Electrostatic interactions play an essential role in DNA repair and cold-adaptation of Uracil DNA Glycosylase.

Olufsen, M., Smalås, A. O. & Brandsdal B. O.*

The Norwegian Structural Biology Centre, Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway.

*Corresponding author:
Bjørn Olav Brandsdal
Email: Bjorn-Olav.Brandsdal@chem.uit.no
Phone: +4777644057
Fax: +4777644765

Running title: Electrostatic interactions in UDG-DNA
Keywords: Uracil DNA glycosylase, protein-DNA binding, continuum electrostatics, MM-PBSA, molecular dynamics

Summary

Uracil DNA glycosylase (UDG) is an important DNA repair enzyme and initiates the base excision repair pathway removing uracil from damaged DNA. Cod UDG (cUDG) and human (hUDG) is used as a model system in the investigation of specific issues related to DNA repair and putative differences in the temperature adapted behavior of these two enzymes. cUDG has previously been shown to display the typical cold-adapted feature, that is, increased catalytic efficiency but at the cost of reduced thermal stability. The increased catalytic efficiency is attributed both to an increased k_{cat} and a reduced K_m, and the source of the different behavior of cUDG and hUDG is explored at the atomic level using a range of computational approaches. Continuum electrostatic calculations reveal that cUDG possesses surface potentials that are more complementary to the DNA potential at and around the catalytic site when compared to hUDG. These calculations also add further support to the view that differences in electrostatics is important for substrate recognition as well as substrate binding. Comparative molecular dynamics simulations combined with free energy calculations using the Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) method were carried out in an attempt to estimate the energetics involved in substrate binding. The absolute binding free energies obtained reveal that the Michaelis-Menten complex is more stable for cUDG when compared to its warm-active counterpart. Decomposition of the binding free energy into contributions from individual residues shows that this is largely a result from improved interactions between the DNA recognition loop and DNA in cUDG. However, the present MM-PBSA calculations also demonstrate the limitations of the methodology. While the binding free energies are stable, they are spurred with large standard deviations, pointing to limitations of free energy calculations where large individual contributions to the free energy occur.

Introduction

Organisms capable of survival in low temperature niches have been known for a long time, and are collectively referred to as psychrophilic (cold-loving). Most of our planet is covered with permanently cold regions, and life has effectively colonized most ecological niches. Survival in extreme environments requires that the organisms adapt their metabolisms to low temperature, including their enzymes. Mechanisms of enzymatic adaptation to cold environments are not presently well understood. Hochachka and Somero [1] suggested that organisms adapted to cold environments need to compensate the reduced temperature by expressing enzymes with increased flexibility to maintain a high catalytic efficiency. More recent investigations suggest that the increased catalytic efficiency of cold-adapted enzymes is not necessarily attributed to an overall increase in structural flexibility but rather to the key components directly involved in the catalytic cycle [2,3]. Crystallographic analysis of uracil DNA glycosylase from cod (cUDG) and human (hUDG) did not provide indications of any differences in molecular flexibility [4]. Subsequent MD simulations do, however, indicate increased flexibility for the DNA binding loop in cUDG as compared to hUDG [5]. Structural analysis of psychrophilic and mesophilic trypsin did not either reveal significant differences in overall flexibility [6], which was also supported by investigations using computer simulations [7]. However, both the crystallographic and the computational study of cold- and warm-active trypsin point towards different dynamic behavior in localized regions as a possible mean for enzymatic adaptation to cold environments. Increased molecular flexibility is not necessarily the only strategy for adaptation to low temperature.

Alteration of the electrostatic potential of key residues has been proposed to play a central role in adaptation of citrate synthase [8]. Psychrophilic citrate synthase has significantly different electrostatic potentials at and around the active site in comparison to its thermophilic counterpart, and focused electrostatic attraction of substrates has been proposed to be a possible source for the enhanced catalytic activity of the cold-active citrate synthase [8]. Kumar and Nussinov [9] also found that electrostatics play different roles in psychrophilic and thermophilic citrate synthase. Qualitative investigations of the electrostatic surface potentials in seven trypsin isoenzymes using continuum electrostatic

calculations showed a more negatively charged substrate binding site in the cold-adapted trypsin when compared to warm-active homologues [10]. Accommodation of small synthetic inhibitors and cognate amino acid side-chains to the specificity pocket of trypsin is electrostatically more favorable in the cold-adapted enzyme [11], suggesting that electrostatics is important in temperature adaptation. Optimization of electrostatics has also been suggested to be an adaptational strategy followed by cod UDG [4,12].

Uracil DNA glycosylase is a DNA repair enzyme and is the first enzyme in the base excision repair pathway [13]. The enzyme catalyzes removal of uracil from single- and double-stranded DNA by cleaving the N-glycosylic bond between the target base and deoxyribose [14]. The crystal structure of the catalytic domain of the family 1 UDG from several species are known: human [15], cod [4], herpes simplex virus type-1 [16], Escherichia coli [17] and Epstein-barr virus [18]. Several crystal structures of hUDG and herpes simplex virus type-1 UDG in complex with DNA have also been determined [16,19-22]. The catalytic domain of cUDG and hUDG consists of 223 residues, and the sequence identity between them is 75%. The overall topology is a typical α/β protein [15]. The four important loops for detection and catalysis are: the 4-Pro loop (^{165}PPPPS^{169}), the Gly-Ser loop (^{246}GS^{247}), the Leu272 loop (^{268}HPSPLSVYR^{276}) and the water-activating loop (^{145}DPYH^{148}) [20]. These loops are conserved among cUDG and hUDG. The amino acids mentioned above are from hUDG, and there are two substitutions in the Leu272 loop in the cUDG sequence: V274A and Y275H. The Leu272 loop is particularly important as it moves into the minor groove of the double-stranded DNA and is involved in the flipping of the uracil base. This movement is essential for bringing the catalytic important residue His268 within hydrogen bonding distance of the oxygen atom (O2) of uracil [20]. Even if the cold-adapted cUDG and the warm-adapted hUDG enzymes have very similar 3D structure, the cUDG enzyme is up to 10 times more catalytic efficient (k_{cat}/K_m) in the temperature range from 15-37°C compared to the human homologue [23]. This is achieved through optimization of both kinetic parameters as k_{cat} is increased and K_m is reduced for cUDG. K_m is most affected and possibly reflects increased substrate interactions in the reaction catalyzed by the cold-adapted enzyme.

Enzyme-substrate interactions and the apparently improved substrate accommodations for cUDG are further explored using a range of computational

techniques, including continuum electrostatics calculations, molecular dynamics simulations and free energy calculations. We find that formation of the Michaelis-Menten complexes is highly favorable for both enzymes, but the cUDG-DNA complex is energetically more stable when compared to hUDG-DNA. Overall, this is attributed to improved electrostatic properties of cUDG, but also superior interactions between key structural areas and the substrate in the cold-adapted enzyme. The present investigations thus point to improved electrostatics as a possible route for cold-adaptation and enhanced catalytic efficiency.

Method

Structural Models

Crystal structures were available for cUDG [4], hUDG [15] and hUDG-DNA [21]. The cUDG-DNA complex was modeled with the hUDG-DNA structure as template. This DNA had originally a 2'-deoxy-pseudouridine-5'monophosphate, but this base was modeled into a 2'-deoxy-uracil-5'monophosphate by switching place for the atoms: C2 ↔ C4, O2 ↔ O4 and N1 ↔ C5, (Fig. 1). The latter is the uracil base, recognized and removed from the DNA by UDG. The double-stranded DNA from the crystal structure of hUDG-DNA consists of 19 bases. These structures were used as starting structures in the MD simulations. All the crystal structures were of the recombinant enzymes with three mutations in the N-terminal end: P82M, V83E and G84F. UDG contains several histidines, all except His148 were considered as neutral, and protonated at the Nε2 atom, in the simulations. These choices were based on data from NMR and continuum electrostatics calculations [24].

Molecular dynamics

The AMBER9 program package [25] with the parm99 force field [26] was used to run and analyze the MD simulations. Water molecules were added to the protein with a 15 Å buffer from the edge of the box and described according to the TIP3P model [27]. Prior to the MD simulations, the molecular systems were subjected to 200 cycles of energy minimization of the water with the protein fixed and then 200 cycles of minimization of the whole system. In the initial phase, the temperature of the system was slowly raised in steps to the final temperature of 300 K, followed by an equilibration period of 110 ps. The production phase of the simulations was carried out in the isothermal-isobaric ensemble (300 K and one atmosphere pressure). Pressure and temperature were maintained by the Berendsen coupling algorithm [28]. A 8 Å cutoff was used for non-bonded interactions and the Particle-Mesh-Ewald (PME) method [29] was used to handle long-range interactions beyond the cutoff. SHAKE [30] was applied to constrain covalent bonds involving all hydrogen atoms. A time step of 2 fs was employed. Coordinates were

written to file every 20 ps during the production phase, and the simulations were carried out for 10 ns for both cUDG and hUDG. The density, total energy, temperature and root-mean-square deviation plotted *vs* time were used to investigate the stability of the simulations, and all four properties are stable throughout the simulations (results not shown).

MM-(GB)PBSA

The MM-PBSA method [31-33] was applied to calculate the binding energy of the protein-DNA complex. The structural ensembles consisting of 500 conformations collected every 20 ps in the MD simulations were post-processed using the MM-PBSA method to estimate the free energy of binding. This method estimates the free energy of each conformation according to:

(1) $$G = H_{MM} + G_{sol} - TS_{Solute}$$

E_{MM} is the total molecular mechanical energy in the gas phase and can be divided into several energy terms:

(2) $$H_{MM} = H_{bond} + H_{angle} + H_{torsion} + H_{elec} + H_{vdW}$$

(3) $$G_{sol} = G_{Pol} + G_{np}$$

(4) $$G_{np} = \gamma \cdot SASA + b$$

where H_{bond}, H_{angel}, $H_{torsion}$, H_{elec} and H_{vdW} are the bond, angle, torsion, electrostatic and van deer Waals energies, respectively. G_{sol} is the solvation free energy and can be divided into two terms according equation 3. G_{pol} is the electrostatic solvation free energy, and is normally calculated with the Poisson-Boltzmann method (PB) [34] or with the GB method [35,36]. G_{np} is the nonpolar solvation free energy and is calculated with equation 4. In equation 1, the T is the temperature and S is the solute entropy. There are different ways to calculate the entropy [37], and the solute entropy is estimated using normal mode analysis [38] as implemented in the AMBER program package. The solute entropy of each snapshot is calculated from the structure minimized in vacuum with a distance-dependent dielectric constant of 4r and the convergence criterion for the energy gradient was set to 0.1 kcal mol^{-1} $Å^{-1}$.

The electrostatic contribution to the solvation free energy was calculated with the PBSA program in AMBER [34] for the binding free energies and with the generalized Born (GB) method [35,36] for the decomposition of binding free energies. We have also calculated the solvation free energies contributions to the binding energy with the GB method, and since the GB and PB gave similar results only the results from the PB calculations are shown. The solute and solvent dielectric constants were set to 1 and 80 in all PB and GB calculations, and the ionic strength was set to zero. In both methods, the parameters used to calculate the non polar contribution to the solvation energy (γ and b) was set to 0.0072 kcal/Å^2 and 0.0 kcal/mol, respectively. These parameters have been developed to be used with the AMBER force field [39]. The lattice spacing was set to 2 grids/Å and a maximum of 1000 iterations were used for the PBSA calculations. The solvent-accessible-surface-area (SASA) was calculated with a probe radius of 1.4 Å both in the PB and GB methods. The molsurf program [40] and the LCPO method [41] was applied to calculate the SASA in the PBSA and GB methods, respectively. When computing the contribution from individual residues to the free energy of binding, the surface area was computed by recursively approximating a sphere around an atom starting from an icosahedra .

The binding free energy was calculated according to the following equation:

(5) $$\Delta G_{bind} = \left\langle \Delta G_{complex} \right\rangle - \left\langle \Delta G_{receptor} \right\rangle - \left\langle \Delta G_{ligand} \right\rangle$$

where $\left\langle \Delta G_{complex} \right\rangle$, $\left\langle \Delta G_{receptor} \right\rangle$ and $\left\langle \Delta G_{ligand} \right\rangle$ are the average free energies of the protein-DNA complex, protein, and the DNA, respectively, averaged over 500 snapshots from the MD simulations.

Continuum electrostatics calculations and surface potentials

The DelPhi program [42,43] was used to calculate the electrostatic potential of cUDG and hUDG. His148 was charged in the continuum electrostatics calculations, in addition to all Lys, Arg, Glu and Asp residues. The calculations were performed using the partial charges and atomic radii of the AMBER force field (parm99) [26]. The electrostatics was calculated using the linear Poisson-Boltzmann equation and a grid size of 165x165x165

points in a 3-dimensional grid. Stepwise focusing was used to increase the accuracy [44]. Initially a rough grid was calculated with the Coulombic boundary conditions. The resulting grid of this calculation was adopted as the boundary condition for two further focused calculations, and in the last calculation the molecule occupied ~85% of the box. A solvent probe of 1.4 Å was used to calculate the molecular surface. These calculations were run with zero ionic strength and the dielectric constants of the protein and the water were set to 20 and 80, respectively.

Results and discussion

Qualitative investigations – continuum electrostatic calculations.

Enzymes need to form a complex with their substrates before they can exert their mode of action. Electrostatic interactions play a key role in virtually all biological systems, and are expected to be of particular importance for enzymes with DNA as substrate. DNA is a highly charged macromolecule and enzymes that bind to DNA often have complementary charged surfaces, yielding a tight enzyme-substrate complex. Uracil DNA glycosylase is an important enzyme involved in DNA repair removing misincorporated uracil from the DNA strand. It has a positively charged surface at and around the active site where the enzymatic action takes place. The electrostatic potential is not only important for catalysis but also when recognizing DNA, as illustrated in Fig. 2. The structural model used to generate the electrostatic isocontours in Fig. 2 was constructed by placing the DNA fragment observed in the crystal structure (1EMH [21]) 15 Å from the crystallographic position. Figure 2 also shows that the positive electrostatic isocontour extends out of the binding site, and interacts favorably with the negative isocontour from the DNA strand. It has been proposed that the increased substrate affinity observed for cUDG when compared to hUDG is due to enhanced positive electrostatic potential at surface areas central to formation of the enzyme-substrate complex [4,12]. The charge of the phosphodiesters on the target nucleotide and the two connected nucleotides has been shown to have large effect on the ground state (K_m effects) value and even greater effect on the ionic transition state (k_{cat}/K_m effects) [45]. It is thus reasonable to expect that altering the charges in the active site region will also affect binding of the DNA.

Continuum electrostatic calculations have over the past years been used extensively to investigate properties of electrostatic origin in macromolecules [46-48], and are now an important computational tool for the exploration of electrostatics of such molecules. The computed electrostatic surface potentials and isocontours for cUDG and hUDG are presented in Fig. 3. Both cUDG and hUDG have, as expected, highly positive electrostatic potentials in the specificity pocket and in nearby areas that are know to interact directly with DNA. As indicated by Moe *et al* [12], differences are observed in the electrostatic surface potentials between cUDG and hUDG, as the psychrophilic

enzyme has a more positive electrostatic potential near the active site. Figure 3A and B show that the psychrophilic enzyme has more positive electrostatic potentials close to both ends of this short DNA strand. Residue 171, which is Glu and Val in hUDG and cUDG respectively, is particularly interesting when it comes to possible differences in substrate binding. This substitution generates a more positive electrostatic potential in this area of the psychrophilic enzyme (Fig. 3A and B). hUDG and cUDG have K_m values of 2.4 μM and 0.8μM at 295 K [12], respectively, corresponding to three times higher K_m for hUDG compared to cUDG. Mutations of residue 171 have large effects on the K_m value, and the K_m values are 0.7 and 1.7 μM for the hUDG-E171V and the cUDG-V171E [12], respectively. Mutation of this residue thus yields a hUDG enzyme with similar K_m as the cUDG, but also visa versa. Since K_m is roughly an inversely measure of the binding strength between the enzyme and its substrate [49], lower K_m values indicate stronger association between the enzyme and substrate. It therefore seems likely that the 171 residue is important for binding DNA to the UDG molecule. Figure 3A and B show that there are larger areas with positive potential at the flanking sides of the catalytic site, suggesting stronger non-specific interactions between the cold-adapted enzyme and DNA. Ultimately, this can lead to increased stability of the UDG-DNA complex. Furthermore, the more positive electrostatic potential where the DNA strand interacts with the enzyme, will accommodate for a more efficient recognition of DNA and orient it in the correct position for catalytic cleavage.

It is also interesting to examine the potentials in other areas of the structures, and particularly at the opposite side of the DNA binding site. Both enzymes have predominantly negative electrostatic surface potentials here, but cUDG possesses a larger area with negative electrostatic potential when compared to its warm-active homologue (Fig. 3C and D). This is expected to be of importance when it comes to recognition of damaged DNA, binding and subsequently catalysis, completing the removal of uracil. Assuming that UDG moves freely in solution searching for DNA, the negative potential on the back of the enzyme will lead to repulsive interactions with the negatively charged DNA. As a response, the enzyme will shift orientation and reorient in such a way that the DNA matches the positively charged specificity pocket. More negative charges at the back of the enzyme will make it more energetic favorable for the enzyme to reorient, and

cUDG can be expected to recognized DNA and find the optimal binding orientation to DNA more easily than hUDG.

Thermodynamic analysis of UDG-DNA interactions.

Qualitative examination of the electrostatic properties of cUDG and hUDG (Fig. 3) encouraged us to initiate more accurate investigations of the enzyme-substrate complexes using free energy calculations. Several methods are currently available to compute the strength of binding between proteins and their binding partners, ranging from computationally expensive methods like free energy perturbation (FEP) and thermodynamic integration (TI) (see [50] for a review of the methods) to various empirical/knowledge-based scoring approaches [51-53]. None of these are however suitable for studying protein-DNA interactions. That is, FEP and variants thereof are currently restricted to study relative binding free energies of similar compounds, while simple rapid scoring functions are too approximate to yield quantitative results of protein-DNA complexes. The linear interaction energy (LIE) method [54] and the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method [31,33] are other popular methods used to calculate the association energy of macromolecular complexes. Both approaches are based on analysis of molecular dynamics trajectories and the generated structural ensembles. Attempts were made to estimate the absolute binding free energy between UDG and DNA using the LIE method, but obtaining converged energies was not possible within reasonable simulation time (results not shown). The MM-PBSA method was initially used to study the stability of various DNA and RNA fragments [33]. In later years, however, the method has also been applied to calculate binding free energies of proteins and small ligand [55,56], protein-protein [57-59], protein-RNA [60] and protein-DNA [61]. Estimation of the absolute free energy of binding was therefore carried out using the MM-PBSA approach. It is challenging to calculate the absolute binding energy for association of highly charged large macromolecules, but the MM-PBSA method has proven to be able to qualitatively reproduce the absolute binding energies for such systems [60].

The binding free energy can be calculated in two ways with the MM-PBSA method, either using a single MD simulation of the complex or using individual simulations of complex, protein and ligand. The former is referred to as the single trajectory MM-PBSA method whereas the latter is referred to as the multiple trajectories MM-PBSA. The single trajectory approach assumes that there are no conformational changes in the protein or in the ligand from the unbound to the bound state, which may in some cases, be a rather drastic assumption. The advantage of using only one simulation of the complex is that convergence in the free energies can in principle be achieved more easily as compared to separate simulations. Estimation of the free energy using individual simulations of protein, ligand and complex is difficult due to convergence problems associated with particularly the intramolecular energy terms which cancel out when only one trajectory is used. Nonetheless, successful application of both single and multi trajectory calculations have been reported [60,62].

Free energies of binding

The individual contributions and the resulting free energies for formation of the Michaels-Menten complex for cUDG and hUDG are presented in Table 1 and 2, respectively. The results show that the psychrophilic enzyme interacts more favorably with DNA irrespective of whether the single or multiple simulations method is used. The total binding free energy to DNA is –39.8 kcal/mol and –28.7 kcal/mol for cUDG using single and multiple simulations, respectively, while the corresponding values for hUDG are –34.2 and –25.3 kcal/mol. The relative binding free energy is thus –5.7 kcal/mol and –3.4 kcal/mol in favor of cUDG with the single and the multiple trajectory methods, respectively. There are presently no experimental association constants available for binding of the DNA fragment studied here to cUDG or hUDG. Experimental studies have, however, shown that hUDG binds to dsDNA containing uracil homologues with a binding strength of ~ –9.0 kcal/mol [63]. Other experimental bindings studies of *Escherichia coli* UDG bound to different DNA fragments show that the binding free energy varies from –8.0 kcal/mol to –15.0 kcal/mol [64-66]. The free energies appear to be somewhat overestimated with both procedures, but the relative difference between cUDG and hUDG is less sensitive to the choice of single *vs* multiple trajectories.

Whether the free energies of binding can be computed from the simulation of only the complex, depend upon the structural changes which the protein and DNA undergo during complex formation. Crystal structures are available of both cUDG and hUDG without DNA present, but structure of the complex is only available for hUDG. Nonetheless, comparison of hUDG with and without DNA bound reveal that there are only minor conformational differences between bound and unbound enzyme. The overall backbone root-mean-squared deviation (r.m.s.d.) between the two structures is 1.43 Å. No experimental structure is, however, available for examination of possible changes in the DNA strand during the binding process. DNA is not a static structure but undergoes rapid unpairing of individual base pairs and slow large cooperative unfolding events [67,68]. There is an ongoing debate on how DNA repair enzymes, such as DNA glycosylases, recognize rare damaged bases in a large background of normal DNA bases. Two views for localization of damaged sites have emerged: the *base sampling model* and the *inherent extrahelicity model*. The *base sampling model* suggests that UDG localizes uracil by breaking base pairs and flip them out to test them against the interactions offered in the specificity pocket [69]. Another view is the *inherent extrahelicity model*, where the base pairs involving uracil is inherently weak and that the uracil will spontaneous flip-out to an extrahelical conformation, complementary to the binding interactions offered by UDG [67,69]. NMR imino proton experiments have also shown that the U·A base pairs rapidly open at room temperature and the opening rates are greater or equal to the rate constants for the kinetic steps of base flipping of UDG [65,67]. It has been shown that UDG does not alter the opening rate of the base but instead slows the closing rate of the A·U base pair [67]. Irrespective of how the enzyme actually localizes the damaged base, the Michealis-Menton complex will be the same for the two proposed mechanisms, but free energies contributions may be left out in our MM-PBSA calculations. The structure of DNA when bound to hUDG indicates distortions from ideal geometry as the DNA strand is bent. As has been pointed out by others [65], the energetic effect of DNA bending is highly unfavorable and constitutes a significant contribution to enzymatic base flipping. The energetic effect of DNA bending is very challenging to capture, and is not fully accounted for in our free energy calculations, particularly when only one simulation is used. If the *base sampling model* is

correct, an additional contribution is missing, corresponding to the free energy required to break the base pair and flip the base into the active site of the enzyme. Thus, neglect of the contribution from DNA bending and possibly flipping of the damaged base will lead to an overestimation of the stability of the Michaelis-Menten complexes. The initial model for the simulations of unbound DNA was constructed from the DNA observed in the crystal structure of hUDG-DNA (1EMH [21]). Both the uracil and the pairing partner adenine are flipped out in an extrahelical conformation in the starting structure. In the MD simulation of unbound DNA, the adenine base which pairs to uracil flips back into the DNA helix, but the uracil base does not. However, the lifetime of an extrahelical base is between 100-800 ns [70] and the present simulations times are 10 ns, thus probably too short time to observe spontaneous base flipping. The DNA structure bound to hUDG is bent to the enzyme surface causing the flanking phosphate bases to be compressed, as judged by the distance between the phosphorus atoms at the nucleotides connected to the uracil nucleotide is compressed from ~12 to 7.7 Å [20,21]. The distance between the same phosphorus atoms is 11.9 Å in the final structure of the simulation of free DNA, showing that the DNA bends back into its favorable relaxed orientation.

Table 1 and 2 show that the enthalpic contribution to the free energy of binding is very favorable in all four ΔG_{bind} values, whereas the entropic contribution opposes binding. The contribution from the solute entropy to the binding free energy varies in the two methods, but is of similar magnitude between all. Estimation of entropy is perhaps the most challenging part of calculating binding energies, due to changes in the degree of freedom of the solutes [62]. Quasiharmonic analysis and normal mode analysis can be used to calculate solute entropies from simulations [38,71]. One limitation with quasiharmonic analysis is to obtain converged energies for the conformational entropy. Conformational entropic studies of the β-heptapeptide did not even show convergence after 150 ns of simulation [72,73]. The normal mode approach requires energy-minimization of the conformations prior to the entropy calculations, and artefactual conformational changes may be introduced during the energy-minimization process [74]. The change in solvent entropy upon binding is not explicitly included in the MM-PBSA method, but included implicitly in the change in SASA associated with binding [60].

The rotational and translational contribution to the entropy is identical for both the single and separate trajectory method and also identical for the two enzymes. Thus, the difference observed in ΔS between the two enzymes is caused by the vibrational part of the entropy. Cold-adapted enzymes are thought to have increased molecular flexibility, and psychrophilic UDG has been shown to possess a more flexible DNA recognition loop compared to mesophilic hUDG [4,5,12]. One should then expect that the cold-adapted cUDG would show the largest difference in entropy upon binding, and this is actually the case in the separate trajectory method (Table 1 and 2). In the single trajectory method, on the other hand, the opposite is observed, as the mesophilic enzyme show the largest loss in entropy upon binding. MD simulations have shown that the DNA recognition loop looses most of its flexibility in the UDG-DNA complex [5]. The conformations for the unbound state in the single trajectory method are extracted from the simulation of the complex, which may explain why the cold-adapted enzyme does not possess a larger loss in entropy with the single trajectory method.

Kinetic experiments have shown that cUDG associates more favorably to DNA compared to its warm-active homologue hUDG (reduced K_m), as discussed previously. When examining the different contributions to the binding free energy, it is interesting to note that the electrostatic contribution to the free energy of binding is much more negative (favorable) for the psychrophilic UDG as compared to its warm active homologue (Table 1 and 2). The continuum electrostatics calculations showed that cUDG has a more positive electrostatic surface potential near the active site (Fig. 3A), thus, it then seems reasonable that this enzyme will have a more favorable electrostatic interactions with the negatively charged DNA.

Statistical considerations

Due to the large energies involved, the protein-DNA interactions are hard to calculate. For example, the free energy of the UDG-DNA complex is ~ –11000 kcal/mol, and even if the individual contribution to the free energy of binding is very high, the binding free energy is usually only a few kcal/mol. Compared to the large numbers, the standard deviation of each individual contribution is rather low. However it is difficult to obtain good statistics for the binding free energies, and as seen for the present calculations of the

standard deviations are rather high. The single trajectory method gave lower standard deviations than the separate trajectory method. Others have also observed large standard deviations for the protein-DNA calculation using the MM-PBSA methodology [61]. When we plot the sum of the gas-phase energies and the solvation free energy for cUDG-DNA complex in the separate simulations, the plot indicates rather stable energies even though large fluctuations in the energy are observed (Fig. 4). To test the internal consistency of the computed binding free energy, the free energies were calculated for the first and the second half of the separate trajectory MM-PBSA method (Table 3). The binding energies from the first and second half of the simulations are quite similar, but with high standard deviations. The long MD simulation and the high number of snapshots used to compute the average free energy should also secure that reliable binding free energy estimate are obtained.

Determinants of binding – decomposition of the binding free energy.

The contribution from individual residues to the interaction free energy has been calculated and analyzed, and are presented in Fig. 5. Decomposition of the binding free energy into contributions from the amino acids is very helpful in determining the residues important to binding and can aid design of enzymes with novel biophysical properties. The binding free energy per residue varies from –9.0 kcal/mol to +2.6 kcal/mol (Fig. 5). It is clear that the four loop regions which have been suggested to be important for detection and catalysis, emerge as the areas with the most favorable contribution to the binding free energy, as indicated in Fig. 5. For hUDG the 4-Pro loop (^{165}PPPPS^{169}), the Gly-Ser loop (^{246}GS^{247}), the Leu272 loop (^{268}HPSPLSVYR^{276}) and the water-activating loop (^{145}DPYH^{148}) have a interaction energy of –11.2 kcal/mol, –6.0 kcal/mol, –25.5 kcal/mol and –6.8 kcal/mol, while the corresponding energies for cUDG are –10.5 kcal/mol, –4.9 kcal/mol, –32.6 kcal/mol and –3.9 kcal/mol, respectively. The four loops contribute with –74.7 kcal/mol and –73.7 for hUDG-DNA and cUDG-DNA, and are thus responsibly for 66.2 % and 70.4 % of the enthalpic contribution to the binding free energy in hUDG and cUDG, respectively.

The catalytic important residue Asp145, located in the water-activating loop, is believed to form unfavorable interactions with the 3`-phosphodiester group of the

deoxyuridine residue of the substrate [75], which is supported by our calculations. This residue actually has the most positive interaction free energy of all residues, +2.3 kcal/mol and +2.6 kcal/mol in cUDG and hUDG, respectively (Fig. 5).

The Leu272 loop or the DNA recognition loop plays an important role in uracil recognition and penetrates into the minor grove of the dsDNA in the complex [76,77]. The Leu272 loop is also believed to play a role in flipping of the uracil base, either to flip out the uracil base from the DNA helix or to work as a "doorstop" to prevent the already flipped-out uracil to flip back into the dsDNA helix [76,77]. The Leu272 loop interacts strongly with DNA in both enzymes and is responsibly for 34.1 % and 44.2 % of the enthalpic contribution to the binding energy in the warm-active and cold-adapted UDG, respectively. Table 4 and 5 show the contributions to the binding free energy from each residue in the Leu272 loop for the two enzyme homologues. Three residues interact very favorably with DNA: His 268, Leu272 and Arg276. The largest differences in the interaction free energy per residue between the cold- and the warm-active enzyme are observed in this loop. Arg276 contribute with –4.7 kcal/mol in hUDG and –9.0 kcal/mol in cUDG to their respective free energies of binding. The side chain of Arg276 is closer to the DNA fragment in cUDG than in hUDG, and is within hydrogen bonding distance (3.40 Å) of two different DNA bases in cUDG (Fig. 6). Arg276 is in contrast not within hydrogen bonding distance of DNA in hUDG. This difference in hydrogen bond distance between the Arg276 in cUDG and hUDG can probably be explained by residue 275. This residue, which is Tyr in hUDG and His in cUDG, is another important residue in the Leu272 loop. While the hydrophobic Tyr side chain is pointing away from the DNA, the polar His275 side chain points towards the DNA and forms a hydrogen bond to O5' atom on the adenine base opposite of the uracil base. Thus, this residue contributes more favorably to binding free energy in cUDG compared to hUDG. As a result, the DNA strand is pulled closer to the enzyme in this area of the structure. The remaining residues in the Leu272 loop of cUDG form similar hydrogen bonds as described for the hUDG-DNA complex [22].

The Leu272 residue which penetrates into the minor grove of the dsDNA has a strong contribution to the binding free energy, which is dominated by hydrophobic or non-polar interactions (van der Waals term in Table 4 and 5). Mutation of Leu272 to Ala

has large effect on the catalytic efficiency when single stranded DNA is used as substrate [77]. For the single stranded DNA the Leu272 loop does not need to flip out the uracil or work as a "doorstop" as in dsDNA. This indicates that Leu272 may also have another task in the catalytic mechanisms. The highly favorable binding energy of the Leu272 residue is important for stabilization the enzyme-DNA complex, and might be important to orient the DNA in the right position for catalytic cleavage. Our calculations also add further support to this. Mutational studies have shown that His268, Ser270, Leu272 and Arg276 are all critical for hUDG activity [63], which may according to the decomposition be a result of their favorable interactions with DNA. Thus, it seems likely that a strong binding between UDG and DNA is important in order to achieve a high catalytic activity. The His275 has also much stronger binding in cUDG compared to the Tyr275 in hUDG, but the cUDG-H272Y mutant show no significance difference in the K_m value compared to cUDG, but the mutant has a reduction in catalytic efficiency caused by a reduction in the k_{cat} [12]. One explanation for this could be that the His275 residue binds stronger to the transition state than to the ground state, affecting k_{cat} instead of K_m. His275 is also thought to be the main contributor to the increased flexibility of the DNA recognition loop in the cold-adapted cUDG [5].

In position 171, which is Glu in hUDG and Val in cUDG, another interesting difference in the binding free energy per residue is observed between the two enzymes. Val171 has a small favorable contribution to the binding energy (–0.3 kcal/mol), while Glu171 has an unfavorable contribution of +2.0 kcal/mol. The positive contribution from the Glu171 residue in the warm-active enzyme is primarily caused by electrostatic repulsion between the DNA and the negatively charged side chain. Figure 3B also shows that hUDG has an area of negative electrostatic potential around Glu171. The distance between the Glu171 Oε1 atom and the closest oxygen at the DNA terminal base is 8.43 Å in the crystal structure (1EMH [21]). The DNA fragment in this crystal structure consists of only 19 nucleotides, but if a longer DNA strand is used, this distance would decrease significantly, leading to more repulsive forces. Experimental studies have shown that residue 171 is very important for catalytic activity in both cUDG and hUDG [12]. hUDG has 3 times higher K_m values than cUDG, but when the Glu171 is mutated to a Val the

hUDG-E171V mutant achieve similar K_m as cUDG. The opposite is observed when the Val171 in cUDG is mutated to a Glu as the K_m values doubles [12]. Since K_m is related to the binding energy, it seems likely that the unfavorably binding energy for Glu171 could explain the observed difference in K_m value between cold- and warm active UDG. Hence, this residue could possibly be a key residue in the explanation of cold-adaptation in UDG.

Concluding remarks.

Enzymes from organisms living at extreme temperatures need to maintain sufficient structural integrity to allow for catalytic efficiency, while at the same time avoid hot and cold denaturation. Uracil DNA glycosylase is a very good model system not only to study environmental adaptation of enzymes, but also to investigate the DNA repair process itself. Comparative investigations using different levels of theory have been applied to explain the increased catalytic efficiency of UDG from cod and human, and to gain a deeper insight into the DNA repair process. The results show that the stability of the Michealis-Menten complex is higher for cUDG when compared to hUDG, and is attributed to improved electrostatic properties on an overall level. Differences in key structural regions, vital to detection of damaged bases and the subsequent catalytic removal of uracil, between cold- and warm-active UDG were identified through decomposition of the free energy of binding into a residual level.

The increased catalytic efficiency observed for cUDG when compared to hUDG is achieved through a combined effect resulting from increased k_{cat} and decreased K_m. While only the ground state of the chemical reaction catalyzed by UDG has been studied here, it would certainly be interesting to investigate the source of the increased k_{cat}. This requires, however, application of even more sophisticated computational approaches, such as hybrid quantum mechanics/molecular mechanical methods, and is left for future studies.

Estimation of the absolute binding free energy is in many cases a difficult task and is especially tricky when large interaction energies are involved. While the literature contains many examples of successful predictions of ΔG_{bind} for protein-small ligand

complexes, few successful studies have been reported that estimate the stability of protein-DNA complexes. The results presented here with the MM-PBSA approach illustrate the need for efforts aiming at improvement of existing methods and development of new methodologies that accurately describe the energetics of macromolecular complexes.

Acknowledgements.

Finacial support from the Research Council of Norway is gratefully acknowledged. The Norwegian Structural Biology Centre is supported by the Functional Genomics Program (FUGE) of the Research Council of Norway.

References

[1] Hochachka, P. W., Somero, G. N., 1984. Temperature adaptation. In. Biochemical adaptations, 355-449

[2] Fields, P. A., Somero, G. N. Hot spots in cold adaptation: Localized increases in conformational flexibility in lactate dehydrogenase a(4) orthologs of antarctic notothenioid fishes. Proc. Natl. Acad. Sci. U. S. A. 1998, 95, 11476-11481.

[3] Georlette, D., Damien, B., Blaise, V., Depiereux, E., Uversky, V. N., Gerday, C., Feller, G. Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. J. Biol. Chem. 2003, 278, 37015-37023.

[4] Leiros, I., Moe, E., Lanes, O., Smalås, A. O., Willassen, N. P. The structure of uracil-DNA glycosylase from atlantic cod (gadus morhua) reveals cold-adaptation features. Acta Crystallogr. D. 2003, 59, 1357-1365.

[5] Olufsen, M., Smalås, A. O., Moe, E., Brandsdal, B. O. Increased flexibility as a strategy for cold adaptation - a comparative molecular dynamics study of cold- and warm-active uracil DNA glycosylase. J. Biol. Chem. 2005, 280, 18042-18048.

[6] Smalås, A. O., Heimstad, E. S., Hordvik, A., Willassen, N. P., Male, R. Cold adaption of enzymes - structural comparison between salmon and bovine trypsins. Proteins. 1994, 20, 149-166.

[7] Brandsdal, B. O., Heimstad, E. S., Sylte, I., Smalås, A. O. Comparative molecular dynamics of mesophilic and psychrophilic protein homologues studied by 1.2 ns simulations. J. Biomol. Struct. Dyn. 1999, 17, 493-506.

[8] Russell, R. J. M., Gerike, U., Danson, M. J., Hough, D. W., Taylor, G. L. Structural adaptations of the cold-active citrate synthase from an antarctic bacterium. Structure. 1998, 6, 351-361.

[9] Kumar, S., Nussinov, R. Different roles of electrostatics in heat and in cold: Adaptation by citrate synthase. Chembiochem. 2004, 5, 280-290.

[10] Gorfe, A. A., Brandsdal, B. O., Leiros, H. K. S., Helland, R., Smalås, A. O. Electrostatics of mesophilic and psychrophilic trypsin isoenzymes: Qualitative evaluation of electrostatic differences at the substrate binding site. Proteins. 2000, 40, 207-217.

[11] Brandsdal, B. O., Smalås, A. O., Åqvist, J. Electrostatic effects play a central role in cold adaptation of trypsin. FEBS Lett. 2001, 499, 171-175.

[12] Moe, E., Leiros, I., Riise, E. K., Olufsen, M., Lanes, O., Smalas, A., Willassen, N. P. Optimisation of the surface electrostatics as a strategy for cold adaptation of uracil-DNA n-glycosylase (ung) from atlantic cod (gadus morhua). J. Mol. Biol. 2004, 343, 1221-1230.

[13] Lindahl, T., Nyberg, B. Heat-induced deamination of cytosine residues in deoxyribonucleic-acid. Biochemistry. 1974, 13, 3405-3410.

[14] Krokan, H. E., Standal, R., Slupphaug, G. DNA glycosylases in the base excision repair of DNA. Biochem. J. 1997, 325, 1-16.

[15] Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., Tainer, J. A. Crystal-structure and mutational analysis of human uracil-DNA glycosylase - structural basis for specificity and catalysis. Cell. 1995, 80, 869-878.

[16] Savva, R., Mcauleyhecht, K., Brown, T., Pearl, L. The structural basis of specific base-excision repair by uracil-DNA glycosylase. Nature. 1995, 373, 487-493.
[17] Ravishankar, R., Sagar, M. B., Roy, S., Purnapatre, K., Handa, P., Varshney, U., Vijayan, M. X-ray analysis of a complex of escherichia coli uracil DNA glycosylase (ecudg) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic udg. Nucleic Acids Res. 1998, 26, 4880-4887.
[18] Geoui, T., Buisson, M., Tarbouriech, N., Burmeister, W. P. New insights on the role of the gamma-herpesvirus uracil-DNA glycosylase leucine loop revealed by the structure of the epstein-barr virus enzyme in complex with an inhibitor protein. J Mol Biol. 2007, 366, 117-131.
[19] Bianchet, M. A., Seiple, L. A., Jiang, Y. L., Ichikawa, Y., Amzel, L. M., Stivers, J. T. Electrostatic guidance of glycosyl cation migration along the reaction coordinate of uracil DNA glycosylase. Biochemistry. 2003, 42, 12455-12460.
[20] Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., Tainer, J. A. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. EMBO J. 1998, 17, 5214-5226.
[21] Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., Tainer, J. A. Uracil-DNA glycosylase-DNA substrate and product structures: Conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. Proc. Natl. Acad. Sci. U. S. A. 2000, 97, 5083-5088.
[22] Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., Tainer, J. A. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. Nature. 1996, 384, 87-92.
[23] Lanes, O., Leiros, I., Smalås, A. O., Willassen, N. P. Identification, cloning, and expression of uracil-DNA glycosylase from atlantic cod (gadus morhua): Characterization and homology modeling of the cold-active catalytic domain. Extremophiles. 2002, 6, 73-86.
[24] Dinner, A. R., Blackburn, G. M., Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. Nature. 2001, 413, 752-755.
[25] Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., Ferguson, D., Seibel, G., Kollman, P. Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. Comput. Phys. Commun. 1995, 91, 1-41.
[26] Wang, J. M., Cieplak, P., Kollman, P. A. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. 2000, 21, 1049-1074.
[27] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 1983, 79, 926-935.
[28] Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., Haak, J. R. Molecular-dynamics with coupling to an external bath. J. Chem. Phys. 1984, 81, 3684-3690.
[29] Darden, T., York, D., Pedersen, L. Particle mesh ewald - an n.Log(n) method for ewald sums in large systems. J. Chem. Phys. 1993, 98, 10089-10092.

[30] Ryckaert, J. P., Ciccotti, G., Berendsen, H. J. C. Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes. J. Comput. Phys. 1977, 23, 327-341.
[31] Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S. H., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., Cheatham, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. Acc. Chem. Res. 2000, 33, 889-897.
[32] Massova, I., Kollman, P. A. Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies. J. Am. Chem. Soc. 1999, 121, 8133-8143.
[33] Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., Case, D. A. Continuum solvent studies of the stability of DNA, rna, and phosphoramidate - DNA helices. J. Am. Chem. Soc. 1998, 120, 9401-9409.
[34] Luo, R., David, L., Gilson, M. K. Accelerated poisson-boltzmann calculations for static and dynamic systems. J. Comput. Chem. 2002, 23, 1244-1253.
[35] Onufriev, A., Bashford, D., Case, D. A. Modification of the generalized born model suitable for macromolecules. J. Phys. Chem. B. 2000, 104, 3712-3720.
[36] Onufriev, A., Bashford, D., Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins. 2004, 55, 383-394.
[37] Peter, C., Oostenbrink, C., van Dorp, A., van Gunsteren, W. F. Estimating entropies from molecular dynamics simulations. J. Chem. Phys. 2004, 120, 2652-2661.
[38] Case, D. A. Normal-mode analysis of protein dynamics. Curr. Opin. Struc. Biol. 1994, 4, 285-290.
[39] Jayaram, B., Sprous, D., Beveridge, D. L. Solvation free energy of biomacromolecules: Parameters for a modified generalized born model consistent with the amber force field. J. Phys. Chem. B. 1998, 102, 9571-9576.
[40] Connolly, M. L. Analytical molecular-surface calculation. J. Appl. Crystallogr. 1983, 16, 548-558.
[41] Weiser, J., Shenkin, P. S., Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (lcpo). J. Comput. Chem. 1999, 20, 217-230.
[42] Rocchia, W., Alexov, E., Honig, B. Extending the applicability of the nonlinear poisson-boltzmann equation: Multiple dielectric constants and multivalent ions. J. Phys. Chem. B. 2001, 105, 6507-6514.
[43] Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., Honig, B. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. J. Comput. Chem. 2002, 23, 128-137.
[44] Moreira, I. S., Fernandes, P. A., Ramos, M. J. Accuracy of the numerical solution of the poisson-boltzmann equation. J Mol Struc-Theochem. 2005, 729, 11-18.
[45] Jiang, Y. L., Ichikawa, Y., Song, F., Stivers, J. T. Powering DNA repair through substrate-electrostatic interactions. Biochemistry. 2003, 42, 1922-1929.
[46] Honig, B., Nicholls, A. Classical electrostatics in biology and chemistry. Science. 1995, 268, 1144-1149.

[47] Koehl, P. Electrostatics calculations: Latest methodological advances. Curr. Opin. Struc. Biol. 2006, 16, 142-151.
[48] Warshel, A., Sharma, P. K., Kato, M., Parson, W. W. Modeling electrostatic effects in proteins. Bba-Proteins Proteom. 2006, 1764, 1647-1676.
[49] Fersht, A. Structure and mechanism in protein science. W.H. Freeman and Company, 1999.
[50] Brandsdal, B. O., Osterberg, F., Almlof, M., Feierberg, I., Luzhkov, V. B., Aqvist, J. Free energy calculations and ligand binding. Adv. Prot. Chem. 2003, 66, 123-158.
[51] Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein ligand complex of known 3-dimensional structure. J. Comput. Aid. Mol. Des. 1994, 8, 243-256.
[52] Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., Mee, R. P. Empirical scoring functions.1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J. Comput. Aid. Mol. Des. 1997, 11, 425-445.
[53] Jain, A. N. Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. J. Comput. Aid. Mol. Des. 1996, 10, 427-440.
[54] Åqvist, J., Medina, C., Samuelsson, J. E. New method for predicting binding-affinity in computer-aided drug design. Protein Eng. 1994, 7, 385-391.
[55] Brigo, A., Lee, K. W., Fogolari, F., Mustata, G. L., Briggs, J. M. Comparative molecular dynamics simulations of hiv-1 integrase and the t66i/m154i mutant: Binding modes and drug resistance to a diketo acid inhibitor. Proteins. 2005, 59, 723-741.
[56] Kuhn, B., Kollman, P. A. A ligand that is predicted to bind better to avidin than biotin: Insights from computational fluorine scanning. J. Am. Chem. Soc. 2000, 122, 3909-3916.
[57] Adekoya, O. A., Willassen, N. P., Sylte, I. The protein-protein interactions between smpi and thermolysin studied by molecular dynamics and mm/pbsa calculations. J. Biomol. Struct. Dyn. 2005, 22, 521-531.
[58] Luo, C., Xu, L. F., Zheng, S. X., Luo, Z., Jiang, X. M., Shen, J. H., Jiang, H. L., Liu, X. F., Zhou, M. D. Computational analysis of molecular basis of 1: 1 interactions of nrg-1 beta wild-type and variants with erbb3 and erbb4. Proteins. 2005, 59, 742-756.
[59] Wang, W., Kollman, P. A. Free energy calculations on dimer stability of the hiv protease using molecular dynamics and a continuum solvent model. J. Mol. Biol. 2000, 303, 567-582.
[60] Reyes, C. M., Kollman, P. A. Structure and thermodynamics of rna-protein binding: Using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. J. Mol. Biol. 2000, 297, 1145-1158.
[61] Zhang, Q., Schlick, T. Stereochemistry and position-dependent effects of carcinogens on tata/tbp binding. Biophys. J. 2006, 90, 1865-1877.
[62] Gohlke, H., Case, D. A. Converging free energy estimates: Mm-pb(gb)sa studies on the protein-protein complex ras-raf. J. Comput. Chem. 2004, 25, 238-250.

[63] Chen, C. Y., Mosbaugh, D. W., Bennett, S. E. Mutations at arginine 276 transform human uracil-DNA glycosylase into a single-stranded DNA-specific uracil-DNA glycosylase. DNA Repair. 2005, 4, 793-805.
[64] Jiang, Y. L., Kwon, K., Stivers, J. T. Turning on uracil-DNA glycosylase using a pyrene nucleotide switch. J. Biol. Chem. 2001, 276, 42347-42354.
[65] Krosky, D. J., Song, F. H., Stivers, J. T. The origins of high-affinity enzyme binding to an extrahelical DNA base. Biochemistry. 2005, 44, 5949-5959.
[66] Stivers, J. T., Pankiewicz, K. W., Watanabe, K. A. Kinetic mechanism of damage site recognition and uracil flipping by escherichia coli uracil DNA glycosylase. Biochemistry. 1999, 38, 952-963.
[67] Cao, C. Y., Jiang, Y. L., Stivers, J. T., Song, F. H. Dynamic opening of DNA during the enzymatic search for a damaged base. Nat. Struct. Mol. Biol. 2004, 11, 1230-1236.
[68] Parker, C. N., Halford, S. E. Dynamics of long-range interactions on DNA - the speed of synapsis during site-specific recombination by resolvase. Cell. 1991, 66, 781-791.
[69] Pearl, L. H. Structure and function in the uracil-DNA glycosylase superfamily. Mutat. Res-DNA Rep. 2000, 460, 165-181.
[70] Cao, C. Y., Jiang, Y. L., Krosky, D. J., Stivers, J. T. The catalytic power of uracil DNA glycosylase in the opening of thymine base pairs. J. Am. Chem. Soc. 2006, 128, 13034-13035.
[71] Karplus, M., Kushick, J. N. Method for estimating the configurational entropy of macromolecules. Macromolecules. 1981, 14, 325-332.
[72] Schafer, H., Daura, X., Mark, A. E., van Gunsteren, W. F. Entropy calculations on a reversibly folding peptide: Changes in solute free energy cannot explain folding behavior. Proteins. 2001, 43, 45-56.
[73] Schafer, H., Mark, A. E., van Gunsteren, W. F. Absolute entropies from molecular dynamics simulation trajectories. J. Chem. Phys. 2000, 113, 7809-7817.
[74] Kuhn, B., Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. J Med Chem. 2000, 43, 3786-3791.
[75] Jiang, Y. L., Drohat, A. C., Ichikawa, Y., Stivers, J. T. Probing the limits of electrostatic catalysis by uracil DNA glycosylase using transition state mimicry and mutagenesis. J. Biol. Chem. 2002, 277, 15385-15392.
[76] Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W., Tainer, J. A. Crystal-structure of human uracil-DNA glycosylase in complex with a protein inhibitor - protein mimicry of DNA. Cell. 1995, 82, 701-708.
[77] Wong, I., Lundquist, A. J., Bernards, A. S., Mosbaugh, D. W. Presteady-state analysis of a single catalytic turnover by escherichia coli uracil-DNA glycosylase reveals a "pinch-pull-push" mechanism. J. Biol. Chem. 2002, 277, 19424-19432.
[78] DeLano, W. L., The pymol molecular graphics system, 2002, DeLano Scientific, San Carlos, CA, USA

Tables

Table 1

Binding free energies of the cUDG-DNA complex[a,b] computed with the MM-PBSA approach using single and multiple trajectories.

Contribution[c]	cUDG-DNA	cUDG	DNA	Delta[d]
H_{elec}(single)	-7913.8 ± 91.7	-6118.4 ± 78.9	350.7 ± 36.9	-2146.1 ± 74.6
H_{vdW}(single)	-1235.2 ± 26.4	-1012.4 ± 23.8	-133.5 ± 9.6	-89.3 ± 6.5
H_{int}(single)	5618.7 ± 45.7	4745.2 ± 41.9	873.5 ± 18.3	0.0 ± 0.0
G_{np}(single)	95.8 ± 1.2	76.7 ± 0.8	30.0 ± 0.6	-10.9 ± 0.6
G_{pol}(single)	-4683.2 ± 80.8	-2615.7 ± 70.5	-4221.3 ± 33.6	2153.8 ± 73.0
$G_{gas+solv}$(single)	-8117.7 ± 50.4	-4924.6 ± 44.9	-3100.6 ± 17.4	-92.5 ± 9.5
TS_{tot}(single)	2955.1 ± 15.8	2498.6 ± 13.7	509.2 ± 5.8	-52.7 ± 20.7
G_{tot}(single)	-11072.8 ± 52.8	-7423.2 ± 46.9	-3609.8 ± 18.4	-39.8 ± 22.7
H_{elec}(multi)	-7913.8 ± 91.7	-6081.7 ± 74.9	355.4 ± 44.9	-2187.5 ± 122.4
H_{vdw}(multi)	-1235.2 ± 26.4	-1001.8 ± 27.3	-139.9 ± 9.4	-93.5 ± 40.3
H_{int}(multi)	5618.7 ± 45.7	4731.6 ± 45.2	871.6 ± 18.4	15.5 ± 64.5
G_{np}(multi)	95.8 ± 1.2	79.1 ± 1.4	29.7 ± 0.6	-13.0 ± 2.1
G_{pol}(multi)	-4683.2 ± 80.8	-2650.6 ± 72.9	-4224.6 ± 41.4	2192.0 ± 114.3
$G_{gas+solv}$(multi)	-8117.7 ± 50.4	-4923.4 ± 44.5	-3107.8 ± 17.3	-86.5 ± 65.2
TS_{tot}(multi)	2955.1 ± 15.8	2504.8 ± 13.7	508.2 ± 5.5	-57.9 ± 21.5
G_{tot}(multi)	-11072.8 ± 52.8	-7428.2 ± 46.5	-3616.0 ± 18.1	-28.7 ± 68.7

[a]All values are given in kcal/mol.

[b]Mean value calculated from 500 snapshots with standard deviations.

[c]H_{elec}: Coulombic energy, H_{vdW}: van der Waals energy, H_{int}: internal energy, G_{np}: nonpolar solvation free energy, G_{pol}: polar solvation free energy, $G_{gas+solv} = H_{elec} + H_{vdW} + H_{int} + G_{np} + G_{pol}$, TS_{tot}: total entropy contribution, $G_{tot} = G_{gas+solv} + TS_{tot}$.

[d]Delta = (UDG-DNA) – (UDG) – (DNA)

Table 2

Binding free energies of the hUDG-DNA complex[a,b] computed from single and multiple trajectories MM-PBSA calculations.

Contribution[c]	hUDG-DNA	hUDG	DNA	Delta[d]
H_{elec}(single)	-7452.0 ± 85.9	-5838.8 ± 75.3	325.8 ± 38.6	-1939.0 ± 65.5
H_{vdW}(single)	-1241.1 ± 24.8	-1014.7 ± 24.4	-130.0 ± 8.7	-96.4 ± 7.6
H_{int}(single)	5671.7 ± 46.3	4796.2 ± 42.9	875.5 ± 18.8	0.0 ± 0.0
G_{np}(single)	98.3 ± 1.2	80.0 ± 1.1	30.4 ± 0.4	-12.1 ± 0.8
G_{pol}(single)	-4880.2 ± 74.9	-2630.8 ± 64.0	-4203.9 ± 36.3	1954.5 ± 66.9
$G_{gas+solv}$(single)	-7803.3 ± 48.0	-4608.1 ± 43.1	-3102.2 ± 17.3	-92.7 ± 10.3
TS(single)	2981.6 ± 17.5	2527.8 ± 14.3	512.2 ± 5.7	-58.4 ± 21.6
G_{tot}(single)	-10784.9 ± 51.1	-7136.2 ± 45.4	-3614.4 ± 18.2	-34.3 ± 24.1
H_{elec}(multi)	-7452.0 ± 85.9	-5804.0 ± 80.9	355.4 ± 44.9	-2003.4 ± 130.9
H_{vdw}(multi)	-1241.1 ± 24.8	-1009.0 ± 24.9	-139.9 ± 9.4	-92.2 ± 36.7
H_{int}(multi)	5671.7 ± 46.3	4802.3 ± 43.0	871.6 ± 18.4	-2.2 ± 64.9
G_{np}(multi)	98.3 ± 1.2	81.1 ± 1.2	29.7 ± 0.6	-12.5 ± 1.7
G_{pol}(multi)	-4880.2 ± 74.9	-2685.4 ± 78.4	-4224.6 ± 41.4	2029.8 ± 119.0
$G_{gas+solv}$(multi)	-7803.3 ± 48.0	-4615.0 ± 44.6	-3107.8 ± 17.3	-80.5 ± 68.0
TS(multi)	2981.6 ± 17.5	2528.6 ± 14.6	508.2 ± 5.5	-55.2 ± 23.1
G_{tot}(multi)	-10784.9 ± 51.1	-7143.6 ± 46.9	-3616.0 ± 18.1	-25.3 ± 71.8

[a]All values are given in kcal/mol.

[b]Mean value calculated from 500 snapshots with standard deviations.

[c]H_{elec}: Coulombic energy, H_{vdW}: van der Waals energy, H_{int}: internal energy, G_{np}: nonpolar solvation free energy, G_{pol}: polar solvation free energy, $G_{gas+solv} = H_{elec} + H_{vdW} + H_{int} + G_{np} + G_{pol}$, TS_{tot}: total entropy contribution, $G_{tot} = G_{gas+solv} + TS_{tot}$.

[d]Delta = (UDG-DNA) – (UDG) – (DNA)

Table 3

Binding free energies[a] calculated for the first[b] and second[c] half of the UDG-DNA simulations and their contributions.

Enzyme	Contributions[d]	First half	Second half
cUDG	ΔH_{gas}	-2231.3 ± 129.9	-2299.7 ± 131.6
cUDG	ΔG_{solv}	2148.2 ± 104.9	2210.3 ± 113.5
cUDG	$T\Delta S_{tot}$	-57.4 ± 20.6	-58.3 ± 22.4
cUDG	ΔG_{tot}	-25.6 ± 69.8	-31.1 ± 67.5
hUDG	ΔH_{gas}	-2071.5 ± 155.2	-2124.3 ± 127.4
hUDG	ΔG_{solv}	1991.4 ± 130.7	2043.8 ± 98.7
hUDG	$T\Delta S_{tot}$	-56.2 ± 23.0	-54.1 ± 23.2
hUDG	ΔG_{tot}	-23.9 ± 71.1	-26.4 ± 72.7

[a]All values are given in kcal/mol.

[b]Mean value calculated using 250 snapshots with standard deviations.

[c]First half includes snapshots 1-250, while second half includes snapshots 251-500.

[d]ΔH_{gas}: gas phase energy, ΔG_{solv}: solvation free energy, TS_{tot}: total entropy contribution, ΔG_{tot}: $G_{gas+solv} + TS_{tot}$.

Table 4

Contributions[a] to the free energy of binding[b,c] from residues in the Leu272 loop in hUDG.

Residue	ΔG_{elec}	ΔG_{vdW}	ΔG_{pol}	ΔG_{np}	$\Delta G_{gas+solv}$
His268	-1.5	-3.1	0.1	-0.2	-4.7 ± 0.9
Pro269	1.8	-0.5	-1.6	0.0	-0.3 ± 0.1
Ser270	-7.2	-2.3	7.3	-0.2	-2.5 ± 0.8
Pro271	-0.3	-3.3	1.5	-0.4	-2.4 ± 0.5
Leu272	-0.2	-7.9	0.6	-1.1	-8.5 ± 0.8
Ser273	-2.8	-1.9	4.4	-0.2	-0.5 ± 0.9
Val274	-3.1	-0.4	2.9	0.0	-0.6 ± 0.2
Tyr275	-4.0	-2.8	6.1	-0.5	-1.2 ± 0.8
Arg276	-219.9	-3.5	219.4	-0.7	-4.7 ± 2.0
Sum	-237.2	-25.7	240.6	-3.3	-25.5 ± 2.8
Average	-26.4	-2.9	26.7	-0.4	-2.8 ± 0.8

[a]H_{elec}: Coulombic energy, H_{vdW}: van der Waals energy, G_{np}: nonpolar solvation free energy, G_{pol}: polar solvation free energy, $G_{gas+solv} = H_{elec} + H_{vdW} + H_{int} + G_{np} + G_{pol}$.

[b]All values are given in kcal/mol.

[c]Mean value calculated from 500 snapshots with standard deviations for the ΔG_{tot}.

Table 5

Contributions[a] to the free energy of binding[b,c] from residues in the Leu272 loop in cUDG.

Residue	ΔG_{elec}	ΔG_{vdW}	ΔG_{pol}	ΔG_{np}	$\Delta G_{gas+solv}$
His268	-4.1	-2.9	1.8	-0.2	-5.4 ± 0.9
Pro269	1.2	-0.4	-1.0	0.0	-0.3 ± 0.2
Ser270	-6.5	-2.4	6.6	-0.2	-2.5 ± 0.9
Pro271	0.0	-2.7	0.7	-0.3	-2.3 ± 0.5
Leu272	0.7	-7.6	-0.9	-1.0	-8.8 ± 0.9
Ser273	-4.3	-1.6	4.8	-0.2	-1.4 ± 1.3
Ala274	-3.1	-0.2	3.1	0.0	-0.2 ± 0.1
His275	-13.0	-1.7	12.2	-0.3	-2.9 ± 0.8
Arg276	-237.7	-2.1	231.4	-0.6	-9.0 ± 1.5
Sum	-266.8	-21.7	258.6	-2.7	-32.6 ± 2.7
Average	-29.6	-2.4	28.7	-0.3	-3.6 ± 0.8

[a]H_{elec}: Coulombic energy, H_{vdW}: van der Waals energy, G_{np}: nonpolar solvation free energy, G_{pol}: polar solvation free energy, $G_{gas+solv} = H_{elec} + H_{vdW} + H_{int} + G_{np} + G_{pol}$.

[b]All values are given in kcal/mol.

[c]Mean value calculated from 500 snapshots with standard deviations for the ΔG_{tot}.

FIGURE LEGENDS

Fig. 1.

Structural differences between uracil and pseudo uracil.

Fig. 2

Electrostatic isosurfaces of cUDG and DNA. The DNA was moved 15 Å out of the specificity pocket as observed in the model of cUDG-DNA. The isocontour surface of cUDG was set to -5kT/e (red) and 5kT (blue), while the isocontour surface of DNA was set to -3kT/e (red) and 3kT (blue) The figure was generated using pymol [78].

Fig. 3

Electrostatic isosurfaces of cUDG and hUDG. A and B show cUDG and hUDG bound to dsDNA, and C and D show cUDG and hUDG from the opposite side of the specificity pocket, respectively. The isocontour surfaces correspond to -2kT/e (red) and 4kT (blue). The figure was generated using pymol [78].

Fig. 4

Sum of energies from gas phase and solvation free energies calculated for 500 snapshots of the cUDG-DNA complex. The solvation free energies were calculated with the PB method. The snapshots were taken from separate trajectories. The circles, triangles and the squares represent DNA, cUDG and cUDG-DNA complex, respectively.

Fig. 5

Free energy of binding per residue of cUDG (black dashed line) and hUDG (gray) in complex with DNA with important residues highlighted.

Fig. 6

Stereographic illustration of the interactions between the Leu272 loop and the DNA as observed in the MD simulations. For simplicity only the DNA nucleotides that interact with the Leu272 loop are shown. The bases are removed for all nucleotides except the

uracil base. All residues in the Leu272 loop (residue 268-276) for both cUDG (blue) and hUDG (red) are shown. The DNA from the cUDG-DNA and the hUDG-DNA simulations are shown in light blue and orange, respectively. Only hydrogen bonds between UDG and DNA shorter than 3.4 Å are shown. The figure was generated in PyMol [78].

FIGURES

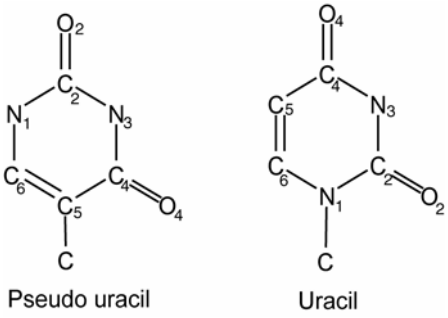

Fig. 1.

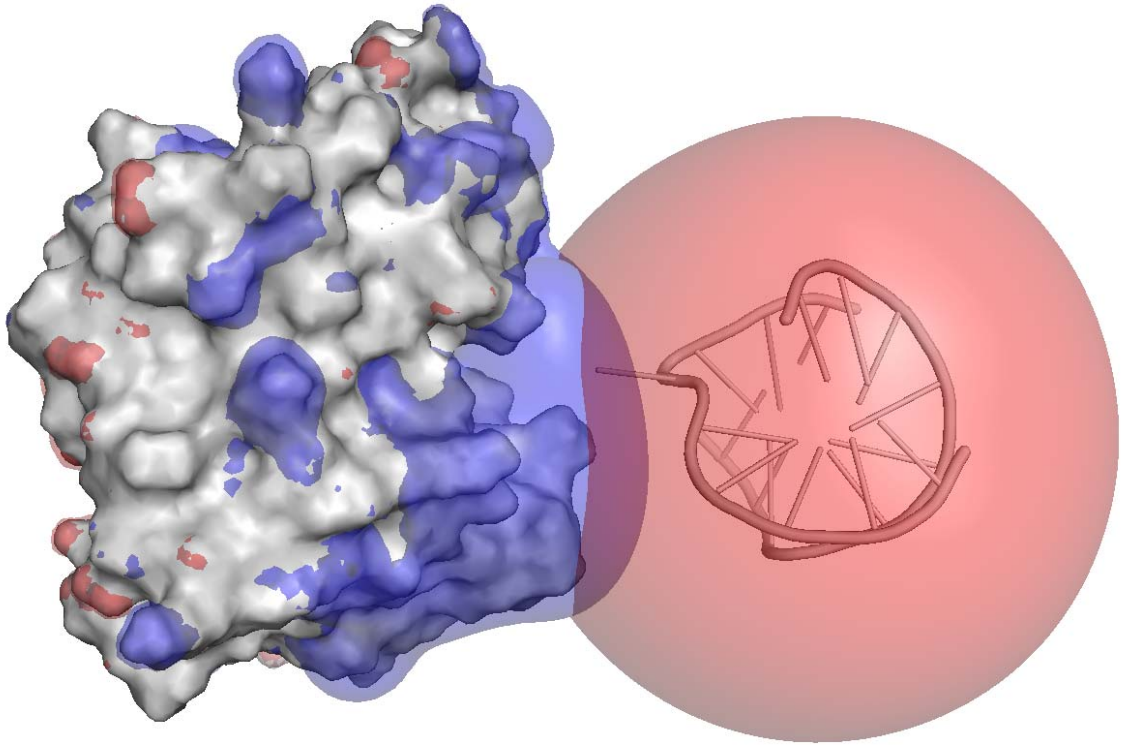

Fig. 2

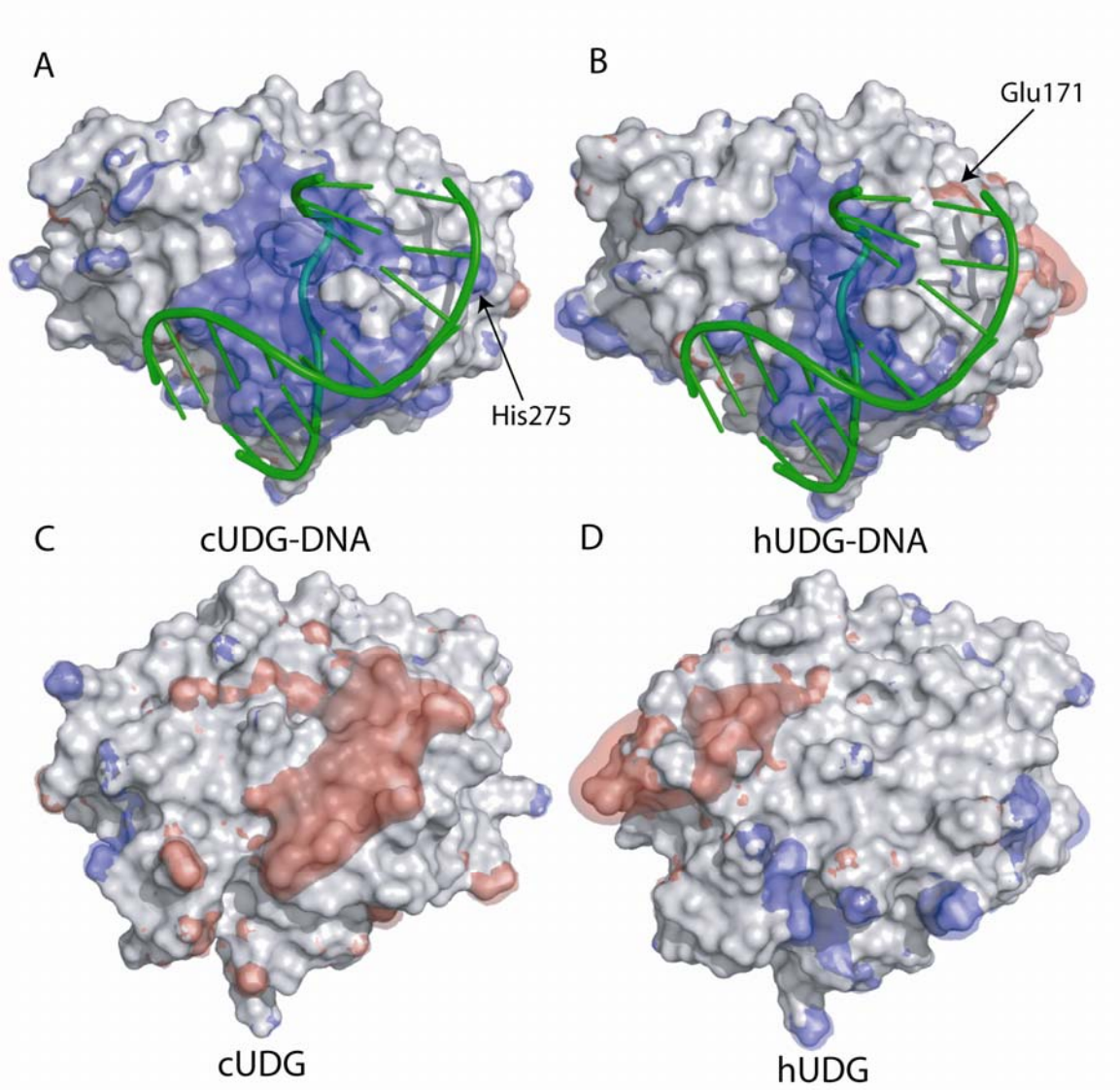

Fig. 3

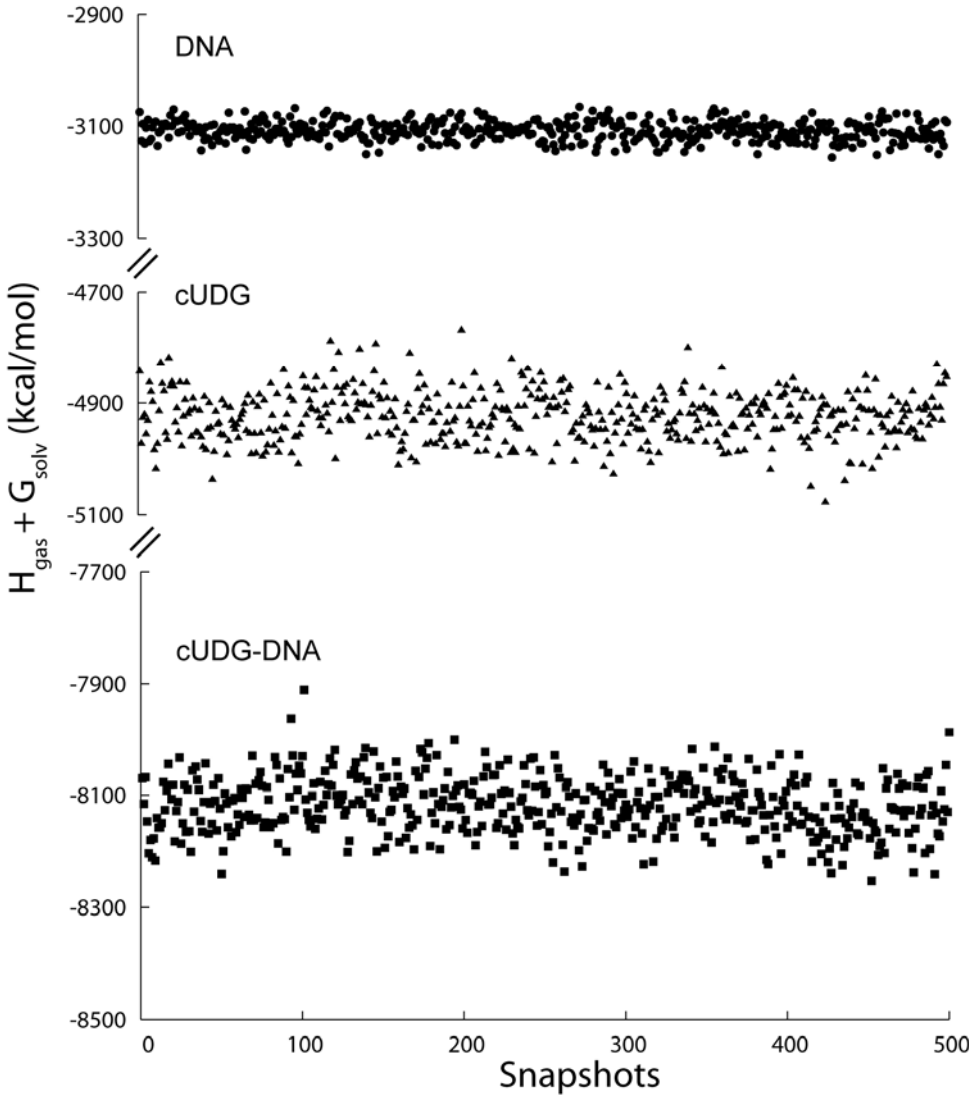
DNA
cUDG
cUDG-DNA
-2900
-3100
-3300
-4700
-4900
-5100
-7700
-7900
-8100
-8300
-8500
0
100
200
300
400
500
Snapshots
H_gas + G_solv (kcal/mol)

Fig. 4

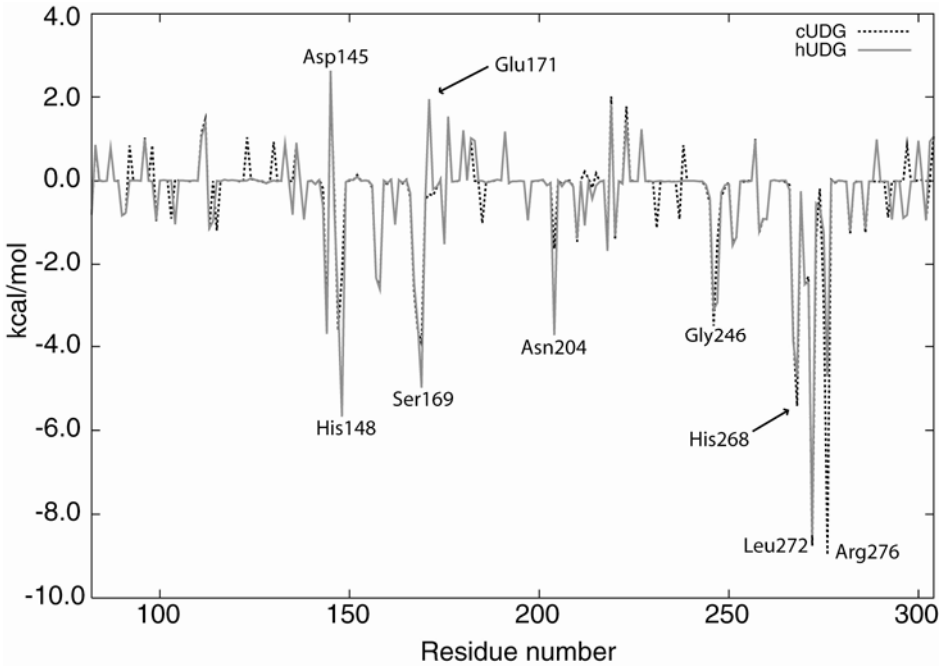
cUDG
hUDG
Asp145
Glu171
Asn204
Gly246
Ser169
His148
His268
Leu272
Arg276
kcal/mol
4.0
2.0
0.0
-2.0
-4.0
-6.0
-8.0
-10.0
100
150
200
250
300
Residue number

Fig. 5

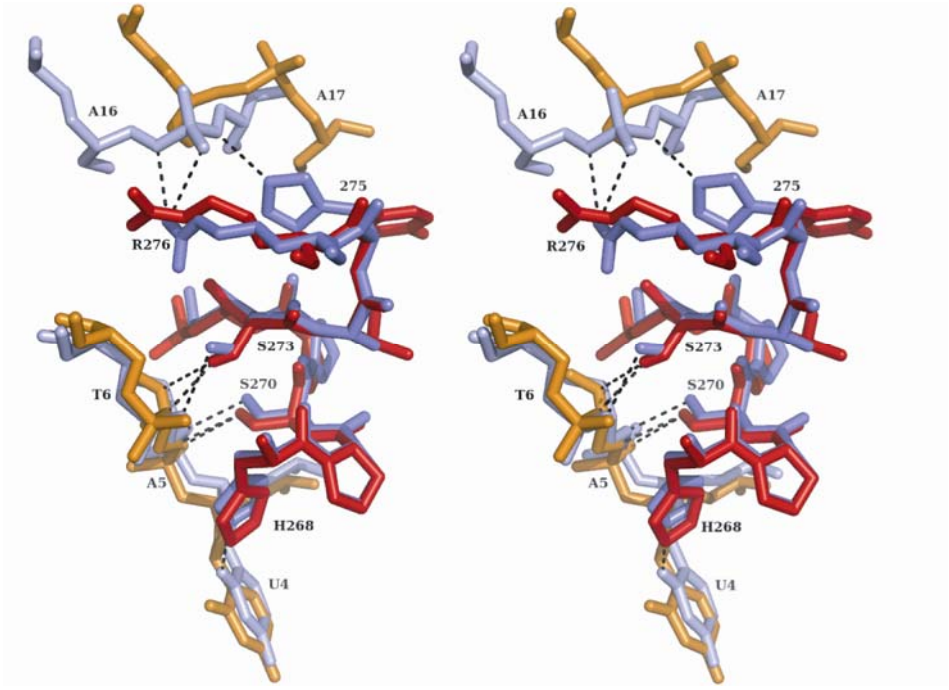
A16
A17
275
R276
S273
S270
T6
A5
H268
U4
A16
A17
275
R276
S273
S270
T6
A5
H268
U4

Fig. 6