

## Sequence analysis

***Haloferax volcanii* PitA: an example of functional interaction between the Pfam chlorite dismutase and antibiotic biosynthesis monooxygenase families?**Elizabeth Bab-Dinitz<sup>1</sup>, Hagay Shmueli<sup>1</sup>, Julie Maupin-Furlow<sup>2</sup>, Jerry Eichler<sup>1</sup> and Boaz Shaanan<sup>1,\*</sup><sup>1</sup>Department of Life Sciences, Ben Gurion University, Beersheva 84105, Israel and <sup>2</sup>Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611-0700, USA

Received on December 11, 2005; revised on and accepted on January 4, 2006

Advance Access publication January 10, 2006

Associate Editor: Chris Stoeckert

**ABSTRACT**

A curious fusion between chlorite dismutase-like and antibiotic biosynthesis monooxygenase-like domains within a single open reading frame has been revealed by both sequence homology and structural modeling in *Haloferax volcanii* PitA and its homologues in other halophilic archaea. While this fusion may reflect an environmental adaptation to life in hypersaline environments and hence one specific to haloarchaea, PitA and its homologues may represent a paradigm of biologically-relevant interplay between these two distinct activities in accordance with the Rosetta Stone approach.

**Contact:** bshaanan@bgu.ac.il**INTRODUCTION**

In the course of structural and functional study of poorly conserved open reading frames from halophilic archaea, an open reading frame (ORF) encoding a unique protein comprising two domains was identified in *Haloferax volcanii*. The encoded protein, which we name PitA, is predicted by the Pfam (Bateman *et al.*, 2004) database to possess both chlorite dismutase-related and antibiotic biosynthesis monooxygenase (ABM)-related activities.

Chlorite dismutases catalyze the disproportionation of poisonous chlorite into chloride and oxygen via a heme-dependent mechanism. Chlorite dismutases characterized thus far are found in perchlorate-reducing organisms and are essential for the complete reduction of the environmental contaminant perchlorate into chloride and oxygen (Rikken *et al.*, 1996; van Ginkel *et al.*, 1996; Bender *et al.*, 2002; Danielsson Thorell *et al.*, 2002). To date, the ability of halophilic archaea to reduce perchlorate has only been reported for a single species, *Haloferax denitrificans* (Okeke *et al.*, 2002). The Pfam chlorite dismutase family, however, also includes proteins from organisms that have not been reported as able to reduce perchlorate, suggesting additional roles for family members. Indeed, the Pfam chlorite dismutase family member *Thermus thermophilus* TT1485 was recently shown to act as only a weak chlorite dismutase yet could generate oxygen from hydrogen peroxide, an activity not observed with previously characterized chlorite dismutases (Ebihara *et al.*, 2005).

Biochemically characterized members of the Pfam ABM family include small monooxygenases involved in the biosynthesis of aromatic polyketide antibiotics (Shen and Hutchinson, 1993; Kendrew *et al.*, 1997; Rafanan *et al.*, 2001; Chung *et al.*, 2002), heme oxygenases (Wu *et al.*, 2005) and a quinol monooxygenase (Adams and Jia, 2005). While ABM family members reveal only moderate sequence homology and catalyze different chemical reactions, they share a high degree of structural similarity and thus probably a similar mechanism of action, as demonstrated by the exclusive use of molecular oxygen in the oxidation reactions they catalyze.

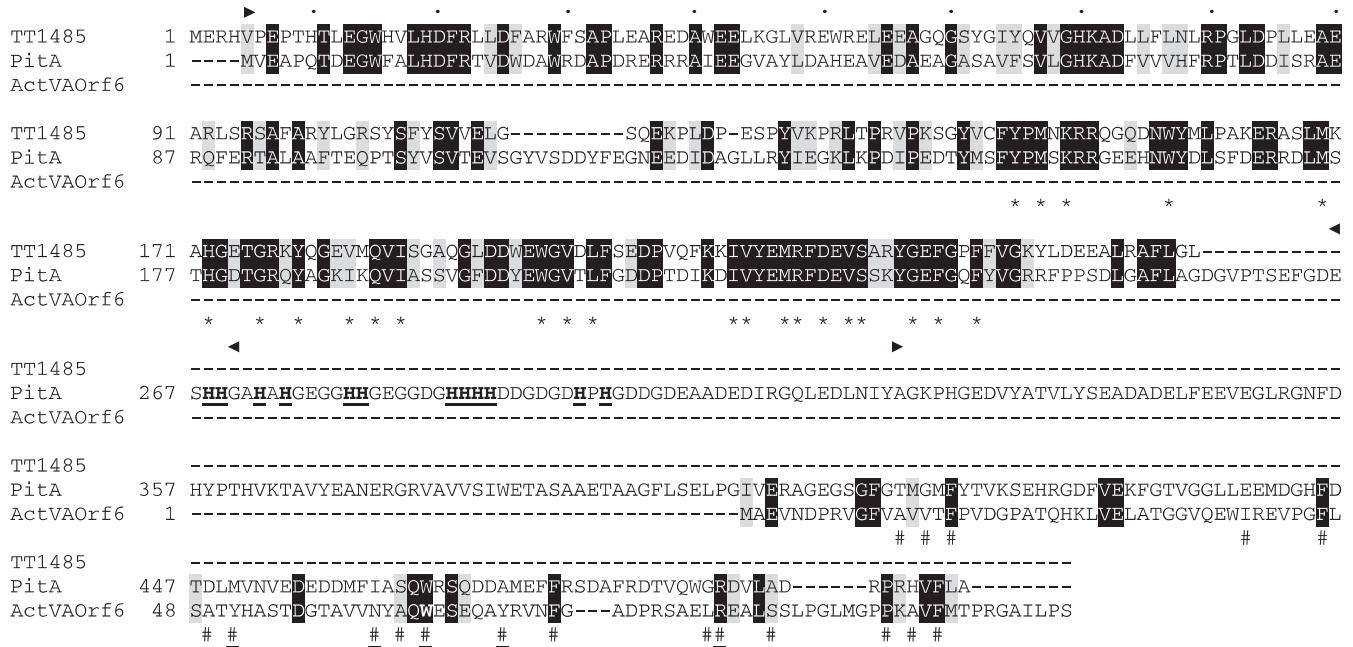
Database searches of known protein sequences revealed that the N- and C-terminal domains of *Hfx.volcanii* PitA, respectively assigned to the Pfam chlorite dismutase and ABM families, are fused on the same ORF only in this and other halophilic archaeal species, namely *Halobacterium* sp. NRC-1 (Ng *et al.*, 2000), *Haloarcula marismortui* (Baliga *et al.*, 2004) and *Natronomonas pharaonis* (Falb *et al.*, 2005). Like *Hfx.volcanii* PitA, the other three homologous haloarchaeal proteins, i.e. *Hbt.* sp. NRC-1 Vng2021c, *Har. marismortui* RmAC3100 and *Nmn.pharaonis* NP2262A possess a similar domain architecture of two highly conserved regions: a chlorite dismutase-like N-terminal domain and an ABM-like C-terminal domain. Furthermore, three of the homologs, i.e. PitA, Vng2021c and RmAC3100, also contain a histidine-rich region predicted to be structurally flexible between the two conserved domains.

Physical and functional interaction may exist between two separate proteins if both encoding genes are found as a fused species in one organism or more (Enright *et al.*, 1999; Marcotte *et al.*, 1999). In the following, we propose functional linkage between members of Pfam chlorite dismutase and ABM families in *Hfx.volcanii* PitA and its homologues in other halophilic archaea. The possible advantage of this functional relationship for microorganisms living in hypersaline environments is considered, along with the possibility that these activities are functionally linked in other organisms.

**METHODS****Protein purification**

PitA was isolated from *Hfx.volcanii* WR341 cells grown at 40°C (Mevarch and Werczberger, 1985). PitA was purified from the soluble cell fraction by affinity chromatography on a Co<sup>2+</sup> chelating column, equilibrated in

\*To whom correspondence should be addressed.



**Fig. 1.** Protein sequence alignment of *Hfx.volcanii* PitA with *T. thermophilus* TT1485 and *S. coelicolor* ActVA-Orf6. TT1485 was aligned with the N-terminal domain of PitA using ClustalW while ActVA-Orf6 was aligned with the C-terminal domain of PitA using Pfam and PSI-BLAST. Fully-conserved residues are shown against a black background while residues conserved in chemical character are shown against a grey background. The arrowheads indicate the first residue of the fragments obtained upon proteolysis with subtilisin as determined by N-terminus sequencing and the last residue of the N-terminal domain (E266 or G270) as determined by MS analysis. Amino acids that form the predicted heme-binding cavity in TT1485 are indicated by asterisks. Amino acids in ActVA-Orf6 found within a 4–5 Å distance from the substrate sancycline are indicated by hatches and those proposed to be involved in the catalysis are indicated by underlined hatches. Histidine residues lying in the region linking the PitA N- and C-terminal domains are in bold and underlined.

2 M NaCl, 50 mM Tris–HCl, pH 7.2 containing 0.02% azide. Elution was achieved using a 20–500 mM imidazole gradient. To confirm the identity of the eluted protein, deemed to be >90% pure by SDS–PAGE and Coomassie staining, the stained band was cut out and subjected to tryptic digestion and matrix-assisted laser desorption ionization time-of-flight mass spectrometry analysis of the resulting fragments (Technion Israel Institute of Technology, Haifa, Israel). The determined molecular weight profiles were probed against a partially completed version (<http://zdna2.umbi.umd.edu/~haloweb/hvo.html>) of the *Hfx.volcanii* genome. The PitA ORF is presented in Figure 1.

### Limited proteolysis of PitA

Purified PitA (20–60 µg) was subjected to proteolysis by either subtilisin or chymotrypsin in reaction buffer containing 10 mM CaCl<sub>2</sub>, 2 M NaCl, 20 mM Tris–HCl, pH 8 at 37°C for 10–30 min. The resulting proteolytic fragments were separated by SDS–PAGE and either visualized by Coomassie staining or transferred to PVDF membranes (BioRad, Hercules, CA) and subjected to N-terminal sequencing (Weizmann Institute of Science, Rehovot, Israel).

### Mass spectrometry

To determine the precise molecular weight of the N-terminal fragment of PitA generated upon subtilisin proteolysis, MS was performed. The proteolytic products were subjected to size exclusion chromatography using a HighPrep 16/60 Sephacryl S-200 column on an ÄKTA Purifier P-10 FPLC apparatus (Amersham Biosciences, Buckingham, UK) in buffer containing 2 M NaCl, 0.02% azide, 50 mM Tris–HCl, pH 7.2, at a flow rate of 0.5 ml/min, at 4°C. Fractions containing the PitA N-terminal fragment (~30 kDa, according to SDS–PAGE) were concentrated, dialyzed against double distilled water and analyzed by MS. The molecular weight value

obtained was then used to predict the C-terminus of the fragment (<http://www.expasy.org/tools/protparam.html>).

## RESULTS

Analysis of the *Hfx.volcanii* PitA sequence using the Conserved Domains Database (<http://www.ncbi.nlm.nih.gov>) revealed the presence of two conserved domains, corresponding to the N- and C-terminal halves of the protein, linked by a histidine-rich region predicted to be structurally unstable (Fig. 1). The N-terminal domain (residues 1–252) was assigned by Pfam to the chlorite dismutase family (PF06778) and by COG (Tatusov *et al.*, 2000) to COG3253, a protein family of unknown function. The C-terminal domain (residues 417–500) was assigned by Pfam to the ABM family (PF03992) and by COG to COG2329, a family predicted to participate in polyketide biosynthesis in a manner related to monooxygenase. PSI-BLAST analysis of the PitA sequence matched the Pfam functional prediction, with biochemically characterized chlorite dismutases appearing as homologues of the PitA N-terminal domain after the second iteration of PSI-BLAST. The PitA N-terminal domain was thus revealed to share 20% sequence identity with *Dechloromonas agitata* chlorite dismutase and 18% sequence identity with the same enzyme from *Ideonella dechloratans* ( $e < 1e-5$ ). The biochemically characterized aromatic polyketide monooxygenases TcmH, AknX and ActVa-Orf6, the heme oxygenases IsdG and IsdI and the quinol monooxygenase QuMo, all members of the Pfam ABM family, were identified as homologues of the PitA C-terminal domain after the fourth

PSI-BLAST iteration, sharing 15–21% sequence identity ( $e < 1e-5$ ). Indeed, all those sequences identified as being homologous ( $e < 1e-5$ ) to either of the PitA domains were assigned by Pfam to either the chlorite dismutase or ABM families.

The two members of the Pfam chlorite dismutase family for which structural information is available display a significant degree of sequence identity to the PitA N-terminal domain. TT1485 from *T. thermophilus* (PDB code 1VDH) shares 42% sequence identity with the PitA N-terminal domain while *Bacillus stearothermophilus* Apc35880 (PDB code 1T0T) shares 46% sequence identity with this portion of PitA. Sharing 54% sequence identity, TT1485 and Apc35880 display a high degree of structural similarity and are both assigned to the chlorite dismutase-like family by the Structural Classification of Proteins (SCOP) database (Murzin *et al.*, 1995). Characterized family members exist as homotetramers or pentamers (van Ginkel *et al.*, 1996; Danielsson-Thorel *et al.*, 2002; Ebihara *et al.*, 2005). Although the amino acids involved in heme binding and catalysis in chlorite dismutases are still unknown, analysis of the TT1485 crystal structure reveals that amino acids conserved in members of the Pfam chlorite dismutase family, including PitA, are organized around a cavity predicted to serve as the heme-binding site, according to the structural model of TT1485 in complex with the co-factor (Ebihara *et al.*, 2005). The model also proposes H172 of TT1485, which is strictly conserved in all the members of the Pfam chlorite dismutase family, as a candidate for heme-binding. Owing to the possible Fe–H172–D220 triad, TT1485 may serve a heme peroxidase function (Ebihara *et al.*, 2005). Indeed, several lines of evidence suggest that TT1485 may serve functions other than chlorite degradation. The TT1485 crystal structure does not contain heme, an essential cofactor for chlorite dismutase activity (Ebihara *et al.*, 2005), and reconstitution of the protein with its heme cofactor resulted in only weak oxygen generating activity using chlorite as a substrate (Ebihara *et al.*, 2005). Given the high degree of shared sequence identity between TT1485, Apc35880 and the N-terminal domain of PitA, and structural identity between TT1485 and Apc35880, it is conceivable that the haloarchaeal protein has a similar structure and fulfills roles other than chlorite degradation.

Homologues to the C-terminal domain of PitA with known structure also exist. These Pfam ABM family members include *Streptomyces coelicolor* monooxygenase ActVA Orf6 (PDB code 1LQ9), *Escherichia coli* quinol monooxygenase YgiN (PDB code 1R6Y), *Staphylococcus aureus* heme oxygenases IsdG and IsdI, (PDB codes 1XBW and 1SQE, respectively), as well as several hypothetical proteins (PDB codes 1IUJ, 1TZ0, 1X7V, 1Y0H), all of which are homodimers. Despite their relatively low degree of sequence similarity (<30% identity), these structures are highly similar. Given that the C-terminal domain of PitA shows a comparable degree of sequence identity with the other members of this group (12–20%), it is probable that it also assumes a similar structure. Catalytically important amino acids vary amongst the characterized members of the ABM family. However, since they all catalyze oxidation reactions exclusively using molecular oxygen, it is probable that their modes of action are similar. Based on amino acid conservation, the C-terminal domain of PitA shows the highest similarity to monooxygenases of aromatic polyketides, although only two of the five residues proposed to participate in the reaction catalyzed by ActVA-Orf-6 monooxygenase (W66 and R86) are conserved in PitA. However, of the other three residues (Y51, N62 and Y72), two (Y51 and Y72) appear to be unique to ActVA-Orf-6

monooxygenase as they are not conserved even in other monooxygenases known to catalyze the same or similar reactions, such as TcmH and ElmH (Sciara *et al.*, 2003), emphasizing the existing variability amongst the functionally important residues of aromatic polyketide monooxygenases.

Structural modeling by fold recognition analysis (<http://fischerlab.bioinformatics.buffalo.edu/inub/>) also supports the assignments of the PitA N- and C-terminal domains to the Pfam chlorite dismutase and ABM families, respectively. Such analysis of the N-terminal domain predicted similarity to the 1T0T structure (*B. stearothermophilus* Apc35880) with a 513 score. Analysis of the C-terminal domain predicted similarity to 1R6Y (score of 121), 1LQ9 (score of 97), 1SQE (score of 76) and to the four hypothetical proteins belonging to the ABM Pfam family mentioned above, with scores in the range of 119–165.

In addition to the various bioinformatics predictions, experiments were conducted on the purified PitA protein. Owing to the presence of a histidine-rich region in the middle of the protein (Fig. 1), PitA displays high affinity to divalent cations. This property was exploited for PitA purification via affinity chromatography on a  $\text{Co}^{2+}$  chelating resin. Structural stability of PitA was tested by limited proteolysis using subtilisin or chymotrypsin. Such proteolysis gave rise to two major fragments. N-terminal sequencing of each fragment revealed one to correspond to the N-terminal domain of PitA and the second to the C-terminal domain beginning at residue 322 (data not shown). Given the stability of the proteolytic pattern even after 30 min at 37°C, it can be concluded that PitA comprises structurally compact N- and C-terminal domains linked by a region predicted to be structurally flexible, that is readily accessible and hence sensitive to proteolysis.

Whereas PSI-Blast analysis revealed numerous homologues to each of the two PitA domains, only three homologues to full-length PitA were identified. Strikingly, the three, annotated as hypothetical proteins, are all found in halophilic archaeal species: Vng2021 in *Hbt. sp. NRC-1*, RnnAC3100 in *Har. marismortui* and NP2262A in *Nmn. pharaonis*. Like PitA, these proteins share the domain architecture described above. The presence of a chlorite dismutase-like moiety fused to an ABM-like domain thus appears to be unique to haloarchaea.

## DISCUSSION

While numerous examples of genes whose products are involved in a given cellular process can be found fused into a single ORF (Tsoka and Ouzounis, 2000; Yanai *et al.*, 2001), the predictive power in addressing fused genes lies in those cases where a sequence encoding a protein of unknown function is found physically linked to a characterized gene within a single ORF (Enright *et al.*, 1999; Marcotte *et al.*, 1999; Yanai *et al.*, 2001). In *Hfx. volcanii* PitA and its homologues, *Hbt. sp. NRC-1* Vng2021c, *Har. marismortui* RnnAC3100 and *Nmn. pharaonis* NP2262A, genes encoding a pair of proteins assigned to the Pfam chlorite dismutase and ABM families exist as a fused, single sequence. The two domains of PitA and its homologues may, therefore, functionally interact.

The nature of a putative functional linkage between the chlorite dismutase-like PitA N-terminal domain and the ABM-like C-terminal domain is not evident. Since such fusion is restricted, thus far, to halophilic archaea, it is possible that their putative

interaction is related to the unique environments in which these organisms thrive. Halophilic archaea live in highly saline millieu such as the Dead Sea (*Hfx.volcanii* and *Har.marismortui*), salt fields (*Hbt. sp. NRC-1*) or African soda lakes (*Nmn.pharaonis*), where salt is found in molar concentrations, often approaching saturation. Consequentially, halophilic archaea maintain a highly saline cytoplasm and have evolved to cope with the challenges hypersalinity presents to DNA stability, membrane integrity, bioenergy generation, protein folding and oxygen availability (Gunde-Cimerman et al., 2005). In addition to the ability of some species to grow anaerobically, i.e. in the absence of oxygen (Bickel-Sandkötter et al., 1996), halophilic archaea display various adaptations related to life in highly saline environments where oxygen solubility is limited. For example, 'purple membrane', a two-dimensional crystalline array of the light-driven proton pump, bacteriorhodopsin, appears in response to low oxygen levels (Shand and Betlach, 1991). The vertical position of cells in the water column is controlled by gas vesicles in some halophilic archaeal species, partially regulated in response to oxygen levels (Pfeifer et al., 2002). The fusion of a chlorite dismutase-like domain with an ABM-like domain in *Hfx.volcanii* PitA and its homologues in other halophilic archaea may thus also represent a modification to limited oxygen availability. As such, the N-terminal chlorite dismutase domain would provide an oxygen-generating function, releasing molecular oxygen as part of its processing of chlorite, hydrogen peroxide or other substrates. The C-terminal ABM domain, catalyzing the oxidation of phenolic compounds to the corresponding quinones or mediating the oxidation of other substrates in a molecular oxygen-dependent manner, would make use of the molecular oxygen generated by the N-terminal domain of PitA. Hence, the presence of a chlorite dismutase-like and ABM-like moieties in PitA and its homologues could serve to couple oxygen formation and utilization.

It is also possible that the two domains of PitA fulfill roles other than those presently assigned. Examples of Pfam chlorite dismutase and ABM family members catalyzing reactions distinct from other members of the same group have been reported (see above). The N-terminal domain of PitA could act as a heme-dependent enzyme catalyzing neutralization of toxic reactive oxygen species, such as hydrogen peroxide or superoxide, possibly formed by the C-terminal domain. Conversely, the PitA C-terminal domain could be involved in neutralization of toxic heme derivatives. It is not inconceivable that such alternative activities would be related to the lifestyle of halophilic archaea, given that such gene fusion appears to be restricted to these microorganisms. Indeed, the fusion of chlorite dismutase- and ABM-like domains, and thus, their proposed functional interaction in PitA and its homologues, may be a unique adaptation to life in hypersaline conditions.

Alternatively, the putative interaction of chlorite dismutase- and ABM-like moieties may reflect a more general phenomenon. It is thought that the occurrence of otherwise distinct genes fused into a single ORF in even a single organism can be used to infer a functional relationship between the physically-linked gene products that can be extrapolated to organisms where the same sequences exist as distinct entities (Enright et al., 1999; Marcotte et al., 1999). As proposed by Marcotte et al. (1999), such ORFs are referred to as 'Rosetta Stone sequences'. *Hfx.volcanii* PitA and its homologues may thus represent 'Rosetta Stone sequences' for chlorite dismutase- and ABM-like activities.

## ACKNOWLEDGEMENTS

J.E. is supported by the ISF (grant 433/03). J.M-F. is supported by NIH (R01 GM057498) and DOE (DE-FG02-05ER15650).

*Conflict of Interest:* none declared.

## REFERENCES

- Adams,M.A. and Jia,Z. (2005) Structural and biochemical evidence for an enzymatic quinone redox cycle in *Escherichia coli*: identification of a novel quinol monoxygenase. *J. Biol. Chem.*, **4**, 8358–8363.
- Baliga,N.S. et al. (2004) Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome Res.*, **14**, 2221–2234.
- Bateman,A. et al. (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, 138–141.
- Bender,K.S. et al. (2002) Sequencing and transcriptional analysis of the chlorite dismutase gene of *Dechloromonas agitata* and its use as a metabolic probe. *Appl. Environ. Microbiol.*, **68**, 4820–4826.
- Bickel-Sandkötter,S. et al. (1996) Conversion of energy in halobacteria: ATP synthesis and phototaxis. *Arch. Microbiol.*, **166**, 1–11.
- Chung,J. et al. (2002) Expression, purification, and characterization of AknX atrhron oxygenase, which is involved in aklavinone biosynthesis in *Streptomyces galilaeus*. *J. Bacteriol.*, **184**, 6115–6122.
- Danielsson Thorell,H. et al. (2002) Cloning, characterisation, and expression of a novel gene encoding chlorite dismutase from *Ideonella dechloratans*. *Biochim. Biophys. Acta*, **1577**, 445–451.
- Ebihara,A. et al. (2005) Structure-based functional identification of a novel heme-binding protein from *Thermus thermophilus* HB8. *J. Struct. Funct. Genomics*, **6**, 21–32.
- Enright,A.J. et al. (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature*, **402**, 86–90.
- Falb,M. et al. (2005) Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis*. *Genome Res.*, **15**, 1336–1343.
- Gunde-Cimerman,N., Oren,A. and Plemenitaš,A. (2005) *Adaptation to Life at High Salt Concentrations in Archaea, Bacteria, and Eukarya*. Springer, Dordrecht.
- Kendrew,S.G. et al. (1997) Identification of a monoxygenase from *Streptomyces coelicolor* A3(2) involved in the biosynthesis of actinorhodin: purification and characterization of the recombinant enzyme. *J. Bacteriol.*, **179**, 4305–4310.
- Marcotte,E.M. et al. (1999) Detecting protein function and protein–protein interactions from genome sequences. *Science*, **285**, 751–753.
- Mevarech,M. and Werczberger,R. (1985) Genetic transfer in *Halobacterium volcanii*. *J. Bacteriol.*, **162**, 461–462.
- Murzin,A.G. et al. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
- Ng,W.V. et al. (2000) Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl Acad. Sci. USA*, **97**, 12176–12181.
- Okeke,B.C. et al. (2002) Reduction of perchlorate and nitrate by salt tolerant bacteria. *Environ. Pollut.*, **118**, 357–363.
- Pfeifer,F. et al. (2002) Regulation of gas vesicle formation in halophilic archaea. *J. Mol. Microbiol. Biotechnol.*, **4**, 175–181.
- Rafanan,E.R. et al. (2001) Cloning, sequencing, and heterologous expression of the elmGHIJ genes involved in the biosynthesis of the polyketide antibiotic elloramycin from *Streptomyces olivaceus* Tu2353. *J. Nat. Prod.*, **64**, 444–449.
- Rikken,G.B. et al. (1996) Transformation of (per)chlorate into chloride by a newly isolated bacterium: reduction and dismutation. *Appl. Microbiol. Biotechnol.*, **45**, 420–426.
- Sciara,G. et al. (2003) The structure of ActVA-Orf6, a novel type of monoxygenase involved in actinorhodin biosynthesis. *EMBO J.*, **22**, 205–215.
- Shand,R.F. and Betlach,M.C. (1991) Expression of the *bop* gene cluster of *Halobacterium halobium* is induced by low oxygen tension and by light. *J. Bacteriol.*, **173**, 4692–4699.
- Shen,B. and Hutchinson,C.R. (1993) Tetracenomycin F1 monoxygenase: oxidation of a naphacenone to a naphacenonequinone in the biosynthesis of tetracenomycin C in *Streptomyces glaucescens*. *Biochemistry*, **32**, 6656–6663.

- Tatusov,R.L. *et al.* (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.*, **28**, 33–36.
- Tsoka,S. and Ouzounis,C.A. (2000) Prediction of protein interactions: metabolic enzymes are frequently involved in gene fusion. *Nat. Genet.*, **26**, 141–142.
- van Ginkel,C.G. *et al.* (1996) Purification and characterization of chlorite dismutase: a novel oxygen-generating enzyme. *Arch. Microbiol.*, **166**, 321–326.
- Wu,R. *et al.* (2005) *Staphylococcus aureus* IsdG and IsdI, heme-degrading enzymes with structural similarity to monooxygenases. *J. Biol. Chem.*, **280**, 2840–2846.
- Yanai,I. *et al.* (2001) Genes linked by fusion events are generally of the same functional category: a systematic analysis of 30 microbial genomes. *Proc. Natl Acad. Sci. USA*, **98**, 7940–7945.