge Isotopes, Andover, MA) was used. The α-ketobutyrate is a metabolic precursor of isoleucine, and α-ketoisovalerate is a precursor of leucine and valine. This procedure results in uniform labeling with ^{2}H, ^{13}C, and ^{15}N except for the δ-methyl group of isoleucine and the γ- and δ-methyl groups of valine and leucine. The protonated methyl and amide groups in a deuterated background allow for additional methyl−amide and methyl−methyl NOEs to be determined in order to provide many long-range distance constraints. For ϵ186 expression, the plasmid pe186 was electroporated into *E. coli* BL21(DE3), and a single colony was selected from an LB plate for growth at 37 °C in a liquid M9 medium. The plasmid pe186 contains codons 1−186 of the *dnaQ* gene under transcriptional control of the bacteriophage T7 promoter and translational control of a consensus ribosome-binding site introduced upstream of the *dnaQ* start codon (*6*). For the production of [U-^{13}C,^{15}N]ϵ186, the M9 contained [U-^{13}C]D-glucose (2 g/L) and $^{15}N-NH_4Cl$ (1 g/L; Isotec, Inc., Miamisburg, OH). For the triple-labeled [U-^{2}H,^{13}C,^{15}N]ϵ186, the log-phase culture grown in $^{15}N-NH_4Cl$, [U-^{13}C]D-glucose M9 was sequentially changed from 100% H_2O to 99% D_2O.

To maintain log-phase growth, cultures were never diluted more than fivefold. When the [U-^{13}C,^{15}N]ϵ186 and [U-^{2}H,^{13}C,^{15}N]ϵ186 cultures (3 × 1 L) reached $A_{595} = 0.4$, isopropyl-β-D-thiogalactopyranoside (IPTG) (0.5 mM) was added, and induction was for 12 h. For the selectively protonated ϵ186, [3,3-$^2H_2$,U-$^{13}C_4$]α-ketobutyrate (80 mg/L) and [3-^{2}H,U-$^{13}C_5$]α-ketoisovalerate (80 mg/L) were added when cultures (3 × 1 L) reached $A_{595} = 0.4$. Sixty minutes later, IPTG (0.5 mM) was added, and induction was for 12 h. All of the labeled ϵ186 proteins were purified as previously

described (*6*). The purified [U-^{2}H,^{13}C,^{15}N]ϵ186 protein was dialyzed against 0.9 M Tris-d_{11} (pH 7.0), 10 mM D,L-1,4-dithiothreitol-d_{10} (Isotec, Inc., Miamisburg, OH), and 1 mM $MgSO_4$. The NMR buffer also contained 1 mM sodium azide to retard bacterial growth, 0.1 mM 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), a protease inhibitor, and 10% D_2O for the NMR lock. The selectively protonated ϵ186 was concentrated to a 0.72 mM solution using a centricon concentrator with a 10 kD cutoff (Millipore, Inc., Bedford, MA).

NMR Studies. The NMR experiments were performed on Varian 500 MHz UNITYplus and 600 MHz UNITYInova spectrometers. Studies on the UNITYplus 500 utilized a Nalorac (Martinez, CA) 5 mm ^{1}H{^{13}C,^{15}N} triple-resonance probe with actively shielded *z*-axis gradients and variable-temperature capability. Studies on the UNITYInova 600 used a 5 mm Varian probe of similar design. The sequential backbone and C$^{\beta}$ resonance assignments were established by the combined analysis of CT-HNCA (*22*), CT-HN(CO)CA, HN(CA)CB, and HN(COCA)CB (*23*) spectra. Backbone carbonyl resonances were assigned from an HNCO experiment (*24*). The NMR data were processed using NMRPipe (*25*), and resonances were assigned using NMRView (*26*). The HNCA, HN(CO)CA, HN(CA)CB, and HN(COCA)CB experiments were acquired on a Varian UNITYplus 500 MHz spectrometer as described by Yamazaki et al. (*22, 23*), with pulse sequences obtained from the Kay group. In the HNCA and HN(CO)CA experiments, 64 × 32 × 512 complex points were acquired, with acquisition times of 19.6, 16.0, and 78.8 ms in t_1-t_3, respectively. In the HN(CA)CB and HN(COCA)CB experiments, 64 × 32 × 512 complex points were acquired, with acquisition times of 9.0, 16.0, and 78.8 ms in t_1-t_3, respectively. The HNCA experiment was acquired with 32 scans per increment and a 1.2 s delay between scans, resulting in a total acquisition time of approximately 93 h. HN(CO)CA, HN(CA)CB, and HN(COCA)CB were acquired with 16 scans per increment and a 1.6 s delay between scans, resulting in total acquisition times of approximately 63 h. The HNCO experiment was acquired on a UNITYplus 500 MHz spectrometer using Varian's ghn_co pulse sequence, with 64 × 32 × 512 complex points and acquisition times of 30.3, 16.0, and 78.8 ms in t_1-t_3, respectively. Sixteen scans per increment were acquired with a 1.2 s delay between scans, resulting in a total acquisition time of approximately 48 h. The data were processed using NMRPipe as described in ref *23*, with 90° shifted squared sine-bell apodization functions in t_3 and 90° shifted sine-bell functions in t_1 and t_2. Secondary structure predictions based on the observed shifts of the $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO resonances were made using the chemical shift indexing (CSI)[1] method (*27*) and the TALOS program (*28*). A four-dimensional (4D) ^{15}N,^{15}N-edited NOESY experiment (*29, 30*) was also performed on the triple-labeled protein. The 4D spectrum was acquired on a Varian UNITYInova 600 MHz spectrometer using the sequence of Venters et al. (*29*), with a mixing time of 400 ms, 64 × 16 × 16 × 512 complex points, and acquisition times of 8.0, 8.9, 8.9, and 64.0 ms in t_1-t_4, respectively. Four scans per increment were acquired, with a 1.2 s delay between scans, resulting in a total acquisition time of approximately 10 days. The acquisition dimension was processed using a polynomial filter to suppress the solvent signal in the acquisition dimension, followed by apodization with an 81° shifted squared sine-bell function, and zero filled to 1024 points before Fourier transformation. The downfield half of the spectrum was extracted in the acquisition dimension and in t_1. The data were apodized with an 81° shifted sine-bell function in t_1 and zero filled to 256 points prior to Fourier transformation. The data were extended with forward−backward linear prediction (*31*) in t_2 and t_3, apodized with 90° shifted squared sine-bell functions, and zero filled to 64 points prior to Fourier transformation. The imaginary data were deleted in all dimensions, resulting in a final spectrum of 128 × 64 × 64 × 512 real points. The 4D NOESY spectrum allowed for the determination of short- and medium-range NOE data connecting the amide protons, which in turn were used to confirm the sequential resonance assignments and the secondary structure predictions of the α-helical regions. The assignment of the long-range H^N-H^N NOEs allowed for the identification of interstrand hydrogen-bonding patterns so that the β-sheet structure could be identified.

The chemical shifts of the Val γ, Leu δ, and Ile δ1 resonances of the selectively protonated sample were assigned from H(CCO)NH−TOCSY and (H)C(CO)NH−TOCSY spectra. The (H)C(CO)NH−TOCSY and H(CCO)NH−TOCSY spectra were acquired on a Varian UNITYInova 600 MHz spectrometer as described by Gardner et al. (*32*), using pulse sequences obtained from the Kay group with mixing times of 20.7 ms. The (H)C(CO)NH−TOCSY experiment was acquired with 64 × 32 × 512 complex points and acquisition times of 7.0, 17.8, and 64.0 ms in t_1-t_3, respectively. The H(CCO)NH−TOCSY experiment was also acquired with 64 × 32 × 512 complex points and acquisition times of 21.3, 17.8, and 64.0 ms in t_1-t_3, respectively. The (H)C(CO)NH−TOCSY experiment was obtained with 128 scans per increment, with a 0.8 s delay between scans, resulting in a total acquisition time of approximately 11 days. The H(CCO)NH−TOCSY experiment was obtained with 64 scans per increment, with a 1.0 s delay between scans, resulting in a total acquisition time of approximately 7 days. The spectra were processed like the triple-resonance experiments described previously. A three-dimensional (3D) CT−^{13}C,^{13}C -edited NOESY experiment and a 4D ^{15}N,^{13}C-edited NOESY experiment were also performed on the selectively protonated sample (*33*−*35*). The 3D CT−^{13}C,^{13}C -edited NOESY spectrum was acquired on a Varian UNITYInova 600 MHz spectrometer using the pulse sequence shown in Figure 1D of ref *33* with a mixing time of 175 ms; 60 × 30 × 512 complex points were acquired, with acquisition times of 24.5, 10.0, and 64 ms in t_1-t_3, respectively. The spectrum was acquired with 64 scans per increment, with a 1.0 s delay between scans, resulting in a total acquisition time of approximately 7 days. The acquisition dimension was processed using a solvent suppression filter (*36*), followed by apodization with a 90° shifted squared sine-bell function, and zero filled to double the number of points before Fourier transformation. The data were apodized with a 90° shifted sine-bell function in t_1 and zero filled to 256 complex points prior to Fourier transformation. The data were extended with

[1] Abbreviations: CSI, chemical shift indexing; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; Tris-d_{11}, perdeuterated tris(hydroxymethyl)aminomethane; (#), indicates that the resonance corresponds to either of the prochiral methyl groups.

forward−backward linear prediction (*31*) in t_3, apodized with a 90° shifted sine-bell function, and zero filled to 128 complex points prior to Fourier transformation. The imaginary data were deleted in all dimensions, resulting in a final spectrum of 256 × 128 × 512 real points. The 4D ^{15}N,^{13}C-edited NOESY spectrum was acquired on a Varian UNITYInova 600 MHz spectrometer using the sequence of Muhandiram et al. (*35*), with a mixing time of 175 ms; 64 × 16 × 16 × 512 complex points were obtained, with acquisition times of 9.7, 5.3, 8.9, and 64.0 ms in t_1-t_4, respectively. Four scans per increment were acquired with a 0.9 s delay between scans, resulting in a total acquisition time of approximately 7 days. The spectrum was processed like the ^{15}N,^{15}N-edited NOESY spectrum, but the acquisition dimension was zero filled to 512 points prior to Fourier transformation, resulting in a final spectrum size of 128 × 64 × 64 × 256 real points. Long-range H^N-CH_3 and CH_3-CH_3 NOEs assigned from these spectra confirmed the β-sheet structure and assisted in the positioning of the α helices for molecular modeling.

Molecular Modeling. Molecular modeling involved a combination of sequence alignment, structural superposition, preliminary manual backbone and side-chain model building, and molecular dynamics refinement of the model in the presence of NMR-derived distance and torsion restraints. Sequence alignments were performed using the ClustalW program (*20*) using both the multiple alignment mode and the profile alignment mode. Default substitution and gap costs were used. Input sequences for multiple alignments were gathered using the standalone Blast suite of programs (*37*). Structural superpositions were performed using a code developed locally to solve the least-squares superposition problem. All of the initial modeling, including loop building and side-chain placement, was performed using the Swiss-PdbViewer program (*38*). The initial model was refined using the AMBER package of simulation software (*39*). The model was solvated and ions added using the LEaP module of the AMBER package. The Cornell et al. force field (*40*) was used, and molecular dynamics with NMR-derived distance and torsion restraints in the presence of explicit solvent using a PME (*41*) treatment of electrostatics was performed using the Sander module of the AMBER package. Graphical figures of the resulting model of ϵ186 were produced using Molscript (*42*) and Raster3D (*43*) as well as Swiss-PdbViewer.

Building a molecular model of ϵ186 using straightforward homology modeling techniques was not possible because of the low sequence identity between ϵ186 and those 3′−5′ exonuclease domains for which structural data are currently available. That is, the reliability of homology models is limited by the quality of the sequence alignments, which is generally poor for sequence pairs having such a low identity. However, using the additional information available from NMR, in particular the β-sheet structure and approximate helix placement via secondary structure assignment and NOE distance restraints, allowed us to more accurately align the residues of ϵ186 to those of existing exonuclease domain structures.

We began with an existing multiple alignment (*9*) which included ϵ, *E. coli* Exonuclease I, and the 3′−5′ exonuclease domains of the Klenow fragment and T4 DNA polymerase. α-Carbon positions of conserved residues were then used to superimpose these structures. A manual inspection of backbone and side-chain positions in the superimposed structures was used to define a structural alignment, thus defining a conserved structural core consisting of a central β sheet with helices packing above and below. The core β sheet was quite similar to that identified by NMR in ϵ186, and the latter could be superposed onto it by aligning the conserved residues of motif I. Next, helices in ϵ186 (assigned by NMR) were aligned with those in the conserved core by a combination of sequence alignment and constraints provided by NOEs. Residues not aligning to the conserved core were aligned to one of the existing structures whenever possible. For example, residues 53−78 of ϵ186 were aligned to 52−80 in the *E. coli* Exonuclease I structure (*21*), whose secondary structure is quite similar to that of ϵ186 in this region.

After assigning the backbone positions of ϵ186 using the backbones of the three template structures, we assigned side-chain conformations using existing side-chain conformations for identical or closely similar residues or manual selection from the rotamer library provided in Swiss-PdbViewer (*38*) otherwise. Side-chain clashes were relieved manually. The SwisPDBViewer program was used for side-chain modeling. A single-strand DNA fragment of three nucleotides, together with two metal ions and three waters coordinated to metal B, were included in the model by superposing the Klenow fragment exonuclease domain onto the model of ϵ186. The final model was refined, with NOE distance restraints and secondary structure restraints in full solvent with careful treatment of electrostatics using the Sander module of the AMBER package.

RESULTS

[U-^{15}N] labeled ϵ186 was obtained by growing the cells in minimal media containing [^{15}N]ammonium chloride. The ^{1}H−^{15}N HSQC spectra of ϵ186 are consistent with the conclusion that it is a well-folded protein. Unfortunately, at a 1 mM concentration, the protein precipitated over a period of hours when observed in a 50 mM phosphate/50 mM NaCl (pH 7) buffer at 15 °C, making it unsuitable for long-term structural NMR studies. The protein precipitation was found to exhibit a strong concentration dependence. However, concentrations at which sufficient stability was obtained were well below the threshold required for NMR analysis.

To identify conditions under which the protein might be more stable, we set up a "precipitation screen" to allow for a comparison of a broad range of buffer conditions. The assay yielded two optimal buffer conditions: very low salt or high concentrations of a Tris buffer (pH 7.0). NMR studies performed on ϵ186 in a low-salt buffer indicated that while the protein did not precipitate, it underwent significant structural changes, possibly involving significant unfolding. In contrast, the structure of the protein in a 1 M Tris buffer (pH 7) was found to be sufficiently stable to allow for long-term, multidimensional NMR studies to be performed at 20 °C, although small amounts of precipitation were observed to occur after periods of several days.

Initial triple-resonance experiments on a double-labeled [U-^{13}C,^{15}N] sample did not yield spectra with sufficient signal-to-noise to allow for the assignment of the backbone resonances (Figure 1A). Among other factors, the relatively

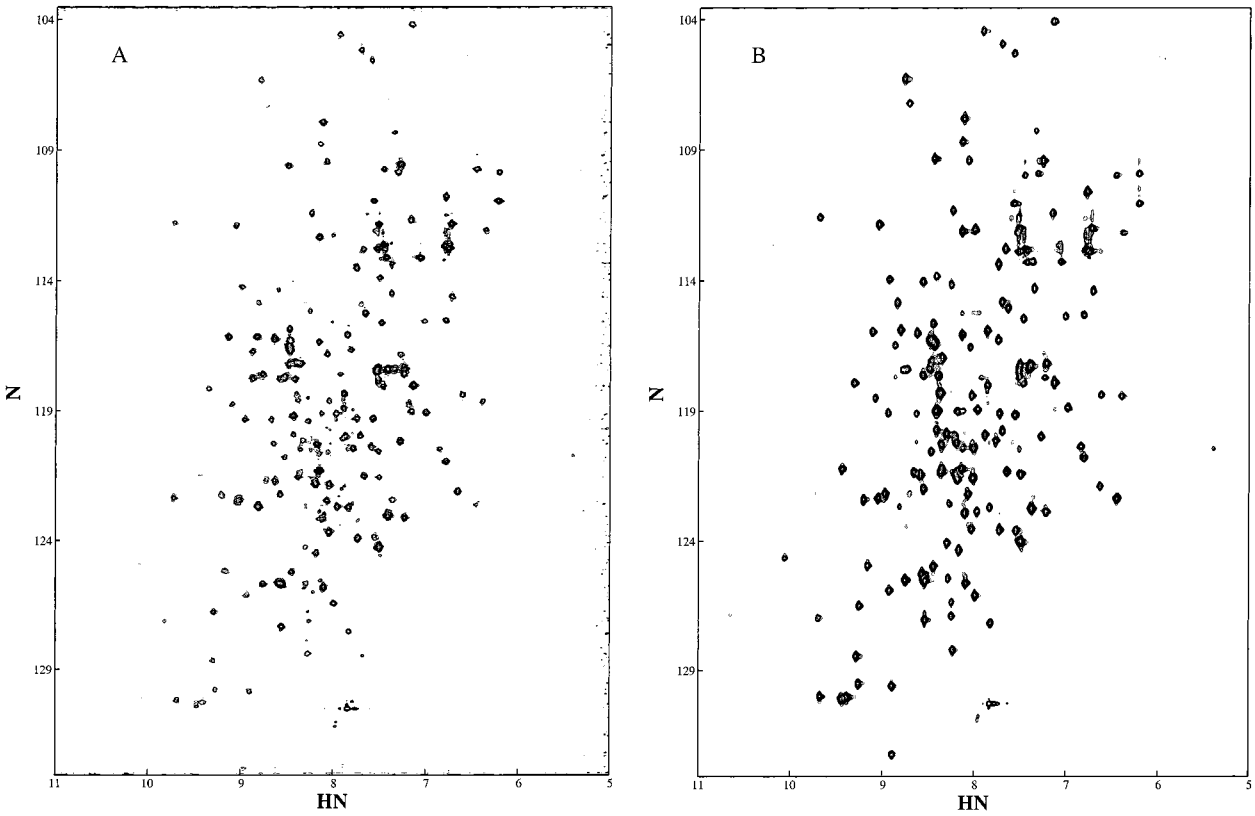

FIGURE 1: ^{1}H-^{15}N HSQC spectra of (A) [U-^{13}C,^{15}N]ϵ186 and (B) [U-^{2}H,^{13}C,^{15}N]ϵ186. The spectra were obtained using the Varian gNhsqc pulse sequence on a UNITYInova 600 MHz spectrometer, with 128 × 512 complex points and acquisition times of 71 and 64 ms in t_1 and t_2, respectively. Eight scans were acquired per increment, with a 1.0 s delay between scans. The enzyme was in a 1 M Tris-d_{11} (pH 7.0) buffer, 1 mM DTT-d_{10}, and 1 mM $MgCl_2$. The spectra were obtained at 20 °C, the temperature at which all of the NMR studies were performed.

low temperature used (20 °C) and perhaps the high buffer concentration increased the rotational correlation time of the protein, resulting in a loss of sensitivity. Figure 1B shows the ^{1}H-^{15}N HSQC spectrum obtained for the triple-labeled [U-^{2}H,^{13}C,^{15}N]ϵ186. This spectrum showed significant improvement over that of the double-labeled enzyme. This sensitivity gain was found to be critical to the success of the 3D experiments used to assign the backbone resonances of ϵ186. Although many regions of the spectra are highly congested, the signal-to-noise was sufficient to allow for the assignment of the chemical shifts of 157 amide resonances (87%), 158 $^{13}C^{\alpha}$ resonances (85%), 146 $^{13}C^{\beta}$ resonances (85%), and 152 ^{13}CO resonances (82%) from the HNCA, HN(CO)CA, HN(CA)CB, HN(COCA)CB, and HNCO experiments. In addition, all of the Val $^{13}C^{\gamma}$-methyl resonances, $^{13}C^{\delta}$-methyl resonances for 16 out of 17 Leu residues, and 12 out of 13 Ile $^{13}C^{\delta 1}$ resonances were assigned from the (H)C(CO)−NH TOCSY spectrum of the selectively methyl protonated sample. The $C^{\gamma}H$ resonances were assigned for 11 out of 12 Val residues, $C^{\delta}H$ resonances were assigned for 16 out of 17 Leu residues, and 12 out of 13 Ile $C^{\delta 1}H$ resonances were assigned from the H(CCO)−NH TOCSY spectrum. Residues near the C terminal gave rise to relatively more intense peaks. These more intense resonances presumably result from a more locally dynamic structure, consistent with the location of these residues in proximity to a "Q-linker" motif which forms the connection between the two domains of ϵ (*5*, *6*, *44*).

Secondary structure predictions derived from an analysis of the $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO chemical shifts using the CSI program of Wishart and Sykes (*27*; Protein Engineering Network Centres of Excellence (PENCE), Medical Research Council of Canada (MRC) Joint Software Centre) and the program TALOS (*28*) are given in Figure 2. According to this figure, ϵ186 contains a total of six α helices and five β strands. The predictions of secondary structure based on the CSI and TALOS analyses are further supported by the large number of $d_{NN(i,i+2)}$ and $d_{NN(i,i+3)}$ NOEs, shown in Figure 2, that provide strong confirmation for the identification of the α helices. Because of the long mixing time of 400 ms and the fact that the sample was deuterated, many sequential amide−amide and some $d_{NN(i,i+2)}$ and $d_{NN(i,i+3)}$ NOEs were observed, even in nonhelical regions of the protein. Although the CSI consensus analysis does not indicate that residues Glu61 through His66 form an α helix, the $^{13}C^{\alpha}$ chemical shifts are indicative of an α helix, as further supported by the observation of four $d_{NN(i,i+2)}$ and two $d_{NN(i,i+3)}$ NOEs in this stretch of residues. In addition, φ and ψ torsion angles as determined from an analysis of the $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO chemical shifts by TALOS for this stretch of residues are also indicative of an α helix. It is unclear whether residues 69−73 form another α helix because the CSI analysis and the $^{13}C^{\alpha}$ chemical shifts do not indicate any secondary structure, while the TALOS and NOE data are consistent with a helical structure.

The structure of the β sheet is apparent from the 25 interstrand H^{N}−H^{N} NOEs measured (Table 1a), which allow for the alignment of the five β strands as shown in Figure 3. Seven methyl−methyl (Table 1b) and 15 long-range amide−methyl (Table 1c) NOEs reinforce the alignment of the β

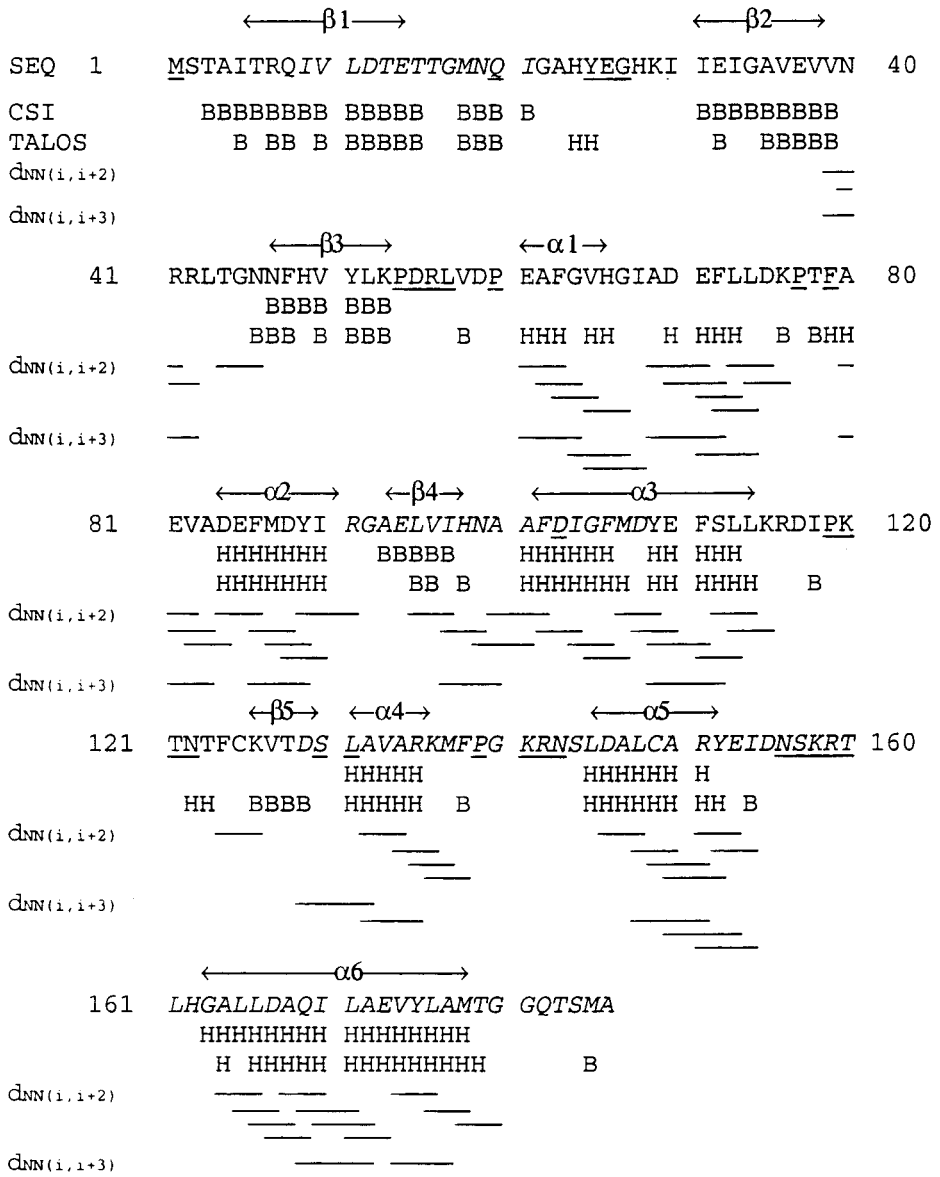

FIGURE 2: Summary of the secondary structure of ϵ186 as predicted by CSI and TALOS analyses. Unassigned residues are underlined. The three exo motifs are indicated by italics and correspond to exo motif I: residues 8−22 (QIVLDTETTGMNQIG); motif II: residues 95−108 (LVIHNAAFDIGFMD), and motif III: residues 128−186 (TDSLAVARKMFPGKRNSLDALCARYEIDNSKRTLHGALLDA QILAEVYLAMTGGQTSMA). NOE connectivities are indicated by horizontal lines which begin and end below the residues which they connect.

strands into a five-stranded β sheet. Although the secondary structure predictions indicate that the first β strand extends from T3 through I21, no interstrand NOEs were observed for β1 beyond residues R7 through E14. On the basis of the modeling described in the following section, it appears that the segment from M18 through I21 is characterized by φ and ψ angles which are generally similar to those of an extended β strand, but the segment is not structurally part of the central β sheet of the protein. As is apparent from Figure 3, the first three β strands, contained in the first half of the protein, are organized into an antiparallel sheet. Strand β4 exhibits a set of amide−amide NOEs with β1, indicating a parallel β1:β4 alignment. We did not observe an amide−amide NOE between Ile31 and Lys52 connecting strands 2 and 3, which was predicted by extending the β strands to the length suggested by the CSI analysis. The failure to observe this NOE may result from the location of these residues near the ends of the corresponding β strands. A short segment extending from Lys125 to Asp129 exhibits multiple NOE interactions with β4 and structurally approximates a fifth, parallel β strand (Figure 3). An additional amide−methyl NOE, Leu95 H^N to Val127 γ# (Table 1c), and two methyl−methyl NOEs, Leu95 δ#−Val127 γ# and Ile97 δ1−Val127 γ# (Table 1b), connect β4 with this fifth β strand. An amide−amide NOE between Leu95 and Cys125 appears to connect β4 to a turn preceding the fifth β strand, while an amide−amide NOE between Ile97 and Leu131 connects the end of β4 to the beginning of α4. We note that the CSI analysis does not support the identification of this stretch of residues as a β strand; however, TALOS predicts φ and ψ torsion angles for residues Lys126 through Asp129 that are consistent with a β-strand secondary structure. Therefore, on the basis of the long-range NOEs between residues Lys126 through Asp129 and β4 and the torsion angles from TALOS, we have assigned this stretch of residues as β5 (Figures 2 and 3) in parallel alignment with respect to β4.

Table 1: Long-Range NOEs

(A) Long-Range Amide−Amide NOEs Observed in $^{2}H,^{13}C,^{15}N$-labeled ϵ186

Arg7 HN	Glu94 HN
Gln8 HN	Va38 HN, Val39 HN, Glu94 HN
Ile9 HN	Glu94 HN, Leu95 HN, Val96 HN
Val10 HN	Val36 HN, Val96 HN
Leu11 HN	Val96 HN, Ile97 HN
Asp12 HN	Gly34 HN
Glu14 HN	Ile31 HN, Glu32 HN
Thr16 HN	Lys29 HN
Ile33 HN	Val50 HN
Ala35 HN	Phe48 HN
Ala93 HN	Lys126 HN
Glu94 HN	Lys126 HN
Leu95 HN	Cys125 HN, Lys126 HN, Val127 HN, Thr128 HN
Val96 HN	Thr128 HN
Ile97 HN	Thr128 HN, Asp129 HN, Leu131 HN

(B) Long-Range Methyl−Methyl NOEs Observed in Selectively Methyl-Protonated ϵ186[a]

Ile5 δ1	Val39 γ#
Val10 γ#	Val38 γ#, Val96 γ#
Leu11 δ#	Leu95 δ#, Ile97 δ1
Ile31 δ1	Val58 γ#, Ile68 δ1, Leu73 δ#
Ile33 δ1	Val50 γ#, Leu52 δ#
Val36 γ#	Leu43 δ#, Leu165 δ#
Val38 γ#	Leu176 δ#
Leu43 δ#	Leu176 δ#
Val50 γ#	Val82 γ#
Leu52 δ#	Val82 γ#
Val58 γ#	Leu73 δ#
Leu95 δ#	Ile104 δ1, Val127 γ#
Ile97 δ1	Ile104 δ1, Val127 γ#
Ile104 δ1	Val127 γ#
Leu114 δ#	Ile118 δ1
Leu145 δ#	Leu148 δ#, Ile170 δ1, Ile171 δ1, Val174 γ#
Leu148 δ#	Ile154 δ1, Val174 γ#
Ile154 δ1	Val174 γ#
Leu161 δ#	Leu165 δ#

(C) Long-Range Amide−Methyl NOEs Observed in Selectively Methyl-Protonated ϵ186

Gln8 HN	Val38 γ#, Val39 γ#,
Ile9 HN	Leu95 δ#, Val96 γ#,
Val10 HN	Val36 γ#, Leu43 δ#
Leu11 HN	Val96 γ#
Ile33 HN	Val50 γ#
Gly34 HN	Leu11 δ#
Ala35 HN	Leu165 δ#
Val36 HN	Val10 γ#, Leu43 δ#
Val38 HN	Ile9 δ1, Val10 γ#, Leu43 δ#
Arg41 HN	Leu176 δ#
Leu43 HN	Val38 γ#, Leu176 δ#
Tyr51HN	Leu73 δ#
Leu52 HN	Leu73 δ#
Ala62 HN	Val58 γ#
Gly64 HN	Val58 γ#
Asp70 HN	Val58 γ#
Ala80 HN	Leu114 δ#
Glu81 HN	Leu52 δ#
Glu94 HN	Ile5 δ1
Leu95 HN	Val127 γ#
Val96 HN	Val10 γ#
Ile97 HN	Val10 γ#, Leu171 δ#
His98 HN	Leu11 δ#, Leu145 δ#
Ile104 HN	Ile97 δ1
Ser112 HN	Ile118 δ1
Lys126 HN	Ile5 δ1
Val127 HN	Ile104 δ1
Asp128 HN	Ile97 δ1, Ile104 δ1
Cys149 HN	Val174 γ#
Asp155 HN	Leu166 δ#
Gln169 HN	Val36 γ#
Leu171 HN	Val10 γ#, Leu145 δ#
Ala172 HN	Val10 γ#, Leu145 δ#
Glu173 HN	Val10 γ#, Leu145 δ#, Ile154 δ1
Val174 HN	Ile154 δ1

[a] (#) Symbol is used to indicate that the prochiral methyl groups have not been assigned.

In addition to the long-range amide−amide NOEs connecting the β strands (Table 1a), long-range NOEs involving the aliphatic methyl groups served to position four of the α helices. Four long-range amide−methyl NOEs (A35 H^N−L165 δ#, Q169 H^N−V36 γ#, L171 H^N−V10 γ#, A172 H^N−V10 γ#, and E173 H^N−V10 γ#) serve to position α6 near the β sheet (Table 1c). Two methyl−methyl NOEs (V36 γ#−L165 δ# and V38 γ#−L176 δ#; Table 1b) also help to orient α6 relative to the β sheet. Three methyl−methyl NOEs (L95 δ#−I104 δ1, I97 δ1−I104 δ1, and I104 δ1−V127 γ#) and two amide−methyl NOEs (I104 H^N−I97 δ1 and V127 H^N−I104 δ1) locate α3 in the vicinity of strands β4 and β5. Four long-range methyl−methyl NOEs (L145 δ#−I170 δ1, L145 δ#−I171 δ1, L145 δ#−V174 γ#, and L148 δ#−V174 γ#) and four amide−methyl NOEs (L149 H^N−V174 γ#, L171 H^N−L145 δ#, A172 H^N−L145 δ#, and E173 H^N−L145 δ#) locate α5 near the center of α6. No long-range NOEs connecting α1, α2, and α4 with other secondary structure elements were assigned.

Despite the rather limited number of long-range NOEs observed, structure calculations were performed using the CNS (version 1.0) program (*45*). The structure calculations started from an extended structure (*46*), using both torsion angle dynamics (*47*) and Cartesian dynamics. The standard CNS protocol was used with the following modifications. The number of steps was increased from 1000 to 4000 in the high-temperature and first slow-cooling annealing stages under torsion-angle dynamics. In addition, the starting temperature was reduced from 2000 to 1000 °C and the number of steps from 3000 to 1500 in the second slow-cooling annealing stage under Cartesian dynamics. A total of 207 amide−amide, 68 methyl−methyl, and 152 amide−methyl NOE distance restraints obtained from these NOESY data were used. In addition, 102 hydrogen-bond restraints implied from the secondary structure analysis and the cross-strand amide−amide NOEs and a total of 232 dihedral-angle restraints obtained from the CSI and TALOS analyses were also included in the calculation. The α-helical hydrogen-bond restraints were 2.4−3.5 Å between the i and $i + 4$ O and N atoms and 1.4−2.5 Å between the O and HN atoms. The cross-strand hydrogen-bond restraints were 2.4−3.5 Å between the O and N atoms and 1.4−2.5 Å between the O and HN atoms of the residues connected by the dotted lines in Figure 3. The NOE-derived distance restraints and the CSI- and TALOS-derived φ and ψ dihedral-angle restraints were set as described by Mueller et al. (*48*). The amide−amide NOE intensities were classified as strong, medium, and weak, corresponding to distance restraints of 1.8−3.5, 1.8−5.0, and 1.8−7.0 Å, respectively, while the amide−methyl and methyl−methyl distances were restrained to 1.8−8.0 and 1.8−8.5 Å. The dihedral angles were taken to be ±2 standard deviations or at least ±20° from the average values predicted by TALOS. CSI predictions were used to generate dihedral-angle restraints for residues that did not meet TALOS acceptance criteria. The dihedral angles were restrained to $\varphi = -70(\pm 50)$ and $\psi = -50(\pm 50)$ and $\varphi = -140(\pm 60)$ and $\psi = -130(\pm 90)$ for α helices and β strands, respectively (*49*). Nine structures of 25 gave no NOE violations >0.5 Å and dihedral-angle violations >5°. These structures superimposed on the average energy-minimized structure (residues 7−178) with an rmsd of 5.3 Å for all of the backbone atoms and 3.7 Å with respect to all of the

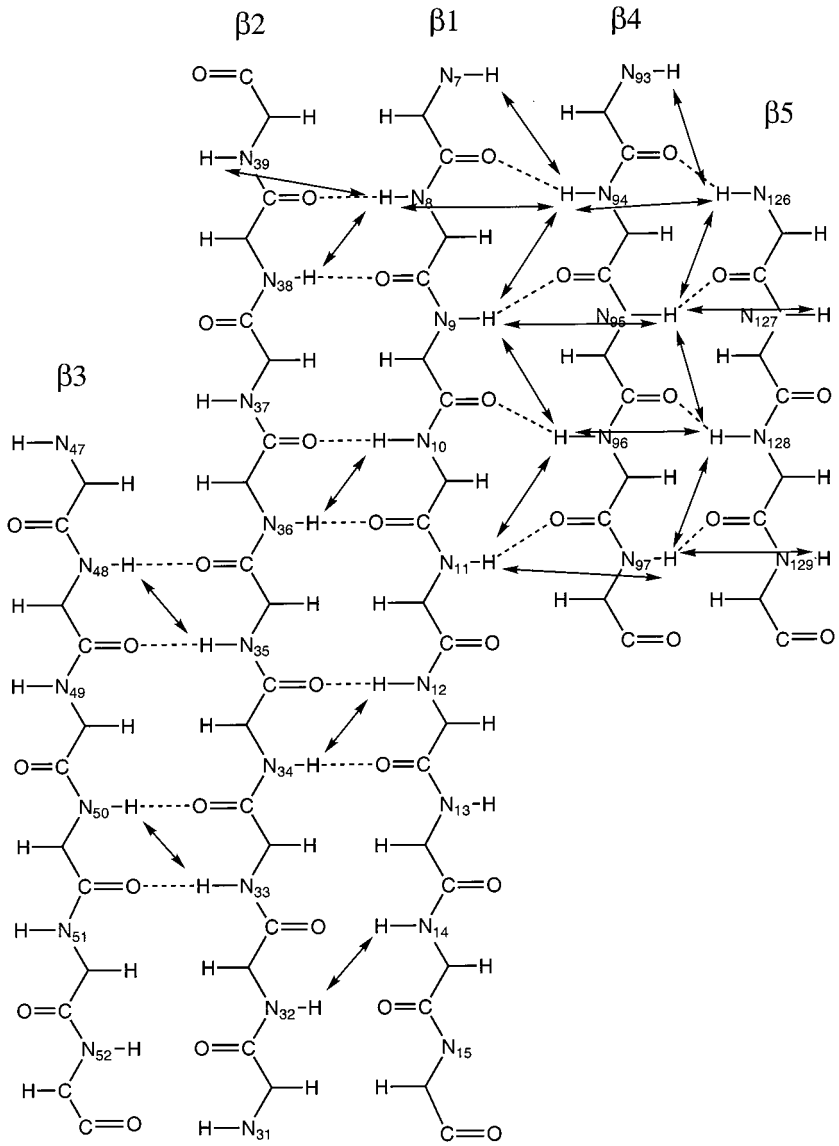

FIGURE 3: Interstrand NOE interactions for the amide protons in the five-stranded β sheet.

secondary structure elements (residues 7−15, 31−39, 47−53, 61−66, 84−90, 93−97, 101−113, 126−130, 131−135, 145−151, and 163−178). The average energy-minimized structure is shown in Figure 4, generated using the program MOLMOL (*50*).

Although the number of experimental restraints was found to be too low to determine the overall fold of the protein, the experimental restraints were sufficient to determine the local fold of the five-stranded β sheet and to localize the α6 helix against one face of the sheet. The rmsd of the five strands and α6 superimposed on the average energy-minimized structure for all of the backbone atoms is 2.1 Å. This local fold serves to position the acidic residues D12, E14, and D167 in close proximity to each other (within 8 Å). These residues, along with D103 in the α3 helix, have been assigned essential catalytic functions (*7, 51*). The number of long-range distance restraints does not localize the α3 helix with the same precision as the α6 helix with respect to the β sheet (i.e., rmsd increases to 2.7 Å when the α3 helix is included), but all five long-range NOEs involving the α3 helix are to the β sheet (see the previous discussion). On the basis of this model, the active site of ϵ186 (Figure 4) is seen to closely resemble the active site of the Klenow exonuclease domain. In particular, both structures (ϵ186 and Klenow) contain four acidic residues: aspartic (D12/D355) and glutamic (E14/E357) residues located on the first β strand and two additional aspartyl residues (D103/D424 and D167/D501) located on α helices positioned next to the β sheet. These residues are known to play a critical catalytic role in the Klenow exonuclease (*10*), as well as in ϵ186 (*7, 51*). Thus, the NMR structure shows that despite the very limited sequence homology, particularly for the Exo III motif, there is substantial structural homology of these exonucleases. This structural homology supports the validity of additional modeling and further defines many structural features which can be used to align the residues of these proteins in addition to sequence homology, as described in the following section.

Molecular Modeling. The NMR data described in the previous section suggest that despite certain distinguishing features of the *E. coli* pol III proofreading system (the separate subunit character of ϵ, the presence of the ϵ-bound θ subunit, and a formally distinct Exo 3 motif), the basic core structure of ϵ is homologous to that of other proofread-

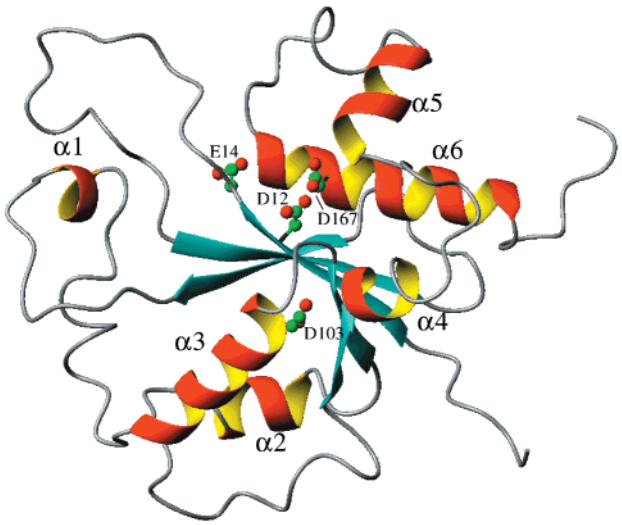

FIGURE 4: Average of the nine structures calculated from the NMR data for ϵ186 using CNS, showing the β sheet, six α helices, and the acidic residues D12, E14, D103, and D167. The structure was calculated using 427 NOE distance restraints, 232 experimental dihedral-angle restraints, and 102 hydrogen-bond restraints. The figure was generated using the program MOLMOL (*50*).

ing 3′ exonucleases, such as those of *E. coli* DNA polymerase I or bacteriophages RB69 and T4. On this premise, we undertook the construction of a molecular model of ϵ186 using homology modeling. Our strategy was to construct a structural template using the known crystal structures of the 3′ → 5′ exonuclease domains of the Klenow fragment (PDB code 1KFS; *11*), T4 DNA polymerase (PDB code 1NOY; *13*), and *E. coli* Exo I (PDB code 1FXX; *21*). Despite the limited sequence identity among them, the three resolved structures are very similar, particularly in the active-site region. In general, in homology modeling, a multiple sequence alignment needs to be constructed of the protein of interest with the structurally defined proteins in order to place the residues of the protein on the structural template. The quality of this sequence alignment is limiting for the accuracy of the model. In the case of ϵ and the 3′ exonucleases, reliable sequence homology is generally limited to the regions of the conserved (catalytic) acidic residues. However, in this study, we additionally use the structural information gained from the NMR experiments, notably the secondary structure elements and NOE restraints, to guide the alignment process. In the following paragraphs, we describe the various steps leading to an alignment of ϵ186 residues with the structurally defined proteins, ultimately permitting the construction of a detailed model for ϵ. This model, which is fully consistent with all of the NMR restraints, resembles in key details the established crystal structures. The model also provides a satisfactory basis by which to rationalize the observed mutator phenotype of many ϵ mutants (*7*). The approach of homology modeling guided by NMR information has been used previously by Kitchen et al. (*19*) to generate a model of Oncostatin M.

An initial sequence alignment was taken from the hidden Markov model alignment of Moser et al. (*9*). This alignment, restricted to the four proteins of interest, is shown in Figure 5. Included in this figure is the alignment of the residues in motifs I−III that include the conserved catalytic residues. Using the α carbons from the aligned residues of these three motifs, we superimposed the exonuclease domains of the Klenow fragment and *E. coli* Exo I onto the exonuclease domain of T4 DNA polymerase. Using this superposition, we manually constructed a multiple structural alignment using the SwissPDBViewer software (*38*). This structural alignment, restricted to regions of defined secondary structure, comprises the first three rows of the alignment in Figure 6A. The proximity of side chains and proximity of α carbons were used as alignment criteria. This structural alignment deviates in several cases from the initial sequence alignment. For example, in Figure 6A, His181 of *E. coli* Exo I is aligned with Tyr497 of the Klenow domain based on the side-chain position, although Moser et al. (*9*) align His181 with Arg496 (Figure 5). The fourth row of Figure 6A represents our modeled alignment of ϵ with this structural alignment, as detailed next.

The structural alignment in the region of strands β1 and β2 (Figure 6A) corresponds well with the sequence alignment in ref *9* (Figure 5). Therefore, as the amide−amide interactions in this region for ϵ186 likewise indicate the presence of the two corresponding β strands, we assumed that the Moser et al. (*9*) alignment in this region was correct. The residues of strand β3 are not included in the alignment in ref *9*, being part of a large unaligned region. Nevertheless, our NMR data are sufficient to accurately determine the backbone structure of ϵ186 within the sheet made up of β strands 1−3. Thus, the ϵ186 residues of strands 1−3 were aligned to the corresponding residues of the three templates by superimposing the α carbons of the ϵ186 residues onto the corresponding residues of T4 on the basis of the Figure 5 alignment in those strands. We also used this NMR structural information to assign the initial positions of ϵ186 in the loop for residues 39−47 between β strands 2 and 3.

Within helix 2, the alignment of Figure 5 does not quite agree with our structural alignment of Figure 6a, being off by a shift of one or two residues. From the NMR constraints of Table 1b, we concluded that Val82 in helix 2 was close to Val50 and Leu52. Because the latter residues reside in strand β3, these constraints were best satisfied by shifting the Moser alignment of ϵ186 by three residues or one helical turn within helix 2 of Figure 6A. This alignment is also more compatible with the NMR secondary structure assignments in Figure 2, which assign a helical structure to residues 84−90 of ϵ186.

The sequence alignment of strand β4 and helix α3 was also modified by the constraints in Table 1b. First, a careful inspection of the local structure of the Klenow fragment revealed that, in comparison to the other template structures, a tighter turn exists between β4 and α3 (the latter containing the conserved aspartate of motif II). This tighter turn is reflected in the one-residue gap in the structural alignment for this region (Figure 6A). In the structural alignment (Figure 6A), the conserved aspartate from motif II is followed by an α helix. From Figure 2, we see that the residues following Asp103 of ϵ186 are also helical. Thus, we aligned Asp103 of ϵ with the conserved aspartate in motif II and aligned ϵ residues 104−109 with the subsequent residues of the structural alignment (helix α3). According to the NOEs of Table 1b, the side chains of ϵ residues Leu95 and Ile97 in β4 are close to that of Ile104 in α3. These constraints are best satisfied by using the alignment of ϵ186 to the Klenow exonuclease in β4 (Figure 6A) but not to the T4 exonuclease or to Exo I, because the latter would force the side chains

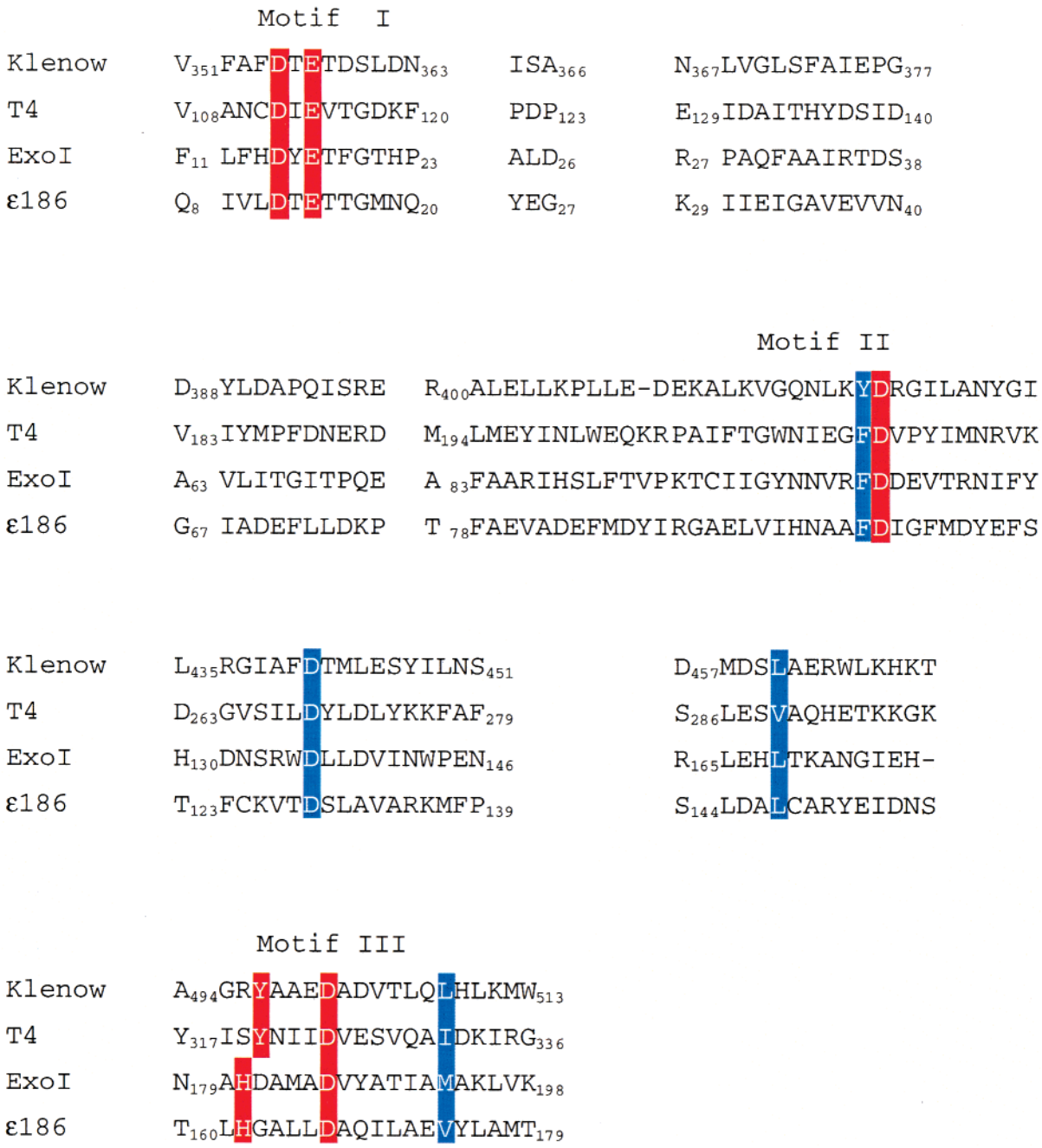

FIGURE 5: Hidden Markov alignment of Moser et al. (*9*) restricted to the exonuclease domains of Klenow (*11*), T4 (*13*), *E. coli* Exonuclease I (*21*), and ϵ186. Color code: (red) conserved, active site residues; (blue) highly conserved residues identified by Moser et al. (*9*).

of residues 95 and 97 to point away from 104 rather than toward it. Interestingly, in the alignment of Figure 6A asparagine at position 99 of ϵ186 is conserved among all four proteins, whereas this is not evident from the alignment of Moser et al. (*9*; Figure 5). In the structures of the Klenow and T4 exonucleases, which contain single-stranded DNA, these asparagine residues are hydrogen bonded to the DNA.

To align ϵ186 strand β5 and helix α4, we noted that, in the structural alignment of the templates, there is a conserved aspartate (Asp441 in the Klenow fragment) located between a β strand and a short helix. In the Klenow fragment, the side chain of this Asp441 hydrogen bonds to the backbone of residues 419 and 420 of motif II, thereby helping to stabilize the structure of the active site. The side chains of the corresponding aspartates in the other template exonucleases likewise hydrogen bond to the backbone of motif II.

In Figure 2, ϵ is shown to contain a similar region of secondary structure containing Asp129. Asp129 is conserved within the ϵ family (*7, 8*), and its alteration results in a mutator phenotype (*8, 51*). Thus, we aligned Asp129 to the indicated conserved aspartate of the structural alignment and aligned the following helical residues sequentially (α4). To align the preceding strand (β5), we noted that, in the structural alignment (Figure 6A), the corresponding strand of the Klenow exonuclease contains a one-residue insertion. We used the NOE information from Table 1b to determine that the side chain of ϵ residue Val127 was directed toward Ile104. This indicated that, proceeding backward from Asp129, ϵ186 should be aligned in this region with Exo I and T4 exonuclease rather than with the Klenow exonuclease.

The alignment of helix α5 of ϵ186 with the templates was based on secondary structure information from Figure 2 and

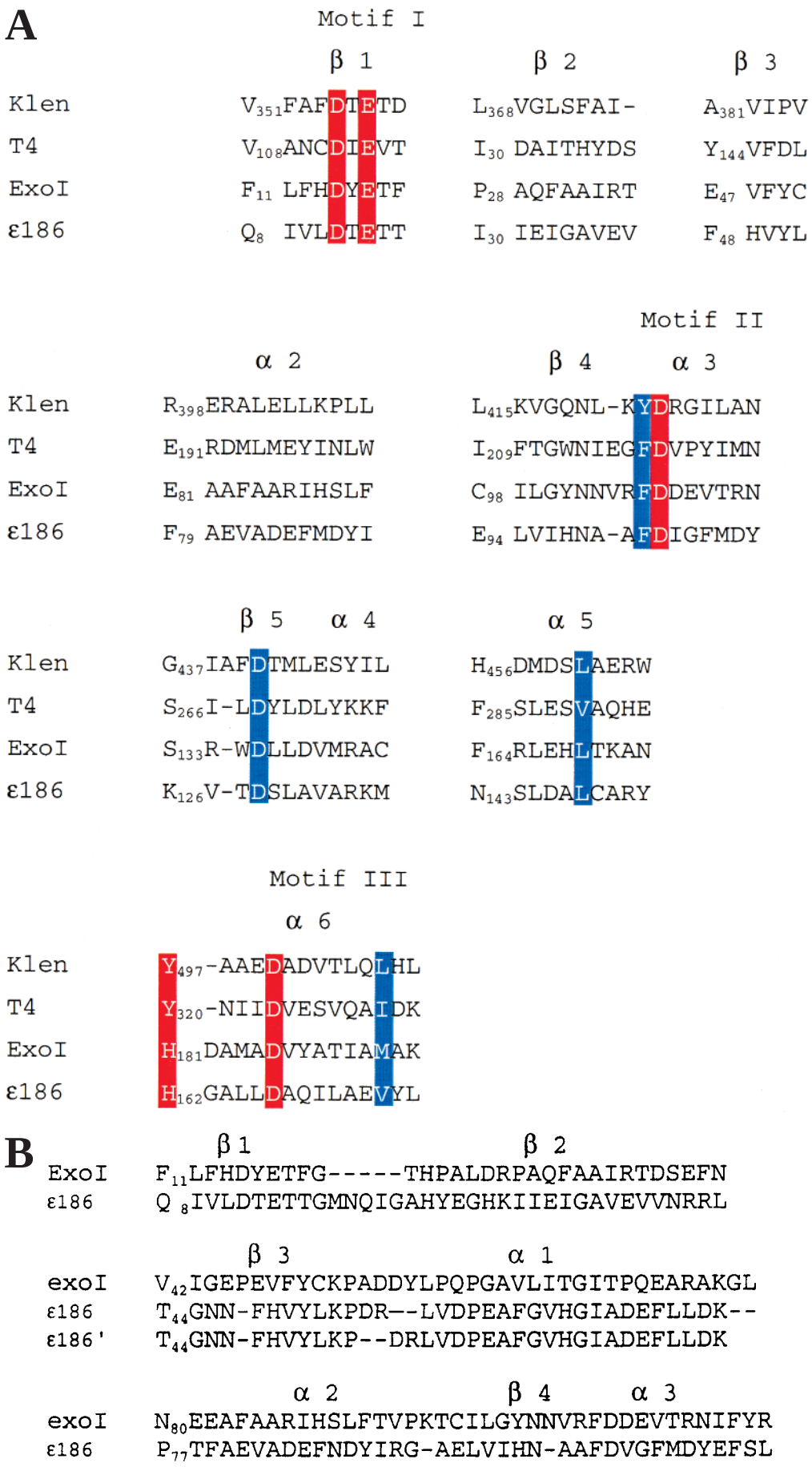

FIGURE 6: (A) Consensus alignment of the ϵ186 to the conserved core of the Klenow exonuclease domain, T4 DNA polymerase, and *E. coli* Exo I. (B) Alternate alignment of ϵ186 with Exonuclease I. Color coded as in Figure 5.

was found to be consistent with NMR constraints involving Leu145 and Leu148 from Table 1b. The alignment here agrees with that of Figure 5. With regard to helix α6, we note that, in the structural alignment, the conserved aspartate from motif III is within an α helix. The NMR secondary structure predictions indicate that this is true as well for the corresponding Asp167 of ϵ (Figure 2). We, thus, aligned Asp167 with the conserved aspartate from motif III and aligned the following ϵ residues 168−179 with the subsequent residues in the structural alignment. We note that within the extended exonuclease family a dichotomy exists with regard to the conserved residues of motif III. Most members, including the Klenow and T4 exonucleases, contain the sequence YxxxD, where the ϵ subfamily, which includes ϵ and Exo I, contains the sequence HxxxxD (HxAxxD; *7*−*9*). We, thus, aligned ϵ His162 with His181 of Exo I. The positions assigned to Ile170, Ile171, and Val174 of ϵ186 are fully consistent with the NMR constraints from Table 1b.

To complete the homology model, we needed to assign backbone and side-chain positions for residues 1−7, 17−29, 39−47, 53−78, 91−93, 110−125, 138−142, 153−161, and 180−186 of ϵ186. These regions, comprised of loops connecting the indicated structural elements as well as one helical region (α1), reside in regions where the three template structures diverge and a structural alignment cannot be constructed. With the exception of segments 17−29 and 53−78, the assignments of these segments were relatively straightforward. Residues 1−7 were modeled using NOE constraints from parts A and B of Table 1 along with dihedral restraints derived from the TALOS analysis. Residues 91−93 and 138−142 were assembled manually, using preliminary database searches for compatible loop conformations. Residues 180−186 were assembled using secondary structural information from Figure 2 together with dihedral restraints. Residues 39−47, 110−125, and 153−161 were modeled using the NOE constraints from parts A and B of Table 1.

From Figure 6A, we see that residues 17−29 of ϵ186 correspond to residues 360−367 of the Klenow fragment, to 117−129 of T4 polymerase, and to 20−27 of Exo I. Residues 360−367 of the Klenow fragment align structurally with residues 20−27 of Exo I. In contrast, residues 117−129 of the T4 polymerase diverge structurally from Klenow and Exo I; however, residues 120−122 of T4 align structurally with residues 361−363 of the Klenow fragment (and with 21−23 of Exo I). Note that the structures of the Klenow fragment and T4 polymerase have been solved with single-stranded DNA in the exonuclease active site. In these structures, Leu361 of the Klenow fragment and Phe120 of T4 polymerase contact the DNA bases on either side of the scissile phosphate bond. Unfortunately, we identified no long-range NOEs involving residues 17−29 of ϵ186, with the exception of an amide−amide NOE involving 16 and 29. However, the secondary structure information (Figure 2) involving these residues is more easily accommodated by a backbone structure close to that of residues 117−129 of T4 polymerase. We, therefore, modeled the backbone of residues 17−29 of ϵ186 on the latter, constraining Ile21 to align with Phe120, because it seemed the most plausible candidate to fit the role of Phe120 of T4 polymerase. A role for Ile21 in the interaction with the DNA is supported by the lack of observed long-range NOEs for the Ile21 methyl group with other aliphatic methyls.

From Figure 6A, we see that residues 53−78 of ϵ186 correspond to residues 386−397 of the Klenow fragment, to 149−190 of T4 polymerase, and to 52−80 of Exo I. The three template structures diverge structurally in these regions. Note that the segment 386−397 of the Klenow fragment contains no regions of defined secondary structure, whereas 149−190 of T4 polymerase contains an α helix for 178−182 followed by a β strand for 184−188. Additionally, the segment 52−80 of Exo I contains an α helix for 61−67 followed by another helix for 71−77. The secondary structure for residues 53−78 of ϵ186 (i.e., an α helix for 61−65 followed by a weakly helical region near 70−74; Figure 2) is most compatible with that of the corresponding residues in Exo I. Furthermore, using a multiple alignment approach, we obtained a reasonable alignment of 53−78 of ϵ186 with 52−80 of Exo I. To align these, we first used the Blast program (*37*) with the ϵ186 sequence as a query, resulting in a set of 63 homologous proteins. These were multiply aligned using the ClustalW program (*20*). Next, the sequence of *E. coli* Exo I was used as a query for Blast, resulting in five homologous proteins. These six sequences were multiply aligned using ClustalW. The two multiple alignments were then aligned with each other, using the profile alignment option of ClustalW. The resulting alignment, restricted to Exo I and ϵ186 and including only residues 8−113 of ϵ186, is shown in Figure 6B. This alignment agrees reasonably well with the alignment of Figure 6A in the regions where they overlap. The agreement deteriorates after residue 113 of ϵ186. Note that, in the middle row of Figure 6B, we provide an alternate alignment involving a small rearrangement of a gap. The latter aligns Arg56 with Tyr57 of Exo I, which interacts with Phe19 and which aligns with Thr16 of ϵ186. That is, this latter alignment would result in an interaction between Arg56 and Thr16 of ϵ186. Note that both residues, when mutated, yield strong mutator phenotypes (*7*).

After backbone positions using the multiple alignment and the three template positions were assigned, side-chain positions were assigned individually using the Swiss PDB View program (*38*). This allows the choice of rotamers for a given residue. The process was iterated manually to avoid steric conflicts and to minimize NOE violations. Following this, the trinucleotide, the two metal ions, and the three water molecules coordinating metal ion B in the Klenow fragment structure were introduced first by rigidly superimposing the ϵ186 structure onto that of the Klenow fragment using α carbons from the consensus alignment and then by copying the nucleic acid, metal, and water coordinates from the latter. The side chains of neighboring ϵ186 residues were adjusted to optimize interactions with the DNA, metal ions, and waters. Following this, the model was prepared for molecular dynamics refinement using the LEaP module of the AMBER package (*39*). The model was immersed in a box of water (TIP3P model) such that a layer at least 10 Å thick surrounded the protein. Five sodium atoms were added to neutralize the system. The final system size was approximately 33 000 atoms. Refinement was accomplished using the Sander module of the AMBER package. The force field of Cornell et al. (*40*) was applied, periodic boundary conditions were employed, and electrostatics were handled by the smooth PME method (*41*) using an 8 Å cutoff in direct

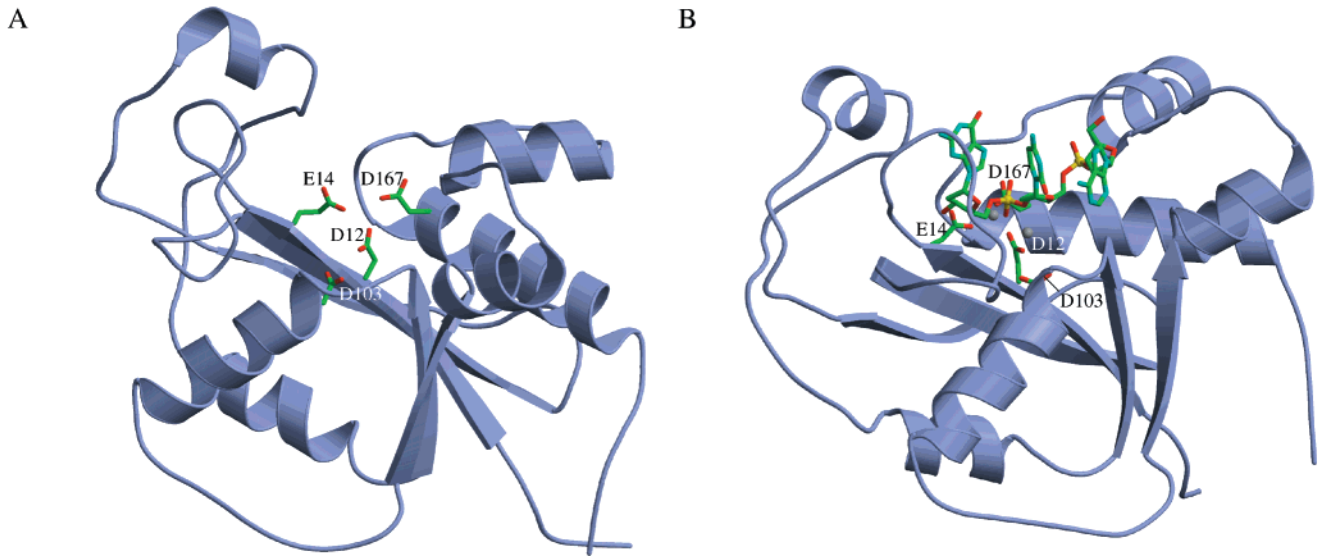

FIGURE 7: (A) Model of the ϵ exonuclease domain, ϵ186, based on the NMR data and the alignment shown in Figure 6. (B) Model of the complex formed between ϵ186 and an A−C−G trinucleotide.

space, a 1 Å grid density, and cubic interpolation. To limit the distortion of the active site during refinement, the van der Waals parameters for Mg^{2+} and Zn^{2+} were modified. We used a value of 1.3 Å for the van der Waals radius parameter of both metals and 0.1 and 0.04 kcal/mol for the van der Waals well-depth parameters for Mg^{2+} and Zn^{2+}, respectively. These values are not necessarily optimal in other circumstances and should be seen as attempts to compensate for the poor modeling of divalent-metal interactions in current empirical force fields. The simulations were performed under conditions of constant volume and temperature (300 K) using a coupling constant of 1 ps. The time step was 2 fs. Initially, the protein atom positions were frozen while the water and ions were equilibrated. The unit cell density was adjusted to 1 g/mL by slowly changing the box dimensions during the equilibration phase. Following this, a series of short 10 ps simulations were performed. During these simulations, the protein, DNA, and metal ions were constrained to their positions in the initial model, with the strength of the constraint on a particular residue dependent on the confidence in the model at that position but, in any case, decreasing for each simulation. The DNA and metal-ion positions were also constrained, strongly at first, to allow the protein to accommodate to the DNA conformation. During these simulations, distance restraints derived from NOE information from parts A−C of Table 1 were also applied. After 100 ps, the positional restraints on the protein were removed. Refinement using only weak NOE distance restraints and weak constraints on DNA and metal-ion positions was performed for an additional 10 ps, followed by 100 ps of unconstrained molecular dynamics.

Structure of ϵ186. The 3D structure of ϵ186 derived from the NMR data and refined by the molecular modeling is shown in Figure 7A. The active site of the enzyme is seen to contain four carboxylate groups contributed by Asp12, Glu14, Asp103, and Asp167, which presumably complex two divalent metal ions. By analogy with the structure of the Klenow exonuclease domain (*10*), the ion in site A is coordinated to three carboxylate oxygen atoms contributed by D12, E14, and D167 and, in the presence of substrate, to one of the nonbridging oxygen atoms of the phosphodiester (Figure 7B). The ion at site B is coordinated to the second oxygen atom of D12, to three waters, and, in the presence of substrate, to the same nonbridging phosphate oxygen atom as metal A, as well as to the 3′ bridging oxygen. Two of the three waters bound to metal B are hydrogen bonded to the side chain of D103, as well as to the backbone carbonyls of T13 and N99, whereas the third water hydrogen bonds to the side chain of H98. His98 also hydrogen bonds to a nonbridging oxygen of the penultimate phosphate group. The Klenow exonuclease was found to contain Zn^{2+} in site A and Mg^{2+} in site B, when crystallized in the presence of both ions. Use of the Amber force field for a Zn^{2+} ion placed in site A and a Mg^{2+} ion placed in site B resulted in minor distortion of this structure, with both carboxylate oxygens of E14 and D167 attempting to coordinate to Zn^{2+}. These distortions are likely due to deficiencies in the treatment of divalent metals by empirical force fields such as Amber. On the basis of structural analogy with the Klenow exonuclease (*10*), metal-ion A stabilizes a hydroxyl ion for nucleophilic attack on the phosphorus, while metal-ion B can stabilize the pentavalent phosphorus in the transition state and stabilize the oxyanion leaving group. As discussed above, the active site of the Klenow exonuclease also contains a tyrosine residue four residues prior to the aspartate in the Exo III helix (YxxxD), which interacts with one of the phosphate oxygens as well as with the hydroxyl ion ligated to metal-ion A (*10*). ϵ and Exo I lack a tyrosine in this region but instead contain a conserved histidine located five residues prior to D167 (HxAxxD), which in our model could fulfill a similar role.

The modeled structure shows extensive interactions between the side chains in the active site and the A−C−G trinucleotide present in the active site (Figures 7B and 8). The sequential bases of the trinucleotide are neither stacked nor coplanar. The cytosine and guanine bases stack against Ile21 in a manner analogous to interactions with Leu361 in the Klenow fragment, while the cytosine deoxyribose appears to stack against the phenyl side chain of the highly conserved Phe102 residue (Figure 6), similar to the role of Tyr423 in the Klenow fragment. The adenine base at the −2 position does not make any stable contacts with residues of ϵ186,

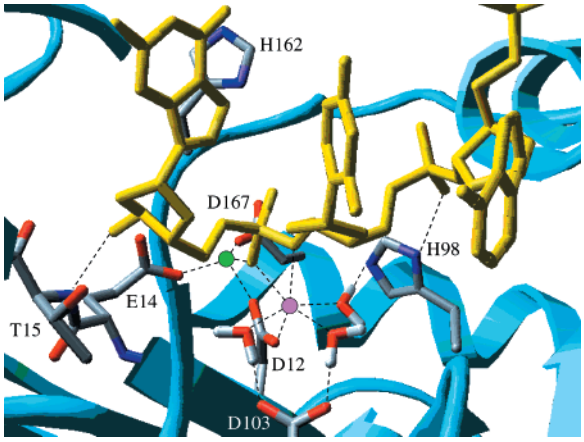

FIGURE 8: Expanded view of the active-site structure illustrating some of the interactions between the nucleotide substrate (yellow) and the enzyme (blue). Side chain and water oxygens are shown in red. Metal-ion A, shown in green, is coordinated to D12, E14, D167, and a nonbridging oxygen atom of the substrate. Metal-ion B, shown in magenta, is coordinated to the second oxygen of D12, to the same nonbridging oxygen as ion A as well as a bridging oxygen, and to three water molecules. Two of the water molecules are within hydrogen-bonding range of D103, but this residue does not coordinate directly to metal-ion B. Also shown are Thr15, His68, and His162. The NH proton of Thr15 is within hydrogen-bonding range of O3′ of the terminal deoxyguanidine and His68. The latter, previously identified as a mutator residue, can H bond with one of the water molecules coordinated to metal-ion B and with a nonbridging phosphate oxygen at the −1 position. Other side chains, as well as the section of the ribbon corresponding to Ile21 and Gly22, have been omitted for clarity.

although during the simulations it transiently hydrogen bonds to the side chain of Asn99. The glycosidic bond of the terminal guanosine is in the anti range ($\chi = -166.9$), the cytosine glycosidic bond ($\chi = -85.0$) is in the high anti range, and the glycosidic bond of the adenosine is also in the anti range ($\chi = -130.4$). In the Klenow fragment structure, the corresponding glycosidic angles are −151.4, −98.3, and +159.4, respectively. The 3′-hydroxyl group of the guanosine deoxyribose is hydrogen bonded to the Thr15 NH group; the side chain of His66 may also interact with this group, although this interaction is not well stabilized in the model. The side chain of His162 is positioned similarly to the side chain of Tyr497 in the Klenow fragment. As noted previously, His162 may also play a role in catalysis, perhaps by positioning a water molecule or assisting in the conversion of this water to the hydroxyl nucleophile. The amide side chain of Asn99 is within hydrogen-bonding distance of the O4′ of the deoxyribose at the −1 position as well as either the cytosine C-2 carbonyl or N3 of the adenine, and the backbone amide of this residue is within hydrogen-bonding distance of the Asp129 side chain. The latter residue is highly conserved and has been shown to yield a mutator phenotype when mutated (*7*). Because it is not in direct contact with the nucleotide substrate, its role appears to be more structural. In addition to Asn99, the side-chain carboxyl oxygens of Asp129 are within hydrogen-bonding distance of the amide groups of Ala100 and Leu131. The penultimate phosphate is hydrogen bonded to His98 and also contacts the side chain of Ser144. As described in the Discussion, the structure presented here provides an intuitive explanation for most of the residue substitutions that result in a mutator phenotype.

DISCUSSION

The ϵ subunit of *E. coli* DNA polymerase III is a separate protein which contains the proofreading 3′ → 5′ exonuclease activity of the polymerase. As such, it differs from its functional analogues, which in most other cases have been shown to be domains of the polymerase protein itself. Although the *E. coli* system is one of the most extensively studied model systems for understanding the physicochemical factors that determine replication fidelity, no structural information has been available for the polymerase or the proofreader subunit. We recently found that ϵ consists of two domains: an N-terminal catalytic domain, ϵ186, and a C-terminal domain, which is required for binding to the α (polymerase) subunit (*5*, *6*). Interestingly, the domain structure of ϵ had previously been predicted by Wootton and Drummond (*44*) based on the identification of a "Q-linker" motif near residue 200. This prediction was verified by the susceptibility of the protein to limited proteolysis with proteinase K (*6*).

Although ϵ itself proved to be an unsatisfactory target for structural analysis, ϵ186 was found to be better behaved. However, this protein also proved to be too unstable for the lengthy NMR experiments that are required to determine the solution structure of a protein of this size. Ultimately, the use of a concentrated Tris buffer and low-temperature measurements (20 °C) combined with [^{2}H,^{13}C,^{15}N] triple labeling and selective methyl protonation were required to obtain analyzable data. These NMR data indicate the presence of six α helices and five β strands. Structurally, most of the β strands are concentrated in the N-terminal portion of ϵ186 and are organized into a single, five-stranded β sheet. A majority of the helical content is located in the C-terminal portion of ϵ186. Despite the absence of sequence identity in the Exo III motif, the model derived here indicates that significant structural homology exists with other exonuclease domains and with the Exo I enzyme. In particular, for all of these systems, the Exo III motif contains a long helical segment that includes an aspartate residue involved in binding one of the metal ions at the active site.

To determine the 3D structure of ϵ186, we initially followed multiple strategies, including the utilization of NMR spectroscopy and homology modeling. At this point, neither strategy independently has yielded a high-resolution structure. The homology modeling effort was limited by the low homology with exonucleases of known structure, particularly for the Exo III motif, while the poor behavior of ϵ186 in solution thwarted attempts to assign the side-chain resonances in the double-labeled [U-^{13}C,^{15}N] protein. However, a combination of approaches has provided a reasonably detailed model for the structure. As noted in the text, this is neither a modeled structure refined by NMR data nor an NMR structure refined by modeling but rather an integrated approach in which many aspects of the alignment were derived from NMR constraints and long-range NOEs were used to select from alternative structures. For example, a structural alignment of the β-sheet regions in ϵ186 with the other exonucleases becomes possible using the β-sheet residues identified from the NMR studies (Figure 3). Using fairly weak NOE distance restraints, there were no NOE violations in the final modeled structure. Further, several key aliphatic residues, such as Ile21 and Val26 which show no

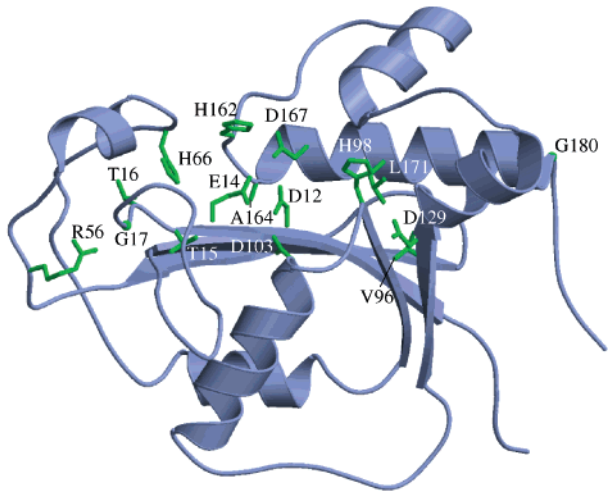

FIGURE 9: Structure of the ϵ186 with the residues previously identified by Taft-Benz and Schaaper (*7*) to yield a mutator phenotype when mutated, drawn using a CPK color scheme. The residues derived from Table 2 of this reference are D12, E14, T15, T16, G17, R56, H66, V96, H98, D103, H162, A164, D167, L171, and G180.

methyl NOEs to other aliphatic residues, are seen be involved in binding to the DNA on the basis of the modeled structure.

The ϵ186 residues previously identified to yield a mutator phenotype when mutated (D12, E14, T15, T16, G17, R56, H66, V96, H98, D103, D129, H162, A164, D167, L171, and G180; *7*, *51*) are indicated in green in Figure 9. Each of these residues must, therefore, perform a critical function in the exonuclease. As is apparent from the figure, nearly all of these residues are located in or near the region of the protein involved in DNA binding. This group includes five acidic residues: D12, E14, D103, D129, and D167. As discussed previously, four of these are involved in binding the divalent cations at the active site, while Asp129 plays a structural role; its side-chain carboxyl groups are within hydrogen-bonding distance of the amide groups of Asn99, Ala100, and Leu131. Thr15, Thr16, and Gly17 are located near the DNA-binding site. The backbone amide of Thr15 is within hydrogen-bonding range of the 3′-hydroxyl group of guanosine deoxyribose. The sole arginine identified, Arg56, is positioned near the above group of residues so that the guanidino side chain can hydrogen bond to the backbone carbonyl and side chain of T16, providing a measure of structural rigidity to this portion of the protein. Three histidine residues are included in this group: His66, His98, and His162. In our model, the His66 side chain appears to interact either with the base or deoxyribose of the terminal guanosine. The substitution H66Y, which produces a mutator phenotype, would result in steric conflict between tyrosine and the terminal guanosine, consistent with reduced exonuclease efficiency. As noted previously, His98 probably contributes to DNA binding through interactions with the penultimate phosphate group as well as with one of the waters coordinating metal-ion B. The H98Y residue substitution, which results in a mutator phenotype, would disrupt the octahedral coordination of metal-ion B and could not hydrogen bond to the penultimate phosphate group, resulting in lower substrate affinity. The side chain of His162 in our model superposes well with His181 of Exo I and Tyr497 in the Klenow fragment. Tyr497 was proposed to play an important catalytic role in contributing to the hydrogen-bonding network that activates and positions the hydroxyl ion bound to metal-ion A for attack on the scissile phosphate (*10*). Presumably, a similar role can be ascribed to His162. However, because of its increased size and the difference in backbone position between residues 162 in ϵ and 497 in Klenow, a tyrosine at 162 would not be able to perform such a role. This is consistent with the mutator phenotype of H162Y (*7*).

Three hydrophobic residues were also identified to yield mutator phenotypes when altered: Val96, Ala164, and Leu171. These residues are not located as close to the DNA and presumably play structural roles. Val96 contributes to a hydrophobic core which also includes Val10, Tyr175, and Leu171. Ala164 is positioned at the beginning of α6, corresponding to Ala499 in Klenow, Ile322 in T4, and Ala183 in Exo I (Figure 6A). In the ϵ model, replacing Ala164 with a residue containing a bulkier side-chain group would disrupt many of the favorable hydrophobic interactions between α6 and the β sheet. Finally, Gly180, the mutator residue most remote from the DNA-binding site, is seen to be located at position 2 of a β turn involving Thr179, Gly180, Gly181, and Gln182. Possibly, the mutator effect which was observed for the G180D and G180N analogues could result from destabilization of this element of the secondary structure. Alternatively, the effect may result from a disruption of the interaction between ϵ186 and one of the other polymerase core proteins. Work currently in progress on the ϵ186−θ complex may help to resolve this question.

In contrast with the previously mentioned results, a number of ϵ mutants have been studied which appeared to have minimal or no significant effects on replication fidelity. These mutants included the series S2A, S112A, S130A, S144A, S157A, and S184A in which each of the serine residues was replaced by alanine (R.M.S., unpublished data) and the series of histidine and aspartate mutants studied by Slupska et al. (*51*) H24G, H28G, H49G, D55G, D59G, D75G, D84G, D88G, D108G, D117G, D146G, and D155G. Analysis of the model presented here indicates that, in all but three cases, the residues are located on the surface of the protein relatively far from the active site. The exceptions are Ser144, Ser130, and Asp146. Ser130 appears to play a structural role, with a side-chain hydroxyl group positioned within hydrogen-bonding distance of the Ile97 carbonyl. The S130A mutation might, thus, be predicted to lead to a small destabilization of the protein but not to dramatically affect fidelity. Ser144 interacts with the penultimate phosphate group. However, its interaction is through the β carbon, so no significant change in affinity would be predicted for the S144A mutant. Asp146 is actually a very weak mutator, decreasing the fidelity by about 4.5-fold in the assay of Slupska et al. (*51*). One of the side-chain oxygen atoms is within hydrogen-bonding distance of the hydroxyl group of Thr160. The second is about 4 Å from one of the oxygen atoms on the cytidyl phosphate. Possibly, this residue could facilitate DNA binding via an intermediate water molecule. Interestingly, the sequence alignment (Figure 6) shows either an aspartate or glutamate residue for all four nucleases.

These studies have demonstrated the feasibility of an integrated NMR and modeling approach in a situation for which either method by itself is insufficient to produce a high-resolution structure. The resulting structure of ϵ186 is consistent with a wide range of biochemical and mutagenesis

data. Despite the complete absence of sequence homology of the Exo III motif of ϵ186 with the related exonucleases, the α-helical secondary structure was found to be similar among these enzymes. More generally, despite the low-sequence homology with the exonucleases for which structural data are available, the general structure of the active-site and metal-ion interactions appears to be well conserved, strongly implying that the hydrolytic mechanism is similarly conserved. The present structure provides a basis for the further investigation of the roles of specific residues in determining the activity of the exonuclease and for the design of studies to evaluate the roles of these residues in replication fidelity.

ACKNOWLEDGMENT

The authors gratefully acknowledge the research group of Dr. Lewis Kay, University of Toronto, and particularly Dr. Ranjith Muhandirum for providing the NMR pulse sequences and advice on their implementation. We also acknowledge Dr. Ronald Venters for providing the 4D $^{15}N/^{15}N$-edited NOESY pulse sequence.

REFERENCES

1. Kelman, Z., and O'Donnell, M. DNA polymerase III holoenzyme: Structure and Function of a chromosomal replicating machine. (1995) *Annu. Rev. Biochem. 64*, 171−200.
2. Schaaper, R. M. Base selection, proofreading, and mismatch repair during DNA-replication in *Escherichia coli*. (1993) *J. Biol. Chem. 268*, 23762−23765.
3. Fijalkowska, I. J., and Schaaper, R. M. Mutants in the Exo I motif of *Escherichia coli dnaQ*: Defective proofreading and inviability due to error catastrophe. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 2856−2861.
4. Lancy, E. D., Lifsics, M. R., Kehres, D. G., and Maurer, R. Isolation and characterization of mutants with deletions in *dnaQ*, the gene for the editing subunit of DNA Polymerase III in *Salmonella typhimurium*. (1989) *J. Bacteriol. 171*, 5572−5580.
5. Taft-Benz, S. A., and Schaaper, R. M. The C-terminal domain of DnaQ contains the polymerase binding site. (1999) *J. Bacteriol. 181*, 2963−2965.
6. Perrino, F. W., Harvey, S., and McNeill, S. M. Two functional domains of the epsilon subunit of DNA polymerase III. (1999) *Biochemistry 38*, 16001−16009.
7. Taft-Benz, S. A., and Schaaper R. M. Mutational analysis of the 3′−5′ proofreading exonuclease of *Escherichia coli* DNA polymerase III. (1998) *Nucleic Acids Res. 26*, 4005−4011.
8. Barnes, M. H., Spacciapoli, P., Li, D. H., and Brown, N. C. The 3′−5′ exonuclease site of DNA Polymerase III from Gram-positive bacteria−Definition of a novel motif structure. (1995) *Nucleic Acids Res. 165*, 45−50.
9. Moser, M. J., Holley, W. R., Chatterjee, A., and Mian, I. S. The proofreading domain of *Escherichia coli* DNA polymerase I and other DNA and/or RNA exonuclease domains. (1997) *Nucleic Acids Res. 25*, 5110−5118.
10. Beese, L. S., and Steitz, T. A. Structural basis for the 3′−5′ exonuclease activity of *Escherichia coli* DNA polymerase I−A 2 metal ion mechanism. (1991) *EMBO J. 10*, 25−33.
11. Brautigam, C. A., and Steitz, T. A. Structural principles for the inhibition of the 3′−5′ exonuclease activity of *Escherichia coli* DNA polymerase I by phosphorothioates. (1998) *J. Mol. Biol. 277*, 363−377.
12. Brautigam, C. A., Sun, S., Piccirilli, J. A., and Steitz, T. A. Structures of normal single-stranded DNA and deoxyribo-3′-S-phosphorothiolates bound to the 3′−5′ exonucleolytic active site of DNA polymerase I from *Escherichia coli*. (1999) *Biochemistry 38*, 696−704.
13. Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H., and Steitz, T. A. Crystal structures of an NH2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. (1996) *Biochemistry 35*, 8110−8119.
14. Shamoo, Y., and Steitz, T. A. Crystal structure of a replication fork DNA polymerase editing complex at 2.7 Å resolution. (1999) *Cell 99*, 155−166.
15. Hamdan, S., Brown, S. E., Thompson, P. R., Yang, J. Y., Carr, P. D., Ollis, D. L., Otting, G., and Dixon, N. E. Preliminary X-ray Crystallographic and NMR Studies on the Exonuclease Domain of the ϵ Subunit of *Escherichia coli* DNA Polymerase III. (2000) *J. Struct. Biol. 131*, 164−169.
16. Rosen, M. K., Gardner, K. H., Willis, R. C., Parris, W. E., Pawson, T., and Kay, L. E. Selective methyl group protonation of perdeuterated proteins. (1996) *J. Mol. Biol. 263*, 627−636.
17. Gardner, K. H., Rosen, M. K., and Kay. L. E. Global folds of highly deuterated, methyl protonated proteins by multidimensional NMR. (1997) *Biochemistry 36*, 1389−1401.
18. Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C., and Kay, L. E. A robust and cost-effective method for the production of Val, Leu, Ile (d1) methyl-protonated ^{15}N-,^{13}C-2H-labeled proteins. (1999) *J. Biomol. NMR 13*, 369−374.
19. Kitchen, D., Hoffman, R. C., Moy, F. J., and Powers, R. Homology Model for Oncostatin M Based on NMR Structural Data. (1998) *Biochemistry 37*, 10581−10588.
20. Thompson J., Higgins, D., and Gibson, T. ClustalW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. (1994) *Nucleic Acids Res. 22*, 4637−4690.
21. Breyer, W. A., and Matthews, B. W. Structure of *Escherichia coli* Exonuclease I Suggests How Processivity is Achieved. (2000) *Nat. Struct. Biol. 7*, 1125−1128.
22. Yamazaki, T., Lee, Revington, M., Mattiello, D. L., Dahlquist, F. W., Arrowsmith, C. H., and Kay, L. E. An HNCA pulse scheme for the backbone assignment of ^{15}N,^{13}C,^{2}H-labeled proteins: Application to a 37-kDa Trp Repressor-DNA complex. (1994) *J. Am. Chem. Soc. 116*, 6464.
23. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., and Kay, L. E. A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N,^{13}C,^{2}H-labeled proteins with high sensitivity. (1994) *J. Am. Chem. Soc. 116*, 11655−11666.
24. Kay, L. E., Guang, Y. X., and Yamazaki, T. Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation. (1994) *J. Magn. Reson., Ser. A 109*, 129−133.
25. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. (1995) *J. Biomol. NMR 6*, 277−293.
26. Johnson, B. A., and Blevins, R. A. NMRView: A computer program for the visualization and analysis of NMR data. (1994) *J. Biomol. NMR 4*, 603−614.
27. Wishart, D. S., and Sykes, B. D. The ^{13}C Chemical-Shift Index: A simple method for the indentification of protein secondary structure using ^{13}C chemical-shift data. (1994) *J. Biomol. NMR 4*, 171−180.
28. Cornilescu, G., Delaglio, F., and Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. (1999) *J. Biomol. NMR 13*, 289−302.
29. Venters, R. A, Metzler, W. J., Spicer, L. D., Mueller, L., and Farmer, B. T., II. Use of $^{1}HN-^{1}HN$ NOEs to determine protein global folds in perdeuterated proteins. (1995) *J. Am. Chem. Soc. 117*, 9592−9593.
30. Grzesiek, S., Wingfield, P., Stahl, S., Kaufman, J. D., and Bax, A., Four-dimensional ^{15}N-separated NOESY of slowly tumbling perdeuterated ^{15}N-enriched proteins. Application to HIV-1 Nef. (1995) *J. Am. Chem. Soc. 117*, 9594−9595.
31. Kay, L. E., Ikura, M., Zhu, G., and Bax, A. 4-Dimensional heteronuclear triple-resonance NMR of isotopically enriched proteins for sequential assignment of backbone atoms. (1991) *J. Magn. Reson. 91*, 422−428.
32. Gardner, K. H., Konrat, R., Rosen, M. K., and Kay, L. E. An (H)C(CO)NH−TOCSY pulse scheme for sequential assign-

ment of protonated methyl groups in otherwise deuterated ^{15}N,^{13}C-labeled proteins. (1996) *J. Biomol. NMR 8*, 351−356.
33. Zwahlen, C., Gardner, K. H., Siddhartha, P. S., Horita, D. A., Byrd, R. A., and Kay, L. E. An NMR experiment for measuring methyl−methyl NOEs in ^{13}C-labeled proteins with high resolution. (1998) *J. Am. Chem. Soc. 120*, 7617−7625.
34. Kay, L. E., Clore, G. M., Bax, A., and Gronenborn, A. M. 4-Dimensional heteronuclear triple-resonance NMR-spectroscopy of interleukin-1-beta in solution. (1990) *Science 249*, 411−414.
35. Muhandiram, D. R., Guang, Y. X., and Kay, L. E. An enhanced-sensitivity pure absorption gradient 4D ^{15}N,^{13}C-edited NOESY experiment. (1993) *J. Biomol. NMR 3*, 463−470.
36. Marion, D., Ikura, M., and Bax, A. Improved solvent suppression in one-dimensional and two-dimensional NMR spectra by convolution of time-domain data. (1989) *J. Magn. Reson. 84*, 425−430.
37. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman D. J. Basic local alignment search tool. (1990) *J. Mol. Biol. 215*, 403−410.
38. Guex, N., and Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. (1997) *Electrophoresis 18*, 2714−2723.
39. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., III, DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal-mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. (1995) *Comput. Phys. Commun. 91*, 1−41.
40. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. A 2nd generation force-field for the simulation of proteins, nucleic acids, and organic molecules. (1995) *J. Am. Chem. Soc. 117*, 5179−5197.
41. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. A Smooth Particle Mesh Ewald Method. (1995) *J. Chem. Phys. 103*, 8577−8593.
42. Kraulis, P. J. MOLSCRIPT: A Program to produce both detailed and schematic plots of protein structures. (1991) *J. Appl. Crystallogr. 24*, 946−950.
43. Merritt, E. A., and Murph, M. E. P. Raster3D Version 2.0: A Program for photorealistic molecular graphics. (1994) *Acta Crystallogr., Sect. D 50*, 869−873.
44. Wootton, J. C., and Drummond, M. H. The Q-linker—A class of interdomain sequences found in bacterial multidomain regulatory proteins. (1989) *Protein Eng. 2*, 535−543.
45. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. Crystallography & NMR system: A new software suite for macromolecular structure determination. (1998) *Acta Crystallogr., Sect. D 54*, 905−921.
46. Nilges, M., Clore, G. M., and Gronenborn, A. M. Determination of 3-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms—circumventing problems associated with folding. (1988) *FEBS Lett. 239*, 129−136.
47. Stein, E. G., Rice, L. M., and Brünger, A. T. Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. (1997) *J. Magn. Reson. 124*, 154−164.
48. Mueller, G. A., Choy, W. Y., Yang, D. W., Forman-Kay, J. D., Venters, R. A., and Kay, L. E. Global folds of proteins with low densities of NOEs using residual dipolar couplings: Application to the 370-residue maltodextrin-binding protein. (2000) *J. Mol. Biol. 300*, 197−212.
49. Luginbuhl, P., Szyperski, T., and Wuthrich, K. Statistical basis for the use of C-13 alpha chemical shifts in protein structure determination. (1995) *J. Magn. Reson., Ser. B 109*, 229−233.
50. Koradi, R., Billeter, M., and Wuthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. (1996) *J. Mol. Graphics 14*, 51−55.
51. Slupska, M. M., King, A. G., Lu, L. I., Lin, R. H., Mao, E. F., Lackey, C. A., Chiang, J.-H., Baikalov, C., and Miller, J. H. Examination of the role of DNA polymerase proofreading in the mutator effect of miscoding tRNAs. (1998) *J. Bacteriol. 180*, 5712−5717.

BI0114170