Genetica **108:** 25–34, 2000.

Specific expansion of protein families in the radioresistant bacterium *Deinococcus radiodurans*

Kira S. Makarova[1,2,3], L. Aravind[2], Michael J. Daly[1] & Eugene V. Koonin[2]
[1]*Uniformed Services University of the Health Sciences, Bethesda, MD 20814-479, USA;* [2]*The National Center for Biotechnology Information National Library of Medicine, The National Institutes of Health, Bethesda, MD 20814, USA;* [3]*Permanent address: Institute of Cytology and Genetics, Russian Academy of Sciences, Novosibirsk 630090, Russia (Phone: (301)-435-5912; Fax: (301)-480-9241; E-mail: makarova@ncbi.nlm.nih.gov)*

Key words: Deinococcus, genome analysis, protein family, specific expansion

Abstract

Computer analysis of the complete genome of *Deinococcus radiodurans* R1 reveals a number of protein families, which are over-represented in this organism, compared to most other bacteria with known genome sequences. These families include both previously characterized and uncharacterized proteins. Most of the families whose functions are known or could be predicted seem to be related to stress-response and elimination of damage products (cell-cleaning). The two most prominent family expansions are the Nudix (MutT) family of pyrophosphohydrolases and a previously unnoticed family of proteins related to *Bacillus subtilis* DinB that could possess a metal-dependent enzymatic activity whose exact nature remains to be determined. Several proteins of the expanded families, particularly the Nudix family, are fused to other domains and form multidomain proteins that are so far unique for *Deinococcus*. The domain composition of some of these proteins indicates that they could be involved in novel DNA-repair pathways. Such unique proteins are good targets for knock-out and gene expression studies, which are aimed to shed light on the unusual features of this interesting bacterium.

Introduction

Specific expansion of some protein families has been detected in the course of analysis of complete genomes in many bacteria and archaea [10, 22, 32]. In many cases, there seems to be a connection between the expansion of a particular protein family and the lifestyle of the corresponding organism. For instance, propagation of ferredoxins has been detected in autotrophic archaea [22], expansion of several families of enzymes required for lipid degradation has been seen in *Mycobacterium tuberculosis* [10], and multiple cytochromes necessary for metal reduction have been identified in the metal-reducing bacterium *Shewanella* [25]. These findings indicate that protein family expansion may play an important role in bacterial evolution and shed light on the pathways or biochemical mechanisms that could be responsible for unique features of different bacterial lineages.

Recently, the complete genome sequence of the radio-resistant bacterium *Deinococcus radiodurans* has become available [36]. The original genome analysis has revealed an unusual propagation of Nudix pyrophosphatases, which could be implicated in 'a cleaning up' of macromolecule damage products, but has not specifically sought to detect potential expansion of other protein families. Taking into account the importance of this bacterium in terms of its extraordinary resistance to various forms of DNA damage [5, 36, 23, 24] and its possible future role in bioremediation programs (Brim, 2000 in press) [20], we were interested in obtaining additional information on unique genome features of *D. radiodurans*, and in particular, specific protein family expansion.

We used profile searches to characterize protein family expansion in *D. radiodurans* as comprehensively as it is feasible with the currently available methods. In addition to the Nudix pyrophosphatase

superfamily, specific expansion of more than 20 other families, both functionally characterized and uncharacterized, was detected. Furthermore, we identified several families of proteins that so far seem to be unique for *Deinococcus* and potentially could be involved in yet unknown mechanisms of DNA damage resistance.

Material and methods

Nucleotide sequence and list of all coding sequences of *Deinococcus radiodurans* are available from TIGR (http://www.tigr.org/tdb/GMR/gdr/htmls/SplashPage.ht ml) and NCBI (http://ncbi.nlm.nih.gov/genebank/genomes/bacteria/Drad/). All *Deinococcus* proteins sequences were compared to the Non-Redundant (NR) protein sequence database at the NCBI (NIH, Behtesda) and among themselves using the gapped BLASTP program [2]. Iterative database searches were performed using the PSI-BLAST program [2]. Additionally, the *Deinococcus* protein sequences were compared to libraries of position-specific scoring matrices (PSSMs) generated from multiple alignments [37, 9] using the IMPALA program [28].

To investigate possible specific protein family expansion as comprehensively as possible, the following multi-step procedure was used. The *Deinococcus* protein set, after filtering for low complexity using the SEG program with default parameters [38] and obtaining all pairwise similarity scores using the gapped BLAST program, was clustered by the single-linkage algorithm (clustering threshold e-value 0.001) using the GROUPER program [35]. One representative sequence from each cluster was selected to generate a PSSM using iterative PSI-BLAST search, first against the *Deinococcus* protein set, and then against the NR database. These PSSMs were used to search for additional members of the families among the *Deinococcus* proteins. Related families, that is, those that were recognized by the same PSSM, were combined into superfamilies. The number of members of each superfamily in other bacteria with large genomes (*Escherichia coli*, *Bacillus subtilis*, *Mycobacterium tuberculosis* and *Synechocystis* sp.) was determined using the previously described PSSM collections [37, 9] or new PSSMs generated in the course of this analysis.

Multiple alignments of protein families were constructed using the ALITRE program [30] and adjusted, if necessary, on the basis of the examination of PSI-BLAST search outputs.

Results and discussion

Specific expansion of protein families and unique multidomain proteins in Deinococcus

Several cases of protein family expansion in *D. radiodurans* seem to show a distinct connection with stress-response and damage-resistance (Table 1). In particular, various families of hydrolases are over-represented. These include MutT-like pyrophosphatases, calcineurine-like phosphoesterases, lipase/epoxidase-like (alpha/beta) hydrolases, subtilisin-like proteases, and sugar deacetylases. In addition to such specifically expanded families, several other families of hydrolases are present in Deinococcus in significant numbers but all of them are also wide spread in other bacteria and are not shown here. Most of these hydrolases are likely to be involved in the decomposition of damage products ('cell cleaning') under stress conditions. Some other expanded families are even more directly related to stress-response systems, for example, proteins involved in tellurium resistance (TerE) and homologs of plant pathogenesis-related proteins (PR1).

Expansion of the Nudix hydrolase family is one of the most prominent features of the *Deinococcus* genome. The prototype member of this family, MutT, is a pyrophosphohydrolase that hydrolyzes 8-oxo-dGTP, releasing pyrophosphate, and is the central component of an antimutagenic system responsible for preventing oxdiative damage to DNA. Subsequently, it has been shown that proteins containing the MutT domain [18] and comprising the so-called Nudix superfamily of pyrophosphohydrolases hydrolyze a variety of substrates [7]. In functional terms, Nudix hydrolases have been loosely defined as house-cleaning enzymes that destroy potentially deleterious damage products. A detailed analysis of Nudix proteins in *Deinococcus* showed that five of them are multidomain proteins, where the MutT domain is combined with other domains (Figure 1). Apparent orthologs for three of these multidomain proteins were detected in other bacteria. In particular, the protein family prototyped by *E.coli* YjaD contains a Zn-ribbon module, which is probably involved in nucleic-acid-binding. Orthologs of this protein are present in many bacteria of the gamma subdivision of proteobacteria, *Mycobacterium* and eu-

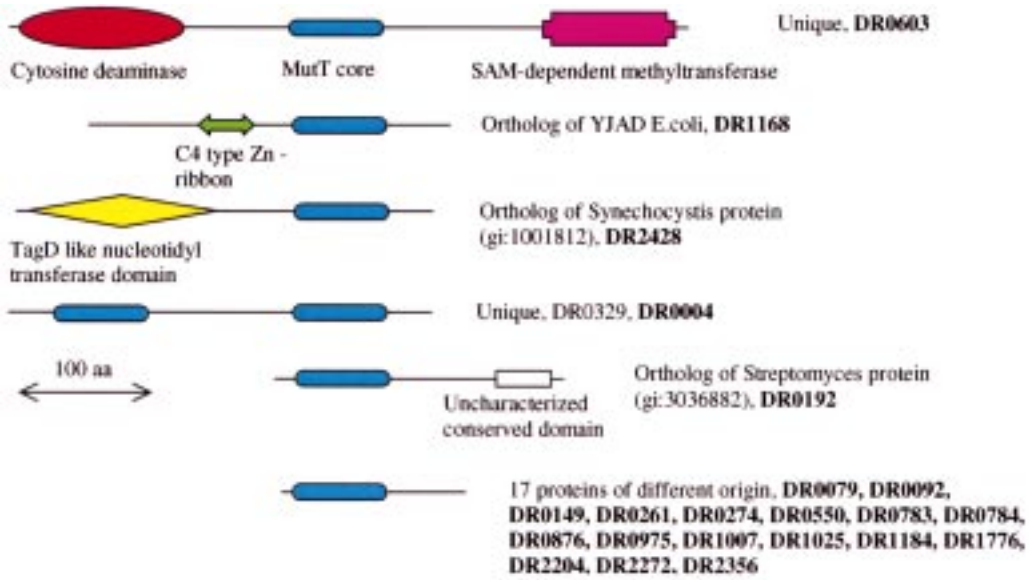

Figure 1. Distinct domain architectures of proteins containing the MutT-like domain.

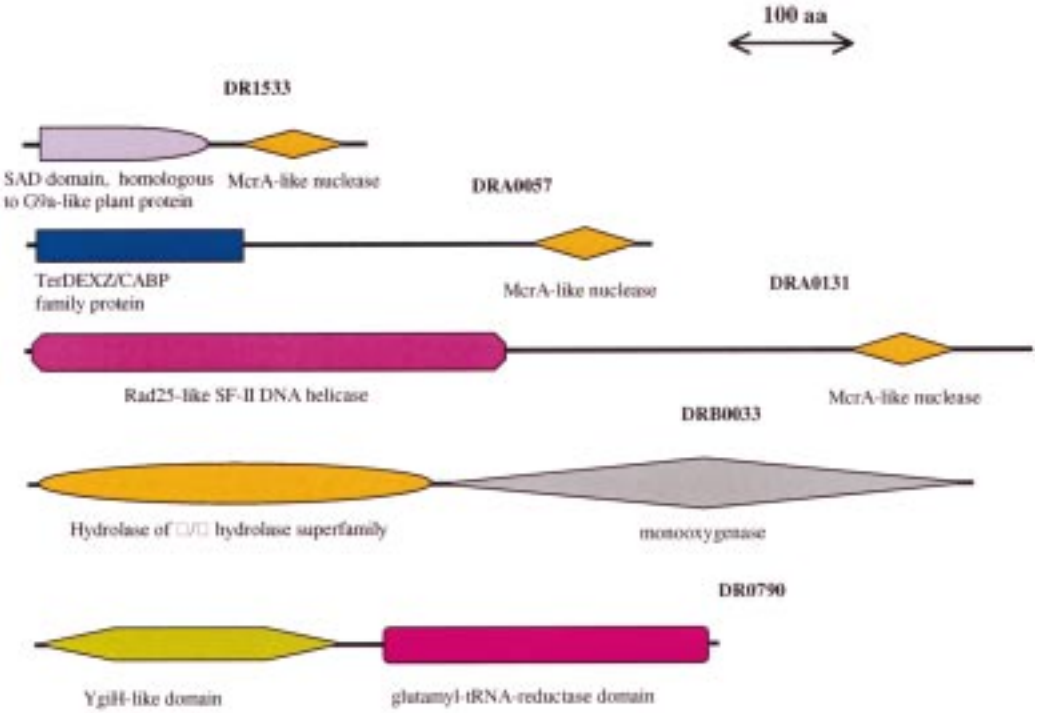

Figure 2. Domain architectures of selected proteins belonging to families expanded in *Deinococcus*.

Table 1. Comparison the number of proteins in the families expanded in *Deinococcus radiodurans* with some other bacteria with the large genome

	Bacillus subtilus	*Escherichia coli*	*Mycobacterium tuberculosis*	*Synechocystis*	*Deinococcus radiodurans*
Proteome size	4099	4289	3918	3169	3204
MutT-like pyrophosphatase	5	13	11	8	23
Calcineurin like phosphoesterase	13	10	5	5	16
Metallo-beta lactamases	16	8	17	14	15
Aminoacetyltransferases	45	23	18	15	45
Subtilisin-like proteases	8	0	7	2	10
DinB/YfiT homologs, predicted metal-dependent enzymes	8	0	7	2	13
Tellurium resistance protein TerE/TerA	3	2	0	1	5
McrA endonuclease family	3	4	6	4	7
Sugar deacetylases, PIG-L family	3	3	3	0	6
Plant pathogenesis related proteins, PR1 family	3	0	1	0	5
Conserved archaeal family	0	0	1	0	5
DegV like protein family	1	0	1	0	5
WD40 repeats	0	0	0	6	5
YIIM_ECOLI family	1	1	1	0	4
YGIH_ECOLI family	1	1	0	1	4
Enterococcus VANW homolog	1	0	0	0	3
Homolog of 2621836_MTH	0	0	0	3	3
Homolog 758294_yeast	0	0	0	0	2

karyotes (human, yeast). In *Mycobacterium* and eukaryotes, the Zn-finger motif is completely or partially disrupted. Another *Deinococcus* protein contain an apparently inactivated (with catalytic motif REXXEE missing) MutT domain combined with a TagD-like nucleotidyltransferase domain. The ortholog of this protein from *Synechocystis* contains an intact catalytic motif and, accordingly, is predicted to possess both the pyrophosphatase and the nucleotidyltransferase activities, whereas in *Deinococcus*, the inactivated MutT domain is likely to perform a regulatory function. A second TagD-like nucleotidyltransferase from *Deinococcus* (DRA0273) is very similar to these two proteins, but the MutT apparently has been disrupted beyond recognition. Orthologs of a third Nudix protein, which contains an uncharacterized C-terminal domain, are present in *Streptomyces*, *Mycobacterium* and *Synechocystis*. In most of them, the Nudix pyrophosphohydrolase seems to be inactivated, suggesting a regulatory function. The unique domain architecture of the DR0603 protein consists of a cytosine deaminase, MutT and SAM-dependent methyltransferase domains, a combination suggestive of a role in an as yet uncharacterized repair pathway. Two closely related *Deinococcus* proteins contain a duplication of the MutT domain, which so far has not been observed in any other organism. Three more Nudix proteins are specifically related to the proteins containing the duplication of the MutT domain, and the genes for two of these are adjacent on the chromosome (DR0783, DR0784). It seems that these seven related MutT domains form a *Deinococcus*-specific subfamily of Nudix hydrolases. Altogether, *Deinococcus* encodes 23 Nudix family proteins that contain 25 individual MutT domains. Some of these proteins are likely to be repair enzymes, including the typical MutT, as well as novel ones, as suggested by the domain combinations discussed above; the functions of others are likely to involve utilization damage products that are formed under stress conditions such as desiccation or irradiation. It is unlikely that all these MutT-containing proteins have been encoded a distant ancestor of the *Deinococcus* lineage. Rather, it appears that the heterogeneous collection of Nudix family pyrophosphohydrolases encoded by *D. radiodurans* had been compiled via the mixed routes of duplication, in the case of the distinct family of seven domains, and horizontal gene transfer. The major role

for the latter is indicated by the scattered phylogenetic distribution of the orthologs of several of the *Deinococcus* Nudix proteins.

Expansion of several other protein families seems to make sense in terms of the unusual stress-resistance capabilities of *D. radiodurans*. For example, *Deinococcus* encodes seven small nuclease domains related to the McrA endonuclease of *E. coli* [16]. The McrA-like nuclease domain is part of several multidomain protein architectures that seem to be unique for *Deinococcus* and, at least in some cases, are suggestive of repair functions (Figure 2). A particularly clear-cut example of such functionally interpretable association is DRA0131 protein in which the endonuclease domain is combined with a RAD25-like helicase. In DR1533, the McrA-like endonuclease is linked to a SAD domain, which so far had been detected only in eukaryotic chromatin-associated proteins [13]. In another protein, the McrA domain is fused to a domain of the TerDEXZ/CABP family, which also is expanded in the *Deinococcus* genome (see below). There seems to be little doubt that this previously unnoticed propagation of McrA-like nucleases makes a contribution to the repair potential of *Deinococcus*. In evolutionary terms, the McrA domain, just as the MutT domain, seems to be mobile, and while two of McrA-like genes in *Deinococcus* may be the results of a relatively recent duplication (DR1312 and DR2483, which are 50% identical), the rest are likely to have been acquired via horizontal gene transfer.

Expansion of the TerDEXZ/CABP family is of interest since in other bacteria, some proteins of this family confer resistance to various damaging agents, including heavy metal anions, methyl methane sulfonate, mitomycin C and UV [4, 17] and other forms of stress [3]. The CABP1 and CABP2 proteins, members of this family that are expressed during starvation in *Dictyostelium*, form a heterodimer which binds cAMP [14]. Thus the proteins of this family may be predicted to participate in cAMP(cGMP)-dependent regulation of stress-response.

Deinococcus encodes five members of the pathogenesis-related (PR) protein family, which is widespread in eukaryotes, but only sporadically seen in bacteria (Table 1). In plants, the expression of PR proteins is induced by viral and fungal infection and unspecific wounds [1, 21]. All PR proteins show high thermal stability and resistance to proteolytic degradation. The recently resolved NMR structure of a PR1 protein [12] reveals a unique, rigid alpha-beta-alpha sandwich fold stabilized by disulfide bonds, which is compatible with the exceptional stability of these proteins. In this case, it seems likely that PR1 proteins confer resistance to various damaging agents in plants and bacteria, particularly *Deinococcus*, via similar mechanisms. The nature of these mechanisms is not yet known and its elucidation will be of major interest.

A particularly notable expansion is seen in the family of uncharacterized, small proteins prototyped by *B. subtilis* DinB, a damage-inducible gene product [8]. *Deinococcus* encodes by far the greatest number of these proteins among bacteria, with a smaller, independent expansion seen in *B. subtilis* and *Mycobacterium* (Table 1). Examination of the multiple alignment of this family (Figure 3) reveals three conserved histidines which could form a catalytic triad of a novel metal-dependent enzyme, perhaps a hydrolase. So far, the only direct connection between this family and stress-response is the induction of *dinB* by DNA damage and of another *B. subtilis* gene of this family, *yfiT*, by general stress [39]. The prediction of enzymatic activity of these proteins raises the interesting possibility that they could be nucleases directly involved in DNA degradation, which occurs in *Deinococcus* immediately after irradiation or other acute DNA damage [6, 34]. It seems that this interesting protein family is particularly amenable to experimental investigation given its expansion in *B. subtilis*, a thoroughly characterized model organism.

Independent expansion of such families as alpha/beta hydrolases and S-adenosylmethionine-dependent methylases in *Deinococcus* and *Mycobacterium*, and subtilisin-like proteases, aminoacetyltransferases and the so-called DinB family in *Deinococcus* and *Bacillus* could point to some similar features in the respective lifestyles of these bacteria. Being an obligatory heterotroph, *Deinococcus* could use its numerous lipases (alpha/beta hydrolases), that presumably have different specificities, for utilizing exogenous lipids as it has been demonstrated for *Mycobacterum* [10]. Similarly, secreted subtilases, digesting exogenous proteins, as in *Bacillus*, could serve for obtaining amino acids. Alpha/beta hydrolases also could be involved in certain stress-response pathways. Enzymes of this superfamily are mainly neutral lipases or acetyl esterases, but some of them possess unusual substrate specificity, for example, heroin esterase from *Rhodococcus* [27] and antibiotic bialaphos acetyl esterase from *Streptomyces* [26], or activity, for example, metal-ion-free oxidoreductase from *Streptomyces* [15]. A role in xenobiotic

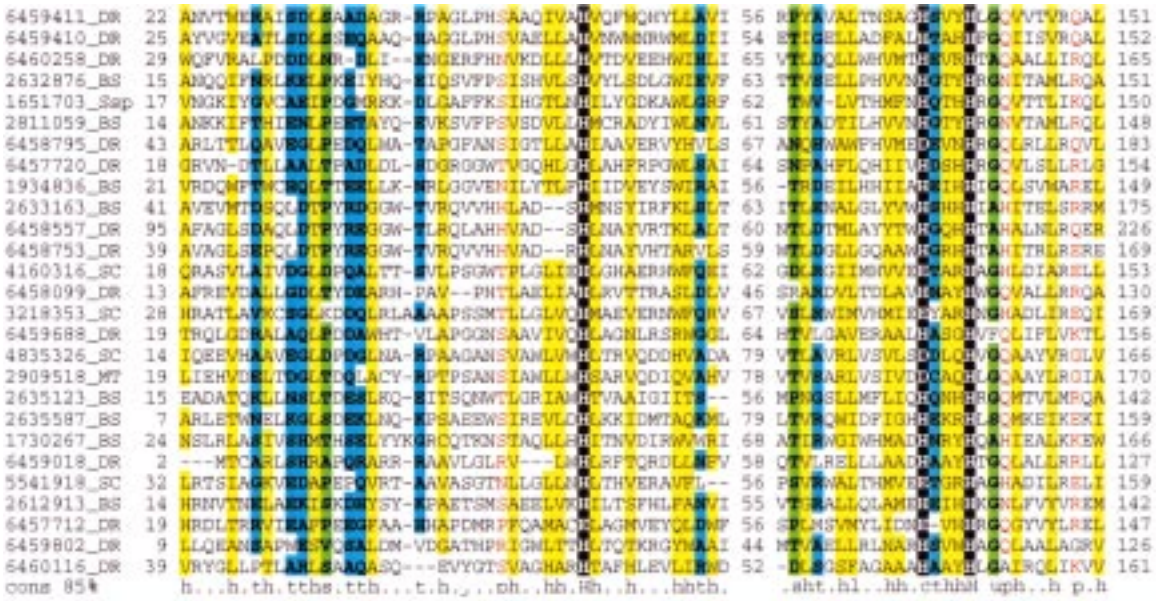

Figure 3. Multiple alignment of the conserved core of the DinB/YfiT protein family. The alignment was generated by parsing the PSI-BLAST HSPs and realigning them with ALITRE [30]. The numbers between aligned blocks indicate the lengths of variable inserts that are not shown; the numbers at the end of each sequence indicate the distances from the protein termini to the proximal and distal aligned blocks. The shading of conserved residues is according to the 85% consensus. The three predicted metal ligand residues are shown in inverse shading (white against a black background); Consensus sequence was obtained by 'consensus' program (http://www.bork.embl-heidelberg.de/Alignment/consensus.html) with default amino acid grouping assignments (**h**, **s**, **t**, **p**, +, etc.). The coloring of conserved position is the following: **h** – hydrophobic residues (yellow background); **s** – small residues (bold with green background); **t** turn-like residues (bold with cyan background); + positively charged and polar (red): In front of each sequence the Genbank identifier number (GI) and a two-letter code of species are shown. DR B *Deinococcus radiodurans*, BS B *Bacillus subtilis*, SC B *Streptomyces coelicolor*, MT – *Mycobacterium tuberculosis*, Ssp B *Synecocystis* sp.

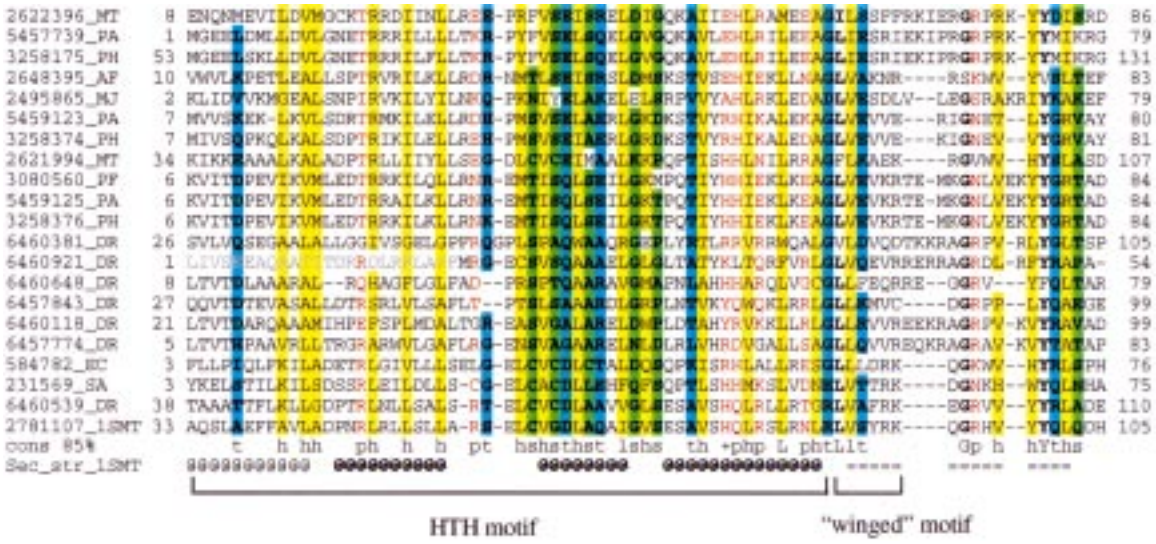

Figure 4. Multiple alignment of the conserved core of the *Deinococcus*-specific winged helix-turn-helix protein family of predicted transcriptional regulators. The numbers and coloring in this alignment are the same as in Figure 3. Secondary structure was predicted using the PDB record 1smt (SMTB repressor from *Synechococcus*) as a template. = indicates ß-strands, and @ indicates a-helix. Two letter code for species is the following: AF – *Archaeaoglobus fulgidus*, MJ B *Methanococcus jannaschii*, MTH – *Methanobacterium thermoauthotroficum*, PH B *Pyrococcus horikoshii*, PA B *Pyrococcus abyssi*, PH B *Pyrococcus furiosus*, EC B *Escherichia coli*, SA B *Staphylococcus aureus*, the rest are as in Figure 3. Corrected sequence fragment of 6460921_DR is shown in gray.

Table 2. Unique protein families in *Deinococcus*

ORFs belong to the family	Range of identity (%)	Approximate length	Sequence features and comments
DRA0346 DRB0145	33	400	α/β proteins
DR1261 DR1348	31	80	DQE/H-rich proteins, predominantly β strand proteins; Present in Caulobacter crescentus unfinished genome
DR1022 DR2185	43	150	Predominantly a helix proteins; N-terminal domain in DR1022 (C-terminal domain B MazG-like protein, related to phosphoribosyl-ATP pyrophosphatase
DR0082 DR2593 DR1748	31–35	160	Repetitive sequences (GRhGG repeats); coiled-coil
DR2532 DR2457	43	120	α/β proteins,Tryptophan-rich
DR0871 DR1920 DR2360	36–44	120	Membrane proteins; CXPXXXC motif; DR0871 has duplication of the domain
DR1814 DR1000	30	150	Predominantly a helical proteins;
DR2179 DR1611	71	150	Possible recent duplication; α/β proteins
DR1251 DR1319 DR1545	26–31	180	Secreted; a/ß proteins
DR1530 DR0419	43	130	Predominantly a helical proteins; G-rich loop
DRA0012 DR2241	43	450	Both have 90 aa N-terminal repeat; contain motif CXXC and CXXXC; predominantly a helical and coiled-coil proteins.
DR0481 DR1195 DR1301	31–44	170	Predominantly a helical proteins; some have trans-membrane segments
DR0387 (DR2038+ DR2039)	38–46	260	Predominantly a helical proteins

metabolism and/or decomposition of damage products seems likely for the expanded families of alpha/beta hydrolases in *Deinococcus*. One of the *Deinococcus* alpha/beta hydrolases is fused to a flavin-containing monooxygenase domain, so far a unique domain configuration (Figure 2). The well-established role of flavin-containing monooxygenases in xenobiotic transformation and oxygen reactivity [29] adds credence to the hypothesis that the two domains function jointly in some form of stress-response.

Amino-group acetyltransferases comprise another family that appears to have undergone independent expansion in *Deinococcus* and in *Bacillus*. Acetyltransferases of this type participate in various metabolic pathways, including lipid biosynthesis, and in regulatory systems. Other bacteria, except for *B. subtilis*, have less than half the number of these enzymes found in *D. radiodurans*. Some of these enzymes in *B. subtilis* are involved in detoxication and drug-resistance pathways [19]. Probably in *Deinococcus* they perform

similar function, especially in the disruption of various toxic products, which arise upon irradiation.

Several other family expanded in *Deinococcus* include only uncharacterized proteins. For example, *Deinococcus* encodes four members of a conserved family of membrane proteins, prototyped by the *E. coli* YgiH. One of these proteins is fused to a domain of glutamyl-tRNA-reductase (Figure 2), which apparently has lost the enzymatic activity as indicated by the disruption of the active center (not shown). Similar inactivated versions of glutamyl-tRNA-reductase are seen in cyanobacteria, suggesting an as yet uncharacterized regulatory pathway.

Some of the uncharacterized families that are expanded in *Deinococcus* are widespread among Archaea or Eukaryotes. For example, WD40 repeats are present in a variety of eukaryotic regulatory proteins, but among bacteria, are detectable only in *Deinococcus* and *Synechocystis*; their functions in these bacteria are unclear. This observation as well as unexpected expansion of an uncharacterized achaea-specific protein family could indicate that *Deinococcus* has acquired some specific physiologic mechanisms through horizontal transfer of eukaryotic and archaeal genes.

Specialized and unique protein families in Deinococcus

Several families of *Deinococcus* proteins are highly specialized and, in the initial analysis, appeared to have no homologs in other species. Database searches with individual sequences of the respective proteins failed to show statistically significant similarity to any proteins other than their paralogs from *Deinococcus*. Only a profile, which includes information about all of them (see Methods), allows the identification of homologs from other organisms. Examples of such specialized families include a group of aminoacetyltransferases and a distinct family of helix-turn-helix (HTH) DNA-binding proteins predicted to function as transcriptional regulators (Figure 4). The latter are of particular interest because at least one protein of this family (DR0171) appears to be associated with radiation resistance [5]. Potential involvement of these proteins in stress-response is emphasized by the fact that they show the closest similarity to the ArsR/AsnC/CadC family of winged-HTH domains, some of which regulate genes involved in heavy metal detoxification [11] or metal transport [33].

An intriguing possibility is that novel unique proteins could be involved in unknown mechanisms of DNA repair or stress response specific for *Deinococcus*. Indeed, about 720 proteins encoded in the genome have no detectable homologs in the current database. Most of these are predicted membrane or non-globular proteins; these tend to evolve rapidly, which hampers detection of sequence similarity. Nevertheless, we identified 26 families with at least two members each that, at this stage, appear to be *Deinococcus*-specific (Table 2). Some of these families possess conserved sequence and structure features that are reminiscent of some well-characterized proteins. For example, the DR2457-like and DR2241-like families contain pairs of conserved cysteines that resemble those present in certain oxidoreductases or Zn-ribbon nucleic acid-binding proteins. Several other uncharacterized globular proteins in *Deinococcus* (e.g. DR1088, DR1486) also contain such cysteine pairs, which suggests metal-binding and perhaps nucleic-acid-binding.

Concluding remarks

Direct connections between protein family expansion observed in a particular organism, in this case the radio- and desiccation-resistant bacterium *Deinococcus radiodurans*, and its lifestyle should be considered tenuous pending the requisite experimental studies. Nevertheless, it appears that at least in the case of the two most outstanding family expansions in *D. radiodurans*, those of the Nudix and DinB/YfiT families, such an association is extremely likely.

A naïve interpretation of protein family expansion in any organism would hold that all the paralogs result from gene duplication within this particular genome. The feasibility of such duplication in *Deinococcus* is supported by the demonstration of gene amplification under positive selection [31]. A more detailed examination of the protein families shows, however, that each of them is a heterogeneous collection of relatively recent duplications and horizontal transfers. The natural competence of *Deinococcus* is compatible with a significant amount of lateral gene acquisition.

Systematic experimental analysis is required to corroborate or refute the hypothesis that expansion of gene families in general is an adaptation to the specific requirements of an organism's niche. If confirmed on multiple occasions, this principle may have considerable heuristic value in future genome analyses.

Acknowledgements

This work was funded by grants DE-FG02-98ER62583, DE-FG02-97ER62492, and DE-FG07-97ER20293 from the U.S. Department of Energy; and grant 5R01-GM39933-09 from the National Institutes of Health.

References

1. Alexander D. et al.: Increased tolerance to two oomycete pathogens in transgenic tobacco expressing pathogenesis-related protein 1a, *Proc. Natl. Acad. Sci. USA* **90** (1993): 7327–7331.
2. Altschul S.F. et al.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* **25** (1997): 3389–3402.
3. Antelmann H. et al.: First steps from a two-dimensional protein index towards a response- regulation map for Bacillus subtilis, *Electrophoresis* **18** (1997): 1451–1463.
4. Azeddoug H. and Reysset G.: Cloning and sequencing of a chromosomal fragment from Clostridium acetobutylicum strain ABKn8 conferring chemical-damaging agents and UV resistance to *E. coli* recA strains, *Curr. Microbiol.* **29** (1994): 229–235.
5. Battista J.R.: Against all odds: the survival strategies of Deinococcus radiodurans, *Annu. Rev. Microbiol.* **51** (1997): 203–224.
6. Battista J.R., Earl A.M. and Park M.J.: Why is deinococcus radiodurans so resistant to ionizing radiation? *Trends Microbiol.* **7** (1999): 362–365.
7. Bessman M.J., Frick D.N. and O'Handley S.F.: The MutT proteins or 'Nudix' hydrolases, a family of versatile, widely distributed, 'housecleaning' enzymes, *J. Biol. Chem.* **271** (1996): 25059–25062.
8. Cheo D.L., Bayles K.W. and Yasbin R.E.: Cloning and characterization of DNA damage-inducible promoter regions from Bacillus subtilis, *J. Bacteriol.* **173** (1991): 1696–1703.
9. Chervitz S.A. et al.: Comparison of the complete protein sets of worm and yeast: orthology and divergence, *Science* **282** (1998): 2022–2028.
10. Cole S.T. et al.: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature* **393** (1998): 537–544.
11. Diorio C. et al.: *An Escherichia coli* chromosomal ars operon homolog is functional in arsenic detoxification and is conserved in gram-negative bacteria, *J. Bacteriol.* **177** (1995): 2050–2056.
12. Fernandez C. et al.: NMR solution structure of the pathogenesis-related protein P14a, *J. Mol. Biol.* **266** (1997): 576–593.
13. Fujimori A. et al.: Cloning and mapping of Np95 gene which encodes a novel nuclear protein associated with cell proliferation, *Mamm. Genome* **9** (1998): 1032–1035.
14. Grant C.E., Bain G. and Tsang A.: The molecular basis for alternative splicing of the CABP1 transcripts in Dictyostelium discoideum, *Nucleic Acids Res.* **18** (1990): 5457–5463.
15. Hecht H.J. et al.: The metal-ion-free oxidoreductase from Streptomyces aureofaciens has an alpha/beta hydrolase fold, *Nat. Struct. Biol.* **1** (1994): 532–537.
16. Hiom K. and Sedgwick S.G.: Cloning and structural characterization of the mcrA locus of Escherichia coli, *J. Bacteriol.* **173** (1991): 7368–7373.
17. Jobling M.G. and Ritchie D.A.: The nucleotide sequence of a plasmid determinant for resistance to tellurium anions, *Gene* **66**: 245–258.
18. Koonin E.V.: A highly conserved sequence motif defining the family of MutT-related proteins from eubacteria, eukaryotes and viruses, *Nucleic Acids Res.* **21** (1993): 4847.
19. Kunst F. et al.: The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*, *Nature* **390** (1997): 249–256.
20. Lange C.C. et al.: Engineering a recombinant Deinococcus radiodurans for organopollutant degradation in radioactive mixed waste environments, *Nat Biotechnol* **16** (1998): 929–933.
21. Linthorst H.J. et al.: Tobacco and tomato PR proteins homologous to win and pro-hevein lack the 'hevein' domain, *Mol. Plant Microbe. Interact.* **4** (1991): 586–592.
22. Makarova K.S et al.: Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell, *Genome Res.* **9** (1999): 608–628.
23. Mattimore V. and Battista J.R.: Radioresistance of Deinococcus radiodurans: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation, *J. Bacteriol.* **178** (1996): 633–637.
24. Minton K.W. and Daly M.J.: A model for repair of radiation-induced DNA double-strand breaks in the extreme radiophile Deinococcus radiodurans, *Bioessays* **17** (1995): 457–464.
25. Myers C.R. and Myers J.M.: Outer membrane cytochromes of Shewanella putrefaciens MR-1: spectral analysis, and purification of the 83-kDa c-type cytochrome, *Biochim. Biophys. Acta* **1326** (1997): 307–318.
26. Raibaud A. et al.: Nucleotide sequence analysis reveals linked *N*-acetyl hydrolase, thioesterase, transport, and regulatory genes encoded by the bialaphos biosynthetic gene cluster of Streptomyces hygroscopicus, *J. Bacteriol.* **173** (1991): 4454–4463.
27. Rathbone D.A. et al.: Molecular analysis of the Rhodococcus sp. strain H1 her gene and characterization of its product, a heroin esterase, expressed in *Escherichia coli*, *Appl. Environ. Microbiol.* **63** (1997): 2062–2066.
28. Schaffer A.A. et al.: IMPALA: Matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices, *Bioinformatics* (2000) (in press).
29. Schlenk D.: Occurrence of flavin-containing monooxygenases in non-mammalian eukaryotic organisms, *Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol.* **121** (1998): 185–195.
30. Seledtsov I.A., Vul'f Iu. I. and Makarova K.S.: Multiple alignment of biopolymer sequences, based on the search for statistically significant common segments, *Mol. Biol. (Mosk)* **29** (1995): 1023–1039.
31. Smith M.D. et al.: Duplication insertion of drug resistance determinants in the radioresistant bacterium Deinococcus radiodurans, *J. Bacteriol.* **170** (1988): 2126–2135.
32. Stephens R.S. et al.: Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*, *Science* **282** (1998): 754–759.
33. Thelwell C., Robinson N.J. and Turner-Cavet J.S.: An SmtB-like repressor from Synechocystis PCC 6803 regulates a zinc exporter, *Proc. Natl. Acad. Sci. USA* **95** (1998): 10728–10733.

34. Udupa K.S. et al.: Novel ionizing radiation-sensitive mutants of Deinococcus radiodurans, *J. Bacteriol.* **176** (1994): 7439–7446.
35. Walker D.R. and E.V. Koonin: SEALS: a system for easy analysis of lots of sequences, *Ismb* **5** (1997): 333–339.
36. White O. et al.: Genome Sequence of the Radioresistant Bacterium Deinococcus radiodurans R1, *Science* **286** (1999): 1571–1577.
37. Wolf Y.I. et al.: Distribution of protein folds in the three superkingdoms of life, *Genome Res.* **9** (1999): 17–26.
38. Wootton J.C.: Non-globular domains in protein sequences: automated segmentation using complexity measures, *Comput. Chem.* **18** (1994): 269–285.
39. Yamamoto M et al.: Cloning and sequencing of a 35.7 kb in the 70 degree-73 degree region of the Bacillus subtilis genome reveal genes for a new two-component system, three spore germination proteins, an iron uptake system and a general stress response protein, *Gene* **194** (1997): 191–199.