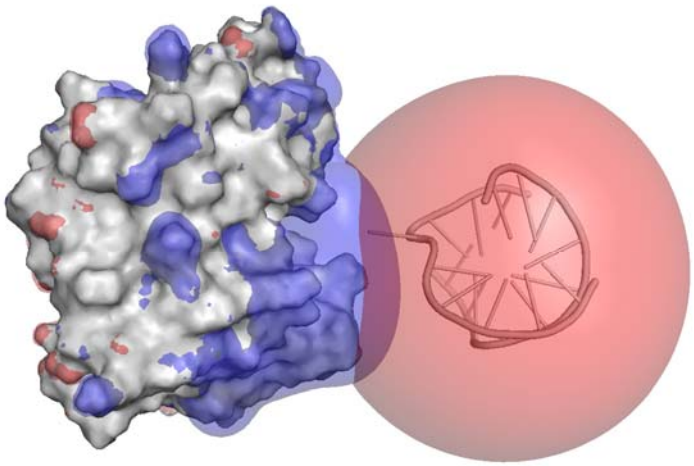
A DISSERTATION FOR THE DEGREE OF PHILOSOPHIAE DOCTOR

The molecular origin of cold adaptation:

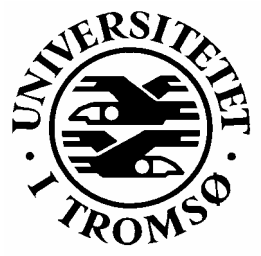
A comparative study of cold- and warm-active uracil DNA glycosylase

Magne Olufsen

April 2007



Department of Chemistry, Faculty of Science,
University of Tromsø, Norway



A DISSERTATION FOR THE DEGREE OF PHILOSOPHIAE DOCTOR

The molecular origin of cold adaptation:

A comparative study of cold- and warm-active uracil DNA glycosylase

Magne Olufsen

April 2007

Department of Chemistry, Faculty of Science,
University of Tromsø, Norway

Acknowledgements

The work leading to this thesis has been carried out at the Norwegian Structural Biology Centre (NorStruct), Department of Chemistry, University of Tromsø, Norway, from January 2003 to April 2007. Financial support was given from the Research Council of Norway (NFR), which is greatly acknowledged.

I wish to express my deepest gratitude to my supervisors Professor Arne O. Smalås and Dr. Bjørn Olav Brandsdal. Arne, your structured leadership, pleasant personality, supportive nature and your wise guidance has made it a pleasure to work for you. Bjørn Olav, your optimistic nature, wise guidance, expertise in the field and our scientific discussions have been very inspiring and of great help through these years.

I would also like to thank all the former and present group members both at Department of Chemistry and at NorStruct for creating a stimulating working environment.

I am grateful to my mother Gunn Robertsen and Odd Valdermo for endless hours of child care, which has made it possibly for me to focus more on my research. A warm gratitude goes to my girlfriend Vibeke Aune. Your love, care and encouragement, especially in down periods, have meant a lot to me. Finally, I want to thank my daughter Martine. You have shown exceptional patience when I had to work late, travel abroad or when you had to stay at the University with me.

Magne Olufsen

Tromsø, April 2007

Abbreviations

Throughout the text the standard 3-letter abbreviations for amino acids are used. Other abbreviations are listed below:

UDG	Uracil DNA glycosylase
cUDG	Cod uracil DNA glycosylase
hUDG	Human uracil DNA glycosylase
MD	Molecular dynamics
MM	Molecular mechanics
MM-PBSA	Molecular-mechanics/Poisson-Boltzmann/surface area
r.m.s.d.	Root mean squared deviations (starting structure as reference)
r.m.s.f.	Root mean squared fluctuations (average structure as reference)
ε_p	Dielectric constant of the protein
PB	Poisson-Boltzmann
GB	Generalized-Born
LIE	Linear interaction energy
E. coli	Escherichia coli

List of papers

I. Moe, E., Leiros, I., Riise, E.K., **Olufsen, M**., Lanes, O., Smalås, A.O. & 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

Comments: My contribution to this paper has been to crystallize and solve the structure of the hUDG-E171V mutant.

II. **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

III. **Olufsen, M**., Brandsdal, B.O. & Smalås, A.O. Comparative unfolding studies of psychrophilic and mesophilic uracil DNA glycosylase: MD simulations show reduced thermal stability of the cold-adapted enzyme. J. Mol. Graph. Model. (2007), in press.

IV. **Olufsen, M**., Papaleo, E., Smalås, A.O. & Brandsdal, B.O. Ion pairs and their role in modulating stability of cold- and warm-active uracil DNA glycosylase. Submitted to Proteins.

V. **Olufsen, M**., Smalås, A.O. & Brandsdal, B.O. Electrostatic interactions play an essential role in DNA repair and cold-adaptation of Uracil DNA Glycosylase. Manuscript.

Reprints were made with permission from the copyright holders

Related publications

i. Mekonnen, S.M., **Olufsen M**., Smalås, A.O. & Brandsdal, B.O. Predicting proteinase specificities from free energy calculations. J. Mol. Graph. Model. (2006). **25**, 176-185.

ii. Papaleo, E., **Olufsen, M**., De Gioia, L. & Brandsdal, B.O. Optimization of electrostatics as a strategy for cold-adaptation: A case study of cold- and warm-active elastases. J. Mol. Graph. Model. (2007), in press.

Table of contents

Background

During the last decade or so, scientists at the Norwegian Structural Biology Centre (NorStruct) have used the enzyme uracil DNA glycosylase (UDG) as a model system in the study of cold-adaptation, protein-DNA recognition and enzyme specificity. The cod UDG (cUDG) and human UDG (hUDG) are thoroughly biologically characterized and there are established well-functioning recombinant expression systems to produce mutants. More than 30 mutants of cUDG and hUDG have been expressed, purified and characterized. The kinetic constants (k_{cat} and K_m) have been obtained for the majority of these mutants. In addition, the crystal structure has been determined for some of them. This Ph.D. project is an extension of earlier work done at NorStruct.

Enzyme kinetics

Enzymes are protein catalysts which speed up rates of chemical reactions by temporarily binding to the substrate (target molecule) and lower the activation energy needed to convert the substrate to product. Michaelis and Menten developed in 1913 the typical scheme for enzymatic reactions [1]:

$$(I) \qquad E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \overset{k_2}{\rightarrow} E + P$$

where E, S and P are enzyme, substrate and product, respectively. In the simple Michaelis-Menten kinetics with only one enzyme-substrate complex and all binding steps are fast, then k_2 is equal to k_{cat}. Enzymatic reactions are usually monitored in terms of the substrate concentration ([S]) and are analyzed in terms of velocity (v):

$$(II) \qquad v = \frac{-d[S]}{dt}$$

The steady-state approximation gives the velocity as a function of substrate consentration:

$$(III) \qquad v = \frac{[E]_0 \cdot [S] \cdot k_{cat}}{K_m + [S]}$$

where k_{cat} and K_m are the turnover number and the Michaelis constant, respectively. The catalytic efficiency of an enzymatic reaction is defined as k_{cat}/K_m, and in order for an enzyme to increase its catalytic efficiency it can either increase k_{cat}, decrease K_m or adjust both parameters [2]. The Michaelis constant is linked to the binding of the enzyme-substrate:

$$(IV) \qquad K_m = K_S + \frac{k_2}{k_1}$$

where K_S is the dissociation constant (inverse of the binding constant). When k_2 is the rate limiting step ($k_2 << k_{-1}$), then K_m is equal to K_s. Equation IV is also valid for Briggs-Haldane kinetics (when $k_2 >> k_{-1}$) [1]. The Michaelis-Menten scheme may be extended to cover a wide range of different cases with additional intermediates in the reaction pathway, but the Michaelis-Menten still applies [1]. In these cases K_m is always less or

equal to K_S. There is still thus an inverse relationship between the binding energy of the enzyme-substrate complex and the K_m value. k_{cat}/K_m is also called the second order rate constant, and with low substrate concentration ([S] << K_m), equation III can be reduced to [1]:

$$(V) \qquad v = \frac{k_{cat}}{K_m} \cdot [E]_0 \cdot [S]$$

Enzymes can also use binding energy to lower the activation energy of an enzymatic reactions, instead of lowering the K_m [1,3]. If the enzyme binds stronger to the transition state relative to the ground state then the binding energy will lower the activation energy. In such cases it is much better to compare k_{cat}/K_m for an enzymatic reaction with two different substrates, because k_{cat}/K_m includes both the activation energy and the binding energy. By using different substrates and comparing the k_{cat}/K_m values, it is possibly to calculate the difference in both activation and binding energy associated with reaction with different substrates:

$$(VI) \qquad \Delta\Delta G = -R \cdot T \cdot \frac{(k_{cat}/K_m)_A}{(k_{cat}/K_m)_B}$$

Extremophiles

On our planet there are many harsh habitats that are considered extreme. Examples of such habitats include: the deep sea with high pressure, salt lakes with high salt concentrations, areas close to volcanoes with high temperatures, dry desert regions and polar regions with low temperatures. Organisms have populated all these different habitats, and in order to survive in the harsh environments the organisms have adapted to different extreme conditions. In some cases they have adapted to more than just one of the extreme conditions. Like for example organisms living in the deep sea, they have adapted to high pressure, low temperatures and salt. Extremophiles can be classified according to their environmental requirements for optimal growth [4], and Table 1 lists the most common subclasses.

Table 1
Examples of extremophiles and the conditions they are adapted to.

Name	Adapted to
Thermophiles	High temperatures
Psychrophiles	Low temperatures
Barophilies/piezophiles	High pressure
Alkaliephiles	High pH
Acidophiles	Low pH
Xerophiles	Dry conditions
Osmophiles	High sugar concentrations

2.1 Cold adaptation

A vast amount of our planet consists of cold environments, like for example the Arctic, Antarctic, mountain regions, glaciers and deep sea waters. Microorganisms and ectotherms (cold-blooded animals, whose body temperature is regulated by their behavior or surroundings) that live in cold areas have to adapt to the surrounding environment. Ectotherms and micro organisms living in cold environments have many physiological adaptations which help them to cope with low temperature. Lipids in the cell membrane of Arctic fish species are for example less saturated than those of southern fish (a

chemical exchange that is equivalent of replacing butter with olive oil), making the lipids more liquid at low temperatures. The cold also affects the metabolism, the complete set of chemical reactions that occur in a living organism to maintain its body. Enzymes are crucial to metabolism as their main task is to catalyze unfavorably chemical reactions in the cells. Enzymes from microorganisms and ectotherms living in cold regions are often referred to as cold-adapted or psychrophilic enzymes (Greek; psychro means cold, cold loving enzymes) [5]. The temperature is one of the most important factors for enzyme activity, and enzymatic reaction rates can be reduced 30-80 times when the temperature decreases from 37 to 0°C [6]. To deal with this temperature dependency, cold-adapted enzymes usually have higher catalytic efficiency at moderate and low temperatures compared to their mesophilic and thermophilic homologues [6]. The temperature dependence of chemical reactions, including enzymatic reactions can be described by the Arrhenius equation [6]:

$$(VII) \qquad k_{cat} = Zp \cdot e^{-E_a / RT}$$

where k_{cat} is the turnover number, E_a is the activation energy, R is the gas constant, T is the temperature, Z is the collision frequency and p is a steric factor. Increased k_{cat} can be achieved by lowering the activation energy for the reaction [7].

Several strategies have been postulated to explain how enzymes adapt to cold environments, but the most widely accepted hypothesis is that increased structural flexibility of components involved in the catalytic cycle in psychrophilic enzymes enhances the catalytic efficiency [5,8,9]. Psychrophilic enzymes are less heat stabile than their mesophilic homologues, and this is thought to be a result of the above mentioned increase in flexibility [5,7]. The stability/flexibility relationship is controversial because cold-adapted organisms are under no selective pressure to stabilize their proteins at elevated temperatures, and it is believed that the stability property has slowly vanished due to genetic drift [10]. Although the flexibility hypothesis has been the dominating theory to explain the increased catalytic efficiency of cold-adapted enzymes, other adaptational strategies should not be ruled out. For example, cold-adapted enzymes can increase their catalytic activity by optimizing the electrostatic properties at and around the active site [5].

2.2 Structural adaptation to low temperatures

Considerable effort has been directed towards defining structural features important for thermal adaptation. Numerous protein homologues from cold and warm environments have been analyzed in order to pinpoint structural and sequential differences to explain thermal adaptation. Unfortunately, differences that are critical for thermal adaptation are often hidden among other differences produced by genetic drift and other adaptational effects, which make it difficult to identify features specific for thermal adaptation [7]. The overall fold and the active site of cold- and warm-active protein homologues are very similar, indicating that the catalytic mechanism and reaction pathway are the same [8]. There are, however, also differences between cold- and warm-active enzymes, and progress has been made, both to explain adaptation to cold and to warm environments.

Weakening of intramolecular non-bonded interactions are often referred to in order to explain the higher molecular flexibility and lower thermal stability of psychrophilic proteins compared to their mesophilic homologues. Several comparative studies have reported that the number of salt-bridges are lower in cold-adapted enzymes [7,11,12]. Psychrophilic enzymes have also shown to lack surface salt-bridges and ion pairs between secondary structure elements and domains compared to mesophilic homologues [13,14]. In most cases salt-bridges are thought to stabilize the protein structure, thus making the psychrophilic enzymes, with fewer salt-bridges, more flexible and thermally unstable compared to mesophilic homologues. The picture is further complicated as the dielectric constant of water changes with temperature (from 55.5 at 100°C to 88.0 at 0°C [15]), consequently, the strength of ionic interactions will decrease with decreasing temperature due to increased screening.

In other cases, psychrophilic enzymes have a general lack of aromatic interactions [13] or fewer hydrogen bonds than their mesophilic counterparts [11,16]. Another trend is that psychrophilic enzymes have smaller and less hydrophobic residues in the core of the protein compared to both mesophilic and thermophilic enzymes [7,9]. This will lead to cavities in the core of the protein, and such structural differences will probably make the psychrophilic enzymes more unstable and more flexible.

The molecular surface of enzymes seems to be important for cold adaptations, and psychrophilic enzymes tend to have a higher proportion of hydrophobic residues at the surface [7]. Hydrophobic residues at the surface will destabilize the protein structure due to a decrease in entropy of the water molecules surrounding the protein [7]. For some cold-adapted enzymes an increase in charged residues at the protein surface has been observed, especially for negatively charged residues [17,18]. For example, charge-charge repulsions is thought to increase flexibility of the linker region of the psychrophilic cellulase [19].

Cold-adapted enzymes have been shown to possess an increased number of glycine residues and a reduced number proline residues in their sequences, particularly in loop regions when compared to their warm-active homologues [2,5]. Glycine, which lacks a side chain, can make the main chain more flexible. The side chain of Pro, on the other hand, forms a five member ring with the Cα atom and the main chain nitrogen, leading to a more rigid main chain. Enzymes form cold environments have also in some cases low relative Arg content [Arg/(Arg+Lys)] compared to the mesophilic counterparts [2,5,7]. The Arg side chain has a much higher hydrogen bond potential compared to Lys and these hydrogen bonds are able to stabilize the protein structure. In addition, it has been reported that psychrophilic enzymes have larger accessibility to the catalytic cavity compared to its warm-active homologues [20,21], giving rise to higher specific activity at low temperatures.

There are also other structural features which are thought to be important for cold adaptation, but the above section covers the most frequently discussed features in the literature. It is important to note that not all cold-adapted enzymes possess all the features listed above, but seem to usually use a few of them in order to achieve efficient catalysis at low temperatures. In an extensive study of structural differences between thermophilic, mesophilic and psychrophilic enzymes it was found that different protein families use different strategies to adapt to low temperatures [11].

Uracil DNA glycosylase (UDG) as a model system

In this study UDG has been used as a model system to study cold adaptation and protein-DNA recognition. The biological function of UDG is well known. In contrast to other nucleotide excision repair systems, UDG does not require any additional co-factors for activity, which makes it an excellent candidate for studies of cold adaptation and protein-DNA recognition.

3.1 DNA repair

DNA can be damaged by a variety of agents and processes, such as spontaneous deamination of bases, radiation, oxidative stress, alkylating agents and replication errors [22]. Faithful maintenance of the genome is crucial to the individual and to species [23]. In humans, DNA damage can be repaired by four major repair pathways, and several proteins are involved in each of the pathways [24]. If the DNA repair system is defect and the DNA is not repaired, severe diseases such as cancer can occur. Uracil, which does not normally occur in DNA, can appear in DNA either if deoxyuridine triphosphate is misincorporated in DNA instead of thymine or as a result of deamination of cytosine [23]. There are at least 11 different mammalian DNA glycosylases which initiate the base excision repair (BER) and removes damaged or inappropriate bases [24]. DNA glycosylases cleave the N-glycosylic bond between the target base and deoxyribose, and releases a free base and leaves an apurinic/apyrimidine site [22].

3.2 Uracil DNA glycosylase

Uracil DNA glycosylase is a DNA-repair enzyme in the base excision repair (BER) pathway and removes uracil from both single and double stranded DNA. All free living organisms express uracil DNA glycosylase, indicating that this is a highly important enzyme [25]. Six members in the uracil DNA glycosylase family are known: Uracil DNA N-glycosylase (Family-1), mismatch-specific DNA glycosylase, single-stranded selective monofunctional uracil DNA glycosylase, thermostable uracil DNA glycosylase, uracil

DNA glycosylase B and MIG protein/endonuclease III/*Methanococcus jannaschii* uracil DNA glycosylase family [26-28]. The different families of uracil DNA glycosylase have very limited sequence similarity and differ also in the makeup of the active site [28,29]. Surprisingly, there are also a lack in conservation of catalytic residues among the different families of uracil DNA glycosylase [27].

Human UDG (hUDG) and cod UDG (cUDG), both from Family-1, have been used as the model system to study cold adaptation. Throughout this thesis the abbreviation UDG will be used for enzymes from the Family-1 uracil DNA glycosylase superfamily. In humans this enzyme occurs both in nuclei and mitochondria [23], and the two enzymes have different sequence in the N-terminal. UDG from the nuclei consists of 313 amino acids while the mitochondria UDG only consists of 304 amino acids, but the catalytic domain is identical for the two variants of the enzyme [30]. Even if the N-terminal part of the enzyme is removed and only the catalytic domain is left, the enzyme is still fully active [30]. Some results also indicate that UDG is able to interact with other proteins [31-33], this interaction seems to be through the presequence (the N-terminal end) of UDG, which is not necessary for catalytic activity [32]. UDG is highly specific for uracil and show negligible activity towards the natural DNA bases or uracil in RNA (Fig. 1). UDG binds the uracil base in a specificity pocket, and the residues forming this pocket is highly conserved in this family [27].

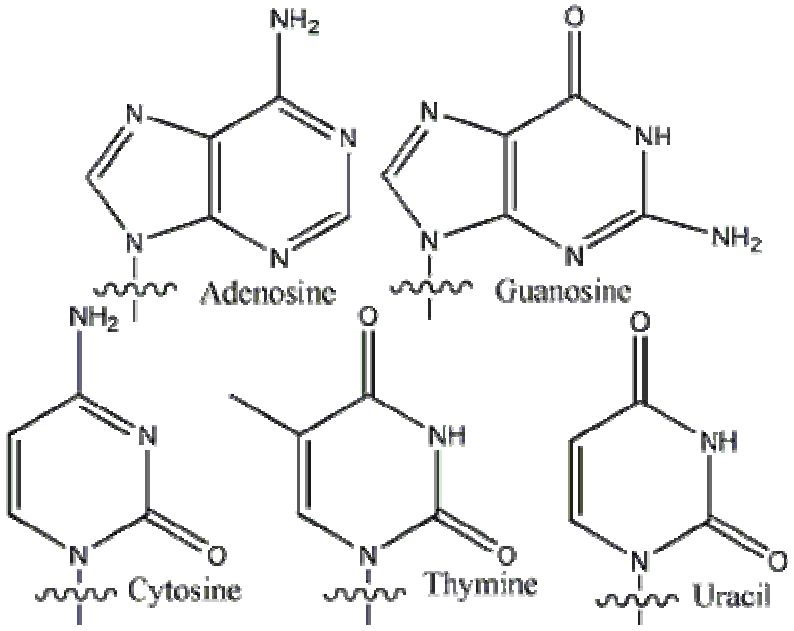


Figure 1
The four natural occurring bases in the DNA and the uracil base.

3.2.1 Recognition of DNA and catalytic mechanism of UDG

Perhaps the least understood stage in the processing of uracil bases in DNA is how UDG recognize the damaged DNA sites within vast stretches of DNA [34]. UDG can search for damaged bases either in a distributive or a processive mechanism. In the distributive mechanism the UDG dissociates from the DNA after removing the uracil base. While in a processive mechanism the UDG locates sequential uracil prior to dissociation [35]. In vitro experiments have indicated that UDG slides along the DNA and scans the strand for uracil residues in a processive manner [35,36]. This mechanism is highly affected by the salt concentration, and already at around 50 mM salt the search is shifted to a distributive search mechanism [36]. There are two different views of how UDG recognize the uracil in the DNA strand. The *base sampling model* suggests that UDG localizes uracil by breaking base pairs in the double stranded DNA and flip them out to test them against the interactions offered in the specificity pocket [27]. Another view is the *inherent extrahelicity model*, suggesting that the base pairs involving uracil are inherently weak and that the uracil will spontaneous flip-out to an extrahelical conformation,

complementary to the binding interactions offered by UDG [27,37]. The UDG then trap the extrahelical uracil base.

UDG catalyses the removal of uracil from both double and single stranded DNA by cleaving the bond between the uracil base and the sugar ring (glycosylic bond) [22]. Studies of the catalytic mechanism of human and *Escherichia coli* (*E. coli*) UDG have shown that UDG removes uracil in a stepwise reaction mechanism (Fig. 2) [38,39]. The first step is the cleavage of the glycosylic bond, where His268 and Asp145 stabilize the first transition state. In the next step, a water molecule bound to His148 and Pro146 attacks the carbon bound to uracil base and one of the hydrogens on the water molecule is transferred to the Asp145 (Fig. 2) [38]. Four negatively charged phosphate groups on the DNA repel the anionic leaving group and stabilize the positive charge on the sugar ring. These four phosphate groups stabilize the rate-determining transition state with more than 20 kcal/mol [38]

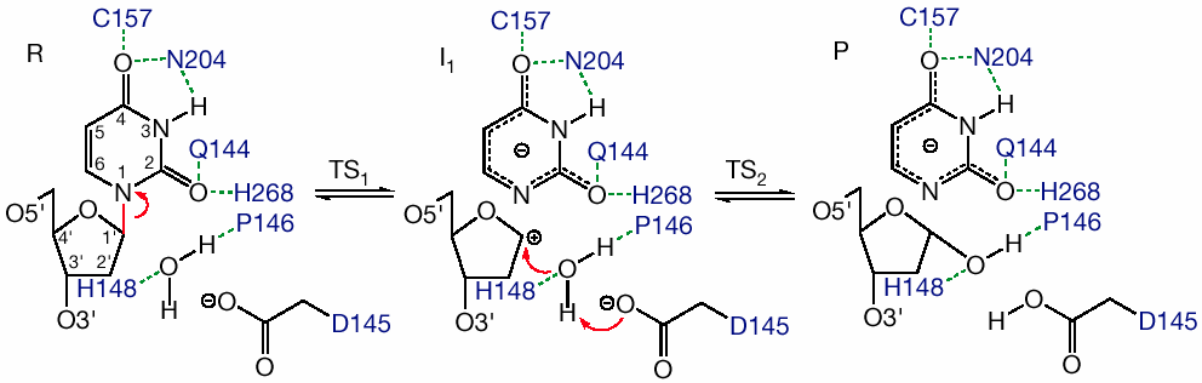


Figure 2
Catalytic mechanism for UDG. Figure adopted from Dinner *et al.* [38].

3.2.2 Comparison of mesophilic hUDG and psychrophilic cUDG

In order to improve our understanding of cold adaptation, cUDG and hUDG are used as a comparative model system. High sequence and structural similarity makes UDG a very good choice as a model system to explore adaptational features. The high amount of characterized mutants and crystal structures available for both cUDG and hUDG also make it an excellent system for theoretical studies. Even though the two enzymes have very similar three-dimensional structure, the cold-adapted enzyme is up to ten times more catalytically active in the temperature range from 288-310 K and is less heat stable compared to the warm-active representative [40]. Thus, cUDG has the typical features of cold adaptation. cUDG and hUDG have a sequence identity of 75%, an overall displacement of main chain atoms of 0.63 Å and similar secondary and tertiary structure [41]. The catalytic domain of both enzymes consists of 223 amino acids and the active site is situated in the C-terminal region of the enzyme. The tertiary structure of both enzymes consists of a β-sheet with 4 parallel strands and 11 helices (Fig. 3A). The side of the enzyme facing the DNA upon binding has a positive electrostatic potential for both enzymes (Fig. 3B). The loops: 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}) are in close contact with the DNA upon binding and are thought to be important for detection and catalysis [42]. The sequences in the loops are from hUDG, and cUDG has two mutations in the Leu272 loop. Residue 274 and 275 are Ala and His in cUDG, respectively.

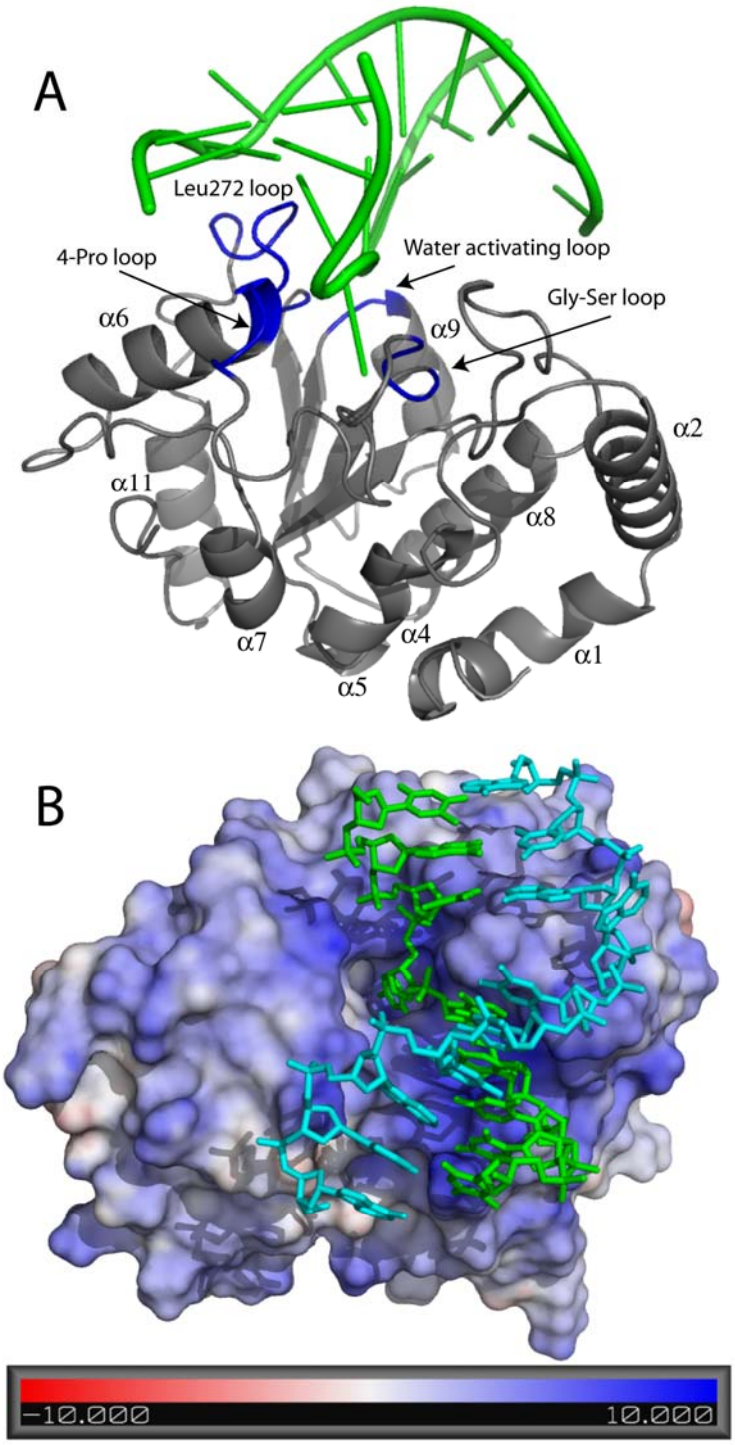


Figure 3
Two different representations of cUDG in complex with DNA. **A:** Ribbon style representation of cUDG bound to DNA. The four loops important for recognition and catalysis are shown in blue. The helices are also numbered. **B:** The electrostatic potential is projected to the surface of cUDG, and colored in kT/q. The electrostatic potential was calculated in DelPhi. The figure was generated using PyMol [43].

3.2.3 Structural adaptations to cold in psychrophilic cUDG

Psychrophilic cUDG has been thoroughly studied to find structural features which could explain the increase in catalytic efficiency and reduced thermal stability compared to its mesophilic homologue. A comparative study of the crystal structure of mesophilic hUDG and cold-adapted cUDG show that the mesophilic enzyme has 11 salt-bridges while the psychrophilic variant of the enzyme has only 5 salt-bridges [41]. The cold-adapted homologue has also three more hydrogen bonds compared to its mesophilic cousin [41]. There is a slight decrease in the size of the hydrophobic core residues in cUDG compared to hUDG, which leads to an increase in cavities in the core [40,41]. Some of the additional cavities are situated close to the active site, which could give rise to increased flexibility of the active site. A more flexible DNA recognition loop (Leu272 loop) has also been proposed to increase the catalytic efficiency and reduce the thermal stability of cUDG compared to hUDG [41].

Theoretical methods

Today computational methods play a central role in many applications in the study of biophysical properties of macromolecules. Since only a limited number of properties of biomolecular systems are actually accessible to measurement by experimental techniques, computer simulations can complement experiments by providing time series, distributions or average values of any definable quantity [44]. For example, the study of folding/unfolding pathways, conformational distributions and interactions between parts of systems are properties that computational methods are well suited for.

4.1 Molecular mechanics

Molecular mechanics (MM) refer to computational approaches that adopt classical mechanics to analyze the structure and energetic of molecular systems [45]. MM ignores the electronic motions and the energy of the system is calculated as a function of the nuclear positions only. The atoms are treated as a set of soft spheres with point charges. These energy functions along with the parameters make up a force field. Today several force fields are available for the study of biological systems, for example: AMBER [46], GROMOS [47], CHARMM [48] and OPLS-AA [49]. Equation VIII shows a typical form of a potential energy function (see also Fig. 4):

$$(VIII) \qquad V\left(r^N\right) = \sum_{bonds} \tfrac{1}{2} k_b \left(b_i - b_{i,0}\right)^2 + \sum_{angles} \tfrac{1}{2} k_\theta \left(\theta_i - \theta_{i,0}\right)^2 + \sum_{torsions} \tfrac{1}{2} V_n \left(1 + \cos(n\phi - \delta)\right)$$
$$+ \sum_{\substack{inproper \\ torsions}} \tfrac{1}{2} k_\zeta \left(\zeta_i - \zeta_{i,0}\right)^2 + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)$$

$V(\boldsymbol{r}^N)$ is the potential energy as a function of the positions ($\boldsymbol{r}$) of all N atoms. In equation VIII bond lengths, angles and improper torsions are evaluated in terms of deviations from their equilibrium values ($b_{i,0}$, $\theta_{i,0}$ *and* $\zeta_{i,0}$), while torsions are evaluated from the minimal value ($cos(n\phi\text{-}\delta)$). The last term in equation VIII is the non-bonded potential, which describes the energy between atoms separated by more than three bonds or atoms in separate molecules. The van der Waals contribution is typically evaluated with the 6-12

Lennard-Jones potential, while the electrostatic energy is obtained by applying Coulombs law. However, more sophisticated force fields contain additional terms than those mentioned above.

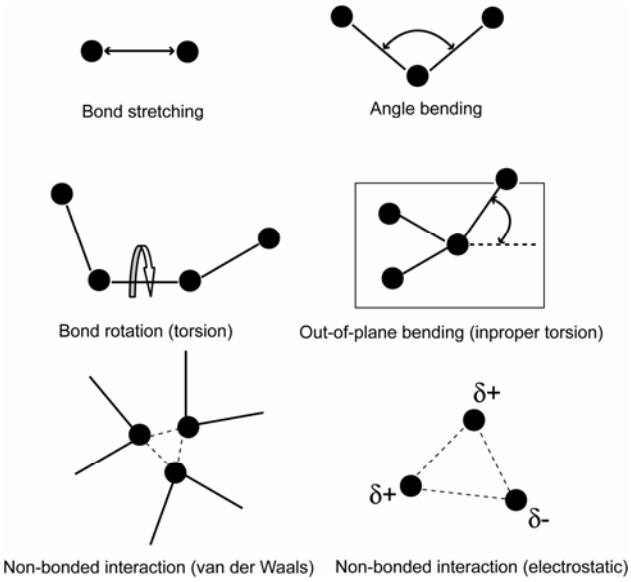


Figure 4
Schematic representations of the main contributors to a typical potential energy function in force fields.

4.2 Molecular dynamics

Molecular dynamics (MD) simulations generate a trajectory or a set of conformations that describes how the system varies over time [45] by integrating Newton`s laws of motions. The resulting trajectory contains information about positions and velocities of all atoms and how they change over time. Prior to a MD simulation the initial velocities and positions of all atoms are needed. The velocities are normally obtained from a random Maxwell distribution, while the initial positions are often taken from X-ray or NMR structures, but could also be computed from a homology model at atomic resolution. In a force field the motions of all particles are coupled, giving rise to a many-body problem that cannot be solved analytically. But by applying the finite difference method, the problem can be solved by breaking the integration into many small steps, all separated in time by a fixed time step (δt). The potential energy function with respect to the atomic

positions give the force acting on the atoms and this force is used to calculate the accelerations of all particles. The accelerations of the particles combined with positions and velocities are used to calculate the new positions and velocities at time (t + δt). The force is assumed to be constant during the time step. In order to keep the temperature stable during the MD simulation, the system is normally coupled to an external bath [50].

In a MD simulation a combination of 3N positions and 3N momenta defines the points in the 6N-dimential phase space, where N is the number of atoms [45]. If the sampling of the phase space is sufficient, the trajectory from the MD simulation can give good estimates of thermodynamic quantities and dynamic properties. MD simulations are often used to sample the conformational space prior to calculations of binding energies [51-54], pK_a values [55,56] and strengths of salt-bridges [57]. In addition, MD simulations are an important tool in the study of the dynamic nature of biological macromolecules and folding/unfolding of proteins [58-60]. They are also crucial in the structure refinement of macromolecules using NMR or X-ray crystallography.

4.3 Free energy calculation methods

The ability to calculate the strength of non-covalent interaction has been an important objective in computational chemistry. There are several methods that are currently used to predict the strength of binding energies. The methods are ranging from computationally expensive methods like free energy perturbation (FEP) and thermodynamic integration (TI) (see e.g. Brandsdal *et al.* [61] for a review of the methods) to various empirical or knowledge-based scoring approaches [62-64] where sampling of the conformational space is neglected. The linear interaction energy (LIE) method [65] and the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method [66,67] are commonly used for calculation of association energies. These approaches are based on analysis of the molecular dynamics or Monte Carlo trajectories.

4.4 MM-PBSA method

The MM-PBSA method [66-68] was initially used to study the stability of various DNA and RNA fragments, but the method has in later years also been applied in the calculation of binding free energies of proteins and small ligand [53,54], protein-protein [69-71] and protein-DNA complexes [72]. This method estimates the free energy of every single conformation according to:

$$(IX) \qquad G = H_{MM} + G_{sol} - TS_{MM}$$

where H_{MM} is the molecular mechanical energy and can be divided into several energy terms:

$$(X) \qquad \mathrm{H}_{MM} = E_{bond} + E_{angle} + E_{torsion} + E_{elec} + E_{vdW}$$

where E is the bond, angle, torsion, electrostatic and the van der Waals term in the molecular mechanical force field. The term $-TS_{MM}$ is the solute entropy and is usually calculated with normal mode analysis. The G_{sol} is the solvation free energy and can be divided into two terms:

$$(XI) \qquad G_{\mathrm{sol}} = G_{pol} + G_{np}$$

where G_{pol} is the electrostatic contributions to the solvation free energy and can be obtained by either solving the Poisson-Boltzmann (PB) equation [73] or by the generalized-Born (GB) method [74,75]. The non-polar solvation free energy, G_{np}, is determined with a solvent-accessible-surface-area (SASA) dependent term [76]. The binding free energy is finally calculated according to:

$$(XII) \qquad \Delta\mathrm{G} = \left\langle \Delta\mathrm{G} \right\rangle_{\mathrm{complex}} - \left\langle \Delta\mathrm{G} \right\rangle_{\mathrm{receptor}} - \left\langle \Delta\mathrm{G} \right\rangle_{\mathrm{ligand}}$$

where $\left\langle \Delta\mathrm{G} \right\rangle_{\mathrm{complex}}$, $\left\langle \Delta\mathrm{G} \right\rangle_{\mathrm{receptor}}$ and $\left\langle \Delta\mathrm{G} \right\rangle_{\mathrm{ligand}}$ are the free energies of the complex, the protein and the ligand, respectively, averaged over a set of snapshots extracted from a MD simulation. The energy contributions to the binding energy can either be extracted from a single trajectory simulation of the complex or it is possible to use separate trajectories for the complex, receptor and the ligand. A drawback with the separate trajectory method is that it is very difficult to get the H_{MM} energy to converge for large molecules within reasonable computing time [61]. The single trajectory method assumes

that there are no changes in the structure of the receptor and ligand upon binding, which is not always the case. Studies using linear interaction energy (LIE) calculations have shown that the ligand, if not very small and rigid, often adopt different conformations when free in solution compared to the conformations in the complex [61]. But in contrast to the LIE method, the MM-PBSA method does not require optimization of any parameters for the calculation of the total energy. In addition, the MM-PBSA method estimates the entropy of the solute.

4.5 Continuum electrostatics

Explicit solvation energies of large macromolecules are very time consuming to calculate with computational methods. An alternative approach is to apply continuum electrostatics, which describe the solvent properties as average values [73]. The advantages of continuum electrostatic models are their low computational cost and the possibility of visualization of the electrostatic potential. In these methods the solvent is represented as a high dielectric medium, while the solute is described as a low dielectric medium, and the molecular surface area is often used as the boundary (Fig. 5). The classical treatment of electrostatics in solution is based on the Poisson-Boltzmann (PB) equation:

$$(XIII) \qquad \nabla\cdot\left[\varepsilon(r)\nabla\cdot\phi(r)\right]-\varepsilon(r)\kappa(r)^2\sinh\left[\phi(r)\right]+\frac{4\pi\rho^f(r)}{kT}=0$$

where ε is the dielectric constant, ϕ(r) is the electrostatic potential in units of kT/q, k is the Boltzmann constant, T is the temperature, q is the charge of a proton, r is the position vector and ρ^f is the fixed charge density (in proton charge units). The term $\kappa^2 = 1/\lambda^2 = 8\pi q^2 I/ekT$, where λ is the Debye length and I is the ionic strength of the bulk solution. The salt effect is described by the second term in equation XIII, and if there are no mobile ions in the system, this term is absent. With no mobile ions present and with a uniform dielectric constant for the entire system, the PB equation is reduced to the Coulomb's law.

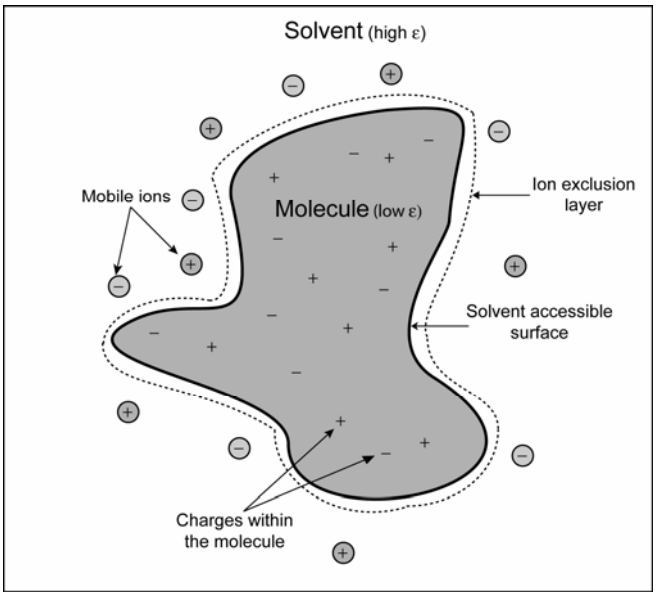


Figure 5
A molecule in a heterogeneous dielectric medium.

The PB equation is solved by numerical methods like finite difference methods, and today several programs have implemented an algorithm to solve the PB equation. Examples of programs that are widely used for this purpose are DelPhi [77] and UHBD [78]. Even though the continuum approaches have provided a wealth of information in different fields, they also have some limitations. Continuum models do not tend to work well when short range effects become important, for example when individual waters molecules bridging functional groups [79]. Such models are also very dependent on the dielectric constant, and it can be very challenging to choose the optimal dielectric constant of the solute [79,80]. The protein dielectric constant (ε_p) is not an universal constant, but is a parameter which depend on the model used [81]. ε_p represents the electrostatic interactions which are not represented explicitly in the model. Such factors can be the effect of water penetration into the protein core or fluctuations of polar groups in the protein [80-82]. Because proteins are structurally heterogeneous and possess flexible areas, the ε_p can also have different optimal values in different regions of the protein [80]. All the above mentioned factors make it difficult to choose the optimal protein dielectric constant. Early calculations used a low dielectric constant for the protein, typically in the range 1-4, while more recent studies indicate that a larger constant is needed to reproduce experimental shifts in pK_a values [56,81].

Aims of study

The main focus of this thesis is to explore the structural features responsible for cold adaptation using uracil DNA glycosylase (UDG) as a model system. Since UDG is a DNA binding protein, general aspects of protein-DNA recognition will also be studied. These objectives will mainly be studied with computational methods.

Subgoal 1:

Is improved flexibility responsible for increased catalytic efficiency of cold-adapted UDG?

In general, cold-adapted enzymes are believed to possess higher flexibility in structural components involved in the catalytic cycle compared to their warm-active homologues. In UDG, increased flexibility in the DNA recognition loop is thought to explain the high catalytic efficiency of cold-adapted cod UDG (cUDG). Molecular flexibility is very difficult to study by experimental methods, thus MD simulations will be used to study differences in flexibility between the cold-adapted cUDG and the warm-active human UDG (hUDG).

Subgoal 2:

Which forces are important for thermal stability of UDG?

The general view is that cold-adapted enzymes are less stable than their warm-active homologues. Experimental results have also shown that psychrophilic cUDG has lower kinetically derived stability than mesophilic hUDG. Because of the reduced stability of cUDG, our hypothesis is that the psychrophilic enzyme will unfold at lower temperatures or have higher unfolding rate than the mesophilic homologue. High temperature MD simulations will be used to analyze the unfolding pathway and find important molecular contacts for stability of cUDG and hUDG.

Comparative studies of cold- and warm-active enzymes have shown that warm-active enzymes often have increased number of salt-bridges compared to cold-adapted homologues. Analysis of the crystal structure of cUDG and hUDG showed that the

mesophilic enzyme has increased number of salt-bridges. Salt-bridges are generally thought to have favorable effect on the protein stability. Continuum electrostatics will be used to calculate the strength of ion pairs in both cold- and warm-active UDG to investigate if the warm-active hUDG are more stabilized by salt-bridges than the cold-adapted homologue.

Subgoal 3:

Is the binding to DNA different for cold- and warm-active UDG?

Kinetic studies have shown that cold-adapted cUDG has lower K_m values compared to the mesophilic counterpart. If the enzyme follows the simple Michaelis-Menten mechanism, K_m will be lowered by a more favorable substrate binding. So far, the binding energy between cold-adapted UDG and DNA has not been studied. In this study MD simulations and free energy calculations methods will be used to calculate the strength of the binding for hUDG and cUDG in complex with double stranded DNA.

Results and discussion

UDG is an excellent model system for the study of enzymatic adaptation to low temperature. Its biological function is well-known, and several crystal structures of both cUDG and hUDG are available, including mutant and native structures, which have been characterized in terms of stability and kinetics [40,41,83-85]. Even though cUDG and hUDG have highly similar secondary structure and 75% sequence identity, cUDG has been shown to be up to 10 times more catalytically efficient (k_{cat}/K_m) in the temperature range from 288-310 K [40]. cUDG has also been found to be much more pH and temperature labile than hUDG [83], indicating that cUDG is generally less stable compared to hUDG. In this study the psychrophilic cUDG and the mesophilic hUDG enzymes have been explored with various methods in order to describe how UDG adapt to cold environments at a molecular level.

Life in the cold has to cope with reduction in chemical reaction rates and lower fluid viscosity. In order to maintain sufficient metabolic fluxes at low temperatures, the psychrophilic organisms produce "cold-adapted" enzymes, which are able to maintain high catalytic efficiency even at low temperature [86]. The relationship between activity-flexibility-stability is probably the leading hypothesis to explain enzyme properties in temperature adaptation [2,5,7]. Increased molecular flexibility will lead to higher activity and less stable enzymes, and on the other hand, a very stable enzyme will be too rigid to perform its catalytic function at high rate, and yielding a less efficient enzyme. Flexibility can be considered either as a static or a dynamic property [87,88]. Static flexibility refers to the number and structural diversity of the different conformers in the equilibrium ensemble. Dynamic flexibility is how quickly the structure can interconvert between the conformers and is a measure of the energy barriers between the conformers in the equilibrium ensemble [87,88]. Static flexibility can be gained from B-factors from crystal structures, from hydrogen/deuterium (H/D) exchange experiments or from conformers from MD simulations. Dynamic flexibility can be measured by dynamic fluorescence quenching or proteolytic nicking, for more information on these methods see e.g. *Siddiqui and Cavicchioli* [7].

5.1 Molecular flexibility

Enzymes are not static structures, but possess a certain dynamic property. Flexibility of regions which are directly involved in the enzyme catalysis is believed to be essential for enzyme activity [89]. Thermophilic enzymes have decreased catalytic efficiency at low temperatures, which is thought to be caused by a too rigid structure at low temperatures [90-92]. The dynamic properties of protein structures decrease with decreasing temperature and cold-adapted enzymes are thought to have improved flexibility in structural components involved in the catalytic cycle in order to maintain high catalytic activity at low temperatures [2]. The increased flexibility would increase the specific activity (k_{cat}), as the increased flexibility will enable good complementary with the substrate [2]. Psychrophilic cUDG has higher specific activity compared to mesophilic hUDG (paper I). Previous studies have postulated that the increased catalytic efficiency of cUDG could be explained by increased flexibility of the DNA recognition loop also called the Leu272 loop [41]. MD simulations has been used with success to reveal flexibility in engineered mesophilic subtilisin [93]. The engineered subtilisin showed typical cold-adapted features and also increased local and global flexibility.

MD simulations on cold- and warm-active UDG and on mutants of the enzymes were performed to get insight into the flexibility of the two homologues enzymes (paper II). The simulations were run for both the UDG-DNA complex and for the uncomplexed enzyme solvated by water. The MD simulations show that the cold-adapted cUDG has higher overall flexibility per residue compared to the mesophilic homologue. This indicates that the structure of the mesophilic enzyme is more rigid than the psychrophilic counterpart. The plot of root-mean-squared fluctuations (r.m.s.f.) per residue shows that especially in the DNA recognition loop, there are large differences in flexibility between the two enzymes (paper II). The psychrophilic enzyme has much higher flexibility in this loop compared to the mesophilic counterpart. The finding fits well with the emerged picture that cold-adapted enzymes have higher flexibility in the active site cleft or in loop structures around the active site in order to maintain sufficient activity at low temperatures [7,93].

The results from the MD simulations of all six different UDG variants show that there seem to be a correlation between catalytic efficiency and flexibility of this loop (paper II). This is especially evident for the cUDG-V171E mutant, which has the lowest flexibility in the DNA recognition loop and lowest catalytic efficiency of the variants in this study. It is interesting to note that the mutation of the 171 residue does not have any effect on the flexibility on nearby residues in the sequence even though it has large effect on the flexibility of the DNA recognition loop. Introducing the negatively charged Glu171 reduces the positive electrostatic potential in the DNA recognition loop area. The Leu272 loop consists of neutral residues and one positively charged Arg residue, and for these residues it would probably be destabilizing to be in a positive electrostatic potential. Reducing the positive electrostatic potential near this loop will maybe stabilize the loop and make it less flexible. Even if the largest differences in flexibility between cUDG and hUDG are seen in the Leu272 loop, there is also increased flexibility of the Gly-Ser loop (residue 246-247) in the psychrophilic enzyme. The Ser in the Gly-Ser loop forms a hydrogen bond to the DNA in the enzyme-DNA complex and this loop is probably involved in orienting the enzyme prior to the DNA scan [42].

5.1.1 Structural features responsible for increased Leu272 loop flexibility in cUDG

If we compare the amino acid sequence of the Leu272 loop of cUDG and hUDG, Val274 and Tyr275 in hUDG are mutated to Ala274 and His275 in cUDG. In addition, Phe279 in hUDG is mutated to Leu in the cUDG structure. Residues Val274, Tyr275 and Phe279 in the mesophilic enzyme are believed to form a hydrophobic cluster which could restrict the motion of the Leu272 loop [41]. In the psychrophilic enzyme the three residues are mutated to smaller amino acids that are not able to form a hydrophobic cluster. The cUDG-H275Y mutant has much lower flexibility in the Leu272 loop compared to cUDG, but similar flexibility to the human variant of the enzyme (paper II). This indicates that His275 in cUDG is the main contributor to the increased flexibility of the Leu272 loop, relative to the human enzyme.

5.2 Thermal stability

Proteins are usually only marginally stable at their physiologically relevant temperature, and their free energy of unfolding typically varies between 5 to 15 kcal/mol [1]. Large opposing energies are involved in the stabilization of proteins, and all contributions, both favorable and unfavorable, are therefore important when considering protein stability. How proteins achieve their stability do not seem to follow any general rules [94], but the emerging picture is, for globular proteins, that the hydrophobic effect and burial of nonpolar side chains stabilizes the native state [95,96]. Disulfide bonds, electrostatic interactions and hydrogen bonds are, however, also important for structural stability and contribute favorably to protein stability [97-99].

Kinetically derived stability of cUDG and hUDG has been measured as half-life times (the time it takes for the enzyme to lose 50% of its activity). The psychrophilic enzyme has shorter half-life for all temperatures examined [40,83], and has also lower temperature optimum (the temperature with highest relative activity) than its mesophilic counterpart. The temperature optimum for cUDG and hUDG is 314 K and 318 K, respectively [83,100]. These results show that the psychrophilic variant loses activity at lower temperature than its mesophilic homologue. Even though this is not a direct measure of thermodynamic stability, it indicates that mesophilic hUDG might be more thermostable than the psychrophilic variant of the enzyme.

In order to investigate the stability of UDG, high temperature MD simulations of cUDG and hUDG were performed, to investigate the unfolding process of these enzymes (paper III). The MD simulations were carried out at three different temperatures: 375 K, 400 K and 425 K. The root-mean-squared deviations (r.m.s.d.) were used to measure unfolding during the MD simulations. For the two lowest temperatures, the psychrophilic enzyme unfolds more rapidly than the mesophilic variant, indicating reduced structural stability. If the time taken to reach r.m.s.d. of 15 Å, is used as a measure of unfolding rate, there are large differences in the unfolding rate at 400 K. The r.m.s.d. of the psychrophilic enzyme reaches a value of 15 Å after ~0.5 ns at 400 K while the mesophilic variant needs 4.2 ns to reach the same degree of unfolding at the same temperature (paper III). At 425

K the psychrophilic enzyme has the same high unfolding rate as observed for the 400 K simulations, while at this temperature hUDG has also reached high unfolding rate, similar to the rate of cUDG. This indicates that the cold-adapted cUDG needs lower temperature to reach a fast unfolding rate compared to the warm-active hUDG. However, once the thermal energy of the systems is sufficiently high (i.e. 425 K in this case), rapid unfolding is observed for both enzymes.

Salt-bridges are also expected to be important for stability of proteins [99], and it is believed that the warm-active enzymes are more stabilized by salt-bridges than the cold-adapted homologues [7]. The crystal structure of the psychrophilic cUDG (pdb code 1OKB) has a reduced number of salt-bridges compared to the mesophilic hUDG (pdb code 1AKZ) [41]. In our study we have analyzed 5000 structures from MD simulations to identify putative salt-bridges in cUDG and hUDG (paper IV). The results from this study show that the two enzymes have similar amount of salt-bridges, cUDG and hUDG have 11 and 12 salt-bridges, respectively. Continuum electrostatics was used to calculate the electrostatic contributions to the stability of each salt-bridge (paper IV). Virtually all ion-pairs present in both cUDG and hUDG have a favorable electrostatic contribution, which probably lead to increased structural stability. This is in accordance with other studies of salt-bridges in other proteins [99,101]. Without taking the physiological temperature of the organisms into considerations, we find comparable electrostatic stability of ion-pairs in cUDG and hUDG. However, if we look at their respective environmental temperatures, the ion-pairs are more stabilizing in hUDG when compared to cUDG (paper IV). When it comes to their net contribution to protein stability, entropy must also be considered. Salt-bridge formations will reduce the available conformational space and the conformational entropy of the folded state will consequently decrease. Hence, ion-pairs destabilize the native state from an entropic point of view, whereas they are enthalpically stabilizing. The entropic effect is intrinsically difficult to estimate through computer simulations, and the fact that it changes with temperature makes it even more challenging to calculate.

5.2.1 Structural features responsible for stability in UDG

Comparative analysis of the ion pairs and unfolding of cold- and warm-active UDG have shown that certain molecular contacts seem to explain the differences in stability between the two enzymes. One of the first events in the unfolding process of cUDG is the melting of the N-terminal, while in hUDG both terminals unfold as an early step in the unfolding pathway. Thus, molecular contacts which stabilize the terminals are probably very important for stability of UDG. In the N-terminal there are especially three hydrogen bonds which are thought to be important for the stability of both enzymes. When the hydrogen bonds: Ser88:Oγ-Asp133:Oδ2, Trp89:Nε1-Cys132:O and Trp89:Nε1-Thr129:O are lost, the N-terminal of both enzymes rapidly unfolds (paper III). These hydrogen bonds connect the N-terminal of helix α1 and the helix α4 and the loop between helix α4 and helix α5 (for helix numbering see Fig 3A). In addition, the side chain of Trp89 packs into a hydrophobic area between helices α4 and α8. Interestingly, the crystal structure of a thermophilic uracil DNA glycosylase from *Thermus thermophilus* HB8 (TthUDG) has a stabilizing [4Fe-4S] cluster close to the N-terminal [102]. TthUDG belongs to the family-4 uracil DNA glycosylases while hUDG and cUDG are from family-1. Even though the amino acid sequence homology is low between TthUDG and hUDG or cUDG, the topology and order of the secondary structure elements are similar between the two uracil DNA glycosylase families [102]. Temperature adaptation of UDG seems to involve additional stabilization of the N-terminal part.

The high temperature MD simulations show that there are large differences in stability in the C-terminal between the two enzymes (paper III). hUDG unfolds in the C-terminal at the two lowest temperatures, while at the highest temperature both enzymes unfold. The C-terminal of both enzymes is situated at the surface, and the side facing away from the solvent forms a hydrophobic cluster with helix α7 and α11. In addition, the Trp301 side chain forms stacking interactions with the side chains of Phe284 and Trp195. When the hydrophobic packing/stacking interactions are lost, the C-terminal starts to melt. The hydrogen bond between Gln198 and residue 300 (Asp in hUDG and Asn in cUDG) might also be responsible for stabilizing the C-terminal. But even if the interactions between Gln198 and Asn300 are lost, the C-terminal seems to resist unfolding. Structural analysis

shows that the ionic interaction between Lys138 and residue 297 is important for stability of the C-terminal. cUDG has a Glu in this position and the distance between the atoms Lys138:Nζ and Glu297:Oε1 is 5.4 Å. Thus, this ionic interaction will stabilize the C-terminal of cUDG. hUDG, on the other hand, has a Lys in position 297 and the positive charges will repel each other and destabilize the C-terminal of hUDG (paper III). To verify the importance of this ionic contact, the Glu297 in cUDG was mutated to Lys and the Lys297 in hUDG was mutated to Glu. MD simulations were performed on these two mutants as well. The cUDG-E297K did unfold in the C-terminal, and this shows that the Lys138-Glu297 ionic interaction is important for C-terminal stability. The C-terminal of the hUDG-K297E mutant did also unfold, indicating that there are also other molecular features than the former mentioned ionic contact that is important for stabilizing the C-terminal. From residue 293 to the C-terminal end, there are four residues in the sequence that make cUDG more hydrophobic than hUDG. The substitutions are K293L, K296T, D300N and E303A (mutation from hUDG to cUDG), all charged residues in hUDG, substituted to hydrophobic or uncharged in cUDG (for a full alignment see Leiros *et al.* [41]). It will be more unfavorable for cUDG to unfold in the C-terminal and expose its more hydrophobic residues to water compared to hUDG which has more hydrophilic residues close to the C-terminal. This could also explain why cUDG have a more stable C-terminal compared to hUDG.

In the comparative salt-bridge study, we identified 12 and 11 salt-bridges in hUDG and cUDG, respectively (paper IV). The strength of the different salt-bridges varies for both enzymes. In the mesophilic enzyme we found 3 very strong salt-bridges at its physiological temperature (ΔG_{tot} > 2.0 kcal/mol), Asp180-Arg282, Asp183-Lys302 and Asp300-Lys302. In the psychrophilic enzyme only one strong salt-bridge was found, Asp180-Lys282. But the mesophilic enzyme has also the three weakest salt-bridges (paper IV). The Asp180-Lys/Arg282 salt-bridge is probably especially important for UDG stability. In the unfolding study of the warm-active UDG this specific salt-bridge seems to be responsible for keeping the 165-190 region folded (paper III).

The warm- and the cold-active UDG have 5 and 3 global salt-bridges, respectively. The global salt-bridges will connect different part of the structure and this probably leads to a more rigid overall structure. Local salt-bridges, on the other hand, can still be intact even if the protein is highly unfolded. Thus, global salt-bridges are most likely more important for protein stability compared to local ones. The warm-active UDG has also a strong ionic network (Asp183, Lys302 and Asp300), which is not observed in the cold-adapted UDG. These factors could probably stabilize the mesophilic UDG more than its cold-adapted homologue.

5.3 UDG-DNA binding and electrostatics

The ability of proteins to interact with other macromolecules in a highly specific manner is an important feature for a variety of biological processes, including DNA repair, antigen-antibody interactions, signal transduction, enzymatic catalysis, drug design among others. It has been proposed that cUDG has increased substrate affinity compared to hUDG due to enhanced positive electrostatic potential at surface areas central to formation of the enzyme-substrate complex [41]. Continuum electrostatics was applied to investigate the difference in electrostatics between cold- and warm-active UDG. Both cUDG and hUDG has a highly positive electrostatic potential in the specificity pocket and in nearby areas that are know to interact directly with DNA. As indicated (paper I), there are differences in the electrostatic surface potential between cUDG and hUDG, as the psychrophilic enzyme has a more positive electrostatic potential near the active site. The 171 residue seems to be a key residue for explaining difference in electrostatic potential (paper V).

The binding studies of cUDG and hUDG with the MM-PBSA method show that there are large energies involved in the binding between UDG and DNA. Even if the standard deviations are very high, especially for the separate trajectory, the energies calculated with both single and the separate trajectories indicate that the psychrophilic enzyme associates stronger to DNA than the mesophilic variant (paper V). All the four loops which are important for detection and catalysis in UDG have favorable interactions with

the DNA, but the DNA recognition loop seems to be especially important for binding, and is responsible for 34.1 % and 44.2 % of the enthalpic contribution to the binding energy in the warm- and cold-active UDG, respectively. Residues 275 and 276 in particular, have a much stronger binding energy per residue in cUDG. Residue 276 is an Arg in both enzymes, and 275 is a Tyr in hUDG and the simulations show that this residue bends away from the DNA in the complex, while cUDG has a His in this position which hydrogen bonds to the DNA.

Concluding remarks

Several structural features important for cold-adaptation have been pinpointed in this study. This study shows that electrostatic surface properties near the active site and flexibility of the DNA recognition loop seem to be important for adaptation to cold for psychrophilic cUDG. A link between high positive electrostatic potential at and around the active site and flexibility of the DNA recognition loop is proposed, but further studies are needed to fully investigate this hypothesis. The UDG-DNA binding study identified several residues which are important for binding and especially the DNA recognition loop form strong interactions with the DNA.

Summary of the thesis

In this study uracil DNA glycosylase (UDG) has been used as a model system for the study of cold adaptation and protein-DNA recognition with computational methods. Cold-adapted, or so-called psychrophilic, enzymes are attractive as targets for commercialization due to their reduced thermal stability which is usually accompanied with an increased catalytic efficiency. UDG removes uracil from the DNA strand as the first step in the base excision repair pathway (DNA repair system). A deficient DNA repair system is associated with serious diseases, such as development of cancer in humans. Uracilation of DNA represents a constant threat to the survival of many organisms, and since UDG is the most efficient of all the enzymes in the UDG superfamily and found in all free living organisms its a vital enzyme.

MD simulations, continuum electrostatics, X-ray crystallography and free energy calculations have been used to study the structural and energetic differences between warm-active human UDG and cold-active cod UDG at atomic level. Analyses of the MD simulations show that the psychrophilic UDG has a highly flexible DNA recognition loop compared to its warm-active homologue. This is thought to explain the observed high catalytic efficiency for cold-adapted UDG. This is in accordance with the emerged picture that cold-adapted enzymes have higher flexibility of components involved in the catalytic cycle. Analyses of several mutants of cod and human UDG indicate that there is a correlation between catalytic efficiency and flexibility in this DNA recognition loop.

Continuum electrostatics calculations has been applied to analyze all ionic contacts in the warm- and cold-active UDG. The electrostatic contribution of the ion-pairs is slightly more favorable in cod UDG at 298 K. This is primarily attributed to more optimized interactions between the ion-pairs and nearby dipoles/charges in cod UDG. When we take the environmental temperatures into account, the electrostatic stability becomes more favorable for the ion-pairs in the mesophilic enzyme. Comparative studies of the electrostatic potential of cod and human UDG show that both enzymes have a positive electrostatic potential near the active site, but the potential is even higher for the cold-

adapted enzyme. This difference in electrostatic potential could probably explain the different K_m values for these two enzymes.

Comparative high temperature MD simulations were used to study the unfolding and structural stability of cod and human UDG. The simulations showed that there are distinct structural differences in the unfolding pathway between the warm- and cold-active UDG, particularly evident in the N- and C-terminals. The results from these MD simulations also showed that at certain temperatures the psychrophilic enzyme has a higher unfolding rate compared to its mesophilic homologue. The MM-PBSA method was used to analyze the binding energy of the UDG-DNA complex. Cod UDG possesses a slightly more favorable DNA binding energy compared to human UDG. Decomposition of the binding energy per residue made it possible to pin-point residues that were important to the binding energy. The DNA recognition loop is responsibly for 34.1 % and 44.2 % of the enthalpic contribution to the binding energy in the warm- and cold- active UDG, respectively. Thus, this loop is very important for UDG-DNA association, and especially important for cod UDG.

References

[1] Fersht, A. Structure and mechanism in protein science. W.H. Freeman and Company, 1999.

[2] Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, J. C., Sonan, G., Feller, G., Gerday, C. Some like it cold: Biocatalysis at low temperatures. FEMS Microbiol. Rev. 2004, 28, 25-42.

[3] Avis, J. M., Fersht, A. R. Use of binding-energy in catalysis - optimization of rate in a multistep reaction. Biochemistry. 1993, 32, 5321-5326.

[4] van den Burg, B. Extremophiles as a source for novel enzymes. Curr Opin Microbiol. 2003, 6, 213-218.

[5] Smalås, A. O., Leiros, H. K. S., Os, V., Willassen, N. P., 2000. Cold adapted enzymes. In. Biotechnol annu rev, 6, 1-57

[6] Lonhienne, T., Gerday, C., Feller, G. Psychrophilic enzymes: Revisiting the thermodynamic parameters of activation may explain local flexibility. Bba-Protein Struct M. 2000, 1543, 1-10.

[7] Siddiqui, K. S., Cavicchioli, R. Cold-adapted enzymes. Annu Rev Biochem. 2006, 75, 403-433.

[8] Feller, G. Molecular adaptations to cold in psychrophilic enzymes. Cell Mol Life Sci. 2003, 60, 648-662.

[9] Feller, G., Gerday, C. Psychrophilic enzymes: Molecular basis of cold adaptation. Cell. Mol. Life Sci. 1997, 53, 830-841.

[10] Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N., Arnold, F. H. Directed evolution study of temperature adaptation in a psychrophilic enzyme. J. Mol. Biol. 2000, 297, 1015-1026.

[11] Gianese, G., Bossa, F., Pascarella, S. Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes. Proteins. 2002, 47, 236-249.

[12] Riise, E. K., Lorentzen, M. S., Helland, R., Smalas, A. O., Leiros, H. K. S., Willassen, N. P. The first structure of a cold-active catalase from vibrio salmonicida at 1.96 angstrom reveals structural aspects of cold adaptation. Acta Crystallogr. D. 2007, 63, 135-148.

[13] Davail, S., Feller, G., Narinx, E., Gerday, C. Cold adaptation of proteins - purification, characterization, and sequence of the heat-labile subtilisin from the antarctic psychrophile bacillus ta41. J. Biol. Chem. 1994, 269, 17448-17453.

[14] Feller, G., Payan, F., Theys, F., Qian, M. X., Haser, R., Gerday, C. Stability and structural-analysis of alpha-amylase from the antarctic psychrophile alteromonas-haloplanctis-a23. Eur. J. Biochem. 1994, 222, 441-447.

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

[16] 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.

[17] Feller, G., Zekhnini, Z., LamotteBrasseur, J., Gerday, C. Enzymes from cold-adapted microorganisms - the class c beta-lactamase from the antarctic psychrophile psychrobacter immobilis a5. Eur. J. Biochem. 1997, 244, 186-191.

[18] Leiros, H. K. S., Willassen, N. P., Smalas, A. O. Structural comparison of psychrophilic and mesophilic trypsins - elucidating the molecular basis of cold-adaptation. Eur. J. Biochem. 2000, 267, 1039-1049.
[19] Violot, S., Aghajari, N., Czjzek, M., Feller, G., Sonan, G. K., Gouet, P., Gerday, C., Haser, R., Receveur-Brechot, V. Structure of a full length psychrophilic cellulase from pseudoalteromonas haloplanktis revealed by x-ray diffraction and small angle x-ray scattering. J. Mol. Biol. 2005, 348, 1211-1224.
[20] Aghajari, N., Van Petegem, F., Villeret, V., Chessa, J. P., Gerday, C., Haser, R., Van Beeumen, J. Crystal structures of a psychrophilic metalloprotease reveal new insights into catalysis by cold-adapted proteases. Proteins. 2003, 50, 636-647.
[21] 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.
[22] Krokan, H. E., Standal, R., Slupphaug, G. DNA glycosylases in the base excision repair of DNA. Biochem J. 1997, 325, 1-16.
[23] Lindahl, T., Wood, R. D. Quality control by DNA repair. Science. 1999, 286, 1897-1905.
[24] Krokan, H. E., Kavli, B., Slupphaug, G. Novel aspects of macromolecular repair and relationship to human disease. J. Mol. Med. 2004, 82, 280-297.
[25] Priet, S., Gros, N., Navarro, J. M., Boretto, J., Canard, B., Querat, G., Sire, J. Hiv-1-associated uracil DNA glycosylase activity controls dutp misincorporation in viral DNA and is essential to the hiv-1 life cycle. Mol. Cell. 2005, 17, 479-490.
[26] Moe, E., Leiros, I., Smalas, A. O., McSweeney, S. The crystal structure of mismatch-specific uracil-DNA glycosylase (mug) from deinococcus radiodurans reveals a novel catalytic residue and broad substrate specificity. J. Biol. Chem. 2006, 281, 569-577.
[27] Pearl, L. H. Structure and function in the uracil-DNA glycosylase superfamily. Mutat. Res-DNA Rep. 2000, 460, 165-181.
[28] Sartori, A. A., Fitz-Gibbon, S., Yang, H. J., Miller, J. H., Jiricny, J. A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. EMBO J. 2002, 21, 3182-3191.
[29] Aravind, L., Koonin, E. V. The alpha/beta fold uracil DNA glycosylases: A common origin with diverse fates. Genome Biol. 2000, 1, RESEARCH0007.
[30] Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T. A., Skorpen, F., Krokan, H. E. Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the ung gene. Nucleic Acids Res. 1997, 25, 750-755.
[31] Handa, P., Acharya, N., Varshney, U. Chimeras between single-stranded DNA-binding proteins from escherichia coli and mycobacterium tuberculosis reveal that their c-terminal domains interact with uracil DNA glycosylases. J. Biol. Chem. 2001, 276, 16992-16997.
[32] Nagelhus, T. A., Haug, T., Singh, K. K., Keshav, K. F., Skorpen, F., Otterlei, M., Bharati, S., Lindmo, T., Benichou, S., Benarous, R., Krokan, H. E. A sequence in the n-terminal region of human uracil-DNA glycosylase with homology to xpa interacts with the c-terminal part of the 34-kda subunit of replication protein a. J. Biol. Chem. 1997, 272, 6561-6566.

[33] Otterlei, M., Warbrick, E., Nagelhus, T. A., Haug, T., Slupphaug, G., Akbari, M., Aas, P. A., Steinsbekk, K., Bakke, O., Krokan, H. E. Post-replicative base excision repair in replication foci. EMBO J. 1999, 18, 3834-3844.
[34] Krokan, H. E., Drabløs, F., Slupphaug, G. Uracil in DNA - occurrence, consequences and repair. Oncogene. 2002, 21, 8935-8948.
[35] Bennett, S. E., Sanderson, R. J., Mosbaugh, D. W. Processivity of escherichia-coli and rat-liver mitochondrial uracil-DNA glycosylase is affected by nacl concentration. Biochemistry. 1995, 34, 6109-6119.
[36] Higley, M., Lloyd, R. S. Processivity of uracil DNA glycosylase. Mutat. Res. 1993, 294, 109-116.
[37] 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.
[38] Dinner, A. R., Blackburn, G. M., Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. Nature. 2001, 413, 752-755.
[39] Werner, R. M., Stivers, J. T. Kinetic isotope effect studies of the reaction catalyzed by uracil DNA glycosylase: Evidence for an oxocarbenium ion-uracil anion intermediate. Biochemistry. 2000, 39, 14054-14064.
[40] 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.
[41] 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.
[42] 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.
[43] DeLano, W. L., The pymol molecular graphics system, 2002, DeLano Scientific, San Carlos, CA, USA
[44] van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D. P., Glattli, A., Hunenberger, P. H., Kastenholz, M. A., Ostenbrink, C., Schenk, M., Trzesniak, D., van der Vegt, N. F. A., Yu, H. B. Biomolecular modeling: Goals, problems, perspectives. Angew. Chem. Int. Edit. 2006, 45, 4064-4092.
[45] Leach, A. R. Molecular modelling. second Ed., Person Education Limited, 2001.
[46] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. J. Am. Chem. Soc. 1995, 117, 5179-5197.
[47] van Gunsteren, W., Berendsen, H. J. C., Molecular simulation (gromos) library manual, 1987, Groningen
[48] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R.,

Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998, 102, 3586-3616.

[49] Weiner, S. J., Kollman, P. A., Nguyen, D. T., Case, D. A. An all atom force-field for simulations of proteins and nucleic-acids. J. Comput. Chem. 1986, 7, 230-252.

[50] 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.

[51] Almlof, M., Brandsdal, B., Aqvist, J. Accurate predictions of protein-protein binding free energies. Abstr. Pap. Am. Chem. Soc. 2005, 229, U784-U784.

[52] Almlof, M., Brandsdal, B. O., Aqvist, J. Binding affinity prediction with different force fields: Examination of the linear interaction energy method. J. Comput. Chem. 2004, 25, 1242-1254.

[53] 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.

[54] 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.

[55] Tarus, B., Straub, J. E., Thirumalai, D. Dynamics of asp23-lys28 salt-bridge formation in a beta(10-35) monomers. J. Am. Chem. Soc. 2006, 128, 16159-16168.

[56] van Vlijmen, H. W. T., Schaefer, M., Karplus, M. Improving the accuracy of protein pk(a) calculations: Conformational averaging versus the average structure. Proteins. 1998, 33, 145-158.

[57] Papaleo, E., Olufsen, M., De Gioia, L., Brandsdal, B. O. Optimization of electrostatics as a strategy for cold-adaptation: A case study of cold- and warm-active elastases. J. Mol. Graph. Model. 2006,

[58] Caflisch, A., Karplus, M. Acid and thermal-denaturation of barnase investigated by molecular-dynamics simulations. J. Mol. Biol. 1995, 252, 672-708.

[59] Li, A. J., Daggett, V. Characterization of the transition-state of protein unfolding by use of molecular-dynamics - chymotrypsin inhibitor-2. Proc. Natl. Acad. Sci. U. S. A. 1994, 91, 10430-10434.

[60] Li, A. J., Daggett, V. Molecular dynamics simulation of the unfolding of barnase: Characterization of the major intermediate. J. Mol. Biol. 1998, 275, 677-694.

[61] 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.

[62] 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.

[63] 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.

[64] 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.
[65] Åqvist, J., Medina, C., Samuelsson, J. E. New method for predicting binding-affinity in computer-aided drug design. Protein Eng. 1994, 7, 385-391.
[66] 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.
[67] 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.
[68] 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.
[69] 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.
[70] 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.
[71] 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.
[72] Zhang, Q., Schlick, T. Stereochemistry and position-dependent effects of carcinogens on tata/tbp binding. Biophys. J. 2006, 90, 1865-1877.
[73] Honig, B., Nicholls, A. Classical electrostatics in biology and chemistry. Science. 1995, 268, 1144-1149.
[74] Onufriev, A., Bashford, D., Case, D. A. Modification of the generalized born model suitable for macromolecules. J. Phys. Chem. B. 2000, 104, 3712-3720.
[75] 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.
[76] Sitkoff, D., Sharp, K. A., Honig, B. Accurate calculation of hydration free-energies using macroscopic solvent models. J. Phys. Chem. 1994, 98, 1978-1988.
[77] Nicholls, A., Honig, B. A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the poisson-boltzmann equation. J. Comput. Chem. 1991, 12, 435-445.
[78] Davis, M. E., Mccammon, J. A. Solving the finite-difference linearized poisson-boltzmann equation - a comparison of relaxation and conjugate-gradient methods. J. Comput. Chem. 1989, 10, 386-391.
[79] Simonson, T. Macromolecular electrostatics: Continuum models and their growing pains. Curr. Opin. Struc. Biol. 2001, 11, 243-252.

[80] Demchuk, E., Wade, R. C. Improving the continuum dielectric approach to calculating pk(a)s of ionizable groups in proteins. J. Phys. Chem. 1996, 100, 17373-17387.
[81] Schutz, C. N., Warshel, A. What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins. 2001, 44, 400-417.
[82] Warshel, A., Sharma, P. K., Kato, M., Parson, W. W. Modeling electrostatic effects in proteins. Bba-Proteins Proteom. 2006, 1764, 1647-1676.
[83] Lanes, O., Guddal, P. H., Gjellesvik, D. R., Willassen, N. P. Purification and characterization of a cold-adapted uracil-DNA glycosylase from atlantic cod (gadus morhua). Comp. Biochem. Physiol. 2000, 127, 399-410.
[84] Leiros, I., Lanes, O., Sundheim, O., Helland, R., Smalås, A. O., Willassen, N. P. Crystallization and preliminary x-ray diffraction analysis of a cold-adapted uracil-DNA glycosylase from atlantic cod (gadus morhua). Acta Crystallogr. D. 2001, 57, 1706-1708.
[85] 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.
[86] Hoyoux, A., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Huston, A. L., Marx, J. C., Sonan, G., Zeng, Y. X., Feller, G., Gerday, C. Extreme catalysts from low-temperature environments. J. Biosci. Bioeng. 2004, 98, 317-330.
[87] Tang, K. E. S., Dill, K. A. Native protein fluctuations: The conformational-motion temperature and the inverse correlation of protein flexibility with protein stability. J. Biomol. Struct. Dyn. 1998, 16, 397-411.
[88] Zecchinon, L., Claverie, P., Collins, T., D'Amico, S., Delille, D., Feller, G., Georlette, D., Gratia, E., Hoyoux, A., Meuwis, M. A., Sonan, G., Gerday, C. Did psychrophilic enzymes really win the challenge? Extremophiles. 2001, 5, 313-321.
[89] Tsou, C. L. The role of active site flexibility in enzyme catalysis. Biochemistry-Moscow. 1998, 63, 253-258.
[90] Jaenicke, R. Stability and stabilization of globular proteins in solution. J. Biotechnol. 2000, 79, 193-203.
[91] Li, W. F., Zhou, X. X., Lu, P. Structural features of thermozymes. Biotechnol. Adv. 2005, 23, 271-281.
[92] Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E. Z., Kern, D. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nat. Struct. Mol. Biol. 2004, 11, 945-949.
[93] Tindbaek, N., Svendsen, A., Oestergaard, P. R., Draborg, H. Engineering a substrate-specific cold-adapted subtilisin. Protein. Eng. Des. Sel. 2004, 17, 149-156.
[94] Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L., Engel, P. C. Protein stability in extremophilic archaea. Front. Biosci. 2000, 5, D787-D795.
[95] Dill, K. A. Dominant forces in protein folding. Biochemistry. 1990, 29, 7133-7155.
[96] Kumar, S., Tsai, C. J., Nussinov, R. Maximal stabilities of reversible two-state proteins. Biochemistry. 2002, 41, 5359-5374.

[97] Pace, C. N. Polar group burial contributes more to protein stability than nonpolar group burial. Biochemistry. 2001, 40, 310-313.
[98] Pace, C. N., Alston, R. W., Shaw, K. L. Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. Protein Sci. 2000, 9, 1395-1398.
[99] Kumar, S., Nussinov, R. Salt bridge stability in monomeric proteins. J. Mol. Biol. 1999, 293, 1241-1255.
[100] Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Helle, N. M., Haug, T., Levine, D. W., Krokan, H. E. Properties of a recombinant human uracil-DNA glycosylase from the ung gene and evidence that ung encodes the major uracil-DNA glycosylase. Biochemistry. 1995, 34, 128-138.
[101] Kumar, S., Nussinov, R. Relationship between ion pair geometries and electrostatic strengths in proteins. Biophys. J. 2002, 83, 1595-1612.
[102] Hoseki, J., Okamoto, A., Masui, R., Shibata, T., Inoue, Y., Yokoyama, S., Kuramitsu, S. Crystal structure of a family 4 uracil-DNA glycosylase from thermus thermophilus hb8. J. Mol. Biol. 2003, 333, 515-526.

Paper I

ARTICLE IN PRESS

doi:10.1016/j.jmb.2004.09.004 J. Mol. Biol. (2004) xx, 1–10

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Optimisation of the Surface Electrostatics as a Strategy for Cold Adaptation of Uracil-DNA *N*-glycosylase (UNG) from Atlantic Cod (*Gadus morhua*)

Elin Moe[1], Ingar Leiros[2], Ellen Kristin Riise[3], Magne Olufsen[3] Olav Lanes[4], Arne O. Smalås[3] and Nils Peder Willassen[1]*

[1]*Department of Molecular Biotechnology, Institute of Medical Biology, Faculty of Medicine, University of Tromsø N-9037 Tromsø, Norway*

[2]*The European Synchrotron Radiation Facility, 38043 Grenoble Cedex, France*

[3]*The Norwegian Structural Biology Centre, University of Tromsø, N-9037 Tromsø Norway*

[4]*Biotec Pharmacon ASA N-9008 Tromsø, Norway*

**Corresponding author*

Cold-adapted enzymes are characterised by an increased catalytic efficiency and reduced temperature stability compared to their mesophilic counterparts. Lately, it has been suggested that an optimisation of the electrostatic surface potential is a strategy for cold adaptation for some enzymes. A visualisation of the electrostatic surface potential of cold-adapted uracil-DNA *N*-glycosylase (cUNG) from Atlantic cod indicates a more positively charged surface near the active site compared to human UNG (hUNG). In order to investigate the importance of the altered surface potential for the cold-adapted features of cod UNG, six mutants have been characterised and compared to cUNG and hUNG. The cUNG quadruple mutant (V171E, K185V, H250Q and H275Y) and four corresponding single mutants all comprise substitutions of residues present in the human enzyme. A human UNG mutant, E171V, comprises the equivalent residue found in cod UNG. In addition, crystal structures of the single mutants V171E and E171V have been determined. Results from the study show that a more negative electrostatic surface potential reduces the activity and increases the stability of cod UNG, and suggest an optimisation of the surface potential as a strategy for cold-adaptation of this enzyme. Val171 in cod UNG is especially important in this respect.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: uracil-DNA *N*-glycosylase; cold-adapted enzymes; electrostatics; stability; crystal structure

Introduction

Cold-adapted enzymes possess lower thermal stability and higher catalytic efficiency than their mesophilic homologues. The dominating hypothesis is that these features originate from an increased structural flexibility.[1] Comparative studies have shown that cold-adapted enzymes have fewer cross-linking interactions, in particular in regions associated with the catalytic site. Although not proven, these findings indicate that increased local flexibility in key regions of the structure is a major determinant for cold activity.[2] However, other strategies for cold adaptation might exist, and an optimisation of the electrostatic surface potential near the active site is observed for some cold-adapted enzymes. Salmon trypsin is reported to have a more negatively charged surface near the catalytic area compared to bovine trypsin.[3] The increased negative potential is believed to cause the increased substrate affinity for positively charged substrates, and the optimised surface potential is suggested to be one of the strategies for adaptation of trypsin to cold environments.[4] Malate dehydrogenase (MD) from *Aquaspirillium arcticum* is reported to have a more positive potential on the surface around the substrate-binding site and a decreased negative potential around the co-factor-binding site compared to thermophilic MD.[5] It is suggested that the altered potentials may facilitate the interaction with the negatively charged substrate (oxaloacetate) and increase the catalytic efficiency at low temperatures. Citrate synthase (CS) from an Antarctic bacterium

Abbreviations used: UNG, uracil-DNA *N*-glycosylase; cUNG, recombinant cod UNG; hUNG, recombinant human UNG.

E-mail address of the corresponding author: nilspw@fagmed.uit.no

0022-2836/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.

ARTICLE IN PRESS

is reported to have a more negatively charged surface potential near the positively charged active site of the enzyme compared to thermophilic CS.[6] The hypothesis is that the negative potential serves to focus the negatively charged substrate (oxaloacetate) towards the positive potential of the active site.

Uracil-DNA *N*-glycosylase (UNG) is a DNA repair enzyme involved in the base excision repair pathway (BER) for removal of uracil in DNA.[7] The structure of the catalytic domain is known from human UNG[8] herpes simplex virus 1 (HSV-1) UNG,[9] *Escherichia coli* UNG[10] and Atlantic cod UNG.[11] The four homologues have similar structures and consist of a classic single-domain α/β-fold with a central four-stranded parallel and twisted β-sheet surrounded by 11 α-helices. The active site is located within a positively charged groove at the C-terminal end of the parallel β-sheet. Four loops are involved in the detection and removal of uracil in DNA: The 4-Pro loop (165-PPPPS-169), the Gly-Ser loop (246-GS-247), the Leu272-loop (268-HPSPLSVY-275) and the water-activating loop (145-DPYH-148). Some of the most important residues for catalysis are the conserved His268 and Asp145. His268 is located in the Leu272-loop, while Asp145 is located in the water-activating loop.[12]

The gene encoding the catalytic domain of cold-adapted UNG from Atlantic cod (*Gadus morhua*) has been identified.[13] Biochemical analysis of recombinant cod UNG (cUNG) revealed typical cold-adapted features such as low temperature stability, low temperature optimum and increased catalytic efficiency compared to recombinant human UNG (hUNG). In addition, the kinetic constants of cUNG showed a reduced K_M value, thus an increased substrate affinity compared to hUNG. A visualisation of the electrostatic surface potential of the three-dimensional structure of cUNG indicates that it possesses a more positive potential near the active site compared to hUNG.[11] The increased positive charge on the surface of cod UNG might explain the increased affinity for the negatively charged substrate DNA.

Further analyses suggested that four amino acid residues on the surface of cUNG were responsible for the differences in surface potentials. In order to study how the surface electrostatics affect the cold-adapted features of cod UNG, one mutant with four substituted amino acids (V171E, K185V, H250Q and H275Y) and four corresponding single mutants were constructed and characterised, and the 3D-structure of the single mutant V171E was determined. In addition, a mutant of human UNG, E171V, was investigated. The results show that the mutations affect both the catalytic efficiency and the stability of cod UNG. In particular, Val171 is shown to be essential in order to maintain the cold-adapted features of cUNG. The substitution of Val with Glu reduced the substrate affinity, with a 50% reduction in activity as a result, and increased the stability by adding one ion-pair.

Results

Construction of mutants

A visualisation of the electrostatic surface potential of the crystal structure of cUNG using SwissPDBViewer[14] and GRASP[15] reveals a more positive surface potential near the DNA binding site compared to hUNG (Figure 1(a) and (b)). Five mutants of cUNG were constructed on the basis of sequence and structural analysis, one mutant with four substitutions, V171E, K185V, H250Q and H275Y (quadruple), and four corresponding single mutants. In addition, one mutant of recombinant hUNG, E171V, was constructed. A model of cUNG with the four amino acid substitutions is shown in Figure 1(c).

Large-scale expression and purification

The cUNG mutants and cUNG were produced in a 15 l Chemap 3000 fermenter and purified by ion-exchange chromatography and gel-filtration to apparent purity as shown in Figure 2.

The molecular mass of the mutants of cod UNG was determined to be approximately 28 kDa, which is the same as observed previously.[13] The hUNG migrates slower than cUNG, with a molecular mass estimated as 29 kDa. The yield varied, but was typically between 2 mg and 4 mg per fermentation. The hUNG E171V mutant was produced and purified using the methods described above (results not shown).

Kinetics

The kinetic constants k_{cat} and K_M were determined for cUNG, hUNG and the mutants at three different temperatures, 15 °C, 22 °C and 37 °C at the optimal condition for cUNG (pH 7.5 and 50 mM NaCl). The results are summarised in Table 1.

The results show that k_{cat} of the quadruple mutant is reduced by approximately 40%, while k_{cat} of the V171E and H275Y mutants is reduced by more than 50% compared to cUNG. k_{cat} of K185V is increased at 37 °C and 15 °C, while it is decreased at 22 °C compared to cUNG. The k_{cat} value of H250Q is increased at 37 °C and reduced at 22 °C relative to cUNG, while it is approximately the same as of cUNG at 15 °C.

The K_M values of the V171E mutant are approximately on the same level as hUNG, indicating a reduced substrate affinity compared to cUNG. Quadruple and H275Y possess slightly lower K_M values than cUNG, while the K185V and H250Q mutants have approximately the same K_M values as cUNG.

Alterations of the kinetic constants of the quadruple, V171E and H275Y mutants result in reduced catalytic efficiencies (k_{cat}/K_M) compared to cUNG. Cod UNG K185V possesses an increased catalytic efficiency at 37 °C and 22 °C and decreased efficiency at 15 °C, while the H250Q mutant

ARTICLE IN PRESS

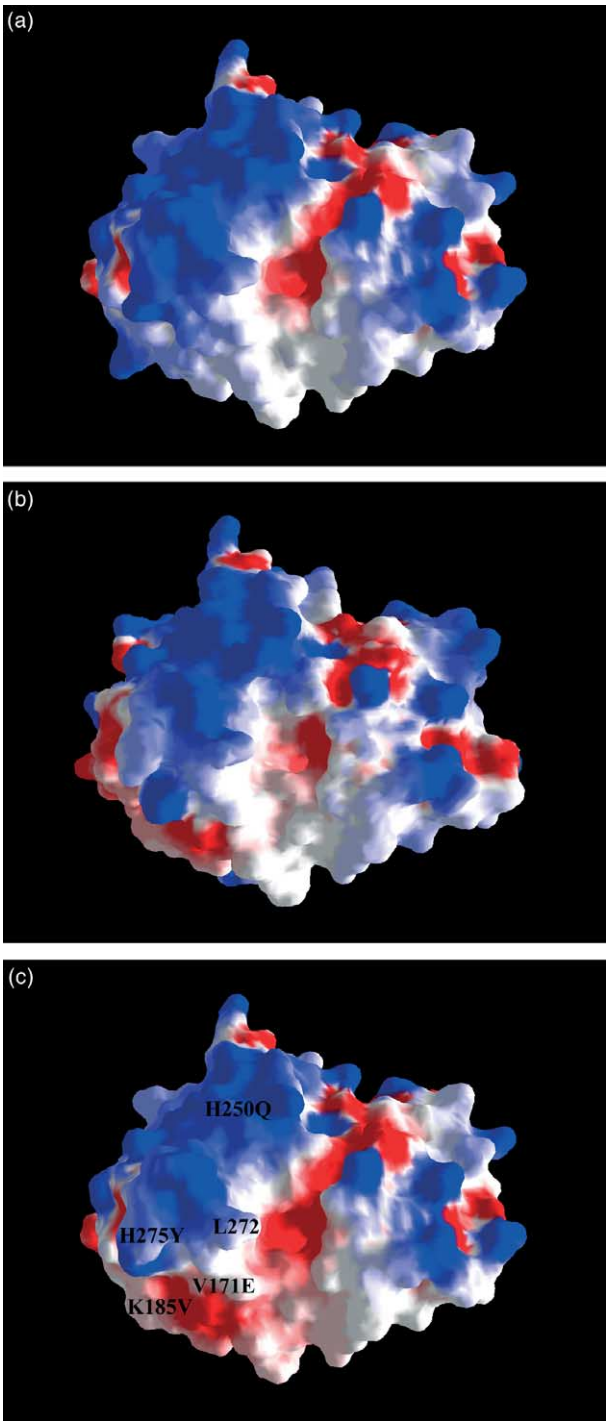


Figure 1. Difference in electrostatic surface potential of cUNG and hUNG. The Figure was made using the SwissPDBViewer,[14] with an electrostatic surface potential imported from GRASP[15] and contoured at ± 3 kT/e, where red describes a negative potential and blue describes a positive potential: (a) cUNG, (b) hUNG, (c) cUNG with substituted residues believed to increase the surface potential.

shows higher catalytic efficiency at 37 °C and lower catalytic efficiency at 22 °C and 15 °C, relative to cUNG. The hUNG E171V mutant possesses an increased catalytic efficiency at all three

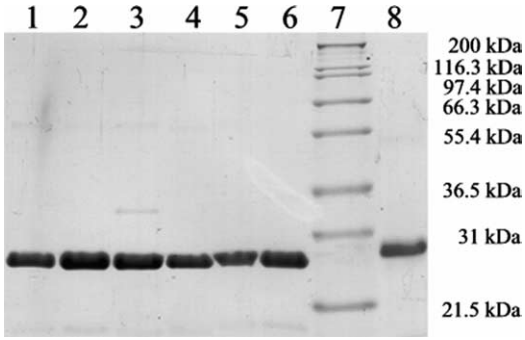


Figure 2. SDS-PAGE of purified cUNG and mutants. Approximately 3 μg of each enzyme was applied to the gel. Lane 1, cUNG; lane 2, rcV171E, lane 3, rcK185V, lane 4, rcH250Q, lane 5, rcH275Y, lane 6, quadruple, lane 7, Mark 12™ wide-range protein standard (NOVEX, San Diego, CA); lane 8, hUNG.

temperatures compared to hUNG due to reduced K_M values.

Temperature stability

The thermal stability of cUNG, hUNG and the mutants were determined at 37 °C. The results are shown in Figure 3. There is an apparent increase in the stability of all the mutants at 37 °C compared to cUNG, and the quadruple and the V171E mutants seem to be as stable as hUNG. Analysis of the temperature stability of the hUNG E171V mutant at 37 °C shows reduced stability compared to hUNG.

Crystal structure of the cUNG V171E and hUNG E171V mutants

The crystal structure of the catalytic domain of the cUNG V171E and hUNG E171V mutants has been determined at 1.7 Å and 1.95 Å resolution, respectively. The overall structures of the mutants are very similar to the catalytic domains of cUNG[11] and hUNG[8] with an overall root-mean-square (r.m.s.) difference of 0.27 Å and 0.32 Å, based on all main-chain atoms.

Electron density near residue 171 in the structure of cUNG V171E confirms that the valine to glutamic acid mutation introduces the expected ion pair interaction between Glu171 and His186. The orientation of Glu171 is slightly different from the corresponding residue in hUNG (Figure 4), but the ion pair distances are comparable, indicating similar strength of the interactions. Distances between the carboxylate oxygen atoms of Glu171 and the His nitrogen atom are 3.05 Å and 3.65 Å in hUNG, and 3.12 Å and 3.33 Å in the cUNG V171E mutant, respectively. The structure of the hUNG E171V mutant, correspondingly show no dramatic changes compared to hUNG at the mutation site other than the expected loss of the 171–186 interactions (Figure 5). The structure becomes more flexible in this region and the local

ARTICLE IN PRESS

Table 1. Kinetic constants for cUNG, hUNG and mutants at 37 °C, 22 °C and 15 °C

UNG	Temperature (°C)	V_{max} (nmol min^{-1} mg^{-1})	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (min^{-1} μM^{-1})
cUNG	37	49,135±1626	1241	0.8±0.09	1499
	22	26,807±555	677	0.8±0.06	892
	15	11,064±374	279	0.5±0.07	607
Quadruple (V171E, K185V, H250Q, H275Y)	37	29,204±2210	737	0.6±0.22	1219
	22	13,244±282	334	0.5±0.05	612
	15	6753±213	171	0.5±0.07	335
cUNGV171E	37	20,189±553	510	2.0±0.12	260
	22	8070±239	204	1.7±0.12	117
	15	4041±143	102	1.4±0.13	72
cUNGK185V	37	63,997±1607	1616	0.8±0.07	1948
	22	25,865±617	653	0.6±0.06	1166
	15	16,159±750	408	0.8±0.13	510
cUNGH250Q	37	52,888±1079	1335	0.8±0.06	1637
	22	20,063±512	507	0.6±0.06	844
	15	10,879±282	275	0.7±0.07	419
cUNGH275Y	37	23,590±859	596	0.6±0.09	1017
	22	11,974±553	302	0.6±0.11	509
	15	6166±384	156	0.7±0.16	224
hUNG	37	25,347±1256	647	2.1±0.23	309
	22	12,530±740	320	2.4±0.29	135
	15	5580±399	142	2.2±0.34	66
hUNGE171V	37	26,529±885	677	1.0±0.1	684
	22	10,256±285	262	0.7±0.07	396
	15	6545±253	167	1.0±0.12	164

environment becomes similar to that of cUNG, although the conformation of the Val171 side-chain is slightly different in the two structures.

Discussion

An optimisation of the electrostatic surface potential is suggested to be a strategy for adaptation of enzymes to cold environments. A qualitative evaluation of the electrostatic surface potential of the 3D structure of cod UNG revealed more positive potential near the active site compared to human UNG. The presence of residues giving rise to more positive electrostatic potential in the active-site region might explain the increased substrate affinity (reduced K_M) and thus, in part, the enhanced catalytic efficiency of this enzyme. In order to investigate this hypothesis, five mutants of cUNG and one mutant of hUNG were produced, characterised and compared to cUNG and hUNG. In addition, the 3D X-ray structure of the cUNG V171E and the hUNG E171V mutants were determined

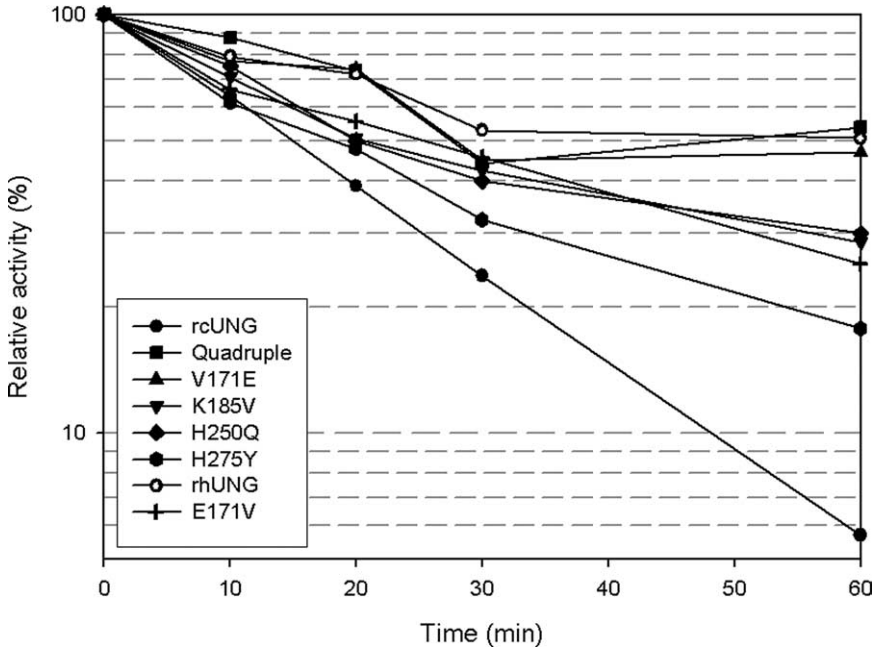


Figure 3. Temperature stability of cUNG, hUNG and six mutants at 37 °C. Stability was measured after incubation at 37 °C in 10, 20, 30 and 60 minutes. Standard deviations are not indicated, but were typically found in the range of 1–5%.

ARTICLE IN PRESS

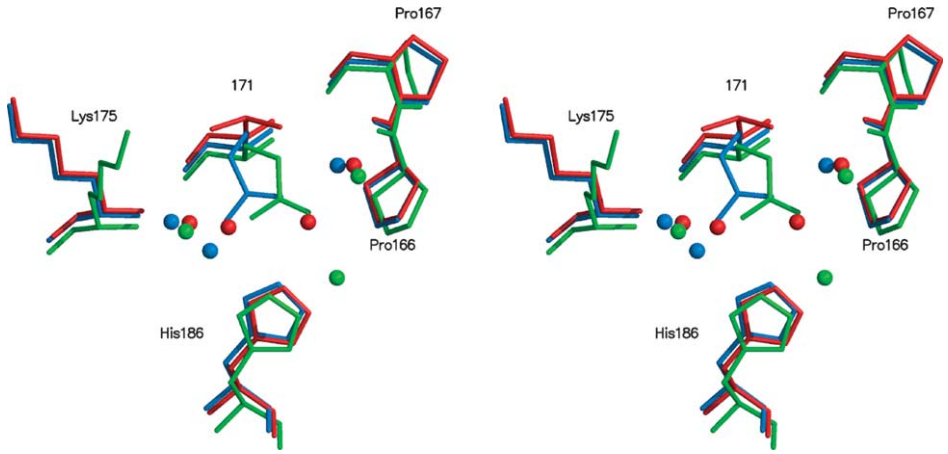


Figure 4. Stereo view of residue 166, 167, 171, 175 and 186 of cUNG V171E (blue), cUNG (red) and hUNG (green). The water molecules near residue 171 are the same colour as the enzyme. One ion-pair, less than 4 Å, is created between Glu171 and His186 in the mutant.

and analysed. The structure of cUNG with locations of the substituted amino acids is illustrated in Figure 6.

The cUNG V171E mutant seems to closely resemble hUNG in its behaviour. Compared to cUNG, the K_M values are increased at all temperatures tested, resulting in a significantly reduced catalytic efficiency. Substituting Val with Glu reduces the surface potential near the active site and leads to a weaker interaction between UNG and the negatively charged substrate DNA. The V171E mutant also shows increased temperature stability at 37 °C. According to the structure of hUNG,[8] Glu171 is involved in two salt-bridges, one weak interaction to Lys175 (5.7 Å) and one to His186 (3.05 Å and 3.65 Å). Residue 171 is located in helix α6, and the ion pair Glu171–Lys175 is intrahelical, while Glu171–His186 runs from α6 to the loop between α6 and α7. The helix α6 is located close to the DNA-binding area and moves upon DNA-binding. The corresponding salt-bridges are not present in cUNG. However, the crystal structure of the cUNG V171E mutant shows that the substitution of Val171 to Glu introduces one salt-bridge between Glu171 and His186 (3.12 Å and 3.33 Å) (Figure 4). This additional salt-bridge probably results in a more rigid structure of this area of the mutant. On the basis of the hypothesis about an inverse relationship between rigidity and activity, it is possible to explain the reduced k_{cat} by a decreased flexibility of this mutant compared to cUNG. The reduced flexibility also explains the increased temperature stability of this mutant.

The hUNG E171V mutant has correspondingly obtained reduced temperature stability and increased catalytic efficiency as cold-adapted features. Substituting Glu171 with Val leads to a removal of the previously described salt-bridges between Glu171 and Lys175 and His186 (Figure 5), giving rise to reduced stability of structural elements close to the active site. The increased catalytic efficiency is caused mainly by an increased substrate affinity (reduced K_M) and might be explained by an enhanced interaction between the

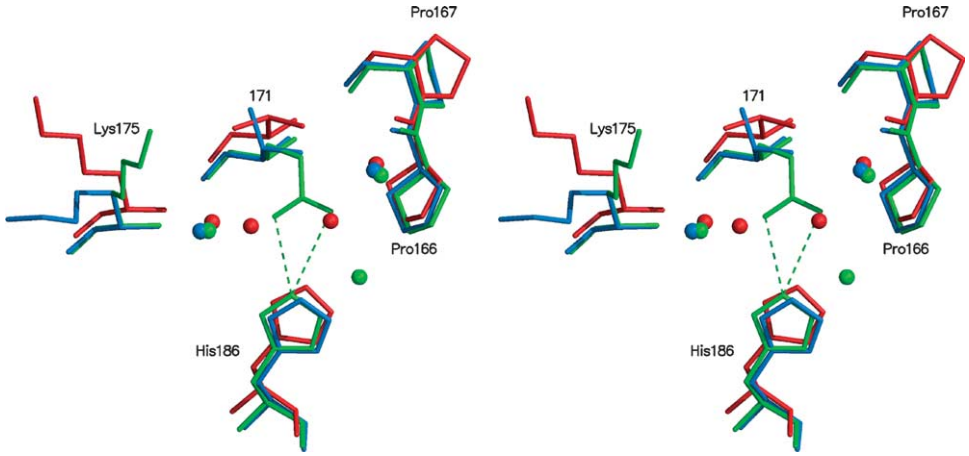


Figure 5. Stereo view of residues 166, 167, 171, 175 and 186 of hUNG E171V (blue), cUNG (red) and hUNG (green). The water molecules near residue 171 are the same colour as the enzyme. The broken line indicates the salt-bridge between 171 OE1 and OE2 to 186 NE2 in hUNG.

ARTICLE IN PRESS

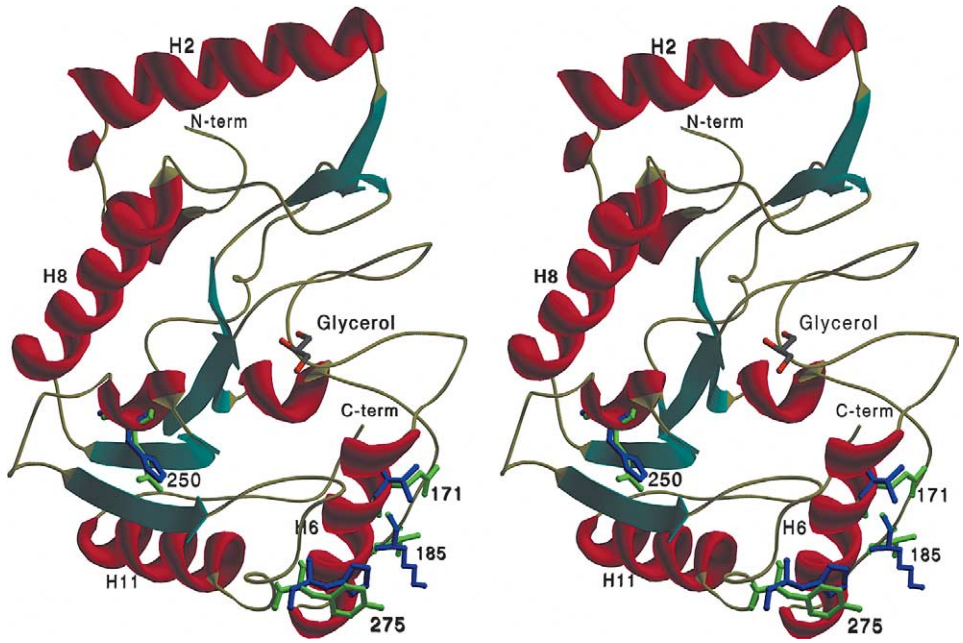


Figure 6. Stereo view of the crystal structure of cUNG with residues believed to affect the electrostatic surface potential (V171, K185, H250 and H275). cUNG residues are shown in blue, while hUNG residues are shown in green. The glycerol molecule found in the active site of cUNG is included for clarity. Secondary structure elements are as defined by Leiros *et al*.[11]

enzyme and the substrate due to the removal of the negatively charged Glu, and the increased flexibility resulting from the removal of an ion-pair interaction. On the basis of the results from analysis of the V171E and the E171V mutants, it seems as if residue 171 is a structural determinant for catalytic activity and stability in both hUNG and cUNG.

The catalytic efficiency of the cUNG K185V mutant is increased at 37 °C and 22 °C, and reduced slightly at 15 °C relative to cUNG, i.e. displays a slight loss of cold-adaptation properties. The structures of cUNG and hUNG reveal that this amino acid is located at the surface of the enzyme, in the loop between α6 and α7 in cUNG and has no interaction with other amino acids (Figure 6). According to the theory about a reduced substrate interaction caused by a more negatively charged surface, one would expect the catalytic efficiency to be reduced for this mutant. However, it is possible that the distance between residue 185 and the substrate is too large to affect the enzyme activity. The results from the temperature stability measurement show an increased stability compared to cUNG, thus Val185 has a stabilising effect at 37 °C.

The catalytic efficiency of the cUNG H250Q mutant is increased at 37 °C, while it is similar at 22 °C and reduced at 15 °C compared to cUNG, and as for the cUNG K185V mutant, this indicates loss of cold-adaptation features. In the structure of cUNG, His250 is located in α9, and forms hydrogen bonds with Ala266 and Leu244. Ala266 is located at the terminus of the β4 strand near the catalytically important Leu272 loop, and Leu244 is located in the β3 strand, near the highly conserved Gly-Ser loop (residues 246–247) (Figure 6). Thus, His250 forms hydrogen bonds to residues close to regions involved in DNA-binding and catalysis. In hUNG, Gln250 forms a hydrogen bond to Thr266, but not to residue 244. The substitution of His250 to Gln in cUNG could remove the hydrogen bond between residues 250 and 244 in cUNG and introduce a more flexible Gly-Ser loop, thereby explaining the increased catalytic efficiency of the H250Q mutant at 37 °C. However, the suggested increased flexibility does not seem to improve the catalytic efficiency at low temperatures.

The catalytic efficiency of the cUNG H275Y mutant is decreased compared to cUNG at all temperatures measured, while the temperature stability is increased slightly. His275 is located in the proximity of the so-called Leu272-loop, or the minor groove intercalation loop. The movement of this loop is important for bringing the catalytically important His268 within hydrogen-bonding distance of uracil O2, and for accomplishing the formation of the uracil-recognition pocket.[12] The substitution of His275 to Tyr probably reduces the flexibility of the Leu272 loop, and hence reduces the catalytic efficiency and increases the temperature stability of cUNG.

The quadruple mutant possesses all the four substitutions as described above, and adopts features such as reduced catalytic efficiency and increased temperature stability compared to cUNG. It has obtained a stronger substrate affinity (lower K_M) than cUNG, while all the single mutants have a weaker or similar substrate affinity than cUNG. The temperature stability of the quadruple mutant (Figure 3) seems to be additive, as it is more stable than any of the mutants, but the activity for the

ARTICLE IN PRESS

quadruple mutant appears to be an average of the four individual mutations (Table 1). However, the total picture indicates that the quadruple mutant displays a greater loss in catalytic efficiency at lower temperatures, whereas the temperature stability is increased, again indicative of a loss of cold-adaptation features.

The difference in catalytic efficiency between most cold-adapted enzymes and homologous mesophilic enzymes increases with decreasing temperatures. This is true also for cUNG and hUNG, where the difference in catalytic efficiency is 80% at 37 °C, and 89% at 15 °C (Table 1). On the basis of this observation it should be possible to evaluate the effect of amino acid substitution on the cold-adapted features of cUNG, by calculating the difference in catalytic efficiency at high and low temperatures between the mutants and the native cold-adapted enzyme as shown in Table 2.

Calculations show that the reduction in catalytic efficiency is most severe for the V171E mutant, but there is also a considerable effect for the quadruple and H275Y mutants, and the difference in catalytic efficiency increases with decreasing temperature for these mutants compared to cUNG. There seems to be no reduction in the catalytic efficiency of the K185V or the H250Q mutants at 37 °C or 22 °C, while some effect is observed at 15 °C. On the basis of this information and the results from the characterisation, it seems as if Val171 is the single most important residue among those investigated in order to maintain the cold-adapted features of cUNG, while the other residues are of a somewhat more moderate importance. However, all the mutants showed a decrease in catalytic efficiency that was most pronounced at low temperature, i.e. there appears to be a rational relationship between charge complementarity and cold adaptation.

Conclusion

The results from the present study show that the increased electrostatic surface potential is important for the catalytic efficiency of cUNG, hence indicating that an optimisation of the potential in the substrate-binding area is a strategy for cold adaptation of this enzyme. Val171 is shown to be particularly important in this respect. Substitution of Val171 to Glu in cUNG resulted in a reduced substrate affinity (increased K_M) that contributes strongly to the reduced catalytic efficiency. An additional salt-bridge was introduced by the mutation and led to an increased temperature stability compared to cUNG. The opposite features were observed for the hUNG E171V mutant, caused by an enhanced substrate interaction (reduced K_M) and removal of a salt-bridge. Thus the presence of Val171 in cUNG seems to optimise the interaction between the enzyme and the substrate, and increase the flexibility of the structure close to the active site, and by that means the enzyme performs a more efficient catalysis at low temperatures. Characterisations of the other surface charge mutants tested in this study (K185V, H250Q, H275Y and the quadruple mutant) confirm the relation between surface charge and cold activity.

The cross-mutation between Val and Glu in cUNG and hUNG at residue 171 demonstrates that local flexibility features and electrostatic environments, especially near the catalytic region of an enzyme, is of vital importance for the catalytic efficiency. The clear correlation between mutations imposing rigidity and poorer electrostatic complementarity between enzyme and substrate, with a reduced catalytic efficiency, an effect that is increasingly notable at reduced temperature, shows that cold activity of enzymes can in fact be manipulated. Additionally, the data on the mutants show a clear correlation between catalytic efficiency and stability. This is very strong support for the classical hypothesis on a close relation between stability and flexibility on one hand, and catalytic efficiency on the other.

Table 2. Reduction (%) in catalytic efficiency at different temperatures for cUNG mutants relative to cUNG

Construct	Temperature (°C)		
	37	22	15
cUNG	0	0	0
Quadruple	20	30	45
V171E	83	87	88
K185V	0	0	16
H250Q	0	0	31
H275Y	32	43	63

Materials and Methods

Materials

Vent DNA polymerase and EcoRI restriction enzyme were obtained from New England Biolabs (Beverly, MA). SalI restriction enzyme, dNTPs and phage T4 DNA ligase were purchased from Promega (Madison, WI). Primers were obtained from MedProbe (Eurogentech SA, Belgium), and the Qiaquick gel extraction kit was purchased from Qiagen (Germany). The GeneAmp 9700 Thermocycler was obtained from Perkin Elmer (Foster City, CA). Expression vector pTRC99A, Äkta Explorer Purification System, purification columns and deoxy [5-^{3}H]uridine 5′-triphosphate (16.2 Ci/mmol) were obtained from Amersham Biosciences (Uppsala, Sweden). The Bio-Rad Protein Assay Dye Reagent Concentrate and the Mini-Protean® 3 electrophoresis system were purchased from Bio-Rad (Hercules, CA), and SigmaPlot was obtained from SPSS Inc. (Chicago, IL). The QuikChange® site-directed mutagenesis kit was purchased from Stratagene (La Jolla, CA).

E. coli TOP10 [F-*mc*rA (*mrr-hs*dRMS-*mc*rBC) 80*lac*Z M15 *lac*X74 *de*oR *rec*A1 *ara*D139 (*ara-le*u) 7697 *ga*lU *ga*lK *rps*L (Str R) *en*dA1 *nu*pG] was purchased from Invitrogen (Carlsbad, CA) and *E. coli* NR8052 [Δ(*pro-lac, thi, ara, trpE9777, ung1*)] and purified recombinant human UNG (UNGΔ84) were kindly provided by Dr Hans E. Krokan, Institute for Cancer Research and Molecular Biology, Norwegian University of Science and Technology.

ARTICLE IN PRESS

Construction of mutants

The mutants of cUNG were constructed as described.[16] In short, two PCR fragments were amplified using the gene upstream primer and mutation downstream primer, and the mutation upstream primer and the gene downstream primer. The fragments were purified from gel using the Qiaquick gel extraction kit, and eluted in 20 μl of elution buffer supplied by the manufacturer. Then 1 μl of each of the purified fragments was applied in a new PCR reaction and ligated using the gene upstream and downstream primers.

PCR was performed in a GeneAmp 9700 Thermocycler. Each reaction had a volume of 50 μl and contained 0.4 unit of Vent DNA polymerase and buffer supplied by the manufacturer, 0.125 mM dNTPs and each of the upstream and downstream primers at a concentration of 1 μM, in addition to the template. The amplification was carried out at 94 °C for two minutes, 80 °C for two minutes, followed by 30 cycles at 94 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 30 seconds, and a final extension step at 72 °C for seven minutes. The polymerase was added to the PCR mixture during the step at 80 °C. The fragments containing the genes encoding mutated cUNG were purified from 1% (w/v) agarose gel, digested with EcoRI and SalI restriction enzymes, and ligated into the pTrc99A expression vector using 200 units of T4 DNA ligase. Competent *E. coli* TOP10 was transformed with the plasmid and screened for positive clones. Plasmid DNA was isolated from the TOP10 cells using the Qiagen plasmid purification kit, and used for transformation of competent *E. coli* NR8052 and sequencing.

The following mutants were constructed: V171E, K185V, H250Q, H275Y, and the quadruple mutant V171E, K185V, H250Q and H275Y. The template for construction of the mutants was the gene encoding the catalytic domain of recombinant cod UNG with three additional residues (MEF) and two altered Arg codons (from AGA to CGT) in the N-terminal part of the sequence (cUNGΔ81o).

The gene upstream and downstream primers for the cod mutants are OPUNG (5′-ATG GAA TTC TTC GGA GAG AC-3′) and NPUNG (5′-CTG CAG GTC GAC TTA GAG TGC-3′) or NPpTrc (5′-GAT GCC TGG CAG TTC CCT AC-3′). OPUNG posesses an EcoRI restriction enzyme site and NPUNG posesses a SalI restriction enzyme site. Upstream primers for the mutants are (the mutation codons are underlined): OPV171E (5′-C AGT CTC GAA AAC ATA TAC-3′), OPK185V (5′-GAT GGC TTC GTG CAT CCT GG-3′), OPH250Q (5′-ca tac gcc CAG aag aag gga g-3′), and OPH275Y (5′-G TCT GCT TAT CGT GGG TTC-3′). The downstream primers were complementary and reverse of the primers listed above.

The human UNG mutant E171V was constructed using the QuikChange® site-directed mutagenesis kit and according to the manual from the manufacturer. The primers used were as follows: 5′-GTT CCG CCT CCG CCC AGT TTG GTG AAC ATT TAT AAA G-3′; 5′-C TTT ATA AAT GTT CAC CAA ACT GGG CGG AGG CGG AAC-3′.

DNA sequencing

DNA sequencing was performed using the Amersham Biosciences Thermo Sequenase Cy5 Dye Terminator Kit, ALFexpress™ DNA sequencer and ALFwin Sequence Analyser version 2.10, and according to the protocol supplied by the manufacturer. Gels were made with Reprogel™ Long Read and Reproset UV-polymeriser. All items were purchased from Amersham Biosciences (Uppsala, Sweden).

Expression

Large-scale expression was performed using a 15 l Chemap 3000 fermenter (Switzerland). A 200 ml preculture of *E. coli* NR8052 transformed with pTrc99A containing genes encoding the cUNG or mutants were used for inoculation of 7 l of 2×LB-medium supplemented with 20 mM glucose and 100 μg/ml of ampicillin. The cells were grown at 30 °C and the expression induced with 1 mM IPTG at A_{600}=2.0: 50 ml of 20% (w/v) glucose was added to the growth medium every hour after induction to avoid glucose starvation. The cells were harvested six hours after induction by centrifugation at 10,000*g* for ten minutes and stored at −20 °C.

Protein purification

The cUNG and the H275Y mutant were purified as described.[13] The quadruple, V171E, K185V, H250Q and E171V mutants were purified without using the Blue Sepharose FF column (1.6/5.0). This purification step was exchanged with purification on a Source 15 S column (2.6/3.0), followed by concentration of the sample to 3 ml and further purification on a Superdex 75 column (2.6/60). Fractions containing UNG after gel-filtration were pooled, concentrated and stored at −20 °C in 50% (v/v) glycerol. All solutions containing enzyme were kept on ice except during the purification steps, which were performed at room temperature. The centrifugations were performed at 4 °C.

Activity measurements

Preparation of substrate ([^{3}H]dUMP DNA made by nick translation) and measurements of UNG activity was performed as described.[17] One unit is defined as the amount of enzyme required for releasing 1 nmol of acid-soluble uracil per minute at 37 °C.

Protein determination

Protein concentration were determined with Bio-Rad Protein Assay Dye Reagent Concentrate based on the Bradford dye-binding procedure[18] according to the microtiter plate protocol described by the manufacturer (Bio-Rad, Hercules, CA), using bovine serum albumine (BSA) as standard. SDS-PAGE was performed as described,[19] with a 12% (w/v) polyacrylamide separating gel and a 5% (w/v) polyacrylamide stacking gel using the Mini-Protean® 3 electrophoresis system (BioRad, Hercules, CA). After electrophoresis, gels were stained with Coomassie brilliant blue G250.

Kinetics

K_M and k_{cat} were measured in the presence of eight different concentrations of [^{3}H]dUMP substrate in the range of 0.56–4.5 μM at 15 °C, 22 °C and 37 °C. The amounts of UNG used were different for each mutant at the different temperatures, giving cpm values between 500 and 5000. Assays were performed in 25 mM Tris–HCl (pH 7.5 at 37 °C), 50 mM NaCl, 1 mM EDTA and 100 μg/ml of BSA. Calculation of the kinetic constants was performed by the enzyme kinetics module in SigmaPlot (SPSS Inc., Chicago, IL).

ARTICLE IN PRESS

Temperature stability measurements

The temperature stability was examined by incubation of diluted enzyme in 10 mM Tris–HCl (pH 8.0), 50 mM NaCl, 1 mM EDTA, 1% (v/v) glycerol in a total volume of 50 µl at 37 °C (pH was adjusted at 37 °C). After different time intervals, as indicated in the Figure legends, the tubes were transferred onto ice. From each tube, 5 µl was used in assay mixtures for measuring residual activity under standard assay conditions.

X-ray structure determination

Crystals of the cod UNG V171E mutant were obtained by the vapour-diffusion, hanging-drop method at room temperature with a protein concentration of 10 mg/ml prior to mixing with equal volumes of reservoir solution (2 µl) of 1.4 M sodium citrate in a 0.1 M Hepes-buffer at pH 7.5. Diffraction quality crystals of maximal size of 0.4 mm appeared in the drop after approximately three days. The X-ray data were collected at 100 K at ID14 EH4 at the European Synchrotron Radiation Facility (ESRF) at a wavelength of 0.933 Å. Reservoir solution containing 20% glycerol was used as cryo-protectant. The selected crystal of the V171E mutant (0.2 mm × 0.15 mm × 0.02 mm) diffracted to 1.7 Å resolution and belongs to the $P2_1$ space group with cell dimensions $a=69.01$ Å, $b=67.77$ Å, and $c=70.42$ Å, $\beta=119.4°$ ($V=286{,}926$ Å^3) and with two molecules in the asymmetric unit. The cell is slightly larger than for the cUNG ($V=274{,}324$ Å^3).[20]

Crystals of the hUNG E171V mutant were obtained by hanging-drop methods from a protein solution of 12.5 mg/ml prior to mixing with equal amounts (1.5 µl) of a reservoir solution made from 0.1 M sodium cacodylate buffer (pH 6.5), 25% (w/v) PEG 8000 and 0.1 M sodium acetate buffer (pH 4.9). X-ray data were collected to 1.95 Å resolution at the Swiss-Norwegian beam line, ESRF, at a wavelength of 0.933 Å. The crystals belong to space group $P2_12_12_1$ with cell dimensions $a=47.51$ Å, $b=54.77$ Å, and $c=78.07$ Å with one molecule in the asymmetric unit.

Both data sets were integrated using DENZO,[21] and scaling and merging was carried out using SCALA of the CCP4 program suite.[22] Due to the difference in cell parameters in cUNG V171 compared to the cUNG, the initial phases were obtained from molecular replacement searches in CNS[23] using cUNG as search model. The search gave two possible rotation-solutions, 180° apart, with the first peak 7.46 times higher than the mean of rotation function, and the other 5.26 times higher than peak number 3. These two peaks correspond to the two molecules in the asymmetric unit. There is 7 Å difference in the *c*-axis between the hUNG E171V and native hUNG[8] unit cells; thus, initial phases were obtained from molecular replacement searches also in this case.

Alternating cycles of model building using the program O[24] and refinement with CNS brought the final *R*-factor to 19.8% and R_{free} to 22.6% for cUNG V171E. The corresponding final *R*-factors for hUNG E171V were 21.0% and 25.4%. For both models, the refinement data were gradually extended to maximum resolution, and 5% of the reflections were kept out of refinement for cross validation.[25] Solvent molecules were added to the models where the difference density exceeded 4σ and within reasonable hydrogen bonding distances (3.4 Å). Details from the data collection and refinement characteristics are given in Table 3.

The quality of the final models is good. The electron density is generally well defined and about 90% of the main-chain dihedral angles are in the most favoured regions of the ϕ/ψ plot as judged from PROCHECK.[26]

Table 3. Data collection and refinement summary

	cUNG V171E	hUNG E171V
A. *Data collection*		
Resolution range (Å)	25–1.70	25–1.95
No. unique reflections	64,933	15,383
Redundancy (multiplicity)	3.2	2.3
R-merge (%)	4.6	13.2
Completeness (%)	94.3	93.2
Average $I/\sigma(I)$	9.7	4.0
B. *Refinement statistics*		
R-value (%)	19.82	20.98
Free *R*-value (%)	22.56	25.44
Deviation from ideal geometry		
Bond lengths (Å)	0.005	0.006
Bond angles (deg.)	1.189	1.233
Average *B*-values (Å^2)		
Main-chain atoms (number)	25.80 (1778)	15.41 (892)
Side-chain atoms (number)	27.99 (1800)	17.47 (914)
Glycerol (2)	38.30[a]	–
Chloride ion (2)	49.26[a]	–
Water molecules (number)	37.75 (365)	21.11 (126)
All atoms	28.01	16.76
Ramachandran plot (%)		
Most-favoured regions	89.4	89.9
Additionally allowed regions	10.1	10.1
Generously allowed regions	0.5[b]	0

[a] The values are calculated from the *B*-values of each atom.
[b] Phe158.

Acknowledgements

The Norwegian research council supported this work. We thank Eva Ryeng and Marit Sjo Lorentzen for technical assistance, and the organisers of the Swiss-Norwegian beam line (SNBL) and ID14 at the European Synchrotron Radiation Facility, Grenoble, France, for providing beam time.

References

1. Hochachka, P. W. & Somero, G. N. (1984). *Biochemical Adaptation*. Princeton University Press, Princeton, NJ.
2. Smalas, A. O., Leiros, H. K., Os, V. & Willassen, N. P. (2000). Cold adapted enzymes. *Biotechnol. Annu. Rev.* **6**, 1–57.
3. Gorfe, A. A., Brandsdal, B. O., Leiros, H. K., Helland, R. & Smalas, A. O. (2000). Electrostatics of mesophilic and psychrophilic trypsin isoenzymes: qualitative evaluation of electrostatic differences at the substrate binding site. *Proteins: Struct. Funct. Genet.* **40**, 207–217.
4. Brandsdal, B. O., Smalas, A. O. & Aqvist, J. (2001). Electrostatic effects play a central role in cold adaptation of trypsin. *FEBS Letters*, **499**, 171–175.
5. Kim, S. Y., Hwang, K. Y., Kim, S. H., Sung, H. C., Han,

ARTICLE IN PRESS

Y. S. & Cho, Y. (1999). Structural basis for cold adaptation. Sequence, biochemical properties, and crystal structure of malate dehydrogenase from a psychrophile *Aquaspirillium arcticum*. *J. Biol. Chem.* **274**, 11761–11767.
6. Russell, R. J., Gerike, U., Danson, M. J., Hough, D. W. & Taylor, G. L. (1998). Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure*, **6**, 351–361.
7. Lindahl, T. & Nyberg, B. (1974). Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry*, **13**, 3405–3410.
8. Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E. & Tainer, J. A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell*, **80**, 869–878.
9. Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995). The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature*, **373**, 487–493.
10. Ravishankar, R., Bidya Sagar, M., Roy, S., Purnapatre, K., Handa, P., Varshney, U. & Vijayan, M. (1998). X-ray analysis of a complex of *Escherichia coli* uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. *Nucl. Acids Res.* **26**, 4880–4887.
11. Leiros, I., Moe, E., Lanes, O., Smalas, A. O. & Willassen, N. P. (2003). The crystal structure of uracil-DNA *N*-glycosylase from Atlantic cod (*Gadus morhua*) reveals cold-adaptation features. *Acta Crystallog. sect. D*, **59**, 1357–1365.
12. Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E. & Tainer, J. A. (1998). Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **17**, 5214–5226.
13. Lanes, O., Leiros, I., Smalas, A. O. & Willassen, N. P. (2002). Identification, cloning, and expression of uracil-DNA glycosylase from Atlantic cod (*Gadus morhua*): characterization and homology modeling of the cold-active catalytic domain. *Extremophiles*, **6**, 73–86.
14. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
15. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
16. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51–59.
17. Lanes, O., Guddal, P. H., Gjellesvik, D. R. & Willassen, N. P. (2000). Purification and characterization of a cold-adapted uracil-DNA glycosylase from Atlantic cod (*Gadus morhua*). *Comp. Biochem. Physiol. B*, **127**, 399–410.
18. Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.
19. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
20. Leiros, I., Lanes, O., Sundheim, O., Helland, R., Smalas, A. O. & Willassen, N. P. (2001). Crystallization and preliminary X-ray diffraction analysis of a cold-adapted uracil-DNA glycosylase from Atlantic cod (*Gadus morhua*). *Acta Crystallog. sect. D*, **57**, 1706–1708.
21. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
22. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
23. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
24. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
25. Kleywegt, G. J. & Brunger, A. T. (1996). Checking your imagination: applications of the free *R* value. *Structure*, **4**, 897–904.
26. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.

Edited by R. Huber

(Received 6 May 2004; received in revised form 1 September 2004; accepted 3 September 2004)

Paper II

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 280, No. 18, Issue of May 6, pp. 18042–18048, 2005
Printed in U.S.A.
© 2005 by The American Society for Biochemistry and Molecular Biology, Inc.

Increased Flexibility as a Strategy for Cold Adaptation

A COMPARATIVE MOLECULAR DYNAMICS STUDY OF COLD- AND WARM-ACTIVE URACIL DNA GLYCOSYLASE*

Received for publication, January 26, 2005, and in revised form, March 3, 2005
Published, JBC Papers in Press, March 3, 2005, DOI 10.1074/jbc.M500948200

Magne Olufsen, Arne O. Smalås, Elin Moe, and Bjørn O. Brandsdal‡

From The Norwegian Structural Biology Centre, University of Tromsø, N-9037 Tromsø, Norway

Uracil DNA glycosylase (UDG) is a DNA repair enzyme in the base excision repair pathway and removes uracil from the DNA strand. Atlantic cod UDG (cUDG), which is a cold-adapted enzyme, has been found to be up to 10 times more catalytically active in the temperature range 15–37 °C as compared with the warm-active human counterpart. The increased catalytic activity of cold-adapted enzymes as compared with their mesophilic homologues are partly believed to be caused by an increase in the structural flexibility. However, no direct experimental evidence supports the proposal of increased flexibility of cold-adapted enzymes. We have used molecular dynamics simulations to gain insight into the structural flexibility of UDG. The results from these simulations show that an important loop involved in DNA recognition (the Leu^{272} loop) is the most flexible part of the cUDG structure and that the human counterpart has much lower flexibility in the Leu^{272} loop. The flexibility in this loop correlates well with the experimental k_{cat}/K_m values. Thus, the data presented here add strong support to the idea that flexibility plays a central role in adaptation to cold environments.

A vast amount of our planet consists of cold environments, like the Arctic, Antarctic, mountain regions, and deep sea waters, and cold-adapted organisms are able to breed and grow successfully in these regions. Organisms that are adapted to these extreme environments are often termed psychrophiles, whereas their warm-active counterparts are referred to as mesophiles and thermophiles. The temperature is one of the most important factors for enzyme activity, and enzymatic reaction rates can be reduced 30–80 times when the temperature decreases from 37 to 0 °C (1). To deal with this strong temperature dependence, cold-adapted enzymes usually have higher catalytic activity at low temperatures and decreased thermostability as compared with their mesophilic and thermophilic counterparts (1). Several strategies have been postulated to explain how enzymes adapt to cold environments, and the main theory, although not proven, has been that increased structural flexibility in psychrophilic enzymes enhances the catalytic efficiency (2). It is not necessarily that the elevated flexibility is attributed to an overall increase but may equally well be a result of increased flexibility for some of the structural elements. Improved flexibility in local areas seems to be a strategy for cold-adapted enzymes to maintain high catalytic activity at lower temperatures (3–5). Increased flexibility of psychrophilic enzymes would also explain the observed decrease in thermal stability (5). The observed increased catalytic activity at low temperatures and the decreased thermostability of psychrophilic enzymes suggest that there is a relationship between stability and activity to maintain the activity at low temperatures. The stabilizing intramolecular forces are consequently weakened, giving rise to a reduction in the thermal stability (6). The stability/flexibility relationship is controversial because cold-adapted organisms are under no selective pressure to stabilize proteins at elevated temperatures, and it is believed that the stability property has slowly vanished due to genetic drift (7).

Enzymes use different strategies to adopt to cold environments. Several cold-adapted enzymes studied so far seem to have low relative arginine content (Arg/(Arg+Lys)) as compared with the mesophilic and thermophilic enzymes (8). A decreased number of proline residues and an increased number of glycine residues are also thought to be a strategy that cold-adopted enzymes use (9). Another difference is that cold-adapted enzymes in some cases have a lower content of ion pairs, aromatic interactions, and/or hydrogen bonds (6). The above mentioned factors may give rise to increased local or overall molecular flexibility.

Although the flexibility hypothesis has been the dominating theory to explain the increased catalytic efficiency of cold-adapted enzymes, other adaptational strategies should not be ruled out. Optimization of the electrostatic potential is of great importance for the higher catalytic efficiency for cold-adapted enzymes (10). The increased binding affinity observed for the cold-adapted anionic salmon trypsin as compared with the mesophilic bovine trypsin is caused by optimization of electrostatic features (11). In the cod uracil DNA glycosylase (cUDG)[1], there are data supporting that the cold-adapted features observed could be caused by optimization of the electrostatic surface potential (12).

Uracil DNA glycosylase (UDG) is a DNA repair enzyme and is involved in the base excision repair pathway removing uracil from the DNA stand (13). This enzyme, which is the first in the base excision repair pathway, catalyzes the hydrolysis of promutagenic uracil residues from single- or double-stranded DNA. The crystal structure of the catalytic domain of UDG from several species are known: human (hUDG) (14), cod (cUDG) (15), virus 1 (16), and *Escherichia coli* (17). The three-dimensional structure of UDG in complex with DNA has also been determined (16, 18–21). Cold-adapted cUDG has been found to be up to 10 times more catalytic efficient (k_{cat}/K_m) in the temperature range 15–37 °C as compared with the human

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 47-77644057; Fax: 47-77644765; E-mail: Bjorn-Olav.Brandsdal@chem.uit.no.

[1] The abbreviations used are: UDG, uracil DNA glycosylase; cUDG, cod UDG; hUDG, human UDG; r.m.s.d., root mean square deviation; MD, molecular dynamics.

This paper is available on line at http://www.jbc.org

Downloaded from www.jbc.org at Universitetsbiblioteket i Trom on April 12, 2007

counterpart (22). cUDG is also found to be less stable than hUDG at all temperatures examined (23). The catalytic domain of hUDG and cUDG consists of 223 amino acids, and the sequence identity between them is 75%. The overall topology is a typical α/β protein (14). The loops involved in detection and catalysis of uracil are conserved in cUDG and hUDG, and the loops important to catalysis are: 4-Pro loop (^{165}PPPPS^{169}), the Gly-Ser loop (^{246}GS^{247}), the Leu^{272} loop (^{268}HPSPLSVYR^{276}), and the water-activating loop (^{145}DPYH^{148}) (20). The amino acids mentioned above are from hUDG, and there are two substitutions in the Leu^{272} loop in the cUDG structure: V274A and Y275H. The Leu^{272} loop is an important loop as it moves into the minor groove of the double-stranded DNA and flips out the uracil base. This movement is essential for bringing the catalytic important residue, His^{268}, within hydrogen-bonding distance of uracil (the O_2 atom) (20).

Visualization of the electrostatic potential of cUDG and hUDG indicated that the former has a more positively charged surface near the active site. The amino acids at positions 171 and 275 seem to be two key residues when explaining the difference in the surface potential (15, 24). Several mutants were subsequently constructed and analyzed in terms of kinetic, thermodynamic, and structural properties (12, 15, 24). In this study, molecular dynamics (MD) simulations have been performed for six UDG variants and were also repeated with DNA bound. The UDG variants included are: cUDG, cUDG-H275Y, cUDG-V171E, and cUDG-loop (A266T, V267A, A274V and H275Y) and hUDG and hUDG-E171V. It has previously been suggested that differences in catalytic activity could be explained by flexibility in the Leu^{272} loop area (15), and in this report, the relationship between the flexibility and catalytic activity is studied by means of MD simulations. The emerged picture is that cold-adapted enzymes have higher flexibility in the catalytic important areas of the structure (6). However, no direct experimental evidence supporting the hypothesis of increased flexibility of cold-adapted enzymes has yet been provided. MD simulations can provide insight into the dynamics of proteins and have been used to study the structural flexibility of UDG homologues.

MATERIALS AND METHODS

Models—Crystal structures were available for all of the UDG variants, except the cUDG-H275Y and the cUDG-loop mutant, and in the latter cases, models were made in the O program (25) from the cUDG structure. These structures were used as starting structures in the MD simulations. All the crystal structures were recombinant enzymes with three mutations in the N-terminal end: P82M, V83E, and G84F. The crystal structure of UDG in complex with double-stranded DNA (Protein Data Bank entry: 1EMH) (19), consisting of 19 bases, was used as a starting structure for the hUDG-DNA simulation and as a template to model the other complexes. This DNA originally had a 2′-deoxypseudouridine-5′monophosphate but was modeled into a 2′-deoxy-uracil-5′monophosphate by switching the places of the atoms: C2 ↔ C4, O_2 ↔ O4 and N1 ↔ C5 (Fig. 1). The latter is the uracil base that UDG recognizes and removes from the DNA strain.

UDG contains several histidines, and these were considered as neutral in the simulations, except for His^{148}. These choices were based on data from NMR and continuum electrostatics calculations (26). The hydrogen-bonding pattern of all histidines was analyzed to determine the position of the proton, and the proton was placed on the Nδ atom on the neutral histidines.

Molecular Dynamics—The AMBER program package (27) and the AMBER95 force field (28) were used to run and analyze all the MD simulations. The systems were energy-minimized by 2000 steps of steepest decent followed by 3000 steps using the conjugate gradients method. 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 50 ps. The time step used was 2 fs. In the production phase, the simulation was run with the isothermal-isobaric ensemble (300 K and one atmosphere pressure), and pressure and temperature were maintained by the Berendsen coupling algorithm (29). A 10 Å cutoff for non-bonded interactions was used in the simulation, and the Particle-Mesh Ewald method (30) was used to handle long range interactions. The SHAKE procedure (31) was applied to restrain covalent bonds involving all hydrogen atoms. During the production phase, the coordinates were written to file every ps. Crystallographic waters were kept, and in addition, water molecules were added, according to the TIP3P model (32), around the protein with a 15 Å buffer from the edge of the periodic box. The total box size for the system was ~85 × 70 × 75 Å with roughly 50,000 atoms in the system. The simulations were run for 1 ns for each mutant except for native cUDG and hUDG, which were run for 2 ns. Average structures were calculated from the snapshots during the simulation. To examine the stability of the protein during the simulation, the root mean square deviation (r.m.s.d.) of Cα atoms *versus* simulation time was calculated. r.m.s.d. based on main chain atoms has also been calculated, but these values were very similar to the Cα r.m.s.d., and the Cα r.m.s.d. is therefore used throughout this text.

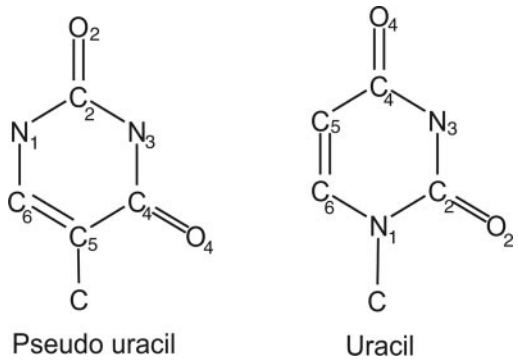


FIG. 1. **Structural differences between uracil and pseudo uracil.**

RESULTS AND DISCUSSION

Thermodynamic and kinetic characterizations of cUDG and hUDG have shown that the former is more catalytic efficient and less thermostable as compared with the latter. To gain insight into the features responsible for these properties, several mutants of cUDG and hUDG have been constructed and subsequently investigated by combining biochemical analysis with three-dimensional structure analysis (12, 15, 24). To gain further insights into the role that molecular flexibility plays in enzymatic adaptation to cold environments, six different UDG variants have been studied by MD simulations. The six different variants are: cUDG, cUDG-V171E, cUDG-H275Y, cUDG-loop mutant (A266T, V267A, A274V, and H275Y), hUDG, and hUDG-E171V. Fig. 2 shows cUDG bound to the DNA fragment with important residues highlighted. The variants studied have been mutated from the amino acid in native cUDG to the amino acid in native hUDG and vice versa. All of them have previously been characterized in terms of kinetics and thermodynamics, and K_m and k_{cat} values are available for each of them (Table I) (12).[2]

Analysis of the crystal structure of hUDG and hUDG in complex with DNA has shown that the Leu^{272} loop moves when the enzyme binds to the DNA strand. This movement is essential for bringing the catalytic His^{268} into hydrogen-bonding distance of the uracil O_2. Residues 275 and 276 in the Leu^{272} loop are involved in DNA repair recognition (20), and it is therefore reasonable to expect that this loop is very important for catalytic activity.

Overall Features of the MD Simulations—To assess the stability of the simulations, the r.m.s.d. has been calculated relative to the average structure (calculated from the coordinates from the production phase) and is presented in Fig. 3*A* for native cUDG and hUDG. Fig. 3*A* shows that the r.m.s.d. for the

[2] E. Moe, I. Leiros, K. Torseth, O. Lanes, A. O. Smalås, and N. P. Willassen, unpublished results.

Downloaded from www.jbc.org at Universitetsbiblioteket i Trom on April 12, 2007

FIG. 2. **A ribbon structure of the cUDG-DNA model used in the MD simulation.** Some residues are drawn in a *ball-and-stick*. The program DSSP (34) was used to find the secondary structure, and the figure was generated by Molscript (35) and Raster3D (33). *C-term*, C terminus.

TABLE I
Average r.m.s.d of Cα atoms after 1 ns of MD simulation and kinetic constants determined at 22 °C for cUDG, hUDG, and four mutants (Ref. 12 and Footnote 2)

UDG	r.ms.d.		k_{cat}/K_m	k_{cat}	K_m
	Average	Residues 268–276			
	Å		$min^{-1}\mu M^{-1}$	min^{-1}	μM
cUDG	0.62	1.38	892	677	0.8
cUDG-DNA	0.58	0.62			
cUDG-loop	0.58	1.04	548	263	0.5
cUDG-loop-DNA	0.52	0.50			
cUDG-H275Y	0.59	0.96	509	302	0.6
cUDG-H275Y-DNA	0.54	0.50			
hUDG-E171V	0.57	0.85	396	262	0.7
hUDG-E171V-DNA	0.48	0.44			
hUDG	0.55	0.97	135	320	2.4
hUDG-DNA	0.50	0.42			
cUDG-V171E	0.56	0.67	117	204	1.7
cUDG-V171E-DNA	0.49	0.54			

cold-adapted cUDG is between 0.5 and 1.3 Å during the simulation, whereas the hUDG simulations display lower overall r.m.s.d. In all 12 MD simulations, the total r.m.s.d. is stable through the simulation (results not shown).

There are only small differences in the average r.m.s.d. among the different UDG variants (Table I), calculated for 1000 MD structures. The UDG simulations without DNA have an r.m.s.d. between 0.56 and 0.62 Å, whereas the corresponding values for the simulations with DNA are between 0.48 and 0.58 Å (Table I). The cold-adapted cUDG has the highest average r.m.s.d., both with and without DNA bound, whereas the warm-active hUDG-E171V and hUDG-DNA display the lowest r.m.s.d. It is interesting to note that cUDG has the highest r.m.s.d. both when DNA is present and when DNA is absent as this indicates that the cold-active UDG has higher overall flexibility as compared with its warm-active counterpart hUDG. This is to our knowledge the first time MD simulations show that a cold-adapted enzyme has increased overall molecular flexibility as compared with its warm-active counterpart. However, based on this, one cannot conclude that the increased flexibility is the basis for the enhanced catalytic activity as observed experimentally, although strong evidence is presented. Comparison of the r.m.s.d. for the DNA recognition loop (Leu^{272} loop, residues 268–276) shows that there are large differences among the UDG variants, and this issue will be discussed in the next sections.

Comparison of Human and Cod UDG—To further investigate the details of the MD simulations, the average r.m.s.d.

Downloaded from www.jbc.org at Universitetsbiblioteket i Trom on April 12, 2007

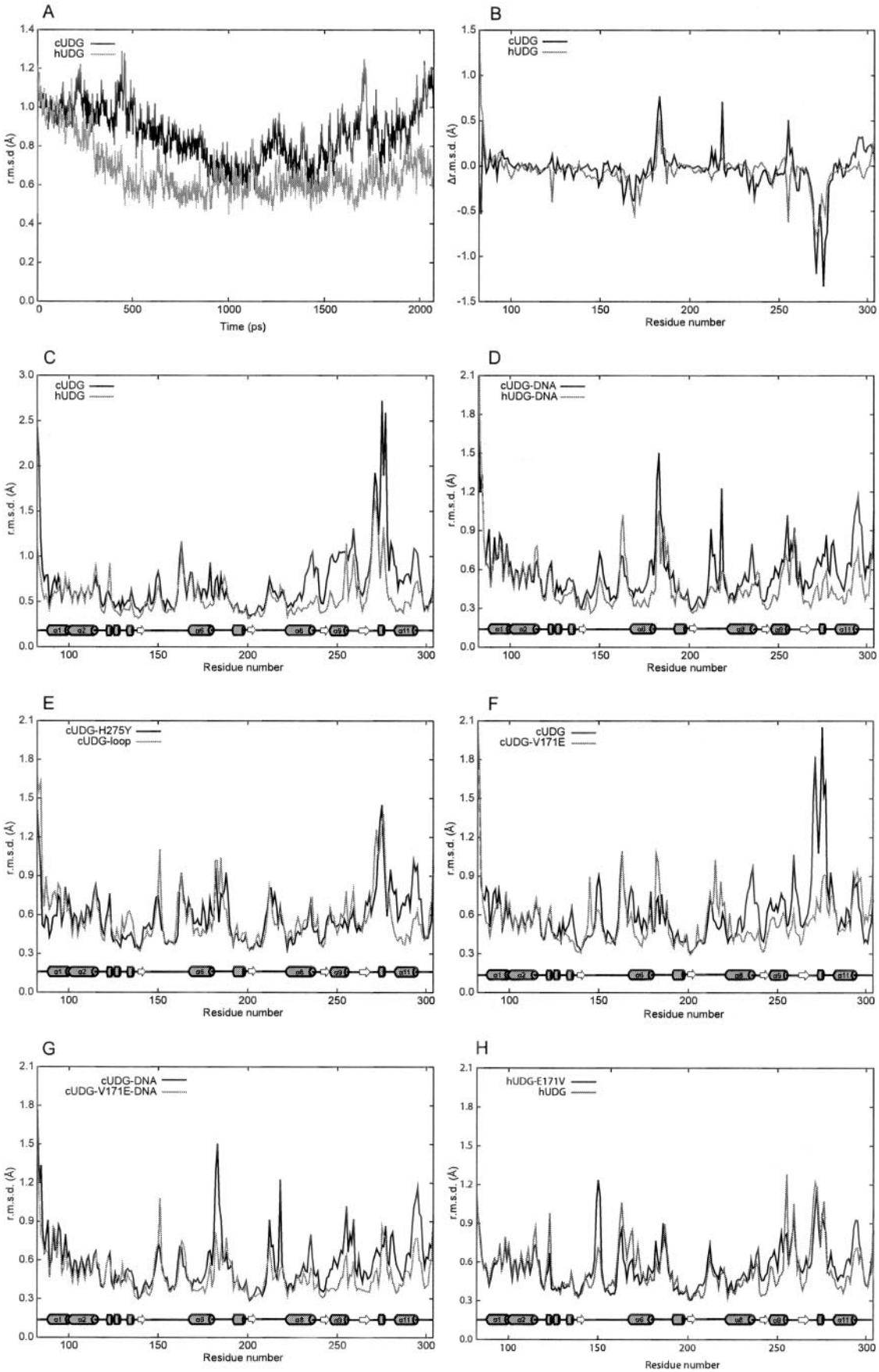


FIG. 3. **Various r.m.s.d. plots from the MD simulations.** *A*, a plot of the r.m.s.d. as a function of time for the simulation of hUDG and cUDG, calculated based on all Cα atoms in the average structure and the snapshot during the simulation. *B*, Δr.m.s.d. ($r.m.s.d._{UDG\text{-}DNA} - r.m.s.d._{UDG}$) per residue of cUDG and hUDG. *C*, r.m.s.d. of cUDG and hUDG calculated from 2 ns of MD simulation at 300 K. *D–H*, r.m.s.d. calculated from 1 ns of MD simulation at 300 K.

Downloaded from www.jbc.org at Universitetsbiblioteket i Trom on April 12, 2007

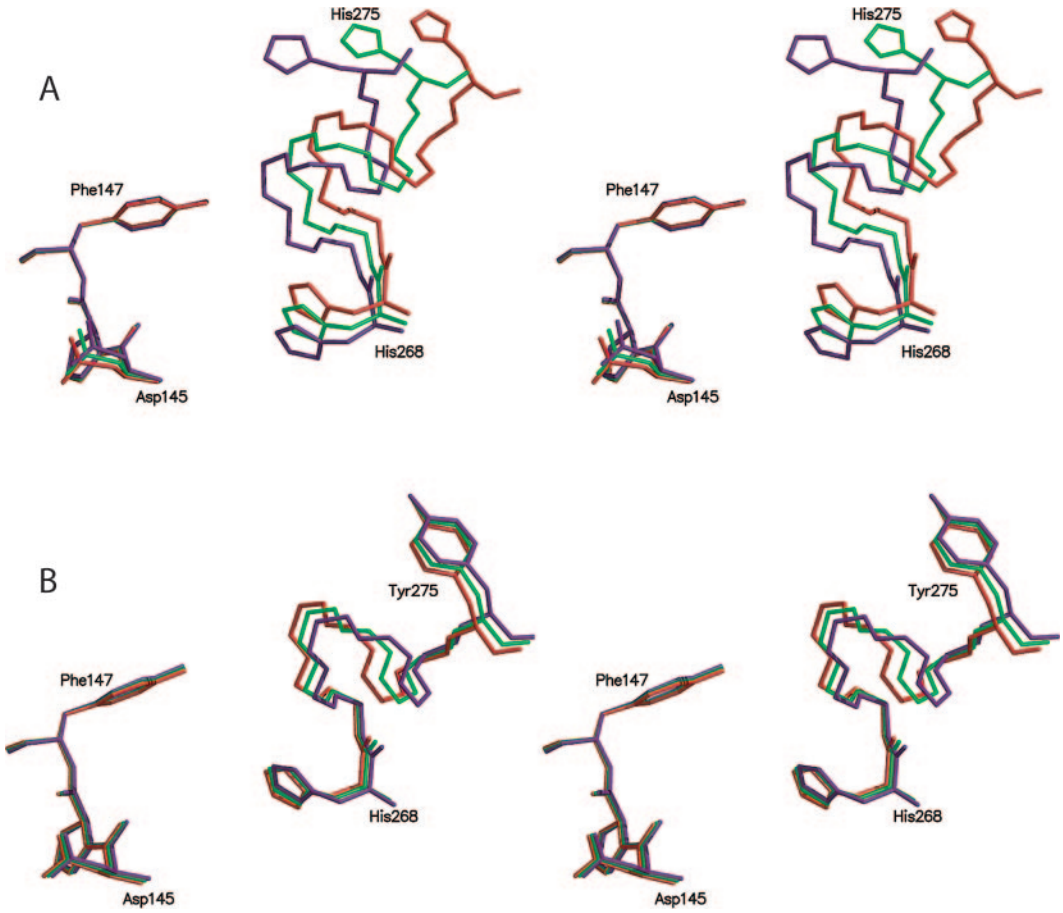


FIG. 4. **Movement of the DNA binding loop from the cUDG and hUDG simulations.** Average structures of the cUDG Leu^{272} loop (*A*) and hUDG Leu^{272} loop (*B*) from 0 to 1 ns (*red*), 0 to 2 ns (*green*), and 1 to 2 (*blue*) ns simulation has been superimposed. From residues 269 to 274, only main chain atoms are drawn, and residue 276 is not shown. The figures were generated by Molscript (35) and Raster3D (33).

values/residue for the simulations have been calculated. Fig. 3*C* shows a comparison of the average r.m.s.d. for the Cα atoms of cUDG and hUDG. Several large differences are observed in the C-terminal end of the protein, especially in the Leu^{272} loop. The residue with highest r.m.s.d. is also found in this region, which is His^{275} in the cUDG simulations and Pro^{271} in hUDG. The Gly-Ser loop (^{246}GS^{247}) is also more flexible for the cod variant with an increased r.m.s.d. of 0.41 Å relative to the human counterpart. Three average structures have been calculated from these two simulations based on the coordinates from 0 to 1, 0 to 2, and 1 to 2 ns of cUDG and hUDG. Analysis of these structures reveals that the Leu^{272} loop of the cod structure moves throughout the simulation (Fig. 4*A*). In contrast, the average structures from the hUDG simulation are much more similar, and consequently, less movement is observed in this loop region (Fig. 4*B*). For the cUDG simulation, the Cα atom in the His^{275} residue moves 5.47 Å from the initial to the final structure, whereas the same atom moves 4.83 Å between the average structure from 0 to 1 ns and the average structure from 1 to 2 ns simulation. During the hUDG simulation, this movement is not observed (Fig. 4, *A* and *B*). The crystal structures of cUDG and hUDG have approximately the same number of charged residues, but the cUDG structure was found to have slightly more hydrogen bonds (15). One should believe that the slight increase in hydrogen bonds in the cUDG structure would make the cUDG structure more rigid than the hUDG structure, but this does not seem to be the case according to the MD simulations.

In the simulations of cUDG and hUDG in complex with DNA, the structure of the former is again more flexible than the human counterpart, especially in the C-terminal part of the protein (Fig. 3*D*). The average residual r.m.s.d. decreases only by 0.04 and 0.05 Å between the unbound UDG simulation and the complex simulation of the cUDG and the hUDG, respectively. We have also calculated the Δr.m.s.d. (r.m.s.d.$_{\text{UDG-DNA}}$ − r.m.s.d.$_{\text{UDG}}$) to investigate the effect of DNA on the molecular flexibility. Fig. 3*B* shows that most of the flexibility is lost in the Leu^{272} loop area for both the cUDG and the hUDG structure. The decrease in flexibility in the Leu^{272} loop upon DNA binding is observed for all the different UDG variants. This loop is in close contact with the DNA, and it therefore seems reasonable that the r.m.s.d. in this area will decrease when DNA is included in the simulations. Around residue 183 (Gly for cUDG and Asp for hUDG), both structures have an increase in r.m.s.d. for the simulation of the complex; residue 183 is not close to the DNA or the DNA binding area but is situated at the outside of the protein. In the average structure of hUDG and hUDG-DNA, the Asp^{183} side chain forms a salt bridge to Lys^{302}, and the distance between the Nζ atom and the Oδ1 and Oδ2 is 2.76 and 3.36 Å in the hUDG and 2.93 and 3.25 Å in the hUDG-DNA structure. In the average structure of cUDG and cUDG-DNA, Gly^{183} does not form any interactions. Therefore, it is difficult to explain this increase in mobility for residue 183 in the complex simulation, but DNA is highly charged and may therefore influence all parts of the protein through long range interactions and changes in electrostatic potentials. This could affect the flexibility of the protein.

Analysis of the cUDG-Loop and the cUDG-H275Y Simulations—The cUDG-loop mutant (A266T, V267A, A274V, and H275Y) and the cUDG-H275Y mutant have similar k_{cat}/K_m values, 548 and 509 min$^{-1}\mu$M^{-1}, respectively, whereas the

Downloaded from www.jbc.org at Universitetsbiblioteket i Trom on April 12, 2007

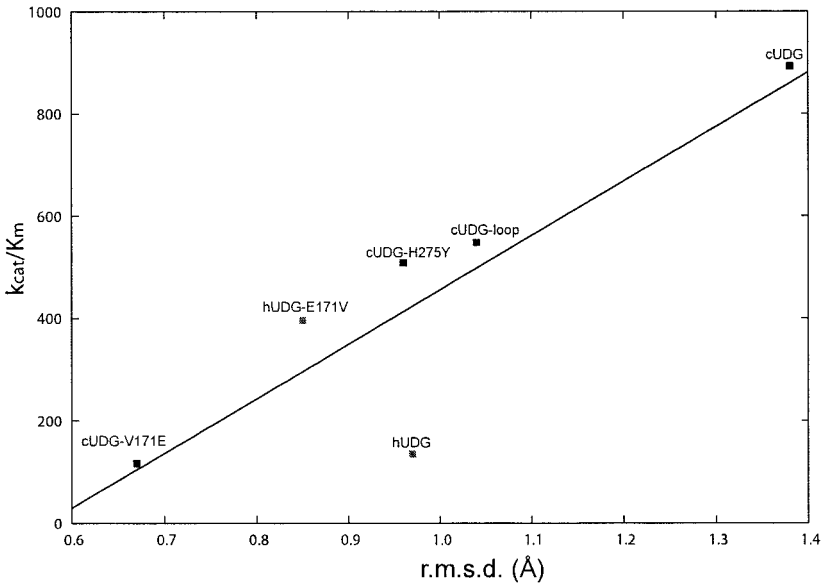


FIG. 5. **Plot of k_{cat}/K_m (catalytic efficiency) *versus* Cα r.m.s.d. of the Leu^{272} loop for the different UDG variants.**

native cod enzyme has a k_{cat}/K_m value of 892 $min^{-1}\mu M^{-1}$ (12). The average Cα r.m.s.d. values of the Leu^{272} loop are 1.04 and 0.96 Å for the cUDG-loop and cUDG-H275Y simulations, respectively, whereas the corresponding value for cUDG is 1.38 Å. Thus, the catalytic efficiency decreases with these substitutions, and there appears to be a concomitant decrease in the flexibility of the Leu^{272} loop for the MD simulations. It therefore seems likely that there is a close connection between the flexibility of the Leu^{272} loop and catalytic efficiency.

The nature of the amino acid at position 275 seems to be a key factor to explain the flexibility of the Leu^{272} loop. It is reasonable to expect that the difference in the side chain volume, size and charge can reduce the flexibility of this residue and subsequently the loop for hUDG. It is found that when mutating H275Y in cUDG, the r.m.s.d. for the Cα atom of residue 275 is reduced by 0.60 Å as compared with the cUDG simulation (Fig. 3, *C* and *E*). If we consider all the residues in the Leu^{272} loop, the r.m.s.d. of the cUDG-loop and the cUDG-H275Y simulations is virtually identical to that observed for hUDG (Fig. 3, *C* and *E*), which also indicates that the His to Tyr mutation is responsible for the decreased flexibility of this loop. It has been suggested that Val^{274}, Tyr^{275}, and Phe^{279} residues in the hUDG structure form a hydrophobic cluster and that this cluster will restrict the motion of the Leu^{272} loop. All three residues are substituted to smaller ones in the cUDG structure (Ala^{274}, His^{275}, and Leu^{279}) (15). The Cα r.m.s.d. of Pro^{271} decreases by 1.09 Å for the cUDG-H275Y average structure as compared with the average structure from the cUDG simulation. The carbonyl oxygen of residue Pro^{271} of cUDG-H275Y is involved in two hydrogen bonds to main chain nitrogens in residues 273 and 274 with a distance of 3.39 and 3.23 Å, respectively, whereas in the average structure of cUDG, the distances between the same atoms are 3.67 and 3.85 Å, respectively. This difference may explain why the native cUDG structure has a more flexible Pro^{271} residue.

Analysis of the cUDG-V171E and the hUDG-E171V Simulations—The residue at position 171 is important for the cold-adapted features of cUDG as k_{cat}/K_m decreases by almost a factor of 8 when Val^{171} is mutated to Glu, which is the amino acid in the mesophilic human sequence. The cUDG-V171E mutant also shows increased temperature stability at 37 °C as compared with cUDG (12). The Cα r.m.s.d. plot of the cUDG-V171E structure is quite different from the cUDG structure, especially in the Leu^{272} loop where the difference in r.m.s.d. is 0.71 Å (Fig. 3*F*). It is interesting to note that the mutation of the 171 residue does not have any effect on the flexibility of the nearby residues but does have a large effect on the residues in the Leu^{272} area. Moe *et al.* (12) have shown that the electrostatic environment near the catalytic region of the enzyme is of vital importance to the catalytic efficiency. Thus, it is likely that the change in electrostatic environment could explain why the 171 substitution has a large effect on the flexibility of the Leu^{272} loop. The r.m.s.d. values of both cUDG-V171E and cUDG in complex with DNA (Fig. 3*G*) show that the cUDG has higher flexibility in the Leu^{272} loop, but the largest differences in r.m.s.d. values between the two structures are for residues 182–185. The average structure of the cUDG-V171E-DNA has a weak salt bridge between the side chain of Glu^{171} and His^{186} (4.40 Å), which can explain the decrease in r.m.s.d. in the sequence 182–185.

The hUDG-E171V mutant has a 3-fold increased catalytic efficiency and also decreased temperature stability as compared with the native human enzyme (12); thus, substitution from an amino acid found in the warm-active enzyme to an amino acid in the cold-adapted cUDG makes the enzyme adopt the typical cold-adapted features. The hUDG-E171V mutant also has a lower flexibility in the Leu^{272} loop as compared with hUDG (Fig. 3*H*), whereas in the complex simulation, more or less the same flexibility is observed.

The cUDG-V171E variant has the lowest r.m.s.d. value in the Leu^{272} loop and also displays the lowest catalytic efficiency in the study, whereas both the k_{cat}/K_m and the r.m.s.d. of the Leu^{272} loop for the hUDG-E171V variant are in the middle between the endpoints drawn by native cUDG and cUDG-V171E. These results fit well with the idea that there is a correlation between catalytic efficiency and flexibility of this loop.

Concluding Remarks—MD simulations of UDG have been used to predict the flexible parts of the enzyme, and the Leu^{272} loop is the most mobile part of the structure for all of the UDG variants included in our study, except for the cUDG-V171E mutant, which has higher flexibility in other parts of the structure (Fig. 3, *C* and *E*–G). In addition, these simulations show that the C-terminal end of cUDG enzyme is more flexible as compared with the human counterpart, as postulated previously (22). The His^{275} residue in cUDG is probably the main contributor to the increased flexibility of the Leu^{272} loop, relative to the human enzyme. The size, volume, and charge of this residue can probably explain the increased flexibility of the Leu^{272} loop of the cod enzyme. Another interesting aspect is

Downloaded from www.jbc.org at Universitetsbiblioteket i Trom on April 12, 2007

that the r.m.s.d. of the Leu^{272} loop correlates well with the k_{cat}/K_m values as cUDG has by far the highest k_{cat}/K_m value and it also has by far the highest flexibility in the Leu^{272} loop, whereas the cUDG-V171E mutant has the lowest r.m.s.d. in the Leu^{272} loop and also the lowest k_{cat}/K_m of the mutants in this study. Five of the six r.m.s.d. values of the Leu^{272} loop from the UDG correlate with the k_{cat}/K_m values (Fig. 5); the correlation factor is 86.4% for all the variants and 99.0% if the outlier hUDG is excluded. In the UDG-DNA simulation, there is less difference in r.m.s.d. among the variants, but still, cUDG-DNA has the highest r.m.s.d. value, and the warm-active variant has the lowest r.m.s.d. values of the Leu^{272} loop. This verifies the hypothesis that states that cold-adapted enzymes are characterized by an improved flexibility of the structural components involved in the catalytic cycle (6) for UDG. By analysis of the activation parameters of cold-adapted enzymes, it has been shown that local flexibility is essential for high activity at low temperatures (1), and these findings fit well with the suggested link between catalytic efficiency and flexibility of the DNA recognition loop of UDG.

REFERENCES

1. Lonhienne, T., Gerday, C., and Feller, G. (2000) *Biochim. Biophys. Acta* **1543,** 1–10
2. Feller, G., and Gerday, C. (1997) *CMLS Cell. Mol. Life Sci.* **53,** 830–841
3. Fields, P. A., and Somero, G. N. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 11476–11481
4. Hochachka, P. W., and Somero, G. N. (1984) *Biochemical Adaptations*, pp. 355–449, Princeton University Press, Princeton, NJ
5. Georlette, D., Damien, B., Blaise, V., Depiereux, E., Uversky, V. N., Gerday, C., and Feller, G. (2003) *J. Biol. Chem.* **278,** 37015–37023
6. Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, J. C., Sonan, G., Feller, G., and Gerday, C. (2004) *FEMS Microbiol. Rev.* **28,** 25–42
7. Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N., and Arnold, F. H. (2000) *J. Mol. Biol.* **297,** 1015–1026
8. Smalås, A. O., Leiros, H. K. S., Os, V., and Willassen, N. P. (2000) *Biotechnology Annual Review*, Vol. 6, pp. 1–57, Elsevier Science B.V., Amsterdam
9. Gerday, C., Aittaleb, M., Bentahir, M., Chessa, J. P., Claverie, P., Collins, T., D'Amico, S., Dumont, J., Garsoux, G., Georlette, D., Hoyoux, A., Lonhienne, T., Meuwis, M. A., and Feller, G. (2000) *Trends Biotechnol.* **18,** 103–107
10. Russell, R. J. M., Gerike, U., Danson, M. J., Hough, D. W., and Taylor, G. L. (1998) *Structure* **6,** 351–361
11. Brandsdal, B. O., Smalås, A. O., and Åqvist, J. (2001) *FEBS Lett.* **499,** 171–175
12. Moe, E., Leiros, I., Riise, E. K., Olufsen, M., Lanes, O., Smalas, A., and Willassen, N. P. (2004) *J. Mol. Biol.* **343,** 1221–1230
13. Lindahl, T., and Nyberg, B. (1974) *Biochemistry* **13,** 3405–3410
14. Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., and Tainer, J. A. (1995) *Cell* **80,** 869–878
15. Leiros, I., Moe, E., Lanes, O., Smalås, A. O., and Willassen, N. P. (2003) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59,** 1357–1365
16. Savva, R., Mcauleyhecht, K., Brown, T., and Pearl, L. (1995) *Nature* **373,** 487–493
17. Ravishankar, R., Sagar, M. B., Roy, S., Purnapatre, K., Handa, P., Varshney, U., and Vijayan, M. (1998) *Nucleic Acids Res.* **26,** 4880–4887
18. Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. (1996) *Nature* **384,** 87–92
19. Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., and Tainer, J. A. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5083–5088
20. Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., and Tainer, J. A. (1998) *EMBO J.* **17,** 5214–5226
21. Bianchet, M. A., Seiple, L. A., Jiang, Y. L., Ichikawa, Y., Amzel, L. M., and Stivers, J. T. (2003) *Biochemistry* **42,** 12455–12460
22. Lanes, O., Leiros, I., Smalås, A. O., and Willassen, N. P. (2002) *Extremophiles* **6,** 73–86
23. Lanes, O., Guddal, P. H., Gjellesvik, D. R., and Willassen, N. P. (2000) *Comp. Biochem. Physiol.* **127,** 399–410
24. Leiros, I., Lanes, O., Sundheim, O., Helland, R., Smalås, A. O., and Willassen, N. P. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57,** 1706–1708
25. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
26. Dinner, A. R., Blackburn, G. M., and Karplus, M. (2001) *Nature* **413,** 752–755
27. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) *Comp. Phys. Commun.* **91,** 1–41
28. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc.* **117,** 5179–5197
29. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) *J. Chem. Phys.* **81,** 3684–3690
30. Darden, T., York, D., and Pedersen, L. (1993) *J. Chem. Phys.* **98,** 10089–10092
31. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) *J. Comp. Phys.* **23,** 327–341
32. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) *J. Chem. Phys.* **79,** 926–935
33. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277,**
34. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22,** 2577–2637
35. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950

Downloaded from www.jbc.org at Universitetsbiblioteket i Trom on April 12, 2007

Paper III

Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling xxx (2006) xxx–xxx
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Comparative unfolding studies of psychrophilic and mesophilic uracil DNA glycosylase: MD simulations show reduced thermal stability of the cold-adapted enzyme

Magne Olufsen, Bjørn Olav Brandsdal, Arne Oskar Smalås *

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

Received 14 June 2006; received in revised form 17 October 2006; accepted 18 October 2006

Abstract

Uracil DNA glycosylase (UDG) is a DNA repair enzyme involved in the base excision repair (BER) pathway, removing misincorporated uracil from the DNA strand. The native and mutant forms of Atlantic cod and human UDG have previously been characterized in terms of kinetic and thermodynamic properties as well as the determination of several crystal structures. This data shows that the cold-adapted enzyme is more catalytically efficient but at the same time less resistant to heat compared to its warm-active counterpart. In this study, the structure–function relationship is further explored by means of comparative molecular dynamics (MD) simulations at three different temperatures (375, 400 and 425 K) to gain a deeper insight into the structural features responsible for the reduced thermostability of the cold-active enzyme. The simulations show that there are distinct structural differences in the unfolding pathway between the two homologues, particularly evident in the N- and C-terminals. Distortion of the mesophilic enzyme is initiated simultaneously in the N- and C-terminal, while the C-terminal part plays a key role for the stability of the psychrophilic enzyme. The simulations also show that at certain temperatures the cold-adapted enzyme unfolds faster than the warm-active homologues in accordance with the lower thermal stability found experimentally.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Molecular dynamics; Uracil DNA glycosylase; Protein unfolding; Cold adaptation; Psychrophilic

1. Introduction

Several cold-adapted enzymes from different organisms living in constantly cold environments have been identified and characterized in the last few decades. Such enzymes possess unique properties, attractive not only for the academic research but also to the biotechnological market. The emerging picture suggests that cold-adapted enzymes have higher catalytic efficiency, lower thermostability and improved flexibility in the structural parts directly involved in the catalytic cycle [1,2]. Indeed, most psychrophilic enzymes studied so far display reduced overall stability when compared to their mesophilic counterparts. The increased catalytic efficiency is usually accompanied by a decrease in thermal stability, which is believed to be a consequence of higher structural flexibility in the psychrophilic enzymes. When the temperature decreases from 37 to 0 °C the reaction rates are generally reduced by 30- to 80-fold [2], and in order to maintain a sufficient metabolic flux at low temperatures it has been suggested that psychrophilic enzymes have higher flexibility in the structural parts directly involved in the catalytic cycle [1]. However, enzymes are shown to use different structural adaptation strategies to achieve the above-mentioned features [3], and psychrophilic enzymes often possess decreased number of proline residues, increased number of glycine residues, low relative arginine content [arg/(arg + lys)] and lower numbers of ion pairs, aromatic interactions and/or hydrogen bonds compared to the mesophilic counterparts [3–5].

Cold-adapted uracil DNA glycosylase from cod (cUDG) has been shown to be up to 10 times more catalytic efficient (k_{cat}/K_m) in the temperature range from 15 to 37 °C compared to the warm-active human counterpart (hUDG) [6]. UDG is the first enzyme in the BER pathway, and catalyzes the hydrolysis of promutagenic uracil residues from single- or double-stranded DNA [7]. The crystal structure of the catalytic domain of UDG from several species are known: human (hUDG) [8], Atlantic

* Corresponding author. Tel.: +47 77644070; fax: +47 77644765.
E-mail address: Arne.Smalas@chem.uit.no (A.O. Smalås).

1093-3263/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2006.10.003

cod (cUDG) [9], virus 1 [10] and *Escherichia coli* [11]. The catalytic domain of hUDG and cUDG consists of 223 amino acids with a sequence identity of 75%, and the overall topology is a typical α/β protein [8]. The r.m.s.d. between the two crystal structures is 0.492 Å (all atoms were included). The higher catalytic efficiency of cUDG compared to hUDG has been postulated to arise from a more flexible DNA-recognition loop [12] and optimization of the electrostatic surface potential near the specificity pocket [9] in the cold-adapted enzyme.

cUDG has also been found to be much more pH and temperature labile than hUDG [13], indicating that cUDG is generally less stable compared to hUDG. Large opposing energies are involved when proteins go from the unfolded to the folded state, but the overall change in energy is small, typically ranging from 1 to 15 kcal/mol [14]. All contributions, both favorable and unfavorable, are therefore important when considering protein stability [14]. How proteins achieve their stability does not seem to follow any general rules [15], but the emerging picture is, for globular proteins, that the hydrophobic effect and burial of non-polar side chains stabilizes the native state [16,17]. Disulfide bonds, electrostatic interactions and hydrogen bonds are, however, also important for structural stability and can contribute favorably to protein stability [18–20]. The majority of reversible two-state folding proteins, including thermophilic, mesophilic and psychrophilic enzymes, show maximum stability around room temperature [17].

High temperature MD simulations have been shown to provide valuable insights into the unfolding process of several proteins: chymotrypsin inhibitor 2 (CI2) [21], barnase [22,23] and engrailed homeodomain (En-HD) [24]. Protein unfolding studies using MD simulations are carried out by increasing the temperature necessary to overcome the enthalpic forces stabilizing the 3D structure. Because of the differences in accessible time scales by experiments and computer studies, higher temperatures are needed when performing computer experiments relative to that observed in experimental unfolding studies. A MD study of CI2 has shown that the unfolding pathway is not widely affected by the simulation temperature, and thus raising the temperature only increases the reaction rate [25].

In this study, we have employed MD simulations to investigate the structure–function relationship further with focus on possible differences in the thermal unfolding pathway between a psychrophilic and a mesophilic UDG. We have especially analyzed the structures to see if there are important molecular contacts, which could explain the observed difference in thermal stability between these enzymes. Comparative high temperature (375, 400, 425 K) simulations for 6 and 9 ns have been carried out for the two enzymes, and the unfolding process has been carefully monitored both at an overall and a detailed structural level.

2. Methods

All MD simulations were carried out using the AMBER7 program package [26] with the recently developed force field parm99 [27]. The generalized Born methodology [28] was applied to describe the electrostatics from the solvent. Implicit solvent has been chosen in this study. The advantage of using implicit solvent models is the increase in computational efficiency. MD simulations with implicit solvent have successfully been used to investigate protein folding [29,30], and recent development in the implicit solvent methodology now provide a level of realism that challenge the results from the explicit solvent simulations [31]. The simulations were initiated with 100 cycles of energy minimization and the protein atoms constrained with a 5 kcal/mol $Å^2$ force constant. The temperature was slowly raised to the final temperature (375, 400 or 425 K) during a heating phase of approximately 20 ps. In this phase, the protein atoms were weakly restrained to their initial positions. SHAKE [32] was employed to constrain all bonds involving hydrogen atoms. A cutoff of 15 Å was used for the non-bonded interactions and every 1000 step the rotational and translational motion was removed. The time step was set to 2 fs, and the temperature was maintained by the Berendsen coupling algorithm [33]. The hydrophobic effect was taken into account by adding a surface energy term to the total energy [34].

Crystal structures were used as starting structures in the MD simulations for both hUDG [8] and cUDG [35] (Protein Data Bank entries 1AKZ and 1OKB, respectively). The crystal structures were recombinant enzymes with three mutations in the N-terminal end: P82M, V83E and G84F. UDG contains several histidines, and in the simulation the His 148 was charged, while the rest of the histidines were considered as neutral with the hydrogen atom placed on the Nε atom. These choices were based on data from NMR and continuum electrostatics calculations [36].

Hydrogen bonds and r.m.s.d. values of the Cα atoms were calculated with the Amber program package. In the hydrogen bond calculations a distance cutoff of 3.4 Å and an angle cutoff of 60° (the angle between: hydrogen-donor–acceptor) were applied. These calculations were also tested with no angle cutoff, but these calculations gave similar results, and the angle cutoff method was used in this study. The DSSP program [37] was used to find the secondary structure elements of each snapshot. The secondary structure elements were calculated for snapshots every 10th ps and for smoothening the plots, the number of residues in secondary elements was averaged every 10 snapshots. The surface area was calculated by the MSMS program [38], and a probe radius of 1.5 Å was applied to calculate the solvent accessible surface area (SASA).

3. Results and discussion

Organisms have adapted to extreme environments and survive at different temperatures by optimization of their proteins at some level and thereby maintain sufficient metabolic fluxes. Increased availability of sequences, kinetic, thermodynamic and structural information of proteins from psychrophilic, mesophilic and thermophilic enzymes have facilitated a deeper understanding of features specific for adaptation at the molecular level. Presently, much deeper insight has been

Please cite this article in press as: M. Olufsen et al., Comparative unfolding studies of psychrophilic and mesophilic uracil DNA glycosylase: MD simulations show reduced thermal stability of the cold-adapted enzyme, J. Mol. Graph. Model. (2006), doi:10.1016/j.jmgm.2006.10.003

achieved on how proteins have adapted to cope with high temperatures without denaturation, as compared to the low temperature adaptation. Higher temperatures involve increased reaction rates and atomic mobility, while low temperature slows down the chemical reaction rates and reduced solvent viscosity. Hence, it may appear that adaptation to high and low temperature requires opposite changes, but several studies have shown that the picture is far more complicated than simply reversing the features responsible for high temperature adaptation when explaining how the challenges are met at low temperature.

Uracil DNA glycosylase is an excellent model system for the study of the mechanisms of enzymatic adaptation to low temperature. Its biological function is well-known, and several crystal structures of both cUDG and hUDG are available, including mutant and native structures, which have also been characterized in terms of stability and kinetics [6,8,9,13,35,39]. It is well-established that cold-adapted enzymes are less resistant to heat as compared to their warm-active homologues and this has also been shown to be the case for cUDG [13]. In light of the small differences revealed by structural analysis, computational simulation of the thermal unfolding process was a natural next step.

In order to gain further insights into the thermal stability of cUDG and hUDG eight comparative MD simulations at three different temperatures (375, 400 and 425 K) have been carried out. Molecular interactions and intramolecular contacts (like salt-bridges, hydrogen bonds and hydrophobic contacts) have been thoroughly monitored in order to search for specific features and structurally important regions that could explain the difference in thermal stability. Even higher temperatures were tested, but increasing the temperature beyond 425 K yielded complete loss of ordered structure within just a few ps. The resulting trajectories were too difficult to analyze as most of the sampling were on the fully unfolded polypeptide chains.

3.1. Structural fluctuation in the MD simulations

Computer aided thermal unfolding of proteins are usually performed at temperatures between 200 and 225 °C (473–498 K) [40], and a commonly used criterion for an unfolded state has been a Cα r.m.s.d. exceeding 10 Å [41]. According to this, the proteins studied here are all fully denaturated after approximately 5–6 ns. There are still some fluctuations in the r.m.s.d. values in the simulations at the two highest temperatures (Fig. 1A and B). However, the protein unfolding proceeds slowly in the 375 K simulations (Fig. 1A), and the structural ensembles show less fluctuations when compared to the 400 and 425 K simulations. The 375 K simulations were therefore extended to 9 ns to reach this threshold. cUDG and hUDG show similar unfolding rates for the first 5 ns of the simulations. However, the psychrophilic enzyme displays

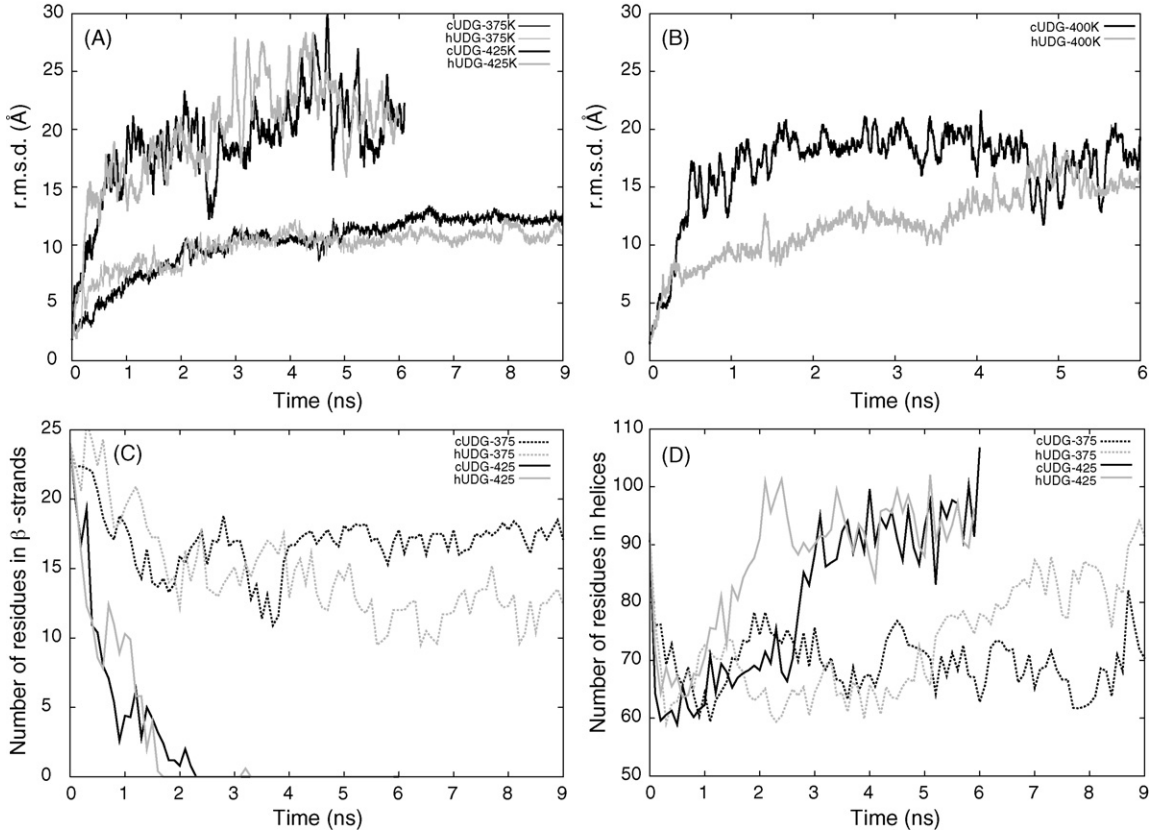


Fig. 1. Various plots from the simulations. (A) and (B) are average r.m.s.d. plots as a function of time during the simulation. The calculated r.m.s.d. values are based on all Cα atoms in the starting structure and from every snapshot during the simulation. (C) and (D) are number of residues in secondary structure elements during the simulation. The secondary structure elements are calculated by DSSP [37].

Please cite this article in press as: M. Olufsen et al., Comparative unfolding studies of psychrophilic and mesophilic uracil DNA glycosylase: MD simulations show reduced thermal stability of the cold-adapted enzyme, J. Mol. Graph. Model. (2006), doi:10.1016/j.jmgm.2006.10.003

higher r.m.s.d. values in the last 4 ns (Fig. 1A), indicating greater structural changes in the cold-adapted enzyme. The unfolding rate is faster when the temperature is raised to 400 K, and the simulations show that the enzymes unfolded to a higher degree at this temperature when compared to the 375 K results. Interestingly, large differences in the unfolding rates between the psychrophilic and the mesophilic enzyme are observed in the 400 K simulations. The psychrophilic enzyme unfolds much more rapidly than the mesophilic counterpart, indicating reduced structural stability (Fig. 1B). Destruction of the native protein structures occurs very fast when the temperature is raised to 425 K, and unfolding of cUDG and hUDG proceeds at a similar rate. Moreover, the simulations show that the mesophilic enzyme requires higher temperatures to achieve similar degree of structural distortions as observed for the psychrophilic enzyme. This already indicates differences in the interactions stabilizing the 3D structure, and is also in agreement with the observed reduced stability of the cold-adapted enzyme.

4. Structural investigations

4.1. Secondary structure

The starting structure of both cUDG and hUDG contain 92 and 24 residues in helices (α-helices, π-helices and 3–10 helices) and β-strands, respectively. The amount of residues in β-strands decreases at approximately the same rate in the simulation at 375 and 400 K for both enzymes (Fig. 1C), while for the simulation at the highest temperature, all β-strands are lost at around 2 ns for both enzymes (Fig. 1C). Both enzymes also have similar extent of residues involved in helices in the 375 and 400 K simulations, while it actually increases for both enzymes the 425 K simulation as the structure get distorted (Fig. 1D) (residues in helices for the 400 K simulation is not shown). Visual inspection of the simulated trajectories shows that the unfolding process generates new helices, originating from areas that were initially loops and β-strands in their respective crystal structures. The catalytic domain of both cUDG and hUDG is distorted already in the beginning of the 425 K simulations, and after 600 ps both enzymes have r.m.s.d. values of more than 15 Å (Figs. 1A and 7). When the domain becomes distorted and the β-sheets are destroyed, solvent gains access to the core β-strands yielding rapid unfolding. On the contrary, helices are stable even after an opening of the catalytic domain. However, both NMR and MD studies have shown that the denatured state may comprise significant amount of secondary structure [42]. There are also indications that the unfolded protein can increase the helical content during the simulation and that helices can be formed in water in the absence of significant tertiary interactions [22].

4.2. Hydrogen bonding pattern

The starting structures of cUDG and hUDG contained 240 and 223 intramolecular hydrogen bonds (shorter than 3.4 Å), respectively, and many are lost early in the simulations. However, after approximately 1 ns the number of hydrogen bonds reaches a stabile plateau, which is steadily maintained throughout the simulations. The average number of hydrogen bonds for the last 5 ns of the cUDG simulations is 198, 179 and 165 for the 375, 400 and the 425 K simulation, respectively, while the corresponding numbers for the mesophilic enzyme are 191, 174 and 156. Thus, as the simulation temperature is increased, there is a concomitant decrease in number of intact hydrogen bonds. This is reasonable as the structures become more distorted as the simulation temperature is raised.

4.3. Accessible surface area

The solvent accessible surface area (SASA) is an important parameter for mapping unfolding. The SASA has been calculated for the crystal structures and of snapshots from the MD simulations. The initial SASA for cUDG and hUDG is 10259 and 10376 $Å^2$, respectively. SASA increases rapidly for the 375 K simulations by approximately 20%, and stays roughly at the same level throughout the simulations (Fig. 2A). In the 400 K simulation the SASA increase to roughly 40% at the end of the simulation (Fig. 2B). While at 425 K the enzyme rapidly unfolds, and already after only 100–200 ps the SASA

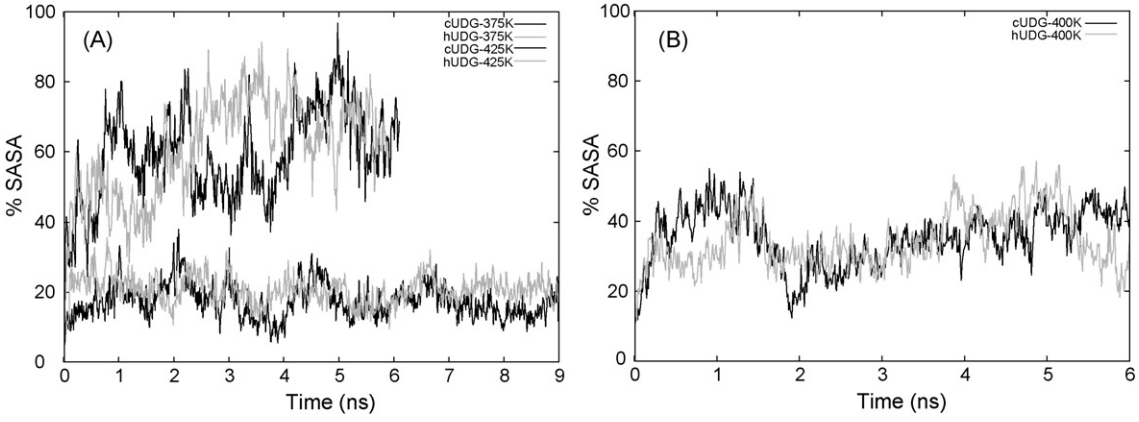


Fig. 2. Solvent-accessible surface area (SASA) as a function of simulation time. (A) shows percent increase in SASA from the 375 and the 425 K simulation. (B) shows percent increase in SASA from the 400 K simulation. The reference is the SASA from the crystal structures. The SASA are calculated by MSMS [38].

has increased by 40% (Fig. 2A). There are only minor differences in SASA among the enzymes as they unfold at different temperatures.

5. Stability of the N- and C-terminals

The N-terminal of cUDG unfolds at all three temperatures and as expected to a greater extent as the temperature is raised (Fig. 3A, C and E). While the N-terminal of the mesophilic variant of the enzyme also unfolds at all three temperatures (Fig. 3B, D and F), difference is observed when compared to cUDG. The N-terminal is more stable in the hUDG simulations as observed in the 375 and 400 K simulations. Structural investigations of the interactions involving the N-terminal show that there are several stabilizing contacts present. Numerous hydrophobic contacts between residues in the N-terminal helices (α1 and α2, α-helix 1 and 2, numbering according to Leiros et al. [9]) and residues in α6 and the loop between α6 and β1 are found. However, three hydrogen bonds (Ser88:Oγ-Asp133:Oδ2, Trp89:Nε1–Cys132:O and Trp89:Nε1–Thr129:O), which are conserved in the two enzymes, appear to be important for the N-terminal stability (Fig. 4). These hydrogen bonds connect the

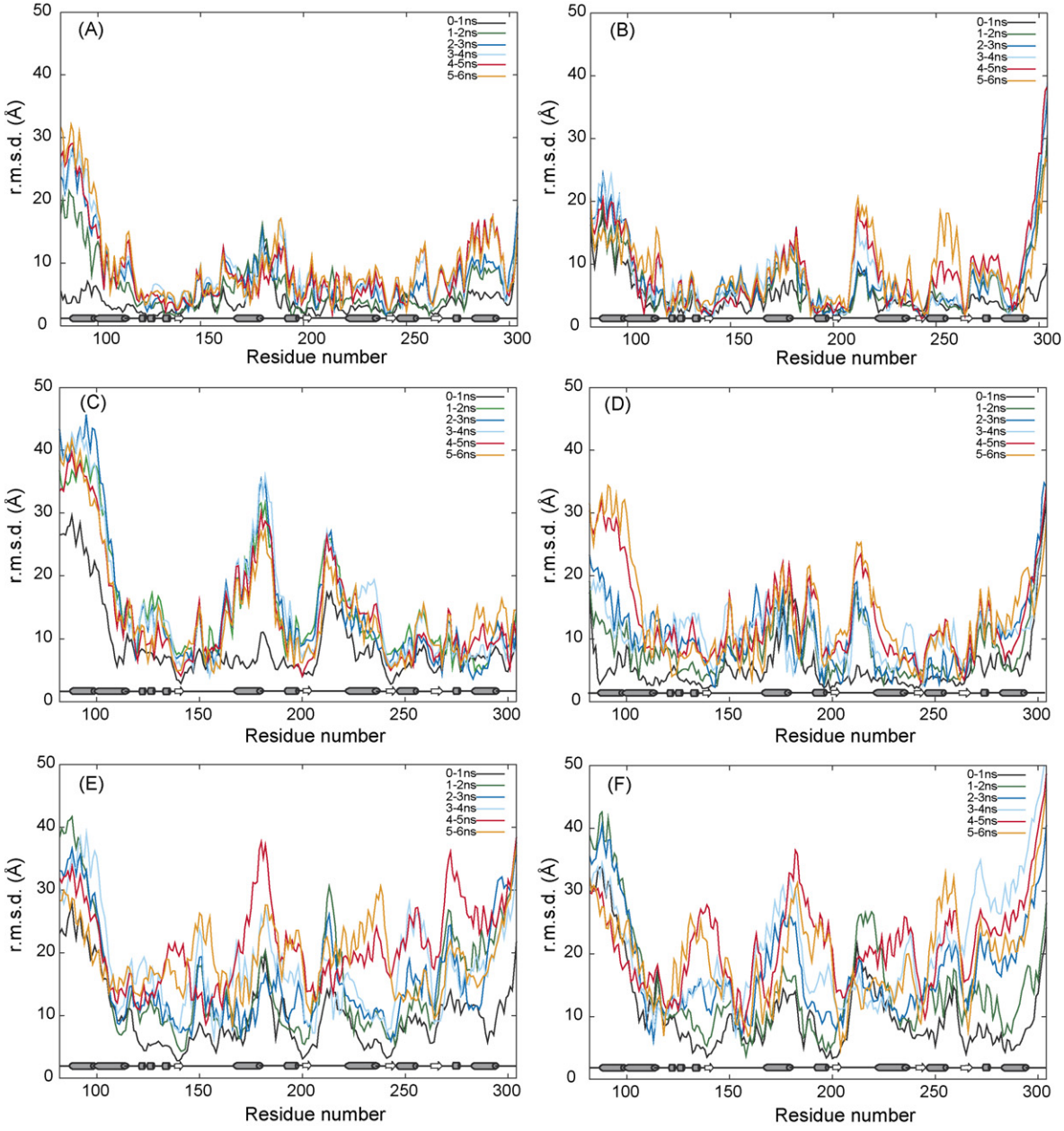


Fig. 3. Flexibility in the structure (r.m.s.d. per residue). The figure shows r.m.s.d. per residue. values in different part of the simulations, 0–1 ns is the r.m.s.d. values from the simulations from 0 to 1 ns, and so on. All the Cα atoms are used to calculate the r.m.s.d. values and the starting structure is used as a reference. The gray cylinder and the white arrow at the bottom of each plot is where helices and β-strands are located in the starting structure, respectively. (A), (C) and (E) is the r.m.s.d. values from the cUDG simulations at 375, 400, and 425 K, respectively. While (B), (D) and (F) is the r.m.s.d. values from the hUDG simulations at 375, 400 and 425 K, respectively.

Please cite this article in press as: M. Olufsen et al., Comparative unfolding studies of psychrophilic and mesophilic uracil DNA glycosylase: MD simulations show reduced thermal stability of the cold-adapted enzyme, J. Mol. Graph. Model. (2006), doi:10.1016/j.jmgm.2006.10.003

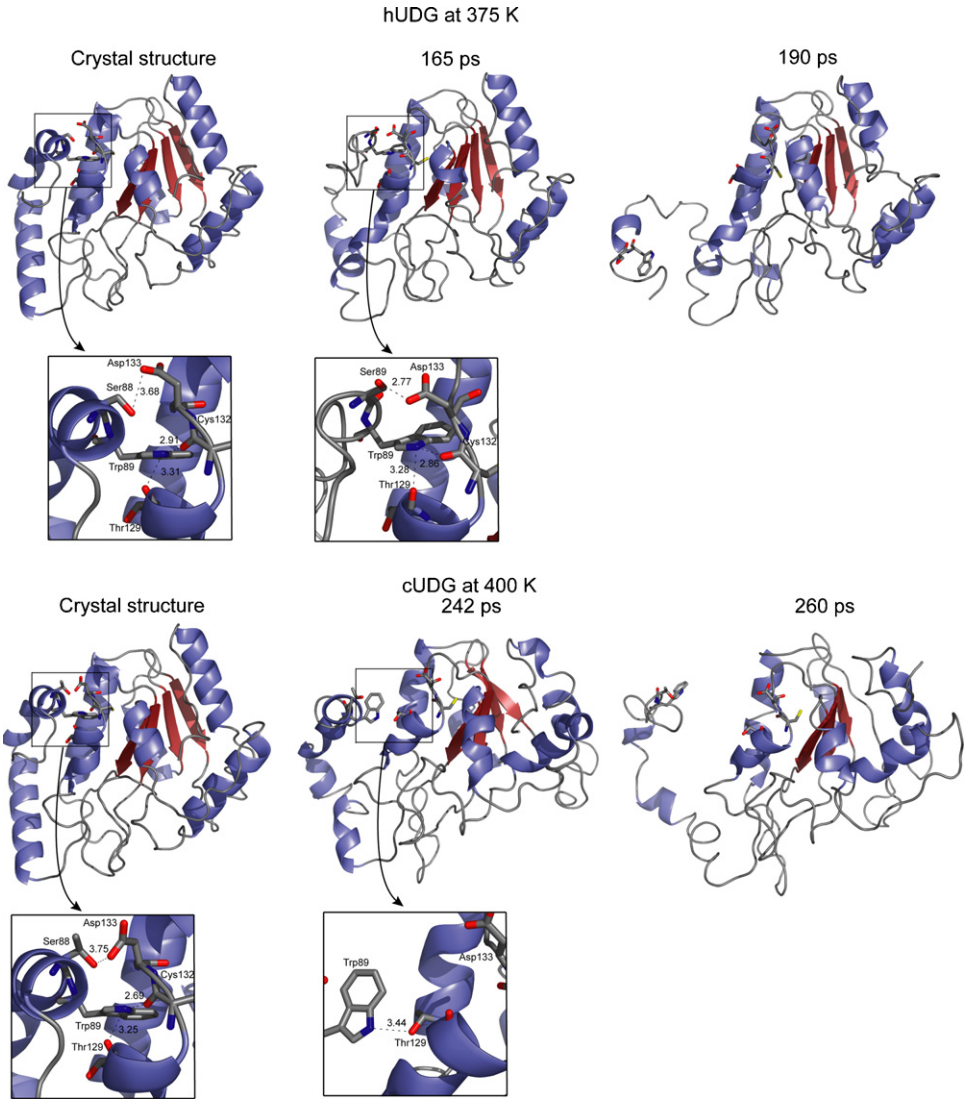


Fig. 4. Snapshots from the unfolding of hUDG at 375 K and cUDG at 400 K. The figure shows how the N-terminal unfolds from the rest of the structure. Some important residues are shown in ball-and-stick. The figure is generated in PyMOL [45].

N-terminal of helix α1 and the helix α4 and the loop between helix α4 and helix α5. In addition, the side chain of Trp89 packs into a hydrophobic area between helices α4 and α8.

Subtle differences in the hydrogen-bonding pattern observed during the simulations of cUDG and hUDG result in strikingly different behavior of their respective N-terminals. In the 375 K cUDG simulation the N-terminal starts to unfold at around 470 ps (Figs. 3A, 5A and 5B), as the previously mentioned hydrogen bond, Trp89:Nε1–Thr129:O, is lost. In contrast, the N-terminal starts to unfold at 177 ps in the 375 K simulation of hUDG, here the hydrogen bond between Ser88:Oγ and Asp133:Oδ2 is lost after 169 ps, and 8 ps later the Trp89:Nε1–Thr129:O hydrogen bond also disappear. Together, an immediate unfolding of the α1 and α2 helices is observed (Fig. 5A and B) in hUDG. Fig. 4 shows a detailed picture of the unfolding events occurring in the N-terminal of hUDG. The mesophilic enzyme starts to unfold earlier than the psychrophilic homologue at this temperature, but at around 2 ns the psychrophilic enzyme is more severely unfolded in the N-terminal.

When the temperature is increased to 400 K, the simulations show that the N-terminal of cUDG becomes highly flexible after only 70 ps. This is again a result of the loss of the Ser88:Oγ-Asp133:Oδ2 hydrogen bond. After the distortion of an additional hydrogen bond, Trp89:Nε1–Thr129:O, complete unfolding of the terminal is observed (Fig. 5D). Figs. 4 and 7 show the unfolding events observed for the N-terminal. Analysis of the simulation of hUDG at 400 K reveals that the hydrogen bond between the Ser88:Oγ and the Asp133:Oδ2 is lost after about 1025 ps (Fig. 5E), while the hydrogen bond between the Trp89:Nε1 and the Thr129:O disappear after 1075 ps (Fig. 5D). These events lead to a melting of the N-terminal in a similar manner as observed for cUDG.

In the 425 K simulation, molecular contacts are lost much faster than at lower temperatures, for example in the hUDG simulation important hydrogen bond in the N-terminal (Trp89:Nε1–Thr129:O) are lost after 110 ps and then the N-terminal starts to unfold (Fig. 5G). The corresponding in hydrogen bond in the psychrophilic enzyme is lost already after

Please cite this article in press as: M. Olufsen et al., Comparative unfolding studies of psychrophilic and mesophilic uracil DNA glycosylase: MD simulations show reduced thermal stability of the cold-adapted enzyme, J. Mol. Graph. Model. (2006), doi:10.1016/j.jmgm.2006.10.003

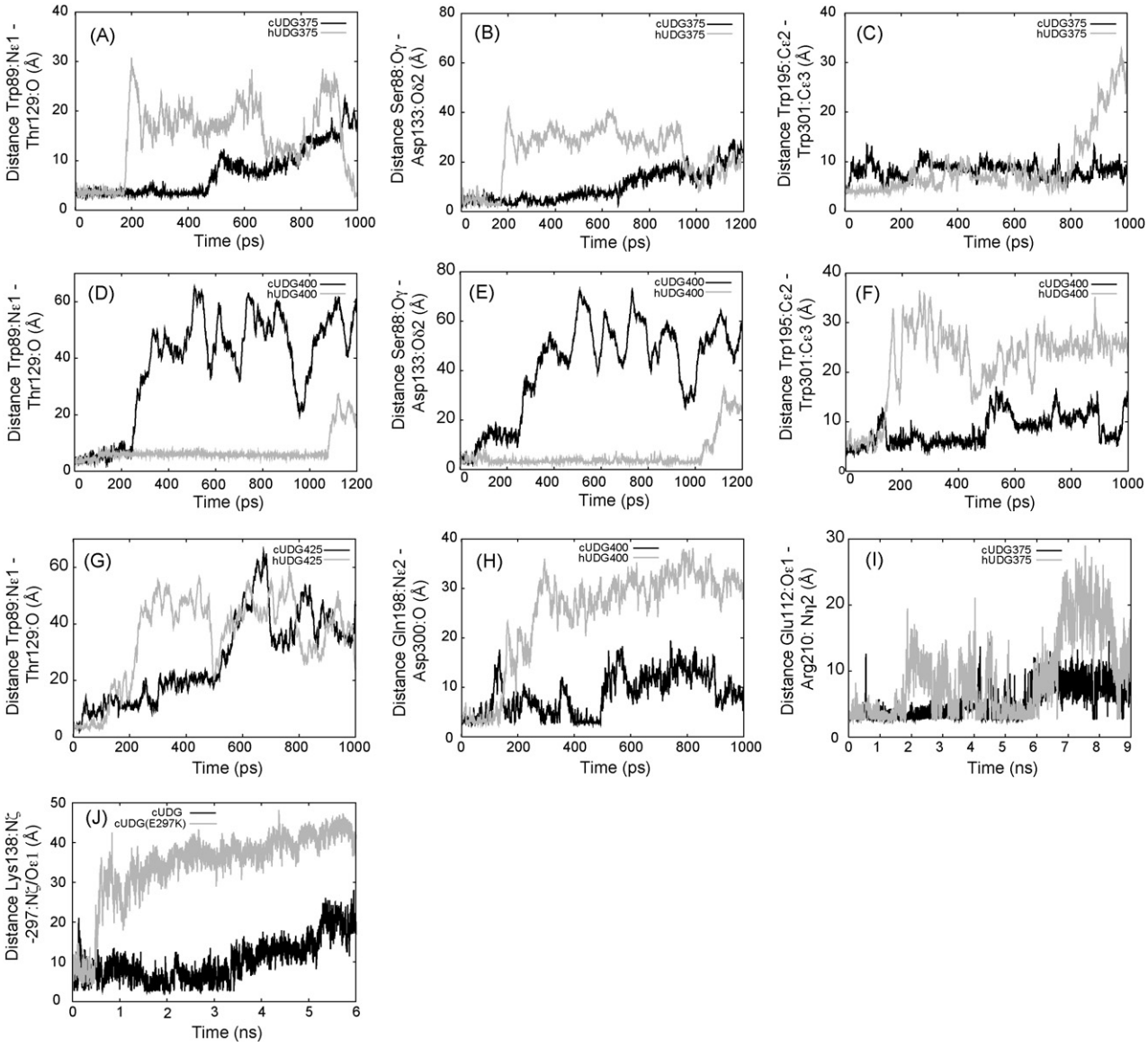


Fig. 5. Atom distance plots. Distances between selected atoms during simulation. (A), (D) and (G): distanced between the atoms Trp89:Nε1 and Thr129:O for the 375, 400 and 425 K simulation, respectively. (B) and (E): distance between the atoms Ser88:Oγ and Asp133:Oδ2 for the 375 and 400 K simulation, respectively. These above-mentioned atom pairs form hydrogen bonds close to the N-terminal. (C) and (F): distance between the atoms Trp195:Cε2 and Trp301:Cε3, for the 375 and 400 K simulation, respectively, which form a hydrophobic contact in the C-terminal. (H): distance between the atoms Gln198:Nε2 and Asp300:O in the 400 K simulation, these residues form a hydrogen bond in the C-terminal. (I): distance between atoms Glu112:Oε1 and Arg 210:Nη2. (J): distance between the atoms Lys138:Nζ and atom (Lys297:Nζ or Glu297:Oε1) in cUDG(E297K) and cUDG, respectively. Both simulations are done at 400 K simulation. There are different scales on the axes in the different plots.

25 ps (Fig. 5G). Many molecular contacts and interactions have been monitored in these simulations as well, and the unfolding of the two enzymes is very similar compared to the lower simulations temperatures. However, it is apparent that the kinetic energy of the system is so high that too many contacts are lost simultaneously making analysis of the simulations more difficult.

The different behavior of the N-terminal of cUDG compared to the hUDG is not obvious, but it is likely that subtle differences in the interactions contribute differently to the stability of the two N-terminals. In particular, the Trp89:Nε1–Thr129:O hydrogen bond seems to be important for N-terminal stability. Interestingly, the recently reported crystal structure of a thermophilic UDG from *Thermus thermophilus* HB8 (TthUDG) has a stabilizing [4Fe-4S] cluster close to the N-terminal [43]. TthUDG belongs to the family 4 UDG while hUDG and cUDG are from family 1. Even though the amino acid sequence homology is low between TthUDG and hUDG or cUDG, the topology and order of the secondary structure elements are very similar between the two UDG families [43]. High temperature adaptation thus seems to involve additional stabilization of the N-terminal part. Hence, increasing the stability of the N-terminal might make the mesophilic and the psychrophilic enzyme more thermostable. MD studies of Chymotrypsin inhibitor 2 have also shown that the N-terminal unfolds from the rest of the structure as one of the first events in

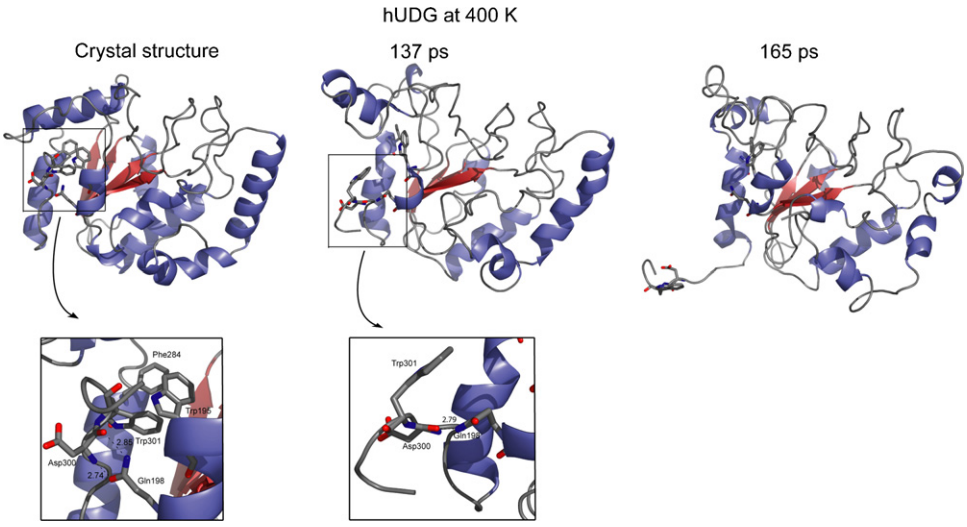


Fig. 6. Snapshots from the unfolding of hUDG at 400 K. The figure shows how the C-terminal unfolds from the rest of the structure. Some important residues are shown in ball-and-stick. The figure is generated in PyMOL [45].

the unfolding process [25,44], acting as an initiator of unfolding.

The very end of the C-terminal part (residue 293–304) is situated at the surface of the protein, and the side of the C-terminal facing away from the solvent is involved in a hydrophobic cluster with helices α7 and α11. In addition, the Trp301 side chain forms stacking interactions with the side chains of Phe284 and Trp195. There are also several hydrogen bonds that could be important for C-terminal stability, especially involving residue Gln198, forming hydrogen bond interaction with residue 300 (Asp in hUDG and Asn in cUDG) (Fig. 6). The fact that Trp301, Phe284, Trp195 and Gln198 are all conserved in the sequences studied by Leiros et al. [9], additionally indicates structural importance. The simulation shows that the C-terminal is far more stable in the cold-adapted enzyme when compared to its warm-active counterpart. The C-terminal of hUDG unfolds at all three simulations temperatures, while the C-terminal of cUDG only melts at 425 K.

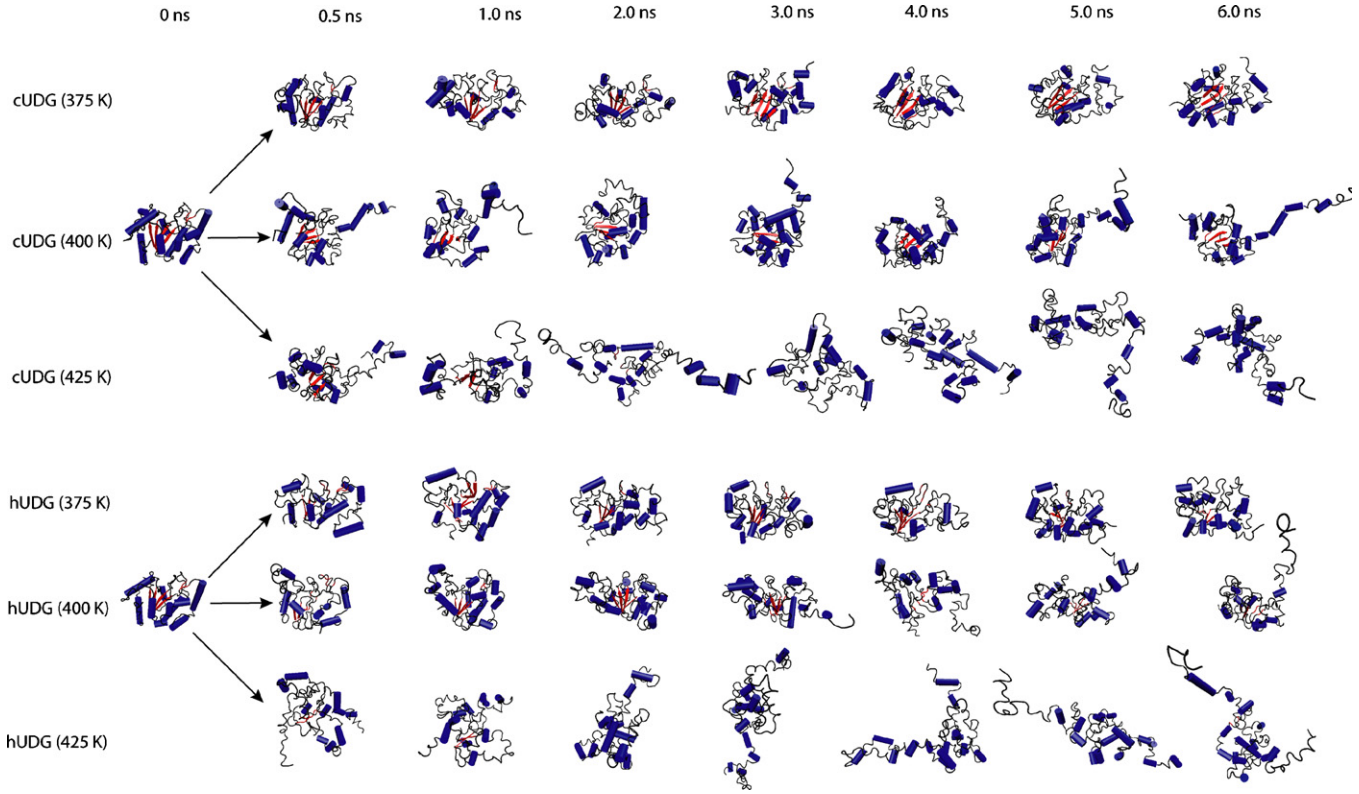


Fig. 7. Snapshots from the thermal unfolding simulations of cUDG and hUDG. The structures are made with the VMD program [46], helices, β-strands and loops are colored blue, red and gray, respectively.

Please cite this article in press as: M. Olufsen et al., Comparative unfolding studies of psychrophilic and mesophilic uracil DNA glycosylase: MD simulations show reduced thermal stability of the cold-adapted enzyme, J. Mol. Graph. Model. (2006), doi:10.1016/j.jmgm.2006.10.003

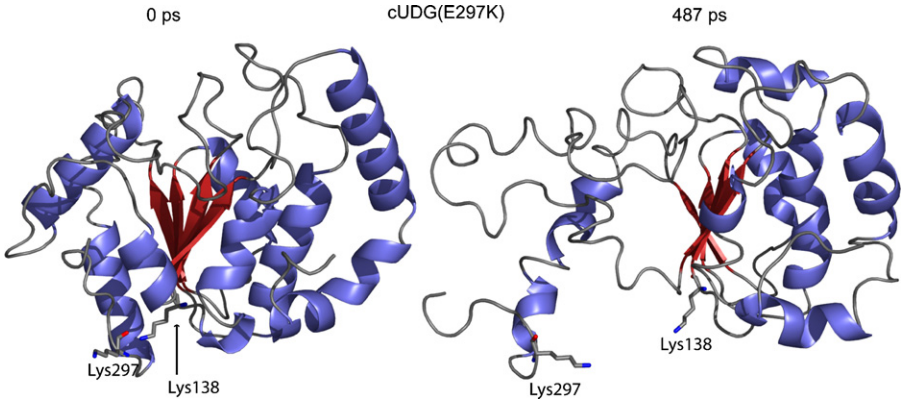


Fig. 8. Snapshots from the unfolding of cUDG(E297K) mutant at 400 K. The figure shows how the C-terminal unfolds from the rest of the structure. Some important residues are shown in ball-and-stick. The figure is generated in PyMOL [45].

Again, slightly different intramolecular interactions are responsible for the observed behavior in the C-terminal of cUDG and hUDG. The 375 K simulation of hUDG shows that around 450–500 ps, the hydrogen bonds between Gln198 and Asp300 disappear. Additional hydrophobic contacts in the C-terminal are also disrupted at around 800 ps, among others the Trp195–Trp301 side chain contact (Fig. 5C). These events seem to initiate melting of the C-terminal. The C-terminal of cUDG does not unfold at all at this temperature, but the last three residues have large fluctuations (Fig. 3A). Increasing the temperature to 400 K result in similar observations of the stability of the C-terminals. After only 140 ps several intermolecular contacts are lost, among others the hydrogen bond between Gln198 and residue 300 (Fig. 5H) and the hydrophobic contact between Trp195 and Trp301 (Fig. 5F). Together, these events lead to a rapid unfolding of the C-terminal as shown in Fig. 6. Again, the C-terminal is stable in the simulation of the psychrophilic enzyme, even though the interactions between residue Gln198 and Asn300 are not present after 500 ps. Favorable hydrophobic packing interactions between several residues in the C-terminal, among others the Trp195–Trp301 residues (Fig. 5F), is a possible explanation of this observation. The 425 K simulations show that both C-terminals unfold, but again with a higher stability in hUDG (Fig. 3E, F and 7).

Structural differences between the two enzymes may explain the differences observed in the stability of the C-terminal. First of all, residue 297 is a Lys in hUDG and a Glu in cUDG. In the crystal structure of cUDG this residue form a weak salt-bridge to Lys138 (5.4 Å between the Lys Nζ atom and the Glu Oε1 atom), while an unfavorable electrostatic interaction between two positively charged residues is observed in hUDG. To further investigate whether this explains the differences in unfolding of the C-terminal, two mutants of the enzymes were modeled from the crystal structures. The Lys297 residue in hUDG was mutated to a Glu and the Glu297 residue in cUDG was mutated to a Lys. Then 6 ns MD simulations of these two mutants were performed at 400 K. The C-terminal of the cUDG(E297K) mutant unfolds after ∼500 ps (Figs. 5J and 8), while the C-terminal of cUDG does not unfold at this temperature at all. This indicates that the weak salt bridge between Lys138 and Glu297 stabilizes the C-terminal in cUDG. The simulation of the hUDG (K297E) mutant shows that the enzyme still unfolds even when the mutations are introduced. Thus, other elements are also likely contributing to the observed difference in the unfolding of the C-terminal. From residue 293 to the C-terminal end, there are four residues in the sequence that make cUDG more hydrophobic than hUDG. The substitutions are K293L, K296T, D300N and E303A (mutation from hUDG to cUDG), all charged residues in hUDG, substituted to hydrophobic or uncharged in cUDG (for a full alignment see Leiros et al. [9]). It seems likely that the hydrophobic contacts are particularly important in the C-terminal, especially the hydrophobic/stacking contact between residue Trp195 and Trp301 (Fig. 5C and F). When this interaction is destroyed the C-terminal easily unfolds (Fig. 6).

6. Other differences in stability

Even if the largest differences in unfolding pathway among the two enzymes are in the terminals, there are also other parts of the structure possessing differences in unfolding pathways. In the 375 K simulation, the mesophilic enzyme unfolds to a high extent in the residues 207–226 region, whereas, the psychrophilic enzyme has a much lower r.m.s.d. per residue in this part of the structure (Fig. 3A and B). One possible reason for this difference is that the salt-bridge between residue Glu112 and Arg210 is not present for most of the simulation of hUDG whereas, it is intact for the first 5.5 ns of the simulation of cUDG (Fig. 5I).

In the 400 K simulation there are also other parts than the terminals of the structure that unfolds. Especially the segment from residue 205 to 220 unfolds rapidly for both enzymes, while in the 165–190 region the cUDG enzyme unfolds to a much higher extent than the mesophilic counterpart (Fig. 3C and D). There are several intermolecular interactions that may explain the difference in unfolding of this area. By analyzing atomic interactions trajectories it seems like residue 183, situated in a sharp turn, could explain much of the observed difference in this area. This residue is an Asp in hUDG and forms a salt bridge with Lys302 that is lost after 80 ps and does therefore probably not explain the observed differences. cUDG has a Gly at position 183 and one can therefore expect that the psychrophilic enzyme

Please cite this article in press as: M. Olufsen et al., Comparative unfolding studies of psychrophilic and mesophilic uracil DNA glycosylase: MD simulations show reduced thermal stability of the cold-adapted enzyme, J. Mol. Graph. Model. (2006), doi:10.1016/j.jmgm.2006.10.003

becomes more flexible in this area of the structure. In addition, there is a salt-bridge between Asp180 and Lys/Arg282 (Lys in cUDG and Arg in hUDG), which is lost at around 85 ps for both enzymes. This interaction is however reformed again in the hUDG structure, and is actually intact for large parts of the simulation. Hence, these specific interactions can explain the differences observed in the 165–190 area between the two enzymes. In the 425 K simulation both enzymes unfolds rapidly. After a short period of simulation, both cUDG and hUDG are highly distorted (Fig. 3E, F and 7).

7. Concluding remarks

The high sequence identity and structural similarity between cUDG and hUDG allows for the construction of various hypotheses that can be tested experimentally. Such studies (covering cloning, protein production, purification, characterization, etc.) are demanding in terms of costs and labor. Integrating computational approaches with different levels of complexity with the biochemical and/or biophysical studies, therefore allows in many cases for a more rationalized experimental process. In this study we have used molecular dynamics simulations at high temperatures to investigate not only the specific features of the unfolding mechanism of two homologous DNA repair enzymes, but also to gain a deeper understanding of the structurally important regions modulating thermal stability of the two. It is important to note that the objective of the present study is not by any means to try to solve the problem of protein folding. Instead, these simulations are used as tools to increase our knowledge of the structure–function relationship in uracil DNA glycosylase.

The simulations presented here (at 375 and 400 K) show that the psychrophilic UDG has lower thermal stability compared to the mesophilic UDG, which is in accordance with experiments [13]. The mesophilic enzyme needs higher temperature to reach the same rate of unfolding compared to cUDG. However, once the thermal energy of the systems is sufficiently high (i.e. 425 K in this case), rapid unfolding is observed for both enzymes. Subsequent analysis of the resulting trajectories provides indication of the structural regions important for protein stability. It is particularly the terminals that have different behavior between the cold- and warm-adapted UDG enzymes. Important molecular contacts, probably decisive for the stability of both terminals have been pin-pointed.

Acknowledgements

This work has been supported by a grant from the Research Council of Norway (RCN) (NFR 154197/432), and the Norwegian Structural Biology Centre (NorStruct) is supported by the National program in Functional Genomics (FUGE) in the Research Council of Norway.

References

[1] G. Feller, Molecular adaptations to cold in psychrophilic enzymes, Cell. Mol. Life Sci. 60 (2003) 648–662.
[2] T. Lonhienne, C. Gerday, G. Feller, Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility, Biochim. Biophys. Acta 1543 (2000) 1–10.
[3] D. Georlette, V. Blaise, T. Collins, S. D'Amico, E. Gratia, A. Hoyoux, J.C. Marx, G. Sonan, G. Feller, C. Gerday, Some like it cold: biocatalysis at low temperatures, FEMS Microbiol. Rev. 28 (2004) 25–42.
[4] C. Gerday, M. Aittaleb, M. Bentahir, J.P. Chessa, P. Claverie, T. Collins, S. D'Amico, J. Dumont, G. Garsoux, D. Georlette, A. Hoyoux, T. Lonhienne, M.A. Meuwis, G. Feller, Cold-adapted enzymes: from fundamentals to biotechnology, Trends Biotechnol. 18 (2000) 103–107.
[5] A.O. Smalås, H.K.S. Leiros, V. Os, N.P. Willassen, Cold Adapted Enzymes, vol. 6, Elsevier Science B.V., Amsterdam, 2000, pp. 1–57.
[6] O. Lanes, I. Leiros, A.O. Smalås, N.P. Willassen, Identification, cloning, and expression of uracil-DNA glycosylase from Atlantic cod (gadus morhua): Characterization and homology modeling of the cold-active catalytic domain, Extremophiles 6 (2002) 73–86.
[7] T. Lindahl, B. Nyberg, Heat-induced deamination of cytosine residues in deoxyribonucleic-acid, Biochemistry 13 (1974) 3405–3410.
[8] C.D. Mol, A.S. Arvai, G. Slupphaug, B. Kavli, I. Alseth, H.E. Krokan, J.A. Tainer, Crystal-structure and mutational analysis of human uracil-DNA glycosylase—structural basis for specificity and catalysis, Cell 80 (1995) 869–878.
[9] I. Leiros, E. Moe, O. Lanes, A.O. Smalås, N.P. Willassen, The structure of uracil-DNA glycosylase from Atlantic cod (gadus morhua) reveals cold-adaptation features, Acta Crystallogr. D 59 (2003) 1357–1365.
[10] R. Savva, K. Mcauleyhecht, T. Brown, L. Pearl, The structural basis of specific base-excision repair by uracil-DNA glycosylase, Nature 373 (1995) 487–493.
[11] R. Ravishankar, M.B. Sagar, S. Roy, K. Purnapatre, P. Handa, U. Varshney, M. Vijayan, X-ray analysis of a complex of *Escherichia coli* uracil DNA glycosylase (ecudg) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG, Nucl. Acids Res. 26 (1998) 4880–4887.
[12] M. Olufsen, A.O. Smalås, E. Moe, B.O. Brandsdal, Increased flexibility as a strategy for cold adaptation—a comparative molecular dynamics study of cold- and warm-active uracil DNA glycosylase, J. Biol. Chem. 280 (2005) 18042–18048.
[13] O. Lanes, P.H. Guddal, D.R. Gjellesvik, N.P. Willassen, Purification and characterization of a cold-adapted uracil-DNA glycosylase from Atlantic cod (gadus morhua), Comp. Biochem. Physiol. 127 (2000) 399–410.
[14] A.R. Fersht, V. Daggett, Protein folding and unfolding at atomic resolution, Cell 108 (2002) 573–582.
[15] R. Scandurra, V. Consalvi, R. Chiaraluce, L. Politi, P.C. Engel, Protein stability in extremophilic archaea, Front. Biosci. 5 (2000) D787–D795.
[16] K.A. Dill, Dominant forces in protein folding, Biochemistry 29 (1990) 7133–7155.
[17] S. Kumar, C.J. Tsai, R. Nussinov, Maximal stabilities of reversible two-state proteins, Biochemistry 41 (2002) 5359–5374.
[18] C.N. Pace, Polar group burial contributes more to protein stability than non-polar group burial, Biochemistry 40 (2001) 310–313.
[19] C.N. Pace, R.W. Alston, K.L. Shaw, Charge-charge interactions influence the denatured state ensemble and contribute to protein stability, Protein Sci. 9 (2000) 1395–1398.
[20] S. Kumar, R. Nussinov, Salt bridge stability in monomeric proteins, J. Mol. Biol. 293 (1999) 1241–1255.
[21] A.J. Li, V. Daggett, Characterization of the transition-state of protein unfolding by use of molecular-dynamics—chymotrypsin inhibitor-2, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 10430–10434.
[22] A.J. Li, V. Daggett, Molecular dynamics simulation of the unfolding of barnase: characterization of the major intermediate, J. Mol. Biol. 275 (1998) 677–694.
[23] A. Caflisch, M. Karplus, Acid and thermal-denaturation of barnase investigated by molecular-dynamics simulations, J. Mol. Biol. 252 (1995) 672–708.
[24] U. Mayor, N.R. Guydosh, C.M. Johnson, J.G. Grossmann, S. Sato, G.S. Jas, S.M.V. Freund, D.O.V. Alonso, V. Daggett, A.R. Fersht, The complete folding pathway of a protein from nanoseconds to microseconds, Nature 421 (2003) 863–867.

[25] R. Day, B.J. Bennion, S. Ham, V. Daggett, Increasing temperature accelerates protein unfolding without changing the pathway of unfolding, J. Mol. Biol. 322 (2002) 189–203.
[26] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman, 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. 91 (1995) 1–41.
[27] J.M. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? J Comput. Chem. 21 (2000) 1049–1074.
[28] A. Onufriev, D. Bashford, D.A. Case, Modification of the generalized born model suitable for macromolecules, J. Phys. Chem. B. 104 (2000) 3712–3720.
[29] J. Karanicolas, C.L. Brooks, Integrating folding kinetics and protein function: Biphasic kinetics and dual binding specificity in a ww domain, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 3432–3437.
[30] Y.Z. Ohkubo, C.L. Brooks, Exploring flory's isolated-pair hypothesis: statistical mechanics of helix-coil transitions in polyalanine and the c-peptide from rnase a, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 13916–13921.
[31] M. Feig, C.L. Brooks, Recent advances in the development and application of implicit solvent models in biomolecule simulations, Curr. Opin. Struc. Biol. 14 (2004) 217–224.
[32] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical-integration of cartesian equations of motion of a system with constraints—molecular-dynamics of *n*-alkanes, J. Comput. Phys. 23 (1977) 327–341.
[33] H.J.C. Berendsen, J.P.M. Postma, W.F. Vangunsteren, A. Dinola, J.R. Haak, Molecular-dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[34] D. Sitkoff, K.A. Sharp, B. Honig, Accurate calculation of hydration free-energies using macroscopic solvent models, J. Phys. Chem. 98 (1994) 1978–1988.
[35] I. Leiros, O. Lanes, O. Sundheim, R. Helland, A.O. Smalås, N.P. Willassen, Crystallization and preliminary X-ray diffraction analysis of a cold-adapted uracil-DNA glycosylase from Atlantic cod (gadus morhua), Acta Crystallogr. D 57 (2001) 1706–1708.
[36] A.R. Dinner, G.M. Blackburn, M. Karplus, Uracil-DNA glycosylase acts by substrate autocatalysis, Nature 413 (2001) 752–755.
[37] W. Kabsch, C. Sander, Dictionary of protein secondary structure—pattern-recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[38] M.F. Sanner, A.J. Olson, J.C. Spehner, Reduced surface: an efficient way to compute molecular surfaces, Biopolymers 38 (1996) 305–320.
[39] E. Moe, I. Leiros, E.K. Riise, M. Olufsen, O. Lanes, A. Smalas, N.P. Willassen, 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. 343 (2004) 1221–1230.
[40] V. Daggett, A. Fersht, The present view of the mechanism of protein folding, Nat. Rev. Mol. Cell Biol. 4 (2003) 497–502.
[41] V. Daggett, Molecular dynamics simulations of the protein unfolding/folding reaction, Acc. Chem. Res. 35 (2002) 422–429.
[42] C.J. Bond, K.B. Wong, J. Clarke, A.R. Fersht, V. Daggett, Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 13409–13413.
[43] J. Hoseki, A. Okamoto, R. Masui, T. Shibata, Y. Inoue, S. Yokoyama, S. Kuramitsu, Crystal structure of a family 4 uracil-DNA glycosylase from *Thermus thermophilus* hb8, J. Mol. Biol. 333 (2003) 515–526.
[44] A.J. Li, V. Daggett, Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations, J. Mol. Biol. 257 (1996) 412–429.
[45] W.L. DeLano, The Pymol Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA, 2002.
[46] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.

Please cite this article in press as: M. Olufsen et al., Comparative unfolding studies of psychrophilic and mesophilic uracil DNA glycosylase: MD simulations show reduced thermal stability of the cold-adapted enzyme, J. Mol. Graph. Model. (2006), doi:10.1016/j.jmgm.2006.10.003

Paper IV

Ion pairs and their role in modulating stability of cold- and warm-active uracil DNA glycosylase.

Magne Olufsen[1], Elena Papaleo[2], Arne Oskar Smalås[1] & Bjørn Olav Brandsdal[1]*

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

[2]Department of Biotechnology and Bioscience, University of Milano-Bicocca, P.za della Scienza 2, 20126, Milan, Italy

*Corresponding author:
Bjørn Olav Brandsdal
Email: bjoerno@chem.uit.no
Phone: +47 77644057
Fax: +47 77644765

Running title: Ion pairs in cold- and warm-active UDG.
Key words: Uracil DNA glycosylase, salt-bridge, ion pairs, molecular dynamics simulations, stability, continuum electrostatics, cold adaptation,

Abstract

MD simulations and continuum electrostatics calculations have been used to study the observed differences in thermostability of cold- and warm-active uracil DNA glycosylase (UDG). The present study focuses on the role of ion pairs and how they affect the thermal stability of the two enzymes. Based on analysis of the MD generated structural ensembles, we find that cod UDG (cUDG) and human UDG (hUDG) have 11 and 12 ion-pairs which are present in at least 30 % of the conformations. The electrostatic contribution of the ion-pairs, computed using continuum electrostatics, is slightly more favorable in cUDG at 298 K. This is primarily attributed to more optimized interactions between the ion-pairs and nearby dipoles/charges in cUDG. When we take the environmental temperatures into account, the electrostatic stability becomes more favorable for the ion-pairs in the mesophilic enzyme. Both enzymes contain one three-member ionic network, but the one found in hUDG is far more stabilizing. Global salt-bridges are generally found to be more stabilizing than local bridges. In order to maintain structural integrity at low temperature, the electrostatic stability of ion-pairs is reduced, which possibly contribute to increased molecular flexibility at low temperature. Our results also suggest that care should be taken when performing statistical analysis of crystal structures with respect to ion-pairs, and that crystallization conditions must be carefully examined when performing such analysis.

Introduction

Life has evolved to cope with a broad range of different environments and is found in many extreme habitats such as low or high temperature niches, high pressure, low or high pH, high salinity etc. Adaptation to thermal extremes surely involves many mechanisms and survival in such environments requires enzymes with specific properties to maintain metabolism and reproductive activities. Interest in extremophilic enzymes has increased in the past years not only from an academic point of view but also as possible targets for commercialization. Cold-adapted, or so-called psychrophilic, enzymes are particularly attractive as targets due to their reduced thermal stability which is usually accompanied with an increased catalytic efficiency. The question of how enzymes can maintain stability in extreme environments continues to attract attention, and is presently not well understood. Much information about protein stability has been gained with the increasing availability of high-resolution structures and through comparative studies of homologous proteins[1,2]. Structural analysis of numerous psychrophilic, mesophilic and thermophilic enzyme homologues have revealed that their structures are often very similar[3], indicating difficulties with identifying features responsible for thermal adaptation by conventional methods. In an extensive study of structural differences between thermally adapted enzymes it was found that different protein families may use different strategies to adopt to low temperatures[4]. Psychrophilic enzymes often have decreased number of proline residues, increased number of glycine and acidic residues, low relative arginine ratio [arg/(arg+lys)] and lower content of ion pairs, aromatic interactions and/or hydrogen bonds compared to mesophilic counterparts[2,5-7]. There is also a trend that cold-adapted enzymes have increased interactions with solvent and higher accessibility to the active site[3,8], and multimeric psychrophilic enzymes often have decreased sub-unit interactions. It is clear that many different factors can contribute to enzymatic temperature adaptation and that no single factor can be invoked to explain adaptation in general.

Alteration of electrostatic interactions is one of the simplest ways nature can manipulate biophysical properties central to thermal stability, catalytic efficiency and substrate binding. It is clear that the stability of proteins stems from the delicate balance between opposing forces. The native folded three dimensional structure of proteins is only marginally stable and is at the edge of thermodynamic stability. Many different factors contribute to the overall stability of proteins, but the hydrophobic effect is generally accepted as the main contributor in protein folding. The hydrophobic effect drives the molecule towards a more condensed structure by minimizing unfavorable contacts between non-polar

residues and solvent. It is also accompanied by the entropically favorable release of water molecules around exposed hydrophobic groups[9-11]. In the initial phase of folding the hydrophobic effect leads to formation of the globular protein structure, often referred to as the molten globule state. The next step involves optimization of specific interactions that shifts the equilibrium towards the native state. Electrostatic interactions are an example of specific interactions and play an important role when a protein reaches its native fold. Other specific interactions that can contribute to the stability are van der Waals interactions and hydrogen bonds between polar groups[12,13]. Electrostatic interactions and their role in temperature adaptation of proteins is the key focus in this study.

There is at present only limited knowledge of how electrostatics is exploited by enzymes obtained from cold-adapted organisms. This is a result of the fact that only a few experimentally determined structures of psychrophilic enzymes are available. Enhanced electrostatic interactions has been found to play an important role in cold-adaptation of trypsin and uracil DNA glycosylase (UDG) with regard to activity[14-16], but the role in stability has not yet been investigated in detail. Cold active citrate synthase has strikingly different electrostatic potential in comparison to its hyperthermophilic counterpart, and focused electrostatic attraction of substrates has been proposed to be a possible source of the enhanced catalytic efficiency of the cold active citrate synthase[8]. Kumar and Nussinov[17] have also investigated the roles of electrostatics in psychrophilic, mesophilic and thermophilic citrate synthase using continuum electrostatics calculations. They found that electrostatics play different roles and appear to stabilize the thermophilic enzyme while ensuring proper protein solvation and active-site flexibility in the psychrophilic citrate synthase.

Salt-bridges or ion pairs can be considered as a special case of hydrogen bonds, and are composed of negative charges from Asp, Glu and C-terminals, and positive charges from Arg, His, Lys and N-terminals. These interactions are predominantly of electrostatic character and are also pH-dependent. Ionic interactions and their role in modulating protein stability have been the focus of many studies in the past two decades[9,18,19]. The increased occurrence of salt-bridges as well as salt-bridge networks in thermophilic proteins compared to their mesophilic homologues is the main reason for the great interest in electrostatics and its role in protein stability[20-22]. Even though the question of whether salt-bridges stabilize the native state of proteins has been raised repeatedly, no definite answer has yet been provided. Experimental and theoretical studies have showed that in some cases salt-bridges stabilize the native structure whereas in others an unfavorable contribution to protein stability is observed[9,23]. This is most likely related to the fact that the location, geometry and

optimization of the electrostatics varies greatly among not only homologous proteins but also between none-related proteins[23,24].

In order to improve our understanding of the role electrostatics play in adaptation to cold environments, uracil DNA glycosylase from cod (cUDG) and human (hUDG) is used as a comparative model system. UDG is an important enzyme in the DNA repair process and is the first enzyme in the base excision repair pathway, catalyzing the removal of promutagenic uracil residues from single- or double-stranded DNA[25]. The crystal structure of the catalytic domain of UDG from several species are known: human (hUDG)[26], Atlantic cod (cUDG)[27], virus 1[28], E.coli[29] and Epstein-barr virus[30]. The catalytic domain of hUDG and cUDG consists of 223 amino acids with a sequence identity of 75%, and the overall topology is a typical α/β protein[26]. High sequence and structural similarity makes UDG a good choice as a model system to explore adaptational features. Further support to this choice is also added by the fact that several crystal structures of mutant enzymes have been solved and subjected to kinetic and thermodynamic characterizations [16,27,31-34]. Moreover, UDG from Atlantic cod has been found to be up to ten times more catalytically active than the human counterpart between 283 and 310 K[32]. The psychrophilic cUDG has also been found to be much more pH and temperature labile than hUDG[31], indicating that cUDG is less stable compared to hUDG. The measured half-life was 0.5 min for cUDG and 8 min for hUDG at 323 K[32]. Even though these two enzymes has similar number of charged residues, similar amount of hydrogen bonds and no obvious differences in hydrophobic packing[27], the psychrophilic enzyme show less thermal stability compared to the mesophilic counterpart. Differences in salt-bridges are however observed in their respective crystal structures. The possible role of salt-bridges in thermal adaptation of UDG is in this study investigated using molecular dynamics simulation and continuum electrostatic calculations.

Methods

MD simulations

The AMBER8 program package[35] with the parm99 force field[36] was used to run and analyze the MD simulations. UDG contains several histidines and were with one exception (His148), considered as neutral in the simulations. This choice was based on existing data from NMR and continuum electrostatics calculations[37]. 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[38]. The total box size for the system was approximately 85x70x75 Å with roughly 50 000 atoms in the system. 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 100 ps. The time step used was 2 fs and for every 1000 step, the rotational and translational motion was removed. In the production phase the simulation was run with the isothermal-isobaric ensemble (300 K and one atmosphere pressure). Pressure and temperature were maintained by the Berendsen coupling algorithm[39]. A 8 Å cutoff for non-bonded interactions was used in the simulation and the Particle-Mesh-Ewald (PME) method[40] was used to handle long-range interactions beyond the cutoff. SHAKE[41] was applied to constrain covalent bonds involving all hydrogen atoms. Coordinates were written to file every ps during the production phase, and the simulations were carried out for 5 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).

The accessible surface area (ASA) of the residues was calculated using WHAT IF[42], with the standard parameters. The ASA of a salt-bridge is the average ASA of the side chains of the two residues involved in the salt-bridge. The ASA was calculated for 10 snapshots from the MD simulations and the crystal structure, but only the average value is presented. The distances between salt-bridge partners were calculated with the AMBER software package. The distances were calculated between the following atoms: (Lys:Nζ, Arg:Cζ and His:Cε1) and (Asp:Cγ and Glu:Cδ). A cutoff of 5.0 Å was used to decide if the salt-bridge is stable with respect to distance. If two oppositely charged residues are within 5.0 Å for more than 30% of the simulation time it is considered to be a salt-bridge and the energy of the salt-bridge was calculated.

Continuum electrostatic calculations

Continuum electrostatic calculations were carried out on 99 structures from the MD simulations on hUDG and cUDG and on the crystal structure of these enzymes. Every 50th structure from the simulations was subjected to continuum calculations. The DelPhi program[43,44] was used to calculate the electrostatic stability of the salt-bridges. The calculations were performed using the partial charges and atomic radii of the AMBER force field (parm99)[36]. The PARSE3 set[45] of atomic radii was also tested together with the partial charges from the AMBER parm99 force field. On average the all AMBER parameter set stabilized each salt-bridge by 0.2 kcal/mol more than the mixed PARSE3 and the AMBER parameter set. Only the results from the all AMBER parameter set are shown. 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[46]. 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 ~90% of the box. Lower boxfill values were also tested (~80%) but this gave similar results, consistent with published results[46]. A solvent probe of 1.4 Å was used to calculate the molecular surface. The dielectric constant of water is not constant with temperature, and the electrostatic energy was calculated for 0°C for the psychrophilic enzyme and 37°C for the mesophilic homologue. In addition, the electrostatic energy was calculated for both enzymes at 25°C. A bulk dielectric constant of 87.9, 80.0 and 75.9 was used for the temperatures 273 K, 298 K and 310 K, respectively. All the calculations were run both with 0.1 M salt and with zero ionic strength and three different dielectric constants of the protein were tested (4, 10 and 20).

The electrostatic stability of the salt-bridges has been calculated with respect to their corresponding hydrophobic isosteres[47]. The hydrophobic isosteres are the corresponding charged salt-bridge residues with all atom charges in the side chain set to zero[47]. The total electrostatic free energy of a salt-bridge ($\Delta\Delta G_{tot}$) can be divided into three terms:

Eq I: $$\Delta\Delta G_{tot} = \Delta\Delta G_{dslv} + \Delta\Delta G_{brd} + \Delta\Delta G_{prt}$$

The $\Delta\Delta G_{dslv}$ is the desolvation energy and in order to calculate this term, four DelPhi runs are needed: two of the folded protein with one of the salt-bridge side chain charged, and two calculations of the unfolded state which is represented with only one the salt-bridge partners charged (Eq. II).

Eq II: $$\Delta\Delta G_{dslv} = \Delta\Delta G_{folded,+} - \Delta\Delta G_{unfolded,+} + \Delta\Delta G_{folded,-} - \Delta\Delta G_{unfolded,-}$$

The $\Delta\Delta G_{brd}$ (bridge) is the electrostatic interactions between the salt-bridge partners. To calculate the bridge term only one calculation is needed, since the frc-mode in DelPhi was applied. In this calculation the frc-file contains the coordinate of one of the salt-bridge partners with the charged side chain atoms, while the pdb file contain the complete protein with the side chain of the other salt-bridge partner charged and the other residues uncharged. The $\Delta\Delta G_{prt}$ (protein) reflects the electrostatic energy between the charges in the salt-bridge and all the other atomic charges in the protein. Also here, only one run is needed when the frc-mode is applied. Then the side chain of the salt-bridge is charged in the frc-file and in the pdb-file all atoms are charged except the side chain atoms of the salt-bridge partners.

Result and discussion

Adaptation of enzymes to different environments occurs through mutations at specific positions in the protein sequence, resulting in enzymes capable of functioning in its specific habitat. Uracil DNA glycosylase from cod and human have been used as our model system to investigate the adaptation to low temperature. Inspection and comparison of their respective crystal structures shows differences in putative salt-bridges[27], but their possible role in providing stability has not yet been investigated. The crystallographic analysis of ion-pair interactions in cUDG and hUDG considered all histidines as protonated (charged), but NMR and previous computational investigations suggest that only His148 is charged[37]. Our simulations and subsequent analysis therefore only treat the His148 as charged, and the total number of ionic interactions is thus significantly lower than suggested by the crystallographic analysis[27]. The crystal structure of the psychrophilic enzyme (PDB entry 1OKB[27]) has 4 salt-bridges shorter than 5.0 Å, whereas the mesophilic homologue (PDB entry 1AKZ[26]) has 10 such salt-bridges (Table I). If we assume that salt-bridges are generally stabilizing, this implies that the mesophilic enzyme is far more stabilized when compared to its cold-active cousin. However, the situation is far more complicated. Several crystal structures of cUDG and hUDG have been determined and deposited to the protein data bank[48]. Five high resolution X-ray structures of hUDG have been examined, and their content of salt-bridges range from 5 to 10, while 4 and 6 are found in the two crystal structures of cUDG (Table I). Many ion-pairs are also located at the protein surface, and the electron density for highly solvent exposed lysines and arginines is often not well-defined. It should also be kept in mind that crystal structures present a static picture of the protein structures with one side-chain orientation for the partners in the salt-bridge, but with variability in the atomic positions as reflected in the crystallographic B-factors. Together, this adds even more complexity to the question of how important salt-bridges are to protein stability. It is clear that a much more reliable picture can be obtained by using computer simulations techniques. Molecular dynamics simulations allows for exploration of the conformational space of the protein and in particular of structural flexibility.

Ionic interactions present in the MD simulations

MD simulations generate an ensemble of conformations and thus include valuable information of the protein dynamics. The r.m.s.f. of the Cα atoms show that the MD simulations of both enzymes is equally stable throughout the simulation time (Fig. 1). In the

present simulations we have sampled conformations every ps yielding a total of 5000 structures for each enzyme that were analyzed with respect to possible salt-bridges interactions. The search for ion-pair interactions can be done in several different ways, but the distance between the two residues forming the ion pair is the first and most important criterion. Kumar and Nussinov[49] have suggested to differentiate between ion-pairs based on not only the distance between the charged-residue side-chain nitrogen and oxygen, but also the distance between the side-chain charged-group centroids[50]. Accordingly, an ion-pair is formed when both distances are less than 4.0 Å, a nitrogen-oxygen (N-O) bridge is formed when the distance between the atoms is less than 4.0 Å but with the centroid distance greater than 4.0 Å. Longer-range ion-pairs are formed when both distances are more than 4.0 Å. In this study we have used the distances between the following atoms: (Lys:Nζ, Arg:Cζ and His:Cε1) and (Asp:Cγ and Glu:Cδ). Fluctuations in the geometry of salt-bridges are observed in the trajectories where they break and reform. The distance between the residues in salt-bridges has been observed to vary in other MD simulation studies[51], NMR studies[52] and also between different crystal structures of the same protein[52]. Table I also shows that this is the case for UDG. The lifetime (or occupancy) of salt-bridges has therefore been used as the key parameter to identify ion pairs. It is defined as the number of structures where the salt-bridge is intact, judged by a distance criterion of 5 Å, divided by the total number of structures (5000). We require that salt-bridges must possess a lifetime greater than 30 % in order to be counted. Using this criterion, the distance between the salt-bridge partners is shorter than 5 Å for at least 30 % of the investigated structures, but the average distance calculated using all structures can be longer than 5 Å.

The environment of all charged amino acids was analyzed with respect to the presence of an oppositely charged amino acid in all sampled structures. If we require that the salt-bridge should possess a lifetime greater than 30 %, there is an increase in the number of salt-bridges in both cUDG and hUDG in the simulations as compared to the crystal structures (Table II). For cUDG seven new salt-bridges are formed during the MD simulations, while for hUDG one salt-bridge observed in the crystal structure disappears and three new are formed (Table II). The relatively higher increase in the number of salt-bridges between the crystal structure and the simulation for cUDG when compared to hUDG can be attributed to different crystallization conditions and particularly to the presence of PEG 4000 in hUDG and 1.4 M sodium citrate in cUDG[26,27]. The hydrophobic conditions in crystallization of hUDG may make it more energetically favorable for surface charges to interact with each

other, while high salt concentrations can lead to increased screening effects and solvent interactions for cUDG. cUDG and hUDG have a total of 11 and 12 salt-bridges that are intact for 30% of the simulation time, respectively (Table II). The average distance between the bridging residues from the MD simulations is 4.63 and 4.88 Å for cUDG and hUDG, respectively. Particularly stable ion-pairs are defined as those with a lifetime greater than 80 %. Table II shows that both cUDG and hUDG have six highly stable pairs, which also have the shortest average distance between the partners. The average distance is 4.06 and 3.93 Å for the six ion-pairs with lifetime > 80 % in cUDG and hUDG, respectively. All salt-bridges observed in the MD simulations are shown in Fig. 2A and B.

Electrostatic contribution to ion-pair stability.

One of the main objectives with this work is to investigate the role of ion-pairs in temperature adaptation of UDG. This requires that we are able to assess the strength of such interactions. We have combined MD simulations with continuum calculations for this purpose. Continuum electrostatics calculations have traditionally been carried out using static structural models, thus neglecting conformational fluctuations within salt-bridges and their interactions with explicit solvent. Full atomistic molecular dynamics simulations with explicit treatment of solvent represent a straightforward way of simultaneously including both effects. It is also possible to use rigorous statistical mechanics to extract thermodynamic information on salt-bridge formation, but these are generally more complicated to use compared to continuum models and requires more simulations time to reach convergence. Potential of mean force for association of ion pairs along user-defined paths has been computed[53], but application of artificial restraints to force association can be a limitation of such methods. Continuum methods based on classical electrostatics treat the solvent in terms of average bulk properties while the solute is represented in full atomic detail. Such methods have been widely used as both qualitative and quantitative tools in studies of electrostatic potentials, pH-dependent properties as well as solvation energies[54]. Ion-pairs and the electrostatic contribution to the free energy change upon salt-bridge formation have previously been investigated with continuum electrostatic calculations[17,20,22,23,47,49,52,55]. The computational method outlined by Hendsch and Tidor[47] has been followed here.

The electrostatic stability of all ion-pairs that survived in more than 30 % of the MD generated structures has been computed (Table II). In order to reduce the computing time only every 50th structure from the trajectories was included, and the calculations thus include 100 representative conformations for each ion-pair. The following energy terms are obtained:

the total electrostatic contribution to the free energy $\Delta\Delta G_{tot}$, the desolvation penalty for the charged residues $\Delta\Delta G_{dslv}$, the bridge energy for the electrostatic interaction between the pairing residues $\Delta\Delta G_{brd}$ and the protein energy term for the electrostatic interaction between the ion-pair and the surrounding protein environment $\Delta\Delta G_{prt}$. The $\Delta\Delta G_{ass}$ is the free energy change associated with forming a salt-bridge but it does not consider the interaction of the salt-bridge with the rest of the protein. $\Delta\Delta G_{ass}$ is the sum of the $\Delta\Delta G_{dslv}$ and the $\Delta\Delta G_{brd}$ [47].

The total electrostatic stability of an ion-pair is computed as the sum of $\Delta\Delta G_{dslv}$, $\Delta\Delta G_{brd}$ and $\Delta\Delta G_{prt}$, and these terms are the average values based on the 100 structures in our sub-set. To obtain the total electrostatic stability, 6 continuum electrostatic calculations are therefore needed, yielding a total of 600 for each of the pairs. While the folded state is modeled with all atoms present, the unfolded state is modeled using a single amino acid representation. How the three dimensional structure of proteins look like when they are unfolded is presently not known, which makes it very challenging to model the unfolded state. We have made attempts to use the structures from our high temperature unfolding studies of cUDG and hUDG[33], but obtaining convergent energies was extremely difficult (results not shown).

Effect of the internal dielectric constant and ionic strength.

Continuum electrostatic calculations have been shown to depend on the choice of the internal dielectric constant describing the macromolecular environment. Early calculations used a low dielectric constant, typically in the range 1-4, while more recent studies indicate that a larger constant is needed to reproduce experimental shifts in pK_a values[56,57]. The nature of the protein dielectric constant has been discussed extensively in the literature (see e.g. Warshel *et al.*[19]). van Vlijimen *et al.*[57] studied pK_a shifts using continuum calculations on 100 MD generated structures, and found that a protein dielectric constant of 20 resulted in best agreement with experimental data. It is not our intention to derive or find optimal values of this constant, but its effect on the strength of the ion pairs in UDG and on our results has, however, been assessed. In order to investigate how the protein dielectric constant (ε_p) affects the computed electrostatic interactions it assumes three different values: 4, 10 and 20. The results from these continuum calculations are summarized in Table III. The individual contributions from each ion-pair in all examined conditions are given in the Supplementary Material (Tables I-XII). Focusing on the $\Delta\Delta G_{tot}$ values presented in Table III, it is

immediately clear that increasing the dielectric constant in the protein effectively reduces the electrostatic stability of the ion-pairs. $\Delta\Delta G_{tot}$ is identical in cUDG and hUDG when ε_p=4, though with different individual contributions. The more favorable electrostatic interactions between the two partners ($\Delta\Delta G_{brd}$) and between the ion-pairs and the surroundings ($\Delta\Delta G_{prt}$) is outweighed by the higher desolvation penalty for cUDG as compared to hUDG (Table III). When the protein dielectric constant increases to 10 and 20, $\Delta\Delta G_{tot}$ becomes 0.5 kcal/mol more stabilizing in the cold-active enzyme at 298 K. The electrostatic stability is on average –2.3 and –2.1 kcal/mol with ε_p=4 for ion-pairs in cUDG and hUDG, respectively. The corresponding values with ε_p=10 are –1.6 and –1.5 kcal/mol and for ε_p=20 –1.4 and –1.2 kcal/mol for cUDG and hUDG respectively. We can use ±1.0 kcal/mol as a threshold to discriminate stabilizing pairs from those with small or negligible contributions. Among all the pairs, five have a stability of less than the threshold in hUDG while only one is observed in cUDG independent of the ε_p used. Thus, increasing the dielectric constant from 4 to 20 have negligible effect on the relative contribution of electrostatics in ion-pairs, and pairs with little or no effect from electrostatics is identified with all three ε_p.

Ionic strength (100 mM) does not affect the results appreciably and only serves to reduce the total stability, but not relatively between cUDG and hUDG. The total electrostatic contribution to the stability of the ion pairs is reduced by 1.1-1.2 kcal/mol with all three protein dielectric constants used. Presence of 100 mM salt decreases the desolvation penalty and the interaction between the ion pair and the rest of the protein, while the bridging term is increased (Table III). Others have also found that the ionic strength plays a minor role in relative values[47]. Ionic strength will of course play an important role when dealing with proteins that are exposed to extreme salinity (< 2 M)[58]. The enzymes studied here are also intracellular proteins and thus not exposed to extreme salinities. Since we are interested in the relative effect of ion-pairs between cUDG and hUDG, ionic strength will not be discussed in more detail.

Temperature and its effect on the electrostatic stability of ion-pairs.

Irrespective of the choice of ε_p comparable contributions to stability from salt-bridges are found in the two protein homologues at 298 K. The physical properties of water changes with temperature, and the dielectric constant of bulk water changes from 87.90 to 55.51 when the temperature is increased from 273 K to 373 K[17]. In order to investigate how the change in

dielectric properties of the solvent influence the stabilities, the calculations with ε_p=20 was reevaluated with ε_w=87.90 and 75.90 for cUDG and hUDG, respectively. Electrostatic interactions will accordingly become more favorable at higher temperatures due to this reduced screening effect, without considering other favorable or unfavorable contributions. The difference in the total electrostatic stability of ion pairs between cUDG and hUDG is largest with ε_p=10 and 20, and the results obtained with ε_p=20 are therefore expected to be valid for ε_p=4 and 10 as well. Table IV and V show the results from these calculations. The total electrostatic stability of the ion pairs is now approximately +3.1 kcal/mol more favorable in hUDG when compared to cUDG. The enhanced stability is almost exclusively attributed to differences in the bridging term, which is 3.6 kcal/mol more favorable in the mesophilic enzyme.

Location of ion pairs plays an important role in temperature adaptation.

Several different factors may influence the stability of ion-pairs, including the structural location, whether they are internal or external, local or global and the surrounding environment. The accessible surface area of the constituent residues has been calculated using a water probe of 1.4 Å[59] to discriminate between internal and external salt-bridges (Table II). If the average ASA of the two residues forming the pair is less than 20 % then the pair is classified as buried in the protein core (internal), otherwise they are defined as external. Only the conserved Glu112-Arg210 interaction is buried in cUDG and hUDG, and similar energetics is observed for this particular bridge (Table IV and V). The average ASA of all ion-pairs is 39 Å^2 and 50 Å^2 for cUDG and hUDG, respectively, and the desolvation penalty also reflects this fact (Table III, IV and V). The study of Hendsch and Tidor[47] suggest that ion pairs and hydrogen bonds destabilize the folded state due to loss of hydration at room temperature, an effect which is not fully compensated for by improved Coulombic interactions in the folded state. Our calculations show, however, that the desolvation penalty is compensated for by favorable Coulombic interactions in the folded state for most of the bridges in the two UDG enzymes (Table III). Only three ion pairs in hUDG have positive $\Delta\Delta G_{tot}$ when the internal dielectric constant is set to 4 (Supplementary Material, Table II), but $\Delta\Delta G_{tot}$ becomes negative for all three when ε_p is increased to 10 or 20. While the absolute value of all contributions to $\Delta\Delta G_{tot}$ is reduced when ε_p increases, Table III also shows that the $\Delta\Delta G_{dslv}$ is most affected.

Ion-pairs can be classified as either local or global depending on the number of residues that are between the two partners in the sequence. If five or less amino acids are between the two, the pair is regarded as local, otherwise a global bridge is formed. The psychrophilic enzyme has three salt-bridges between partners that are far apart in sequence (global salt-bridges) and all three link helices together. The mesophilic enzyme contains five global salt-bridges, the corresponding three observed in the psychrophilic enzyme and in addition two unique ones (Table II). The unique interactions are: Asp183-Lys302 and Glu197-Lys303. The majority of the ion-pairs are local (Table II), which is consistent with the general properties of salt-bridges in monomeric proteins[49]. These findings are also in agreement with the hierarchical view of how protein folding proceeds[60,61]. The energetics of local and global salt-bridges show an interesting trend, global bridges are more stabilized by electrostatics when compared to local. Global bridges have on average 0.3 to 1.5 kcal/mol more negative $\Delta\Delta G_{tot}$ values, and they are slightly more favorable in cUDG at 298 K. One of the two unique pairs observed in hUDG (Asp183-Lys302) is a highly stable salt-bridge, while the other (Lys197-Glu303) is destabilizing or weakly stabilizing. The three conserved pairs are of comparable stability in the cold- and warm-active UDG. If we take the respective physiological temperature into account, global ion-pairs are more stabilizing in hUDG while local pairs are of comparable stability (Table IV and V).

All but two salt-bridges in hUDG have a favorable interaction with the rest of the protein ($\Delta\Delta G_{prt}$, Supporting Material, Table II, IV and VI). The Lys197-Glu303 and Glu289-Lys293 are located at the protein surface and the side-chains extend into solution, which can explain the unfavorable $\Delta\Delta G_{prt}$. The local environment is generally more stabilizing in the psychrophilic enzyme regardless of the protein dielectric constant used. $\Delta\Delta G_{ass}$ which reflects the stability of ion-pairs in the absence of charges in the rest of the protein is more positive in cUDG when compared to hUDG. Together, this indicates more optimized dipoles and/or location of charges in the vicinity of ion-pair residues in cUDG[24]. Other studies have also shown that nearby charged residues can give rise to a strongly favorable $\Delta\Delta G_{prt}$ [23].

Putative ionic networks.

Our structural inspection revealed the presence of one ionic network in both hUDG and cUDG, each with three members (Table II). Networks of ionic interactions occur when more than two ionic residues interact, and an increased occurrence has been suggested to be

essential in explaining the enhanced thermal stability of thermophilic proteins[21,22]. Usher *et al.*[62] did not find a correlation between stability and the number of ion pairs when investigating the CheY protein. Lack of correlation between electrostatic contribution to the stability of hyperthermophilic proteins and the number of charged groups, ion pairs or ion pair networks was also observed by Xiao and Honig[22]. Location of the ionizable residues within the protein structures is, however, dictating whether or not specific interactions are stabilizing. Asp227 is interacting with Lys231 and Lys252 in cUDG (Fig. 3A) and their net electrostatic stability is –2.14 kcal/mol, while Lys302 interacts with Asp183 and Asp300 (Fig. 3B) with a net stability of –4.91 kcal/mol in hUDG (Table IV and V). The ionic network in hUDG is thus far more stabilizing than the one observed in cUDG, and this difference may be of structural importance. The network observed in hUDG appears to have a more important role, as it stabilizes the C-terminal (Fig. 3B), while the network in cUDG is between two helices.

Concluding remarks

We have here investigated the electrostatic stability of ion-pairs in the context of temperature adaptation of cold- and warm-active uracil DNA glycosylase by using MD simulations and continuum electrostatics calculations. The MD generated structural ensemble of cUDG differs significantly from the crystal structure with respect to the number of ion-pairs, which is primarily attributed to crystallization conditions. Seven new salt-bridges are formed during the simulations, while for its mesophilic counterpart only one more is found. Consequently, statistical analysis of the number of ion-pairs in crystal structures of proteins should be interpreted with care. Our comparison of crystal structures of UDG adds further support to this. It is not only lack of information on protein dynamics from crystal structures, but also effects from different crystallization conditions can cause differences in ion-pairs. Conformational ensembles obtained from NMR experiments have a significant degree of overlap with MD generated ensembles (see e.g. Kumar and Nussinov[52]), and NMR ensembles can give more information on protein dynamics.

The strength and the number of ion-pairs present in a protein and whether they are involved in networks or not are important for the overall structural stability. Their presence can have a large impact on the structural integrity modulating molecular plasticity. The main hypothesis for cold-adaptation is that increased structural flexibility is the cause of the

observed enhanced catalytic efficiency and reduced thermal stability of cold-active enzymes when compared to their warm-active cousins. Based on chemical intuition, ion-pairs should have a stabilizing effect on the protein and favor the folded state. Virtually all ion-pairs present in cUDG and hUDG have a favorable electrostatic contribution possibly leading to increased structural stability. When it comes to their net contribution to protein stability, entropy must also be considered. Strong interactions will reduce the available conformational space and the conformational entropy of the folded state will consequently decrease. Hence, ion-pairs destabilize the native state from an entropic point of view, whereas they are enthalpically stabilizing. The entropic effect is intrinsically difficult to estimate through computer simulations, and the fact that it changes with temperature makes it even more challenging to calculate. Enthalpic contributions also change with temperature, and we have investigated how the change in the solvent dielectric constant influences the stability of ion-pairs. Without taking the living temperature of the organisms into considerations, we find comparable electrostatic stability of ion-pairs in cUDG and hUDG. Proteins exhibit generally only marginal stability in their natural environment due to the delicate balance between stabilizing and destabilizing interactions. Medium sized globular proteins have average values for the Gibbs free energy of stabilization on the order of 50 kJ/mol[63], which is equivalent to only a small number of weak non-covalent interactions[11,12,64]. If we look at the respective environmental temperatures, the ion-pairs are more stabilizing in hUDG when compared to cUDG. The mesophilic enzyme also has more global bridges and a strong ionic network in the C-terminal, which may contribute to the enhanced thermal stability of hUDG.

Insight into the structural features promoting thermal stability is of great importance, particularly in design of enzymes with altered kinetic and thermodynamic properties. UDG from cod is presently commercialized, and its main application area is in PCR experiments removing contamination from previous PCR amplifications. cUDG is completely and irreversible inactivated by moderate heat treatment. The present study identifies some key residues involved in stabilization through ion pair network formation. In order to investigate the role of these in further detail mutagenesis combined with kinetic, thermodynamic and structural studies is needed.

Acknowledgement

Funding 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. Siddiqui KS, Cavicchioli R. Cold-adapted enzymes. Annu Rev Biochem 2006;75:403-433.
2. Smalås AO, Leiros HK, Os V, Willassen NP. Cold adapted enzymes. Biotechnol Annu Rev 2000;6:1-57.
3. D'Amico S, Collins T, Marx JC, Feller G, Gerday C. Psychrophilic microorganisms: challenges for life. Embo Rep 2006;7:385-389.
4. Gianese G, Bossa F, Pascarella S. Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes. Proteins 2002;47:236-249.
5. Feller G. Molecular adaptations to cold in psychrophilic enzymes. Cell Mol Life Sci 2003;60:648-662.
6. Georlette D, Blaise V, Collins T, D'Amico S, Gratia E, Hoyoux A, Marx JC, Sonan G, Feller G, Gerday C. Some like it cold: biocatalysis at low temperatures. FEMS Microbiol Rev 2004;28:25-42.
7. Gerday C, Aittaleb M, Bentahir M, Chessa JP, Claverie P, Collins T, D'Amico S, Dumont J, Garsoux G, Georlette D, Hoyoux A, Lonhienne T, Meuwis MA, Feller G. Cold-adapted enzymes: from fundamentals to biotechnology. Trends Biotechnol 2000;18:103-107.
8. Russell RJM, Gerike U, Danson MJ, Hough DW, Taylor GL. Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. Structure 1998;6:351-361.
9. Bosshard HR, Marti DN, Jelesarov I. Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. J Mol Recognit 2004;17:1-16.
10. Jaenicke R. Protein Folding - Local Structures, Domains, Subunits, and Assemblies. Biochemistry 1991;30:3147-3161.
11. Jaenicke R. Stability and stabilization of globular proteins in solution. J Biotechnol 2000;79:193-203.
12. Dill KA. Dominant Forces in Protein Folding. Biochemistry 1990;29:7133-7155.
13. Pace CN. Polar group burial contributes more to protein stability than nonpolar group burial. Biochemistry 2001;40:310-313.
14. Brandsdal BO, Smalås AO, Åqvist J. Electrostatic effects play a central role in cold adaptation of trypsin. FEBS Lett 2001;499:171-175.
15. Gorfe AA, Brandsdal BO, Leiros HKS, Helland R, Smalås AO. Electrostatics of mesophilic and psychrophilic trypsin isoenzymes: Qualitative evaluation of electrostatic differences at the substrate binding site. Proteins: Struct, Funct, Genet 2000;40:207-217.
16. Moe E, Leiros I, Riise EK, Olufsen M, Lanes O, Smalas A, Willassen NP. 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.

17. Kumar S, Nussinov R. Different roles of electrostatics in heat and in cold: Adaptation by citrate synthase. Chembiochem 2004;5:280-290.
18. Dominy BN, Minoux H, Brooks CL. An electrostatic basis for the stability of thermophilic proteins. Proteins 2004;57:128-141.
19. Warshel A, Sharma PK, Kato M, Parson WW. Modeling electrostatic effects in proteins. Bba-Proteins Proteom 2006;1764:1647-1676.
20. Elcock AH. The stability of salt bridges at high temperatures: Implications for hyperthermophilic proteins. J Mol Biol 1998;284:489-502.
21. Goldman A. How to make my blood boil. Structure 1995;3:1277-1279.
22. Xiao L, Honig B. Electrostatic contributions to the stability of hyperthermophilic proteins. J Mol Biol 1999;289:1435-1444.
23. Kumar S, Nussinov R. Relationship between ion pair geometries and electrostatic strengths in proteins. Biophys J 2002;83:1595-1612.
24. Hwang JK, Warshel A. Why Ion-Pair Reversal by Protein Engineering Is Unlikely to Succeed. Nature 1988;334:270-272.
25. Lindahl T, Nyberg B. Heat-Induced Deamination of Cytosine Residues in Deoxyribonucleic-Acid. Biochemistry 1974;13:3405-3410.
26. Mol CD, Arvai AS, Slupphaug G, Kavli B, Alseth I, Krokan HE, Tainer JA. Crystal-Structure and Mutational Analysis of Human Uracil-DNA Glycosylase - Structural Basis for Specificity and Catalysis. Cell 1995;80:869-878.
27. Leiros I, Moe E, Lanes O, Smalås AO, Willassen NP. The structure of uracil-DNA glycosylase from Atlantic cod (Gadus morhua) reveals cold-adaptation features. Acta Crystallogr D 2003;59:1357-1365.
28. 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.
29. Ravishankar R, Sagar MB, 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.
30. Geoui T, Buisson M, Tarbouriech N, Burmeister WP. 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.
31. Lanes O, Guddal PH, Gjellesvik DR, Willassen NP. Purification and characterization of a cold-adapted uracil-DNA glycosylase from Atlantic cod (Gadus morhua). Comp Biochem Physiol 2000;127:399-410.
32. Lanes O, Leiros I, Smalås AO, Willassen NP. 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.
33. Olufsen M, Brandsdal BO, Smalas AO. Comparative unfolding studies of psychrophilic and mesophilic uracil DNA glycosylase: MD simulations show reduced thermal stability of the cold-adapted enzyme. J Mol Graph Model 2006.
34. Olufsen M, Smalås AO, Moe E, Brandsdal BO. 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.
35. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, 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.

36. Wang JM, Cieplak P, Kollman PA. 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.
37. Dinner AR, Blackburn GM, Karplus M. Uracil-DNA glycosylase acts by substrate autocatalysis. Nature 2001;413:752-755.
38. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. J Chem Phys 1983;79:926-935.
39. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-Dynamics with Coupling to an External Bath. J Chem Phys 1984;81:3684-3690.
40. 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.
41. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. J Comput Phys 1977;23:327-341.
42. Vriend G. What If - a Molecular Modeling and Drug Design Program. J Mol Graph 1990;8:52-56.
43. 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.
44. 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.
45. Sitkoff D, Sharp KA, Honig B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. J Phys Chem 1994;98:1978-1988.
46. Moreira IS, Fernandes PA, Ramos MJ. Accuracy of the numerical solution of the Poisson-Boltzmann equation. J Mol Struc-Theochem 2005;729:11-18.
47. Hendsch ZS, Tidor B. Do Salt Bridges Stabilize Proteins - a Continuum Electrostatic Analysis. Protein Sci 1994;3:211-226.
48. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. Protein Data Bank - Computer-Based Archival File for Macromolecular Structures. J Mol Biol 1977;112:535-542.
49. Kumar S, Nussinov R. Salt bridge stability in monomeric proteins. J Mol Biol 1999;293:1241-1255.
50. Barlow DJ, Thornton JM. Ion-Pairs in Proteins. J Mol Biol 1983;168:867-885.
51. Papaleo E, Olufsen M, De Gioia L, Brandsdal BO. Optimization of electrostatics as a strategy for cold-adaptation: A case study of cold- and warm-active elastases. J Mol Graph Model 2006.
52. Kumar S, Nussinov R. Fluctuations in ion pairs and their stabilities in proteins. Proteins: Struct, Funct, Genet 2001;43:433-454.
53. Masunov A, Lazaridis T. Potentials of mean force between ionizable amino acid side chains in water. J Am Chem Soc 2003;125:1722-1730.
54. Honig B, Nicholls A. Classical Electrostatics in Biology and Chemistry. Science 1995;268:1144-1149.
55. Lounnas V, Wade RC. Exceptionally stable salt bridges in cytochrome P450cam have functional roles. Biochemistry 1997;36:5402-5417.
56. Schutz CN, Warshel A. What axe the dielectric "constants" of proteins and how to validate electrostatic models? Proteins: Struct, Funct, Genet 2001;44:400-417.

57. van Vlijmen HWT, Schaefer M, Karplus M. Improving the accuracy of protein pK(a) calculations: Conformational averaging versus the average structure. Proteins: Struct, Funct, Genet 1998;33:145-158.
58. Elcock AH, McCammon JA. Electrostatic contributions to the stability of halophilic proteins. J Mol Biol 1998;280:731-748.
59. Lee B, Richards FM. Interpretation of Protein Structures - Estimation of Static Accessibility. J Mol Biol 1971;55:379-380.
60. Baldwin RL, Rose GD. Is protein folding hierarchic? II. Folding intermediates and transition states. Trends Biochem Sci 1999;24:77-83.
61. Tsai CJ, Kumar S, Ma BY, Nussinov R. Folding funnels, binding funnels, and protein function. Protein Sci 1999;8:1181-1190.
62. Usher KC, De la Cruz AFA, Dahlquist FW, Swanson RV, Simon MI, Remington SJ. Crystal structures of CheY from Thermotoga maritima do not support conventional explanations for the structural basis of enhanced thermostability. Protein Sci 1998;7:403-412.
63. Pfeil W. Protein Stability and Folding: A Collection of Thermodynamic Data. Berlin: Springer; 1998.
64. Jaenicke R. Protein Stability and Molecular Adaptation to Extreme Conditions. Eur J Biochem 1991;202:715-728.
65. Parikh SS, Mol CD, Slupphaug G, Bharati S, Krokan HE, Tainer JA. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. EMBO J 1998;17:5214-5226.
66. Parikh SS, Walcher G, Jones GD, Slupphaug G, Krokan HE, Blackburn GM, Tainer JA. 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.
67. Slupphaug G, Mol CD, Kavli B, Arvai AS, Krokan HE, Tainer JA. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. Nature 1996;384:87-92.
68. DeLano WL. The PyMOL Molecular Graphics System: DeLano Scientific, San Carlos, CA, USA; 2002.

Tables

Table I
Salt-bridges and their distances[a] (Å) observed in crystal structures of cUDG and hUDG.

	cUDG	**V171E[b]**	**hUDG**	**E171V[c]**	**hUDG[d]**	**hUDG[d]**	**hUDG[d]**
PDB-code	1OKB	-	1AKZ	1YUO	1SSP	1EMJ	4SKN
Glu87-Lys90				4.30	3.78		4.06
Arg91-Glu92	4.39						
Glu96-Lys99	3.52	3.61	3.49	3.64	3.60	3.60	3.78
Glu111-Lys114			4.08			4.11	
Asp111-Arg115		4.36					
Glu112-Arg210	3.95	4.14	4.34	4.05	4.13	4.12	4.95
Asp133-Lys136				4.45			
Lys135-Asp136			4.85	3.86			
Lys138-Glu297		3.32					
Asp180-Lys/Arg282[e]		3.82	4.42	4.18	4.56	4.65	
Asp183-Lys302			4.16	4.28		4.94	5.00
Lys218-Glu219			4.63				
Asp227-Lys252	3.51	3.56	3.66	3.34	3.59	3.53	3.50
Asp257-Lys259							4.85
Asp257-Arg260						4.43	
Glu289-Lys293			4.34	4.25			
Asp300-Lys302			3.45	3.38			
Average distance	**3.84**	**3.80**	**4.14**	**3.97**	**3.93**	**4.20**	**4.36**
Number of salt-bridges	**4**	**6**	**10**	**10**	**5**	**7**	**6**

[a]The distances between the following atoms has been measured: (Lys:Nζ, Arg:Cζ and His:Cε1) and (Asp:Cγ and Glu:Cδ), and only pairs with a distance less than 5 Å is reported.

[b]cUDG with one mutant: V171E[16], no pdb code available

[c]hUDG with one mutant: E171V[16]

[d]hUDG in complex with different DNA fragments[65-67]

[e]Residue 282 is a Lys in cUDG and a Arg in hUDG.

Table II
Salt-bridges observed in the crystal structure and in the MD simulations[a] of cUDG and hUDG.

	cUDG				hUDG			
Salt-bridge	Cryst[b] (Å)	MD[b] (Å)	Occupancy[c] (%)	ASA[d] (Å^2)	Cryst[b] (Å)	MD[b] (Å)	Occupancy[c] (%)	ASA[d] (Å^2)
Glu87-Arg90	5.04	5.26	35	59				
Arg91-Glu92	4.39	4.36	92	46				
Glu96-Lys99	3.52	3.54	99	33	3.49	3.54	100	37
Glu111-Lys114					4.08	5.62	41	64
Asp111-Arg115	5.09	3.96	100	52				
Glu112-Arg210	3.95	3.99	100	16	4.34	4.02	100	19
Asp133-Lys135						6.33	40	55
Lys135-Asp136					4.85	-		
Asp136-Lys138	7.97	4.70	56	45				
Asp145-His148	7.10	5.35	65	29				
Asp180-Lys/Arg282*	5.03	4.26	82	30	4.42	4.32	99	31
Asp183-Lys302					4.16	3.85	90	51
Lys197-Glu303					12.60	6.32	36	58
Lys218-Glu219					4.63	4.81	60	65
Asp227-Lys231	5.24	5.82	49	45				
Asp227-Lys252	3.51	4.25	88	23	3.66	4.25	83	41
Asp257-Lys259	7.32	4.46	76	54	7.44	6.10	39	63
Glu289-Lys293					4.34	5.85	44	66
Asp300-Lys302					3.45	3.57	98	46
Average	5.29	4.63	77	39	5.12	4.88	69	50

[a]The MD distance is the average calculated using all simulated structures.

[b]The distances between the following atoms has been measured: (Lys:Nζ, Arg:Cζ and His:Cε1) and (Asp:Cγ and Glu:Cδ).

[c]Occupancy is percentage of the structures where the salt-bridge distance is less than 5 Å.

[d]Accessible surface area (ASA) is the accessible surface area calculated using a probe of 1.4 Å, and is the average of the two constituent residues.

Table III
Stabilities[a] of the salt-bridges observed in the MD simulations[b] of cUDG and hUDG

Protein	Salt (mM)	Dielec (prot)	$\Delta\Delta G_{dslv}$[c] kcal/mol	$\Delta\Delta G_{brd}$[d] kcal/mol	$\Delta\Delta G_{prt}$[e] kcal/mol	$\Delta\Delta G_{tot}$[f] kcal/mol	$\Delta\Delta G_{ass}$[g] kcal/mol
cUDG	0	4	71.1	-54.3	-42.5	-25.7	16.8
hUDG	0	4	58.5	-50.7	-33.5	-25.7	7.8
cUDG	0	10	29.6	-29.4	-18.1	-18.0	0.1
hUDG	0	10	24.5	-28.1	-13.9	-17.5	-3.6
cUDG	0	20	14.5	-20.2	-9.6	-15.3	-5.8
hUDG	0	20	12.2	-19.9	-7.1	-14.8	-7.7
cUDG	100	4	70.5	-51.5	-43.4	-24.4	19.0
hUDG	100	4	57.9	-47.6	-34.8	-24.5	10.3
cUDG	100	10	29.1	-26.8	-19.1	-16.8	2.3
hUDG	100	10	24.1	-25.2	-15.1	-16.3	-1.4
cUDG	100	20	14.1	-17.8	-10.6	-14.2	-3.6
hUDG	100	20	11.8	-17.1	-8.4	-13.7	-5.3

[a]The dielectric of the water was set to 80 in all calculations.

[b]The ion-pairs included in these calculations are given in Table 2, and their individual contributions are presented in the Supplementary material.

[c] $\Delta\Delta G_{dslv}$ is the desolvation penalty for the charged residues.

[d] $\Delta\Delta G_{brd}$ is the bridge energy for the electrostatic interaction between the pairing residues.

[e] $\Delta\Delta G_{prt}$ is the protein energy term for the electrostatic interaction between the ion-pair and the surrounding protein environment.

[f] $\Delta\Delta G_{tot}$ is the sum of these three terms.

[g] $\Delta\Delta G_{ass}$ is the sum of the $\Delta\Delta G_{dslv}$ and the $\Delta\Delta G_{brd}$.

Table IV

Average electrostatic free energy contributions[a] to salt-bridges[b] in cUDG at 273 K.

Salt-bridge	Occupied[d] %	ASA[e] Å^2	$\Delta\Delta G_{dslv}$[f] kcal/mol	$\Delta\Delta G_{brd}$[g] kcal/mol	$\Delta\Delta G_{prt}$[h] kcal/mol	$\Delta\Delta G_{tot}$[c,i] kcal/mol	$\Delta\Delta G_{ass}$[j] kcal/mol
Glu87-Arg90	35	59	0.74	-0.66	-0.98	-0.90 (0.37)	0.07
Arg91-Glu92	92	46	1.01	-1.81	-0.33	-1.13 (0.31)	-0.80
Glu96-Lys99	99	33	1.30	-1.99	-0.71	-1.41 (0.22)	-0.70
Asp111-Arg115	100	52	1.00	-1.96	-0.17	-1.12 (0.17)	-0.96
Glu112-Arg210	100	16	1.75	-2.60	-0.41	-1.26 (0.18)	-0.85
Asp136-Lys138	56	45	1.00	-1.22	-0.73	-0.94 (0.27)	-0.21
Asp145-His148	65	29	1.27	-1.58	-0.44	-0.76 (0.28)	-0.32
Asp180-Lys282	82	30	1.80	-1.76	-2.05	-2.01 (0.42)	0.04
Asp227-Lys231	49	45	1.00	-1.15	-1.06	-1.21 (0.38)	-0.15
Asp227-Lys252	88	23	1.56	-1.31	-1.18	-0.93 (0.37)	0.24
Asp257-Lys259	76	54	0.98	-1.65	-0.68	-1.35 (0.41)	-0.67
Sum	-	-	13.41	-17.70	-8.73	-13.03 (1.06)	-4.29
Average	77	39	1.22	-1.61	-0.79	-1.18 (0.31)	-0.39

[a]The dielectric constant of water and protein was 87.90 and 20, respectively.

[b]All salt-bridges that were inside a radii of 5 Å for 30 percent of the simulation time is included (Table 2).

[c]The standard deviations for the average values are indicated in parenthesis for the total energy.

[d]Occupancy is percentage of the structures where the salt-bridge distance is less than 5 Å.

[e]Accessible surface area (ASA) is the accessible surface area calculated using a probe of 1.4 Å, and is the average of the two constituent residues.

[f] $\Delta\Delta G_{dslv}$ is the desolvation penalty for the charged residues.

[g] $\Delta\Delta G_{brd}$ is the bridge energy for the electrostatic interaction between the pairing residues.

[h] $\Delta\Delta G_{prt}$ is the protein energy term for the electrostatic interaction between the ion-pair and the surrounding protein environment.

[i] $\Delta\Delta G_{tot}$ is the sum of these three terms.

[j] $\Delta\Delta G_{ass}$ is the sum of the $\Delta\Delta G_{dslv}$ and the $\Delta\Delta G_{brd}$.

Table V
Average electrostatic free energy contributions[a] to salt-bridges[b] in hUDG at 310 K.

Salt-bridge	Occupied[d] %	ASA[e] Å^2	$\Delta\Delta G_{dslv}$[f] kcal/mol	$\Delta\Delta G_{brd}$[g] kcal/mol	$\Delta\Delta G_{prt}$[h] kcal/mol	$\Delta\Delta G_{tot}$[c,i] kcal/mol	$\Delta\Delta G_{ass}$[j] kcal/mol
Glu96-Lys99	100	37	1.23	-2.34	-0.38	-1.50 (0.27)	-1.11
Glu111-Lys114	41	64	0.50	-1.20	-0.24	-0.94 (0.45)	-0.70
Glu112-Arg210	100	19	1.87	-3.02	-0.52	-1.67 (0.28)	-1.15
Asp133-Lys135	40	55	1.21	-1.49	-0.99	-1.28 (0.45)	-0.28
Asp180-Arg282	99	31	1.89	-2.48	-1.78	-2.37 (0.85)	-0.59
Asp183-Lys302	90	51	1.03	-2.04	-1.46	-2.47 (0.27)	-1.01
Lys197-Glu303	36	58	0.53	-1.01	0.10	-0.38 (0.42)	-0.48
Lys218-Glu219	60	65	0.53	-1.33	-0.01	-0.81 (0.42)	-0.80
Asp227-Lys252	83	41	1.29	-1.69	-0.11	-0.51 (0.42)	-0.39
Asp257-Lys259	39	63	0.78	-1.35	-0.71	-1.28 (0.42)	-0.57
Glu289-Lys293	44	66	0.47	-1.02	0.04	-0.50 (0.47)	-0.55
Asp300-Lys302	98	46	1.26	-2.33	-1.37	-2.44 (0.31)	-1.07
Sum	-	-	12.60	-21.30	-7.43	-16.13 (1.54)	-8.70
Average	69	50	1.05	-1.78	-0.62	-1.34 (0.42)	-0.73

[a]The dielectric constant of water and protein was 75.90 and 20, respectively.

[b]All salt-bridges that were inside a radii of 5 Å for 30 percent of the simulation time is included (Table 2).

[c]The standard deviations for the average values are indicated in parenthesis for the total energy.

[d]Occupancy is percentage of the structures where the salt-bridge distance is less than 5 Å.

[e]Accessible surface area (ASA) is the accessible surface area calculated using a probe of 1.4 Å, and is the average of the two constituent residues.

[f] $\Delta\Delta G_{dslv}$ is the desolvation penalty for the charged residues.

[g] $\Delta\Delta G_{brd}$ is the bridge energy for the electrostatic interaction between the pairing residues.

[h] $\Delta\Delta G_{prt}$ is the protein energy term for the electrostatic interaction between the ion-pair and the surrounding protein environment.

[i] $\Delta\Delta G_{tot}$ is the sum of these three terms.

[j] $\Delta\Delta G_{ass}$ is the sum of the $\Delta\Delta G_{dslv}$ and the $\Delta\Delta G_{brd}$.

Figures legends

Fig. 1

The r.m.s.f. of the Cα atoms as a function of time during the simulation of cUDG (grey) and hUDG (black).

Fig. 2

Salt-bridges present for more than 30 % of the MD generated structures of cUDG (A) and hUDG (B). The figure was generated using Pymol[68].

Fig. 3

The salt-bridge network in cUDG (A) and hUDG (B). The figure was generated using Pymol[68].

Figures

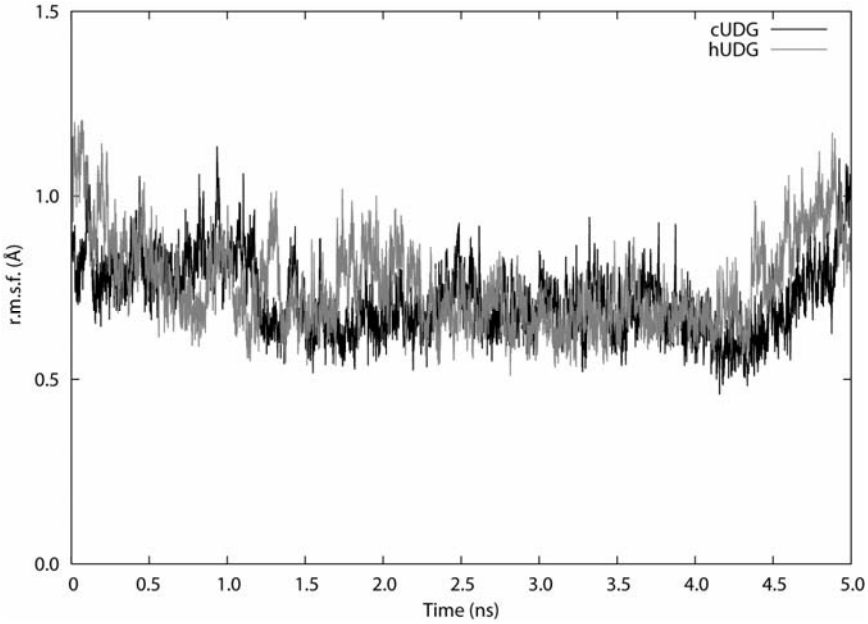


Fig. 1

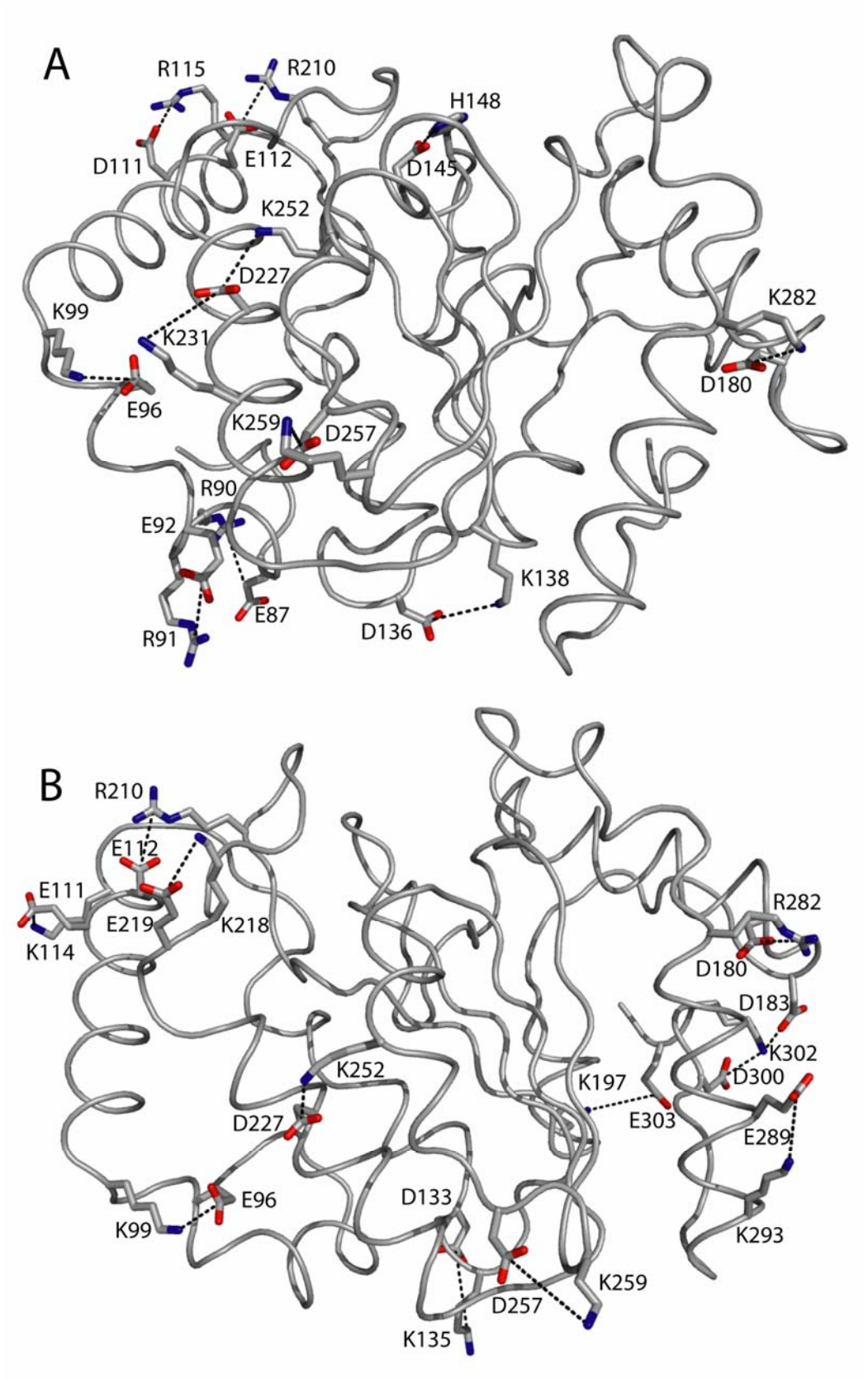

A
R115
R210
H148
D111
E112
D145
K252
D227
K282
K99
K231
E96
D180
K259
D257
R90
E92
K138
E87
R91
D136
B
R210
E112
E111
E219
K218
K114
R282
D180
D183
K302
K252
K197
D300
D227
E303
E289
E96
D133
K99
K293
K259
D257
K135

Fig. 2

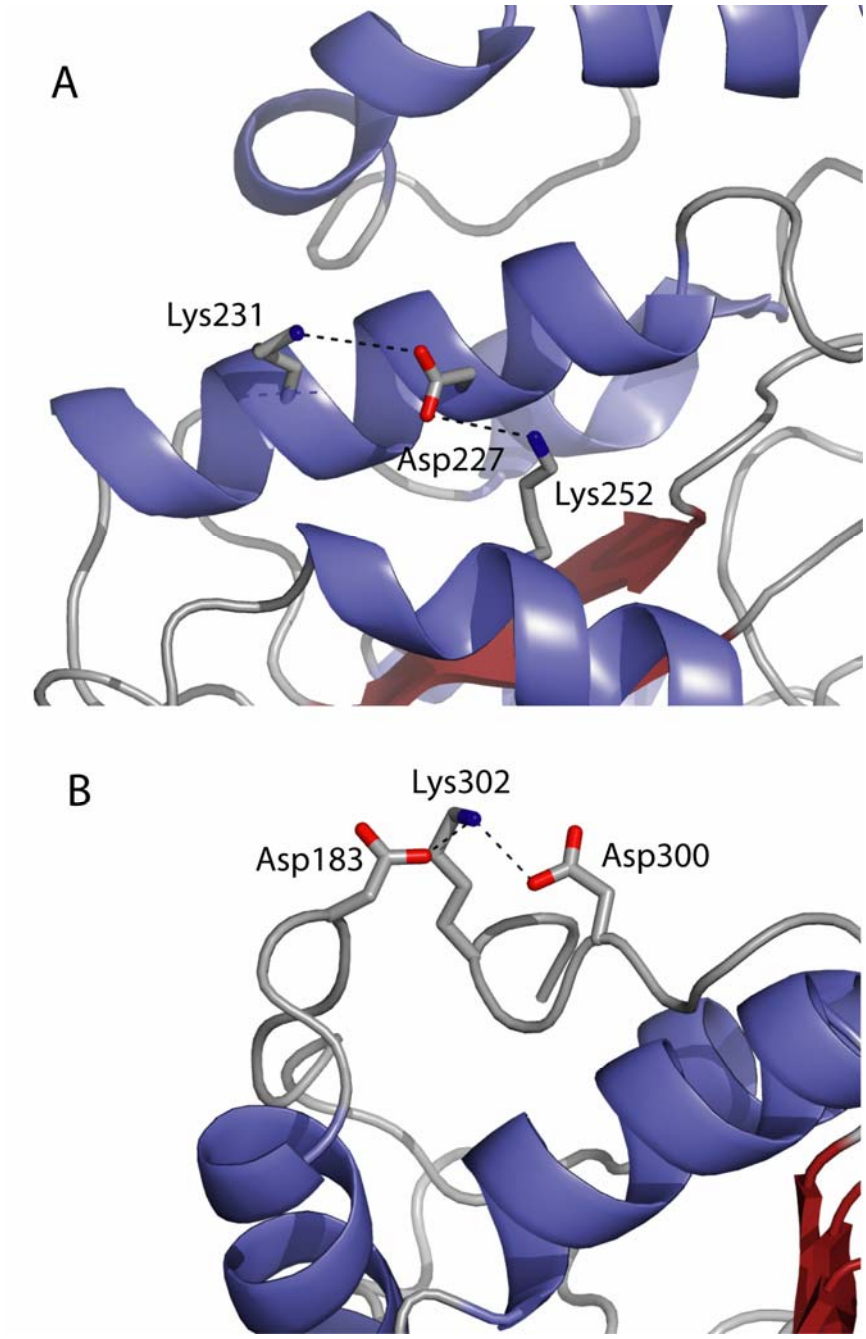


Fig. 3

Supporting information

Information for all the tables in supporting information. All salt-bridges that were inside a radii of 5 Å for 30 percent of the simulation time is included in the tables. The dielectric constant of water was set to 80. The standard deviations for the average values are indicated in parenthesis for the total energy. $\Delta\Delta G_{dslv}$ is the desolvation penalty for the charged residues, $\Delta\Delta G_{brd}$ is the bridge energy for the electrostatic interaction between the pairing residues, $\Delta\Delta G_{prt}$ is the protein energy term for the electrostatic interaction between the ion-pair and the surrounding protein environment and $\Delta\Delta G_{tot}$ is the sum of these three terms. $\Delta\Delta G_{ass}$ is the sum of the $\Delta\Delta G_{dslv}$ and the $\Delta\Delta G_{brd}$. Occupancy is percentage of the structures where the salt-bridge distance is less than 5 Å. Accessible surface area (ASA) is the accessible surface area calculated using a probe of 1.4 Å, and is the average of the two constituent residues.

Table I

Average electrostatic free energy contributions to salt-bridges in cUDG with $\varepsilon_p = 4$ and zero salt added.

Salt-bridge	**Occupied** %	**ASA** Å^2	**$\Delta\Delta G_{dslv}$** kcal/mol	**$\Delta\Delta G_{brd}$** kcal/mol	**$\Delta\Delta G_{prt}$** kcal/mol	**$\Delta\Delta G_{tot}$** kcal/mol	**$\Delta\Delta G_{ass}$** kcal/mol
Glu87-Arg90	35	59	3.97	-1.05	-4.55	-1.63 (1.23)	2.92
Arg91-Glu92	92	46	5.42	-5.87	-1.11	-1.56 (1.18)	-0.45
Glu96-Lys99	99	33	6.54	-5.93	-3.65	-3.04 (0.89)	0.61
Asp111-Arg115	100	52	5.72	-6.59	-1.13	-1.99 (0.73)	-0.87
Glu112-Arg210	100	16	9.85	-9.89	-2.86	-2.90 (0.87)	-0.04
Asp136-Lys138	56	45	4.65	-2.63	-3.03	-1.01 (0.96)	2.02
Asp145-His148	65	29	6.16	-5.09	-1.98	-0.91 (1.20)	1.08
Asp180-Lys282	82	30	10.81	-5.96	-10.77	-5.92 (1.96)	4.85
Asp227-Lys231	49	45	4.80	-2.85	-4.13	-2.18 (1.73)	1.95
Asp227-Lys252	88	23	7.83	-3.45	-5.80	-1.42 (1.51)	4.38
Asp257-Lys259	76	54	5.34	-4.96	-3.51	-3.12 (1.65)	0.38
Total		-	71.09	-54.26	-42.52	-25.7 (4.38)	16.83
Average	77	39	6.46	-4.93	-3.87	-2.34 (1.26)	1.53

Table II

Average electrostatic free energy contributions to salt-bridges in hUDG with $\varepsilon_p = 4$ and zero salt added.

Salt-bridge	Occupied %	ASA Å^2	$\Delta\Delta G_{dslv}$ kcal/mol	$\Delta\Delta G_{brd}$ kcal/mol	$\Delta\Delta G_{prt}$ kcal/mol	$\Delta\Delta G_{tot}$ kcal/mol	$\Delta\Delta G_{ass}$ kcal/mol
Glu96-Lys99	100	37	5.44	-5.53	-2.40	-2.49 (1.07)	-0.09
Glu111-Lys114	41	64	2.03	-2.14	-0.72	-0.82 (1.07	-0.10
Glu112-Arg210	100	19	9.53	-9.59	-2.83	-2.89 (1.14)	-0.06
Asp133-Lys135	40	55	5.61	-3.41	-5.38	-3.18 (1.21)	2.20
Asp180-Arg282	99	31	10.59	-7.81	-8.59	-5.81 (2.70)	2.78
Asp183-Lys302	90	51	4.87	-4.80	-4.58	-4.51 (0.96)	0.07
Lys197-Glu303	36	58	1.81	-1.57	0.01	0.25 (0.77)	0.25
Lys218-Glu219	60	65	1.99	-2.36	-0.57	-0.94 (0.94)	-0.37
Asp227-Lys252	83	41	5.42	-3.42	-1.52	0.47 (1.19)	1.99
Asp257-Lys259	39	63	3.56	-2.73	-2.74	-1.91 (1.15)	0.83
Glu289-Lys293	44	66	1.66	-1.55	-0.07	0.04 (0.92)	0.11
Asp300-Lys302	98	46	6.02	-5.80	-4.12	-3.90 (1.15)	0.22
Total	-	-	58.55	-50.71	-33.53	-25.7 (4.43)	7.84
Average	69	50	4.88	-4.23	-2.79	-2.1 (1.01)	0.65

Table III

Average electrostatic free energy contributions to salt-bridges in cUDG with $\varepsilon_p = 10$ and zero salt added.

Salt-bridge	Occupied %	ASA Å^2	$\Delta\Delta G_{dslv}$ Kkcal/mol	$\Delta\Delta G_{brd}$ kcal/mol	$\Delta\Delta G_{prt}$ kcal/mol	$\Delta\Delta G_{tot}$ kcal/mol	$\Delta\Delta G_{ass}$ kcal/mol
Glu87-Arg90	35	59	1.63	-0.87	-1.98	-1.22 (0.63)	0.76
Arg91-Glu92	92	46	2.23	-3.06	-0.57	-1.41 (0.56)	-0.84
Glu96-Lys99	99	33	2.80	-3.28	-1.51	-1.99 (0.41)	-0.49
Asp111-Arg115	100	52	2.28	-3.36	-0.42	-1.50 (0.33)	-1.08
Glu112-Arg210	100	16	3.98	-4.75	-1.06	-1.83 (0.37)	-0.77
Asp136-Lys138	56	45	2.08	-1.77	-1.43	-1.12 (0.65)	0.31
Asp145-His148	65	29	2.71	-2.72	-0.90	-0.91 (0.53)	-0.01
Asp180-Lys282	82	30	4.21	-3.05	-4.40	-3.23 (0.84)	1.16
Asp227-Lys231	49	45	2.11	-1.76	-1.97	-1.63 (0.75)	0.35
Asp227-Lys252	88	23	3.36	-2.08	-2.45	-1.17 (0.68)	1.28
Asp257-Lys259	76	54	2.18	-2.70	-1.46	-1.97 (0.75)	-0.52
Total		-	29.56	-29.41	-18.13	-17.99 (2.03)	0.15
Average	77	39	2.69	-2.67	-1.65	-1.64 (0.59)	0.01

Table IV

Average electrostatic free energy contributions to salt-bridges in hUDG with $\varepsilon_p = 10$ and zero salt added.

Salt-bridge	**Occupied** %	**ASA** Å^2	**$\Delta\Delta G_{dslv}$** kcal/mol	**$\Delta\Delta G_{brd}$** kcal/mol	**$\Delta\Delta G_{prt}$** kcal/mol	**$\Delta\Delta G_{tot}$** kcal/mol	**$\Delta\Delta G_{ass}$** kcal/mol
Glu96-Lys99	100	37	2.34	-3.10	-0.88	-1.64 (0.46)	-0.76
Glu111-Lys114	41	64	0.91	-1.38	-0.36	-0.84 (0.58)	-0.48
Glu112-Arg210	100	19	3.84	-4.61	-1.09	-1.87 (0.49)	-0.78
Asp133-Lys135	40	55	2.35	-1.94	-2.07	-1.66 (0.60)	0.42
Asp180-Arg282	99	31	4.07	-3.75	-3.43	-3.11 (1.29)	0.32
Asp183-Lys302	90	51	2.02	-2.68	-2.22	-2.88 (0.43)	-0.65
Lys197-Glu303	36	58	0.89	-1.12	0.04	-0.19 (0.47)	-0.23
Lys218-Glu219	60	65	0.93	-1.54	-0.16	-0.77 (0.53)	-0.61
Asp227-Lys252	83	41	2.42	-2.12	-0.48	-0.19 (0.58)	0.30
Asp257-Lys259	39	63	1.50	-1.65	-1.21	-1.36 (0.59)	-0.15
Glu289-Lys293	44	66	0.80	-1.11	0.01	-0.31 (0.56)	-0.32
Asp300-Lys302	98	46	2.49	-3.15	-2.03	-2.68 (0.51)	-0.65
Total	-	-	24.55	-28.14	-13.88	-17.5 (2.18)	-3.60
Average	69	50	2.05	-2.35	-1.16	-1.5 (0.59)	-0.30

Table V

Average electrostatic free energy contributions to salt-bridges in cUDG with $\varepsilon_p = 20$ and zero salt added.

Salt-bridge	**Occupied** %	**ASA** Å^2	**$\Delta\Delta G_{dslv}$** kcal/mol	**$\Delta\Delta G_{brd}$** kcal/mol	**$\Delta\Delta G_{prt}$** kcal/mol	**$\Delta\Delta G_{tot}$** kcal/mol	**$\Delta\Delta G_{ass}$** kcal/mol
Glu87-Arg90	35	59	0.79	-0.78	-1.08	-1.07 (0.42)	0.01
Arg91-Glu92	92	46	1.09	-2.07	-0.37	-1.35 (0.35)	-0.98
Glu96-Lys99	99	33	1.40	-2.28	-0.78	-1.65 (0.24)	-0.87
Asp111-Arg115	100	52	1.08	-2.23	-0.17	-1.33 (0.20)	-1.15
Glu112-Arg210	100	16	1.88	-2.92	-0.43	-1.48 (0.20)	-1.04
Asp136-Lys138	56	45	1.09	-1.41	-0.81	-1.13 (0.30)	-0.32
Asp145-His148	65	29	1.38	-1.80	-0.49	-0.91 (0.31)	-0.42
Asp180-Lys282	82	30	1.93	-2.00	-2.24	-2.31 (0.46)	-0.07
Asp227-Lys231	49	45	1.09	-1.33	-1.18	-1.42 (0.43)	-0.24
Asp227-Lys252	88	23	1.69	-1.51	-1.29	-1.12 (0.42)	0.18
Asp257-Lys259	76	54	1.06	-1.89	-0.75	-1.58 (0.46)	-0.83
Total		-	14.48	-20.23	-9.59	-15.34 (1.19)	-5.75
Average	77	39	1.32	-1.84	-0.87	-1.39 (0.34)	-0.52

Table VI

Average electrostatic free energy contributions to salt-bridges in hUDG with $\varepsilon_p = 20$ and zero salt added.

Salt-bridge	Occupied %	ASA Å^2	$\Delta\Delta G_{dslv}$ kcal/mol	$\Delta\Delta G_{brd}$ kcal/mol	$\Delta\Delta G_{prt}$ kcal/mol	$\Delta\Delta G_{tot}$ kcal/mol	$\Delta\Delta G_{ass}$ kcal/mol
Glu96-Lys99	100	37	1.19	-2.19	-0.37	-1.38 (0.26)	-1.00
Glu111-Lys114	41	64	0.48	-1.11	-0.23	-0.86 (0.42)	-0.63
Glu112-Arg210	100	19	1.81	-2.85	-0.50	-1.54 (0.27)	-1.03
Asp133-Lys135	40	55	1.17	-1.39	-0.96	-1.18 (0.42)	-0.22
Asp180-Arg282	99	31	1.84	-2.33	-1.71	-2.21 (0.80)	-0.50
Asp183-Lys302	90	51	1.00	-1.91	-1.38	-2.28 (0.26)	-0.91
Lys197-Glu303	36	58	0.51	-0.93	0.09	-0.33 (0.38)	-0.42
Lys218-Glu219	60	65	0.51	-1.23	-0.01	-0.73 (0.39)	-0.72
Asp227-Lys252	83	41	1.25	-1.57	-0.12	-0.45 (0.39)	-0.33
Asp257-Lys259	39	63	0.75	-1.26	-0.67	-1.18 (0.40)	-0.50
Glu289-Lys293	44	66	0.45	-0.94	0.04	-0.45 (0.43)	-0.49
Asp300-Lys302	98	46	1.22	-2.18	-1.29	-2.25 (0.29)	-0.96
Total	-	-	12.18	-19.88	-7.13	-14.84 (1.44)	-7.71
Average	69	50	1.01	-1.66	-0.59	-1.24 (0.39)	-0.64

Table VII

Average electrostatic free energy contributions to salt-bridges in cUDG with $\varepsilon_p = 4$ and 100 mM salt added.

Salt-bridge	Occupied %	ASA Å^2	$\Delta\Delta G_{dslv}$ kcal/mol	$\Delta\Delta G_{brd}$ kcal/mol	$\Delta\Delta G_{prt}$ kcal/mol	$\Delta\Delta G_{tot}$ kcal/mol	$\Delta\Delta G_{ass}$ kcal/mol
Glu87-Arg90	35	59	3.93	-0.80	-4.64	-1.51 (1.23)	3.13
Arg91-Glu92	92	46	5.38	-5.61	-1.05	-1.28 (1.17)	-0.24
Glu96-Lys99	99	33	6.47	-5.65	-3.71	-2.89 (0.88)	0.82
Asp111-Arg115	100	52	5.69	-6.34	-1.25	-1.90 (0.73)	-0.65
Glu112-Arg210	100	16	9.80	-9.64	-3.02	-2.86 (0.87)	0.16
Asp136-Lys138	56	45	4.58	-2.36	-3.10	-0.88 (0.96)	2.23
Asp145-His148	65	29	6.09	-4.85	-2.02	-0.78 (1.22)	1.24
Asp180-Lys282	82	30	10.77	-5.73	-10.90	-5.86 (1.96)	5.04
Asp227-Lys231	49	45	4.73	-2.58	-4.11	-1.97 (1.71)	2.14
Asp227-Lys252	88	23	7.75	-3.19	-6.02	-1.46 (1.51)	4.56
Asp257-Lys259	76	54	5.30	-4.70	-3.61	-3.01 (1.65)	0.60
Total		-	70.48	-51.45	-43.43	-24.40 (4.37)	19.03
Average	77	39	6.41	-4.68	-3.95	-2.22 (1.26)	1.73

Table VIII

Average electrostatic free energy contributions to salt-bridges in hUDG with $\varepsilon_p = 4$ and 100 mM salt added.

Salt-bridge	Occupied %	ASA Å^2	$\Delta\Delta G_{dslv}$ kcal/mol	$\Delta\Delta G_{brd}$ kcal/mol	$\Delta\Delta G_{prt}$ kcal/mol	$\Delta\Delta G_{tot}$ kcal/mol	$\Delta\Delta G_{ass}$ kcal/mol
Glu96-Lys99	100	37	5.38	-5.25	-2.54	-2.40 (1.07)	0.13
Glu111-Lys114	41	64	2.00	-1.89	-0.75	-0.64 (1.08)	0.11
Glu112-Arg210	100	19	9.47	-9.34	-2.96	-2.83 (1.14)	0.14
Asp133-Lys135	40	55	5.55	-3.16	-5.55	-3.15 (1.19)	2.39
Asp180-Arg282	99	31	10.55	-7.57	-8.67	-5.70 (2.65)	2.98
Asp183-Lys302	90	51	4.82	-4.54	-4.59	-4.30 (0.96)	0.28
Lys197-Glu303	36	58	1.75	-1.30	-0.10	0.35 (0.75)	0.45
Lys218-Glu219	60	65	1.94	-2.09	-0.72	-0.87 (0.92)	-0.15
Asp227-Lys252	83	41	5.34	-3.16	-1.72	0.47 (1.18)	2.19
Asp257-Lys259	39	63	3.53	-2.48	-2.86	-1.81 (1.16)	1.04
Glu289-Lys293	44	66	1.62	-1.29	-0.26	0.06 (0.85)	0.32
Asp300-Lys302	98	46	5.97	-5.54	-4.07	-3.64 (1.15)	0.44
Total	-	-	57.93	-47.61	-34.78	-24.46 (4.38)	10.32
Average	69	50	4.83	-3.97	-2.90	-2.04 (1.18)	0.86

Table IX

Average electrostatic free energy contributions to salt-bridges in cUDG with $\varepsilon_p = 10$ and 100 mM salt added.

Salt-bridge	Occupied %	ASA Å^2	$\Delta\Delta G_{dslv}$ kcal/mol	$\Delta\Delta G_{brd}$ kcal/mol	$\Delta\Delta G_{prt}$ kcal/mol	$\Delta\Delta G_{tot}$ kcal/mol	$\Delta\Delta G_{ass}$ kcal/mol
Glu87-Arg90	35	59	1.59	-0.62	-2.07	-1.11 (0.63)	0.97
Arg91-Glu92	92	46	2.19	-2.81	-0.51	-1.14 (0.55)	-0.62
Glu96-Lys99	99	33	2.74	-3.02	-1.57	-1.85 (0.39)	-0.28
Asp111-Arg115	100	52	2.25	-3.12	-0.54	-1.41 (0.33)	-0.87
Glu112-Arg210	100	16	3.94	-4.52	-1.21	-1.79 (0.37)	-0.58
Asp136-Lys138	56	45	2.03	-1.53	-1.52	-1.02 (0.46)	0.50
Asp145-His148	65	29	2.66	-2.51	-0.95	-0.80 (0.54)	0.14
Asp180-Lys282	82	30	4.17	-2.82	-4.53	-3.17 (0.83)	1.35
Asp227-Lys231	49	45	2.05	-1.52	-1.97	-1.43 (0.73)	0.54
Asp227-Lys252	88	23	3.30	-1.84	-2.66	-1.20 (0.69)	1.46
Asp257-Lys259	76	54	2.15	-2.46	-1.56	-1.86 (0.75)	-0.31
Total		-	29.09	-26.78	-19.08	-16.78 (1.96)	2.30
Average	77	39	2.64	-2.43	-1.73	-1.53(0.57)	0.21

Table X

Average electrostatic free energy contributions to salt-bridges in hUDG with ε_p = 10 and 100 mM salt added.

Salt-bridge	Occupied %	ASA Å^2	$\Delta\Delta G_{dslv}$ kcal/mol	$\Delta\Delta G_{brd}$ kcal/mol	$\Delta\Delta G_{prt}$ kcal/mol	$\Delta\Delta G_{tot}$ kcal/mol	$\Delta\Delta G_{ass}$ kcal/mol
Glu96-Lys99	100	37	2.29	-2.83	-1.01	-1.56 (0.45)	-0.54
Glu111-Lys114	41	64	0.88	-1.14	-0.39	-0.66 (0.59)	-0.27
Glu112-Arg210	100	19	3.80	-4.38	-1.23	-1.81 (0.49)	-0.58
Asp133-Lys135	40	55	2.30	-1.70	-2.24	-1.63 (0.58)	0.61
Asp180-Arg282	99	31	4.03	-3.52	-3.51	-3.00 (1.24)	0.51
Asp183-Lys302	90	51	1.99	-2.43	-2.24	-2.68 (0.43)	-0.44
Lys197-Glu303	36	58	0.84	-0.86	-0.07	-0.09 (0.44)	-0.03
Lys218-Glu219	60	65	0.88	-1.28	-0.30	-0.70 (0.52)	-0.40
Asp227-Lys252	83	41	2.37	-1.88	-0.67	-0.19 (0.57)	0.48
Asp257-Lys259	39	63	1.47	-1.41	-1.32	-1.26 (0.60)	0.06
Glu289-Lys293	44	66	0.76	-0.86	-0.17	-0.28 (0.49)	-0.11
Asp300-Lys302	98	46	2.45	-2.90	-1.99	-2.43 (0.51)	-0.45
Total	-	-	24.05	-25.21	-15.14	-16.29 (2.12)	-1.15
Average	69	50	2.00	-2.10	-1.26	-1.36 (0.58)	-0.10

Table XI

Average electrostatic free energy contributions to salt-bridges in cUDG with ε_p = 20 and 100 mM salt added.

Salt-bridge	Occupied %	ASA Å^2	$\Delta\Delta G_{dslv}$ kcal/mol	$\Delta\Delta G_{brd}$ kcal/mol	$\Delta\Delta G_{prt}$ kcal/mol	$\Delta\Delta G_{tot}$ kcal/mol	$\Delta\Delta G_{ass}$ kcal/mol
Glu87-Arg90	35	59	0.77	-0.55	-1.18	-0.96 (0.42)	0.22
Arg91-Glu92	92	46	1.06	-1.83	-0.33	-1.10 (0.34)	-0.77
Glu96-Lys99	99	33	1.37	-2.04	-0.84	-1.51 (0.23)	-0.68
Asp111-Arg115	100	52	1.06	-2.00	-0.29	-1.23 (0.19)	-0.94
Glu112-Arg210	100	16	1.85	-2.71	-0.58	-1.44 (0.20)	-0.86
Asp136-Lys138	56	45	1.06	-1.18	-0.91	-1.03 (0.29)	-0.13
Asp145-His148	65	29	1.34	-1.62	-0.55	-0.82 (0.32)	-0.27
Asp180-Lys282	82	30	1.90	-1.79	-2.37	-2.26 (0.46)	0.11
Asp227-Lys231	49	45	1.05	-1.11	-1.19	-1.25 (0.42)	-0.05
Asp227-Lys252	88	23	1.65	-1.30	-1.49	-1.14 (0.42)	0.35
Asp257-Lys259	76	54	1.04	-1.66	-0.85	-1.47 (0.46)	-0.62
Total		-	14.15	-17.79	-10.57	-14.21 (1.18)	-3.64
Average	77	39	1.29	-1.62	-0.96	-1.29 (0.34)	-0.33

Table XII

Average electrostatic free energy contributions to salt-bridges in hUDG with $\varepsilon_p = 20$ and 100 mM salt added.

Salt-bridge	Occupied %	ASA Å^2	$\Delta\Delta G_{dslv}$ kcal/mol	$\Delta\Delta G_{brd}$ kcal/mol	$\Delta\Delta G_{prt}$ kcal/mol	$\Delta\Delta G_{tot}$ kcal/mol	$\Delta\Delta G_{ass}$ kcal/mol
Glu96-Lys99	100	37	1.15	-1.94	-0.51	-1.30 (0.25)	-0.79
Glu111-Lys114	41	64	0.46	-0.87	-0.27	-0.69 (0.43)	-0.41
Glu112-Arg210	100	19	1.79	-2.63	-0.63	-1.48 (0.27)	-0.84
Asp133-Lys135	40	55	1.13	-1.17	-1.12	-1.16 (0.39)	-0.04
Asp180-Arg282	99	31	1.81	-2.12	-1.79	-2.10 (0.75)	-0.31
Asp183-Lys302	90	51	0.97	-1.67	-1.41	-2.11 (0.26)	-0.70
Lys197-Glu303	36	58	0.47	-0.70	-0.03	-0.25 (0.35)	-0.22
Lys218-Glu219	60	65	0.48	-0.99	-0.14	-0.65 (0.38)	-0.51
Asp227-Lys252	83	41	1.21	-1.36	-0.30	-0.45 (0.39)	-0.15
Asp257-Lys259	39	63	0.73	-1.03	-0.79	-1.08 (0.41)	-0.29
Glu289-Lys293	44	66	0.42	-0.70	-0.13	-0.41 (0.37)	-0.28
Asp300-Lys302	98	46	1.19	-1.95	-1.26	-2.02 (0.29)	-0.76
Total	-	-	11.82	-17.13	-8.39	-13.69 (1.38)	-5.31
Average	69	50	0.99	-1.43	-0.70	-1.14 (0.38)	-0.44

Paper V

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

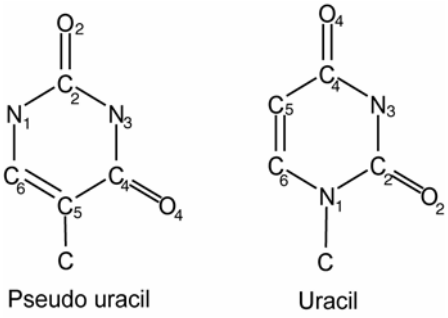


Fig. 1.

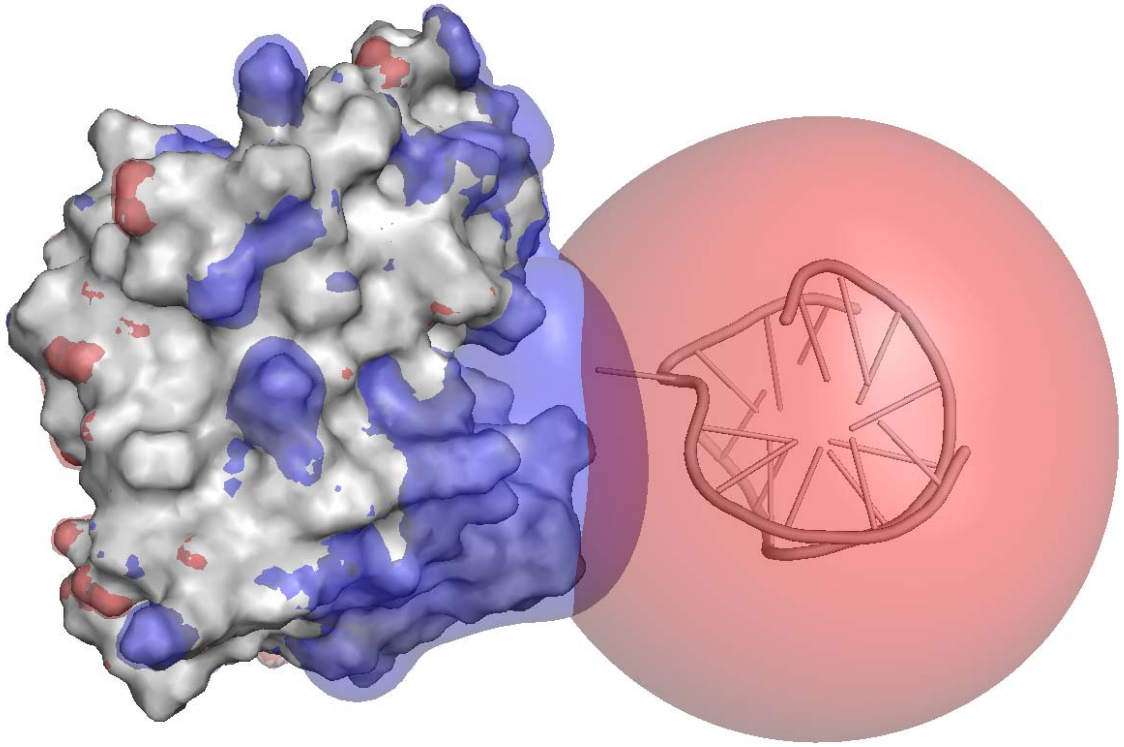


Fig. 2

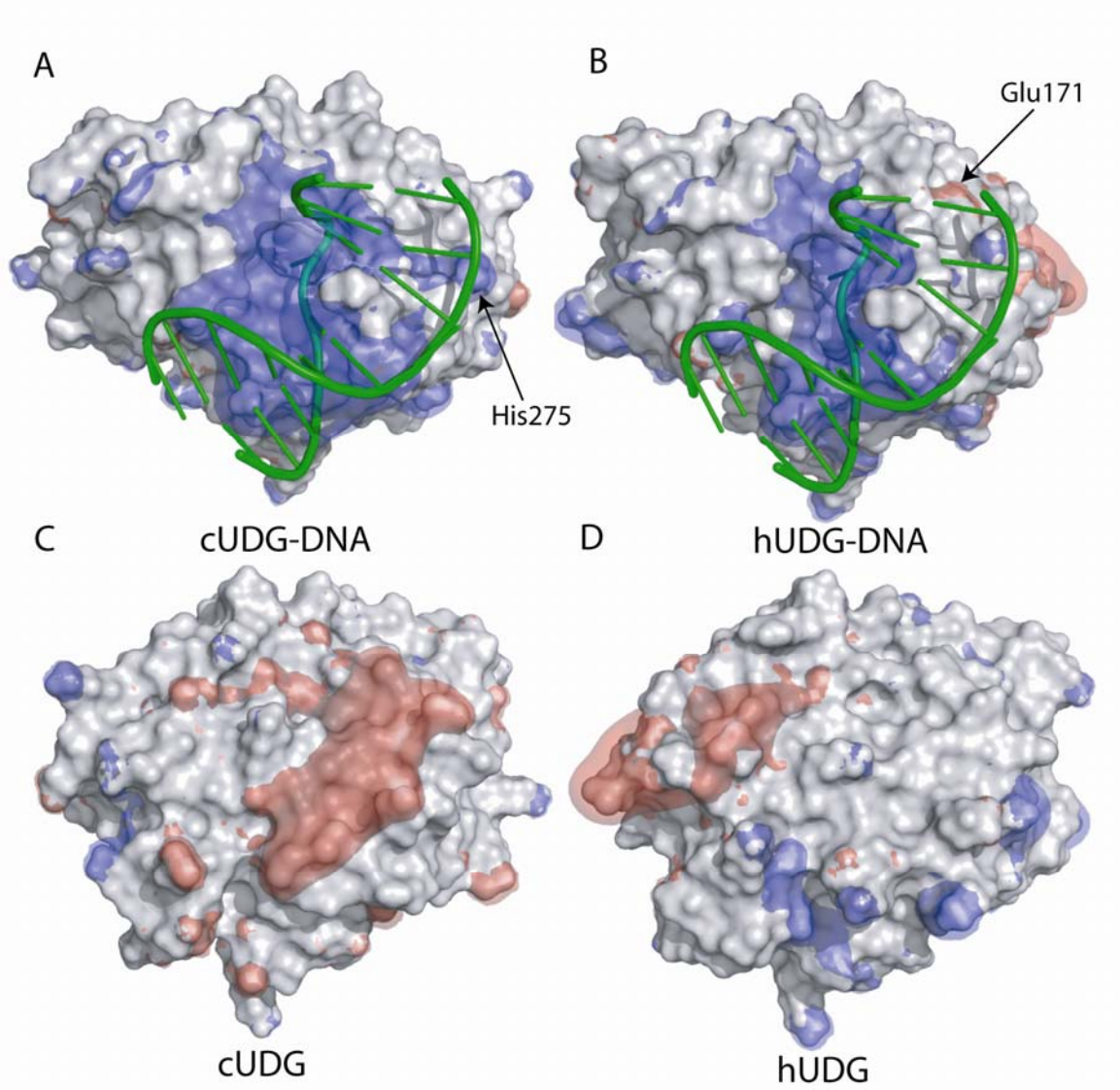


Fig. 3

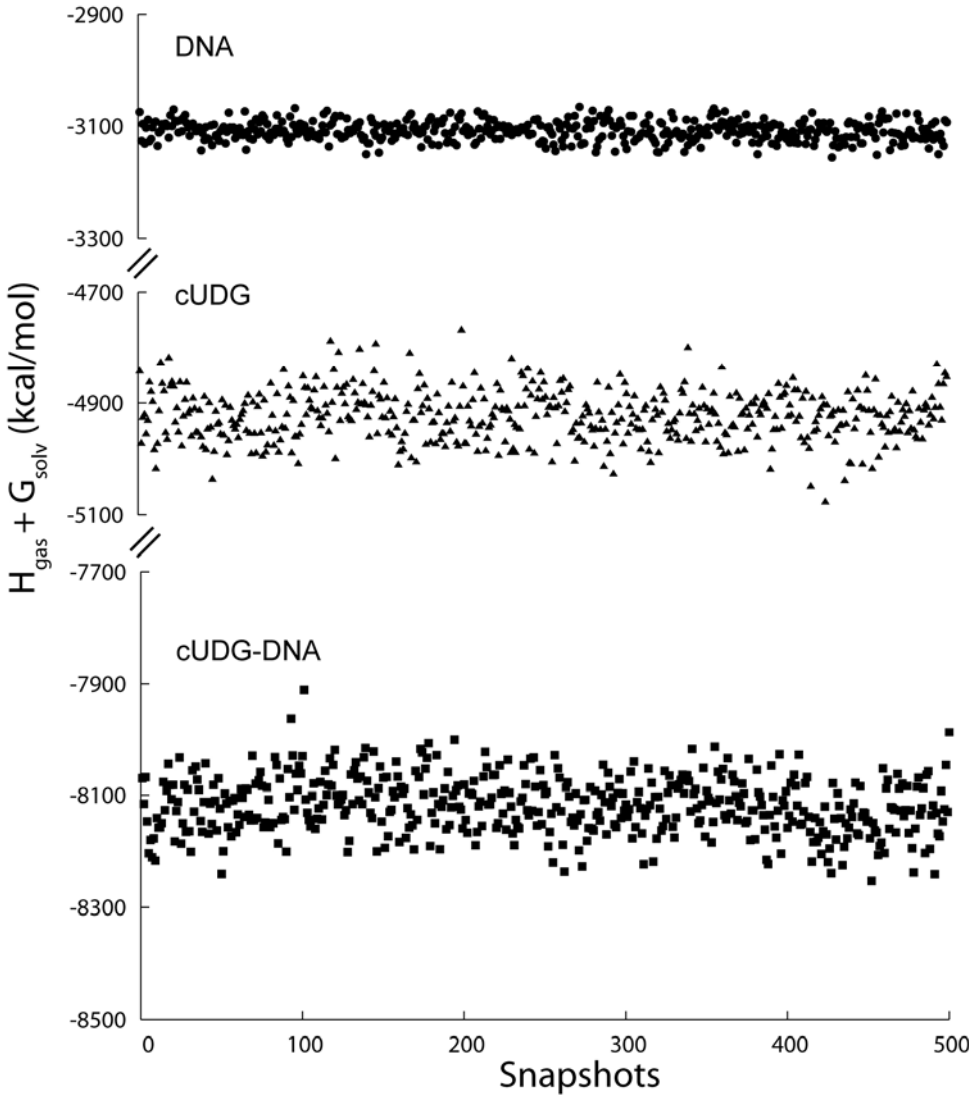

DNA
cUDG
cUDG-DNA
-2900
-3100
-3300
-4700
-4900
-5100
-7700
-7900
-8100
-8300
-8500
0
100
200
300
400
500
Snapshots
H_gas + G_solv (kcal/mol)

Fig. 4

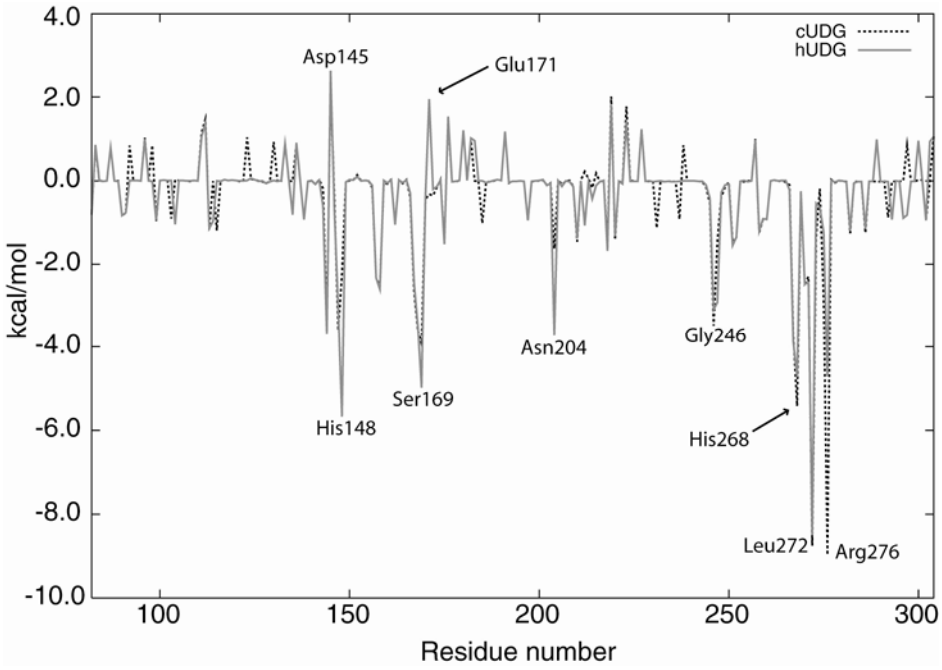

cUDG
hUDG
Asp145
Glu171
Asn204
Gly246
Ser169
His148
His268
Leu272
Arg276
kcal/mol
4.0
2.0
0.0
-2.0
-4.0
-6.0
-8.0
-10.0
100
150
200
250
300
Residue number

Fig. 5

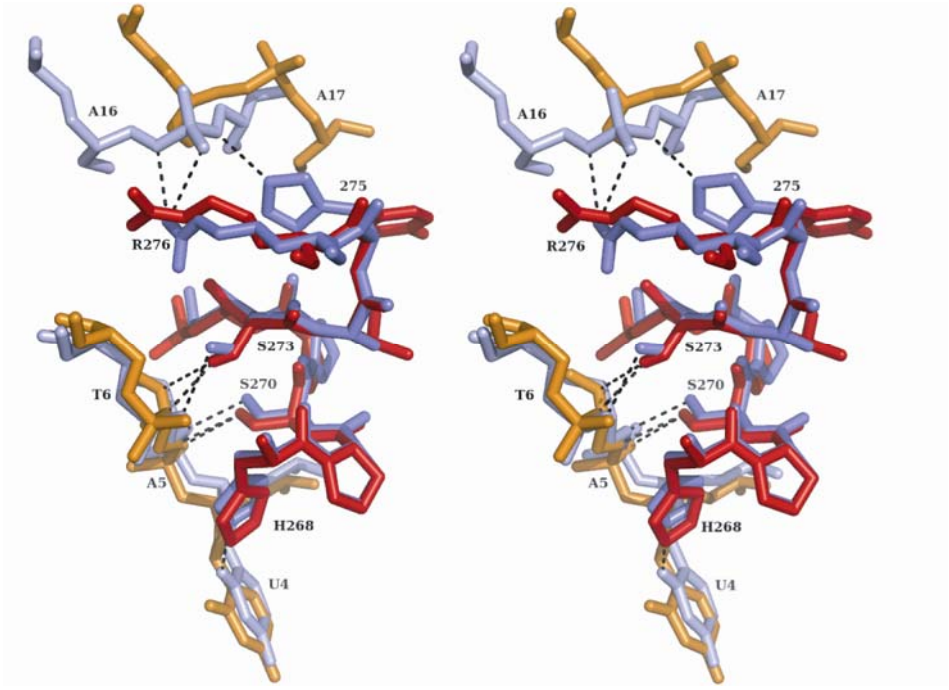

A16
A17
275
R276
S273
S270
T6
A5
H268
U4
A16
A17
275
R276
S273
S270
T6
A5
H268
U4

Fig. 6