BMC
Bioinformatics

RESEARCH ARTICLE

Open Access

LUD, a new protein domain associated with lactate utilization

William C Hwang[1,2*], Constantina Bakolitsa[2], Marco Punta[3], Penelope C Coggill[3], Alex Bateman[3,4],
Herbert L Axelrod[1,5], Neil D Rawlings[3], Mayya Sedova[1,2], Scott N Peterson[2], Ruth Y Eberhardt[3,4], L Aravind[6],
Jaime Pascual[7] and Adam Godzik[1,2,8,9*]

Abstract

Background: A novel highly conserved protein domain, DUF162 [Pfam: PF02589], can be mapped to two proteins: LutB and LutC. Both proteins are encoded by a highly conserved LutABC operon, which has been implicated in lactate utilization in bacteria. Based on our analysis of its sequence, structure, and recent experimental evidence reported by other groups, we hereby redefine DUF162 as the LUD domain family.

Results: JCSG solved the first crystal structure [PDB:2G40] from the LUD domain family: LutC protein, encoded by ORF *DR_1909*, of *Deinococcus radiodurans*. LutC shares features with domains in the functionally diverse ISOCOT superfamily. We have observed that the LUD domain has an increased abundance in the human gut microbiome.

Conclusions: We propose a model for the substrate and cofactor binding and regulation in LUD domain. The significance of LUD-containing proteins in the human gut microbiome, and the implication of lactate metabolism in the radiation-resistance of *Deinococcus radiodurans* are discussed.

Keywords: LUD, DUF162, LutB, LutC, Domain of unknown function, *Deinococcus radiodurans*

Background

We are now in an era when we can routinely sequence the complete genomes of microbes and rapidly identify their protein coding complements. The sequences of millions of proteins are now known. Despite this wealth of information we are still far from understanding how all of these proteins operate to give rise to a living organism. At present, in a consistent percentage of proteins the predicted function remains unknown [1,2]. From our analysis of 23 million proteins in the Pfam sequence database (Pfam release 27.0), 20% of them have no associated Pfam domain [3] and more are classified into DUF (Domains of Unknown Function) families [2]. This uncharacterized set of proteins potentially contains novel biological systems. Therefore, it is important to uncover these hidden functions through analysis of protein sequence, protein

structure, and finally through directed experimental analyses [4-7].

There have been various attempts to classify the multitude of protein sequences into families to facilitate an improved understanding of the functional repertoire of proteins. In addition, there is a growing number of protein families defined for which no protein has ever been previously experimentally characterized. These families have been called DUFs [2] or Uncharacterized Protein Families (UPFs) [8]. The Pfam database contains one of the largest collections of such families with over 4,000 defined to date.

A novel domain, DUF162 [Pfam: PF02589] [COG: COG1556] [eggNOG: COG1556] [CDD: 224473], was found predominantly in Bacteria, and to a lesser extent in Archaea and Eukaryota. Recently, one protein (YvbY from *Bacillus subtilis*) in this DUF162 family was identified as lactate-utilization protein C (LutC), which was homologous to the YkgG protein in *E. coli*, hinting at a possible role in lactate utilization [9,10]. Indeed, DUF162 domain is a constituent domain of two proteins (LutB and LutC) encoded by the conserved LutABC operon in

* Correspondence: wchwang@sanfordburnham.org; adam@sanfordburnham.org
[1]Joint Center for Structural Genomics, La Jolla, CA 92037, USA
[2]Sanford Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
Full list of author information is available at the end of the article

bacteria. This operon has been linked to lactate utilization [9,10] and is implicated in the oxidative conversion of L-lactate into pyruvate [9]. Based on our analysis of its sequence, structure, and recent experimental evidence reported by other groups, we hereby redefine DUF162 domain as the LUD domain.

Here, we report the first crystal structure [PDB: 2G40] of the LUD domain family: LutC protein (encoded by ORF DR_1909) from *Deinococcus radiodurans* [11,12] at 1.70 Å resolution. We propose a model for the substrate and cofactor binding and regulation.

Results and discussion

LUD domain structure

The Joint Center for Structural Genomics (JCSG) determined the first crystal structure of the LUD domain family: LutC protein from *Deinococcus radiodurans*. The LutC protein structure is a mixed alpha-helix and beta-sheet protein (Figure 1). The protein core is made up of two orthogonal beta-sheets, each consisting of four beta-strands. The alpha-helices are packed against the two solvent-facing surfaces of the beta-sheets as well as against the side openings of the protein core.
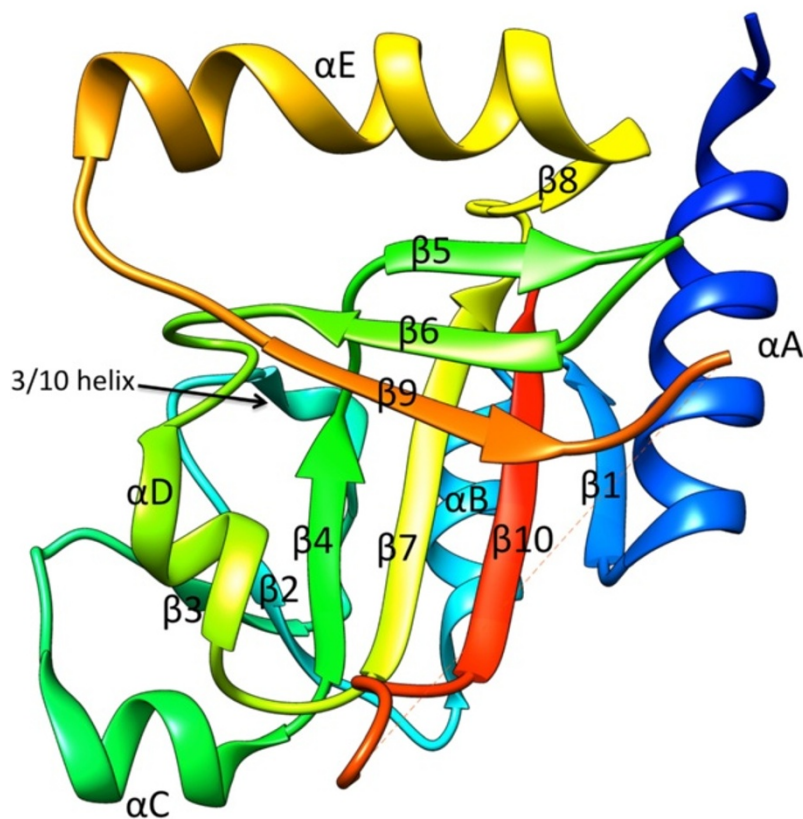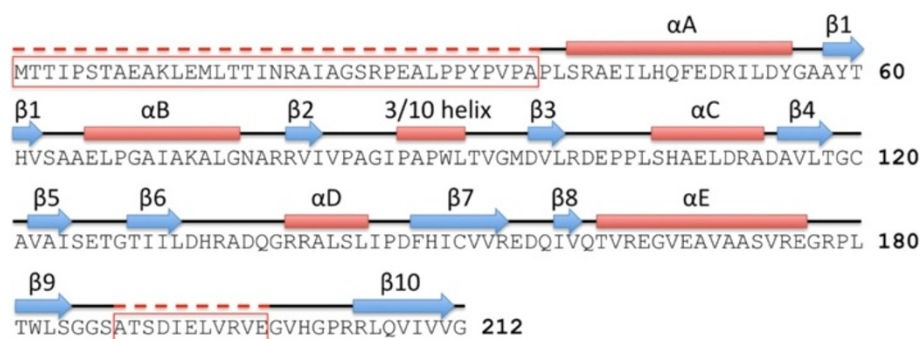

Figure 1 Structure of LutC protein from *Deinococcus radiodurans*. The protein structure is shown in cartoon style and colored in rainbow format (N-terminus being blue and C-terminus red). The dashed line in the figure represents a break in the protein polypeptide chain as a result of missing electron density in the protein structure.

Some regions of the LutC protein sequence are highly conserved as assessed by ConSurf. The conserved areas are concentrated on one side of the structure and form a groove about 20 Å in length (Figure 2), which might be functionally important. LutC protein appears to be dimeric, with a buried surface of 1721 Å^2 at the dimer-interface. The highly conserved area coincides with parts of the dimer interface.

Structural alignment with other protein structures present in the Protein Data Bank, using the program DALI [13,14], suggests LutC protein is structurally akin to proteins found in the ISOCOT superfamily [15]. This is consistent with its classification in SCOP [16] as part of the NagB/RpiA/CoA transferase-like fold and superfamily. The ISOCOT superfamily is known to comprise proteins of diverse functions including sugar isomerases, translation factor eIF2B, ligand-binding domains of the DeoR-family transcription factors, acetyl-CoA transferases, and methenyltetrahydrofolate synthetase [15].

Domain organization

While predominantly found to exist by itself, LUD domain is also frequently found together with domains such as the 4Fe-4S dicluster domain Fer4_8 [Pfam: PF13183], DUF3390 [Pfam: PF11870], and cysteine-rich iron-sulfur binding cluster domain CCG [Pfam: PF02754] [17]. Figure 3 shows the most common domain architectures featuring the LUD domain according to Pfam release 27.0.

LUD domain-containing proteins encoded by the highly conserved LutABC Operon

LUD domain is a protein domain of approximately 160 residues in length (Figure 4, and Additional file 1). It is found in two proteins encoded by the highly conserved LutABC operon (Figures 5 and 6), which appears in a wide variety of Gram-positive and Gram-negative bacteria [9]. The LutABC operon was found to be important for growth and biofilm formation in *Bacillus subtilis* [9]. The LUD domain is found in both LutB and LutC
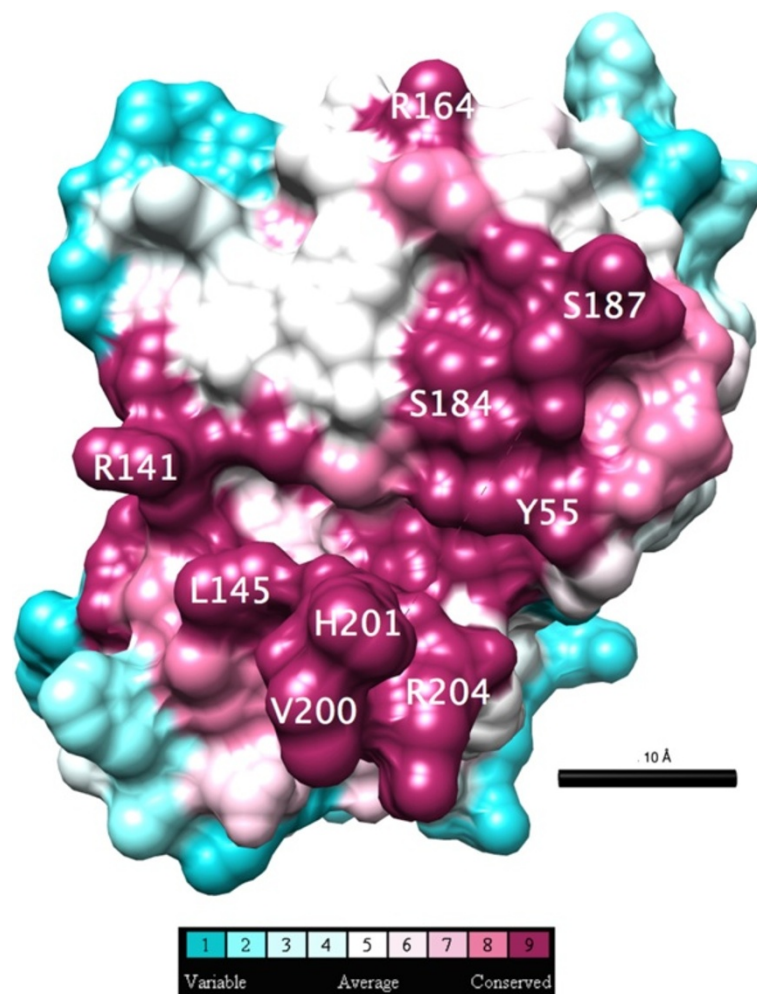

Figure 2 Conservation of residues in the LUD domain family projected on the three dimensional structure of the LutC protein from *Deinococcus radiodurans.*
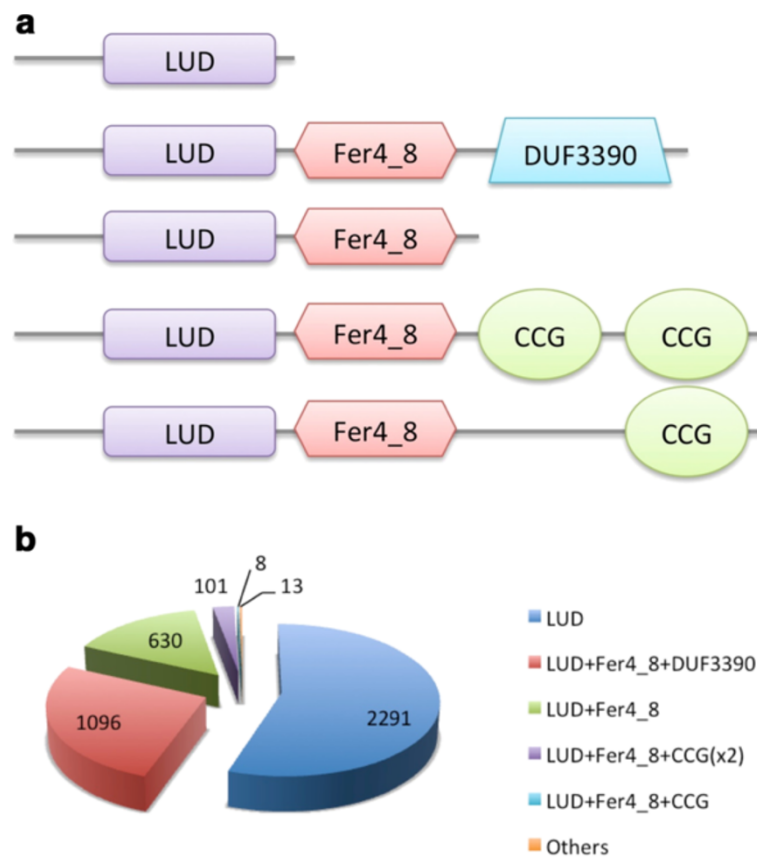
Figure 3 Domain organization of LUD domain. a. The most common domain organizations of LUD domain are shown. While predominantly found to exist by itself, LUD domain is also frequently found together with domains such as 4Fe-4S dicluster domain Fer4_8 [Pfam:PF13183], DUF3390 [Pfam:PF11870], cysteine-rich iron-sulfur binding cluster domain CCG [Pfam:PF02754]. **b**. Pie chart showing the frequency of common LUD domain organizations in known proteins.

proteins encoded by the LutABC operon. In the vast majority of cases, the LUD domain is the only constituent domain of LutC proteins, whereas in LutB proteins it is often associated with protein families Fer4_8, CCG, or DUF3390 (Figure 5). Indeed, in Pfam release 27.0 there is just one instance of LutB protein being made of DUF162 alone, which occurs in *Deinococcus radiodurans* (Figure 6). However, searching the section of DNA in *Deinococcus radiodurans* from the start of *lutB* to the start of *lutC* finds a frame-shift and a copy of DUF3390 on the opposite strand, though no apparent Fer4_8, implying possible poor quality sequencing in this region. Finally, LutA protein is most often made of two copies of CCG domains. Both Fer4_8 and CCG domains are likely iron-sulfur cluster binding domains [17]. LutA protein is a putative iron-sulfur heterodisulfide reductase; LutB protein a putative iron-sulfur oxidoreductase; LutC protein a putative subunit of an iron-sulfur protein. Together, they are thought to mediate the oxidation of lactate via a cytochrome-like electron transfer chain, though the precise roles played by LutABC remain unclear [9].

Presence in gut microbiome

It is worth noting that LUD domain has an increased abundance in gut microbiome. From our comparative genomics analysis of the metahit human gut microbiome of 124 human subjects (unpublished result, data not shown), the average ratio of number of homologs from the metahit human gut microbiome versus those found in UniProtKB is about 0.07. The ratio for LUD domain is ten times higher at 0.72, suggesting it plays a significant role in the gut microbiome, possibly related to its role in anaerobic metabolism. Interestingly, lactic acid bacteria (LAB) are being used as probiotics [18]. Lactate metabolism is integral to human health and host-pathogen interactions. Pathogenic bacteria have been shown to decrease local pH in hosts, through an increase in lactate production, so as to facilitate the release of iron from host transferrin [19]. In other species, acquisition of lactate is necessary for bacteremia [20] and colonization [21]. Lactate is also a potent signaling molecule in inflammatory pathways and has emerged as a critical regulator of cancer development, maintenance and metastasis [22]. By modulating lactate concentrations in the host's environment
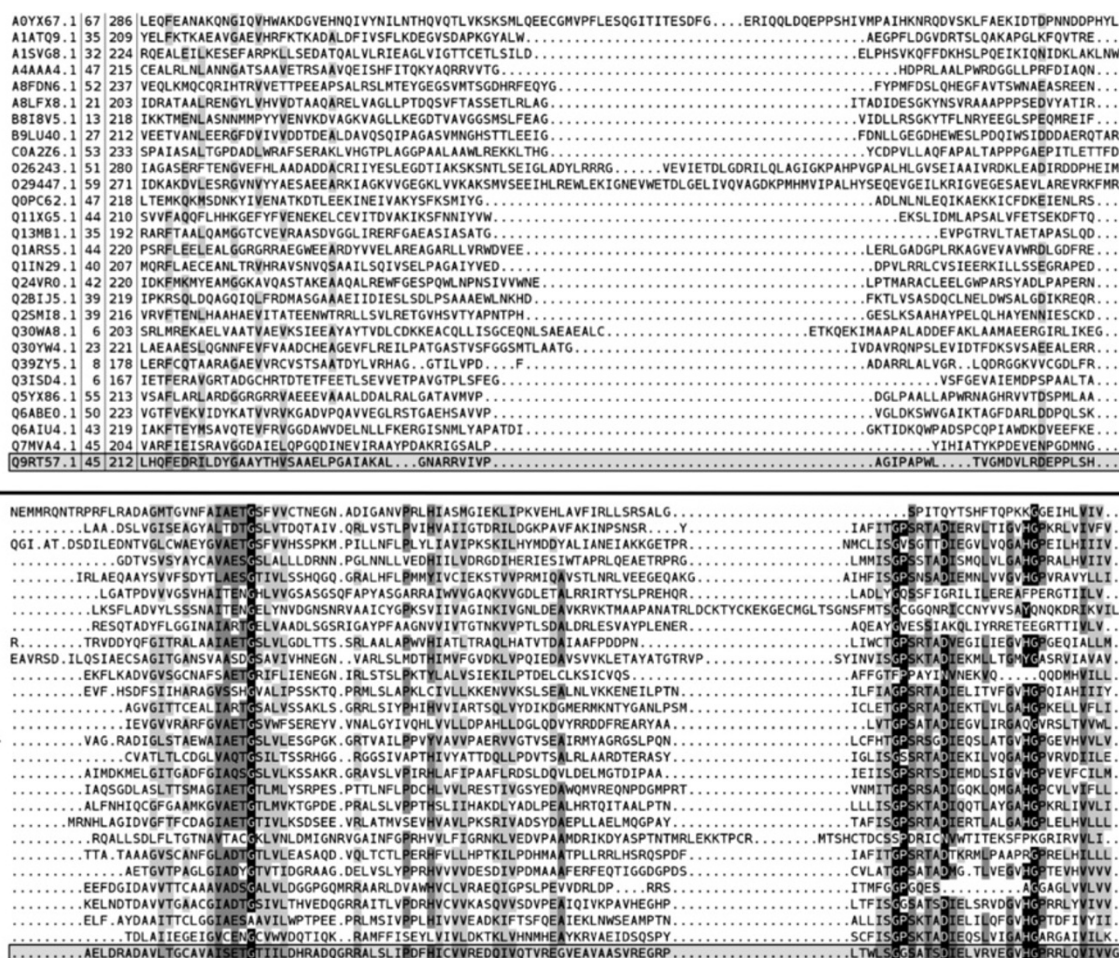
Figure 4 Alignment with representative sequences of LUD family (Pfam DUF162-PF02589). a. N-terminal part of the alignment. **b**. C-terminal part of the alignment. Shades of grey reflect average similarity.
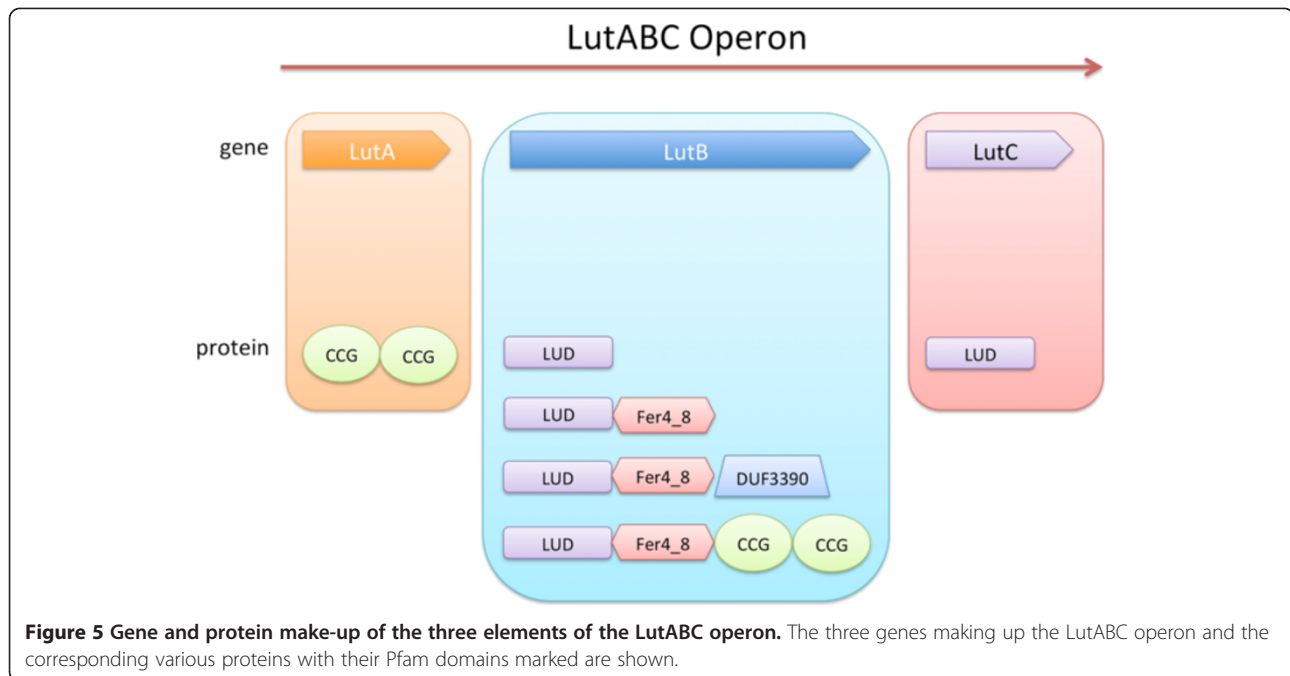
through LUD domains and other lactate-related pathways, lactobacilli could thus influence the outcomes of both pathogenicity and disease [23].

Model for LUD domain substrate-cofactor binding and regulation

Inspection of the LutC protein dimer structure identified a highly conserved cavity (lined by residues Y55, H201, and R204) near the dimer interface. We proposed this cavity to be the putative active site (Figure 7), where the oxidative conversion of lactate into pyruvate occurs [9], based on the following observations: First, the residues surrounding this cavity are highly conserved, suggesting they are functionally important. Second, this cavity is large enough to accommodate both NAD + and lactate, hypothetical cofactor and substrate (Figure 8). NAD is among the top 5 possible ligands for LutC dimer as predicted by IsoCleft [24]. Top ligand predicted by Isocleft predicted was NDP (NADPH). Third, in the

docking model the highly conserved H201 in LutC protein is located close to the substrate-cofactor reaction site and could hence serve as the catalytic histidine. Fourth, the 11-residue disordered loop (between S187 and G199) near this cavity could function as a substrate binding regulator, analogous to the role played by the disordered loop in the active site of lactate dehydrogenase (LDH), which converts pyruvate to lactate [25]. Taken together, it is likely that this pocket is indeed the active site.

Another moderately conserved cavity lined by residues R155, C120, and D137 (Figure 7), roughly coincides with the ISOCOT superfamily primary binding site. Docking of NAD to this shallow and small cavity leaves it not fully embedded and partially exposed. Thus, it is unlikely to form the active site. Nevertheless, this cavity could bind smaller molecules and is a good candidate for allosteric regulation. Allosteric regulation has been reported for certain proteins of the ISOCOT superfamily [26,27].
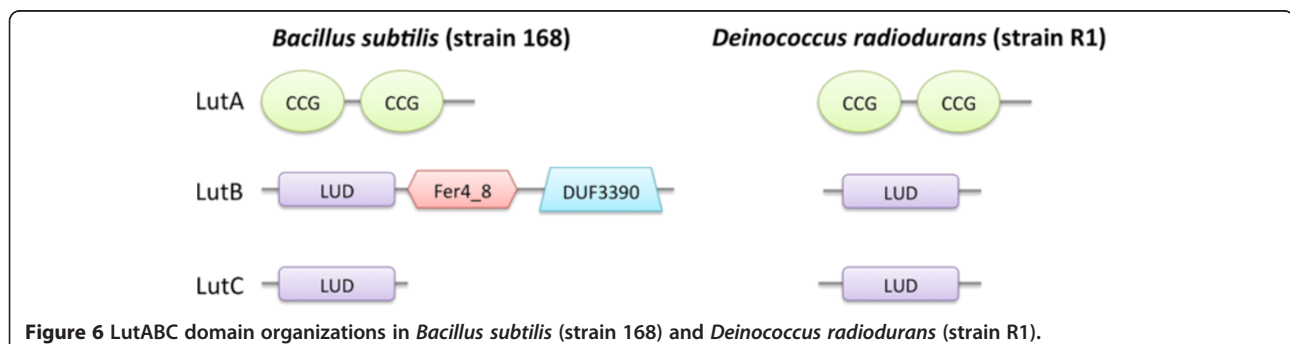
Figure 5 Gene and protein make-up of the three elements of the LutABC operon. The three genes making up the LutABC operon and the corresponding various proteins with their Pfam domains marked are shown.

Functional implications in Deinococcus radiodurans

The LutC protein was selected as a target because of the interest in *Deinococcus radiodurans* by JCSG. *Deinococcus radiodurans* is the most radiation-resistant bacterium known to date [12]. It can survive 4000 Gray (Gy) of irradiation, a dose hundreds of times greater than that considered lethal for most organisms. How it accomplishes such a remarkable feat remains enigmatic. A study examining global gene expression following ionizing radiation exposure and desiccation allowed a dissection of the response to double strand breaks (induced by both ionizing radiation and desiccation) and oxidative stress associated with reactive oxygen species (ROS). LutC protein was not induced in either treatment but was constitutively expressed [11]. Free radicals, in particular ROS, generated when cells are exposed to ionizing radiation, are cytotoxic. The unpaired electrons of free radicals render them highly reactive with biological molecules. Unsaturated fatty acids present in the membrane are particularly susceptible to

free radicals. Furthermore, free radical-oxygen will deplete oxygen in the cytosol and abolish aerobic metabolism. Anaerobic lactate metabolism can be an indispensable alternative energy source. Moreover, lactate can function as a scavenger of free radicals [28]. Thus, lactate utilization may contribute to the radiation-resistance of the *Deinococcus radiodurans*. As the LutC protein from *Deinococcus radiodurans* represents a prototypical LUD domain in lactate utilization, it could be contributing towards radiation-resistance in this bacterium.

Conclusions

Lactate metabolism is integral to human health, and may play a role in the radiation resistance in *Deinococcus radiodurans*. The LUD domain is a highly conserved protein domain that has recently been identified to play a role in lactate metabolism. In this report, we described the crystal structure of the *Deinococcus radiodurans* LutC protein, the first for a member of the LUD domain
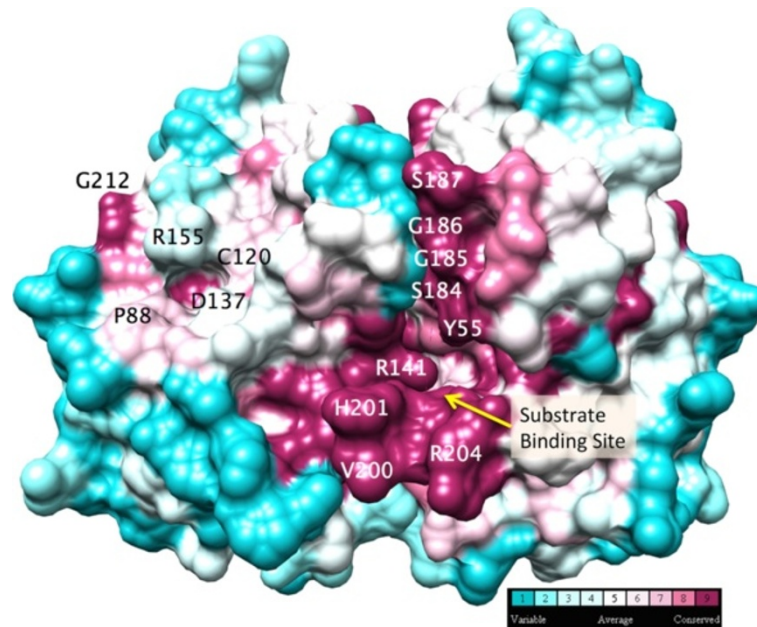

Figure 6 LutABC domain organizations in *Bacillus subtilis* (strain 168) and *Deinococcus radiodurans* (strain R1).

Figure 7 The highly conserved cavity near the dimer interface as the possible active site.

family. Using sequence and structure analysis, we proposed a model for the substrate and cofactor binding and regulation in LUD domains. We also analyzed possible implications for radiation resistance in *Deinococcus radiodurans*. Further experimental characterization will be needed to test these hypotheses.

Methods

Sequence analysis

Alignment of representative sequences of LUD family (Pfam DUF162-PF02589) was built by taking the SEED sequences of the family, reducing redundancy at 40% sequence identity and finally realigning the remaining

Figure 8 Docking of NADH to the hypothetical active site near the dimer interface. The monomers are colored in cyan and brown, respectively. Highly conserved residues, Y55, H201, R204 nearby are highlighted in green and labeled.

sequences plus the sequence of 2G40 (UniProtKB id: Q9RT57) with ClustalW [29]. For better visualisation the alignment has been split in two parts (a) and (b). In (a) we show the N-terminal part of the alignment that continues toward the C-terminus in (b). Shades of grey reflect average similarity as calculated from the BLOSUM62 amino acid substitution matrix (black most conserved, white least conserved). Dashes (-) represent deletions, dots (.) represent insertions and lower case letters represent inserted residues. For each sequence, we report the UniProtKB id (e.g. F9YU00), the position along the protein sequence of first and last residue in the alignment (in the case of Q9RT57, for example, aligned residues range from 45 to 212) and, finally, the amino acid sequence. 2G40 (Q9RT57) sequence is highlighted by a shaded box. The alignment is visualized with Belvu [30] (sonnhammer.sbc.su.se/Belvu.html). More sequence and domain analysis for the LUD domain family can be found in the Additional file 1.

Structure determination

Structure determination of LutC protein was carried out by the JCSG high-throughput structural biology pipeline [31]. Diffraction data were collected at Stanford Synchrotron Radiation Lightsource (SSRL) beamline 1-5. The crystal structure was determined by MAD phasing using seleno-methionine-derivatized protein. The structure was validated using the JCSG Quality Control server (http://smb.slac.stanford.edu/jcsg/QC). Experimental details as well as structural and refinement statistics can be found in the Additional file 2.

Atomic coordinates and experimental structure factors have been deposited into the Protein Data Bank (http://www.rcsb.org) with PDB ID: 2G40.

Structure analysis

LutC protein dimer was generated by symmetry-related positions in Pymol [32]. Dimer interface was assessed by PISA [33]. Conservation of LutC protein amino acid residues was assessed by ConSurf [34], which obtained close homologous sequences through BLAST. Molecular docking was performed with MVD [35] using default parameters. Structure graphics were prepared in Chimera [36].

Additional files

Additional file 1: Sequence and Domain Analysis. This section contains additional sequence and domain analysis of LUD domain family.

Additional file 2: Experimental Details [PDB:2G40]. This section contains experimental details as well as structural and refinement statistics.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Joint Center for Structural Genomics, La Jolla, CA 92037, USA. [2]Sanford Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA. [3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. [4]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK. [5]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [6]National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894, USA. [7]Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA. [8]Center for Research in Biological Systems, University of California, San Diego, La Jolla, CA 92093-0446, USA. [9]Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah 21589, Kingdom of Saudi Arabia.

References
1. Jaroszewski L, Li Z, Krishna SS, Bakolitsa C, Wooley J, et al: **Exploration of uncharted regions of the protein universe.** *PLoS Biol* 2009, **7**:e1000205.
2. Bateman A, Coggill P, Finn RD: **DUFs: families in search of function.** *Acta Crystallogr Sect F: Struct Biol Cryst Commun* 2010, **66**:1148–1152.
3. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al: **The Pfam protein families database.** *Nucleic Acids Res* 2012, **40**:D290–D301.
4. Roberts RJ: **Identifying protein function–a call for community action.** *PLoS Biol* 2004, **2**:E42.
5. Roberts RJ, Chang YC, Hu Z, Rachlin JN, Anton BP, et al: **COMBREX: a project to accelerate the functional annotation of prokaryotic genomes.** *Nucleic Acids Res* 2011, **39**:D11–D14.
6. Galperin MY, Koonin EV: **From complete genome sequence to 'complete' understanding?** *Trends Biotechnol* 2010, **28**:398–406.

7. Hanson AD, Pribat A, Waller JC, De Crecy-Lagard V: 'Unknown' proteins and 'orphan' enzymes: the missing half of the engineering parts list–and how to find it. *Biochem J* 2010, **425**:1–11.
8. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, et al: The Pfam protein families database. *Nucleic Acids Res* 2004, **32**:D138–D141.
9. Chai Y, Kolter R, Losick R: A widely conserved gene cluster required for lactate utilization in Bacillus subtilis and its involvement in biofilm formation. *J Bacteriol* 2009, **191**:2423–2430.
10. Smaldone GT, Antelmann H, Gaballa A, Helmann JD: The FsrA sRNA and FbpB protein mediate the iron-dependent induction of the Bacillus subtilis lutABC iron-sulfur-containing oxidases. *J Bacteriol* 2012, **194**:2586–2593.
11. Schmid AK, Howell HA, Battista JR, Peterson SN, Lidstrom ME: Global transcriptional and proteomic analysis of the Sig1 heat shock regulon of Deinococcus radiodurans. *J Bacteriol* 2005, **187**:3339–3351.
12. Makarova KS, Aravind L, Wolf YI, Tatusov RL, Minton KW, et al: Genome of the extremely radiation-resistant bacterium Deinococcus radiodurans viewed from the perspective of comparative genomics. *Microbiol Mol Biol Rev* 2001, **65**:44–79.
13. Holm L, Rosenstrom P: Dali server: conservation mapping in 3D. *Nucleic Acids Res* 2010, **38**:W545–W549.
14. Hasegawa H, Holm L: Advances and pitfalls of protein structural alignment. *Curr Opin Struct Biol* 2009, **19**:341–348.
15. Anantharaman V, Aravind L: Diversification of catalytic activities and ligand interactions in the protein fold shared by the sugar isomerases, eIF2B, DeoR transcription factors, acyl-CoA transferases and methenyltetrahydrofolate synthetase. *J Mol Biol* 2006, **356**:823–842.
16. Murzin AG, Brenner SE, Hubbard T, Chothia C: SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995, **247**:536–540.
17. Hamann N, Mander GJ, Shokes JE, Scott RA, Bennati M, et al: A cysteine-rich CCG domain contains a novel [4Fe-4S] cluster binding motif as deduced from studies with subunit B of heterodisulfide reductase from Methanothermobacter marburgensis. *Biochemistry* 2007, **46**:12875–12885.
18. Ljungh A, Wadstrom T: Lactic acid bacteria as probiotics. *Curr Issues Intest Microbiol* 2006, **7**:73–89.
19. Friedman DB, Stauff DL, Pishchany G, Whitwell CW, Torres VJ, et al: Staphylococcus aureus redirects central metabolism to increase iron availability. *PLoS Pathog* 2006, **2**:e87.
20. Herbert MA, Hayes S, Deadman ME, Tang CM, Hood DW, et al: Signature tagged Mutagenesis of Haemophilus influenzae identifies genes required for in vivo survival. *Microb Pathog* 2002, **33**:211–223.
21. Exley RM, Wu H, Shaw J, Schneider MC, Smith H, et al: Lactate acquisition promotes successful colonization of the murine genital tract by Neisseria gonorrhoeae. *Infect Immun* 2007, **75**:1318–1324.
22. Doherty JR, Cleveland JL: Targeting lactate metabolism for cancer therapeutics. *J Clin Invest* 2013, **123**:3685–3692.
23. Maudsdotter L, Jonsson H, Roos S, Jonsson AB: Lactobacilli reduce cell cytotoxicity caused by Streptococcus pyogenes by producing lactic acid that degrades the toxic component lipoteichoic acid. *Antimicrob Agents Chemother* 2011, **55**:1622–1628.
24. Najmanovich R, Kurbatova N, Thornton J: Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites. *Bioinformatics* 2008, **24**:i105–i111.
25. Clarke AR, Wigley DB, Chia WN, Barstow D, Atkinson T, et al: Site-directed mutagenesis reveals role of mobile arginine residue in lactate dehydrogenase catalysis. *Nature* 1986, **324**:699–702.
26. Horjales E, Altamirano MM, Calcagno ML, Garratt RC, Oliva G: The allosteric transition of glucosamine-6-phosphate deaminase: the structure of the T state at 2.3 A resolution. *Structure* 1999, **7**:527–537.
27. Rudino-Pinera E, Morales-Arrieta S, Rojas-Trejo SP, Horjales E: Structural flexibility, an essential component of the allosteric activation in Escherichia coli glucosamine-6-phosphate deaminase. *Acta Crystallogr D Biol Crystallogr* 2002, **58**:10–20.
28. Groussard C, Morel I, Chevanne M, Monnier M, Cillard J, et al: Free radical scavenging and antioxidant effects of lactate ion: an in vitro study. *J Appl Physiol* 2000, **89**:169–175.
29. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, et al: Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 2003, **31**:3497–3500.
30. Sonnhammer EL, Hollich V: Scoredist: a simple and robust protein sequence distance estimator. *BMC Bioinforma* 2005, **6**:108.
31. Elsliger MA, Deacon AM, Godzik A, Lesley SA, Wooley J, et al: The JCSG high-throughput structural biology pipeline. *Acta Crystallogr Sect F: Struct Biol Cryst Commun* 2010, **66**:1137–1142.
32. *The PyMOL molecular graphics system.* Version 12r3preth edition. Schrödinger, LLC. http://pymol.sourceforge.net/faq.html#CITE.
33. Krissinel E, Henrick K: Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 2007, **372**:774–797.
34. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N: ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 2010, **38**:W529–W533.
35. Thomsen R, Christensen MH: MolDock: a new technique for high-accuracy molecular docking. *J Med Chem* 2006, **49**:3315–3321.
36. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al: UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* 2004, **25**:1605–1612.