Role of Active Site Residues in the Glycosylase Step of T4 Endonuclease V. Computer Simulation Studies on Ionization States†

Monika Fuxreiter,‡ Arieh Warshel,§ and Roman Osman*,‡

Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, New York 10029, and Department of Chemistry, University of Southern California, Los Angeles, California 90089

Received January 26, 1999; Revised Manuscript Received May 17, 1999

ABSTRACT: T4 Endonuclease V (EndoV) is a base excision repair enzyme that removes thymine dimers (TD) from damaged DNA. To elucidate the role of the active site residues in catalysis, their pK_a's were evaluated using the semimicroscopic version of the protein dipoles−Langevin dipoles method (PDLD/S). Contributions of different effects to the pK_a such as charge−charge interactions, conformational rearrangement, protein relaxation, and DNA binding were analyzed in detail. Charging of the active site residues was found to be less favorable in the complex than in the free enzyme. The pK_a of the N-terminus decreased from 8.01 in the free enzyme to 6.52 in the complex, while the pK_a of Glu-23 increased from 1.52 to 7.82, which indicates that the key residues are neutral in the reactant state of the glycosylase step. These pK_a's are in agreement with the optimal pH range of the reaction and support the N-terminus acting as a nucleophile. The Glu-23 in its protonated form is hydrogen bonded to O4′ of the sugar of 5′ TD and can play a role in increasing the positive charge of C1′ and, hence, accelerating the nucleophilic substitution. Furthermore, the neutral Glu-23 is a likely candidate to protonate O4′ to induce ring opening required to complete the glycosylase step of EndoV. The positively charged Arg-22 and Arg-26 provide an electrostatically favorable environment for the leaving base. To distinguish between S_N1 and S_N2 mechanisms of the glycosylase step the energetics of protonating O2 of 5′ TD was calculated. The enzyme was found to stabilize the neutral thymine by ∼3.6 kcal/mol, whereas it destabilizes the protonated thymine by ∼6.6 kcal/mol with respect to an aqueous environment. Consequently, the formation of a protonated thymine intermediate is unlikely, indicating an S_N2 reaction mechanism for the glycosylase step.

UV light can induce modifications in DNA bases, such as formation of covalent bonds between adjacent pyrimidines (*1*). These photoproducts are mutagenetic and lethal for the cells, often resulting in cancer (*2, 3*). The cells have elaborate protective mechanisms to eliminate these lesions by either base (*4*) or nucleotide excision pathways (*5, 6*). Endonuclease V (EndoV)[1] from bacteriophage T4 is a base excision enzyme, which initiates the repair of a thymine dimer lesion in DNA (*7, 8*).

The catalytic activity of endonuclease V comprises two steps (*9, 10*): (i) cleavage of the glycosidic bond and (ii) incision of the phosphodiester bond at the apyrimidinic site. The enzyme was studied extensively by biochemical (*11*−*13*) and structural methods (*14*−*17*), providing a wealth of information on its binding properties to DNA as well as on the catalytic mechanism. Nevertheless, the molecular basis for selective recognition of thymine dimers and for the role of various active site residues has not been unequivocally established.

It has been shown that the enzyme first associates with the DNA by nonspecific electrostatic interactions (*18*) and locates the damaged site by a scanning mechanism (*19*−*21*). The positively charged residues on the interacting surface of the enzyme were demonstrated by site-directed mutagenesis and substrate analogue binding to be significant for specific binding (*22, 23*). The three-dimensional crystal structure of the enzyme in complex with DNA exhibits a kink of 60° near the thymine dimer (TD) (*16*). Molecular dynamics (MD) simulations and NMR studies showed that the DNA in solution has a smaller kink of ∼20° around the damaged site (*24*−*28*). Thus, the protein induces a large distortion in DNA structure when the complex is formed. In the crystal structure, the adenine opposite the 5′ thymine of the dimer was found in an extrahelical position embedded inside the protein, presumably relaxing the strain introduced by the bending. These observations suggest that conformational changes in DNA could play a role in specific binding, although this relationship remains to be established.

The proposed catalytic mechanism of endonuclease V is displayed in Scheme 1. Capturing the reduced intermediate of the product of the glycosylase step established the role of the N-terminus as the nucleophile (*29*). It has been proposed that the reaction is facilitated by a protonation on O2 of the

† This work was supported by USPHS Grant CA 63317 (R.O.) and also by National Institutes of Health Grant GM24492 (A.W.).

* To whom correspondence should be addressed. Telephone: 212-241-5609. Fax: 212-860-3369. Email: osman@inka.mssm.edu.

‡ Mount Sinai School of Medicine.

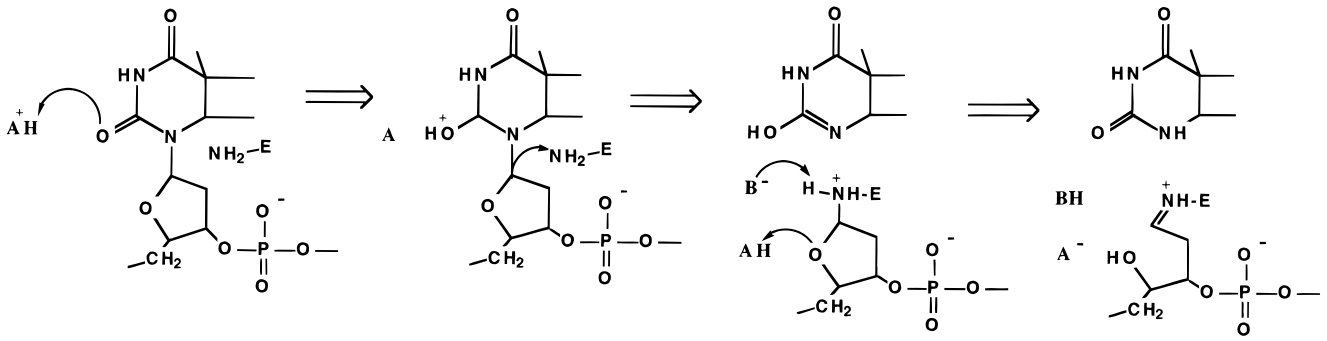
§ University of Southern California.

[1] Abbreviations: EndoV, T4 Endonuclease V; TD, thymine dimer; PDLD, protein dipoles−Langevin dipoles method; PDLD/S, semimacroscopic protein dipoles−Langevin dipoles method; MD, molecular dynamics simulations; LRA, linear response approximation; SCAAS, surface-constrained all-atom solvent model; RHF, restricted Hartree−Fock calculation; ur, unrelaxed model; rel, relaxed model.

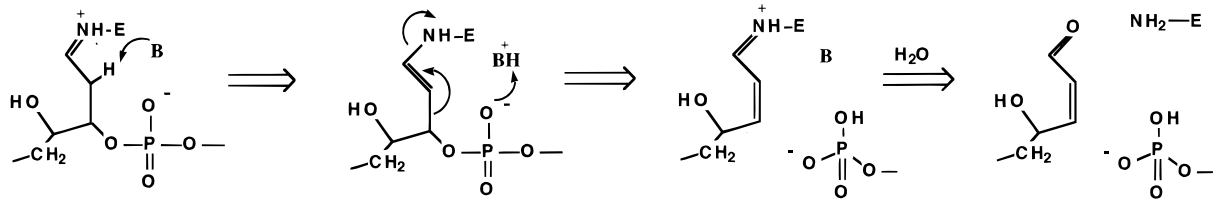
10.1021/bi9901937 CCC: $18.00
Published on Web 07/09/1999

Scheme 1

1. Glycosylase step



2. Endonuclease step



5′ thymine of the dimer (*29*, *30*). Such a mechanism requires an unprotonated N-terminus around pH 7 and assumes the existence of a nearby residue that can donate a proton to the thymine dimer. The pK_a of the N-terminus in the complex with DNA has not been measured. Furthermore, in the structure of the complex, Arg-26 is found close to O2 of 5′ TD, but its ability to donate a proton is not known. The scission of the glycosidic bond is followed by an opening of the sugar ring, which is assisted by a general base and a general acid that remove a proton from the amine and protonate the O4′ of the sugar moiety, respectively. The product of this reaction is a protonated Schiff base, which was experimentally identified via reductive methylation (*29*). The identity of the general acid and general base that play a role in the formation of the Schiff base has also not been determined.

The cleavage of the phosphodiester bond occurs via β-elimination (*31*, *32*), followed by the hydrolysis of the Schiff base leading to the dissociation of the enzyme and the formation of an α,β-unsaturated aldehyde. In the presence of high concentrations of external nucleophile the DNA backbone remains intact (*33*), because hydrolysis of the Schiff base is preferred.

The presented reaction mechanism is not fully confirmed, since several intermediates cannot be investigated experimentally. The catalytic effect of the enzyme, however, cannot be explained without the knowledge of the detailed reaction mechanism. To study such a mechanism computationally, it is essential to determine the charge state of the ionizable residues, which also helps to elucidate the role of the catalytic residues. The present work focuses on calculating the pK_a's of the active site residues of endonuclease V: the N-terminus, Glu-23, Arg-22, Arg-26, and Arg-3.

The N-terminus and Glu-23 were shown to be indispensable for the glycoslysis of thymine dimer (*13*, *30*). Since the N-terminus is the attacking nucleophile, its pK_a has a primary influence on the pH profile of this reaction. The replacement of Glu-23 by neutral residues (e.g., Gln) was found to abolish the catalytic activity of the enzyme (*11*). This suggests that the negative charge of this carboxylate group has special importance in the course of the scission of the glycosidic bond. The substitution of Glu-23 by Asp, however, resulted in an inactive enzyme, which indicates that Glu-23 plays a more complex role in this reaction (*13*). Mutations of Arg-3, Arg-22, and Arg-26 also reduce the rate constant of glycolysis, although to a smaller extent (*11*). Arg-26 is of special interest, since the NH1 of this residue is positioned 2.6 Å from the O2 of 5′ thymine of the dimer, making it a likely candidate to donate a proton prior to the cleavage of the glycosidic bond. Earlier quantum chemical calculations carried out in vacuo indicated that Arg-26 remains protonated during the reaction (*34*).

The complexation with DNA did not cause significant changes in the overall structure of the enzyme; the RMSD between the free enzyme and the complex is 0.54 Å for main chain atoms and 1.13 Å for heavy atoms, although some active site residues are considerably moved. For example, the distance between the NH2 atoms of Arg-22 and Arg-26 decreased by almost 3 Å. To understand the effect of these conformational changes on the pK_a of the active site residues, an artificial model was constructed by removing the DNA from the complex structure. The pK_a's calculated in this artificial complex enzyme model were compared to those in the free enzyme as well as in the complex. In this way the effects of structural changes in the enzyme induced by DNA binding and the interactions with the DNA groups could be studied separately.

An important step in enzymatic catalysis of the hydrolysis of the glycosidic bond is the stabilization of the leaving group, i.e., the 5′ thymine of TD. Since the pyrimidine anion is not a good leaving group, two possible pathways could be envisaged. In one, a protonation of the anion will greatly enhance the rate of the glycosidic bond breaking. In the other, a positively charged group could provide electrostatic stabilization of the leaving anion, which will be neutralized in a subsequent step. Indeed, Arg-26 is positioned in the

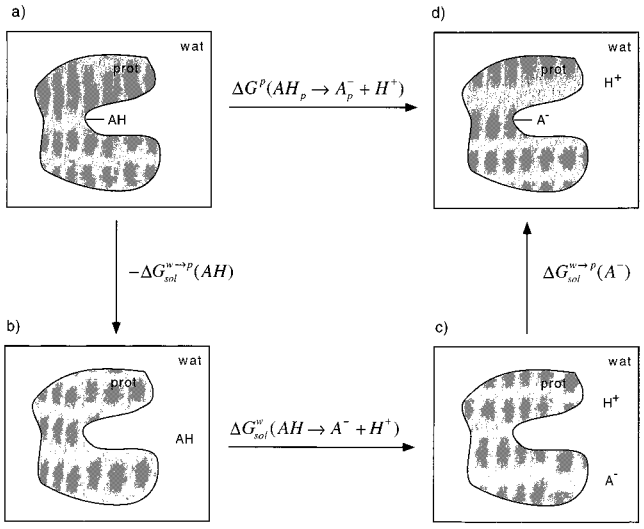


FIGURE 1: Thermodynamic cycle for calculation of pK_a of ionizable residues. ΔG^{w}_{sol} and ΔG^{p}_{sol} are the free energies of solvation in water and in protein, respectively, whereas $\Delta G^{w\rightarrow p}_{sol}$ is the free energy of moving the relevant group from water to the protein site.

proximity of O2 of the 5′ thymine, and it can serve in either of the two proposed roles. Alternatively, Glu-23 was hypothesized to participate in protonation of the pyrimidine ring (*13*, *16*). To test this hyphothesis, we calculated the stability of protonated 5′ thymine in the enzyme compared to aqueous environment according to different protonation schemes using the PDLD/S method (*35*).

METHODS

Calculation of pK_a. The pK_a of an ionizable group can be calculated according to the thermodynamic cycle presented in Figure 1 (*36*):

$$\Delta G^{p}(\mathrm{AH_p} \rightarrow \mathrm{A_p^-} + \mathrm{H_w^+}) = \Delta G^{w}(\mathrm{AH_w} \rightarrow \mathrm{A_w^-} + \mathrm{H_w^+}) + \Delta G^{w\rightarrow p}(\mathrm{A^-}) - \Delta G^{w\rightarrow p}(\mathrm{AH}) \quad (1)$$

Here p and w denote the protein and water environment, respectively, $\Delta G^{w\rightarrow p}$ represents the free energy of moving the particular group from water to its protein site.

Equation 1 can be rewritten as

$$\mathrm{p}K^{p}_{a,i} = \mathrm{p}K^{w}_{a,i} + \frac{1}{2.3RT}\Delta\Delta G^{w\rightarrow p}(\mathrm{AH}_i \rightarrow \mathrm{A}_i^-) \quad (2)$$

where $\mathrm{p}K^{p}_{a,i}$ and $\mathrm{p}K^{w}_{a,i}$ are the pK_a's of the *i*th group in protein and water, respectively. $\Delta G^{w\rightarrow p}(\mathrm{AH}_i \rightarrow \mathrm{A}_i^-)$ is the difference between the solvation free energy of the *i*th residue in water and in protein in its charged and uncharged state, respectively. Since the pK_a in water is often experimentally known, the determination of the pK_a in protein is reduced to the calculation of the second term. As the effect of the protein environment is significantly larger on the ionized state than on the neutral state, the $\Delta G^{w\rightarrow p}(\mathrm{AH})$ term is usually ignored.

Since in general more than one charged group is considered, it is convenient to express the pK_a of each protein group in two terms: (i) the free energy of charging the specific ionized group at the protein site when all other residues are neutral and (ii) the interaction with other ionized residues.

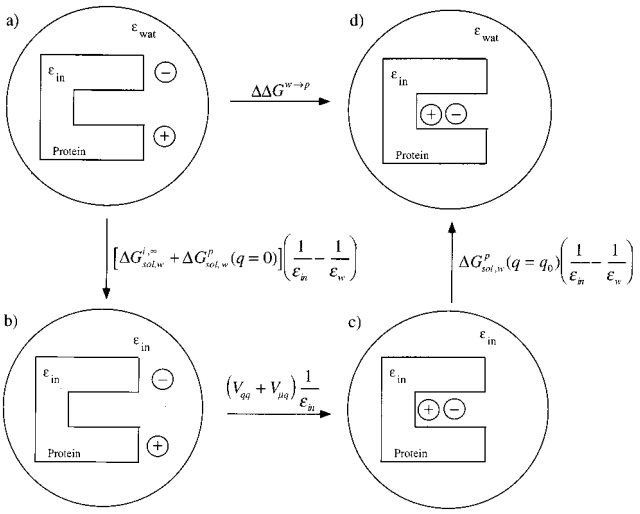


FIGURE 2: Thermodynamic cycle used in PDLD/S calculations. $\Delta G^{i,\infty}_{sol,w}$ is the free energy of solvation in water, when the relevant groups are infinite separation. $\Delta G^{p}_{sol,w}(q = q_0)$ and $\Delta G^{p}_{sol,w}(q = 0)$ are the solvation energies of the protein in water when the relevant groups are charged and uncharged, respectively. V_{qq} and $V_{\mu q}$ are the interactions of the relevant groups with the charges and permanent dipoles of the protein. ϵ_w is the dielectric constant in water, and ϵ_{in} is a scale factor that accounts for those effects which were considered implicitly in the calculations.

$$\mathrm{p}K^{p}_{a,i} = \mathrm{p}K^{p,int}_{a,i} + \sum_{j\neq i}(\Delta \mathrm{p}K^{p}_{a,i})_j = \mathrm{p}K^{w}_{a,i} + \frac{1}{2.3RT}(\Delta\Delta G^{w\rightarrow p}_{self})_i + \frac{1}{2.3RT}\sum_{j\neq i}\Delta G^{p}_{ij} \quad (3)$$

where $\mathrm{p}K^{p,int}_{a,i}$ are the intrinsic pK_a in protein when all other residues are kept neutral in the process of charging and $(\Delta \mathrm{p}K^{p}_{a,i})_j$ is the pK_a shift induced by the effect of other *j*th ionizable residue. To evaluate this equation, two terms are calculated: the self-energy $(\Delta\Delta G^{w\rightarrow p}_{self})_i$ when moving the *i*th residue from water to the protein site, keeping all other ionizable residues neutral, and ΔG^{p}_{ij}, the interaction with other ionizable groups.

Determination of Self-Energy. The self-energy was calculated as the difference between the solvation free energy in water and in protein by the semimicroscopic version of the PDLD model (PDLD/S) (*35*) using the thermodynamic cycle represented in Figure 2.

$$\Delta\Delta G^{w\rightarrow p} = [-\Delta G^{i,\infty}_{sol,w} + (\Delta G^{p}_{sol,w}(q = q_0) - \Delta G^{p}_{sol,w}(q = 0))]\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_w}\right) + (V_{qq} + V_{\mu q})\frac{1}{\epsilon_{in}} \quad (4)$$

where $\Delta G^{i,\infty}_{sol,w}$ is the solvation energy of the *i*th residue in water, when the charged groups are infinitively separated from each other. $\Delta G^{p}_{sol,w}$ $(q = q_0)$ and $\Delta G^{p}_{sol,w}$ $(q = 0)$ are the solvation energies of the protein in water when the studied group is in the charged and uncharged form, respectively. V_{qq} and $V_{\mu q}$ are the interactions with the protein charges and polar groups in a vacuum. ϵ_w is the dielectric constant of water, and ϵ_{in} is a scale factor that represents the contributions which are not considered explicitly in the

model. As induced dipole effects of the protein and water penetration were included implicitly, $\epsilon_{in} = 4$ was used throughout this work.

The solvation energy in the PDLD model (*37*, *38*) is composed of the terms:

$$\Delta G^{p}_{sol} = \Delta G_{L} + \Delta G_{Born} + \Delta G_{\mu q} + \Delta G_{\alpha q} + \Delta G_{vdw} + \Delta G_{hydro} \quad (5)$$

where ΔG_L is the Langevin term, the interaction of the water dipoles with the charges of the studied group, ΔG_{Born} is the bulk solvation energy, $\Delta G_{\mu q}$ and $\Delta G_{\alpha q}$ are the interaction terms with the protein permanent and induced dipoles, respectively, ΔG_{vdw} is the van der Waals interaction term, and ΔG_{hydro} is an estimation of the hydrophobic contribution.

The effect of protein structural relaxation upon formation of charges was studied by using a linear response approximation (LRA) (*35*), by averaging the PDLD/S results over configurations generated by MD simulations with the relevant group charged and uncharged, respectively:

$$\Delta\Delta G^{w\rightarrow p} = \frac{1}{2}(\langle \Delta\Delta G^{w\rightarrow p}\rangle_{q=q_0} + \langle \Delta\Delta G^{w\rightarrow p}\rangle_{q=0}) \quad (6)$$

where the $\langle \Delta\Delta G^{w\rightarrow p}_{q=q_0}\rangle$ term is the average solvation energy of configurations generated when the relevant groups are charged and $\langle \Delta\Delta G^{w\rightarrow p}_{q=0}\rangle$ term when they are uncharged.

Interaction between Ionizable Residues. Since the interaction between two ionizable groups depends on the ionization state of all the other ionizable residues, to calculate the ΔG^{p}_{ij} interaction term, the free energy of a given *m* charge configuration has to be evaluated:

$$\Delta G^{m} = \sum_{i} -q_i^m \left\{ 2.3RT(pK^{int}_{a,i} - pH) + \frac{1}{2}\sum_{i\neq j}\Delta G^{p}_{ij} \right\} = \sum_{i=1}^{N} \left\{ -q_i^m\, W_i^0 + \frac{1}{2}\sum_{i\neq j} W_{ij} q_i^m\, q_j^m \right\} \quad (7)$$

where q_i^m and q_j^m are the actual charges of *i*th and *j*th group in a charge configuration *m*. The coupling term is defined as $W_{ij} = 332(1/R^m_{ij}\,\epsilon^m_{ij})$, where R^m_{ij} is the distance between the centers of *i*th and *j*th group and ϵ^m_{ij} is the effective dielectric constant between the two residues in this particular charge configuration.

Using the free energy of all possible charge configurations, the average charge of each residue can be calculated:

$$\langle q_i\rangle = \frac{\sum_m q_i^m \exp\{-\Delta G^m\beta\}}{Z} \quad (8)$$

where $\langle q_i\rangle$ is the mean value of the charges of the *i*th residue and $Z = \sum_m \exp\{-\Delta G^m\beta\}$. Equations 7 and 8 are solved self-consistently. The pK_a of the *i*th residue is defined as the pH at which $|\langle q_i\rangle| = 0.5$, and $\Delta pK_{a,i}$ is obtained by subtracting $pK^{int}_{a,i}$ from pK_a according to eq 3.

The solution of eqs 7 and 8 is computationally demanding in the case of a large number of ionized residues; therefore, this description was combined with the effective charge approximation of Tanford and Roxby (*39*). In this hybrid approach (*40*, *41*), residues beyond a certain $R_s > R_{cutoff}$ were represented by a two-state charge model.

In this simplified model, the free energy of the charged state is expressed as

$$\Delta G(AH_i \rightarrow A^-) = \bar{q}_i\{W_i^0 + \sum_{i\neq j} W_{ij}\langle q_i\rangle\} \quad (9)$$

where $\bar{q}_i$ is the charge of the *i*th group in its ionized state (-1 or 1).

The average charge of the two-state charge model is calculated by

$$\langle q_i\rangle = \frac{q_i \exp\{-\Delta G(AH_i \rightarrow A_i^-)\beta\}}{\exp\{-\Delta G(AH_i \rightarrow A_i^-)\beta\} + 1} \quad (10)$$

The free energy of a given charge configuration by the hybrid approach can be obtained by combining eqs 7 and 9:

$$\Delta G^{m_s} = \sum_{i=1}^{N_s}\{q_i^{m_s}[-W_i^0 + \sum_{j\neq i}^{N_s} W_{ij} q_j^{m_s}] + q_i^{m_s}\sum_{j>N_s}^{N} W_{ij}\langle q_j\rangle\} \quad (11)$$

where m_s is the charge configuration of the residues within the cutoff radius, *N* is the total number of ionizable groups, and N_s is the number of groups within the cutoff range, for which the partition function is calculated self-consistently. In this work a 10 Å cutoff radius was used.

Models. The pK_a calculations were performed using three different models: (i) free enzyme (2END.pdb) (*14*), (ii) enzyme as in the complex, without the DNA (coordinates derived from 1VAS.pdb) (*16*), and (iii) enzyme in complex with DNA (1VAS.pdb). Since the available complex crystal structure is that of the E23Q inactive mutant (1VAS.pdb) (*16*), the model for our studies had to be constructed by replacing the glutamine by glutamate. Hydrogens were added to the PDB coordinates, and the water structures were rebuilt completely by the ENZYMIX program (*35*).

Calculation of Self-Energy. The three models which were described above were subjected to 2 ps relaxation (using a 0.5 fs step size), followed by 20 ps simulation using a 0.5 fs time step at 300 K with 0.3 kcal mol^{-1} Å^{-2} contraints on the protein and DNA atoms. The calculations were performed using the spherical boundary conditions by the SCAAS model (*42*). The side chains of the studied residues were included in region I, and the center of each system was taken as the center of region I. Region II included the rest of the protein and the DNA. This was immersed in a grid of Langevin dipoles of 18 Å radius, which was surrounded by bulk solvent represented by a dielectric continuum. To treat the long-range interactions properly, the local reaction field model was used (*43*). The pK^{int}_a results were obtained by averaging the PDLD/S energies of 10 configurations, which were collected at each 2 ps interval of the MD simulation. Since the goal of this work was to calculate the difference in solvation free energy in protein and aqueous environment rather than explore conformational equilibrium of the protein, the convergence criteria were set according to $\Delta\Delta G^{w\rightarrow p}$. In previous studies PDLD/S simulations were demonstrated to converge faster than all-atom LRA and FEP treatment (*44*, *45*). The PDLD/S calculations were performed by the

Table 1: pK_a Shifts of Arg-26 Due to the Presence of Arg-22 and Glu23 Obtained by Explicit PDLD/S Calculations and Macroscopic Approach

	free enzyme	complex
PDLD/S	0.33	0.27
macro ($\epsilon = 40$)	0.13	0.19
macro ($\epsilon = 30$)	0.18	0.52
macro ($\epsilon = 20$)	0.87	1.00
macro ($\epsilon = 40$)	2.00	3.98

program POLARIS 6.30 (*35, 45*), and the configuration generation for PDLD/S was done by ENZYMIX (*35*).

Calculation of Interactions with Other Ionizable Groups. The pK_a shifts due to other ionizable residues were calculated using eq 11 on the original models as well as on each configuration generated by MD simulation for evaluation of pK_a^{int} (see above). It is important to note that, for configuration generation, all ionizable groups were kept neutral with the exception of the one, for which pK_a^{int} was calculated. Thus, the relaxation reflects the structural changes due to reorganization of permanent dipoles around a particular charged group.

The effective dielectric constant for interactions between ionized groups was the subject of several previous studies (*46*−*49*). It was shown by both experimental and theoretical approaches that this term is large inside proteins (*50*−*53*). Recently, it was concluded that a more microscopic treatment results in larger ϵ_{ij}, especially if the reorganization of protein dipoles around the interacting ion pair is taken into account (*54*). Several studies demonstrated the robustness of using an effective dielectric constant of 40 to explain mutation effects (*55, 56*). On the basis of these findings, an ϵ_{ij} of 40 was used in our studies to calculate the interactions between ionized residues. The resulting pK_a shifts were in reasonable agreement with ΔpK_a between strongly coupled residues (e.g., Arg-22, Arg-26, and Glu-23) obtained by an explicit evaluation (*54*) as shown in Table 1. In the complex, where the pK_a shifts are of major importance, the error caused by using the macroscopic dielectric of 40 does not exceed 0.1 unit.

The ΔpK_a of the N-terminus due to interaction with other ionizable residues could not be calculated by eq 11, since this residue is not yet implemented in the Polaris 6.30 program. This technical problem was circumvented in the following way: First the ionization states of the nearby residues (within ~15 Å from the N-terminus) were determined as a sum of pK_a^{int} and ΔpK_a computed using eqs 4 and 11, respectively. In the free enzyme and in the complex enzyme model Arg-3, Arg-33, Arg-26, Glu-20, Glu-23, and Glu-120 were ionized, whereas in the full complex Glu-23 was considered neutral. To keep electroneutrality in the N-terminus environment, Asp-87 was ionized instead of Glu-23. The pK_a of the N-terminus was calculated using eq 4 in the presence of the above-mentioned charged residues with $\epsilon_{in} = 20$ to account for the relaxation of ionized groups and to obtain results comparable to those with the macroscopic treatment. The ΔpK_a is obtained by subtracting the intrinsic pK_a computed with uncharged ionizable groups from the pK_a obtained with ionized residues averaged over 10 MD generated configurations (according to eq 3).

The pK_a shifts due to the charges of DNA phosphates were calculated separately on the crystal structure of the complex using eq 4. As the phosphate groups are largely exposed to solvent, with the exception of those of the thymine dimer, only the influence of the latter two was taken into account. Neglecting the other ionized phosphate groups is based on the hypothesis that their charges are largely screened by the surrounding ions. The pK_a shifts of the active site residues due to interaction with DNA phosphates were obtained as a difference between pK_a's computed with charged and uncharged phosphates, respectively, in the presence of ionized Arg-3, Arg-22, Arg-26, and Glu-120 residues. Charging more groups did not result in considerable changes in pK_a's. Because of neglecting the charges of other DNA phosphates and also to compensate the missing effect of protein dipole relaxation, a high effective dielectric of 40 was used (*54*).

Table 2: Charges of TD 5′ Atoms Included in Region I To Calculate the Stability of the Protonated Thymine Dimer Intermediate[a]

	charge set 1		charge set 2	
	n	p	n	p
N1	−0.176	−0.176	−0.260	−0.136
C2	0.502	0.802	0.797	0.782
O2	−0.540	−0.340	−0.611	−0.501
HO2		0.500		0.512
N3	−0.330	−0.330	−0.755	−0.750
H3	0.286	0.286	0.422	0.435
C4	0.411	0.411	0.639	0.602
O4	−0.484	−0.484	−0.552	−0.454

[a] In charge set 1, the charges of the thymine dimer are taken from the Polaris standard charge set, and the protonation was represented in a simplified way. In charge set 2, charges were obtained by ESP calculations on neutral and protonated thymine dimers using the 6-31G* basis set. n and p correspond to neutral and protonated thymine, respectively. The complete set of charges is available upon request from osman@inka.mssm.edu.

Calculation of the Stability of Protonated 5′ Thymine. The stability of 5′ thymine in the neutral and protonated state was evaluated by comparing the solvation free energies in protein and in water, respectively. To assess the error range of the calculations, two charge sets were used. The charges of the relevant groups are displayed in Table 2. The complete set of charges is available from the authors upon request. Charge set 1 was derived from the original Polaris charge set (*35*), but in the protonated state the +1 charge was distributed on C2, O_2, and H atoms. Charge set 2 was obtained by fitting the electrostatic potential obtained from a single-point RHF calculation with a 6-31G* basis set, according to the Merz−Singh−Kollman scheme (*57, 58*). The calculation was performed using the whole thymine dimer in which the sugars were replaced by two methyl groups. The consistency with the Polaris model was maintained by modifying the charges of the N1 atoms in charge set 2 to compensate the difference between the charges of the methyl carbons and C1′ charges of the Polaris charge set (*35*). In this way the protonation of 5′ thymine only influenced the charges of the TD bases, not the sugar atoms. In this model the positive charge delocalized on both bases; therefore, the charges of the relevant atoms (shown in Table 2) reflect only the major changes.

The difference between the free energy of solvation in protein and in water ($\Delta\Delta G^{w\rightarrow p}$) was calculated by the PDLD/S method (see eq 4), averaging the results over

Table 3: Intrinsic pK_a's Calculated for (1) Free Enzyme, (2) Enzyme in Complex Conformation without the DNA, and (3) Complex[a]

	pK_a^{int}(1)		pK_a^{int}(2)		pK_a^{int}(3)	
	ur	rel	ur	rel	ur	rel
Arg-3	9.71	9.97 (0.49)	8.45	10.12 (0.42)	5.51	6.18 (0.94)
Arg-22	10.90	11.15 (0.20)	9.93	10.75 (0.22)	6.58	8.53 (0.30)
Arg-26	11.15	11.72 (0.12)	7.28	10.94 (0.60)	6.59	9.96 (0.50)
Glu-23	2.96	4.32 (0.44)	4.32	5.59 (0.53)	10.20	9.67 (0.74)
N-ter	7.00	7.03 (0.48)	6.80	6.50 (0.32)	1.84	1.88 (0.57)

[a] ur designates the results obtained on the original unrelaxed model; rel stands for the results obtained by averaging over 10 MD generated configurations. The standard deviations are displayed in parentheses.

configurations collected in 5 and 10 ps MD simulations sampling at 1 ps intervals. After 2 ps relaxation of the initial models, the MD simulations were run at a constant temperature of 300 K with a 1 fs time step using SCAAS boundary conditions. In the SCAAS model, region I comprised N1, C2, O2, HO2*, N3, H3, C4, and O4 atoms of the 5′ thymine of the TD, whereas Arg-26 was kept in region II.

RESULTS

The pK_a's of the ionizable residues are controlled by two main effects, the difference between the solvation by protein and water environments characterized by pK_a^{int} and the interaction with other charged groups in the protein reflected by ΔpK_a. To reveal the origins of the pK_a changes in the course of complex formation with the DNA, these effects were studied separately. Therefore, the results of pK_a calculations of the active site residues of EndoV are presented in three sections. First the pK_a^{int} values are described in the free enzyme, the complex enzyme model without DNA, and the full complex. The pK_a shifts induced by the surrounding ionized residues are analyzed in the next section. The overall changes in pK_a's of the active site residues are reported in the third part. This presentation of the results can lead to a better understanding of the effects controlling the pK_a of ionizable residues in the active site of endonuclease V.

pK_a^{int} *of the Active Site Residues*. The intrinsic pK_a's of the active site residues of endonuclease V calculated in the free enzyme, the complex enzyme model (without the DNA), and the entire complex are presented in Table 3. The convergence of the pK_a^{int} values computed for the complex as a function of the generated configurations is displayed in Figure 3. The results appear to reach convergence after approximately six configurations. The fluctuations in the pK_a^{int} values are less than 1 pH unit as indicated by the standard deviations shown in Table 3 in parentheses.

In the free enzyme the intrinsic pK_a's of residues located on the surface of the binding site are close to the experimental pK_a's determined in water (see pK_a^{int}(1) rel in Table 3). For example, pK_a^{int} of Glu-23 is 4.3, which is identical to its pK_a in water. Also, the pK_a^{int} of 11.7 for Arg-26 is close to its pK_a of 12.5 measured in water. Other residues partly buried in the protein show larger deviations from the pK_a in water. For example, the pK_a^{int} of Arg-3 is 2.5 units below the experimental value obtained in water. These results are a direct consequence of the definition of pK_a^{int} in eq 3, because $\Delta\Delta G^{w\rightarrow p}$ describes the difference in solvation of the

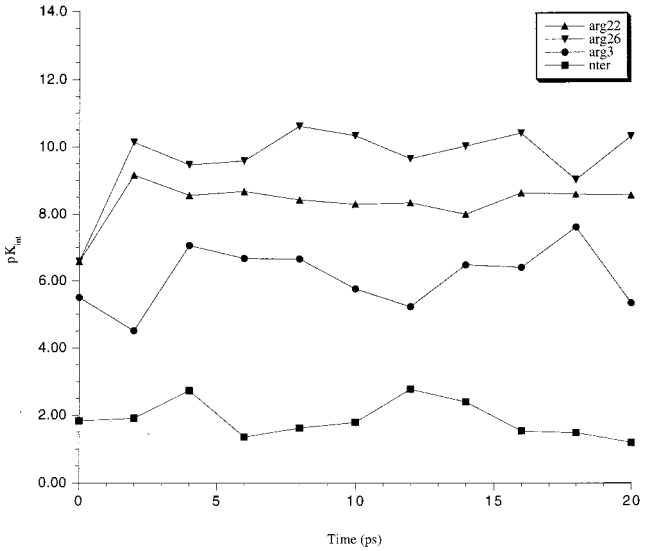


FIGURE 3: Intrinsic pK_a of the N-terminus, Arg-3, Arg-22, and Arg-26 computed in the complex as a function of the generated configurations.

residues in water and in protein. The differences between the pK_a^{int} values and the pK_a's measured in water increase with the conformational change induced in the protein by DNA binding, as represented by the artificial enzyme model, where the DNA was removed from the complex. The discrepancy between pK_a^{int} and $\Delta\Delta G_{self}^{w\rightarrow p}$(A→AH$^+$) in water increases more by the actual presence of the DNA in the complex.

In general, the intrinsic $\Delta\Delta G_{self}^{w\rightarrow p}$(AH → A$^-$), can be observed to decrease for Arg-3, Arg-22, Arg-26, and the N-terminus but to increase for Glu-23. According to eq 3 increase in pK_a^{int} corresponds to positive $\Delta\Delta G_{self}^{w\rightarrow p}$(AH → A$^-$), which means that the enzyme destabilizes the charging of that residue compared to water. The decrease in pK_a^{int} corresponds to negative pK_a^{int}, which is equivalent to positive $\Delta\Delta G_{self}^{w\rightarrow p}$(A → AH$^+$) for bases. For all studied residues it was found that the interactions with DNA result in positive $\Delta\Delta G^{w\rightarrow p}$ for the charged groups, indicating a less favorable solvation in protein than in water when the other ionizable residues are neutral. To find the rationale for this effect, the electrostatic free energy calculated in the protein with linear response approximation was decomposed into three terms: (i) interaction with the protein (or protein + DNA) dipoles, (ii) interaction with water dipoles (Langevin term), and (iii) bulk solvation energy (Born term). The results averaged over 10 configurations are presented in Table 4. Out of the three terms the Langevin term was found to be most sensitive to protein structural changes and to interactions with the DNA. Since the Langevin term is dominant in $\Delta\Delta G_{self}^{w\rightarrow p}$, its increase caused by substituting the water dipoles with less efficiently interacting protein and DNA dipoles has a significant impact on pK_a results.

Interestingly, the change in pK_a^{int} values does not correlate with the extent of structural rearrangement. Arg-22 and Arg-26 exhibited considerable distortion upon complexation with the DNA, decreasing the distance between their NH2 atoms from 7.0 to 4.2 Å. The pK_a^{int} of Arg-22 and Arg-26, however, changed only by 0.4 and 0.8 pH unit, respectively (see pK_a^{int}(1) rel and pK_a^{int}(2) rel in Table 3). In

Table 4: Decomposition of PDLD/S Solvation Free Energies Calculated in the Presence of Uncharged Ionizable Residues, Averaging over 10 MD Generated Configurations of Models: (1) Free Enzyme, (2) Enzyme in Complex Conformation without the DNA, and (3) Complex[a]

		$\Delta G^{p}_{self}(1)$	$\Delta G^{p}_{self}(2)$	$\Delta G^{p}_{self}(3)$
Arg-3	$V_{\mu q}$	−0.05	−0.16	−2.79
	V_L	−12.47	−12.73	−5.48
	V_{bulk}	−1.93	−1.88	−0.86
Arg-22	$V_{\mu q}$	1.66	1.49	−1.17
	V_L	−15.10	−14.61	−8.51
	V_{bulk}	−2.03	−2.06	−2.35
Arg-26	$V_{\mu q}$	0.18	−0.77	−3.72
	V_L	−14.93	−12.64	−9.95
	V_{bulk}	−2.04	−2.03	−0.50
Glu-23	$V_{\mu q}$	−3.11	−3.22	−3.12
	V_L	−13.62	−12.00	−5.25
	V_{bulk}	−2.00	−1.97	−3.37
N-ter	$V_{\mu q}$	−0.32	1.48	−3.72
	V_L	−15.35	−16.15	−7.36
	V_{bulk}	−1.97	−1.98	−0.58

[a] $V_{\mu q}$ is the interaction between the charges of the studied group with the protein permanent dipoles, V_L is the interaction with water dipoles (Langevin term), and V_{bulk} is the bulk solvation energy.

contrast, the pK_a^{int} of Glu-23 increased by much more, although the conformation of this residue is not affected by DNA binding in our model. The movement of the arginines, however, involved the displacement of two water molecules, which were previously hydrogen bonded to Glu-23. The exclusion of these water molecules decreased the V_L term of Glu-23 by 1.62 kcal/mol (see Table 4), which was not compensated by the interactions with the protein dipoles, leading to an increase of pK_a^{int} of Glu-23 by 1.3 units. This example illustrates that different protein conformations do not necessarily lead to large differences in pK_a^{int}. Rather, a more important change is due to the replacement of waters, which cannot be predicted from structural changes alone. This emphasizes the importance of simulations as a necessary step in evaluating pK_a changes in proteins or protein−DNA complexes.

Intrinsic pK_a's computed in the crystal structure of the free enzyme and in the complex enzyme model (corresponding columns $pK_a^{int}(1)$ ur and $pK_a^{int}(2)$ ur in Table 3) indicated large changes in pK_a^{int} due to conformational rearrangement induced by DNA binding. Arg-26 exhibited a particularly large decrease in pK_a^{int} by 3.9 pH units. However, when the relaxation of the local dipoles upon ionization was taken into account (here it was done by the LRA approach) the pK_a^{int}'s in different conformations became more similar. This demonstrates that a large influence of protein conformation on pK_a can come from the improper treatment of the protein relaxation (*45*).

Intrinsic pK_a's especially of the N-terminus and Glu-2, were affected dramatically by the presence of the DNA (see $pK_a^{int}(3)$ ur in Table 3). In the complex the pK_a^{int} of the N-terminus is 5.1 pH units lower than in the free enzyme, corresponding to 7.1 kcal/mol destabilization of the charged state. For Glu-23 the intrinsic pK_a increased by 5.3 pH units, which means that ionizing this residue is less favorable by 7.4 kcal/mol in the complex than in the free enzyme. In the case of Arg-22 and Arg-26, the relaxation of the complex structure greatly reduced the destabilization of the charged state caused by DNA binding, resulting in 2.6 and 1.8 pH

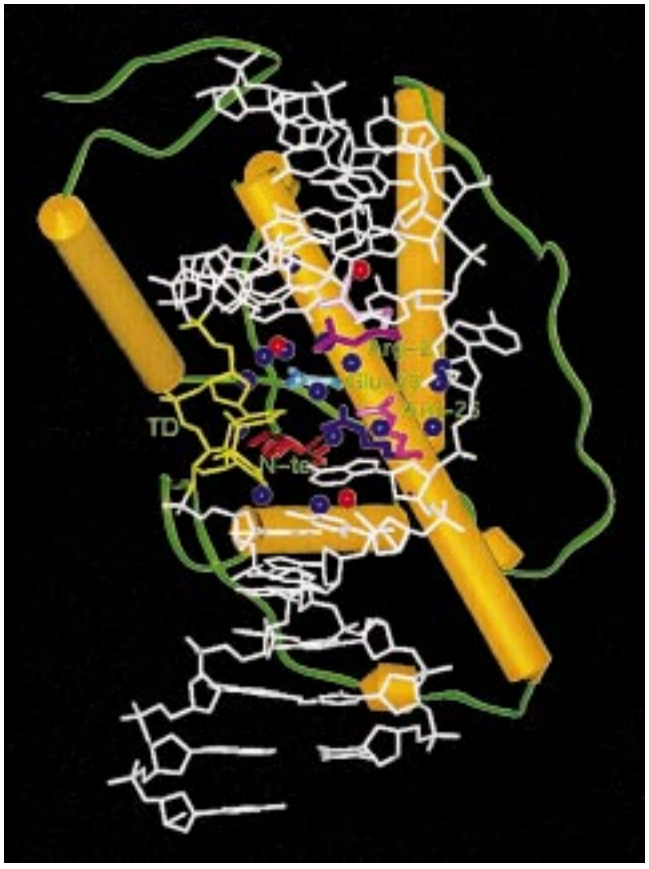


FIGURE 4: Replacement of water molecules upon complex formation. The secondary structure of endonuclease V is represented by helices and ribbons. The DNA is shown in white, and the thymine dimer lesion is colored yellow. The active site residues of the free enzyme (2END.pdb) and the complex (1VAS.pdb) are superimposed on each other and displayed by sticks. The free enzyme residues are in lighter colors, and those of the complex enzyme are in darker colors. The N-terminus is displayed in red, Glu-23 in blue, Arg-26 in magenta, and the Arg-22 in pink. The crystallographic water molecules within 5 Å from the active site residues in the free enzyme are colored blue and in the complex red. Upon complexation with the DNA, 12 water molecules were excluded from the active site, resulting in a significant decrease in the self-energy (due to the reduction in the Langevin term).

unit decreases in pK_a^{int}, respectively, compared to the free enzyme (see $pK_a^{int}(1)$ rel and $pK_a^{int}(3)$ rel in Table 3). The $pK_a^{int}(3)$ rel of Arg-3, which directly interacts with the DNA, increased by a small extent upon relaxation, giving rise to a 3.8 pH unit decrease in its $pK_a^{int}(3)$ rel upon complex formation. Inspection of the Langevin component of the solvation energy helps to explain these observations. The V_L terms increased significantly for all active site residues upon complex formation (see $\Delta G^{p}_{self}(1)$ and $\Delta G^{p}_{self}(3)$ in Table 4). This correlates with the change in the number of water molecules in the enzyme active site, located within 5 Å of the catalytic residues. Due to conformational changes in the protein induced by DNA binding, two water molecules were displaced by Arg-22 and Arg-26. The association with the DNA resulted in displacement of 10 additional water molecules from the active site, leaving only 3 of them in the close environment of the catalytic residues (Figure 4). The exclusion of the total 12 water molecules from the active site due to DNA binding resulted in a significant increase in the Langevin term in the complex compared to the free enzyme: 7.0 kcal/mol for Arg-3, 6.6

Table 5: Interaction between Ionizable Residues in (1) Free Enzyme, (2) Enzyme in Complex Conformation without the DNA, and (3) Complex[a]

	$\Delta pK_a(1)$		$\Delta pK_a(2)$		$\Delta pK_a(3)$		$\Delta pK_{a\ DNA}$
	ur	rel	ur	rel	ur	rel	ur
Arg-3	1.02	1.50	0.13	0.37	0.13	1.02	1.48
Arg-22	1.02	1.02	1.02	0.50	1.02	0.99	0.62
Arg-26	1.02	1.02	0.08	1.50	0.08	1.50	0.56
Glu-23	0.00	−2.80	0.00	−2.80	0.00	−2.80	0.95
N-ter	0.66	0.98	0.50	1.19	3.55	3.79	0.85

[a] ur designates the results obtained on the X-ray structure; rel stands for the results obtained by averaging over 10 configurations. The ΔpK_a of the N-terminus was calculated using eq 4.

kcal/mol for Arg-22, 5.0 kcal/mol for Arg-26, 8.1 kcal/mol for Glu-23, and 8.0 kcal/mol for the N-terminus. Although the interaction with the permanent dipoles displayed the opposite trend, the change in the $V_{\mu q}$ term is smaller than that of the Langevin term. This behavior of the V_L and $V_{\mu q}$ terms is the consequence of replacing the unconstrained water dipoles that solvate the free protein with a set of DNA dipoles constrained by a relatively rigid helix structure. This again emphasizes the importance of reorganization of local dipoles around a particular ionized group in the course of charging.

Interaction between Ionizable Residues. The pK_a shifts due to interactions with other ionizable residues are presented in Table 5. The pK_a shifts computed in the crystal structures are referred to as unrelaxed (ur) ΔpK_a, whereas the results averaged over 10 configurations are referred to as relaxed (rel) ΔpK_a. Since the protein coordinates are identical in the entire complex and the complex enzyme model from which the DNA was removed, the results on unrelaxed structures are equivalent in the two cases. The effect of DNA charges were calculated separately (see Methods) and are displayed as $\Delta pK_{a\ DNA}$. In the case of $pK_a(3)$ rel, the presence of the DNA only influenced the relaxation process.

Although some pK_a shifts are particularly large, especially those of the N-terminus and Glu-23, the corresponding ΔG_{ij}^{p} interaction energies are smaller in each case than the $\Delta\Delta G_{self}^{w\rightarrow p}$ self-energies, which can be obtained from pK_a^{int}'s as defined in eq 3 (compare Table 3 and Table 5).

The pK_a shifts were not influenced to a large extent by structural changes in the protein induced by DNA binding, as shown by the result obtained on unrelaxed structures of the free enzyme and the complex enzyme model (see $\Delta pK_a(1)$ ur and $\Delta pK_a(2)$ ur in Table 5). For example, while Arg-22 and Arg-26 approached each other by almost 3 Å during complex formation, the ΔpK_a of Arg-26 decreased by 1 pH unit and the ΔpK_a of Arg-22 remained unchanged. Similarly to Arg-22, the pK_a shift of Glu-23 also remained unaffected by the protein conformational rearrangement. It suggests that Glu-23 shields Arg-22 from the unfavorable approach of Arg-26 by compensating the repulsion between the two arginines. The pK_a shift of Arg-3 decreased by 0.9 pH unit due to a conformational change of Lys-121, which brings the two positively charged groups close to each other.

Protein relaxation increased the interaction term with other ionized residues, since the reorganization of dipoles around a single charged group always lowers the solvation energy. The inclusion of the DNA in the relaxation process considerably increases the pK_a shift of the interacting residues: Arg-3, Arg-26, and especially the N-terminus, which changed by 2.6 pH units in the presence of the DNA (see $\Delta pK_a(2)$ rel and $\Delta pK_a(3)$ rel in Table 5). Ionizing the TD phosphates had a large influence of 1.5 pH units on the pK_a shift of Arg-3, which is in direct contact with the 5′ phosphate of these groups. The other active site residues, which are beyond 5 Å from the phosphates, had weaker interactions, resulting in less than 1 pH unit increase in ΔpK_a.

Table 6: pK_a Values for Active Site Residues of Endonuclease V in (1) Free Enzyme, (2) Enzyme in Complex Conformation without the DNA, and (3) Complex[a]

	$pK_a(1)$	$pK_a(2)$	$pK_a(3)$
Arg-3	11.47	10.49	8.68
Arg-22	12.17	11.25	10.14
Arg-26	12.74	12.44	12.02
Glu-23	1.52	2.79	7.82
N-ter	8.01	7.69	6.52

[a] pK_a was calculated according to eq 2 as a sum of the self-energy and pK_a shift due to interaction with other ionizable groups.

Overall pK_a Values. The pK_a's constructed from the sum of the corresponding pK_a^{int} and ΔpK_a values calculated on the relaxed structures and the pK_a shifts induced by the interactions with DNA phosphates are presented in Table 6. Since the N-terminus and Glu-23 are critical in the glycosylase step, their pK_a's should be of primary importance to the reaction mechanism. The pK_a of the amino terminus was found to decrease by 1.5 pH units upon DNA binding. The pK_a of 6.5 in the complex predicts the presence of a neutral amino terminus in the pH range of 6−8.5 where the glycolysis reaction is most efficient. This finding is consistent with the role of the amino terminus as the nucleophile in the glycosylase step. Remarkably, the pK_a of Glu-23 has increased in the complex by 6.3 pH units compared to the free enzyme. This is primarily the consequence of the increase in self-energy ($\Delta\Delta G_{self}^{w\rightarrow p}$) to more positive values caused by the loss of interactions with the surrounding water molecules, which could not be compensated by interactions with the protein and DNA dipoles. The negatively charged phosphates of the DNA were found to have a much smaller effect on the pK_a of Glu-23 than the water dipoles. The pK_a of 7.8 suggests that a neutral Glu-23 is present in the complex in the beginning of the glycosylase step.

In the course of complex formation the pK_a's of Arg-3 and Arg-22 were reduced by 2.8 and 2.0 pH units, respectively. The conformational rearrangement is responsible for 1.0 pH unit decrease, whereas the interactions with the DNA account for the rest of the change. The major part of the pK_a reduction upon complex formation is due to burying Arg-3 and Arg-22, which is only partially compensated by the interactions with ionizable groups of the protein and DNA. The change in the pK_a of Arg-26 due to DNA binding is also to lower values, albeit only by 0.7 unit compared to its value in the free enzyme. The pK_a of 12 obtained in the complex makes it unlikely for this residue to donate a proton to the 5′ TD, contradicting the hypothesis that Arg-26 would act as a general acid in the glycolysis reaction.

Stability of Protonated 5′ Thymine. The glycosylase step—as a nucleophilic substitution reaction—can follow either an S_N1 or an S_N2 mechanism. In a monomolecular reaction (S_N1) the protonation of the base leads to a fast cleavage of the N1−C1′ bond, followed by the attack of the nucleophile

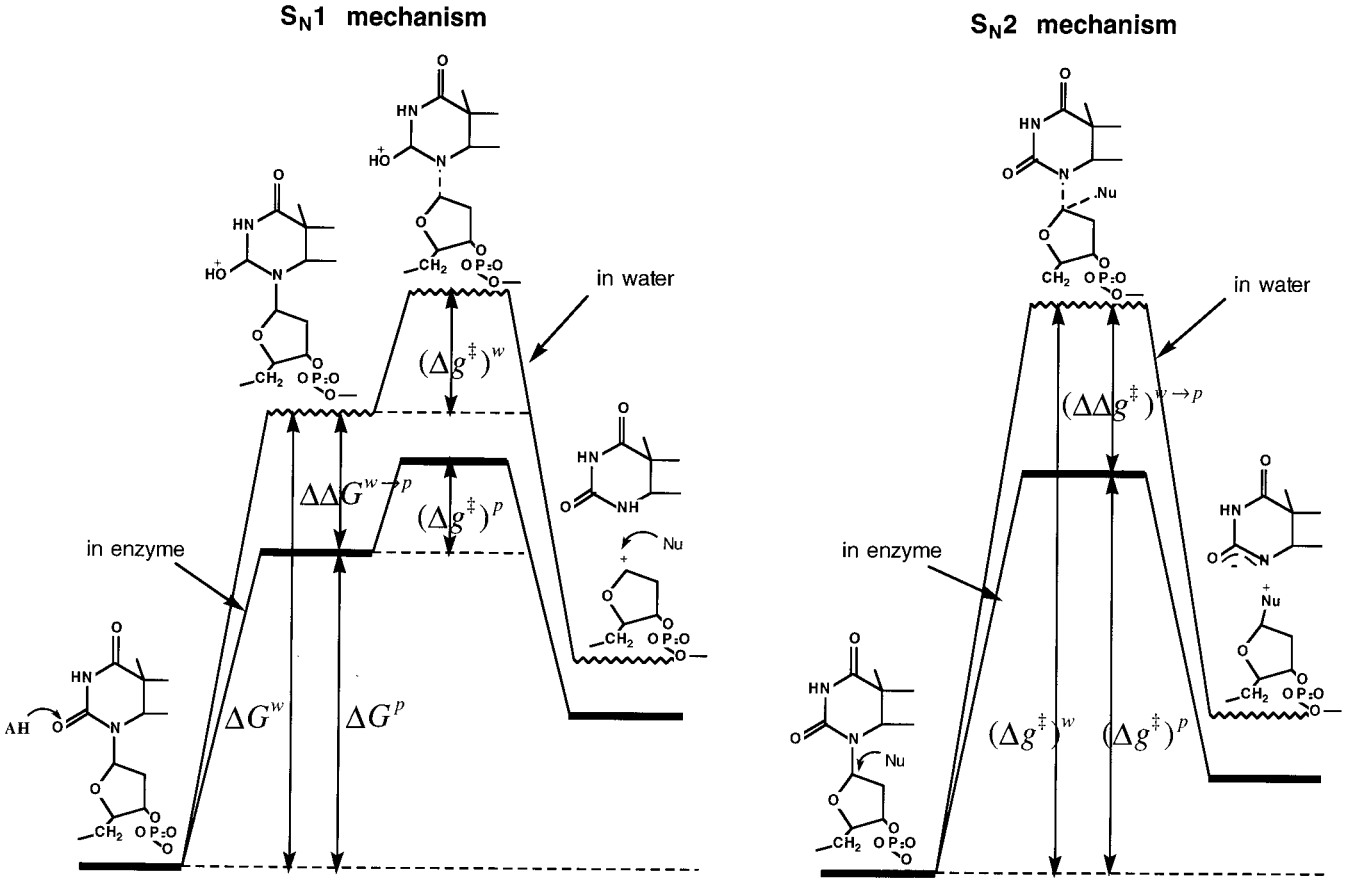


FIGURE 5: Qualitative energy diagrams of the S_N1 and S_N2 mechanisms of the glycosylase step, with respect to water. ΔG^{w} and ΔG^{p} are solvation free energies of the protonated thymine in water and in protein, respectively. The free energy required to transfer the protonated thymine from water to the protein site is denoted by $\Delta\Delta G^{w\to p}$. The activation energy of breaking the glycosidic bond by either mechanism is $(\Delta g^{\ddagger})^{w}$ in water and $(\Delta g^{\ddagger})^{p}$ in protein, and the difference of the two energy barriers is denoted by $(\Delta\Delta g^{\ddagger})^{w\to p}$. The enzyme is known to reduce the overall barrier of the reaction by ~11 kcal/mol compared to water.

on the resultant positively charged C1′ atom. In an S_N2 reaction the attack of the nucleophile initiates the dissociation of the base, resulting in a pentacoordinate transition state around C1′. The schematic energy diagrams of these two mechanisms in protein and in water are presented in Figure 5. The rate constant of the glycolysis reaction by endonuclease V is 7.2×10^{-2} s^{-1} (*11*), which corresponds to an activation energy of 19.2 kcal/mol. The rate constant of the reference reaction can only be approximated from the neutral hydrolysis of thymine in water as 5.8×10^{-10} s^{-1} (*59*), corresponding to an energy barrier of 30.4 kcal/mol. According to these data, the enzyme accelerates the glycosylase reaction by 8 orders of magnitude, corresponding to a reduction of ~11 kcal/mol in the energy barrier compared to an aqueous environment. In the case of an S_N2 reaction this is due to a reduction in the $[(\Delta\Delta g^{\ddagger})^{w\to p}]$ activation energy of breaking the glycosidic bond, while in the case of an S_N1 reaction this is composed of two terms: a decrease in the free energy of formation $(\Delta\Delta G^{w\to p})$ of the protonated TD and a reduction in the activation energy $(\Delta\Delta g^{\ddagger})^{w\to p}$ of the scission of the glycosidic bond. If the glycosylase reaction were to take place by an S_N1 mechanism, the environment must stabilize the formation of the protonated intermediate; therefore, the $\Delta\Delta G^{w\to p}$ term must be negative. Consequently, based on the evaluation of the $\Delta\Delta G^{w\to p}$ term, it is be possible to distinguish between the two enzymatic mechanisms.

The results of the $\Delta\Delta G^{w\to p}$ calculations are summarized in Table 7. The $\Delta\Delta G^{w\to p}$ was evaluated in three states (see columns i, ii, and iii in Table 7) corresponding to two different protonation schemes: (i) neutral thymine and charged Arg-26 (reactant state model), (ii) 5′ thymine protonated on O2 and neutral Arg-26 (representing the case when Arg-26 is the proton donor), and (iii) both thymine and Arg-26 positively charged (corresponding to a proton donation from another residue, e.g., Glu-23). The solvation free energy in the enzyme and in water was computed according to the termodynamic cycle presented on Figure 2. The effect of thymine charges on the results was studied by using two different charge sets, as described in Methods. Simulations 1−3 were performed with charge set 1, whereas simulations 4−9 were carried out using charge set 2 (see Table 7).

Several initial conditions were used to assess the error range of the calculations. The effect of the ionized residues was investigated by charging three different sets of ionizable groups in calculating $\Delta\Delta G^{w\to p}$. In the minimal model only Arg-26 was considered ionized in the protein (see simulations 1, 4, and 7 in Table 7). In the second set Arg-3, Arg-22, Arg-26, Glu-20, and Glu-23 residues were charged (simulations 3, 6, and 9 in Table 7), while in the third case according to the results of the pK_a calculations, all the active site residues were ionized with the exception of Glu-23 (simulations 2, 5, and 8 in Table 7). In each set Arg-26 was ionized

Table 7: $\Delta\Delta G^{w\rightarrow p}$ Calculated for (i) Neutral Thymine, (ii) Protonated Thymine and Neutral Arg-26, and (iii) Protonated Thymine in the Presence of Charged Arg-26[a]

simul	charge set	time (ps)	no. of config	residues ionized in region 2	i	ii	iii
1	1	5	5	Arg-26*	−1.77	3.93	7.69
2	1	5	5	Arg-3, Arg-22, Arg-26*, Glu-20	−3.85	8.45	11.76
3	1	5	5	Arg-3, Arg-22, Arg-26*, Glu-20, Glu-23	−0.89	3.90	7.04
4	2	5	5	Arg-26*	−1.22	2.76	5.63
5	2	5	5	Arg-3, Arg-22, Arg-26*, Glu-20	−3.65	6.26	8.01
6	2	5	5	Arg-3, Arg-22, Arg-26*, Glu-20, Glu-23	−0.93	3.91	5.02
7	2	10	10	Arg-26*	−1.41	2.98	5.73
8	2	10	10	Arg-3, Arg-22, Arg-26*, Glu-20	−3.60	6.56	8.29
9	2	10	10	Arg-3, Arg-22, Arg-26*, Glu-20, Glu-23	−0.87	3.82	5.33

[a] N1, C2, O2, HO2*, N3, H3, C4, and O4 atoms of 5′ thymine were included in region I, whereas Arg-26 Was Kept in region II. The length of the trajectory and the number of configurations used for averaging the results are indicated in the Table. Arg-26* Means that Arg-26 is only ionized in columns i and iii.

Table 8: $\Delta\Delta G^{w\rightarrow p}$ Computed for (i) Neutral Thymine, (ii) Protonated Thymine and Neutral Arg-26, and (iii) Protonated Thymine in the Presence of Charged Arg-26 without Relaxation

simul	charge set	residues ionized in region 2	i	ii	iii
1	1	Arg-26*	−5.64	21.59	35.90
2	1	Arg-3, Arg-22, Arg-26*, Glu-20	−7.49	26.87	40.94
3	1	Arg-3, Arg-22, Arg-26*, Glu-20, Glu-23	−1.99	17.92	32.22
4	2	Arg-26*	−5.46	16.54	27.97
5		Arg-3, Arg-22, Arg-26*, Glu-20	−7.22	20.30	31.55
6	2	Arg-3, Arg-22, Arg-26*, Glu-20, Glu-23	−1.80	14.95	26.37

in two cases, when the 5′ TD was neutral and when both 5′ TD and Arg-26 were charged (columns i and iii in Table 7), according to the protonation schemes described above. The solvation free energies were averaged over 5 and 10 configurations generated in MD runs.

The neutral thymine and charged Arg-26 pair is more stable in the enzyme than in water by any simulation conditions as can be seen by the negative $\Delta\Delta G^{w\rightarrow p}$. The most favorable environment for this reactant state model is when the neighboring residues are ionized and Glu-23 is neutral. Hence, the solvation free energies are in agreement with the pK_a calculations supporting the case that Glu-23 is neutral in the reactant state of the glycosylase reaction. The $\Delta\Delta G^{w\rightarrow p}$ values show that the formation of a protonated 5′ thymine and neutral Arg-26 intermediate is unfavorable in the enzyme by 2.8−8.4 kcal/mol. The other possible intermediate, in which both thymine and Arg-26 are charged, is even more destabilized by 5.0−11.7 kcal/mol in the enzyme than in aqueous environment. This is not surprising, since in the complex structure these two positively charged groups are only separated by ∼2.5 Å from each other.

Ionization states of the active site residues significantly influence the $\Delta\Delta G^{w\rightarrow p}$ results. The largest differences between the solvation free energies of the protonated thymine intermediates in enzyme and in water were obtained in the presence of charged active site residues and neutral Glu-23. In this environment charging of the 5′ TD requires 6.3−8.5 kcal/mol more free energy in enzyme than in water (see state (ii) in simulations 2, 5, and 8 in Table 7). $\Delta\Delta G^{w\rightarrow p}$ values computed with the other sets of ionized residues indicated less destabilization of the protonated states. However, regardless of the charge set or the environment, the protonated thymine intermediates were unambiguously found to be less favored in the enzyme than in an aqueous environment by at least ∼3 kcal/mol. The simplified representation of the protonation by using charge set 1 overestimated the destabilization effect by ∼2−3 kcal/mol, but the results are consistent with those obtained using a more realistic charge set 2. Increasing the simulation length resulted in minor changes in the averaged $\Delta\Delta G^{w\rightarrow p}$ values.

The free energies required to transfer neutral and protonated models from water to a rigid protein are presented in Table 8. The $\Delta\Delta G^{w\rightarrow p}$ results obtained on unrelaxed models compared to the relaxed data (Table 7) demonstrate the significance of reorganization of protein dipoles around a given charged group. In the case of protonated 5′ thymine intermediates the protein relaxation can reduce the solvation energy by as much as 20 kcal/mol.

The results of stabilization calculations for protonated thymine dimer intermediates established that the protein presents an environment which opposes the proton transfer to the 5′ thymine of the TD. It thus appears that the role of Arg-26 is to provide an electrostatic stabilization of the leaving group rather than an actual proton-donating role.

DISCUSSION

The calculation of pK_a's in proteins is fundamental in simulations of enzymatic reactions and in understanding the molecular basis of enzymatic catalysis. The electrostatic mechanisms that determine the charged state of a protein group are the same as those that stabilize the transition state of an enzymatic reaction. The protein environment consists of dipoles, which are less polarizable than water but are preoriented to solvate the transition state of the particular enzymatic reaction or control the ionization state of a functional residue. The results presented in this work not only illustrate the complexity of such mechanisms but also elucidate the major unifying concepts derived from electrostatics.

Although the importance of the pK_a calculations is widely accepted, to our knowledge it has not yet been applied to protein−DNA complexes. This is partly due to experimental difficulties in measuring pK_a values in these systems.

Changes in pK_a values in enzymes when they interact with DNA are particularly interesting, because the high density of negative charge on DNA and the burying of the residues in the enzyme−DNA surface could significantly affect the ionization states of protein residues and hence modify their role in catalysis. In this work we have computed pK_a's for catalytic site residues of endonuclease V with the semi-microscopic approach, which enabled us to analyze different effects in pK_a changes in a consistent way.

The evaluation of the self-energies (pK_a^{int}) was demonstrated as having a crucial role in pK_a calculations. The free energy of charging the active site residues, the N-terminus, Glu-23, Arg-3, Arg-22, and Arg-26 of endonuclease V in the neutral protein, was found to increase upon DNA binding due to both conformational changes in the enzyme and interactions with the DNA. The decomposition of the self-energies showed that the loss of interaction with the water dipoles is responsible for most of this effect. The reduction of the Langevin term is especially dramatic for the catalytically indispensable residues: the N-terminus and Glu-23. Comparison of the results in the unrelaxed crystal structures to those obtained by averaging over MD generated configurations illustrated that large deviations in self-energies in different protein conformations can be the consequence of improper treatment of protein relaxation. This includes structural relaxation and, more importantly, reorganization of dipoles around the charged group.

The interaction energies with other ionizable groups were calculated with a macroscopic approximation, since the evaluation of the charge−charge interactions with a microscopic approach would have been an expensive proposal. The value of the chosen effective dielectric constant was validated by a number of test calculations as well as by computing the ΔG_{ij} term explicitly for the strongly coupled residues. The interaction energies derived from the resulting pK_a shifts are smaller than the corresponding self-energies, emphasizing the importance of a consistent evaluation of the intrinsic pK_a's.

The pK_a's of the N-terminus and Glu-23 were most substantially influenced by the presence of other ionizable residues; the pK_a of the N-terminus was increased by 3.8 pH units and that of Glu-23 decreased by 2.8 pH units. The conformational change of Arg-22 and Arg-26, induced by DNA binding, was found not to have a major effect on the pK_a shifts. As expected, the charges of the TD phosphates increased the pK_a of all active site residues, most considerably that of Arg-3 by 1.5 pH units.

The pK_a's of Arg-3, Arg-22, Arg-26, and the N-terminus in the complex decreased, while that of Glu-23 increased compared to the results in the free enzyme. Thus, charging of these residues is less favorable in the complex than in the free enzyme. The increase in $\Delta\Delta G^{w\rightarrow p}$ is mostly due to the displacement of 12 water molecules from the catalytic site in the course of DNA binding. The reduction of the interaction with the solvent Langevin dipoles was not compensated by the interactions with the dipoles and charges in the protein and the DNA. This can be rationalized by the restrained orientation of the DNA dipoles resulting in a significant increase in self-energies. The conformational rearrangement of Arg-22 and Arg-26 at the active site due to DNA binding does not have a major effect on the pK_a's. The pK_a shift due to this structural rearrangement is actually caused by the displacement of two water molecules, which increases the Langevin term.

The calculated pK_a's indicate that in the free enzyme at neutral pH the N-terminus is protonated, while in the complex the pK_a is lowered by 1.5 pH units, resulting in a neutral group. The presence of the neutral N-terminus in the complex is in agreement with experimental results, which identified it as the nucleophile attacking the C1′ of the 5′ TD sugar. The other indispensable element of the glycosylase step, Glu-23, exhibits a remarkable pK_a increase by more than 6 pH units upon DNA binding. This suggests that Glu-23 is not ionized in the reactant state of the glycosylase reaction. The solvation free energy obtained for the neutral 5′ thymine with different ionized residues also confirmed this result, since the environment containing a neutral Glu-23 provided a better stabilization for the substrate than the one with charged Glu-23.

In the crystal structure of the E23Q mutant of endonuclease V the NE2 of Gln-23 is 2.7 Å from the O4′ of the sugar of 5′ TD. If Glu-23 is protonated in the complex of the native enzyme, this hydrogen bond can be assumed to exist also in the wild-type enzyme. The hydrogen bond between the sugar of 5′ TD and Glu-23 can increase the susceptibility of C1′ of the 5′ TD sugar to nucleophilic attack. Furthermore, the hydrogen bond between Glu-23 and O4′ of 5′ TD might help to stabilize the transition state, and hence it accelerates the reaction by reducing the partial negative charge on the C1′ atom. The protonated Glu-23 is a likely candidate to act as a general acid by donating a proton to O4′ to induce the opening of the sugar ring. Although we did not calculate the pK_a of this intermediate of the reaction, the mutagenesis data seem to support our hypothesis. The Asp-23 mutant of endonuclease V was found to have negligible glycosylase activity (*13*). Since Asp is a shorter residue than Glu, it could be geometrically incompetent to protonate the sugar ring. Moreover, the effect of the protein and DNA may not raise its pK_a enough to become neutral in the complex. On the basis of calculated results and the above considerations, Glu-23 is proposed to have a triple role in the glycosylase step of thymine dimer repair: (i) increasing the electrophilicity of C1′ and thus stabilizing the transition state, (ii) acting as a general acid to induce ring opening of the 5′ TD sugar, and (iii) stabilizing the product, the protonated Schiff base, of the reaction by an ionic interaction.

Arg-3, Arg-22, and Arg-26 considered in this study are less critical elements of catalysis as shown by mutagenesis data (*11*). Although burying these groups upon DNA binding decreased their pK_a's, these groups remain positively charged in the complex in the optimal pH range of the glycosylase reaction. The pK_a of Arg-26 changed only by 0.7 pH unit compared to the free enzyme. This is particularly interesting, since this residue is located in the appropriate position to donate a proton to O2 of 5′ thymine of TD. The calculated pK_a clearly indicates that Arg-26 is unlikely to participate in such protonation process. Moreover, the positive charge of this residue is important since it can contribute to the stabilization of the developing negative charge on the 5′ thymine in the transition state. This is in agreement with experimental data that the substitution of Arg-26 by a neutral residue results in a rate decrease of the glycosylase reaction by 2 orders of magnitude (*11*). Similarly to Arg-26, Arg-22

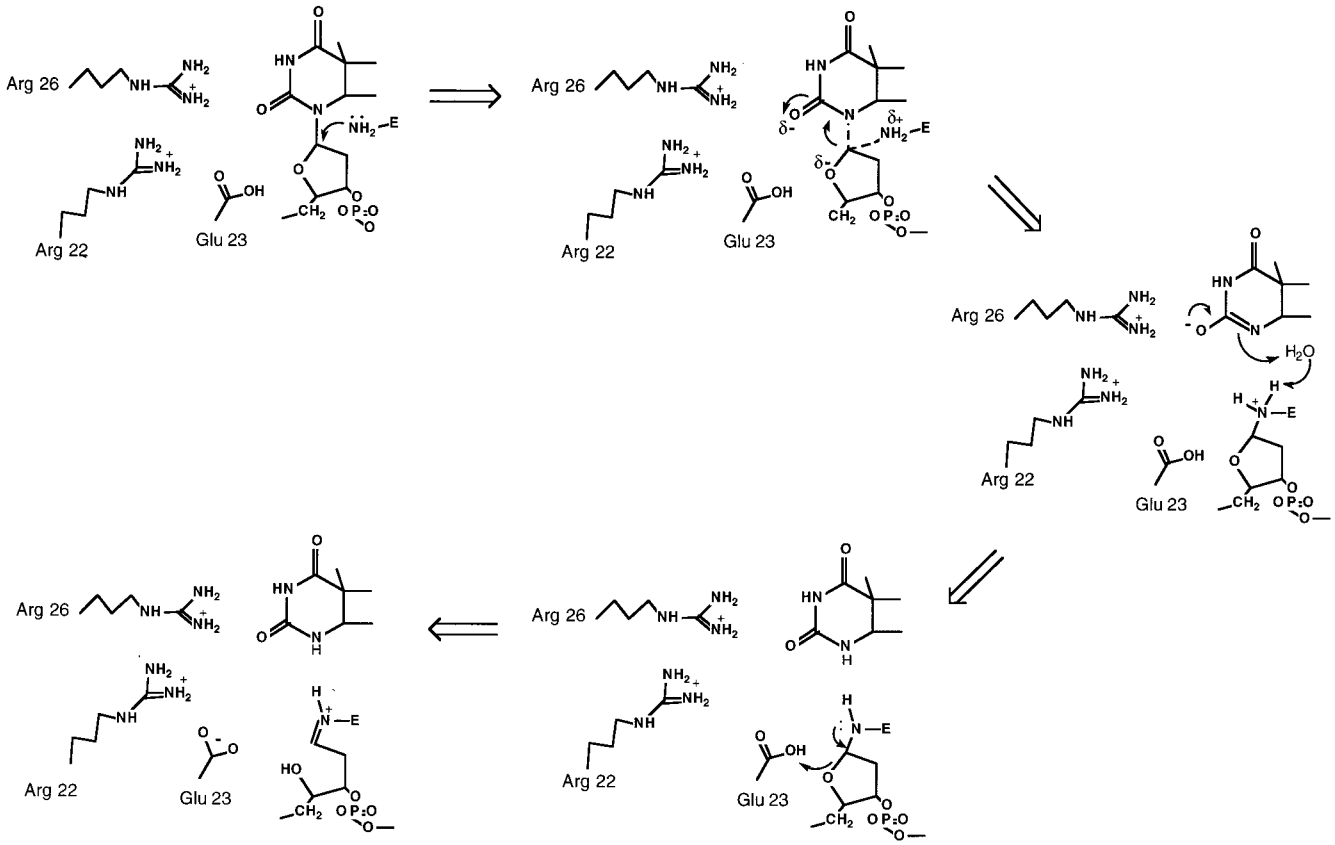


FIGURE 6: Proposed reaction mechanism of the glycosylase step of endonuclease V based on the present work. The reaction follows an S_N2 pathway; the glycosidic bond breaks upon the nucleophilic attack of the N-terminus. The transition state is stabilized by a hydrogen bond between O4′ of the thymine sugar and the neutral Glu-23 as well as by the positively charged Arg-22 and Arg-26. The nearby arginines also provide a favorable electrostatic environment for the leaving base. The ring opening of the sugar is induced by a proton donation to O4′ from Glu-23. The proton from the N-terminus is shuttled to the 5′ base of TD by the assistance of a proposed water molecule. The product of the reaction is a protonated Schiff base, which is stabilized by the negatively charged Glu-23.

also plays a role in providing a favorable electrostatic environment for the leaving base. The replacement of Arg-22 by a neutral residue, Gln, reduces the k_{cat} by ~10-fold (*11*).

The S_N1 mechanism of the glycosylase step was probed by calculating the stability of two protonated thymine intermediates in the enzyme compared to aqueous environment. The free energy required to transfer a protonated thymine from water to the protein site indicates that charging 5′ TD is less favorable in the enzyme than in water. This result is consistent with a lower proton affinity of thymine and the high pK_a of Arg-26, which excludes the proton donation by this residue. The protonation by another residue, e.g., Glu-23, is even less likely due to the proximity of the positive charge of the guanidino group of Arg-26 to the $O2H^+$ of 5′ TD.

Endonuclease V reduces the energy barrier of the glycosylase reaction by ~11 kcal/mol compared to the same reaction in water. If the glycosylase step were to follow the S_N1 pathway, the protonated thymine intermediate should be stabilized by ~10 kcal/mol compared to an aqueous environment (see considerations in Results). The protonated 5′ thymine, however, was found to have a positive $\Delta\Delta G^{w\rightarrow p}$ at all simulation conditions used in this study; therefore, it is less stable in the enzyme than in water. The difference between the solvation free energies is ~6.6 kcal/mol in the most realistic protein environment. If the glycosylase reaction were to proceed by an S_N1 mechanism, the enzyme should be able to reduce the activation energy of breaking the glycosidic bond by ~17.6 kcal/mol to compensate for the increase in $\Delta\Delta G^{w\rightarrow p}$ (see Figure 5). Since this would be inconsistent with the optimization of the reaction by endonuclease V by 11 kcal/mol, the glycosylase step is unlikely to take place by an S_N1 mechanism.

Instead, we propose that the reaction proceeds by an S_N2 mechanism; a detailed scheme is displayed in Figure 6. As previously suggested by experimental results, the neutral N-terminus performs the nucleophilic attack on C1′ of the 5′ thymine of the dimer. No protonation on O2 of 5′ TD precedes this reaction step. The susceptibility of C1′ to nucleophilic attack is actually enhanced by a hydrogen bond between Glu-23 and O4′ of the sugar. This hydrogen bond helps to stabilize the transition state as well. The positively charged guanidino groups of Arg-22 and Arg-26 are important in providing a favorable electrostatic potential for the leaving base. Molecular dynamics simulations of the complex indicated a presence of a water molecule in the proximity of the N-terminus and the base of 5′ thymine (to be published in a subsequent paper). We suggest that this water molecule shuttles the proton from the N-terminus to the negatively charged leaving base. Protonation of the O4′ sugar atom by Glu-23 induces the opening of the sugar ring. This leads to the formation of a protonated Schiff base, a product of the glycosylase reaction, which is stabilized by the nearby negatively charged Glu-23. The described mechanism is in full accord with the results of mutagenesis experiments.

ACKNOWLEDGMENT

M. Fuxreiter thank Z. T. Chu and Y. Y. Sham for helpful discussions and M. Mezei for critical reading of the manuscript.

REFERENCES

1. Cadet, J., and Vigny, P. (1990) in *Bioorganic Photochemistry* (Morrison, H., Ed.) pp 1−272, John Wiley & Sons, New York.
2. Friedberg, E. C., Walker, G. C., and Siede, W. (1995) *DNA Repair and Mutagenesis*, ASM Press, Washington, DC.
3. Modrich, P., and Lahue, R. (1996) *Annu. Rev. Biochem. 65*, 101−133.
4. Krokan, H. E., Standal, R., and Slupphaug, G. (1997) *Biochem. J. 325*, 1−16.
5. Hoeijmakers, J. H. J. (1993) *Trends Genet. 9*, 173−177.
6. Hoeijmakers, J. H. J. (1993) *Trends Genet. 9*, 211−217.
7. Yasuda, S., and Sekiguchi, M. (1970) *Proc. Natl. Acad. Sci. U.S.A. 67*, 1839−1845.
8. Friedberg, E. C., and King, J. J. (1971) *J. Bacteriol. 106*, 500−507.
9. Radany, E. H., and Friedberg, E. C. (1980) *Nature 286*, 182−185.
10. Nakabeppu, Y., and Sekiguchi, M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2742−2746.
11. Doi, T., Recktenwald, A., Karaki, Y., Kikuchi, M., Morikawa, K., Ikehara, M., Inaoka, T., Hori, N., and Ohtsuka, E. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9420−9424.
12. Hori, N., Doi, T., Karaki, Y., Kikuchi, M., Ikehara, M., and Ohtsuka, E. (1992) *Nucleic Acids Res. 20*, 4761−4764.
13. Manuel, R. C., Latham, K. A., Dodson, M. L., and Lloyd, R. S. (1995) *J. Biol. Chem. 270*, 2652−2661.
14. Morikawa, K., Matsumoto, O., Tsujimoto, M., Katayanagi, K., Ariyoshi, M., Doi, T., Ikehara, M., Inaoka, T., and Ohtsuka, E. (1992) *Science 256*, 523−526.
15. Morikawa, K., Ariyoshi, M., Vassylyev, D. G., Matsumoto, O., Katayanagi, K., and Ohtsuka, E. (1995) *J. Mol. Biol. 249*, 360−375.
16. Vassylyev, D. G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E., and Morikawa, K. (1995) *Cell 83*, 773−782.
17. Lee, B. J., Sakashita, H., Ohkubo, T., Ikehara, M., Doi, T., Morikawa, K., Kyogoku, Y., Osafune, T., Iwai, S., and Ohtsuka, E. (1994) *Biochemistry 33*, 57−64.
18. von Hippel, P. H., and Berg, O. G. (1989) *J. Biol. Chem. 264*, 675−678.
19. Lloyd, R. S., Hanawalt, P. C., and Dodson, M. L. (1980) *Nucleic Acids Res. 8*, 5113−5127.
20. Gruskin, E. A., and Lloyd, R. S. (1986) *J. Biol. Chem. 261*, 9607−9613.
21. Dowd, D. R., and Lloyd, R. S. (1990) *J. Biol. Chem. 265*, 3424−3431.
22. Dowd, D. R., and Lloyd, R. S. (1989) *Biochemistry 28*, 8699−8705.
23. McCullough, A. K., Scharer, O., Verdine, G. L., and Lloyd, R. S. (1996) *J. Biol. Chem. 271*, 32147−32152.
24. Kemmink, J., Boelens, R., Koning, T. M. G., Kaptein, R., van der Marel, G. A., and van Boom, J. H. (1987) *Eur. J. Biochem. 162*, 37−43.
25. Miaskiewicz, K., Miller, J., Cooney, M., and Osman, R. (1996) *J. Am. Chem. Soc. 118*, 9156−9163.
26. Spector, T. I., Cheatham, T. E., III, and Kollman, P. A. (1997) *J. Am. Chem. Soc. 119*, 7095−7104.
27. McAteer, K., Jing, Y., Kao, J., Taylor, J.-S., and Kennedy, M. A. (1998) *J. Mol. Biol. 282*, 1013−1032.
28. Yamaguchi, H., van Aalten, D. M., Pinak, M., Furukawa, A., and Osman, R. (1998) *Nucleic Acids Res. 26*, 1939−1946.
29. Schrock, R. D. d., and Lloyd, R. S. (1991) *J. Biol. Chem. 266*, 17631−17639.
30. Schrock, R. D. d., and Lloyd, R. S. (1993) *J. Biol. Chem. 268*, 880−886.
31. Manoharan, M., Mazumder, A., Ransom, S. C., Gerlt, J. A., and Bolton, P. H. (1988) *J. Am. Chem. Soc. 110*, 2690−2691.
32. Kim, J., and Linn, S. (1988) *Nucleic Acids Res. 16*, 1135−1141.
33. Liuzzi, M., Weinfeld, M., and Paterson, M. C. (1987) *Biochemistry 26*, 3315−3321.
34. Krauss, M., Luo, N., Nirmala, R., and Osman, R. (1999) in *Transition State Modeling for Catalysis* (Truhlar, D. G., and Morokuma, K., Eds.) pp 424−438, American Chemical Society, Washington, DC.
35. Lee, F. S., Chu, Z. T., and Warshel, A. (1993) *J. Comput. Chem. 14*, 161−185.
36. Warshel, A. (1981) *Biochemistry 20*, 3167−3177.
37. Russell, S. T., and Warshel, A. (1985) *J. Mol. Biol. 185*, 389−404.
38. Warshel, A., and Levitt, M. (1976) *J. Mol. Biol. 103*, 227−249.
39. Tanford, C., and Roxby, R. (1972) *Biochemistry 11*, 2192−2198.
40. Yang, A. S., Gunner, M. R., Sampogna, R., Sharp, K., and Honig, B. (1993) *Proteins: Struct., Funct., Genet. 15*, 252−265.
41. Bashford, D., and Karplus, M. (1990) *Biochemistry 29*, 10219−10225.
42. King, G., and Warshel, A. (1989) *J. Chem. Phys. 91*, 3647−3661.
43. Lee, F. S., and Warshel, A. (1992) *J. Chem. Phys. 97*, 3100−3107.
44. Langen, R., Brayer, G. D., Berghuis, A. M., McLendon, G., Sherman, F., and Warshel, A. (1992) *J. Mol. Biol. 224*, 589−600.
45. Sham, Y. Y., Chu, Z. T., and Warshel, A. (1997) *J. Phys. Chem. B 101*, 4458−4472.
46. Warshel, A., and Russel, S. T. (1984) *Q. Rev. Biol. 17*, 283−422.
47. Warshel, A., Russell, S. T., and Churg, A. K. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4785−4789.
48. Cutler, R. L., Davies, A. M., Creighton, S., Warshel, A., Moore, G. R., Smith, M., and Mauk, A. G. (1989) *Biochemistry 28*, 3188−3197.
49. Hwang, J. K., and Warshel, A. (1988) *Nature 334*, 270−272.
50. Rees, D. C. (1980) *J. Mol. Biol. 141*, 323−326.
51. Alden, R. G., Parson, W. W., Chu, Z. T., and Warshel, A. (1995) *J. Am. Chem. Soc. 117*, 12284−12298.
52. Mehler, E. L. (1996) *J. Phys. Chem. 100*, 16006−16018.
53. Svensson, B., and Jönsson, B. (1995) *J. Comput. Chem. 16*, 370−377.
54. Sham, Y. Y., Muegge, I., and Warshel, A. (1998) *Biophys. J. 74*, 1744−1753.
55. Muegge, I., Schweins, T., Langen, R., and Warshel, A. (1996) *Structure 4*, 475−489.
56. Muegge, I., Tao, H., and Warshel, A. (1997) *Protein Eng. 10*, 1363−1372.
57. Besler, S. H., Merz, K. M. J., and Kollman, P. A. (1990) *J. Comput. Chem. 11*, 431.
58. Singh, U. C., and Kollman, P. A. (1984) *J. Comput. Chem. 5*, 129.
59. Shapiro, R., and Kang, S. (1969) *Biochemistry 8*, 1806−1810.

BI9901937