Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure of uracil-DNA glycosylase from Atlantic cod (*Gadus morhua*) reveals cold-adaptation features

Ingar Leiros,[a] Elin Moe,[b] Olav Lanes,[c] Arne O. Smalås[a] and Nils P. Willassen[b]*

[a]Department of Chemistry, Faculty of Science, University of Tromsø, N-9037 Tromsø, Norway, [b]Department of Molecular Biotechnology, Institute of Medical Biology, Faculty of Medicine, University of Tromsø, N-9037 Tromsø, Norway, and [c]Biotec Pharmacon, Strandgt. 3, N-9008 Tromsø, Norway

Correspondence e-mail: nilspw@fagmed.uit.no

Received 13 December 2002
Accepted 19 May 2003

PDB Reference: uracil-DNA glycosylase, 1okb, r1okbsf.

Uracil-DNA glycosylase (UDG; EC 3.2.2.3) is a DNA-repair protein that catalyses the hydrolysis of promutagenic uracil residues from single- or double-stranded DNA, generating free uracil and abasic DNA. The crystal structure of the catalytic domain of cod uracil-DNA glycosylase (cUDG) has been determined to 1.9 Å resolution, with final *R* factors of 18.61 and 20.57% for the working and test sets of reflections, respectively. This is the first crystal structure of a uracil-DNA glycosylase from a cold-adapted species and a detailed comparison with the human enzyme is performed in order to rationalize the cold-adapted behaviour of the cod enzyme at the structural level. The catalytic domain of cUDG comprises 223 residues, with a sequence identity to the human UDG of 75%. The tertiary structures of the two enzymes are also similar, with an overall displacement in main-chain atomic positions of 0.63 Å. The amino-acid substitutions and the differences in intramolecular hydrogen bonds, hydrophobic interactions, ion-pair interactions and electrostatic potentials are compared and discussed in order to gain insight into the factors that cause the increased activity and reduced thermostability of the cod enzyme. In particular, the reduced number of strong ion-pair interactions in the C-terminal half of cUDG is believed to greatly affect the flexibility and/or stability. Increased positive electrostatic surface potential on the DNA-facing side of cUDG seems to be responsible for increasing the affinity for the negatively charged DNA compared with that of hUDG.

1. Introduction

The so-called psychrophilic enzymes are proteins from organisms that inhabit areas with temperatures lower than about 293 K. The general characteristics of cold-adapted enzymes are a high specific activity at low temperature, low-temperature optima and reduced thermal stability compared with their mammalian counterparts (Smalås *et al.*, 2000). Interest in extremophilic enzymes has increased significantly in recent years, both from an academic point of view and as targets for possible commercialization. In particular, cold-adapted enzymes have become attractive for their high specific activity. The reduced thermal stability found in most cases to accompany the increased activity is recognized as an additional attractive feature in industrial and biotechnological applications. Control, *i.e.* inactivation, of enzyme activity in a predicable manner is highly preferable in most commercial applications.

The research community has been studying the molecular basis of these unique features for several decades, but until recently the lack of three-dimensional structural data has limited insight into the apparently delicate molecular adap-

tation. Since the first crystal structure of a cold-adapted enzyme became available in 1993 (Atlantic salmon trypsin; Smalås & Hordvik, 1993), many crystal structures describing cold-active enzymes have been determined, both of bacterial origin (Aghajari *et al.*, 1998; Russell *et al.*, 1998; Kim *et al.*, 1999; Alvarez *et al.*, 1998) and from vertebrates (Smalås & Hordvik, 1993; Berglund *et al.*, 1995; Karlsen *et al.*, 1998). To date, the numbers of well characterized cold-adapted enzymes have been relatively equally distributed between the bacterial and vertebrate/invertebrate kingdoms. However, no principal differences in the degree of cold-adaptation or adaptation strategies that depend on the type of cold habitat or the taxonomy of the organisms have thus far been identified (Smalås *et al.*, 2000). Detailed analysis and comparative studies with mesophilic and thermophilic counterparts have revealed structural cold-activity descriptors for separate enzyme classes, but there are few observations of features that can be generalized. Despite the lack of direct evidence for most target enzymes, the proposed correlation between high specific activity and increased flexibility is still the most dominant theory for rationalization of cold-adaptation (Hochachka & Somero, 1984). Reduced stability is in turn regarded as a consequence of the need to increase the molecular flexibility to reduce the free energy of activation at low temperatures. Most comparative studies have therefore focused on exploring differences that can account for reduced stability, *e.g.* core packing, intramolecular interactions and amino-acid compositions. Generally, the total number of intramolecular interactions seems to be similar in both mesophiles and their cold-adapted counterparts, but several case studies have revealed a reduced number of such interactions in certain regions of the cold enzymes (Smalås *et al.*, 1994; Leiros *et al.*, 2000; Genicot *et al.*, 1996). In particular, interactions between domains and subunits of the proteins seem to be weaker for many cold-adapted proteins. Factors such as less well optimized cores or increased loop lengths have been observed for some cold-adapted enzymes (Genicot *et al.*, 1996; Feller *et al.*, 1994, 1997), but do not seem to be a general feature. Increased flexibility may however not be the only adaptation strategy to deal with cold environments. Optimization of electrostatic potentials of residues involved in substrate binding has been shown to play a significant role in the cold-adaptation of trypsin (Gorfe *et al.*, 2000; Brandsdal *et al.*, 2001).

Uracil-DNA glycosylase (UDG) is a central component of the DNA-repair machinery as it initiates the DNA base-excision repair pathway (BER) by removing misincorporated uracil from DNA. It is the first enzyme in this pathway and catalyzes the hydrolysis of promutagenic uracil residues from single- or double-stranded DNA. In contrast to other nucleotide-excision repair systems, the DNA-glycosylases require no additional cofactor for activity and are therefore excellent candidates for the characterization of protein–DNA interactions, damage recognition and the removal of damaged bases from DNA at an atomic level. The enzyme is widespread and is found in almost all organisms examined (Mol *et al.*, 1999). Uracil-DNA glycosylase isolated from Atlantic cod (cUDG) was found to be up to ten times more catalytically active in the temperature range 288–310 K than the human counterpart (Lanes *et al.*, 2002). The higher physiological catalytic efficiency (k_{cat}/K_M) is achieved by the combined effect of a fourfold lower K_M and an approximately twofold increase in k_{cat}. cUDG is unstable at elevated temperatures, a feature shared with most other cold-active enzymes. The measured half-life at 323 K was 0.5 min for cUDG compared with 8 min for recombinant human UDG (rhUDG; Lanes *et al.*, 2002). The crystal structure of the catalytic domain of cUDG has been found in general to be very similar to the already published crystal structure of the human enzyme (hUDG), which is in accordance with the fact that the respective sequences are similar also (sequence identity of 75% in the catalytic domains). A homology model has previously been built based on the crystal structure of the unbound form of hUDG (Lanes *et al.*, 2002). However, in order to identify the differences responsible for the cold-adaptation features observed for cUDG, the need for examination at a more detailed level becomes obvious and a more thorough comparison is needed. Therefore, the crystal structure of cUDG has been determined and a detailed sequence and structural analysis has been performed. From the present study, the most pronounced differences between the two structures are the ion-pair interactions in the C-terminal halves of the two proteins. hUDG is significantly stabilized by five strong salt bridges that are not present in this region of cUDG. Also, cUDG seems to be further destabilized compared with hUDG by the increase in the volumes of buried cavities. Furthermore, the electrostatic surface potentials of the DNA-binding regions are higher for cUDG, allowing increased interactions with the DNA substrate. Optimization of the electrostatic potential in the substrate-binding site has previously been identified as a cold-adaptation strategy (Brandsdal *et al.*, 2001).

2. Materials and methods

2.1. Protein purification and crystallization

Recombinant Atlantic cod uracil-DNA glycosylase (cUDG) expressed in *Escherichia coli* was purified by the following procedure. *E. coli* cell paste was disrupted using a Bio-Neb Cell Disruptor and centrifuged. The cell extract was then precipitated using protamine sulfate and UDG was purified by tandem anionic/cationic ion-exchange, affinity and gel-filtration chromatography and finally cationic ion-exchange chromatography as a final polishing step. The purified cUDG was homogeneous as judged by an SDS–PAGE gel. Crystals were grown at room temperature using the vapour-diffusion method with protein at a concentration of 10 mg ml^{-1} prior to mixing with an equal amount of reservoir solution (2 µl of each). The crystallization solution consisted of 1.4 *M* sodium citrate buffered with 0.1 *M* HEPES pH 7.5. Clusters of crystals useable for X–ray intensity data collection appeared in the drop after about one week. Details of the purification and

crystallization procedures can be obtained elsewhere (Lanes *et al.*, 2002; Leiros *et al.*, 2001).

2.2. Data collection

X-ray intensity data were collected from a single crystal of approximate dimensions 0.1 × 0.3 × 0.05 mm at the Swiss–Norwegian beamline (SNBL) at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. A crystal was flash-cooled by transferring it into a well containing the crystallization solution with 20% glycerol added as a cryoprotectant. The crystal was then placed in a loop and flash-cooled in liquid nitrogen. Intensity data were collected on a MAR345 image-plate detector in the 180 mm mode at a wavelength of 0.873 Å and with a crystal-to-detector distance of 150 mm. 180 images were collected using an oscillation of 1°. The data were indexed with *DENZO* (Otwinowski & Minor, 1997) and processed using programs from the *CCP*4 program package (Collaborative Computational Project, Number 4, 1994). The crystal was indexed in space group $P2_1$, with unit-cell parameters a = 68.58, b = 67.19, c = 68.64 Å, β = 119.85°. There were two molecules in each asymmetric unit and the water content was about 50%. Data-collection and processing statistics are shown in Table 1.

Table 1
Data collection and refinement summary.

Values in parentheses refer to the highest resolution bin.

Data collection	
Resolution range (Å)	12–1.9 (2.0–1.9)
No. of unique reflections	42162
Redundancy	3.3 (2.9)
R_{merge} (%)	9.6 (52.2)
Completeness (%)	98.7 (97.5)
Average $I/\sigma(I)$	10.2 (2.0)
Refinement statistics	
R value (%)	18.61
Free R value (%)	20.57
Deviations from ideal geometry	
Bond lengths (Å)	0.005
Bond angles (°)	1.189
Average B values (Å^2)	
Main-chain atoms	19.47
Side-chain atoms	22.38
Glycerol	26.0
Water molecules	35.5
All atoms	22.4
Ramachandran plot (%)	
Most favoured regions	91.1
Additional allowed regions	8.9

2.3. Structure determination

The crystal structure of cUDG was determined to 1.9 Å resolution by molecular replacement using the crystal structure of human uracil-DNA glycosylase (hUDG; PDB code 1akz) as a search model, with the side chains mutated in order to fit the deposited sequence of cUDG (Lanes *et al.*, 2002). Molecular replacement and the subsequent model refinement were performed using *CNS* (Brünger *et al.*, 1998). Data in the resolution range 12–4 Å were used in all molecular-replacement runs. An initial self-rotation search confirmed the presence of two molecules in the asymmetric unit, as assumed during data processing. The cross-rotation search showed two peaks about 8σ above the mean value and 6σ above the highest non-related peak. The subsequent translation search again gave two individual solutions about 5.2σ and 4.5σ above the highest non-related peak. The translation search to complete the dimer, with the best solution fixed, gave one solution only. Molecular replacement was followed by a run of rigid-body refinement using all data in the resolution range 12–2.45 Å, which brought the R factors for the working and test sets down to 38.0 and 38.3%, respectively. The refined model was obtained by alternating cycles of manual model building using the program *O* (Jones *et al.*, 1991), followed by bulk-solvent correction, torsion-angle simulated annealing, conjugate-gradient energy minimization and individual B-factor refinement. Water molecules were added automatically using the water-picking routines of *CNS* (Brünger *et al.*, 1998), with some minor adjustments to the default values set in the program. As the two protein molecules in the asymmetric unit were very similar according to the manual inspection, a restraint of 300 kcal mol^{-1} Å^{-2} (1255 kJ mol^{-1} Å^{-2}) was used during refinement. Throughout the refinement, the progress was monitored by the R_{work} and R_{free} sets of reflections, as well as the quality checks performed by the *PROCHECK* (Laskowski *et al.*, 1993) and *WHAT_CHECK* (Hooft *et al.*, 1996) programs. This procedure was followed until the refinement reached convergence. The final refined model includes residues 82–304 (two crystallographically independent molecules) of the catalytic domain of cUDG, two chlorine ions, two glycerol molecules bound in the active site of each monomer and a total of 343 ordered water molecules. The final refinement statistics can be seen in Table 1.

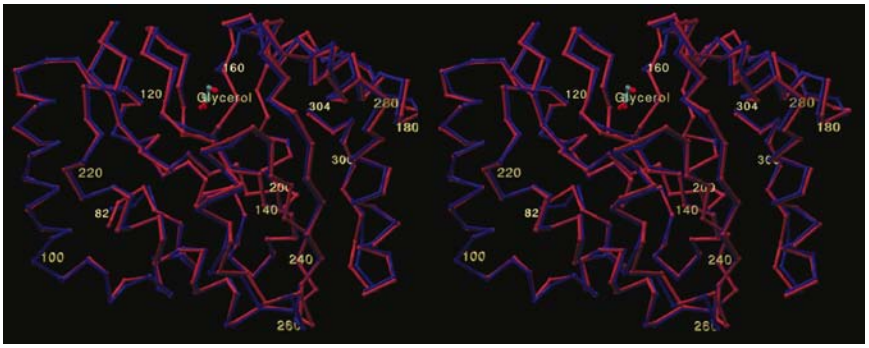
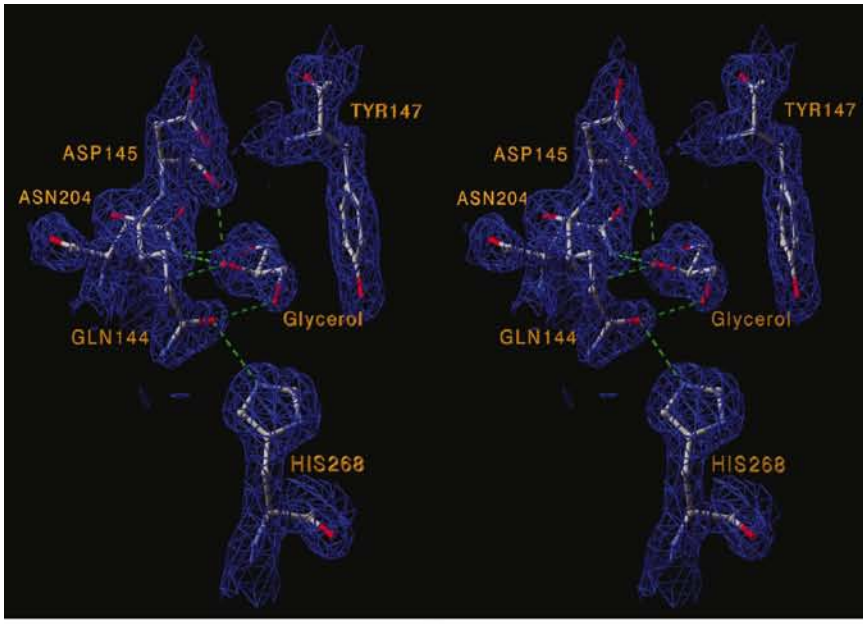


Figure 1
Superpositioning of the crystal structures of cUDG (blue) and hUDG (red; PDB code 1akz; Mol *et al.*, 1995). The glycerol molecule bound in the active site of cUDG is included for clarity.

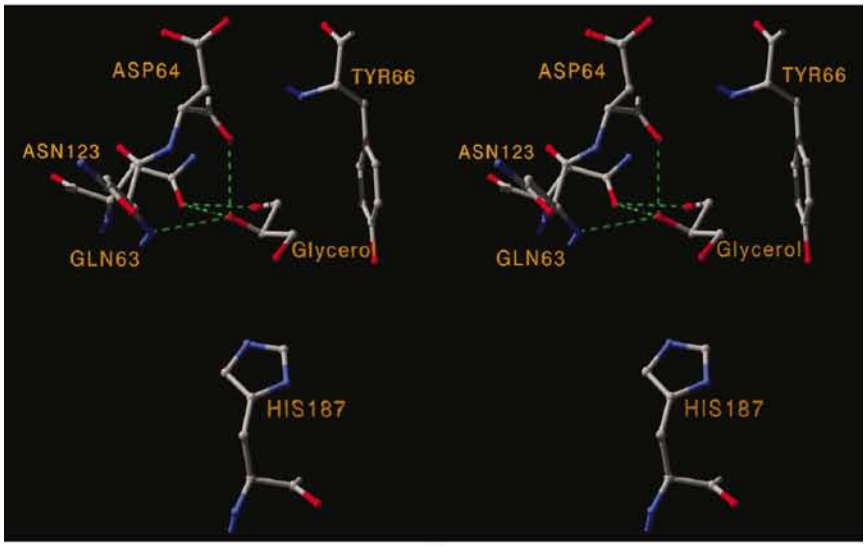
3. Results and discussion

3.1. Refinement and overall structure of cUDG

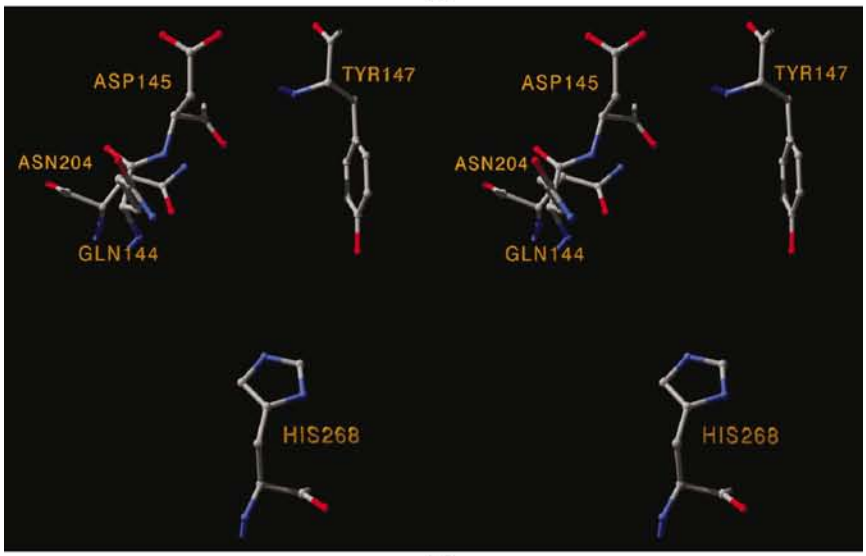
The overall topology of the catalytic domain of uracil-DNA glycosylases is that of a typical α/β protein (Mol *et al.*, 1995). The overall dimensions are approximately 30 ×



(*a*)



(*b*)



(*c*)

35 × 48 Å. There is a central parallel four-stranded β-sheet surrounded on both sides by a total of 11 α-helices. The two termini are situated on opposite sides of the β-sheet. The N-terminus is poorly stabilized, as seen from the high thermal *B* factors and poor electron density in this region. The overall structures of the catalytic domains of cod uracil-DNA glycosylase (cUDG) and human uracil-DNA glycosylase (hUDG; Mol *et al.*, 1995) are similar, with an overall root-mean-square (r.m.s.) deviation of 0.63 Å based on all main-chain atoms. A superpositioning of the C^{α} traces of the cUDG and hUDG structures can be seen in Fig. 1. As the overall structures of the two monomers in the asymmetric unit of cUDG were judged to be very similar from manual inspection of the superimposed molecules, they were tightly restrained during refinement, resulting in a low r.m.s. deviation between them. The restraint refinement also gave the lowest $R_{\rm free}$ value. They are regarded as identical molecules and the following discussion is based on molecule *A* of the asymmetric unit.

3.2. Glycerol binding and the active-site region

The active-site region of the uracil-DNA glycosylases is situated on the C-terminal side of the central β-sheet. In this region, there is a pronounced conical cleft ideally built to accommodate double-stranded DNA. The diameter of this cleft is about 16 Å across and the length of the groove is about 28 Å. In the middle of the DNA-binding groove there is a cavity designed to accommodate the flipped-out uracil base. The amino-acid residues responsible for exclusive binding and cleavage of the flipped-out uracil are the conserved residues Gln144 (backbone amide), Tyr147 (hydrophobic), Phe158 (stacking), Asn204 (two side-chain hydrogen bonds) and His268 (hydrogen bond; Parikh *et al.*, 1998). In addition, Pro146 (carbonyl oxygen), Asp145 and His148 (side chains) coordinate a water

Figure 2
Stereoviews of the active-site environments in the crystal structures of (*a*) cUDG–glycerol, (*b*) *E. coli* UDG–glycerol (Xiao *et al.*, 1999) and (*c*) free hUDG (Mol *et al.*, 1995). The electron density in (*a*) is viewed as a σ_A-weighted $2F_o - F_c$ map contoured at 1.5σ. The conformational shifts of residues Gln144 (Gln63 in *E. coli* UDG) and Asp145 (Asp64 in *E. coli* UDG) can be seen.

molecule assumed to be catalytically critical and this region is therefore denoted as the water-activating loop (residues 144–149; Parikh *et al.*, 1998).

Glycerol has already been found in the active site of a crystal structure of *E. coli* UDG (Xiao *et al.*, 1999). However, the reported structures of the glycerol-bound and free states of *E. coli* UDG are very similar, with an r.m.s. deviation of 0.25 Å on all main-chain atoms in the enzyme; therefore, it is not believed that binding of glycerol will induce any changes to the active-site region in cUDG that are not present in the free form. However, there are other differences in the active-site region that have not previously been observed, which will be discussed further. The glycerol molecule situated in the active site of the crystal structure of cUDG is bound in a similar fashion to glycerol in *E. coli* UDG. The most pronounced differences in the active-site region are the conformational shifts of the side chains of Gln144 and Asp145, two of the residues found to be critical for catalysis by mutational studies (Mol *et al.*, 1995). Conformational shifts have been observed for Asp145 when interacting with DNA and it appears that the cUDG in the cUDG–glycerol complex is in a state not observed in unbound hUDG or in complex with 6-aminouracil. However, in the hUDG–DNA complexes Asp145 adopts a similar conformational state to the glycerol-inhibited cUDG. Representative views of the active-site regions of cUDG–glycerol, *E. coli* UDG–glycerol and free hUDG can be seen in Fig. 2. The fact that Asp145 adopts this conformation is also interesting in view of the positioning of the proposed catalytically critical water molecule. A water molecule can be found in cUDG that adopts exactly the same position and hydrogen-bonding network as in the crystal structure of an uncleaved substrate analogue complexed to hUDG (Parikh *et al.*, 1998).

3.3. Amino-acid substitutions

In order to search for amino-acid residues responsible for the observed cold-adaptation features of cUDG, several aspects have to be taken into account. A sequence alignment (Fig. 3) of the complete catalytic domains of some available vertebrate UDGs [human, GI:14767852; mouse, GI:8571956 and GI:6755941; frog, TC16750 (Tigr); chicken, TC2339

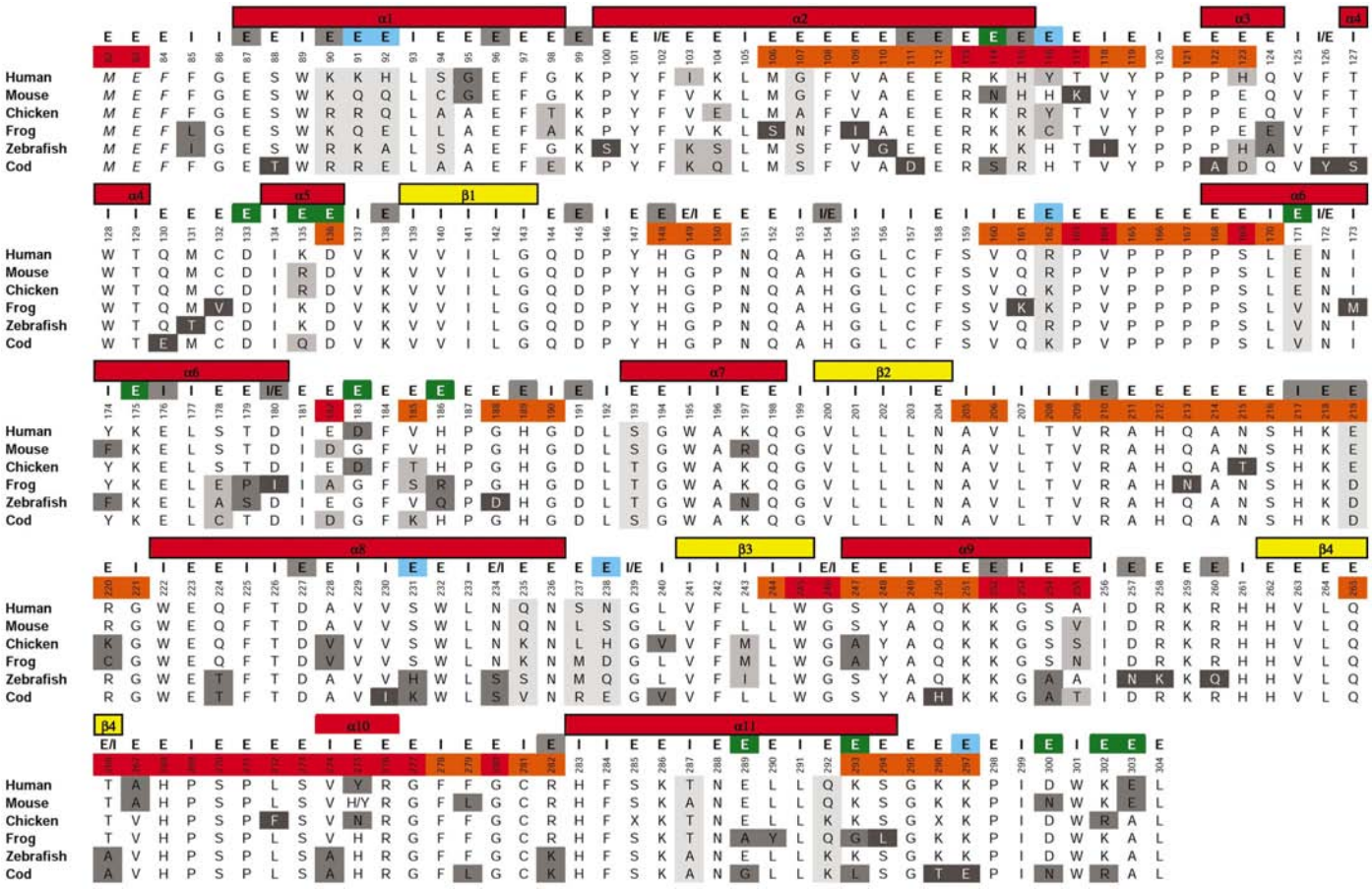


Figure 3
Alignment of the catalytic domain (82–304) of some vertebrate UDG sequences. Residues are shaded according to lack of conservation (dark grey shows a unique residue differing from the consensus, medium grey shows two residues differing from the consensus and light gray denotes no consensus). α-Helices and β-strands are marked as red and yellow tubes above the alignment, respectively. The three N-terminal amino acids M, E and F were added to simplify the comparison and the residue numbers used match the numbering scheme used for the catalytic domain of hUDG. Residues that move upon DNA binding are coloured red (>1.5 Å) or orange (1.0–1.5 Å). External (E) and internal (I) amino-acid residues are based on the water-accessible surface of cUDG (greater or less than 10 Å^2, respectively) and ion-pair interactions shorter than 6 Å are coloured green, blue and grey, denoting hUDG, cUDG and common ion pairs, respectively.

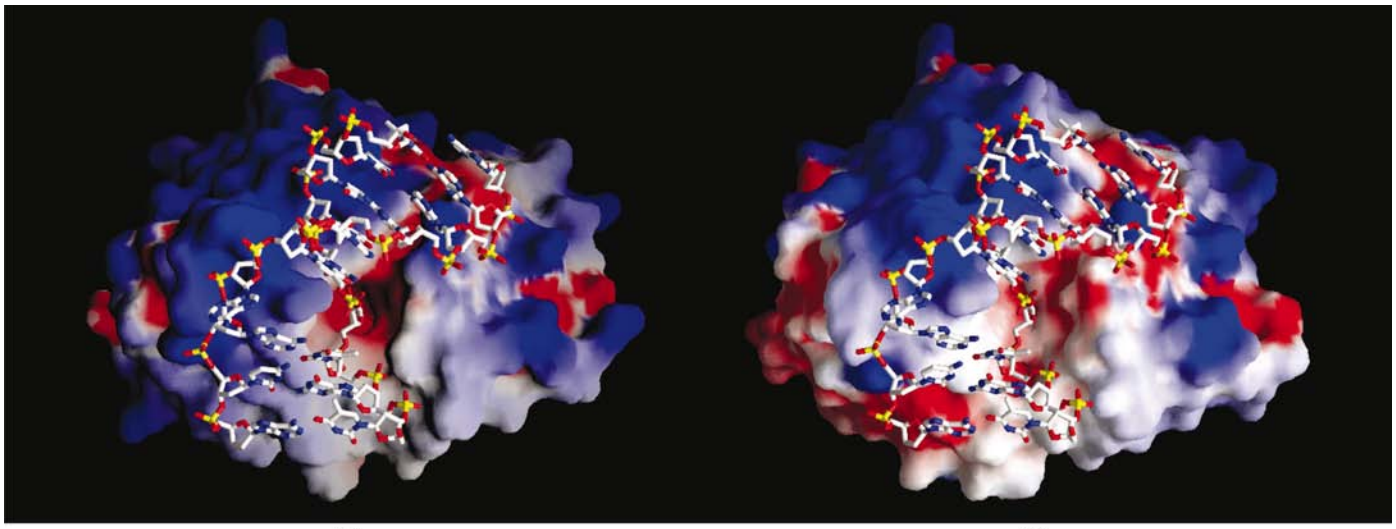


Figure 4
Estimated electrostatic surface potentials of (*a*) the crystal structure of cUDG with DNA modelled in and (*b*) the crystal structure of hUDG–DNA (PDB code 1emh; Parikh *et al.*, 2000).

cUDG < 4Å						hUDG < 4Å					
Donor	Atom	Acceptor	Atom	Distance	From-to	Donor	Atom	Acceptor	Atom	Distance	From-to
Arg91	N^{ε}	Glu92	$O^{\varepsilon2}$	2.80	α1-α1						
Lys99	N^{ζ}	Glu96	$O^{\varepsilon2}$	2.76	α1/α2-α1	Lys99	N^{ζ}	Glu96	$O^{\varepsilon2}$	2.66	α1/α2-α1
						Lys114	N^{ζ}	Glu111	$O^{\varepsilon1}$	3.09	α2-α2
						His115	$N^{\varepsilon2}$	Glu111	$O^{\varepsilon2}$	2.75	α2-α2
						His186	$N^{\varepsilon2}$	Glu171	$O^{\varepsilon1}$	3.05	α6/α7-α6
Arg210	$N^{\eta2}$	Glu112	$O^{\varepsilon2}$	2.76	β2/α8-α2	Arg210	N^{ε}	Glu112	$O^{\varepsilon1}$	2.73	β2/α8-α2
His217	$N^{\varepsilon2}$	Glu112	$O^{\varepsilon1}$	2.72	β2/α8-α2	His217	$N^{\varepsilon2}$	Glu112	$O^{\varepsilon1}$	2.78	β2/α8-α2
Lys252	N^{ζ}	Asp227	$O^{\delta1}$	2.67	α9-α8	Lys252	N^{ζ}	Asp227	$O^{\delta1}$	2.87	α9-α8
						Arg282	N^{ε}	Asp180	$O^{\delta1}$	2.81	α11-α6
						Lys293	N^{ζ}	Glu289	$O^{\varepsilon2}$	3.19	α11-α11
						Lys302	N^{ζ}	Asp183	$O^{\delta2}$	3.14	C-term-α6/α7
						Lys302	N^{ζ}	Asp300	$O^{\delta1}$	3.06	C-term-C-term
cUDG < 6Å						hUDG < 6Å					
Donor	Atom	Acceptor	Atom	Distance	From-to	Donor	Atom	Acceptor	Atom	Distance	From-to
Arg90	$N^{\eta2}$	Glu87	$O^{\varepsilon1}$	4.73	α1-α1	Lys90	N^{ζ}	Glu87	$O^{\varepsilon2}$	4.57	α1-α1
Arg91	N^{ε}	Glu92	$O^{\varepsilon2}$	2.80	α1-α1						
Lys99	N^{ζ}	Glu96	$O^{\varepsilon2}$	2.76	α1/α2-α1	Lys99	N^{ζ}	Glu96	$O^{\varepsilon2}$	2.66	α1/α2-α1
						Lys114	N^{ζ}	Glu111	$O^{\varepsilon1}$	3.09	α2-α2
Arg115	$N^{\eta2}$	Asp111	$O^{\delta1}$	5.09	α2-α2	His115	$N^{\varepsilon2}$	Glu111	$O^{\varepsilon2}$	2.75	α2-α2
His116	$N^{\delta1}$	Glu112	$O^{\varepsilon1}$	4.12	α2-α2						
						Lys135	N^{ζ}	Asp133	$O^{\delta1}$	5.50	α5-α5
						Lys135	N^{ζ}	Asp136	$O^{\delta1}$	4.16	α5-α5
						Lys138	N^{ζ}	Asp136	$O^{\delta1}$	5.52	α5/β1-α5
Lys138	N^{ζ}	Glu297	$O^{\varepsilon1}$	5.40	α5/β1-C-term						
His148	$N^{\delta1}$	Asp145	$O^{\delta2}$	5.71	β1/α6-β1/α6	His148	$N^{\delta1}$	Asp145	$O^{\delta2}$	5.09	β1/α6-β1/α6
His154	$N^{\varepsilon2}$	Asp191	$O^{\delta1}$	4.23	β1/α6-α6/α7	His154	$N^{\varepsilon2}$	Asp191	$O^{\delta1}$	4.17	β1/α6-α6/α7
Lys162	N^{ζ}	Asp191	$O^{\delta2}$	5.10	β1/α6-α6/α7						
						Lys175	N^{ζ}	Glu171	$O^{\varepsilon1}$	5.70	α6-α6
						His186	$N^{\varepsilon2}$	Glu171	$O^{\varepsilon1}$	3.05	α6/α7-α6
His189	$N^{\delta1}$	Asp191	$O^{\delta2}$	5.50	α6/α7-α6/α7	His189	$N^{\delta1}$	Asp191	$O^{\delta2}$	5.49	α6/α7-α6/α7
Arg210	$N^{\eta2}$	Glu112	$O^{\varepsilon2}$	2.76	β2/α8-α2	Arg210	N^{ε}	Glu112	$O^{\varepsilon1}$	2.73	β2/α8-α2
His217	$N^{\varepsilon2}$	Glu112	$O^{\varepsilon1}$	2.72	β2/α8-α2	His217	$N^{\varepsilon2}$	Glu112	$O^{\varepsilon1}$	2.78	β2/α8-α2
Lys218	N^{ζ}	Asp219	$O^{\delta2}$	4.77	β2/α8-β2/α8	Lys218	N^{ζ}	Glu219	$O^{\varepsilon2}$	4.13	β2/α8-β2/α8
Lys231	N^{ζ}	Asp227	$O^{\delta2}$	5.15	α8-α8						
Lys252	N^{ζ}	Asp227	$O^{\delta1}$	2.67	α9-α8	Lys252	N^{ζ}	Asp227	$O^{\delta1}$	2.87	α9-α8
Arg260	$N^{\eta1}$	Glu238	$O^{\varepsilon2}$	4.47	α9/β4-α8/β3						
Arg260	N^{ε}	Asp257	$O^{\delta2}$	4.61	α9/β4-α9/β4	Arg260	N^{ε}	Asp257	$O^{\delta2}$	5.27	α9/β4-α9/β4
Lys282	N^{ζ}	Glu176	$O^{\varepsilon2}$	4.98	α11-α6	Arg282	N^{ε}	Glu176	$O^{\varepsilon2}$	5.52	α11-α6
Lys282	N^{ζ}	Asp180	$O^{\delta1}$	4.32	α11-α6	Arg282	N^{ε}	Asp180	$O^{\delta1}$	2.81	α11-α6
						Lys293	N^{ζ}	Glu289	$O^{\varepsilon2}$	3.19	α11-α11
						Lys302	N^{ζ}	Asp183	$O^{\delta2}$	3.14	C-term-α6/α7
						Lys302	N^{ζ}	Asp300	$O^{\delta1}$	3.06	C-term-C-term
						Lys302	N^{ζ}	Glu303	$O^{\varepsilon1}$	5.32	C-term-C-term

Figure 5
Ion-pair interactions calculated for cUDG and hUDG. Unique ion-pair interactions are shown in red and blue for hUDG and cUDG, respectively, while ion-pairs coloured purple are conserved substitutions.

(Tigr); zebrafish, GI:17035127 and GI:17034292; cod; GI:7413914] and in particular a comparison between the sequences of cUDG and hUDG shows that there are 53 amino-acid substitutions in the cod sequence (compared with hUDG). 24 of these substitutions are unique for cUDG compared with all other vertebrate sequences, *i.e.* they are residues that may be involved the cold-adaptation of cUDG; however, other substitutions should not be ruled out. Almost half of the substitutions are conserved, *i.e.* the substituted residues have similar properties in both proteins.

3.4. Charged residues and electrostatic potentials

The number of charged residues is similar in hUDG and cUDG, but there are three fewer lysine residues and three more arginines in cUDG compared with hUDG. In general, an increase in the arginine content and a higher Arg/Arg + Lys ratio is believed to lead to a more stable structure with an increased ability to form ion-pair and hydrogen-bond interactions (Mrabet *et al.*, 1992). However, this stabilizing effect has been questioned, as arginine most often is found to be external and thereby to interact mainly with water molecules, unless the arginines form ion-pair interactions (Leiros *et al.*, 1999). In addition, an elevated number

of arginine residues (ten) is also found in the sequence of mouse UDG compared with hUDG, which has eight arginines. Therefore, it is most likely that the increased arginine content is not responsible for cold-adaptation in cUDG. The electrostatic surface potentials have been estimated from a homology model of cUDG (Lanes *et al.*, 2002) and important substitutions in the active-site region were identified from that study. In particular, the substitutions Glu171Val, Val185Lys, Gln250His and Tyr275His were indicated as being important in optimizing the electrostatic surface potential of cUDG and may possibly be responsible for the increased activity observed for cUDG compared with hUDG. Re-evaluation of the electrostatic surface potentials based on the present crystal structure (Fig. 4) supports the hypothesis of increased potential of the DNA-binding region of cUDG, resulting in an optimized protein–substrate interaction compared with hUDG.

3.5. Ion-pair interactions

The number of ion-pair interactions in the crystal structures of hUDG and cUDG were calculated and analyzed using the program *CONTACT* from the *CCP*4 program package (Collaborative Computational Project, Number 4, 1994). Histidine residues were included in the calculations and distance cutoffs of 4 and 6 Å were used. The results are summarized in Fig. 5. It is immediately striking that the number of short (<4 Å) ion-pair interactions is lowered in cUDG compared with hUDG. cUDG has only five salt bridges shorter than 4 Å, while hUDG has 11 (see Fig. 6). Only one such short salt bridge is unique to cUDG, while seven are unique to hUDG. The number of salt bridges shorter than 6 Å is 19 and 23 for cUDG and hUDG, respectively. Especially noteworthy are the three clusters of ion pairs in hUDG that are not present in cUDG. In particular, the salt bridges unique to hUDG stabilize the secondary-structure element α6 and its surroundings (see also Fig. 3). Residue Glu171 participates in two ion-pair interactions (Glu171–Lys175 and Glu171–His186), where one is an intrahelical ion pair and the other runs from α6 to the α6/α7 loop. This region is close to the DNA-binding region and also close to the proline-rich loop (4Pro loop; residues 165–169). Both the 4Pro loop and several residues in the α6/α7 loop are mobile upon DNA binding in hUDG, *i.e.* they move more than 1 Å. Residue Asp183 in the α6/α7 loop forms an ion-pair interaction with Lys302 in the C-terminal loop, which is further stabilized by two additional ion-pair interactions (Lys302–Asp300 and Lys302–Glu303).

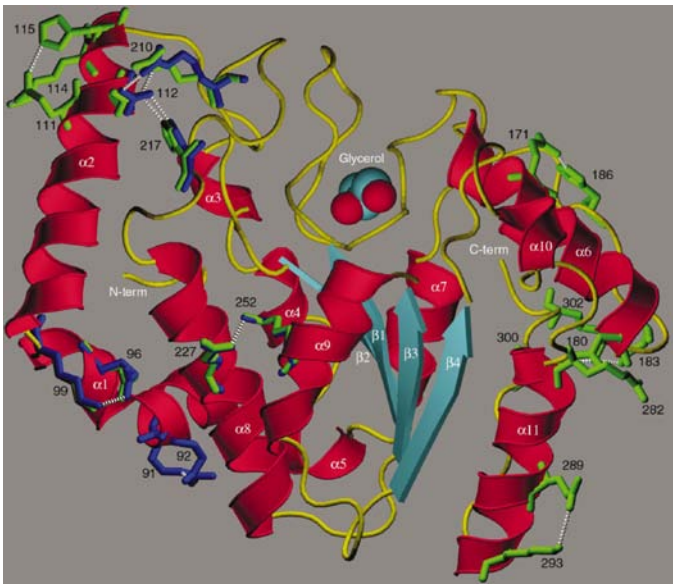


Figure 6
Overall view of the crystal structure of cUDG. Ion-pair interactions (<4 Å) present in the cUDG structure are coloured blue, whereas those found in hUDG are shown in green.

None of the mentioned ion pairs are found in cUDG, in all cases because of amino-acid substitutions. The only ion pairs common to both structures in this region are the two salt bridges connecting the helices α6 and α11. These two ion pairs run from the positively charged residue 282 (Arg in hUDG, Lys in cUDG) to Glu176 and Asp180. Five of the seven unique ion-pair interactions shorter than 4 Å found in the structure of hUDG are located in the C-terminal half of the protein. These are both intrahelical (Glu289–Lys293, both in α6), interhelical [Asp180 (α6)–Arg282 (α11)], helix–loop [Glu171 (α6)–His186 (α6/α7 loop)] and loop–loop [Asp183(α6/α7 loop)–Lys302 (C-terminal loop) and Asp300–Lys302 (both in the C-terminal loop)] interactions. It is likely that the lack of these short unique ion-pair interactions is at least in part responsible for the reduced thermal stability of cUDG. The involved structural elements will also possess increased flexibility in the absence of the close intramolecular interactions. The unique short ion-pair interaction found in cUDG (Arg91–Glu92) is intrahelical and is not believed to influence either the stability or the activity. An additional cluster of ion-pair interactions unique to hUDG is located in the secondary-structure element α5, where three ion pairs are found (Asp133–Lys135, Lys135–Asp136 and Asp136–Lys138). They are all relatively weak, with distances of 4.16–5.52 Å, and are also relatively far away from DNA-interaction sites; however, Asp136 is one of the residues that are mobile upon DNA binding. It should be added that cUDG also has a weak unique ion-pair interaction in this region (Lys138–Glu197), which runs from α5 to the C-terminal loop.

3.6. Hydrogen-bonding interactions

The number of hydrogen-bonding interactions was found to be similar in the two structures, with slightly more hydrogen bonds in cUDG (215) compared with hUDG (201). There are 35 hydrogen bonds only found in cUDG and 21 only found in hUDG. Inspection of the respective crystal structures made it apparent that for 15 of the hydrogen bonds only found in cUDG, the donor and acceptor atoms are also present in hUDG with in most cases only minor offsets in distances and angles. The same number for hUDG is four. This leads to an estimate of 219 and 216 hydrogen bonds in the structures of cUDG and hUDG, respectively. Based on this, a list of unique hydrogen bonds is included in Table 2. From the list, some important hydrogen bonds can be found. The active-site

Table 2
Unique hydrogen bonds in cUDG and hUDG.

The distances marked with asterisks are ion-pair interactions within hydrogen-bonding distances.

Donor	Atom	Acceptor	Atom	Distance
Unique cUDG hydrogen bonds				
Glu87	N	Glu87	$O^{\varepsilon1}$	2.70
Arg90	$N^{\eta2}$	Phe85	O	2.82
Arg91	N^{ε}	Glu92	$O^{\varepsilon2}$	2.80*
Arg91	$N^{\eta2}$	Gln135	$O^{\varepsilon1}$	2.70
Ser114	O^{γ}	Asp111	O	2.76
Lys138	N^{ζ}	Glu297	O	2.73
Gln161	$N^{\varepsilon2}$	Ala153	O	2.93
Trp232	$N^{\varepsilon1}$	Glu92	$O^{\varepsilon1}$	2.96
Asn236	$N^{\delta2}$	Trp232	O	2.91
Arg237	$N^{\eta1}$	Ile134	O	3.04
Arg237	$N^{\eta1}$	Val137	O	3.00
Arg237	$N^{\eta2}$	Val137	O	3.15
His250	$N^{\varepsilon2}$	Leu244	O	3.13
Arg258	$N^{\eta1}$	Gly253	O	3.16
Arg258	$N^{\eta1}$	Ile256	O	2.97
Arg260	N^{ε}	Ser234	O	3.01
His268	$N^{\varepsilon2}$	Gln144	$O^{\varepsilon1}$	3.02
Arg276	$N^{\eta1}$	Leu272	O	3.07
Thr296	$O^{\gamma}1$	Ser294	O	3.42
Unique hUDG hydrogen bonds				
Glu83	N	Glu83	$O^{\varepsilon1}$	2.88
Lys114	N^{ζ}	Glu111	$O^{\varepsilon1}$	3.09*
His115	$N^{\varepsilon2}$	Glu111	$O^{\varepsilon2}$	2.75*
Lys135	N	Asp133	$O^{\delta1}$	3.11
Lys138	N^{ζ}	Asp136	O	3.32
Lys138	N^{ζ}	Gln198	O	2.94
Gln161	$N^{\varepsilon2}$	Asp191	$O^{\delta2}$	3.07
His186	$N^{\varepsilon2}$	Glu171	$O^{\varepsilon1}$	3.05*
Ser231	O^{γ}	Gln235	$O^{\varepsilon1}$	2.90
Ser254	O^{γ}	Lys251	O	2.68
Arg282	$N^{\eta2}$	Asp180	O	2.91
Arg282	N^{ε}	Asp180	$O^{\delta1}$	2.81*
Arg282	$N^{\eta2}$	Asp180	$O^{\delta1}$	3.02*
Thr287	$O^{\gamma1}$	His283	O	2.73
Ser294	O^{γ}	Leu291	O	3.15
Lys296	N^{ζ}	Asn238	O	2.92
Lys302	N^{ζ}	Asp183	$O^{\delta2}$	3.14*

residue His268 $N^{\varepsilon2}$ is hydrogen bonded to Gln144 $O^{\varepsilon1}$ owing to a conformational change in cUDG compared with hUDG (Gln144 adopts a different conformation). Arg276 $N^{\eta1}$ is hydrogen bonded to the carbonyl O atom of Leu272 owing to a positional shift of residue Arg276, which is forced by the increased size of the side chain of residue 267, which is Ala in hUDG and the larger Val in cUDG.

3.7. DNA interaction

Several crystal structures of UDG complexed with DNA have been determined for both hUDG (Parikh *et al.*, 1998; Slupphaug *et al.*, 1996) and herpes simplex virus UDG (Savva *et al.*, 1995). Residues whih were mobile in response to DNA binding were found by superimposing the structures of hUDG with and without DNA (PDB codes 1akz and 1emh, respectively). These amino-acid residues are also indicated in Fig. 3. If the amino-acid residues either in contact with DNA during catalysis (Parikh *et al.*, 1998) or immediately before or after these are considered, almost all are totally conserved in all mammalian sequences. However, some exceptions were found, particularly regarding residues close to or in the minor-groove intercalation loop (Leu272 loop; residues 268–276). Residues 266, 267, 274 and 275 are all substituted in cUDG compared with hUDG. Ala266 and Ala274 (Thr and Val, respectively, in most other vertebrates) are only found in cUDG and zebrafish UDG, while Val267 is found in cod, chicken, frog and zebrafish UDG; most of the other sequences have Ala in this position. Residue 275 is His in cUDG, which is also observed in UDG from frog, zebrafish and in one of the sequences of mouse UDG, while it is Tyr in hUDG. An interesting finding is that chicken UDG has Asn in this position. Tyr275 makes hydrophobic interactions with Val274 and Phe279, while all these residues are substituted by smaller residues in cUDG (His275, Ala274 and Leu279), allowing increased motion during DNA binding and catalysis. The Leu272 loop is essential in all stage of catalysis, as it is involved in damage recognition, DNA compression and nucleotide flipping and contains the essential catalytic residue His268. Upon DNA binding, the whole residue range 265–282 moves more than 1 Å, with the largest movement seen for residues 271–276, with shifts of about 2 Å. It is therefore very likely that these substitutions are important for the differences in catalytic activities observed between hUDG and cUDG.

For the other regions of the protein either close to or in contact with DNA during catalysis, no substitutions are observed for cUDG. However, other sequences have substitutions in these positions: UDG from chicken has a unique Leu272Phe substitution and chicken and frog UDG have Ser247Ala. The latter substitution is in the Gly-Ser loop (residues 246–247) that contacts DNA during binding. However, it is the backbone of residue 247 that contributes to the interaction; therefore, this is not a critical substitution. Many residues have a large movement upon DNA binding, but the major movement is observed in three main interaction regions: the 4-Pro loop (residues 165–169), the Gly-Ser loop (residues 246–247) and the Leu272 loop (residues 268–276). In addition to these regions, mobility is observed in the α-helices $\alpha2$, $\alpha9$ and $\alpha10$ (see also Fig. 4). $\alpha9$ is immediately after the Gly-Ser loop, but the largest differences are actually seen at the C-terminal end of this helix, on the opposite side of the DNA-binding site. $\alpha10$ succeeds the Leu272 loop and contains the DNA-contacting residue Tyr/His275; therefore, movement is expected in this region. Of particular interest is the mobility seen at the C-terminal end of $\alpha2$, where two short and unique ion-pair interactions are found in hUDG. Although long ion-pair interactions are also found in cUDG in this region, the stabilization is believed to be greater for hUDG. Unfortunately, no success in obtaining crystal structures of cUDG–DNA complexes has been achieved thus far.

3.8. Hydrophobic residues and core packing

The number of internal residues and the size of the hydrophobic core are in some cases been less optimized in the cold-adapted enzymes compared with their mammalian counterparts (Smalås *et al.*, 1994; Leiros *et al.*, 2000; Genicot *et al.*, 1996; Aghajari *et al.*, 1998). Based on the water-accessible surface (WAS) of the structures of hUDG and cUDG, 77 and

76 residues are considered as internal, respectively (see also Fig. 3). Of the 53 amino-acid substitutions in cUDG compared with hUDG, seven of these are considered internal, *i.e.* having a water-accessible surface of less than 10 Å^2, in the crystal structure of either hUDG or cUDG. The following internal residues are substituted in cUDG compared with hUDG: Thr127Ser, Val230Ile, Leu240Val, Val274Ala and Thr287Ala. In addition, Phe126Tyr is internal in hUDG but not in cUDG, and Asn234Ser and Thr266Ala are internal in cUDG but not in hUDG. As previously reported (Lanes *et al.*, 2002), there is a decrease in the size of the hydrophobic core for cUDG compared with hUDG. This is confirmed by calculating the volume of the residues considered internal in both structures: a small decrease in the calculated volume for cUDG can be seen compared with hUDG. However, as the substituted internal residues are among the 'outer-core' residues, *i.e.* the outermost layer of internal residues, and not in the central core of the protein, the importance of these substitutions is therefore not clear. On the other hand, considering that the reduction of the size of the core residues will generate cavities in the protein, it is a well known fact that such cavities lead to a decreased stability of the protein. The cavities in the two structures were calculated using routines implemented in the *MolSoft ICM* program (Abagyan *et al.*, 1994). cUDG has an increase in both the number and the volume of the internal cavities. The volumes of the enclosed cavities for cUDG and hUDG are 482 and 341 Å^3, respectively. In particular, there are additional cavities formed in cUDG both close to the active site and in the 4Pro loop involved in DNA binding.

4. Conclusion

The crystal structure of cUDG and its comparison with the already published crystal structures of human and *Escherichia coli* UDGs pinpoints important differences regarding the cold-adaptation features observed for cUDG (Lanes *et al.*, 2002). In particular, the reduction of strong ion-pair interactions in the C-terminal half of cUDG are believed to largely affect the flexibility and/or stability. The optimization of the electrostatic surface potential in the DNA-facing side of cUDG (increase in the number of basic amino acids) in terms of increasing the affinity for the negatively charged DNA is likely to increase the activity of cUDG compared with hUDG. Mutational studies are now in progress in order to test the significance of these observations.

This work has been supported by the Research Council of Norway and Biotec-Pharmacon ASA, Tromsø, Norway. The organizers at the Swiss–Norwegian Beamline (SNBL) at the European Synchrotron Radiation Facility (ESRF) are acknowledged for supplying beamtime.

References

Abagyan, R. A., Totrov, M. M. & Kuznetsov, D. N. (1994). *J. Comput. Chem.* **15**, 488–506.
Aghajari, N., Feller, G., Gerday, C. & Haser, R. (1998). *Protein Sci.* **7**, 564–572.
Alvarez, M., Zeelen, J. P., Mainfroid, V., Rentier-Delrue, F., Martial, J. A., Wyns, L., Wierenga, R. K. & Maes, D. (1998). *J. Biol. Chem.* **273**, 2199–2206.
Berglund, G. I., Willassen, N. P., Hordvik, A. & Smalås, A. O. (1995). *Acta Cryst.* D**51**, 925–937.
Brandsdal, B. O., Smalas, A. O. & Aqvist, J. (2001). *FEBS Lett.* **499**, 171–175.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.*, D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst* D**50**, 760–763.
Feller, G., Payan, F., Theys, F., Qian, M., Haser, R. & Gerday, C. (1994). *Eur. J. Biochem.* **222**, 441–447.
Feller, G., Zekhnini, Z., Lamotte-Brasseur, J. & Gerday, C. (1997). *Eur. J. Biochem.* **244**, 186–191.
Genicot, S., Rentier-Delrue, F., Edwards, D., VanBeeumen, J. & Gerday, C. (1996). *Biochim. Biophys. Acta*, **1298**, 45–57.
Gorfe, A. A., Brandsdal, B. O., Leiros, H. K., Helland, R. & Smalas, A. O. (2000). *Proteins*, **40**, 207–217.
Hochachka, P. & Somero, G. N. (1984). *Biochemical Adaptation*, pp. 355–449. New Jersey: Princeton University Press.
Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Karlsen, S., Hough, E. & Olsen, R. L. (1998). *Acta Cryst.* D**54**, 32–46.
Kim, S. Y., Hwang, K. Y., Kim, S. H., Sung, H. C., Han, Y. S. & Cho, Y. (1999). *J. Biol. Chem.* **274**, 11761–11767.
Lanes, O., Leiros, I., Smalås, A. O. & Willassen, N. P. (2002). *Extremophiles*, **6**, 73–86.
Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leiros, H. K., Willassen, N. P. & Smalas, A. O. (1999). *Extremophiles*, **3**, 205–219.
Leiros, H. K. S., Willassen, N. P. & Smalas, A. O. (2000). *Eur. J. Biochem.* **267**, 1039–1049.
Leiros, I., Lanes, O., Sundheim, O., Helland, R., Smalås, A. O. & Willassen, N. P. (2001). *Acta Cryst.* D**57**, 1706–1708.
Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E. & Tainer, J. A. (1995). *Cell*, **80**, 869–878.
Mol, C. D., Parikh, S. S., Putnam, C. D., Lo, T. P. & Tainer, J. A. (1999). *Annu. Rev. Biophys. Biomol. Struct.* **28**, 101–128.
Mrabet, N. T., Van den Broeck, A., Van den brande, I., Stanssens, P., Laroche, Y., Lambeir, A. M., Matthijssens, G., Jenkins, J., Chiadmi, M., van Tilbeurgh, H., Rey, F., Janin, J., Quax, W. J., Lasters, I., De Mayer, M. & Wodak, S. J. (1992). *Biochemistry*, **31**, 2239–2253.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E. & Tainer, J. A. (1998). *EMBO J.* **17**, 5214–5226.
Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M. & Tainer, J. A. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 5083–5088.
Russell, R. J., Gerike, U., Danson, M. J., Hough, D. W. & Taylor, G. L. (1998). *Structure*, **6**, 351–361.
Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995). *Nature (London)*, **373**, 487–493.
Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E. & Tainer, J. A. (1996). *Nature (London)*, **384**, 87–92.
Smalås, A. O., Heimstad, E. S., Hordvik, A., Willassen, N. P. & Male, R. (1994). *Proteins*, **20**, 149–166.
Smalås, A. O. & Hordvik, A. I. (1993). *Acta Cryst.* D**49**, 318–330.
Smalås, A. O., Leiros, H.-K. S., Os, V. & Willassen, N. P. (2000). *Biotechnol. Annu. Rev.* **6**, 1–57.
Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T. & Gilliland, G. L. (1999). *Proteins*, **35**, 13–24.