Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the uracil-DNA *N*-glycosylase (UNG) from *Deinococcus radiodurans*

Ingar Leiros,[a] Elin Moe,[b] Arne O. Smalås[b] and Sean McSweeney[a]*

[a]The European Synchrotron Radiation Facility, 38043 Grenoble CEDEX 9, France, and [b]The Norwegian Structural Biology Centre, University of Tromsø, N-9037 Tromsø, Norway

Correspondence e-mail: mcsweeney@esrf.fr

Received 31 January 2005
Accepted 2 May 2005

PDB Reference: drUNG, 2boo, r2boosf.

Uracil-DNA glycosylases are DNA-repair enzymes that catalyse the removal of promutagenic uracil from single- and double-stranded DNA, thereby initiating the base-excision repair (BER) pathway. Uracil in DNA can occur by misincorporation of dUMP in place of dTMP during DNA synthesis or by deamination of cytosine, resulting in U–A or U–G mispairs. The radiation-resistant bacterium *Deinococcus radiodurans* has an elevated number of uracil-DNA glycosylases compared with most other organisms. The crystal structure of dr0689 (uracil-DNA *N*-glycosylase), which has been shown to be the major contributor to the removal of misincorporated uracil bases in crude cell extracts of *D. radiodurans*, is reported.

1. Introduction

The maintenance of genomic integrity is crucial to all organisms. DNA damage can be caused through a multitude of factors, *e.g.* UV radiation, γ radiation, chemical mutagens and the intrinsic error rate of the DNA-replication machinery. The natural selection for organisms capable of maintaining genetic information intact is strong and the genomes of all cellular life forms encode many proteins whose primary function is to repair damaged DNA (Friedberg *et al.*, 1995).

The non-pathogenic soil bacterium *Deinococcus radiodurans* is well known for its extreme ability to withstand UV and ionizing radiation, as well as desiccation. *D. radiodurans* tolerates ionizing radiation at doses that are lethal to other organisms and is capable of surviving 5000–30000 Gy of ionizing radiation (Minton, 1994), whereas most other organisms cannot survive radiation doses of greater than 50 Gy. This massive irradiation dose is estimated to induce hundreds of double-strand breaks, thousands of single-strand gaps and about 1000 sites of DNA-base damage per chromosome (Battista, 1997, and references therein). In the case of acute irradiation, *D. radiodurans* repairs its DNA efficiently and without error within some hours. Typically, other organisms cannot tolerate more than two to three radiation-induced double-strand breaks per chromosome (Daly & Minton, 1996).

Uracil-DNA glycosylases (UDGs) are widespread enzymes that are found in all living organisms (Kavli *et al.*, 2002). It has been shown that *Escherichia coli* or yeast cells carrying mutations in the gene for UDG show a several-fold increased spontaneous mutation rate (Duncan & Weiss, 1982; Impellizzeri *et al.*, 1991). UDGs form a central part of the DNA-repair machinery since they initiate the DNA base-excision repair pathway (BER) by hydrolysing the N-glycosidic bond between uracil and the deoxyribose sugar, thereby catalysing

© 2005 International Union of Crystallography
Printed in Denmark – all rights reserved

the removal of mis-incorporated uracil from DNA (Friedberg *et al.*, 1995). Four main members of the uracil-DNA glycosylase (UDG) family have so far been identified. These are (1) the uracil-DNA *N*-glycosylases (UNGs), (2) the mismatch-specific uracil DNA glycosylases (MUGs, dsUDGs), (3) the single-strand selective mono-functional UDGs (SMUGs, ssUDGs) and (4) the [4Fe–4S] cluster-containing members of the uracil-DNA glycosylases, typified by the UDG from *Thermus thermophilus* (ttUDG; Aravind & Koonin, 2000; Pearl, 2000). Whereas most prokaryotes possess either a UDG of type 1 or of type 4, *D. radiodurans* is exceptional in that it contains four putative UDG enzymes (Aravind & Koonin, 2000), of which UDG activity has been confirmed for three: the UDG enzymes of type 1, 2 and 4 (Sandigursky *et al.*, 2004).

In *D. radiodurans*, the gene encoding the type 1 UDG (also known as UNG) is dr0689 (drUNG). This gene encodes a 247-amino-acid protein with a deduced molecular weight of 27.7 kDa and a predicted pI of 6.97. A recent study has confirmed that drUNG is able to excise uracil from both U–A and U–G double-stranded DNA and single-stranded DNA (Sandigursky *et al.*, 2004). drUNG and topoisomerase 1B have been suggested to form a DNA-repair operon in *D. radiodurans*, acquired by horizontal transfer from a eukaryote or *via* a eukaryotic virus (Makarova *et al.*, 2001).

The uracil-DNA glycosylase inhibitor protein (Ugi) from the *Bacillus subtilis* bacteriophage PBS1/2 (Cone *et al.*, 1980) has previously been reported to be incapable of inhibiting type 2 MUGs (Barrett *et al.*, 1998), type 3 SMUGs (Haushalter *et al.*, 1999; Nilsen *et al.*, 2001) or type 4 UDGs (Sandigursky *et al.*, 2004), whereas type 1 UNGs are readily inhibited by Ugi (Mol, Arvai, Slupphaug *et al.*, 1995). The addition of Ugi to *D. radiodurans* crude cell extracts reduced the overall enzyme activity by greater than 95%, indicating that drUNG is responsible for the major uracil-DNA glycosylase activity in *D. radiodurans* (Sandigursky *et al.*, 2004).

In this study, we have determined the crystal structure of drUNG to 1.8 Å resolution and performed a biochemical and structural comparison with the *E. coli* and human UNG enzymes.

2. Materials and methods

2.1. Cloning

The gene encoding the uracil-DNA *N*-glycosylase (UNG) from *D. radiodurans* (dr0689; drUNG) was inserted into the pDEST14 expression vector using Gateway technology (Invitrogen). The gene was amplified using PCR in a GeneAmp 9700 Thermocycler (Perkin Elmer). Each reaction had a volume of 50 µl and contained 1 U Platinum Pfx DNA polymerase (Invitrogen), buffer and $MgCl_2$ supplied by the manufacturer, 0.1 m*M* dNTPs and 1 µ*M* of each of the upstream and downstream primers in addition to the template (genomic DNA of *D. radiodurans*). The amplification was carried out at 369 K for 5 min followed by 30 cycles at 369 K for 30 s, 323 K for 1 min and 345 K for 2 min and a final extension step at 345 K for 7 min. The primers used for amplification of the gene were as follows (Sigma Genosys): FDRUNG, 5′-**CAT CAC CAT CAC CAT CAC** ACC GAC CAA CCC GAC CTG-3′; RDRUNG, 5′-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC CTA TTC CTC CGT CAC CGT GGC-3′; FDRHISTAG, 5′-G GGG ACA AGT TTG TAC AAA AAA GCA GGC TTC GAA GAT AGA ACC ATG **CAT CAC CAT CAC CAT CAC**-3′. The drUNG gene, with 18 additional nucleotides encoding an N-terminal hexahistidine tag at the 5′ end (in bold in the FDRUNG primer), was first amplified using the FDRUNG and RDRUNG primers. The resulting PCR product was purified from a 1% agarose gel using the Qiaquick Gel Extraction Kit (Qiagen) and used in a new PCR reaction using the FDRHISTAG (the nucleotides encoding the His tag are in bold) and the RDRUNG primers. The pDEST14 expression vector does not contain nucleotides encoding an N-terminal His tag; they were thus included in the forward primers used for amplification of the gene. The PCR product from the second PCR reaction was also purified from a 1% agarose gel and used in the ligase-free cloning procedure using the Gateway technology (Invitrogen) according to the manual from the manufacturer. The donor vector used in this procedure was pDONR201. The insertion of the gene was verified by PCR using the FDRUNG and RDRUNG primers and sequencing of the pDONR201 vector with the drUNG gene using primers specific to the vector applied by the manufacturer. The sequencing was performed using the ABI sequencing chemistry and an ABI 377 DNA sequencer (Amersham Biosciences).

2.2. Expression and purification

The BL21(DE3)pLysS expression strain (Stratagene) was transformed with the pDEST14 vector containing the drUNG gene using heat-shock. A resulting transformant was used in large-scale expression in 1 l LB medium containing 100 µg ml^{-1} ampicillin. The cells were grown by vigorous shaking at 310 K until the optical density at 600 nm (OD_{600}) reached 0.6 and expression was induced by adding 0.5 m*M* isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were harvested 3 h post-induction by centrifugation at 5000*g* for 30 min, resuspended in lysis buffer (50 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl and 1 m*M* $MgCl_2$); proteinase inhibitors (Roche), DNase I (Sigma-Aldrich) and lysozyme (Sigma-Aldrich) were added and the cells were disrupted using a cell disruptor (Stansted Fluid Power Ltd, UK). The recombinant protein was isolated from the cell debris by centrifugation at 21 000*g* for 30 min. The resulting protein extract was loaded onto a HiTrap HP Chelating Sepharose column (5 ml) charged with 100 m*M* $NiSO_4$ and equilibrated with buffer *A* (50 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl). The protein was eluted from the column using a gradient of 0–100% buffer *B* (50 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl and 500 m*M* imidazole) and contained approximately 200 m*M* imidazole. Coomassie-stained SDS–PAGE showed that the eluted protein was virtually pure, with an estimated molecular weight of

approximately 28 kDa, as would be expected for drUNG. The protein was then concentrated to 10 mg ml^{-1} and stored at 277 K.

2.3. Activity measurements

Nick-translated calf thymus DNA (Sigma-Aldrich) with deoxy[5-^{3}H]uridine 5′-triphosphate (Amersham Biosciences) was used as a substrate for measurements of general enzyme activity. Preparation of substrate and measurement of drUNG activity was performed as previously described (Lanes *et al.*, 2000). The standard assay procedure is as follows. The enzyme was diluted in cold dilution buffer (5 m*M* Tris–HCl, 10 m*M* NaCl, 1% glycerol pH 8.0). The activity was measured in a total volume of 20 µl in 70 m*M* Tris–HCl, 10 m*M* NaCl, 1 m*M* EDTA pH 8.0, 100 µg ml^{-1} BSA and 230 ng substrate (^{3}H-dUMP DNA). The reaction mixture was incubated for 10 min at 310 K and terminated by the addition of 20 µl ice-cold single-stranded calf thymus DNA (1 mg ml^{-1}) and 500 µl 10% TCA. After incubation on ice for 15 min, free ^{3}H uracil was separated from precipitated material by centrifugation at 16000*g* for 10 min and analysed using a liquid-scintillation counter. One unit of enzyme activity is defined as the amount of enzyme required for the release of 1 nmol acid-soluble uracil per minute at 310 K. K_{m} and k_{cat} were measured in the presence of eight different substrate concentrations in the range 0.56–4.5 µ*M* at 310 K. Calculation of the kinetic constants was performed using the enzyme-kinetics module in *SigmaPlot* (SPSS Inc.).

2.4. Crystallization and data collection

drUNG was crystallized using the hanging-drop vapour-diffusion method. The best crystals were grown by mixing 1 µl drops of 10 mg ml^{-1} protein solution with a solution containing 0.2 *M* ammonium nitrate and 17%(*w*/*v*) PEG 3000. The drops were equilibrated at 277 K and after a few days stacks of crystalline plates fanning out from a single nucleation point appeared. Despite several attempts, these conditions could not be improved and although of relatively poor quality these crystals were used for data collection. The crystals used had overall dimensions of about 10 × 50 × 150 µm. 20%(*v*/*v*) glycerol added to the reservoir solution sufficed as a cryoprotectant for flash-cooling the crystals in a nitrogen cold stream (Oxford Instruments) operating at 100 K. A data set (see Table 1) was collected at the macromolecular crystallography beamline ID23-1, reaching a resolution of 1.8 Å.

Table 1
Data-collection and refinement summary.

Values in parentheses are for the highest resolution shell.

Data collection	
Resolution range (Å)	12–1.8 (1.9–1.8)
No. of unique reflections	40425
Redundancy	2.5 (2.4)
R_{merge}† (%)	10.1 (27.1)
Completeness (%)	93.7 (95.3)
Mean $I/\sigma(I)$	6.5 (3.5)
Refinement statistics	
R value (%)	23.6
Free *R* value (%)	27.2
Deviation from ideal geometry	
Bond lengths (Å)	0.022
Bond angles (°)	1.841
ESU (Å) (R_{work}/R_{free})‡	0.11/0.12
Average *B* values (Å^{2})	
Main-chain atoms	24.5
Side-chain atoms	26.8
Nitrates (3)	28.5
Water molecules (154)	32.6
All atoms	26.2
Ramachandran plot (%)	
Most favoured	89.6
Additionally allowed	9.4
Generously allowed	1.0

† $R_{merge} = \sum_h \sum_i |I_i(h) - \langle I(h)\rangle| / \sum_h \sum_i I(h)$, where $I_i(h)$ is the *i*th measurement of reflection *h* and $\langle I(h)\rangle$ is the weighted mean of all measurements of *h*. ‡ Estimated overall coordinate error from *REFMAC*5 based on maximum likelihood.

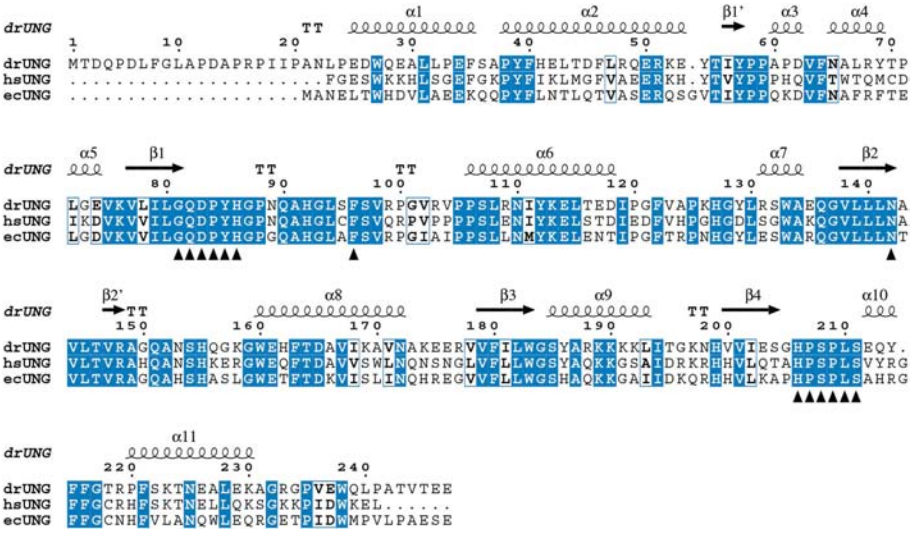

Figure 1
Sequence alignment of *D. radiodurans*, human and *E. coli* UNGs. The secondary-structure elements of drUNG are indicated above the alignment: α-helices are indicated as spirals, β-strands as arrows and turns as 'TT'. Identical residues are indicated as white letters on a blue background, while residues with similar properties are shown as bold letters in blue boxes. Residues important for catalysis and substrate specificity are indicated with black triangles below the sequences.

2.5. Structure determination and refinement

The data were indexed, integrated and scaled using the *XDS* program package (Kabsch, 1993) before being converted to structure factors using the *CCP*4 program *TRUNCATE* (Collaborative Computational Project, Number 4, 1994). The data-collection statistics are presented in Table 1. The crystals were orthorhombic, with unit-cell parameters $a = 63.5$, $b = 85.1$, $c = 85.9$ Å, $\alpha = \beta = \gamma = 90°$. The systematic absences in the collected data set only allowed

the determination of one twofold screw axis, along the *b* axis. The solvent content was estimated to be around 70%, with a Matthews coefficient as high as 4.2 Å^3 Da^{-1}, assuming one protein molecule per asymmetric unit. The crystal structure of drUNG was determined by molecular-replacement methods using *MOLREP* (Collaborative Computational Project, Number 4, 1994). The crystal structure of human UNG (Mol, Arvai, Slupphaug *et al.*, 1995) was used as the search model and the automated program functions in *MOLREP* were applied in order to create the model that presumably had the best fit to the sequence of drUNG. Reflections were used to a high-resolution limit of 3.5 Å. The correct space group was determined by trial-and-error by analysis of the molecular-replacement results run in all possible orthorhombic space groups. Only one well resolved solution could be found, in space group $P2_12_12$, having a correlation coefficient of 0.393 and an *R* factor of 45.3%. A rigid-body fitting of the model using a high-resolution cutoff at 3.5 Å resulted in an R_{work} of 41.4% (R_{free} of 40.5%). Automated model building with *ARP*/*wARP* (Perrakis *et al.*, 1999) including all reflections to 1.8 Å built a total of 221 amino-acid residues in two chains into electron density. After manual intervention using *O* (Jones *et al.*, 1991), the model was refined in *REFMAC*5 (Murshudov *et al.*, 1999), resulting in *R* factors of 24.4 and 28.7% for the working and test sets of reflections, respectively. Subsequent cycles of refinement interspersed with manual rebuilding gave final R_{work} and R_{free} values of 23.6 and 27.2%, respectively, with acceptable protein geometry. The final model of drUNG consists of 230 amino-acid residues in a single polypeptide chain comprising residues 17–246 in the amino-acid sequence; in addition there are three nitrates and 154 water molecules. The structure has been deposited in the Protein Data Bank with accession code 2boo. No electron density is visible for the C-terminal residue Glu247 or for the N-terminal residues 1–16 in addition to the six-histidine tag. For an overview of the refinement statistics, see Table 1.

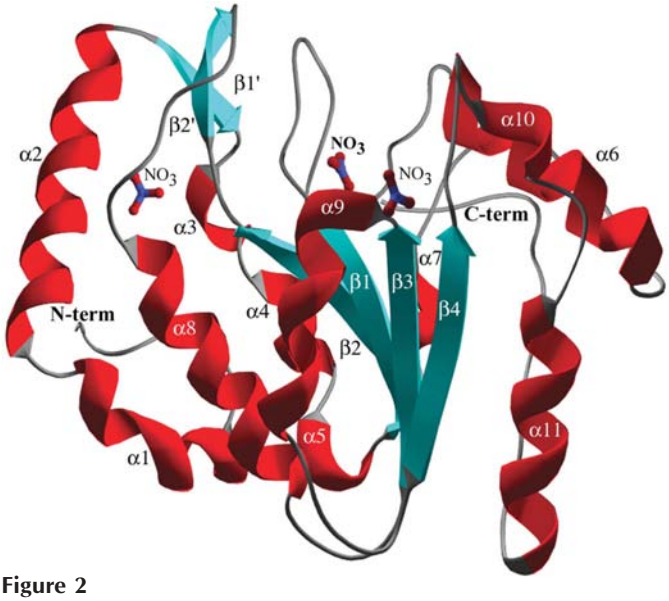

Figure 2
Ribbon illustration of the crystal structure of drUNG. The nitrate molecules are shown as ball-and-stick representations in atom colour, with the one bound in the active site labelled in bold.

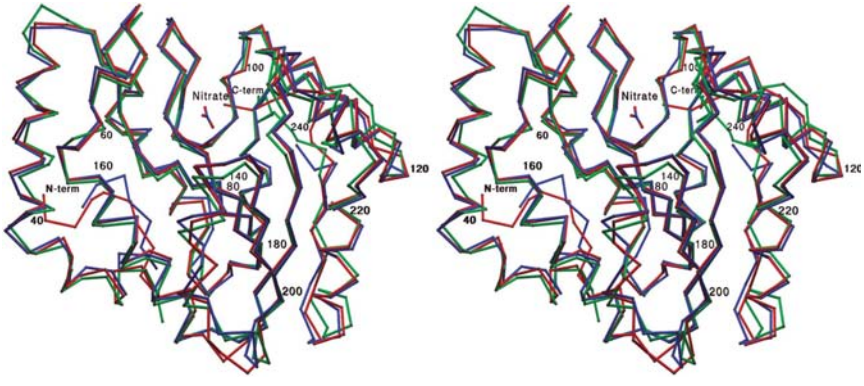

Figure 3
Stereoview of the superpositioning of the crystal structures of drUNG (red), hUNG (blue; PDB code 1akz; Mol, Arvai, Sanderson *et al.*, 1995) and ecUNG (green; PDB code 2eug; Xiao *et al.*, 1999).

3. Results and discussion

3.1. Overall structure

The overall dimensions of drUNG are approximately 28 × 38 × 46 Å. The UNG family of proteins exhibit a common fold that is well conserved between the crystal structures determined to date, *i.e.* human (Mol, Arvai, Sanderson *et al.*, 1995), *E. coli* (Xiao *et al.*, 1999), herpes simplex virus (Savva *et al.*, 1995) and Atlantic cod (Leiros *et al.*, 2003) uracil-DNA glycosylases. UNG is an α/β protein and in drUNG the overall topology is as described for human UNG (Mol, Arvai, Sanderson *et al.*, 1995). Briefly, a central four-stranded parallel β-sheet ($\beta1$–$\beta4$ in Figs. 1 and 2) with strand order 2-1-3-4 is surrounded on both sides by a total of 11 α-helices. In addition, two short β-strands ($\beta1'$ and $\beta2'$ in Figs. 1 and 2) form an antiparallel interaction. The N- and C-termini of drUNG are situated on opposite sides of the central β-sheet. The sequence identity between drUNG and the catalytic domain of the human UNG (hUNG) and *E. coli* UNG (ecUNG) is 50% in both cases. The overall root-mean-square (r.m.s.) deviation between drUNG and the catalytic domain of hUNG (Mol, Arvai, Sanderson *et al.*, 1995) is 0.90 Å based on the main-chain atoms of the 211 residues that could be superpositioned. For the alignment of ecUNG (Xiao *et al.*, 1999) onto drUNG, the r.m.s. deviation was calculated to be 0.86 Å for the 209 residues that could be superpositioned. The overall similarity between these proteins can also be seen in Fig. 3.

3.2. Consideration of DNA binding, the active-site environment and uracil specificity

The active site of the uracil-DNA glycosylases is positioned on the C-terminal side

Table 2
Kinetic constants for drUNG and hUNG at 310 K.

UDG	T (K)	V_{max} (U mg^{-1})	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} μM^{-1})
drUNG	310	34054 ± 1412	935	0.7 ± 0.11	1263
hUNG†	310	25347 ± 1256	647	2.1 ± 0.23	309

† From Moe *et al.* (2004).

of the central four-stranded β-sheet (Slupphaug *et al.*, 1996). In this region, there is a pronounced conical cleft shaped to bind double-stranded DNA. In the middle of this cleft, there is a cavity designed to accommodate a uracil base flipped-out of the double-stranded DNA. Upon DNA binding, about 1000 Å^2 of the protein is buried in the DNA–protein interface (Slupphaug *et al.*, 1996). Structural studies of hUNG in complex with DNA have revealed substantial conformational shifts in the proximity of the DNA-binding region (Parikh *et al.*, 1998, 2000; Slupphaug *et al.*, 1996). Compression of the DNA phosphates flanking the uracil has been suggested to be the first step in damage detection, followed by the damaged nucleotide being pushed out of the minor groove and into the active site of the enzyme (Parikh *et al.*, 1998). The amino-acid residues responsible for these steps have been identified as three Ser-Pro-rich loops, where the corresponding residues in drUNG are 105-PPS-107, 184-GS-185 and 206-HPSPLS-211. Residues in the 206-HPSPLS-211-loop (often referred to as the Leu-loop) also push the damaged nucleotide out of the minor groove, where Leu272 in hUNG (Leu210 in drUNG) replaces the flipped-out uracil nucleotide in the DNA substrate (Slupphaug *et al.*, 1996).

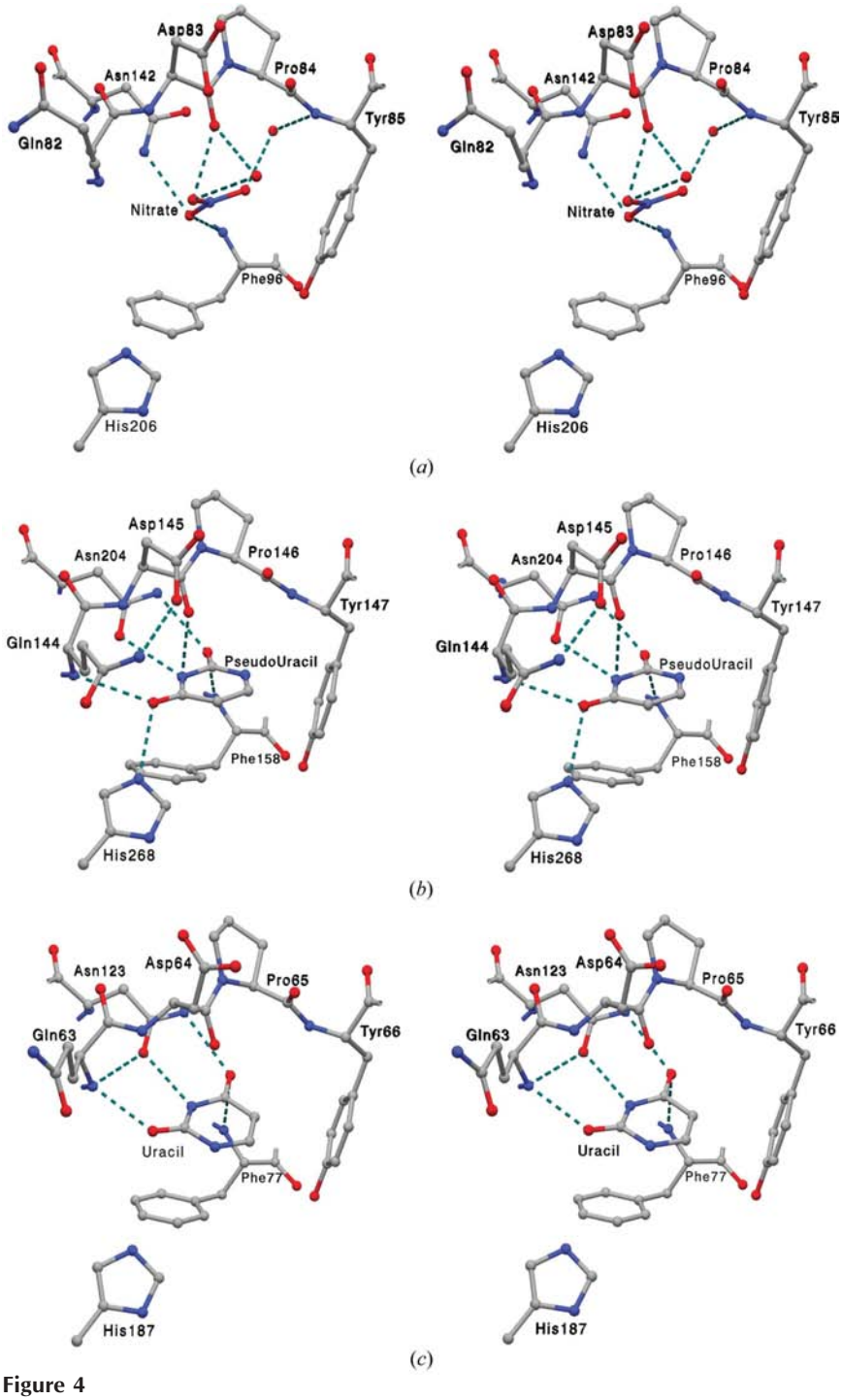

Figure 4
Stereoview showing the active-site environment. (*a*) drUNG with nitrate bound in the active site, (*b*) hUNG in complex with the substrate analogue pseudouracil (PDB code 1emh; Parikh *et al.*, 2000), (*c*) ecUNG in complex with uracil (PDB code 2eug; Xiao *et al.*, 1999).

The final step of substrate binding is the orienting of the flipped-out uracil residue into the active site of UNG. The amino-acid residues making up the uracil-binding environment in drUNG are Gln82 (main-chain hydrogen bond), Tyr85 (hydrophobic interaction), Phe96 (stacking interaction and main-chain hydrogen bond), Asn142 (side-chain hydrogen bonds), His206 (hydrogen bond), Asp83 (side chain), Pro84 (carbonyl oxygen) and His86 (side chain). The three latter residues coordinate a water molecule assumed to be catalytically critical; for this reason, the residue range 82–86 is frequently referred to as the water-activating loop (Parikh *et al.*, 1998). In the crystal structure of drUNG, three nitrate molecules were identified, one of which is situated in the active-site pocket in a position almost overlapping that of the uncleavable substrate analogue pseudouracil in the crystal structure of hUNG and also to the uracil bound in the active site of ecUNG (Fig. 4). When the residues in the uracil-binding environment of drUNG are superpositioned with the corresponding residues in the crystal structures of free hUNG (PDB code 1akz) and ecUNG-uracil (PDB code 2eug), the r.m.s. deviations for main-

chain atoms are 0.35 and 0.37 Å, respectively, which reflects the strong conservation of this region in UNG enzymes from different organisms.

It has previously been shown that as for hUNG and ecUNG, drUNG removes uracil both from single-stranded DNA, as well as G–U and A–U mismatches in double-stranded DNA (Sandigursky *et al.*, 2004). The difference in efficiency of uracil excision from A–U mismatches between drUNG and hUNG has been measured using a standard UDG assay system. The results obtained (Table 2) show that the specific activity of drUNG is four times higher than of hUNG and that the difference in activity is mainly caused by a threefold increased substrate affinity (reduced K_m). An increased substrate affinity could possibly be caused in part by an increased positive surface charge close to the active site of the enzyme: the part of the enzyme mainly involved in contacting the negatively charged DNA substrate. A comparison of amino-acid properties from the gene sequences shows that hUNG and drUNG possess 26 and 27 positively charged residues (Arg and Lys) in total, respectively. However, the distribution of positively charged residues is different. For instance, drUNG possesses two additional arginines (Arg103 and Arg109) in the α6-helix area comprising the DNA-interacting 105-PPS-107 residues, an additional Lys (Lys126) in the α6-α7-loop region. In addition there is a cluster of five positive residues (188-RKKKK-192) in drUNG in the α9-helix that follows the catalytically important 184-GS-185 loop. The equivalent helix in hUNG and ecUNG has only two lysine residues.

The difference in the distribution of positively charged residues between drUNG and hUNG was further analysed by visualization of the surface potential of the proteins using *GRASP* (Nicholls *et al.*, 1991). Although the resulting surface potentials are similar, drUNG possesses a pronounced more positive surface patch close to the active site (lower half of Figs. 5*a* and 5*b*). From structural alignments, it is apparent that the increased positive surface patch in drUNG is caused by amino-acid substitutions (hUNG to drUNG) Pro103Arg and Glu109Arg. In order to evaluate whether the increased positive surface potential of drUNG is caused by these substitutions, a double mutant of drUNG where Arg103Pro and Arg109Glu were substituted was modelled (Fig. 5*c*). The resulting surface-potential estimation resembles more closely the charge distribution seen in hUNG. It thus seems possible that the increased substrate affinity of drUNG compared with hUNG is caused by an increased positive surface potential in the DNA-binding area.

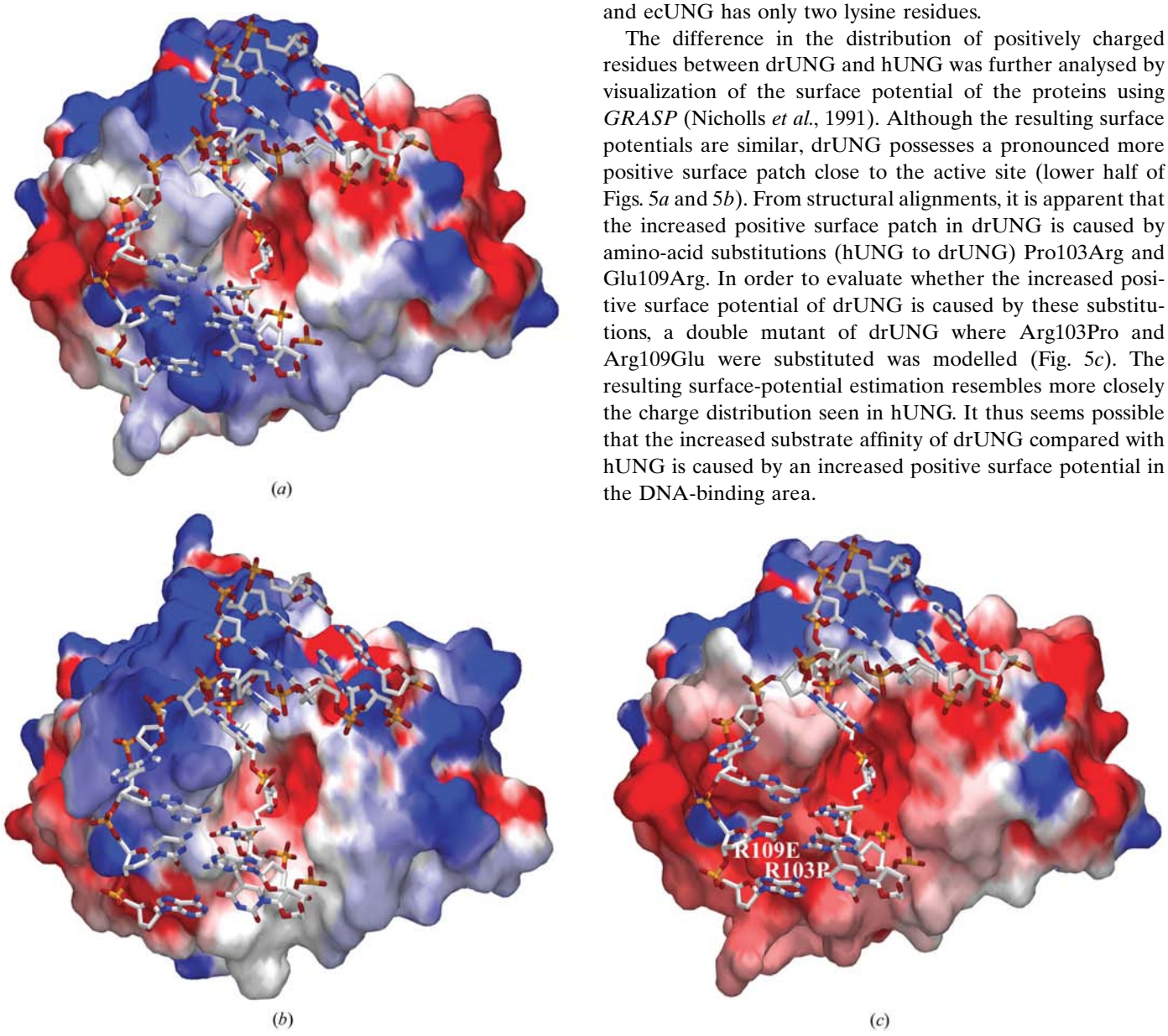

Figure 5
Differences in electrostatic surface potential between drUNG and hUNG. The figure was made using *SwissPDBViewer* (Guex & Peitsch, 1997) with an electrostatic surface potential imported from *GRASP* (Nicholls *et al.*, 1991) and contoured at $\pm 3kT/e$, where red describes a negative potential and blue a positive potential. (*a*) drUNG, (*b*) hUNG complexed with DNA, (*c*) drUNG with substituted residues believed to affect the surface potential. To illustrate the DNA interaction, hUNG in complex with DNA was used to model DNA interacting with the surface of drUNG.

In a similar fashion to drUNG, the cold-adapted cod UNG (cUNG) has an improved binding affinity compared with hUNG. In a recently published mutational analysis study in which cUNG was compared with hUNG (Moe *et al.*, 2004), the estimated surface potentials of the two proteins were studied and residues that affected the binding affinities both in cUNG and hUNG were identified. The results of this study strongly indicated that the residue having the most pronounced effect on binding affinity was Val171 in cUNG (Glu171 in hUNG). Mutating this residue in cUNG into the corresponding residue in hUNG resulted in a weakened binding affinity, with values closely resembling hUNG, and likewise the mutation of Glu171 in hUNG into the Val found in cUNG gave a considerably stronger binding affinity than wild-type hUNG. The rationale for the increased affinity in cUNG was explained as a change towards charge complementarity between the negatively charged DNA substrate and an increasingly positively charged protein-interaction surface (Moe *et al.*, 2004). The corresponding residue in drUNG is Arg109, which is positioned in the positive patch seen in Fig. 5(*a*), perhaps a further indication of its involvement in the DNA-binding and catalytic efficiency of drUNG.

In addition to initial uracil-damage detection by DNA-phosphate backbone compression caused by the residues in the Ser-Pro-rich loops, a Tyr and an Arg in hUNG immediately following the Leu-loop have been suggested to be important in forming a DNA minor-groove reading head (Parikh *et al.*, 1998). Interestingly, in drUNG this specific region is not conserved and the segment 274-VYRG-277 in hUNG is substituted into 212-EQY-214 in drUNG with one deletion and no sequence similarity. Superpositioning of these residues clearly illustrates the local differences between the crystal structures of hUNG and drUNG (Fig. 6). A similar lack of conservation in this region is seen for ecUNG (193-AHRG-196) compared with hUNG, although for the latter two the residue properties are conserved. A recent study (Chen *et al.*, 2004) has focused on mutational analysis of Arg276 in hUNG into 18 other amino acids (Lys was not introduced). The outcome of their study was that all amino-acid substitutions gave reduced catalytic efficiency, mainly caused by a drastic reduction in binding affinity for all mutants.

Following these results and in light of the changes in the Leu-loop region and the amino-acid substitutions into positively charged residues in drUNG compared with hUNG, there exists the clear possibility that drUNG has a somewhat modified way of recognizing and binding substrate DNA, as the binding affinity of drUNG is threefold stronger than for hUNG.

3.3. General structural considerations

Phylogenetic analysis shows that the *Thermus* and *Deinococcus* lineages are related (White *et al.*, 1999, and references therein). In contrast to *T. thermophilus*, which through a genomic sequencing project (Henne *et al.*, 2004) has been found to possess only a UDG enzyme of type 4, *D. radiodurans* is exceptional in that it has three enzymes possessing UDG activity (Sandigursky *et al.*, 2004). These are dr0689, a eukaryotic-like UDG of type 1, dr0715, a MUG-like enzyme of type 2, and dr1751, an archea-like UDG of type 4. As *D. radiodurans* is a poly-extremophile bacterium adapted to survival under conditions both of increased levels of radiation and prolonged periods or cycles of desiccation/rehydration, it has been speculated that increased levels of cytosine deamination (causing a G–C→A–T transition mutation for half of the DNA strands in the next cycle of replication if left unrepaired) would result from these conditions (Sandigursky *et al.*, 2004); however, this assumption will have to be verified. There is only one UDG enzyme in *T. thermophilus* and it is therefore unlikely that the intrinsic high G–C content [66.6 and 69.4% for *D. radiodurans* (White *et al.*, 1999) and *T. thermophilus* HB27 (Henne *et al.*, 2004), respectively] of these related organisms potentially will cause an increased level of cytosine deamination, justifying the elevated number of UDG enzymes in *D. radiodurans*. This could therefore be a specific adaptation to its ability to withstand radiation and desiccation, although drUNG has been suggested to be responsible for the majority of the UDG activity in *D. radiodurans* crude cell extracts (Sandigursky *et al.*, 2004). Another possibility for the elevated number of UDG enzymes in *D. radiodurans* is that uracil, as already mentioned, can be incorporated into DNA during synthesis by incorporation of dUTP instead of dTTP. There is a possibility that *D. radiodurans* has more error-prone DNA polymerases than other organisms and therefore requires more DNA-repair enzymes for this reason.

4. Conclusions

The three-dimensional structure of uracil-DNA *N*-glycosylase (UNG) from *Deinococcus radiodurans* has been determined to a resolution of 1.8 Å. The structure is overall the same as those of human UNG and *E. coli* UNG; however, the catalytic efficiency is four times higher for *D. radiodurans* UNG compared with human UNG. Substitutions into positively

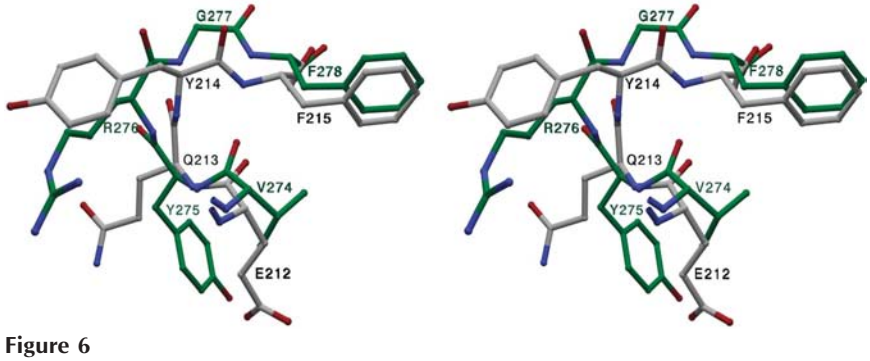

Figure 6
Stereoview showing the superpositioning of non-conserved residues in the α10 region of hUNG (green C atoms) and drUNG (grey C atoms).

charged residues in the DNA-binding region suggest a rational explanation of the much-improved substrate binding.

This work was financed by the National Functional Genomics Programme (FUGE) from The Research Council of Norway and by the European Synchrotron Radiation Facility (ESRF) and was performed in the laboratory and at the beamlines of the Macromolecular Crystallography Group at the ESRF. The members of the Macromolecular Crystallography Group are acknowledged for their help and assistance in this work.

References

Aravind, L. & Koonin, E. V. (2000). *Genome Biol.* **1**, 1–8.
Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J. & Pearl, L. H. (1998). *Cell*, **92**, 117–129.
Battista, J. R. (1997). *Annu. Rev. Microbiol.* **51**, 203–224.
Chen, C. Y., Mosbaugh, D. W. & Bennett, S. E. (2004). *J. Biol. Chem.* **279**, 48177–48188.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cone, R., Bonura, T. & Friedberg, E. C. (1980). *J. Biol. Chem.* **255**, 10354–10358.
Daly, M. J. & Minton, K. W. (1996). *J. Bacteriol.* **178**, 4461–4471.
Duncan, B. K. & Weiss, B. (1982). *J. Bacteriol.* **151**, 750–755.
Friedberg, E. C., Walker, G. C. & Siede, W. (1995). *DNA Repair and Mutagenesis.* Washington, DC: ASM Press.
Guex, N. & Peitsch, M. C. (1997). *Electrophoresis*, **18**, 2714–2723.
Haushalter, K. A., Stukenberg, P. T., Kirschner, M. W. & Verdine, G. L. (1999). *Curr. Biol.* **9**, 174–185.
Henne, A. *et al.* (2004). *Nature Biotechnol.* **22**, 547–553.
Impellizzeri, K. J., Anderson, B. & Burgers, P. M. (1991). *J. Bacteriol.* **173**, 6807–6810.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P. A., Hagen, L., Krokan, H. E. & Slupphaug, G. (2002). *J. Biol. Chem.* **277**, 39926–39936.
Lanes, O., Guddal, P. H., Gjellesvik, D. R. & Willassen, N. P. (2000). *Comput. Biochem. Physiol. B*, **127**, 399–410.
Leiros, I., Moe, E., Lanes, O., Smalas, A. O. & Willassen, N. P. (2003). *Acta Cryst.* D**59**, 1357–1365.
Makarova, K. S., Aravind, L., Wolf, Y. I., Tatusov, R. L., Minton, K. W., Koonin, E. V. & Daly, M. J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 44–79.
Minton, K. W. (1994). *Mol. Microbiol.* **13**, 9–15.
Moe, E., Leiros, I., Riise, E. K., Olufsen, M., Lanes, O., Smalås, A. & Willassen, N. P. (2004). *J. Mol. Biol.* **343**, 1221–1230.
Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W. & Tainer, J. A. (1995). *Cell*, **82**, 701–708.
Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E. & Tainer, J. A. (1995). *Cell*, **80**, 869–878.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins*, **11**, 281–296.
Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L. & Lindahl, T. (2001). *EMBO J.* **20**, 4278–4286.
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.
Pearl, L. H. (2000). *Mutat. Res.* **460**, 165–181.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Sandigursky, M., Sandigursky, S., Sonati, P., Daly, M. J. & Franklin, W. A. (2004). *DNA Rep.* **3**, 163–169.
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.
White, O. *et al.* (1999). *Science*, **286**, 1571–1577.
Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T. & Gilliland, G. L. (1999). *Proteins*, **35**, 13–24.