View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

FEBS 29201 FEBS Letters 579 (2005) 591–596

Hypothesis

Combining data from genomes, Y2H and 3D structure indicates that BolA is a reductase interacting with a glutaredoxin

Martijn A. Huynen[a,b,*], Chris A.E.M. Spronk[b], Toni Gabaldón[a,b], Berend Snel[a,b]

[a] CMBI, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Toernooiveld 1, 6525ED Nijmegen, The Netherlands
[b] Center for Molecular and Biomolecular Informatics, Radboud University Nijmegen, Toernooiveld 1, 6525ED Nijmegen, The Netherlands

Received 15 September 2004; revised 22 October 2004; accepted 9 November 2004

Available online 29 December 2004

Edited by Robert B. Russell

Abstract **Genomes, functional genomics data and 3D structure reflect different aspects of protein function. Here, we combine these data to predict that BolA, a widely distributed protein family with unknown function, is a reductase that interacts with a glutaredoxin. Comparisons at the 3D structure level as well as at the sequence profile level indicate homology between BolA and OsmC, an enzyme that reduces organic peroxides. Complementary to this, comparative analyses of genomes and genomics data provide strong evidence of an interaction between BolA and the mono-thiol glutaredoxin family. The interaction between BolA and a mono-thiol glutaredoxin is of particular interest because BolA does not, in contrast to its homolog OsmC, have evolutionarily conserved cysteines to provide it with reducing equivalents. We propose that BolA uses the mono-thiol glutaredoxin as the source for these.**
© 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: BolA; Comparative genomics; Mono-thiol glutaredoxin; PICOT-HD; Protein function prediction; Uvi31+

1. Introduction

In the genomics era, we obtain many correlates of protein function, such as a protein's 3D structure, its gene expression, its physical interaction partners, the location of its gene on the genome and its phylogenetic distribution. The various sources of information are stored in dedicated databases that provide invaluable resources for the prediction of the biological function of a protein. Nevertheless, none of these give a direct answer to the question: what does the protein do at the molecular level? However, by combining on the one hand homology information, that can be deduced from 3D structure comparisons and that provides information about the molecular function of a protein, with on the other hand genomics context data that provide information about the interaction partners or substrates of a protein, one can derive specific hypotheses about its function.

Here, we combine genome sequences, physical interaction data and 3D structures to provide a specific prediction for the function of BolA, a protein family that is widespread among proteobacteria and eukaryotes including *Homo sapiens.* Despite a considerable amount of research on the function of BolA, its molecular function remains unknown. Originally, *bolA* been identified as a gene that causes round morphology in *Escherichia coli* when overexpressed [1]. No phenotype for strains lacking *bolA* has been found for cells growing on a rich medium. Consistent with this is that *bolA* is mainly expressed under the stress conditions like the stationary phase, osmotic shock, carbon starvation and oxidative stress [2]. The gene *bolA* is under control of RpoS, which also regulates the expression of other stationary-phase-induced stress genes [3]. A homolog of *bolA* in *Schizosaccharomyces pombe*, uvi31+, is expressed under UV radiation [4], which is known to stimulate the intracellular synthesis of reactive oxygen species. Although no direct evidence about BolA's function is available, its associated phenotypes link BolA to cell morphology and cell division. Under nutrient-restrictive conditions *bolA* in *E. coli* is required for normal cell morphology [2], while uvi31+ in *S. pombe* has been implicated in the regulation of septation and cytokinesis [5]. How the link between BolA and cell division and morphology is effectuated is however not clear. Overexpression of *bolA* in *E. coli* leads to the upregulation of cell wall synthesis genes *dacA* (PBP5), *dacC* (PBP6) and *ampC* (AmpC), and BolA's effect on the cell morphology appears dependent on PBP5 and PBP6 [6]. Bola has therewith been proposed to be a regulator of cell wall biosynthetic enzymes [6]. Recently, the structure of the BolA homolog in *Mus musculus* has been determined [7]. BolA has a class II KH fold, instances of which are known to bind DNA and RNA and therewith support a regulatory role for BolA.

Here, we show that the accumulating wealth of genomics data indicates however a different molecular function for BolA. We show that BolA is homologous to the peroxide reductase OsmC and, relative to other class II KH-fold proteins, most closely related to it, and that BolA has a very strong genomic association with the mono-thiol glutaredoxins/PICOT-HD [8] family. These data point to a role of BolA as a reductase that functions in conjunction with a mono-thiol glutaredoxin. The thiol group of the latter would potentially be used by BolA to reduce and/or deglutathionylate substrates.

*Corresponding author. Fax: +31-24-3652977.
E-mail addresses: huynen@cmbi.ru.nl (M.A. Huynen), c.spronk@cmbi.ru.nl (C.A.E.M. Spronk), t.gabaldon@cmbi.ru.nl (T. Gabaldón), snel@cmbi.ru.nl (B. Snel).

0014-5793/$30.00 © 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2004.11.111

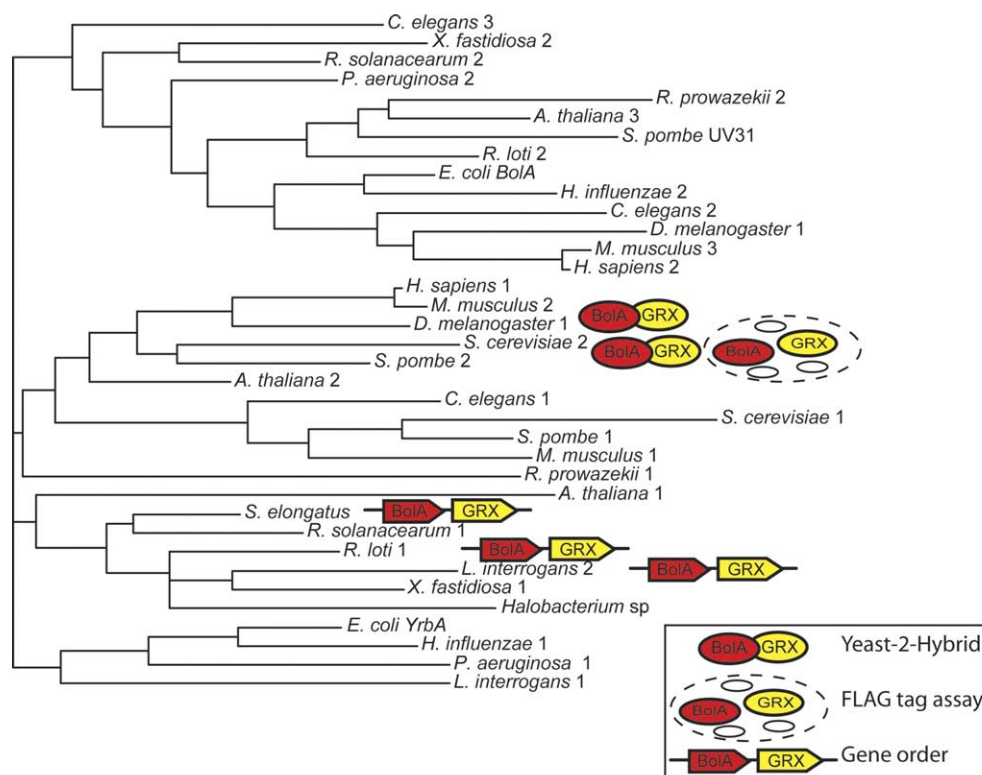

Fig. 1. Phylogeny of the BolA family with interaction predicting genomics data. Different types of genomic contexts are indicated, "GRX" indicates the mono-thiol glutaredoxin family. Physical interaction between BolA and a mono-thiol glutaredoxin is supported by various types of genomics data across a wide phylogenetic range.

2. Methods

The phylogeny of the BolA family (Fig. 1) was reconstructed with PhyML [9] based on a sequence alignment created with muscle [10], the sequence identifiers are: Q9VC53 Q9D8S9 YCE3_HUMAN YN1I_CAEEL Q92ME9 Q84W65 Q9ZDT0 UV31_SCHPO Q87BI5 Q8XYL3 BOLA_ECOLI Q7VKT0 Q88M18 ENSCBRP00000005226 Q8CEI1 Q9USK1 YAE6_YEAST Q8F4C9 MY16_HUMAN MY16_MOUSE Q8ML41 YGX0_YEAST O14280 Q9FIC3 Q9XVJ0 Q98NN9 Q9HQ28 Q7P9Z5 YRBA_ECOLI YA82_HAEIN Q9HVW6 Q8DKI4 Q87DN9 Q8F4D1 Q8XV77 Q9LF68.

Sequence alignments for the comparison of BolA with the other class II KH-fold families were obtained from PFAM [11]. The E values for the profile comparison were calculated with Compass, the database size used (-d option in compass) was the combined length of all the KH-fold alignments.

3. Results

3.1. BolA is homologous to the peroxide reductase OsmC

Comparative sequence analyses indicate that homologs of *bolA* are present in Bacterial and Eukaryotic genomes. Consistent with earlier findings from large-scale phylogenetic analyses [12], the eukaryotic representatives of *bolA* appear derived from the proteobacteria (Fig. 1) and presumably have hitchhiked along with the endosymbiosis of an α-proteobacterium that became the mitochondrion. The molecular function of BolA is unknown and using sequence-to-profile searches (PSI-Blast) [13], we could not detect homology of BolA to sequences with a known molecular function. Recently, however, the 3D structure of a *Mus musculus* member of the BolA family has been published [7]. According to the Dali 3D structure classification system, BolA is homologous to OsmC (Table 1). BolA and the N-terminal domain of OsmC have a class II KH fold. They distinguish themselves from other members of this fold because they both miss the GxxG element [7], which otherwise is well conserved in the class II KH fold and which appears essential

Table 1
Similarity of BolA to other class II KH-fold proteins

Protein family (PDB entry)	3D similarity to Bola, Z scores	Sequence profile similarity to Bola. SW score (E value)
OsmC (1ml8A/1lqlA)	5.8/5.5	73 (2.4E − 5)
Ohr (1n2fA)	5.2	
KH 1 (1hnxC)	5.3	46 (9.4E − 3)
DUF150 (1ib8A)	3.7	44 (4.2E − 2)
GMP synthase C (1gpmA)	2.9	57 (7.0E − 4)
KH 2 (1egaB)	3.8	35 (2.7E − 1)
RBFA (1kkgA)	4.2	40 (9.6E − 2)

The similarities of the BolA structure to other protein structures of the class II KH fold are indicated at the 3D structure level as well as at the sequence profile level. The Z scores are calculated with DALI [35], using pairwise comparisons of the BolA structure with other class II KH folds in the DALI database. The higher the Z score, the less likely it is that the similarity between the 3D structures is "random". The Smith–Waterman scores and E values for the comparison of sequence profiles were calculated with Compass [15], the higher the score, the more similar the sequence profiles. For the profile similarity analysis, one combined alignment was used of the OsmC and Ohr subfamilies, as they form one protein family. Both structure comparison as well as sequence profile comparison clearly indicate that BolA is homologous to the OsmC/Ohr family as well as most similar to the OsmC/Ohr family relative to other members of the class II KH fold.

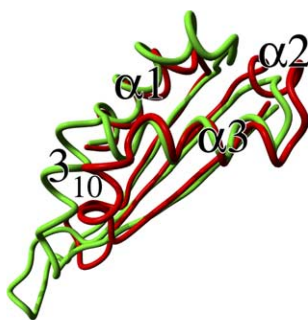

Fig. 2. Superposition of the 3D structures of BolA (red) and OsmC (green). The structures, BolA of *M. musculus* (1v9j) and OsmC of *E. coli* (1ml8), were aligned with SHEBA [36]. Indicated are the α1, α2, α3 helix and the 3_{10} helix of BolA, the nomenclature is adapted from Kasai et al. [7]. The main difference between the structures is that the α3 helix of OsmC is one turn longer than the one of BolA.

for nucleic acid binding [14]. In terms of 3D structure similarity, BolA is most similar to OsmC (Table 1), relative to other proteins with a class II KH fold and the structures can readily be superimposed (Fig. 2). The main difference between the BolA and OsmC structures is that the α3-helix of BolA is one turn shorter than the corresponding helix of OsmC (Fig. 2). Compass-based comparisons [15] of the sequence profile of BolA with the profiles of the other members of the class II KH fold indicate that also at the sequence level BolA is most similar to the OsmC family that also includes the thiol-dependent reductase Ohr (Table 1). OsmC is a hyperperoxide reductase that can detoxify hydroperoxides by reducing them into alcohols [16]. The active site of OsmC has two conserved cysteines, which have been proposed to carry the reducing equivalents for OsmCs reductive activity [16]. These cysteines are however not conserved in BolA. If BolA functions as a reductase, as the

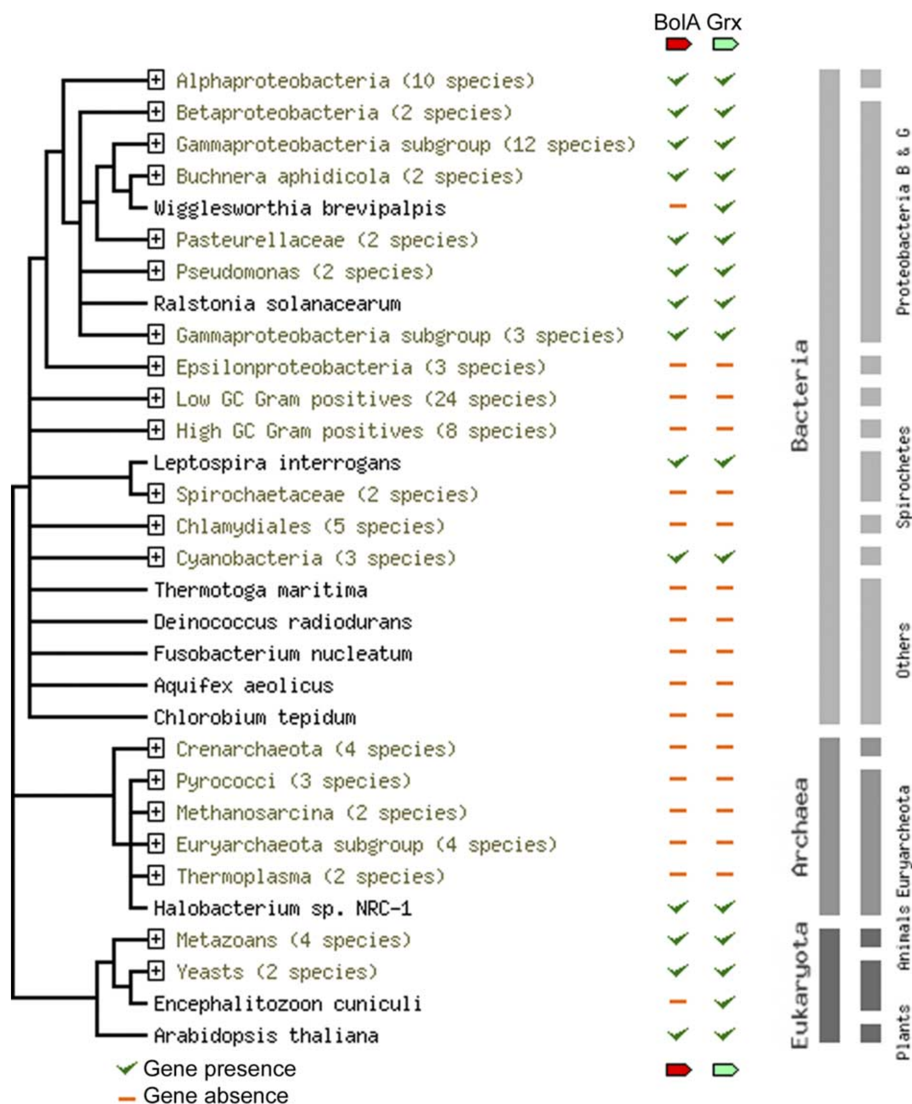

Fig. 3. Phylogenetic distribution of BolA and of the mono-thiol glutaredoxin family (Grx). Data were obtained with STRING [18]. Among 145 sequenced genomes, the phylogenetic distribution of the genes is virtually identical. The only exception is *Encephalotozoon cuniculi*. *Wigglesworthia brevipalpis* does not contain *bolA* in the COG orthology database [37] that is used in STRING, it does however contain a homolog of *bolA*, which in COGs has been classified as *bolA*'s paralog *yrbA*.

homology with OsmC indicates, it has to obtain its reducing equivalents from another source.

3.2. Gene order conservation and gene co-occurrence indicate an interaction of BolA with a mono-thiol glutaredoxin

Potential interaction partners for BolA, which could provide these reducing equivalents, can be found by comparative genome analysis. For *bolA*, we have analyzed a number of so-called "genomic context" [17] types (gene fusion, gene-order conservation, and the co-occurrence of genes among sequenced genomes) using the genomic context server STRING [18]. Two types of genomic context indicate an interaction between BolA and a mono-thiol glutaredoxin/PICOT-homology domain [8]: Their conserved occurrence as neighbors in Bacterial genomes (Fig. 1) and the co-occurrence of their genes across virtually all sequenced genomes (Fig. 3). Either type of genomic context has been successfully used to predict functional interactions in the past [19] and their combined presence leads to an estimated likelihood of interaction of BolA with a mono-thiol glutaredoxin of 97% [18]. In general, genomic context data do not indicate what type of interaction (metabolic, physical or regulatory) two proteins have, although when, as in this case, the association is very strong, the interaction does tend to be a physical one [17].

3.3. An evolutionary conserved physical interaction of BolA with Grx3

We examined genomics databases of physical interaction experiments to establish whether BolA has indeed a physical interaction with a mono-thiol glutaredoxin. In *S. cerevisiae*, both in the yeast-2-hybrid experiments [20] as well in the FLAG tag experiments [21], an interaction of Grx3 (a *S. cerevisiae* homolog of the bacterial mono-thiol glutaredoxins) with YGL220w (a *S. cerevisiae* BolA homolog) has been detected. The detection of such an interaction with two independent methods increases the likelihood that they are indeed biologically relevant [22]. Furthermore, this interaction has also been observed in another species. In a yeast-2-hybrid assay on *Drosophila melanogaster*, CG16804 and CG6523, the orthologs of YGL220w and Grx3, respectively, have also been shown to interact with each other [23]. Such evolutionary conservation of a yeast-2-hybrid interaction increases the likelihood of biological relevance to 100% [24,25]. A physical interaction of BolA with a mono-thiol glutaredoxin thus appears to be more than likely.

3.4. The mono-thiol glutaredoxin/PICOT-HD family

Glutaredoxins are redox enzymes that use glutathione to catalyze disulfide reductions [26]. Consistent with their interaction with BolA is that glutaredoxins are, like BolA, involved in oxidative stress response [26]. Within the glutaredoxins, the mono-thiol glutaredoxins or PICOT-HD [8] form a separate group that lack one of the two conserved cysteines of the dithiol glutaredoxins, and that are frequently fused with thioredoxins in eukaryotes [8]. Their exact function has not been elucidated. It has been suggested that they are involved in the deglutathionylation of protein-GS mixed sulfides [27,28]. For this process dithiol glutaredoxins only require their N-terminal cysteine thiol [29], which is the one cysteine that is conserved in mono-thiol glutaredoxins. Indeed the best characterized member of this family, the mitochondrial protein Grx5 from yeast, is able to deglutathionylate proteins [27], and is involved in defense against oxidative stress [30].

3.5. Interpreting the interaction between BolA and Grx3

Given the strong link between BolA and Grx3, they likely function as a complex and are both involved in the same process. BolA could, e.g., use the reducing equivalents from Grx3 to reduce and/or deglutathionylate some substrate. In that case, the reducing equivalents from Grx3 would "replace" the ones that are carried by the conserved cysteines from OsmC and that are absent from BolA. An alternative is that one of the proteins is a target for the other and the interaction is "transient". In that case, BolA could play a role in the deglutathionylation of Grx3 itself. Here, it should be noted that glutathione is able to deglutathionylate Grx5, albeit at a 20-fold lower rate than for the reduction of dithiolic glutaredoxin from *E. coli* [27].

3.6. The link with cell division

Genomic context data do also indicate other interaction partners of BolA (Fig. 4). Given BolA's phenotypic links to cell division, most interesting are conservation of gene order with the gene for intracellular septation protein A (ISPA) [31] and with the gene for MurA that catalyzes the first step of peptidoglycan synthesis. Neither of these proteins can however be detected in eukaryotes. A link to the cell wall that does carry to the eukaryotes is with a Parvulin-like peptidyl–prolyl isomerase D (PPIase D) whose gene has a conserved gene order with *bolA* in prokaryotes, and is co-expressed with it in *S. cerevisiae* [18]. PPIase D is, in *E. coli*, a periplasmic chaperone that is required for folding of outer membrane proteins [32]. Finally, there is conservation of gene order with an ABC-transporter system that is involved in tolerance to toluene [33], and that has been predicted to be an efflux system [33], and there is gene co-occurrence with Glutathione-*S*-transfer-

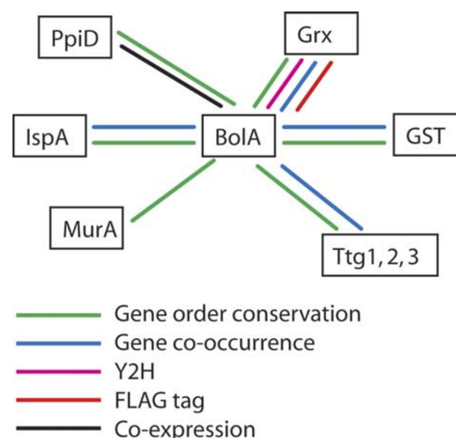

Fig. 4. Predicted interactions of the BolA family. The various types of data that support the interactions are indicated. The proteins on the left can be linked to the cell wall or cell division: MurA is involved in cell wall synthesis, IspA is involved in cell division, and PPIase D is a periplasmic chaperone. The proteins on the right are more generally involved in (oxidative) stress response: Ttg1, Ttg2 and Ttg3 form an ABC transporter involved in resistance to Toluene, GST is Glutathione-*S*-transferase, and Grx is mono-thiol glutaredoxin. The results were obtained with STRING [18]. In the figure, the results for BolA/COG0271 and for its paralog YrbA/COG5007 are combined because the COG subclassification of the BolA family does not match the phylogenetic tree (data not shown) and therefore appears unreliable.

ase, that functions in defense against oxidative stress. The links with MurA, IspA and PPIase D, although significant in themselves and consistent with the phenotypic data on *bolA*, are not as strongly supported by any type of data, nor are they supported by so many types of data as the link between BolA and the mono-thiol glutaredoxins. They suggest that the targets of the BolA-glutaredoxin pair are proteins or other organic compounds that are either part of the membrane or cell wall or that are involved in its generation.

4. Discussion

Protein function predictions have been made based on similarity at the level of the sequence or 3D structure (reviewed in [34]), or on combinations of homology with either functional genomics data, e.g. [23] or genomic context data [17]. Here we combine 3D structure data with both genomic context data and functional genomics data. The increasing pace at which all these types of data are becoming available for hypothetical proteins calls for a further integration of genomic context databases like STRING [18] with homology databases like DALI [35] to facilitate the making of specific hypotheses about protein function. The challenge will thereby include functional information that assists in observing a potential link between "BolA is homologous to a peroxide reductase" and "BolA interacts with a mono-thiol glutaredoxin". Such a level of integration would be of tremendous value for reaping the benefits of genomics for the understanding of the cell at the molecular level.

Acknowledgment: We thank Sander Nabuurs for assistance with the protein structure alignment. This work was supported in part by a grant from the Netherlands Organization for Scientific Research (NWO).

References

[1] Aldea, M., Hernandez-Chico, C., de la Campa, A.G., Kushner, S.R. and Vicente, M. (1988) Identification, cloning, and expression of bolA, an ftsZ-dependent morphogene of *Escherichia coli*. J. Bacteriol. 170, 5169–5176.
[2] Santos, J.M., Freire, P., Vicente, M. and Arraiano, C.M. (1999) The stationary-phase morphogene bolA from *Escherichia coli* is induced by stress during early stages of growth. Mol. Microbiol. 32, 789–798.
[3] Hengge-Aronis, R., Lange, R., Henneberg, N. and Fischer, D. (1993) Osmotic regulation of rpoS-dependent genes in *Escherichia coli*. J. Bacteriol. 175, 259–265.
[4] Lee, J.K., Park, E.J., Chung, H.K., Hong, S.H., Joe, C.O. and Park, S.D. (1994) Isolation of UV-inducible transcripts from *Schizosaccharomyces pombe*. Biochem. Biophys. Res. Commun. 202, 1113–1119.
[5] Kim, M.J., Kim, H.S., Lee, J.K., Lee, C.B. and Park, S.D. (2002) Regulation of septation and cytokinesis during resumption of cell division requires uvi31+, a UV-inducible gene of fission yeast. Mol. Cells 14, 425–430.
[6] Santos, J.M., Lobo, M., Matos, A.P., De Pedro, M.A. and Arraiano, C.M. (2002) The gene bolA regulates dacA (PBP5), dacC (PBP6) and ampC (AmpC), promoting normal morphology in *Escherichia coli*. Mol. Microbiol. 45, 1729–1740.
[7] Kasai, T., et al. (2004) Solution structure of a BolA-like protein from *Mus musculus*. Protein Sci. 13, 545–548.
[8] Isakov, N., Witte, S. and Altman, A. (2000) PICOT-HD: a highly conserved protein domain that is often associated with thioredoxin and glutaredoxin modules. Trends Biochem. Sci. 25, 537–539.
[9] Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.
[10] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
[11] Bateman, A., et al. (2004) The Pfam protein families database. Nucleic Acids Res. 32 (database issue), D138–D141.
[12] Gabaldon, T. and Huynen, M.A. (2003) Reconstruction of the proto-mitochondrial metabolism. Science 301, 609.
[13] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
[14] Baber, J.L., Levens, D., Libutti, D. and Tjandra, N. (2000) Chemical shift mapped DNA-binding sites and 15N relaxation analysis of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K. Biochemistry 39, 6022–6032.
[15] Sadreyev, R. and Grishin, N. (2003) COMPASS: a tool for comparison of multiple protein alignments with assessment of statistical significance. J. Mol. Biol. 326, 317–336.
[16] Rehse, P.H., Ohshima, N., Nodake, Y. and Tahirov, T.H. (2004) Crystallographic structure and biochemical analysis of the Thermus thermophilus osmotically inducible protein C. J. Mol. Biol. 338, 959–968.
[17] Huynen, M., Snel, B., Lathe III, W. and Bork, P. (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome Res. 10, 1204–1210.
[18] von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. and Snel, B. (2003) STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 31, 258–261.
[19] Huynen, M.A., Snel, B., von Mering, C. and Bork, P. (2003) Function prediction and protein networks. Curr. Opin. Cell Biol. 15, 191–198.
[20] Ito, T., et al. (2000) Toward a protein–protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc. Natl. Acad. Sci. USA 97, 1143–1147.
[21] Bhan, A., Galas, D.J. and Dewey, T.G. (2002) A duplication growth model of gene expression networks. Bioinformatics 18, 1486–1493.
[22] von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P. (2002) Comparative assessment of large-scale data sets of protein–protein interactions. Nature 417, 399–403.
[23] Giot, L., et al. (2003) A protein interaction map of *Drosophila melanogaster*. Science 302, 1727–1736.
[24] Kelley, B.P., Sharan, R., Karp, R.M., Sittler, T., Root, D.E., Stockwell, B.R. and Ideker, T. (2003) Conserved pathways within bacteria and yeast as revealed by global protein network alignment. Proc. Natl. Acad. Sci. USA 100, 11394–11399.
[25] Huynen, M.A., Snel, B. and Van Noort, V. (2004) Comparative genomics for reliable protein-function prediction from genomic data. Trends Genet. 20, 340–344.
[26] Fernandes, A.P. and Holmgren, A. (2004) Glutaredoxins: glutathione-dependent redox enzymes with functions far beyond a simple thioredoxin backup system. Antioxid. Redox Signal. 6, 63–74.
[27] Tamarit, J., Belli, G., Cabiscol, E., Herrero, E. and Ros, J. (2003) Biochemical characterization of yeast mitochondrial Grx5 monothiol glutaredoxin. J. Biol. Chem. 278, 25745–25751.
[28] Shenton, D., Perrone, G., Quinn, K.A., Dawes, I.W. and Grant, C.M. (2002) Regulation of protein *S*-thiolation by glutaredoxin 5 in the yeast *Saccharomyces cerevisiae*. J. Biol. Chem. 277, 16853–16859.
[29] Bushweller, J.H., Aslund, F., Wuthrich, K. and Holmgren, A. (1992) Structural and functional characterization of the mutant *Escherichia coli* glutaredoxin (C14–S) and its mixed disulfide with glutathione. Biochemistry 31, 9288–9293.
[30] Rodriguez-Manzaneque, M.T., Ros, J., Cabiscol, E., Sorribas, A. and Herrero, E. (1999) Grx5 glutaredoxin plays a central role in protection against protein oxidative damage in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 19, 8180–8190.

[31] Mac Siomoin, R.A., Nakata, N., Murai, T., Yoshikawa, M., Tsuji, H. and Sasakawa, C. (1996) Identification and characterization of ispA, a *Shigella flexneri* chromosomal gene essential for normal in vivo cell division and intracellular spreading. Mol. Microbiol. 19, 599–609.
[32] Dartigalongue, C. and Raina, S. (1998) A new heat-shock gene, ppiD, encodes a peptidyl–prolyl isomerase required for folding of outer membrane proteins in *Escherichia coli*. EMBO J. 17, 3968–3980.
[33] Kim, K., Lee, S., Lee, K. and Lim, D. (1998) Isolation and characterization of toluene-sensitive mutants from the toluene-resistant bacterium *Pseudomonas putida* GM73. J. Bacteriol. 180, 3692–3696.
[34] Whisstock, J.C. and Lesk, A.M. (2003) Prediction of protein function from protein sequence and structure. Q. Rev. Biophys. 36, 307–340.
[35] Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. J. Mol. Biol. 233, 123–138.
[36] Jung, J. and Lee, B. (2000) Protein structure alignment using environmental profiles. Protein Eng. 13, 535–543.
[37] Tatusov, R.L., et al. (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4, 41.