doi:10.1016/j.jmb.2004.01.065 J. Mol. Biol. (2004) 339, 103–116

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Structural Studies of the Nudix Hydrolase DR1025 From *Deinococcus radiodurans* and its Ligand Complexes

Wasantha Ranatunga[1]†, Emma E. Hill[2]†, Jana L. Mooster[1] Elizabeth L. Holbrook[1,3], Ursula Schulze-Gahmen[1], WenLian Xu[4] Maurice J. Bessman[4], Steven E. Brenner[1,2] and Stephen R. Holbrook[1]*

[1]*Physical Biosciences Division Lawrence Berkeley National Laboratory, 1 Cyclotron Road Berkeley, CA 94720, USA*

[2]*Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720-3102, USA*

[3]*Department of Chemistry University of California Berkeley, CA 94720, USA*

[4]*Department of Biology, Johns Hopkins University, Baltimore MD 21218, USA*

**Corresponding author*

We have determined the crystal structure, at 1.4 Å, of the Nudix hydrolase DR1025 from the extremely radiation resistant bacterium *Deinococcus radiodurans*. The protein forms an intertwined homodimer by exchanging N-terminal segments between chains. We have identified additional conserved elements of the Nudix fold, including the metal-binding motif, a kinked β-strand characterized by a proline two positions upstream of the Nudix consensus sequence, and participation of the N-terminal extension in the formation of the substrate-binding pocket. Crystal structures were also solved of DR1025 crystallized in the presence of magnesium and either a GTP analog or Ap_4A (both at 1.6 Å resolution). In the Ap_4A co-crystal, the electron density indicated that the product of asymmetric hydrolysis, ATP, was bound to the enzyme. The GTP analog bound structure showed that GTP was bound almost identically as ATP. Neither nucleoside triphosphate was further cleaved.

Keywords: Nudix hydrolase; MutT-like; *Deinococcus radiodurans*; X-ray crystallography

Introduction

Enzymes belonging to the Nudix family hydrolyze a variety of substrates containing a nucleoside diphosphate linked to some moiety, *x* (Nudix).[1] This enzyme family is characterized by a conserved sequence motif called the Nudix box, $GX_5EX_7REUXEEXGU$,[1] where U is usually a hydrophobic residue, Ile, Leu, or Val, and X is any amino acid. It has been proposed that these proteins act as surveillance and housecleaning enzymes because their substrates are potentially toxic substances, signaling molecules, or metabolic intermediates whose concentrations require modulation during the cell cycle.[1]

Several important physiological functions have been identified with these enzymes: MutT, which hydrolyzes a mutagenic form of deoxyguanosine triphosphate to the monophosphate, preventing its incorporation into DNA;[2,3] hDcp2, involved in messenger RNA decapping;[4] human ADP-ribose hydrolase, LTRPC2, a calcium-permeable channel modulator;[5] and YgdP and its orthologs, an Ap_nA hydrolase, involved in cellular invasion.[6–9] These enzymes are widely distributed and have been found in all three kingdoms. A recent Pfam[10] search identified over 1200 members of this family from more than 250 species of organisms ranging from viruses to humans, only about 70 of which have been characterized. Nudix family proteins have low overall sequence identity: Pfam shows the average sequence identity between pairs as 17%, and we found the average pairwise sequence identity between the 24 *Deinococcus radiodurans* Nudix proteins is 22%. Each of the enzymes hydrolyzes one or more nucleoside

† W.R. and E.E.H. contributed equally to this work.

Abbreviations used: HMM, hidden Markov model; GMPPNP, GNP guanosine 5′-[β,γ-imido] triphosphate; Ap_4A, P^1,P^4-di(adenosine-5′) tetraphosphate; 1025ATP, DR1025-ATP complex; 1025GNP, DR1025-GNP complex; PDB ID, Protein Data Bank identity code.

E-mail address of the corresponding author: srholbrook@lbl.gov

diphosphate derivatives. These include different ribo- and deoxyribonucleoside triphosphates, nucleoside diphosphate sugar molecules, dinucleoside polyphosphates, and several coenzymes including NADH, FAD, and CoA. It is worth noting that in addition to these canonical substrates, some Nudix hydrolases can hydrolyze non-nucleoside diphosphate derivatives such as diphosphoinositol[11] and 5-phosphoribosyl-1-pyrophosphate,[12] most likely due to the similarities of the pyrophosphate linkage in these compounds.

Structural studies by crystallography and NMR have determined the tertiary structure of eight Nudix proteins (see Table 1). These have revealed a conserved core Nudix fold with highly variable peripheral loops and extensions around a central β-grasp motif.[13] All the structures solved so far also contain an α–β–α sandwich, where the conserved Nudix box is located in the helix of the β-grasp motif. The Nudix fold is the central site for the binding of substrates and metal ions, whereas the substrate specificity of each enzyme is determined by amino acid differences, variation in the loops, C-terminal and N-terminal sequences, and dimerization state.

Some Nudix structures are monomers and others are homodimers (Table 1). At this time, the functional role of the oligomeric state is unclear. For example, Ap_4A hydrolase is monomeric,[14] whereas ADPRase is dimeric in bacteria[15,16] and monomeric in humans[17]. Another dimeric Nudix hydrolase from the hyperthermophilic archaeon *Pyrobaculum aerophilum* has been postulated to assume this form for thermostability.[18]

The extremely radiation resistant bacterium *D. radiodurans*[19] encodes 21 Nudix proteins containing a fully conserved Nudix box, and three others with a partially conserved Nudix motif;[10,20,21] this is one less than in *Bacillus cereus*,[22] and six less than in *Bacillus anthracis*[23] which has 30, the most of any bacterial genome sequenced to date. As a comparison, *Escherichia coli* with a much larger genome than *D. radiodurans* has only 11 Nudix proteins.[24] All 21 of the fully conserved *D. radiodurans* Nudix proteins have been cloned and expressed in soluble form in *E. coli* and subjected to a preliminary screen in order to disclose their enzymatic activities.[25] DR1025 was found to be a Mg^{2+} activated nucleoside triphosphatase and a dinucleoside polyphosphate pyrophosphatase.[25] However, DR1025 did not complement an *E. coli* MutT strain. This enzyme is of special interest, because it is the first Nudix enzyme catalyzing the hydrolysis of these two very different classes of substrate, and it also shares the unique cleavage pattern of some other Nudix hydrolases from *D. radiodurans* by liberating P_i from both nucleoside triphosphates and diphosphates.[26] Other Nudix nucleoside triphosphatases including MutT, *E. coli* Orf17 and Orf135, and the riboTTPase of *Agrobacterium tumefaciens* all liberate PP_i from nucleoside triphosphates and are inactive on nucleoside diphosphates.[3,27–29]

The large number of *D. radiodurans* Nudix proteins suggests that they may play a role in maintaining viability in the face of high levels of radiation: they likely hydrolyze damaged NTPs and other toxic compounds such as ADP-ribose,

Table 1. Structural studies of proteins containing the Nudix fold (data from SCOP and the Protein Data Bank[13,34])

Protein	EC	PDB ID and reference	Structure (resolution Å)	Species	Oligomeric state
Nucleoside triphosphate pyrophosphohydrolase-MutT	3.6.1.-	1mut[47]	NMR	*E. coli*	Monomer
		1tum[48]	NMR	*E. coli*	Monomer
		1ppx, 1pun, 1puq, 1pus[49]	NMR	*E. coli*	Monomer
ADP-ribose pyrophosphatase	3.6.1.13	1g0s, 1g9q, 1ga7[15]	Cryst. (1.9)	*E. coli*	Dimer
		1khz[50]	Cryst. (2.0)	*E. coli*	Dimer
		1mk1, 1mp2, 1mqe, 1mqw, 1mr2[16]	Cryst. (2.0)	*M. tuberculosis*	Dimer
		1q33, 1qvj[17]	Cryst. (1.8)	Human	Monomer
Diadenosine tetraphosphate hydrolase	3.6.1.17	1f3y[51]	NMR	*L. angustifolius*	Monomer
		1jkn[14]	NMR	*L. angustifolius*	Monomer
		1ktg, 1kt9[52]	Cryst. (1.8)	*C. elegans*	Monomer
CoA pyrophosphatase	3.6.1.9	1nqz, 1nqy[30]	Cryst. (1.7)	*D. radiodurans*	Monomer
Isopentyl diphosphate isomerase (IPP)	5.3.3.2	1hzt, 1hx3[53]	Cryst. (1.5)	*E. coli*	Monomer
		1i9a[54]	Cryst. (2.5)	*E. coli*	Monomer
		1q54 (replaced 1n2u)[55]	Cryst. (1.9)	*E. coli*	Monomer
Function uncertain					
PAE3301	–	1k2e, 1jrk, 1k26[18]	Cryst. (1.8)	*P. aerophiluym*	Dimer
DR0079	–	1q27[56]	NMR	*D. radiodurans*	Monomer
DR1025	–	1sjy, 1sol, 1su2, 1sz3	Cryst. (1.4)	*D. radiodurans*	Dimer

EC refers to the enzyme classification number.[57]

Table 2. Data collection statistics for DR1025

	Sm derivative	Native 1	Native 2	ATP	GNP
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_1$	$P4_1$
a	53.17	53.39	53.21	53.23	53.07
b	53.17	53.39	53.21	53.23	53.07
c	121.97	121.59	122.52	122.36	122.20
$\alpha = \beta = \gamma$ (deg.)	90	90	90	90	90
Resolution (Å)	100–1.8	100–1.6	100–1.39	100–1.6	100–1.6
Total observations	453,884	558,558	879,484	327,440	217,732
Unique reflections	16,652[a]	24,286	35,334	44,109	43,122
Completeness (%)	98.2 (98.3)	100. (99.8)	98.4 (89.5)	99.9 (99.9)	99.8(97.9)
$I/\sigma(I)$	43.2 (18.9)	50.6 (9.0)	56.1 (4.4)	45.7 (8.7)	47.4 (3.8)
R_{sym} (%)[b]	4.8 (9.6)	7.5 (28.1)	5.8 (34.4)	4.3 (17.0)	4.4 (33.2)
X-ray wavelength (Å)	1.20	1.00	1.00	1.00	1.00

Values in parenthesis refer to the highest resolution shell (i.e. 1.39–1.44 Å for DR1025 Native 2; 1.80–1.84 Å for Sm derivative).
[a] Anomalous reflections not merged.
[b] $R_{sym} = \sum|(I_{hkl} - \langle I_{hkl}\rangle)|/\sum I_{hkl}$, where, $\langle I_{hkl}\rangle$ is the average of I_{hkl} over all symmetry equivalents.

and they may modulate the accumulation of metabolic intermediates during the repair of its damaged genome.[25] We are conducting structural studies of the *D. radiodurans* Nudix proteins to help identify their cellular functions, to relate their structural diversity to functional diversity and to place them in the context of the broader Nudix family. Here, we have chosen DR1025 for study because of its unique properties described above, and we report its crystal structure alone and in complexes with its Ap_4A substrate and a GTP substrate analogue, GMPPNP (GNP).

Results

Models of the native DR1025 structure (Sm derivative and Native 2), as well as the DR1025 ligand complexes (with a GTP analog and ATP) were determined and refined as described in Materials and Methods and summarized in Tables 2 and 3.

Structure of the DR1025 apoenzyme

The final model of the DR1025 apoenzyme (Native 2) consists of all 159 residues except for the N-terminal methionine and the surface loop residues 35–39 that connect β strands 2 and 3 (Figure 1). The statistics of the final refinement indicate that more than 93% of the residues fall into the most favored regions in the Ramachandran plot. There are no metal ions specified in the final model, although some water molecules may correspond to low occupancy metals.

One chain of the DR1025 dimer (see Figure 1) contains a β-sheet composed of six highly twisted β strands (labeled β1–β6) that separates two α helices at the C-terminal (α2, α3) from α1, which is part of the Nudix box. The secondary structure and solvent accessibility are shown in Figure 2. Strand β1 (13–22) is parallel with β5 (90–103) and antiparallel to β3b (48–52) while β2 (26–32) is antiparallel to both β3a (44–46) and β6 (117–122). The strand β4 (73–87) is antiparallel to β5 and together with β6 forms the outer edges of the β sheet. A large angle between the strands β1, β4, and β5 and the strands β2 and β6 is the result of a kink in β3 around Pro48 dividing this strand into β3a and β3b (see Figure 2). This proline is semi-conserved among aligned sequences and structures of Nudix proteins and is only two residues upstream from the conserved Nudix box.

The secondary structural elements of DR1025 are linked by both short (L1, L3, L4, L6) and long (L2, L5) loop regions. The tightest loops, L1 and L4 are β-hairpin turns; they consist of two charged amino acid residues (E R and E D) followed by a glycine that forms a hydrogen bond within the β-strand. L1 (Glu23–Gly25) connects β1 to β2. L4 (Pro88–Gly90) connects β4 and β5 (Pro88 may be instrumental in creating this turn). L6 (Gln135–Arg137) links α2 to α3. This loop is anchored to

Table 3. Statistics of atomic refinement for the DR1025 datasets

	Sm derivative	Native 2	ATP	GNP
Resolution range for refinement (Å)	20–1.8	20–1.4	50–1.6	50–1.6
No. of protein atoms	1229	1219	2517	2454
No. of water molecules identified	103	102	296	225
No. of metal ions identified	3	0	6	2
Average *B*-factor (Å^2)	18.4	17.1	17.75	20.31
RMSD bond length (Å)	0.004	0.004	0.111	0.014
RMSD bond angle (°)	1.28	1.28	2.000	1.762
R_{free} (%)	25.06	24.21	22.49	23.43
R_{work} (%)	21.82	22.82	20.44	21.30

Refinement of Native 1 was not completed and therefore is not included. All refinement was done with the CNS package.[39]

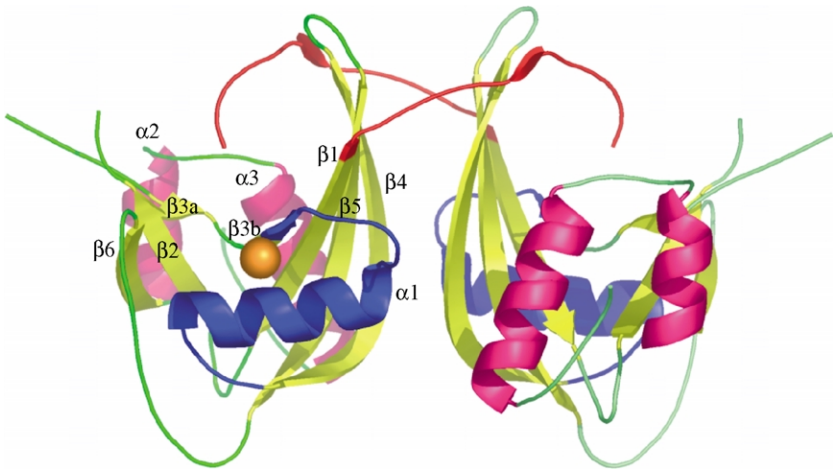

Figure 1. The structure of the DR1025 dimer as observed in the crystal. The Nudix box is colored in blue. Secondary structure elements are labeled. The β strands are colored yellow. The other α helices are in pink and the samarium bound to the Nudix helix (Sm1) is in orange. The intertwining N-terminal extension is in red. The proteins sit on a crystallographic 2-fold axis that relates the monomer to the dimer. Only one samarium of one monomer (Sm1) is shown.

the β-sheet through two hydrogen bonds (Gln135 O–Trp45 N and Arg137 N–Trp45 O) and a salt bridge between Arg137 and Glu32. The short loop, L3 (Glu53–Glu56), contains part of the Nudix consensus sequence and is therefore a good candidate for being functionally important. Structurally this loop is well conserved among all Nudix structures joining β3b to α1 (our unpublished data). The Nudix loop is fixed in place by a main-chain–main-chain and a main-chain–side-chain hydrogen bond between Glu53 and Glu56. There is also a stabilizing salt bridge formed between OE1 of Glu56 and NH2 of Arg64.

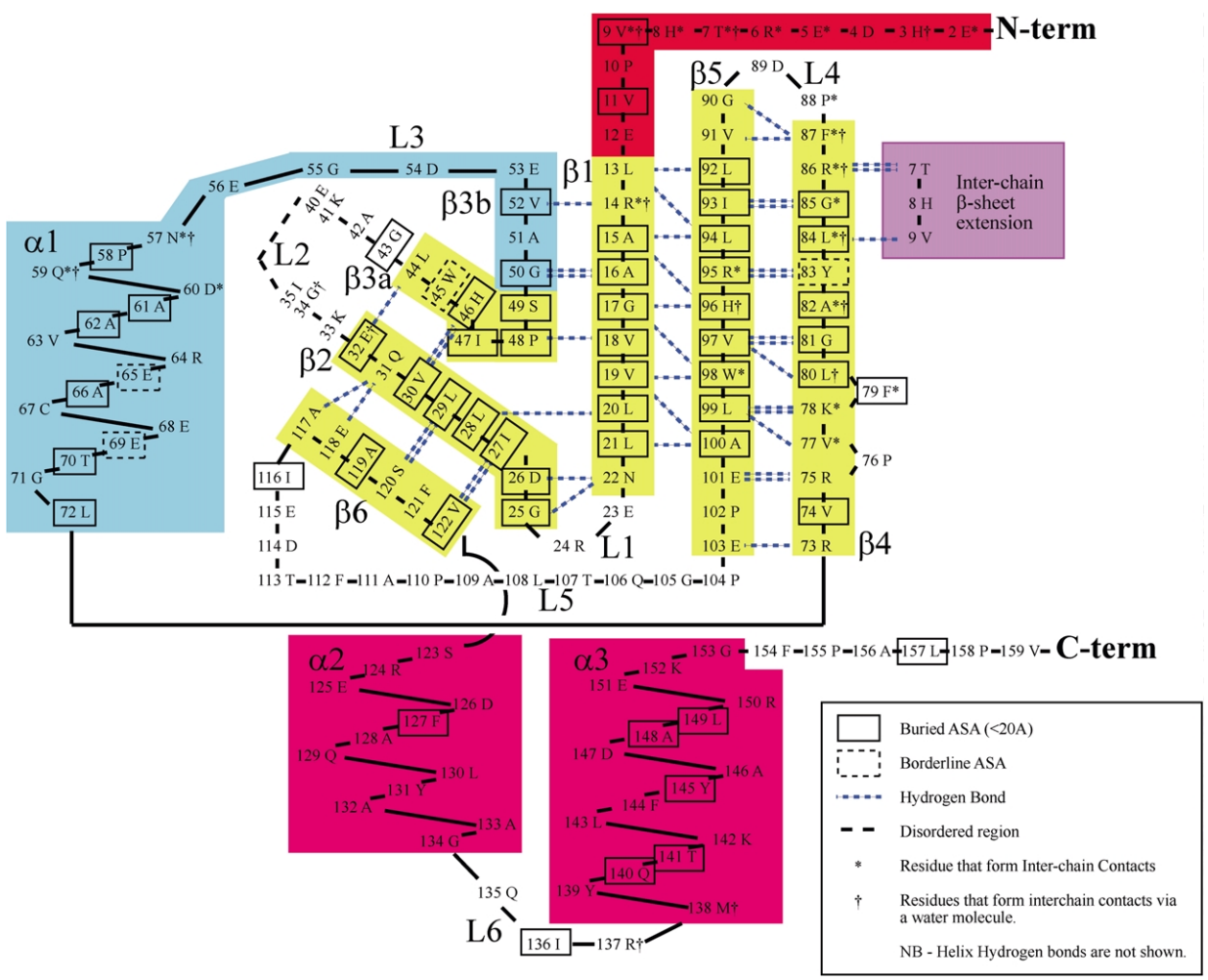

Figure 2. The secondary structure arrangement and surface accessibility of residues in the DR1025 structure. β-Strands (yellow), α-helices (pink) and loops (no color) are labeled sequentially. The Nudix box is in light blue and the residues from the other chain forming an inter-subunit β structure with strand 4 are indicated in purple.

The disordered region of the structure, residues 36–39, lies within loop L2 (Lys33–Gly43) between strands β2 and β3a. This entire loop is highly mobile in the crystal as judged by the thermal parameters. We note that the recently solved Nudix structures of coenzyme A pyrophosphatase (PDBIDs 1nqy and 1nqz) from *D. radiodurans*,[30] also have a disordered region in the equivalent loop. This region extends away from the protein and is therefore a candidate for interaction with another biomolecule that would serve to order its structure. We also note that this region is ordered in our crystal-complex with ATP (see below). The longest loop is L5 (13 amino acid residues, Pro104–Ile116), all residues of which are solvent accessible except for Ile116. L5 is anchored by two hydrogen bonds to the C-terminal end of α1 (Ala111 N–Glu69 O and Ala109 N–Thr70 O) and also contains one internal hydrogen bond (Thr113 N–Ile116 O).

Dimer formation

Based on gel-filtration and dynamic light scattering experiments, the DR1025 protein exists as a dimer in solution.[31] A homodimer is formed in the crystal (as shown in Figure 1) from two chains related by a crystallographic 2-fold axis. The refined crystal structure shows extensive intermolecular contacts and 3113 Å^2 (17% of total) buried surface area in the association of the dimer. The dimer of DR1025 forms mostly by swapping of the amino-terminal extension (residues 1–12), interaction between α1 and β1, β4, β5, and loop L3, and water-bridged interactions between β2 and loop L6. Residues Thr7 and Val9 of the N terminus extend the β-sheet through contacts with Leu84 and Arg86 of β4 of the other chain (Figure 2).

Of the 20 residues that form inter-chain contacts (Table 4), 17 are located in either the N-terminal intertwined region and the beginning of β1, or in the two dimer-facing edge strands β4 and β5 or their linker L3 (see Figure 2). Interestingly, the other three inter-chain contacting residues are located within the Nudix box, although none of these are the highly conserved residues of the Nudix motif. These three residues are all within α1 but are located on the opposite side of the helix from the Nudix substrate-binding cleft.

Sixteen residues form inter-chain contacts bridged by a water molecule (Table 5). Nine of these are in positions that also make direct inter-chain contacts (see Figure 2). Of the remaining seven, four are located in the N-terminal intertwined region, β4, and β5 close to the dimer interface. The other three are dispersed through the structure (this is expected given the larger distance possibility for contacts when a water molecule is bridging them), located within β2 and L6.

Nudix box

Residues 50–72 of DR1025 contain the Nudix signature sequence $G^{50}X_5E^{56}X_7R^{64}E^{65}UXE^{68}E^{69}XG^{71}U$, where the superscript numbers are the residue numbers in DR1025. This sequence belongs to part of β3b, L3 and the α1 helix. This topology is similar to that found in all other Nudix structures (Table 1). As discussed above, the dimer contacts of DR1025 are different from other dimers in that they include part of the Nudix box. Specifically, Asn57 and Gln59 interact directly with Lys78 and Phe79 and indirectly through water with Val78 and Leu80 (Tables 4 and 5), while Asp60 also makes some contacts with Lys78. There are also many interactions with water molecules around the Nudix box, which could be displaced by substrate binding (not shown).

Table 4. Contacts between dimer subunits

Source atoms		Target atoms		Distance
Residues	Atoms	Residues	Atoms	
Glu2A	OE2	Arg95B	NH1	2.88[a]
Glu5A	O	Phe87B	CD1	3.49
	O	Pro88B	CD	3.38
Arg6A	CA	Arg86B	O	3.21
	CB		O	3.24
	C		O	3.46
	NH2	Phe87B	CE1	3.43
Thr7A	O	Gly85B	C	3.44
	O		CA	3.14
	N	Arg86B	O	2.79[a]
	OG1		O	3.34[b]
	O		N	2.81[a]
His8A	ND1	Leu84B	O	3.48[b]
Val9A	N	Leu84B	O	3.00[a]
Arg14A	NH2	Ala82B	O	2.84[a]
Asn57A	OD1	Lys78B	CE	3.46
	ND2	Phe79B	O	2.78[a]
Gln59A	OE1	Phe79B	N	3.01[a]
Asp60A	CG	Lys78B	NZ	3.40
	OD1		CE	3.37
	OD1		NZ	3.21[a]
	OD2		NZ	2.86[a]
Phe79A	CD2	Trp98B	CZ2	3.45

Inter-chain contacts from chain A to chain B of the DR1025 dimer–N.B. The converse contacts (same positions B to A) also exist because of symmetry in the dimer but are not shown here.
[a] Indicates the strong possibility of a hydrogen bond (<3.3 Å).
[b] Indicates a weaker possibility of a hydrogen bond (>3.3 Å).

Water and metal ions

The apoenzyme structure (Native 2, 1.4 Å resolution) contains 102 unique water molecules (204 per homodimer). We did not identify any specific metal ions in this model. In the Sm derivative (1.8 Å resolution), 103 water molecules and three unique Sm ions were located. Some water molecules were absent or moved position in the Sm derivative structure compared to the apoenzyme as outlined below. Additional water molecules in the apoenzyme structure replace the Sm ions in the derivative structure.

Table 5. Water molecules acting as a bridge between the two chains of the DR1025 apoenzyme dimer

	Chain A H-bonds			Chain B H-bonds		
Water number	Residues	Atom	Distance (Å)	Residues	Atom	Distance (Å)
711	Glu32	OE1	2.84[a]	His3	ND1	2.95[a]
	Gly34	O	2.80[a]			
	Arg137	NH2	3.77[b]			
716	Arg14	NE	2.82[a]	Ala82	N	3.01[a]
	His96	NE2	3.51[b]		O	3.20[a]
757	Thr7	OG1	2.74[a]	Arg86	O	2.84[a]
				Phe87	O	3.22[a]
730	Leu84	N	2.74[a]	Val9	O	2.94[a]
		O	3.51[b]			
744	Gln59	NE2	2.83[a]	Val77	O	2.73[a]
783	Leu80	O	2.81[a]	Asn57	ND2	2.93[a]

Again we only show one set of the two symmetry related contacts; there is another set with the chains reversed.
[a] Indicates the strong possibility of a hydrogen bond (<3.3 Å).
[b] Indicates a weaker possibility of a hydrogen bond (>3.3 Å).

One Sm (Sm1, Figure 1) is coordinated by six atoms, including Glu65 and Glu68 of the Nudix box (three water molecules, Glu68 OE1, and Glu65 OE1 and OE2). In the apoenzyme, two water molecules bridging the glutamic acid residues replace this site. Arg64, which was shifted away in the samarium derivative, now forms H-bonds to these water molecules. The second Sm (Sm2) is located near Asp147 and Arg150 and has three water molecules coordinated with it. In the apoenzyme, a water molecule replaces Sm2. The third Sm (Sm3) ion is coordinated with Pro10 and Gly90 carbonyl oxygen atoms, the Glu12 carboxyl side-chain, and two water molecules. This Sm3 ion has only partial occupancy with a high *B* value and is replaced by a water molecule in the apoenzyme structure.

Mobility and flexibility

The DR1025 apoenzyme is generally well ordered as judged by the high-resolution diffraction and an average *B*-factor of 17.1 Å^2. However, as discussed above, part of loop 2 (residues 36–39, connecting β2 and β3) is not visible in the native electron density maps, reflecting local disorder, while the *B*-factors for the rest of the loop range from 20 Å^2 to 50 Å^2. In the Sm derivative density maps, we do not observe residues 36 and 37, but there is weak density for residues 35, 38 and 39. In both the apoenzyme and Sm derivative structures, the N-terminal residue is disordered and not visible in the maps. The highest *B*-factors for the other residues (~40 Å^2) are located at the N terminus (residues 2–5), C terminus (residues 158–159), and Glu40 adjacent to the disordered loop residues.

In addition, Arg14 and Arg124 have alternate side-chain conformations in the apoenyzme, whereas Arg14 has only one conformation in the samarium derivative structure. The two conformations of Arg14 are stabilized mainly by hydrogen bonding of two water molecules, Wat716 and Wat764, that are absent in the Sm derivative, and His96. In Arg124, one conformation has higher occupancy in the apoenzyme structure than that of the Sm derivative structure. Arg124 is coordinated with the carbonyl group of Glu24, Wat720, Wat724 and the carboxylic acid group of Glu125.

Electrostatic surface

We calculated the electrostatic potential of the DR1025 dimer using the Poisson–Boltzmann distribution given by the GRASP program.[32] Figure 3 shows the charge distribution for the surface of DR1025 in approximately the same orientation as in Figure 1. One side of each monomer chain is

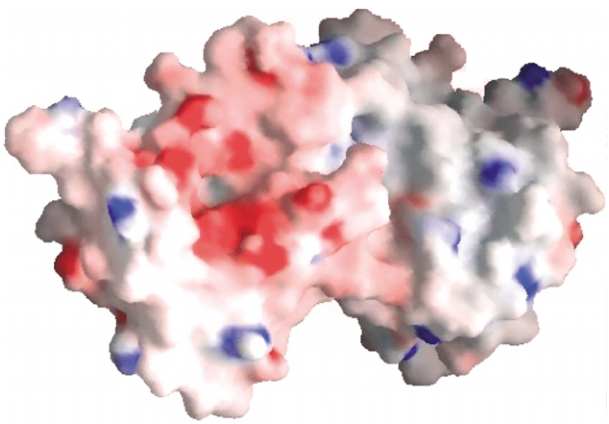

Figure 3. Electrostatic potential surface of the DR1025 dimer as calculated by the GRASP program.[32] The areas of high negative charge are shaded red, while the positively charged regions are shaded blue.

mainly negatively charged (red), while the opposite side is mainly positively charged (blue). The substrate-binding region (see below), including the divalent ion (Sm1)-binding site and the Nudix box, is highly negatively charged. This distribution is unexpected, since Nudix proteins bind polyphosphates as substrates, but may reflect binding sites for metal ligands.

Structures of DR1025 substrate complexes

Since preliminary assays[25] indicated that DR1025 hydrolyzes dGTP or Ap_4A,[25] we co-crystallized DR1025 with a GTP analog (guanosine 5′[β,γ-imido triphosphate]; GNP, GMPPNP) and with Ap_4A (see Materials and Methods) each in the presence of magnesium ions. The resulting crystals diffracted to 1.6 Å resolution and their structures were determined by molecular replacement (MR) and refined as described in Materials and Methods and summarized in Tables 2 and 3. The resulting electron density maps clearly revealed bound ligands and magnesium ions. The space group changes from $P4_12_12$ for the apoenzyme and Sm derivative to $P4_1$ for the ligand complex structures. The dimer that formed around the 2-fold crystallographic axis in the apoenzyme is preserved in the ligand complexes that have two monomers per asymmetric unit related by 2-fold non-crystallographic symmetry. In the complex structures, loop L2 becomes partially (1025GNP residues 36–38 remain disordered) or fully (1025ATP) ordered apparently due to crystallographic packing interactions between L2 (Glu40 OE2) and a symmetry related magnesium hydrate (coordinated water of Mg 201). Otherwise, only small, local changes are observed between the native monomer and dimer and the ligand-bound monomers and dimers. Superposition of all atoms for the following pairs of dimer structures gave RMSD values of: apoenzyme and 1025GNP: 0.74 Å, apoenzyme and 1025ATP: 0.84 Å, 1025ATP and 1025GNP: 0.84 Å, 1025ATP chain A and 1025ATP chain B: 0.91 Å, and 1025GNP chain A and 1025GNP chain B: 0.84 Å, indicating that no significant overall conformational change occurs on ligand binding either

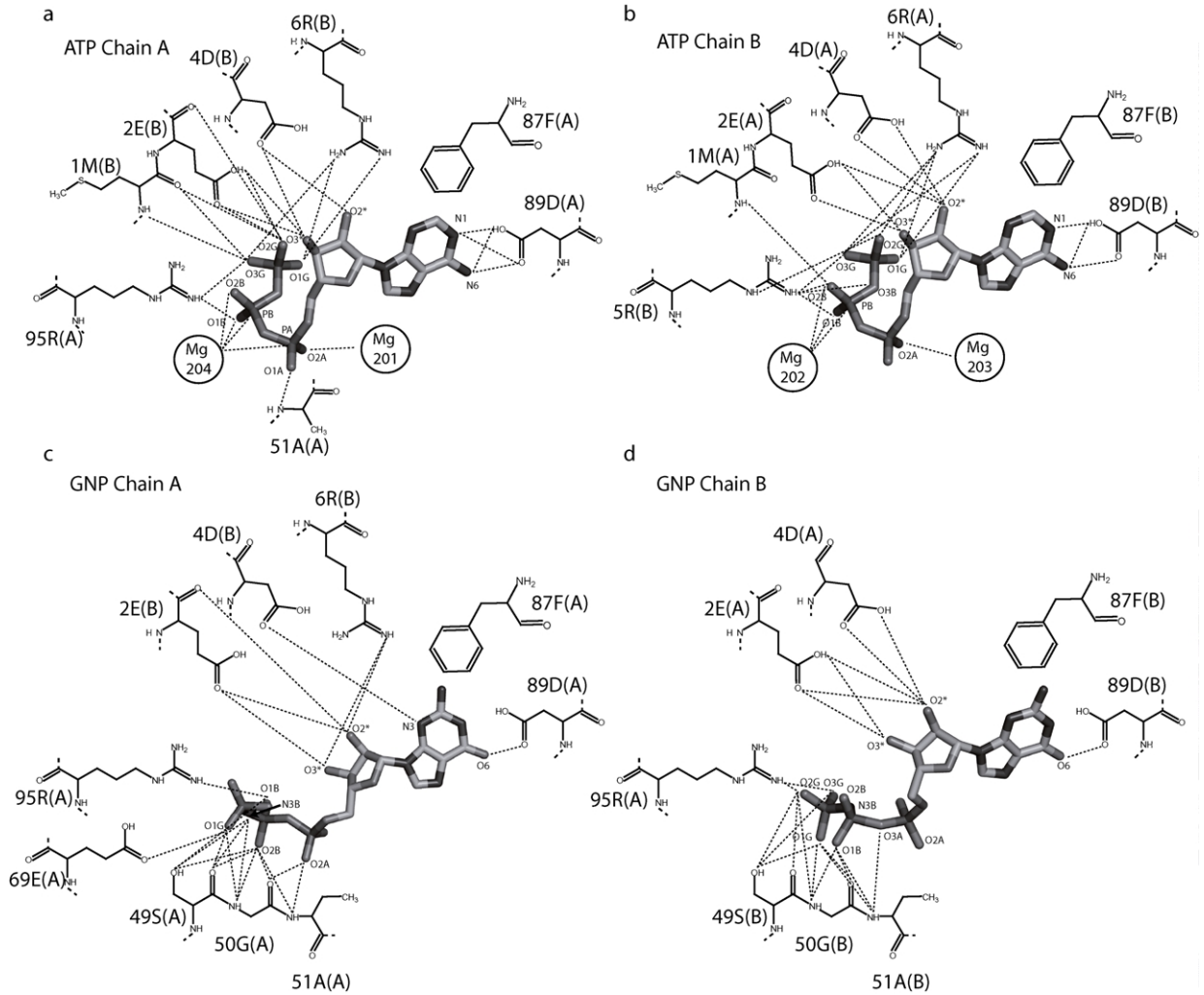

Figure 4. A schema showing the interactions of ATP and GNP for both chains of their respective structures. Dotted lines connect oxygen and nitrogen atoms within 3.5 Å, implying hydrogen bonds and salt bridges. F87 is stacked on the nucleotide bases.

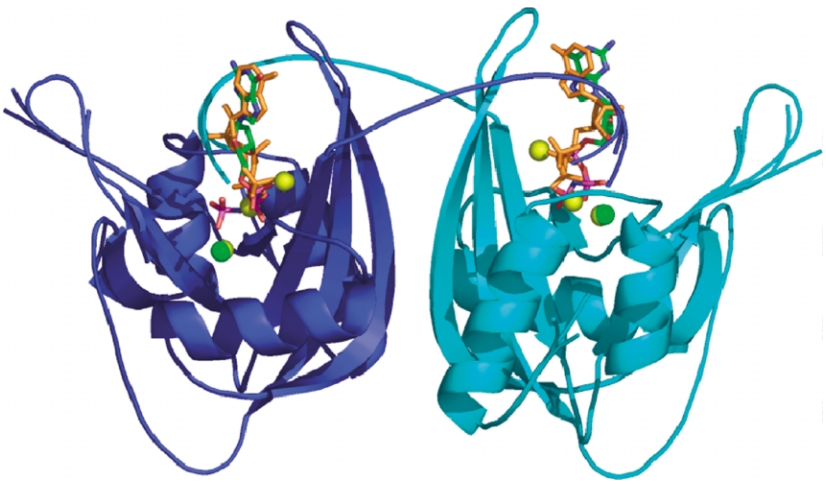

Figure 5. Superposition of the apoenzyme and ligand bound structures including substrates and identified magnesium ions. ATP is colored orange, GNP is colored by atom type, the magnesium ions from the ATP bound structure are yellow and the magnesium ion in the GNP structure is green. The main-chain superposition is indistinguishable for most of the protein; however, differences can be see for the apoenzyme disordered loop L2 and the more ordered ligand structures.

within the monomer or in the orientation of the monomers forming the dimer. Figure 4 gives a schema showing the detailed interactions of ATP and GNP for both chains of their respective structures. Figure 5 shows a superposition of the apoenzyme, GNP, and ATP bound structures and ligands.

Co-crystallization with Ap_4A (P^1,P^4-di(adenosine-5′)tetraphosphate)

Diffraction data collected from crystals of DR1025 grown in the presence of Ap_4A and Mg^{2+} gave rise to electron density maps clearly showing a bound ATP and three bound magnesium ions per monomer. The adenine base stacks directly above Phe87 in L4 and is open on its other face. Asp89 orients the base by hydrogen bonding to the N6 amino group (Figure 4). Glu2 and Asp4 from the other chain of the dimer interact with the ribose hydroxyls. The largest differences (>1.5 Å) between the apoenzyme and the ATP bound protein (chain A/B) are at the N terminus—Glu2 (3.4 Å/3.2 Å); the poorly ordered L2—Glu40 (3.5 Å/5.6 Å) and Lys41 (3.4 Å/4.7 Å); Arg75 (1.8 Å/2.3 Å) and the C-terminal Val159 (5.4 Å/5.9 Å). The relatively equal occupancy and *B*-factors of the three phosphate groups indicate that the ATP is not further cleaved by the enzyme to AMP or ADP.

Co-crystallization with GMPPNP (guanosine 5′[β,γ-imido] triphosphate)

Crystals of DR1025 grown in the presence of GMPPNP (GNP) and Mg^{2+} also provided electron density maps clearly showing the GTP analog bound in the active site of the protein, in this case with a single magnesium ion observed. The GTP ligand is located in almost an identical position as the ATP ligand described above (Figures 4 and 5) and the three phosphate groups are of roughly equivalent electron density as reflected in occupancy and *B*-factor. The largest differences between the native and GNP bound protein (chain A/B >1.5 Å) are at the N terminus—Glu2 (3.2 Å/2.9 Å) and the poorly ordered L2—Glu40 (1.6 Å/6.7 Å) and Lys41 (4.2 Å/3.8 Å), while the C-terminal Val159 is not ordered.

Comparison to other Nudix structures

Enzyme structure

Comparison of the structures of functionally

```
DR1025     PSGAVEDGENPQDAAVREACEETGLR
1g0s       VAGMIEEGESVEDVARREAIEEAGLI
1ktg       PKGHVDPGEDEWQAAIRETKEEANIT
1f3y       PQGGIDEGEDPRNAAIRELREETGVT
1k2e       PGGHVEHNETPIEAVKREFEEETGIV
1hzt       VCGHPQLGESNEDAVIRRCRYELGVE
1mut       PGGKIEMGETPEQAVVRELQEEVGIT
DR0092     PGGGSEGAEAIEETTPRELREETELT
DR0784     VAGIAELGEPLEQTLRREVQEELGLT
DR0783     PGGIAELGEALEDTLRRELQEETGLR
DR0261     PGGRVESGETLQDAARREVREEIGVE
DR0603     PGGGIEPGETPEQAAVREAWEEVGAR
DR0274     PGGGVEMGESHAEALARELREELGAE
DR1168     LAGFVEPSETLEAAVHREVGEEVGVK
DR2356     PGGGIEEGETPQGACAREVLEEVNLT
DR0876     PGGKAQPSETLADAAVRQVLVETGLH
DR0550     PKGGLEAGELIQDGARRETFEETGLV
DR0975     PGGVVEHGETLPEALYREIYEETGLK
DR0192     PGGFVQPGEELHEAALRELRTETSVS
DR1184     PGGSLDAGETPTQAALREAQEEVALD
DR0079     VGGAVQSGETYEEAFREAREELNVE
DR2428     AGVEAHTGGSRDVALVGHIKDESSYY
DR0149     VAGGLEAGETPEQAALREAGEETGQD
DR1776     PGGALSTDEDAATCARREWEEETGLL
DR2272     PGGAVQTGESSAAAAQREWHEETGLR
DR1007     PAGLIDEGETPEAAARRELQEEVGLD
DR2204     VAGGVEKGEDLGAAAARELLEEVGGA
DR0004_1   LGGGLEPGEDFLIAAHRELLEETGLR
DR0004_2   PGGSLEPGESFAECAARELHEETGLR
DR0329_1   PGGGLDPGEAFLAAAQRELKEETGLE
DR0329_2   PGGKLEPGESFEECAARELLEETGLR
```

Figure 6. The Nudix consensus sequence from solved Nudix structures and 23 of the Nudix proteins of *D. radiodurans*. We excluded only one more divergent Nudix from *D. radiodurans* (DR1211). Typical consensus residues are highlighted in yellow. Additional positions of conservation are shown in cyan. In each case, either one representative PDB identifier is given or the DR gene number.

and evolutionarily diverse Nudix proteins shows overall conservation of the Nudix fold, with specific differences in the loops and termini. The most variable parts of the Nudix structures are the C-terminal helices, which have different lengths and absolute position. Some Nudix structures have additional α-helices (diadenosine tetraphosphate hydrolase) or β-strands (ADP-ribose tetraphosphate hydrolase and isopentyl diphosphate isomerase). We noticed in β3a the strong conservation of a proline at a position two amino acid residues upstream of the first Gly of the Nudix sequence box. This Pro is present in DR1025, MutT, diadenosine tetraphosphate hydrolase (*Lupinus angustifolius* and *Caenorhabditis elegans*), CoA pyrophosphatase, and a Nudix protein of unknown function (PDB IDs: 1k2e, 1jrk and 1k26), as well as many of the other *D. radiodurans* Nudix proteins (Figure 6). This Pro is likely to be instrumental in permitting the extreme bend in β3 resulting in a kinked β-sheet.

Metal binding

Another common feature we observe is the mode of divalent metal ion binding. A high occupancy divalent ion frequently occupies a common location and is coordinated by analogous residues of the Nudix consensus sequence and other structurally conserved amino acid residues (Table 6). To some extent, it is the high conservation of the residues responsible for divalent cation binding that is reflected in the Nudix signature sequence despite the fact that the main substrate does not bind here. The metal ion tends to be positioned directly above the highly conserved first Gly of the Nudix box. We note also that frequently another Gly precedes this Gly; allowing space for the metal ion (Figure 6). One other well-conserved position within the Nudix box, not typically considered part of the Nudix consensus sequence, is a Gly directly preceding the first Glu. This is probably necessary to allow the bend from L3 into α1.

Substrate-binding site

We have compared the structures of the DR1025 GNP and ATP complexes determined here with the ligand complexes of MutT and the Ap_4A hydrolases from *L. angustifolius* and *C. elegans*. Superpositions of the conserved Nudix fold were made between PDBID: 1tum (MutT-Mg(2 +)-Ampcpp-Mg(2 +) complex) and DR1025–GNP; and PDB ID: 1jkn (*L. angustifolius* Ap_4A hydrolase–ATP-Mgfx); and PDB ID: 1ktg (*C. elegans* Ap_4A hydrolase–AMP-Mg) with DR1025–ATP as shown in Figure 7. From these superpositions, it is clear that none of the known nucleoside triphosphate ligands bind in the same location or to the corresponding residues as observed for the DR1025 ligands, suggesting that this is a new class of Nudix hydrolase. The DR1025 substrates bind in a largely pre-formed pocket consisting of residues from the N terminus of one chain and several charged residues from the other. Although the overall conformational change of the enzyme binding site is small upon ligand binding, some

Table 6. Conserved glutamic acid residues involved in coordinating metal ion binding in various Nudix structures

			Metal ion coordinating amino acid residues nudix consensus sequence													
Nudix structure	PDB ID	Down stream other	G	x_5	E	x_7	R	E	U	X	E	E	x	G	U	Upstream other
DR1025 Sm	1sol							65E			68E					
DR1025 ATP	1su2	49S						65E								
DR1025 GNP	1sz3							65E								
MutT	1tum		38G					53E			56E	57E				98E
	1ppx		38G					53E			56E	57E				
CoA pyrophosphatase	1nqz							86E								
ADP-ribose pyrophosphatase	1ga7							112E			115E					164E
	1khz	96A						112E				116E				164E
	1mqe							93E								
	1mqw	76A						93E				97E				142E
	1mr2	76A														
	1qvj	214G						230E								
Diadenosine tetraphosphate hydrolase	1ktg	36K						52E				56E				103E 111E
Isopentyl diphosphate isomerase	1q54	67A										87E				

As the MutT structures were solved by NMR these contacts are not present in every model but each is at least present in one of the models. Note in some other cases contacts depend on the chain and or on the specific metal ion when more than one metal ion and/or chain is present. Note that contacts are the same for 1ppx (shown) as they are for 1pun, 1puq, and 1pus (not included).

Figure 7. The relative positions of the DR1025 ATP (red) ligand, the Ap_4Ase ligands from PDB IDs 1ktg (pink) and 1jkn (green), and the MutT ligand from PDB ID 1tum (orange). The sites were compared by superposition of the common protein folds. Only the DR1025 protein chain is shown (chain A in blue with its Nudix box cyan, chain B in green).

significant local changes are observed at the N terminus where Met1 becomes ordered and Glu2 moves by over 3 Å compared to the apoenzyme.

DR1025 sequence similarities

In an attempt to determine which existing Nudix protein (for which we have structural and functional information) DR1025 is most similar to in terms of sequence, we performed some sequence comparisons. A BLAST (version 2.2.4)[33] search with the sequence of DR1025 against the PDB[34] identified the plant *L. angustifolius* Ap_4A hydrolase as the top scoring match with an *E*-value of 6×10^{-5}.

We also created highly specific HMMs for each of the proteins of known structure (see Materials and Methods). These HMMs were run against the protein sequences of the *D. radiodurans* genome in order to identify which sub-types if any match the sequence of DR1025. The most significant matches made to DR1025 were those made by the plant Ap_4A hydrolase HMMs, the best matching with an *E*-value of 3.9×10^{-6}. Interestingly, the individual HMMs for the worm Ap_4A hydrolase HMM did not match well. Several other *D. radiodurans* proteins were also consistently matched by the different Ap_4A hydrolase HMMs.

We note that DR1025 was also matched (with less significant *E*-values) by the Mut-T HMM, one of the ADP-ribose pyrophosphatase HMMs, the CoA pyrophosphatase and the isopentyl diphosphate isomerase HMMs. Clearly, there are overlapping segments of sequence homology for Nudix proteins with different functions, as exemplified by the matches made to the same proteins by several models here. The consistent and more significant matches made by the Ap_4A hydrolase models to DR1025 point to a closer evolutionary relationship to these than any of the others, which is supported by the complex of DR1025 with ATP, the asymmetric hydrolysis product of Ap_4A.

Discussion

The large number of Nudix proteins in the radiation resistant bacterium *D. radiodurans* strongly suggests a role for these proteins in protecting genome integrity through housecleaning and DNA maintenance. It has been shown that at least five proteins of this family are induced in *D. radiodurans* on γ-irradiation.[35] We have determined the high resolution crystal structure of a member of this family alone and in complex with two putative substrates. DR1025 forms an intertwined homodimer with the 12 amino-terminal residues from one monomer forming part of the substrate-binding pocket of the other. Co-crystallization with a non-hydrolyzable GTP analog and with Ap_4A showed both ligands bind to a common pocket in a similar orientation in the complex structures. In the Ap_4A co-crystals, ATP is found bound to the enzyme, suggesting that asymmetric hydrolysis of Ap_4A has occurred. Since no further hydrolysis to ADP or AMP is suggested, the inference is that DR1025 is an asymmetric Ap_4A hydrolase. Sequence comparison to known Nudix structures also shows DR1025 to be most similar to the plant Ap_4A hydrolase. A comparison of the manner of substrate binding of DR1025 to plant and animal Ap_4A hydrolase structures, however, clearly shows a different mode of binding and inferred mechanism. The biological specificity of function is being further explored by computational and experimental approaches.

A comparison of DR1025 with the three-dimensional structures of the functionally and evolutionary diverse Nudix proteins revealed several characteristic features in addition to the conserved Nudix fold.[15] First, as frequently found in Nudix structures, DR1025 has an N-terminal extension located prior to the Nudix fold and involved in forming the substrate-binding pocket. Second, we identified a specific metal-binding motif in and following the Nudix box. This conserved metal-binding motif is essential to coordinate the divalent metal ion necessary for substrate binding and subsequent catalysis. Finally, we located a conserved proline in β3 of DR1025

that creates a kink in the strand connecting β2 and β3 and preceding the Nudix box. This strand kink may be critical in the overall architecture of the Nudix fold.

While one focus of some recent structural genomics initiatives is discovering new folds and ancient evolutionary relationships of all proteins, these projects may overlook the structural and functional diversity amongst proteins with sequence similarity. For the Nudix protein family, this diversity is obtained by amino acid substitutions at structurally equivalent positions important for substrate binding, differences in loops linking secondary structure elements, extensions on either the N-terminal or C-terminal ends of the fold, subtle changes in the length, curvature or orientation of the secondary structure elements and oligomerization of the folding unit. Such structural diversity within a fold supplies great functional latitude, and results in high sequence divergence such that it becomes difficult to align homologous sequences even with the most elegant automated methods. In order to explore this diversity, we need detailed structural information from a large number of proteins with the same fold, but differing in sequence and function. Similar studies have been carried out such as with the four-helical cytokines and cytochromes.[36]

The Nudix proteins represent an ideal family for comparative structural studies, as they comprise a large and phylogenetically disperse family, observed in all kingdoms of life.

Materials and Methods

Expression and purification

The clone (pET24a/Tuner (DE3)) containing the DR1025 gene was constructed as described.[25] Protein was purified to homogeneity (as judged by gel electrophoresis and dynamic light scattering) from induced cultures as described.[31] Gel filtration chromatography and dynamic light scattering suggest the protein forms a dimer in solution.[31]

Crystallization and data collection

Bipyramidal crystals of DR1025 measuring 200–500 μm in the longest dimension formed in hanging drops from 100 mM sodium acetate (pH 4.5), 2 M sodium formate and 15 mg/ml protein after one to two days at 22 °C.[31] Crystals were cryo-cooled by soaking in mother liquor containing 10%, 20% and 30% (v/v) glycerol and flash-frozen in liquid nitrogen. Diffraction data for the native apoenzyme crystal (Native 1; Table 2) were collected at the Lawrence Berkeley National Laboratory Advanced Light Source (ALS) beamline 5.0.1–1.6 Å resolution.

In order to test samarium as a potential derivative, native apoenzyme crystals were soaked in 0.5 mM $SmCl_3$ for 24 hours or 250 mM $Sm(NO_3)_3$ for ten minutes (rapid soak). The incorporation of samarium into the DR1025 crystals was examined by fluorescence scans. Crystals soaked for 24 hours in 0.5 mM $SmCl_3$ did not show any incorporation of samarium. However, one such crystal diffracted to 1.4 Å resolution at beamline 5.0.2 of the ALS and the data collected is subsequently referred as the "native" dataset (Native 2; Table 2) of the apoenzyme. A crystal soaked for ten minutes in 250 mM $Sm(NO_3)_3$ showed the presence of samarium in a fluorescence scan and diffracted to a limit of 1.8 Å resolution at ALS beamline 5.0.2. Data collected by the inverse beam procedure for Friedel pairs gave an R_{sym} of 4.8% (Sm derivative; Table 2). The anomalous *R*-factor for this data set was 5.01% on F_o (7.3% for highest resolution range).

Co-crystals of DR1025 were grown in the presence of magnesium and the potential substrates GMPPNP or Ap_4A using the same conditions described above for apoenzyme crystallization with the addition of 12–18-fold molar excess of both $MgCl_2$ and substrate. Stock solutions of 0.1 M $MgCl_2$, 60 mM Ap_4A and 90 mM GMPPNP were prepared and added to the protein solution directly in the crystallization plate. Crystals of the same morphology as native grew within one to two days. Cryoprotection was the same as for the apoenzyme crystals.

Diffraction data were collected on 1025GNP and 1025ATP to a limit of 1.6 Å resolution at ALS beamline 5.0.2. Data collection and reduction statistics are given for these datasets in Table 2. The space group for both ligand co-crystals was $P4_1$, while the native and samarium datasets belonged to space group $P4_12_12$.

All diffraction images were indexed, integrated and scaled with the programs HKL and HKL2000.[37] The truncate program from the CCP4 package[38] was used to convert intensities to structure factors.

Phasing and structure determination of DR1025

Initial heavy atom positions were identified from an anomalous difference Patterson map calculated with CNS[39] using the Sm derivative data set (Table 2). Two strong heavy atom positions were found in the Harker sections and a heavy atom search was used to determine one additional site. All three positions were included in SAD phasing. The heavy atom positions were refined and density modification applied using CNS. Statistics for the phase calculation were found to be good with the following parameters: Cullis *R*-value (all data 0.47, acentric 0.44); Kraut *R*-value, 0.0301; phasing power 2.9; and figure of merit, 0.46, indicating that the Sm derivative had sufficient phasing power to solve the structure by the SAD method.

Initial electron density maps after density modification with solvent flattening were calculated using CNS. At this point enantiomorph ambiguity between $P4_12_12$ and $P4_32_12$ was investigated. This ambiguity was resolved by independently refining both choices after generating inverse images. Continuity of the electron density in the maps indicated $P4_12_12$ was the correct space group. An initial map calculated at 2.0 Å resolution was used for automatic model building with ARP/wARP.[40] About 93% of the protein main-chain and most side-chains were fit by this procedure. The resulting model and the density map were visualized with the program O.[41] At this point, the four N-terminal residues, the four C-terminal residues, and five internal residues were found to be missing.

A map was calculated to 1.6 Å from the Native 2 data and phases from the Sm derivative model and used to construct a native model. The main-chain was added for some missing residues and side-chains were manually

adjusted. This native model was then refined using CNS with standard stereochemical restraints and restrained atomic thermal parameters.

Difference maps ($2F_o - F_c$ and $F_o - F_c$) were calculated after each cycle of CNS refinement and the native model was checked and manually refitted or built when necessary until the refinement converged with the 1.6 Å data. Then, the resolution was extended to 1.4 Å and several cycles of refinement and model building were performed. At this point, water molecules were also added and refined in each cycle. This step improved the R_{free} from 30.4% to 24.5%. Densities for alternate conformations of two residues were identified and modeled (Arg14 and Arg124). Density around the loop region (35–42) remained very weak and thus it was difficult to build residues in this region. All the residues except residue 1 (Met) and residues in the loop region (36–39) were built and refined in the final model. The final *R*-factor for the native model was 22.82% and the final R_{free} was 24.21% (see Table 3).

This high-resolution native model was then refined *versus* the Sm derivative data and metals and water molecules were added as the refinement progressed. This procedure resulted in the final refined model of the Sm derivative with R_{work} 21.8% and R_{free} 25.1% (Table 3).

Structure determination of DR1025 ligand complexes

The MR method was used to phase both the 1025GNP and 1025ATP complexes. The DR1025 monomer structure was used as the search model for MR. Using the CNS program, solutions of the rotation function were calculated and the best solutions utilized for a translation search. The monomer coordinates were then fixed and a second monomer translation search conducted for the remaining rotation solutions. The resulting solution for the second monomer translation showed high packing values and correlation coefficients (for 1025ATP correlation 0.500; packing 0.5770 and for 1025GNP correlation 0.476; packing 0.5834). The correct solutions for the dimer (two monomers) were checked for overlaps using the GRASP program. After rigid body refinement, R_{work} and R_{free} values were sharply reduced (R_{free}: 45.6% to 27.8% for 1025ATP and 43.8% to 28.3% for 1025GNP), which indicated that the solutions were correct. After group *B*-factor refinement and positional refinement, R_{free} was further reduced to 27% for both structures. At this point, the density of the ligand was apparent for modeling. Ligand structures were modeled using available PDB files (HIC-UP)† and a few additional cycles of positional and *B*-factor refinement were performed. Addition of water molecules and cycles of simulated annealing reduced the R_{free} and R_{work} further to 24% and 21.5% for 1025ATP and 23% and 21.2% for 1025GNP, respectively. Finally, refinement after addition of alternate conformations for arginine residues at positions 14, and 124 reduced the *R* values further as indicated in Table 3.

Structural analyses

Structural visualization was performed using PyMOL.‡ We manually produced the secondary structure hydrogen bond diagram using hydrogen bond data from HBPLUS[42] and surface solvent accessibility data using the Lee & Richards method[43] as implemented in Naccess (S. J. Hubbard)§. We produced structural alignments by fitting the protein backbone of the DR1025 subunit with the backbone of the protein in question. We performed the fitting of the two proteins using the McLachlan algorithm[44] as implemented in ProFit (A.C.R. Martin)‖. For superposition with proteins other than itself, we carried out iterative rounds of fitting based on Nudix consensus sequence, hydrogen bond data from HBPLUS,[42] the CCP4 program CONTACT,[38] solvent accessible surface area calculated using NACCESS and visual inspection. We calculated protein–metal contacts using the CCP4 program, CONTACT.

Specific HMM construction and searching

Using the individual sequences of the Nudix proteins of known structure we collected homologs from the snr database[45] using BLAST[33] with an *E*-value cut-off of $\leq 10^{-12}$. This cut-off permitted collection of only very close homologs to construct a highly specific HMM. In cases where one or more structures have been solved for the same protein in different organisms, an alignment including all representative structure sequences and their homologs was also created. Each individual alignment was then used to create an HMM using HMMER (version 2.3.2.)[46] these HMMs were calibrated and searched against the *D. radiodurans* genome. We only considered matches made to *Deinococcus* sequences by HMMER with an *E*-value of $\leq 10^{-3}$.

Protein Data Bank accession numbers

The data and coordinates for all models have been deposited in the Protein Data Bank under accession numbers RCSB 021774 (1sjy), RCSB 021876 (1sol), RCSB 021990 (1su2) and RCSB 022110 (1sz3).

Acknowledgements

This work was funded by a grant from the US Department of Energy, Office of Biological Energy Research, contract no. DE-AC03-76SF00098. S.E.B. is a Searle scholar (01-L-116) and is supported by NIH 1-K22-HG00056. M.J.B was funded by an NIH grant, GM 18649. Diffraction data were collected at beamlines 5.0.1 and 5.0.2 of the Macromolecular Crystallography Facility at the Advanced Light Source, Lawrence Berkeley National Laboratory.

References

1. Bessman, M. J., Frick, N. & O'Handley, S. F. (1996). The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "Housecleaning" enzymes. *J. Biol. Chem.* **271**, 25059–25062.
2. Treffers, H. P., Spinelli, V. & Belser, N. O. (1954). A

† http://xray.bmc.uu.se/hicup/

‡ DeLano, W. L. (2002). The PyMOL Molecular Graphics System.

§ http://wolf.bms.umist.ac.uk/naccess/

‖ http://www.bioinf.org.uk/software/profit/

factor (or mutator gene) influencing mutation rates in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **40**, 1064–1071.
3. Bhatnagar, S. K., Bullions, L. C. & Bessman, M. J. (1991). Characterization of the mutT nucleoside triphosphatase of *Escherichia coli*. *J. Biol. Chem.* **266**, 9050–9054.
4. Wang, Z., Jiao, X., Carr-Schmid, A. & Kiledjian, M. (2002). The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proc. Natl Acad. Sci. USA*, **99**, 12663–12668.
5. Perraud, A. L., Fleig, A., Dunn, C. A., Bagley, L. A., Launay, P., Schmitz, C. *et al.* (2001). ADP-ribose gating of the calcium-permeable LTRPC2 channel revealed by Nudix motif homology. *Nature*, **411**, 595–599.
6. Bessman, M. J., Walsh, J. D., Dunn, C. A., Swaminathan, J., Weldon, J. E. & Shen, J. (2001). The gene ygdP, associated with the invasiveness of *Escherichia coli* K1, designates a Nudix hydrolase, Orf176, active on adenosine (5′)-pentaphospho-(5′)-adenosine (Ap5A). *J. Biol. Chem.* **276**, 37834–37838.
7. Conyers, G. B. & Bessman, M. J. (1999). The gene, ialA, associated with the invasion of human erythrocytes by *Bartonella bacilliformis*, designates a nudix hydrolase active on dinucleoside 5′-polyphosphates. *J. Biol. Chem.* **274**, 1203–1206.
8. Gaywee, J., Xu, W., Radulovic, S., Bessman, M. J. & Azad, A. F. (2002). The *Rickettsia prowazekii* invasion gene homolog (invA) encodes a Nudix hydrolase active on adenosine (5′)-pentaphospho-(5′)-adenosine. *Mol. Cell. Proteomics*, **1**, 179–185.
9. Ismail, T. M., Hart, C. A. & McLenna, A. G. (2003). Regulation of dinucleoside polyphosphate pools by the YgdP and ApaH hydrolases is essential for the ability of *Salmonella enterica serovar Typhimurium* to invade cultured mammalian cells. *J. Biol. Chem.* **278**, 32602–32607.
10. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R. *et al.* (2002). The Pfam protein families database. *Nucl. Acids Res.* **30**, 276–280.
11. Safrany, S. T., Caffrey, J. J., Yang, X., Bembenek, M. E., Moyer, M. B., Burkhart, W. A. & Shears, S. B. (1998). A novel context for the "MutT" module, a guardian of cell integrity, in a diphosphoinositol polyphosphate phosphohydrolase. *EMBO J.* **17**, 6599–6607.
12. Fisher, D. I., Safrany, S. T., Strike, P., McLennan, A. G. & Cartwright, J. L. (2002). Nudix hydrolases that degrade dinucleoside and diphosphoinositol polyphosphates also have 5-phosphoribosyl 1-pyrophosphate (PRPP) pyrophosphatase activity that generates the glycolytic activator ribose 1,5-bisphosphate. *J. Biol. Chem.* **277**, 47313–47317.
13. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
14. Fletcher, J. I., Swarbrick, J. D., Maksel, D., Gayler, K. R. & Gooley, P. R. (2002). The structure of Ap(4)A hydrolase complexed with ATP-MgF(x) reveals the basis of substrate binding. *Structure*, **10**, 205–213.
15. Gabelli, S. B., Bianchet, M. A., Bessman, M. J. & Amzel, L. M. (2001). The structure of ADP-ribose pyrophosphatase reveals the structural basis for the versatility of the Nudix family. *Nature Struct. Biol.* **8**, 467–472.
16. Kang, L. W., Gabelli, S. B., Cunningham, J. E., O'Handley, S. F. & Amzel, L. M. (2003). Structure and mechanism of MT-ADPRase, a nudix hydrolase from *Mycobacterium tuberculosis*. *Structure (Camb)*, **11**, 1015–1023.
17. Shen, B. W., Perraud, A. L., Scharenberg, A. & Stoddard, B. L. (2003). The crystal structure and mutational analysis of human NUDT9. *J. Mol. Biol.* **332**, 385–398.
18. Wang, S., Mura, C., Sawaya, M. R., Cascio, D. & Eisenberg, D. (2002). Structure of a Nudix protein from *Pyrobaculum aerophilum* reveals a dimer with two intersubunit beta-sheets. *Acta Crystallog. sect. D*, **58**, 571–578.
19. Mattimore, V. & Battista, J. R. (1996). Radioresistance of *Deinococcus radiodurans*: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. *J. Bacteriol.* **178**, 633–637.
20. Makarova, K. S., Aravind, L., Daly, M. J. & Koonin, E. V. (2000). Specific expansion of protein families in the radioresistant bacterium *Deinococcus radiodurans*. *Genetica*, **108**, 25–34.
21. White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D., Dodson, R. J. *et al.* (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science*, **286**, 1571–1577.
22. Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V. *et al.* (2003). Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature*, **423**, 87–91.
23. Read, T. D., Peterson, S. N., Tourasse, N., Baillie, L. W., Paulsen, I. T., Nelson, K. E. *et al.* (2003). The genome sequence of *Bacillus anthracis Ames* and comparison to closely related bacteria. *Nature*, **423**, 81–86.
24. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M. *et al.* (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1474.
25. Xu, W., Shen, J., Dunn, C. A., Desai, S. & Bessman, M. J. (2001). The Nudix hydrolases of *Deinococcus radiodurans*. *Mol. Microbiol.* **2001**, 286–290.
26. Xu, W., Shen, J., Dunn, C. A. & Bessman, M. J. (2003). Substrate switching of the Nudix hydrolases of *Deinococcus radiodurans*. *FASEB J. (Part 1 suppl.)*, **17**, A574.
27. O'Handley, S. F., Frick, D. N., Bullions, L. C., Mildvan, A. S. & Bessman, M. J. (1996). *Escherichia coli* orf17 codes for a nucleoside triphosphate pyrophosphohydrolase member of the MutT family of proteins. Cloning, purification, and characterization of the enzyme. *J. Biol. Chem.* **271**, 24649–24654.
28. O'Handley, S. F., Dunn, C. A. & Bessman, M. J. (2001). Orf135 from *Escherichia coli* is a Nudix hydrolase specific for CTP, dCTP, and 5-methyl-dCTP. *J. Biol. Chem.* **276**, 5421–5426.
29. Xu, W., Shen, J., Dunn, C. A. & Bessman, M. J. (2003). A new subfamily of the Nudix hydrolase superfamily active on 5-methyl-UTP (ribo-TTP) and UTP. *J. Biol. Chem.* **278**, 37492–37496.
30. Kang, L. W., Gabelli, S. B., Bianchet, M. A., Xu, W. L., Bessman, M. J. & Amzel, L. M. (2003). Structure of a coenzyme A pyrophosphatase from *Deinococcus radiodurans*: a member of the Nudix family. *J. Bacteriol.* **185**, 4110–4118.
31. Holbrook, E. L., Schulze-Gahmen, U., Buchko, G. W., Ni, S., Kennedy, M. A. & Holbrook, S. R. (2003). Crystallization and preliminary X-ray analysis of two nudix hydrolases from *Deinococcus radiodurans*. *Acta Crystallog. sect. D*, **59**, 737–740.
32. Nicholls, A., Sharp, K. & Honig, B. (1991). Protein

folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
33. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res*,, **25**, 3389–3402.
34. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.
35. Liu, Y., Zhou, J., Omelchenko, M. V., Beliaev, A. S., Venkateswaran, A., Stair, J. *et al.* (2003). Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. *Proc. Natl Acad. Sci. USA*, **100**, 4193–4196.
36. Hill, E. E., Morea, V. & Chothia, C. (2002). Sequence conservation in families whose members have little or no sequence similarity: the four-helical cytokines and cytochromes. *J. Mol. Biol.* **322**, 205–233.
37. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, C. W. & Sweet, R. M., eds), vol. 276A, pp. 307–326, Academic Press, New York.
38. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
39. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
40. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
41. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard (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.
42. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.
43. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
44. McLachlan, A. D. (1982). Rapid comparison of protein structures. *Acta Crystallog. sect. A*, **38**, 871–873.
45. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E. *et al.* (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucl. Acids Res.* **31**, 365–370.
46. Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
47. Abeygunawardana, C., Weber, D. J., Gittis, A. G., Frick, D. N., Lin, J., Miller, A. F. *et al.* (1995). Solution structure of the MutT enzyme, a nucleoside triphosphate pyrophosphohydrolase. *Biochemistry*, **34**, 14997–15005.
48. Lin, J., Abeygunawardana, C., Frick, D. N., Bessman, M. J. & Mildvan, A. S. (1997). Solution structure of the quaternary MutT∇M2 + −AMPCPP−M2 + complex and mechanism of its pyrophosphohydrolase action. *Biochemistry*, **36**, 1199–1211.
49. Massiah, M. A., Saraswat, V., Azurmendi, H. F. & Mildvan, A. S. (2003). Solution structure and NH exchange studies of the MutT pyrophosphohydrolase complexed with Mg(2 +) and 8-oxo-dGMP, a tightly bound product. *Biochemistry*, **42**, 10140–10154.
50. Gabelli, S. B., Bianchet, M. A., Ohnishi, Y., Ichikawa, Y., Bessman, M. J. & Amzel, L. M. (2002). Mechanism of the *Esherichia coli* ADP-ribose pyrophosphatase, a Nudix hydrolase. *Biochemistry*, **41**, 9279–9285.
51. Swarbrick, J. D., Bashtannyk, T., Maksel, D., Zhang, X. R., Blackburn, G. M., Gayler, K. R. & Gooley, P. R. (2000). The three-dimensional structure of the Nudix enzyme diadenosine tetraphosphate hydrolase from *Lupinus angustifolius* L. *J. Mol. Biol.* **302**, 1165–1177.
52. Bailey, S., Sedelnikova, S. E., Blackburn, G. M., Abdelghany, H. M., Baker, P. J., McLennan, A. G. & Rafferty, J. B. (2002). The crystal structure of diadenosine tetraphosphate hydrolase from *Caenorhabditis elegans* in free and binary complex forms. *Structure*, **10**, 589–600.
53. Durbecq, V., Sainz, G., Oudjama, Y., Clantin, B., Bompard-Gilles, C., Tricot, C., Caillet, J. *et al.* (2001). Crystal structure of isopentenyl diphosphate: dimethylallyl diphosphate isomerase. *EMBO J.* **20**, 1530–1537.
54. Bonanno, J. B., Edo, C., Eswar, N., Pieper, U., Romanowski, M. J., Ilyin, V., Gerchman, S. E. *et al.* (2001). Structural genomics of enzymes involved in sterol/isoprenoid biosynthesis. *Proc. Natl Acad. Sci. USA*, **98**, 12896–12901.
55. Wouters, J., Oudjama, Y., Ghosh, S., Stalon, V., Droogmans, L. & Oldfield, E. (2003). Structure and mechanism of action of isopentenylpyrophosphate-dimethylallylpyrophosphate isomerase. *J. Am. Chem. Soc.* **125**, 3198–3199.
56. Buchko, G. W., Ni, S., Holbrook, S. R. & Kennedy, M. A. (2003). 1H, (13)C, and (15)N NMR assignments of the hypothetical Nudix protein DR0079 from the extremely radiation-resistant bacterium *Deinococcus radiodurans*. *J. Biomol. NMR*, **25**, 169–170.
57. Karp, P. D., Riley, M., Paley, S. M., Pellegrini-Toole, A. & Krummenacker, M. (1999). Eco Cyc: encyclopedia of *Escherichia coli* genes and metabolism. *Nucl. Acids Res.* **27**, 55–58.

Edited by R. Huber

(Received 22 January 2004; accepted 29 January 2004)