A Molecular Model of the Complete Three-Dimensional Structure of the Klenow Fragment of *Escherichia coli* DNA Polymerase I: Binding of the dNTP Substrate and Template-Primer[†,‡,§]

Prem N. S. Yadav,[∥] Janardan S. Yadav,[⊥] and Mukund J. Modak*[,∥]

Department of Biochemistry and Molecular Biology, New Jersey Medical School, and Academic Computing Center, University of Medicine and Dentistry of New Jersey, 185 South Orange Avenue, Newark, New Jersey 07103

Received June 27, 1991; Revised Manuscript Received December 18, 1991

ABSTRACT: A complete three-dimensional structure of the Klenow fragment of *Escherichia coli* DNA polymerase I (pol I) has been proposed on the basis of molecular modeling and molecular mechanics studies using available Cα coordinates. The structure seems quite reliable because the overall surface of electrostatic potentials calculated for the molecularly modeled enzyme closely resembles that reported for the X-ray structure. The modeled structure is then used in developing a ternary complex of dTTP and $(dA)_{25}$–$(dT)_{14}$ poised in its active site. The orientation of both substrates in the ternary complex was primarily guided by the amino acid residues which had been known to interact with dNTP and DNA substrates from earlier studies. The proposed model (a) explains the geometrical and physicochemical relationship of the two substrates with the various critical amino acid residues involved in the binding process and (b) suggests possible roles for additional residues in the binding and/or polymerization reaction. Furthermore, the ternary complex appears to satisfy many biochemical and genetic data concerning catalytic requirements known to exist for the polymerization reaction.

DNA polymerase I of *Escherichia coli* (pol I)[1] is a monomer of approximately 1000 amino acid residues and has a molecular mass of 103 000 Da (Jovin et al., 1969). A multifunctional enzyme, its principal activities may be described as DNA polymerization as well as 3′–5′ and 5′–3′ exonucleolytic degradation activities (Kornberg, 1980). Limited proteolysis of pol I generates a 68 000-Da fragment popularly called the "Klenow fragment" or the large fragment, which retains the polymerase and 3′–5′ exonuclease activities (Brutlag et al., 1969; Klenow & Henningsen, 1970; Joyce & Grindley, 1983). A successful crystallization of the Klenow fragment and resolution of the X-ray crystal structure at 3.3 Å show that the 605 amino acid long polypeptide is folded into two distinct structural domains (Ollis et al., 1985). A combination of crystallographic and molecular genetic data have shown that the smaller amino-terminal domain contains the catalytic center for 3′–5′ exonuclease activity while the large C-terminal region represented by residues 515–928 contains the active site for polymerase reaction (Ollis et al., 1985; Freemont et al., 1986; Derbyshire et al., 1988; Joyce, 1991). The polymerase domain of the protein has been found to contain a large cleft suitable for accommodating a DNA template-primer. Further architectural details of the actual amino acid residues in and around the active site cleft that participate in the binding of the template-primer as well as dNTP substrates are clarified through chemical modification studies using site-specific reagents as well as by site-directed mutagenesis. For example, using pyridoxal phosphate as a dNTP substrate binding site-directed labeling reagent, Basu and Modak (1987) identified Lys-758 as a reactive residue. Pandey et al. (1987), using UV-mediated cross-linking of dTTP to the Klenow fragment, found that His-881 is the site for dNTP cross-linking. Photoaffinity labeling experiments by Rush and Konigsberg (1990) showed that a dNTP analogue, 8-azido-dATP, cross-linked to Tyr-766. In another affinity labeling experiment, Pandey and Modak (1988) found that Arg-682 is the probable site for binding of the β- and/or γ-phosphate group as revealed by the examination of the site of cross-linking of 5′-[(fluorosulfonyl)benzoyl]deoxyadenosine (FSBdA), an analogue of dNTP with an affinity label present at the triphosphate moiety. Therefore, these residues appear to be involved in the binding of dNTP substrates. The site-directed mutagenesis studies which produced mutants with altered DNA polymerization capabilities (Joyce et al., 1985, 1986; Polesky et al., 1990) and additional chemical modification studies (Mohan et al., 1988; Basu et al., 1988) have further shown that the DNA template-primer also interacts within the large cleft. Tyr-766, Arg-841, Asn-845, Glu-849, Arg-668, and Asp-882 are some of the important residues revealed by these studies. Furthermore, DNA footprinting experiments have indicated that the Klenow fragment, when bound at a primer terminus, covers about 8 base pairs of the duplex DNA upstream of the primer terminus (Joyce et al., 1986).

In spite of these advances, molecular mechanisms of template-dependent catalysis of DNA synthesis by pol I remain to be clarified. Furthermore, the nature of the interactions involved in the recognition of DNA and dNTP substrates by the specific amino acid residues is still unknown. Recent advances in computer hardware, molecular modeling, molecular simulation and dynamics, and graphics software in visu-

[†] Supported in part by a grant from the NIH (NIGMS 36307).

[‡] A part of this work was presented at a symposium, "Expanding Frontiers in Polypeptide and Protein Structural Research", held in Whisler, Canada (Yadav et al., 1991).

[§] We dedicate this paper to Dr. Arthur Kornberg on the occasion of his 74th birthday.

* Address correspondence to this author.

[∥] Department of Biochemistry and Molecular Biology.

[⊥] Academic Computing Center.

[1] Abbreviations: pol I, *Escherichia coli* DNA polymerase I; dNTP, deoxynucleoside triphosphate; FSBdA, 5′-[(fluorosulfonyl)benzoyl]-deoxyadenosine.

0006-2960/92/0431-2879$03.00/0

alizing and manipulating molecules on computer monitors prompted us to undertake a model-building exercise which would permit prediction of a complete structure of the Klenow fragment of DNA polymerase I based on Cα coordinates. It could then serve as a starting point to develop a ternary complex of pol I with its DNA and dNTP substrates. The major objective of the present study is to (i) gain insight into the structural, chemical, and enzymatic involvement of the amino acid residues, template-primer, and deoxynucleoside triphosphate in the recognition and polymerization reaction, (ii) provide a working model which will be useful in prediction of important amino acid residues, and (iii) aid in the interpretation of new data that would come from future chemical and/or genetic experiments.

Materials and Methods

(*a*) *Software.* The molecular modeling study was carried out using the SYBYL molecular modeling package (version 5.2) obtained from Tripos Associates Inc. This package is a state-of-the-art program to assist in the description and prediction of molecular behavior. Its database includes a standard fragment library containing hundreds of molecular parts from which novel molecules can be assembled. It can access the crystal structure data directly. In the absence of the appropriate fragment or crystal data, a molecule can be drawn using the SKETCH facility and subsequent access of the CONCORD program for converting the rough drawing into an accurate three-dimensional structure. The software is capable of generating small and/or big molecular structures and performing energy minimization, conformational search, and molecular dynamics calculations and permits vizualization of molecular properties on graphical screen.

The electrostatic field calculations were carried out using Del Phi (version 2.2.0) (Klapper et al., 1986; Gilson, 1988), and the contour map was displayed using Insight II (version 2.2.0) obtained from BIOSYM Technologies Inc.

(*b*) *Hardware.* The HP9000/845 main frame running the UNIX operating system is interfaced with a Silicon Graphics personal IRIS work station. SYBYL, Del Phi, and Insight II softwares are loaded on IRIS, and files are stored in the main frame. Calculations were performed with a personal IRIS, and the interactive graphics was utilized in visualizing molecules. A Tektronix 4693DX printer was directly attached to the personal IRIS to get a hard copy of the figures on the screen.

(*c*) *Data.* Cα coordinates of the Klenow fragment of DNA polymerase I were taken from the Protein Data Bank. Coordinates of the template-primer were generated by SYBYL using its own standard fragment library. The natural and modified nucleotide structures required in this modeling were taken from the crystal structure of B-DNA. The selection of specific amino acids in the polymerase active sites in orienting the dNTP substrates was based on the chemical modification results described by Modak and associates (Pandey et al., 1990).

(*d*) *Computer-Assisted Modeling Protocol.* (*i*) *Structure of the Klenow Fragment.* Cα coordinates of the Klenow fragment of pol I, taken from the Protein Data Bank (Ollis et al., 1985), served as the starting point in developing an atomic model of the enzyme. The complete coordinates of the backbone atoms were generated, and the side chains were added using the SYBYL molecular modeling software (Tripos Associates Inc.). The backbone completion procedure starts from the N-terminal end of the Klenow fragment with a reference fragment consisting of four consecutive residues. The program considers the end-to-end distance of all the fragments present in its own database and saves those fragments whose distances are within a maximum deviation of 0.87 Å from the reference fragment. The retained fragments are further screened by comparing all the inter-Cα distances within the fragments, and the reference fragment and the best 50 fragments matching the individual reference fragment are saved. Finally, the program performs a least-squares fit of each saved fragment onto the reference and selects the ones with the lowest rms deviation (≈0.4 Å). The side-chain generating procedure uses the most commonly occurring (in the database of crystallized proteins) side-chain rotamer for each residue type. The entire structure was scanned to remove bad contacts among the side-chain atoms. This was accomplished by rotating some of the side chains of the residues around Cα and Cβ bonds to reduce local steric hindrances. In the region of local unfavorable contacts, the side chains were oriented to maintain favorable interactions (i.e., appropriate hydrogen-bond and/or van der Waals interactions etc.). For H-bond distance parameters, with A and B as electronegative atoms, the values of A---B $\leq$ 3.4 Å and B---H $\leq$ 2.4 Å and the bond angle A-H---B $\geq$ 120° were used. The distances between nonbonded atoms were those described by Zimmermann et al. (1977). Using the above approach, we first completed the entire structure of bovine pancreatic trypsin inhibitor (BPTI) starting with its Cα coordinates. The computer-built model of the BPTI coincided well (rms deviation 0.36, data not shown) with the Cα backbone and the position of Cβ reported in the crystal structure of BPTI (file name 4pti; Marquardt et al., 1983). These results gave us the confidence to build the structure of the Klenow fragment from the Cα coordinates of pol I (Ollis et al., 1985) in the manner described above. The Klenow structure obtained in this manner was further minimized by the MAXMIN2 procedure using Tripos force field parameters, keeping the positions of Cα fixed. This minimization eliminated some of the bad contacts present in the modeled structure. Figure 1 depicts the stereoview of the complete pol I structure obtained after the above manipulation.

(*ii*) *Electrostatic Potential of pol I.* The charges on the amino acids of the Klenow fragment of pol I were taken from the AMBER force field (Weiner et al., 1984). In the electrostatic calculations, the ionic strength of 0.14 M and the pH value of 7.1 was used. The whole complex was mapped onto a grid size of 65 × 65 × 65. The dielectric constant of 2 was used for the enzyme protein while 80 was used for the solvent. The resulting electrostatic potential contour map is shown in Figure 2.

(*iii*) *Geometry of dNTP and DNA Substrates.* The initial coordinates for deoxynucleoside monophosphate were those found in the consensus structure of B-DNA. The additional pyrophosphate group was added to obtain deoxynucleoside triphosphate (dNTP) using the interactive computer graphics technique following the geometry scheme provided by Saenger (1983). All the structures were then energy refined using the molecular mechanics program MM2 (Bucker & Allinger, 1982), which is best suited for providing good geometry of small molecules. This method also yielded excellent results when comparison of the structure–activity relationship of various anti-HIV-RT nucleoside derivatives was made (Yadav et al., 1990). In addition to MM2, we also calculated the structures of nucleotides with AMBER force fields and found that the minimized structures by both MM2 and AMBER parameters do not deviate significantly from the B-DNA form. Therefore, we used the minimized structure for conformational energy search of all dNTPs along the glycosidic and the phosphodiester bonds. The overall search analyses predicted the

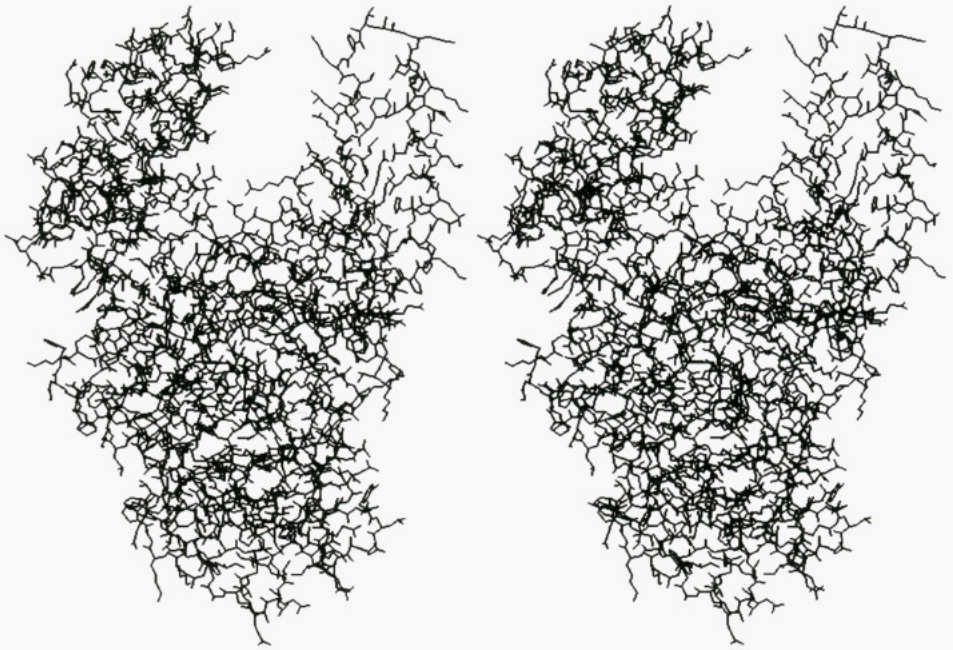

FIGURE 1: Stereoview of the complete structure of the Klenow fragment generated using the SYBYL modeling package.

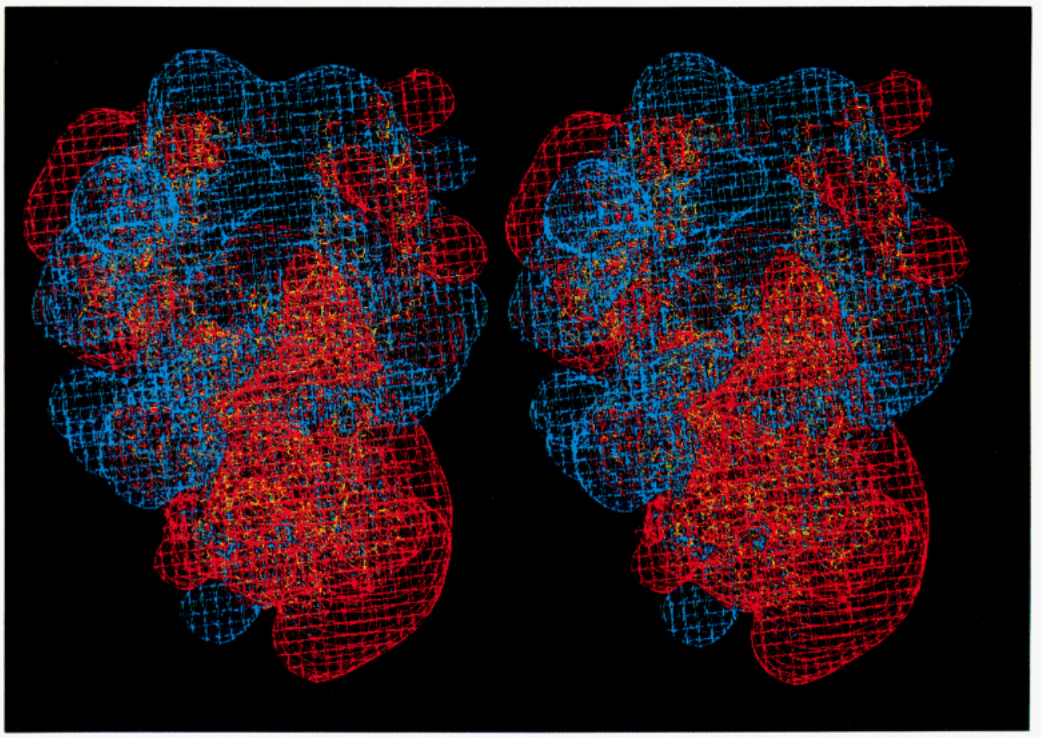

FIGURE 2: Stereoreview of the electrostatic potential field of the Klenow fragment. Positive potential contours are shown in blue, while negative field contours are shown in red. The backbone of the Klenow fragment is shown in yellow.

staggered conformation for the phosphodiester chain. However, the sugar ring remained in the $C_{2'}$ endo conformation, and the torsion angle about the glycosidic linkage ($O_{4'}$–$C_{1'}$–N_1–C_6) was found to have a value of 80°. This arrangement of the base plane being nearly perpendicular to the sugar ring is well suited to adopt the proper Watson–Crick base-pairing scheme reported by Arnott and Hukins (1972). The conformation of the nucleoside moiety in our model is consistent with the solution structure of the nucleotide–enzyme complex reported by Mildvan and colleagues (Sloan et al., 1975; Ferrin & Mildvan, 1986).

B-DNA was read from the data stored in the SYBYL interactive computer graphics technique. First, we made a duplex of $(dA)_{25}$-$(dT)_{25}$ in the B-DNA conformation (Arnott & Hukins, 1972). We then removed 11 dT nucleotides from the 3′ end of the $(dT)_{25}$ strand to make the $(dA)_{25}$-$(dT)_{14}$ template-primer strand. The single-stranded template was allowed to remain in the B-DNA helical conformation. Since this portion of DNA clearly steers away from the polymerase active center, we have left it "as is". The annealing option was used to avoid the unfavorable contacts between the single-stranded DNA and the Klenow fragment. It should be pointed out that the position of the single-strand hangover does not have physiological relevance in our model, and therefore no conformational search of this strand has been carried out.

(*iv*) *Strategy of Making the dTTP and Enzyme Binary Complex*. It has been shown that His-881 is the site of cross-linking of the base moiety of the dNTP (Pandey &

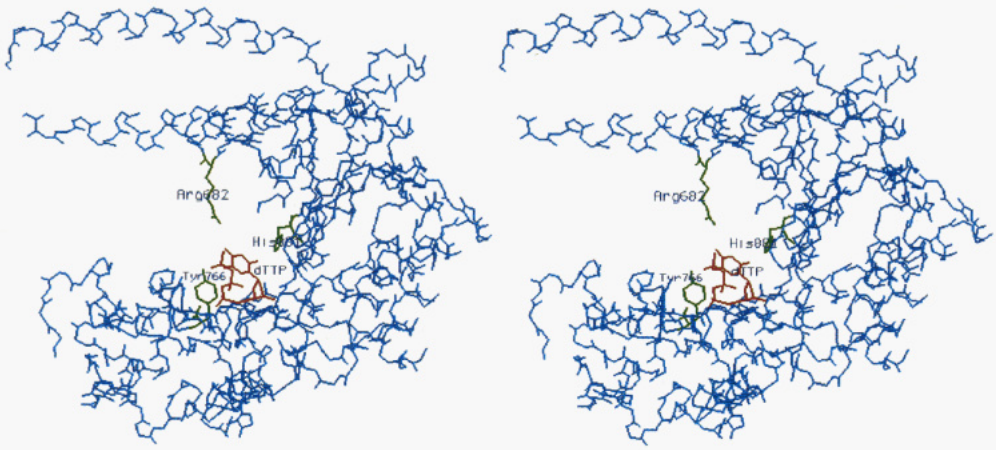

FIGURE 3: Stereoview of the backbone of the Klenow fragment (cyan) complexed with dTTP (red). Only side chains (green) of amino acids involved in the binding of dTTP are shown. The exonuclease portion of the Klenow fragment has not been included in this figure.

Modak, 1988), while Arg-682 was concluded to be the site of affinity labeling in pol I by FSBdA (a dNTP substrate analogue). These results suggested that His-881 and Arg-682 are in close proximity of the base and phosphate moieties, respectively. We selected dTTP as a representative of dNTPs, since it was used as a substrate of choice in the cross-linking experiments. We started by orienting the base and triphosphate ester chain toward His-881 and Arg-682, respectively, in the active site cleft in such a way that optimal favorable contacts with minimum steric hindrance between atoms of the two molecules should exist. Since Lys-758 and Tyr-766 are also reported to participate in the dNTP binding (Basu & Modak, 1987; Rush & Konigsberg, 1990), we attempted to orient these residues toward the sugar and base parts of dTTP. The position of the side chain of Lys-758 does not seem suitable to make contact with any part of the dTTP and was, therefore, excluded from further considerations. Tyr-766 was oriented to make a hydrogen-bond contact with the base of dTTP. The complex so obtained was energy minimized and is depicted in Figure 3. This is one of the many dTTP–enzyme binary complexes that we have made where dTTP was hydrogen bonded to His-881 in a variety of orientations.

(*v*) *Strategy of Making the dTTP, DNA, and pol I Ternary Complex*. In the next phase of model building, we placed the $(dA)_{25}$-$(dT)_{14}$ template-primer in the cleft. The binding of the DNA template-primer to the pol I enzyme is a complex process because multiple contact points exist between the two molecules. Therefore, we decided to use the dTTP–Klenow enzyme complex as a guide to position the primer nucleotide in such a manner that base pairing of the template nucleotide with the bound dTTP is achieved. A number of dTTP–enzyme complexes were tested in this exercise, and only one binary complex, shown in Figure 3, appears to fulfill (i) the Watson–Crick base pairing criterion and (ii) the least steric hindrance between the cleft surface and the double-stranded region of the template-primer. Thus, the position of dTTP in the binary complex should also qualify to be its position in the ternary complex consisting of enzyme, template-primer, and substrate dNTP. In the X-ray crystallographic study of a cocrystal of enzyme–dCTP, a similar orientation of the dCTP substrate was noted by Beese and Steitz (personal communication and figure presented at the Fidelity Meeting at Research Triangle Park, NC, 1989).

The 3′-OH of the primer terminus was then brought in the vicinity (~3 Å) of the α-phosphate of dTTP. Furthermore, the oxygen of the cleavable pyrophosphate linked to the P(α) of dTTP was oriented linearly opposite the oxygen of the 3′-OH of the primer terminus on the basis of the expected stereochemistry of the reaction. Orientation of the triphosphate moiety in this manner in an RNA polymerase system is also suggested by Beal et al. (1990). The distance between the 3′-OH and P(α) was adjusted to a short contact distance in order to poise the groups for the ensuing phosphodiester bond formation. The template-primer $(dA)_{25}$-$(dT)_{14}$ was then fitted into the cavity of pol I in a manner that permits interaction/vicinity of some of the residues reported to be involved in the binding of the template-primer to the enzyme (e.g., Arg-841, Gln-849, Arg-690) and makes minimum steric hindrance with the surrounding.

Since we treated the template-primer as a rigid body, the rest of the template-primer model was automatically positioned into the cleft. The single-stranded template overhang moved in the direction of the 3′–5′ exonuclease portion of pol I, and 7 base pairs of the B-DNA duplex were accommodated by the cleft of the polymerase binding site. This is in agreement with the earlier modeling findings of Steitz and colleagues (Joyce et al., 1986). The position and the orientations of the template-primer–dTTP were further manually adjusted within the active site cleft of the enzyme to optimize the contacts between the two molecular surfaces, which seem to satisfy the requirements for the formation of the prepolymerization complex. In the next step, a conformational energy search of the amino acid side chains at the surface of the template-primer and dTTP substrates was carried out. In most cases we obtained at least four energetically favorable conformational states. Among these, the conformer which had relatively less energy and had either hydrogen-bonding or van der Waals interaction with the substrates was selected for the next step of modeling. In some cases the selected conformation was found to be nearly 30–40 kcal/mol higher than the minimum energy conformation; however, our choice of using these conformers was partly based on good hydrogen-bond geometry. We would like to point out here that the electrostatic term has not been included in the energy calculation. To further refine the geometry of the amino acid side chains around the template-primer and dTTP, restricted local energy minimization around both the template-primer and dTTP substrate was carried out. The annealing scheme dimension was set to 6 and 12 Å around the atoms of both substrates (i.e., the geometry within 6-Å sphere around the substrates was minimized while the atoms within 12 Å were included in the energy calculation). During the annealing, the overall geometry of the template-primer and dNTP substrates was not

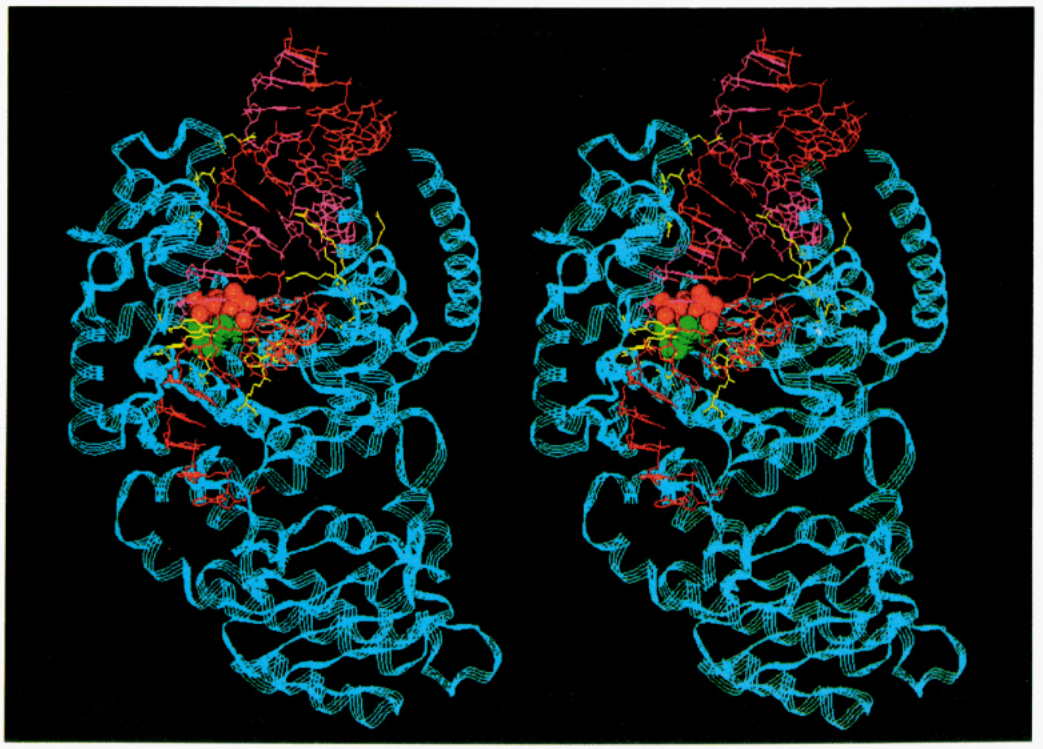

FIGURE 4: Stereoscopic model of the Klenow fragment with $(dA)_{25}$-$(dT)_{14}$ and dTTP. The backbone of the enzyme is shown in the ribbon model (cyan), while dTTP is shown in the space-filling model. The template and primer are shown in red and magenta, respectively. Side chains of the amino acid residues involved in the binding of both substrates are presented in yellow.

disturbed. The resulting energy-refined complex is depicted in Figure 4.

Results and Discussion

A Complete Structure and Electrostatic Potential of the Klenow Fragment. In our overall strategy of building a model for the complete structure of the Klenow fragment, we used Cα coordinates from the crystal structure to establish the backbone and then completed the entire structure using the SYBYL package (Figure 1). Since electrostatic interactions play a very dominant role in the presence of a number of charged residues on both pol I and dNTP and DNA substares, we calculated the electrostatic potentials of the modeled structure of pol I using the DelPhi package of BIOSYM. The stereoview of the electrostatic field of the Klenow fragment is depicted in Figure 2. From the figure, it is clear that most of the positive field contours are centered in the enzyme cleft region and negative field contours are spread all over the enzyme surface. This pattern of positive and negative field contours is in accordance with the one reported by Steitz and colleagues, using the crystal coordinate of the Klenow fragment (Warwicker et al., 1985). Also, close inspection of the figure reveals that the positive electrostatic fields are largely spread in the proposed DNA binding cleft as well as the outer surface of the cleft near the proposed dNTP binding region.

Binding of the DNA Template-Primer. A homopolymeric template-primer, $(dA)_{25}$-$(dT)_{14}$, was aligned into the active site cleft of the Klenow fragment as described under Materials and Method. The amino acid residues within the 4-Å sphere around the DNA template-primer are given in Table I. Of these, Arg-841 and Arg-690 in the DNA duplex binding region of the template-primer obtained in this model are in agreement with the DNA binding sites suggested by chemical modification (Mohan et al., 1988) and mutation studies (Joyce et al., 1985). Furthermore, the positive contour of the electrostatic potential of this region seems quite suitable for binding to the electronegative charge cloud of the DNA molecule

Table I: Amino Acid Residues Falling in the 4.0-Å Region around the DNA Duplex

Ile-627	Leu-628	Glu-917	His-928	Ile-765	Leu-764
Arg-631	Lys-635	Asp-732	Arg-735	Met-768	Phe-762
Gln-677	Asn-678	His-734	Arg-755	Leu-773	Met-848
Val-681	Asn-683	Ile-760	Ser-756	Arg-841	Gln-708
Glu-684	Arg-687	Glu-752	Gln-753	Ile-709	Glu-710
Arg-690	Glu-883	Gln-849	Tyr-766		

(Figure 2). These results are also in agreement with earlier results of Steitz and colleagues (Warwicker et al., 1985).

Binding of the Substrate dNTP and DNA Template-Primer. The position of the dTTP substrate in the binding pocket of the Klenow fragment was tentatively determined by orienting the side chains of the desired amino acids toward the base, sugar, and triphosphate moieties of dTTP (Figure 3). This binary complex was then used as a recipient for the template-primer structure. The resulting prepolymerization complex consisting of substrate, template-primer, and enzyme is shown in Figure 4. The amino acid residues involved in the interaction/vicinity of the substrates are highlighted. The amino acid residues in close atomic proximity of dTTP are His-881, Asn-845, and Tyr-766. Two other residues, Arg-682 and Asp-882, known to be important in substrate binding/polymerization, are also shown. The side chain of Arg-682 is pointing toward the triphosphate while the carboxyl group of Asp-882 is at a distance of 6 Å directly facing the β-phosphate oxygen, suggesting the possible binding site for the divalent metal ion. In this model, the role of His-881 may be explained in two ways. In one, it is stacked against the base of the dNTP, and in the other, it could form a hydrogen bond with the base. The former possibility is more attractive since His-881 may be proposed to interact with all four nucleotide bases in this manner. The interesting feature of the minimization of the complex is the position of Tyr-766. In the begining, we had used this residue to make a hydrogen bond with the base of dNTP, but after minimization, it clearly

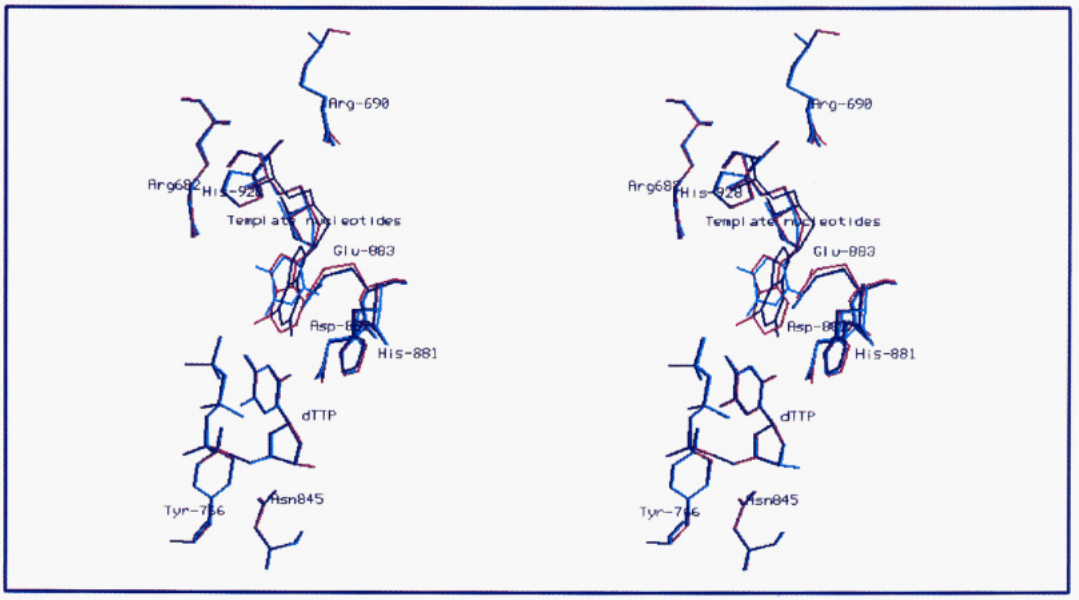

FIGURE 5: Superimposed structure (in stereo) of the dTTP–enzyme complex with different template nucleotide units: dA (magenta), dG (dark blue), and dT (cyan). Only those residues defining the substrate binding pocket and those around the template nucleotide are shown. The local geometry around the template nucleotide was minimized, keeping the position of dTTP fixed. The proper Watson–Crick base pairing occurs only with dA as the template nucleotide (shown in magenta). Note the different spatial positions that residues assume in the presence of different template nucleotides in the template.

stabilized to a position between the sugar moieties of the primer terminus and that of dNTP in the van der Waals distance region. This position of Tyr-766 explains its importance in binding both the primer and dNTP. This also seems well suited for the polymerization reaction where the site of binding of the primer terminus and dNTP must be close to each other. Such a role of Tyr-766 seems consistent with earlier observations that Tyr-766 was the site for binding of both the substrate analogue, 8-azido-dATP (Rush & Konigsberg, 1990), and the primer terminus (Catalano et al., 1990). Kinetic studies of a mutant of DNA pol I in which Tyr-766 is replaced by serine or phenylalanine residues have shown reduced fidelity of DNA synthesis (Carroll et al., 1991). The degree of misincorporation increased in the Tyr → Ser mutant while the fidelity of the Tyr → Phe mutation did not show significant change. This study indicates the important role for the phenyl ring of Tyr-766 in the selection of the proper substrate. The possible role of the phenolic ring of Tyr-766 might be to correctly align the α-phosphorus of the appropriate dNTP with respect to the 3′-OH of the primer terminus. Yet another residue, Asn-845, which has been implicated in the binding of dNTP through mutagenesis studies (Polesky et al., 1990) also appears within a 4-Å distance pointing toward the deoxyribose moiety of dNTP in our model.

In the minimized structure, the position of the Arg-682 side chain did not shift from its original position. In the pre-polymerization complex, there is no direct contact between Arg-682 and the triphosphate chain; however, during nucleophilic attack by the primer terminus, bond length and bond angles around the P(α) of the triphosphate undergo transition in its geometry from tetrahedral to trigonal bipyramidal (Boyd, 1969). At this time, the only available positive residue in the vicinity is Arg-682, which is well suited to accept the leaving pyrophosphate group of the incorporated dNTP. In this context, it should be pointed out that the polymerization reaction is dependent on the presence of a divalent metal ion. For simplicity, we have not yet included a metal ion in this model. However, there is sufficient space to accommodate either a Mg^{2+} or Mn^{2+} ion to interact with the β- and/or γ-phosphate groups with one of the acidic amino acids adjacent to His-881 in the β-sheet structure which may provide the contact point through a metal bridge. Work in this direction is in progress.

Template-Dependent Substrate Binding: Formation of a Specific Pocket. Since DNA replication is dictated by the template nucleotide, i.e., the single nucleotide unit immediately following the double-stranded region of DNA, we attempted to define the relative positions of the side chains of the amino acid residues which were adjacent to the template base. In this study, all four bases were individually inserted as a template base, keeping the preceding double-stranded DNA intact. Local energy minimization calculations were then performed for all four template bases. Figure 5 represents the superimposed structure of the complex having different nucleosides as the template base. The side chains of residues 882–883 (Asp, Glu) and to a lesser extent His-881 adjust to different positions for the four systems. Furthermore, the enzyme cavity appears unaffected regardless of the template base change, but the local side-chain conformation around the template nucleotide does change in response to the template nucleotide base. This change is likely to be responsible for the selection of the proper dNTP substrate binding. In our model Glu-883 is the most proximal to the template nucleotide. Its binding to the template base may cause a different side-chain orientation of Asp-882. This, in turn, influences the position of His-881. We therefore propose that residues 881–883 together with the template nucleotide (hydrogen-bonding center) dictate the selection of the proper dNTP by forming a specific shape of the pocket. Other contributing components of this pocket may include Asn-845 and Tyr-766. While all four dNTPs may recognize some part of the pocket, only the dNTP with the base complementary to the template nucleotide can interact with all the components of the pocket, resulting in stable binding. The stably bound dNTP is positioned with its α-phosphate directly across from the 3′-OH of the primer terminus and forms a base pair with the template, and the subsequent catalytic step can then proceed.

Biochemical Properties of the Enzyme Reaction and the Model. One of the basic properties of DNA polymerase is the requirement of deoxynucleotides as a substrate while ri-

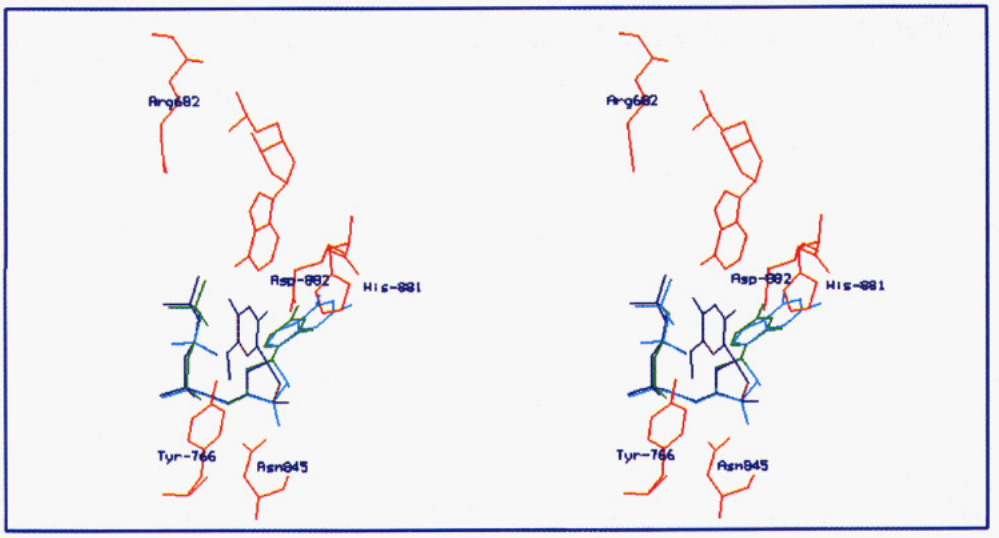

FIGURE 6: Stereoview of the template–enzyme complex (red) with rUTP (green), rGTP (cyan), and an ethyl-substituted dTTP (dark blue). In this case, the enzyme–template-primer complex as seen in Figure 4 was used in conjunction with the desired substrate. For clarity, only residues around the substrate binding pocket are shown. Note that the proper base pairing occurs only with dTTP and ethyl-dTTP (red and blue) but not with ribonucleotides.

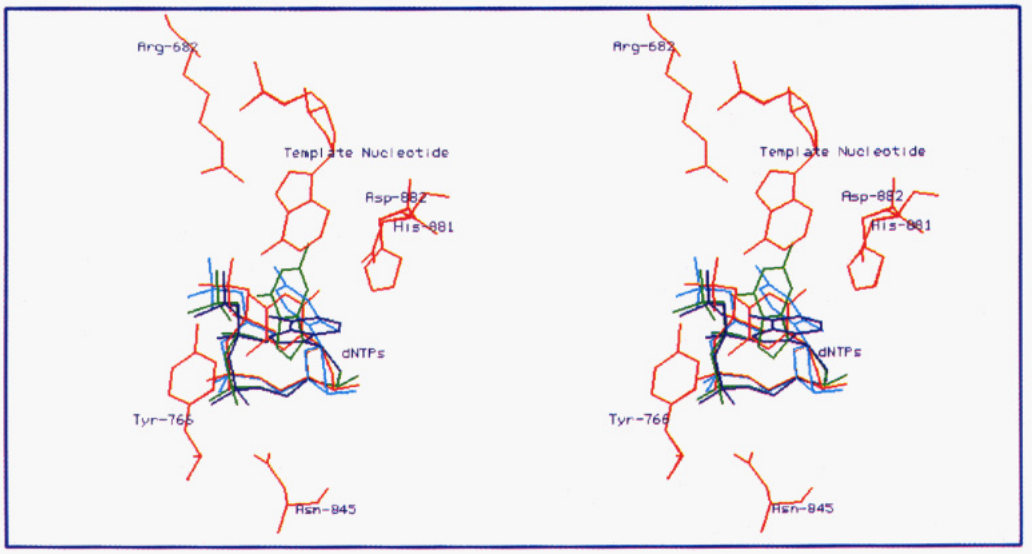

FIGURE 7: Stereoview of the enzyme–template complex with different dNTPs. Only the residues involved in the dNTP binding (binding pocket) are shown here. All the dNTPs are superimposed along $C_{1'}$–$O_{4'}$–$C_{4'}$ of the sugar moiety. The positions of dGTP (green), dATP (dark blue), dCTP (cyan) are indicated. Note that the Watson–Crick base pairing exists only with dTTP (red).

bonucleotides are neither accepted nor compete with dNTPs during the polymerization reaction. We have therefore compared the geometries of different deoxynucleotides and their ribose counterparts. Figure 6 represents how the substitution at the base and sugar affect the interaction of dNTP with the active site pocket in the presence of the template base. The base modification at C_5 does not affect the position of the base ring and sugar and that of the triphosphate orientation. But the substitution at $C_{2'}$ of the sugar moiety drastically changes the ring puckering and hence the orientation of the nucleotide base, which diminishes base pairing and also affects the nature of attracting forces generally found between dNTP and the critical residues around it. This appears to be the major reason why the enzyme does not accept ribonucleotides in the presence of the template. As we discussed earlier, the specificity of selecting the proper nucleotide by the enzyme is based on the pocket developed by the template base bound to Glu-883, Asp-882, His-881, Asn-845, and Tyr-766. Among these five components required for stable dNTP binding, four may be accessible to all natural dNTPs in the absence of the template-primer. However, in the presence of the template nucleotide, binding is restricted to a complementry nucleotide which binds stably in the pocket. The other nucleotides, in spite of their ability to recognize some parts of the binding pocket, would cause higher energy of the system because of their inability to form Watson–Crick base pairing with the template base and would therefore be expelled. Using dA as a template nucleotide, we show the overlapping positions of all four dNTPs in Figure 7.

Concluding Remarks

The most interesting and exciting result we have derived here is the working model for the Klenow fragment of DNA polymerase I which is based on Cα coordinates and many biochemical and genetic data. The model structure seems to be reasonably good as it reproduces the electrostatic behavior of pol I as reported for the crystal structure.

In the preparation of the dTTP–enzyme complex, the choice of the dNTP structure within the binding pocket was primarily dictated by the earlier affinity labeling and NMR observations. Addition of DNA to the dTTP–enzyme complex required selection of the template-primer conformation, i.e., the B-form

of DNA, since it is the most commonly occurring canonical form of DNA found in biological systems. Because of the enormous size of the complex, the side-chain path of the DNA surface was traced manually, and the conformational search was done on one residue at a time. The resulting product is what we have termed a model for the "prepolymerization complex". The model also suggests participation of hitherto unknown amino acid residues in the binding process. The putative roles of these amino acid residues could be clarified through appropriate genetic and biochemical experiments. Experimental results, in turn, will guide further refinements. Similarly, the model should prove useful in explaining the molecular details of the participation of new (or substituted) amino acid residues that may become known in the future. Availability of the complete structure based on crystal data of the enzyme would certainly show initial correspondence, if any, of our model to the actual structure. We must caution, however, that the model we have presented here, without the treatment of solvent and other physiological media components, provides only a qualitative idea of the interactions between the enzyme protein and its substrates. Further refinement of this model would certainly require the molecular dynamics calculation under physiological conditions.

It is our hope that molecular modeling approaches described here for the Klenow fragment of *E. coli* DNA polymerase I would stimulate efforts to use these tools in understanding catalytic mechanisms of other replication enzymes. Since DNA polymerases from diversified sources catalyze identical reactions, i.e., template-dependent dNTP polymerization, it should be possible to extrapolate some of our observations to these systems. Perhaps molecular modeling may provide an alternative approach to predict the structure of replicative enzymes whose crystal structures have not been and may not be solved.

Acknowledgments

We thank Drs. Leslie Michelson and Swamy Laxminarayan for encouragement and for providing us molecular modeling facilities of the academic computing services of the University. We are grateful to Drs. Edward Arnold and Williams Rogers for many useful suggestions in the initial stage of the work.

References

Arnott, S., & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504–1511.

Basu, A., & Modak, M. J. (1987) *Biochemistry 26*, 1704–1709.

Basu, S., Basu, A., & Modak, M. J. (1988) *Biochemistry 27*, 6710–6716.

Beal, R. B., Pillai, R. P., Chuknyisky, P. P., Levy, A., Tarien, E., & Eichhorn, G. L. (1990) *Biochemistry 29*, 5994–6002.

Boyd, D. B. (1969) *J. Am. Chem. Soc. 91*, 1200–1205.

Brutlag, D., Atkinson, M. R., Setlow, P., & Kornberg, A. (1969) *Biochem. Biophys. Res. Commun. 37*, 982–989.

Buckert, U., & Allinger, N. L. (1982) *Molecular Mechanics*, ACS Monograph 117 (MM2 Version 2, QCPE Program 395), American Chemical Society, Washington, DC.

Carroll, S. S., Cowart, M., & Benkovic, S. J. (1991) *Biochemistry 30*, 804–813.

Catalano, C. E., Allen, D. J., & Benkovic, S. J. (1990) *Biochemistry 29*, 5994–6002.

Derbyshire, V., Freemont, P. S., Sanderson, M. R., Beese, L. S., Friedman, J. M., Steitz, T. A., & Joyce, C. M. (1988) *Science 240*, 199–201.

Ferrin, L. J., & Mildvan, A. S. (1986) *Biochemistry 25*, 5131–5145.

Freemont, P. S., Ollis, D. L., Steitz, T. A., & Joyce, C. M. (1986) *Proteins 1*, 66–73.

Gilson, M., Sharp, K., & Honig, B. (1988) *J. Comput. Chem. 9*, 327–335.

Jovin, T. M., Englund, P. T., & Bertsch, L. L. (1969) *J. Biol. Chem. 224*, 2996–3007.

Joyce, C. M. (1991) *Curr. Opin. Struct. Biol. 1*, 123–129.

Joyce, C. M., & Grindley, N. D. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1830–1834.

Joyce, C. M., Fujii, D. M., Laks, H. S., Hughes, C. M., & Grindley, N. D. F. (1985) *J. Mol. Biol. 186*, 283–293.

Joyce, C. M., Ollis, D. L., Rush, J., Steitz, T. A., Konigsberg, W. H., & Grindley, N. D. F. (1986) in *Protein Structure, Folding and Design. UCLA Symposia on Molecular and Cellular Biology* (Oxender, D., Ed.) pp 197–205, Liss, New York.

Klapper, I., Hagstrom, R., Sharp, K., & Honig, B. (1986) *Proteins*, 47–59.

Klenow, H., & Henningsen, I. (1970) *Proc. Natl. Acad. Sci. U.S.A. 65*, 168–175.

Kornberg, A. (1980) *DNA Replication*, Freeman, San Francisco.

Marquardt, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983) *Acta Crystallogr., Sect. B 39*, 480–485.

Mohan, P. M., Basu, A., Basu, S., Abraham, K. I., & Modak, M. J. (1988) *Biochemistry 27*, 226–233.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., & Steitz, T. A. (1985) *Nature 313*, 762–766.

Pandey, V. N., & Modak, M. J. (1988) *J. Biol. Chem. 263*, 6068–6073.

Pandey, V. N., Williams, K. R., Stone, K. L., & Modak, M. J. (1987) *Biochemistry 26*, 7744–7748.

Pandey, V. N., Kaushik, N. A., Pradhan, D. S., & Modak, M. J. (1990) *J. Biol. Chem. 265*, 3679–3684.

Polesky, A. H., Steitz, T. A., Grindley, N. D. F., & Joyce, C. M. (1990) *J. Biol. Chem. 265*, 14579–14591.

Rush, J., & Konigsberg, W. H. (1990) *J. Biol. Chem. 265*, 4821–4827.

Saenger, W. (1983) in *Principles of Nucleic Acid Structure* (Cantor, C. R., Ed.) pp 82–89, Springer-Verlag, New York.

Sloan, D. L., Loeb, L. A., Mildvan, A. S., & Feldmann, R. J. (1975) *J. Biol. Chem. 250*, 8913–8920.

Sundaralingam, M. (1969) *Biopolymers 7*, 821–860.

Warwicker, J., Ollis, D. L., Richards, F. M., & Steitz, T. A. (1985) *J. Mol. Biol. 186*, 645–649.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765–784.

Yadav, J., Yadav, P. N., Laxminarayan, S., Michelson, L., Arnold, E., & Modak, M. J. (1990) *Ann. N.Y. Acad. Sci. 616*, 624–630.

Yadav, J., Yadav, P., Laxminarayan, S., Arnold, E., & Modak, M. (1991) in *Protein: Structure, Dynamics and Design* (Renugopalkrishanan, V., Corey, P. R., Smith, I. C. P., Haung, S. G., & Storer, A. C., Eds.) ESCOM Science Publishers B. V., Leiden.

Zimmermann, S. S., Marcia, S. P., Nemethy, G., & Scheraga, H. A. (1977) *Macromolecules 10*, 1–12.