J Biol Inorg Chem (2010) 15:1071–1078
DOI 10.1007/s00775-010-0666-6

ORIGINAL PAPER

The annotation of full zinc proteomes

Ivano Bertini · Leonardo Decaria · Antonio Rosato

Received: 15 January 2010 / Accepted: 16 April 2010 / Published online: 5 May 2010

Abstract We obtained an extended functional annotation of zinc proteins using a combination of bioinformatic methods. This work was performed using a number of available predicted zinc proteomes of various representative organisms, leading to the almost complete annotation of, among others, the predicted human zinc proteome. The computational tools exploited included sequence-based and, when possible, structure-based functional predictions. We assigned a hypothetical function to 74% of the 1,472 sequences analyzed that lacked annotation in the starting dataset. We also added new functional categories, not described in the reference dataset, such as ubiquitin binding and DNA replication. As a general conclusion, we can state that the quality of each functional prediction parallels the amount of information for the sequence analyzed: the larger the amount of information, the more detailed and reliable is the proposed functional prediction. Among the findings, we have propose a zinc binding site for archaeal zinc-importing proteins. Furthermore, we propose two groups of transcriptional regulators that are involved in fatty acid metabolism.

Keywords Zinc · Metalloproteomics · Metalloproteome · Zinc finger

Introduction

Zinc is essential for life and is the second most abundant transition metal ion in living organisms after iron. In contrast to other transition metal ions, such as copper and iron, zinc is present in cells in a single oxidation state, zinc(II), which does not undergo redox reactions owing to its filled *d* shell. In the present work, we want to further our understanding of the biochemical functions that underlie the requirement of organisms for zinc. The present work constitutes a bioinformatic contribution toward mapping the many cellular processes involving zinc.

The zinc proteomes of 57 representative living organisms including members of Archea, Bacteria, and Eukarya are available [1]. These 57 zinc proteomes were previously predicted to encode cumulatively 18,336 potential zinc-binding proteins, which had been grouped into ensembles on the basis of sequence similarity [1]. Functional information, either based on available experimental data or on computational biology methods, was described in the annotation of most of these protein sequences, which was relevant for all proteins in a given ensemble. Functional hints for additional protein ensembles were provided by the Gene Ontology (GO) database [2]. Of the 18,336 zinc proteins, 1,472 (784 from Eukarya, 212 from Archea, and 476 from Bacteria) did not have a defined functional annotation and no hit was retrieved from the GO database; their annotations typically described them as hypothetical,

Electronic supplementary material The online version of this article (doi:10.1007/s00775-010-0666-6) contains supplementary material, which is available to authorized users.

I. Bertini (✉) · L. Decaria · A. Rosato
Magnetic Resonance Center (CERM),
University of Florence,
Via L. Sacconi 6,
50019 Sesto Fiorentino, Italy
e-mail: bertini@cerm.unifi.it

I. Bertini · A. Rosato
Department of Chemistry,
University of Florence,
Via della Lastruccia 3,
50019 Sesto Fiorentino, Italy

putative, or predicted proteins [1]. Of these sequences, 932 grouped into 204 ensembles, whereas the remaining 540 did not cluster into any group.

In the present work, we attempted to make functional predictions for these proteins by using sequence- and/or structure-based approaches, exploiting online as well as stand-alone bioinformatics databases and software tools [3–5] (see "Materials and methods"). We assigned a hypothetical function to 1,090 sequences (74% of 1,472), of which 721 constituted 132 ensembles of homologs (65% of the 204 ensembles of putative zinc proteins). For the 382 remaining sequences there were not enough data to assign a function even at a low level of confidence. The coverage of functional annotation originally included about 92% of all zinc proteins, which, after our contribution, increased to 98%. It was found that hydrolytic activity is the most represented zinc-related function in our dataset (33% of the total), followed by transcription (24% of the total). Specifically for eukaryotic zinc proteins, a role in transcription is proposed for more than 37% of these proteins. The present data confirmed the dominant role of zinc fingers in the regulation of expression in eukaryotes, within both activators and repressors. This research shows that an essentially complete annotation of the zinc proteome can be achieved for every living organism whose genome sequence is available.

Materials and methods

The prediction of the zinc proteomes [1] from which the present work started to obtain an extended functional annotation of all zinc proteins was essentially based on the use of a list of known zinc-binding domains available in the Pfam database [6, 7], filtered [8] with the available zinc-binding patterns (ZBPs). Below, we briefly recapitulate the procedure that was used in [1] to obtain these data. The Pfam domains that had been retrieved using "Zn" and "zinc" as query keywords were analyzed manually to collect a list of physiological zinc-binding domains. In parallel, all the structures in the Protein Data Bank (PDB) [9] that physiologically bind at least one zinc ion were selected. In both cases, physiological and nonphysiological zinc binders were separated by manually checking the literature for each of the systems under analysis (protein domains or individual structures). The composition in terms of Pfam domains was determined using the HMMER 2.0 [10] program for all the proteins of known structure where zinc binding is of physiological relevance. The residues coordinating the zinc ion(s) defined the ZBP for each structure analyzed [11–13]. When a ZBP was made up of residues contained in a known Pfam domain, the latter was associated with the ZBP. This resulted in a list of zinc-binding Pfam domains that was as extended as possible; one or more ZBPs were assigned to all zinc-binding Pfam domains having a structurally characterized representative [1]. The physiological relevance of zinc binding was typically supported by the available scientific literature. Andreini et al. [1] obtained the list of predicted zinc proteins by making use of the aforementioned Pfam domains and the associated ZBPs, when available, to scan the proteomes of 57 selected organisms using HMMER 2.0. No experimental verification of these predictions has been carried out. The overall procedure was recently reviewed in [8].

Sequence-based methods

We started our search by analyzing all of the sequences of interest against the entire Pfam database (release 23.0) [6, 7, 14] as we deemed that defining the composition in terms of functional domain(s) (including also those not endowed with zinc-binding capability) for each unknown sequence constitutes the most reasonable starting point for any subsequent analysis (see also our workflow for metalloprotein prediction in [8] and the diagram for comparative genomic analyses of trace elements in [15]). Unfortunately, not all the Pfam domains feature a detailed functional description. For example, in various cases only structural similarities to other proteins are reported. For this reason, we complemented the information available from Pfam with queries to the COG database [16, 17]. Each cluster of an orthologous group of proteins (COG) consists of individual proteins or groups of paralogs from at least three lineages and thus corresponds to an ancient conserved domain. In other words, each COG contains protein sequences or groups thereof that, in different species, evolved from a common ancestral gene by speciation and are thus orthologous. Usually, orthologous proteins have the same Pfam domain composition, thereby making the two methods complementary for the identification of orthology relationships. The identification of orthologs is important for reliable predictions of protein function because they normally retain the same function in the course of evolution. In contrast to the Pfam-based methods, which are mainly dependent on sequence similarity, the use of COGs, which take into account the phylogenetic distance between homologs, is more appropriate to identify relationships between distant proteins with low sequence similarity [16, 17]. To predict putative transmembrane regions in the sequences analyzed, we used TMHMM [18–21] as a stand-alone program, whereas we used pSORT [18, 22, 23], an online tool, to predict the cellular localization. For eukaryotic sequences, these were the only sequence-based tools employed. In contrast, for prokaryotic proteins we additionally exploited the STRING [24–26] and

ShOPs [27] tools. The former shows the possible functional partners of the target protein, on the basis of its COG classification, taking into account a variety of experimentally validated data (such as physical interactions, occurrence in the same metabolic pathways, gene fusions, and co-occurrence or coexpression in different organisms). ShOPs is an online server that allows the visualization of operons. The latter analysis can give indirect but very useful hints about the hypothetical function of a protein when it is codified within an operon containing genes that code for other proteins of known function [28].

Structure-based approach

For each of the 204 functionally unassigned ensembles of homologs that were contained in the starting dataset of predicted zinc proteins [1], we created a hidden Markov model (HMM) profile [10] and queried the entire PDB to identify related proteins with known structures to be used as templates in homology modeling, performed using Modeller 6v2 [29–31]. With a structural model of the target protein, the most conserved residues within the ensemble and/or the Pfam domain HMM profile (which reflects the information of a greater number of sequences) to which the target protein belongs could be mapped onto the protein surface. This, in turn, allowed us to define potential functionally important regions such as catalytic pockets or binding sites. As we are dealing with proteins predicted to bind zinc, it is also important to verify whether the proposed binding residues are close in space in the proposed model, i.e., whether they define a reasonable zinc binding site.

No HMM profile was created for the 540 sequences that did not cluster with other proteins in an ensemble in the original dataset. For these we therefore queried the PDB using each individual amino acidic sequence. When the PDB templates corresponded to proteins lacking a functional characterization (e.g., for structures determined within structural genomics projects), we used bioinformatics tools for structural analysis, such as ProFunc [32], WHISCY [33], ProMate [34], and CastP, to obtain additional functional hints. ProFunc is an online server providing clues on a protein's likely or possible function from its 3D structure. This analysis is based on the use of various databases, including the PDB and UniProt [35, 36] for the identification of structurally characterized clefts, folds, or binding motifs on the protein surface. WHISCY, ProMate, and CastP can identify active residues on the target protein surface and consequently define potential functional areas, such as protein–protein binding sites, enzyme active sites, or small-ligand binding pockets.

When homology modeling was not applicable, we performed protein threading, also known as fold recognition, using the online tool Phyre [37, 38]. Protein threading is a method to predict protein structures that aims at assigning known folds to proteins that do not have homologs with known structure. The prediction is made by aligning each amino acid of the target sequence to a position in the template structure, taken from a library of diverse folds, and evaluating how well the target fits the template, typically using a simplified potential for energy calculation. After the best-fit template has been selected, the structural model of the sequence is built on the basis of the alignment with the chosen template. The quality of the final structural prediction is measured through an expectation value (*E* value) [37, 38], which estimates quantitatively the number of possible errors given the size of the library used (thus, the lower the better). The functional predictions derived from the analysis of structural models obtained through threading have a lower degree of confidence with respect to the case of structural models obtained by homology modeling, because the models are inherently of lower quality.

For each ensemble we manually checked the consistency between the results given by all the tools used, and the support provided by the relevant scientific literature.

Results and discussion

Overview of functional annotations

We have complemented the already-available information on homology relationships among the predicted zinc proteins contained in our reference dataset using the COG database [16, 17], which, however, does not cover all of the organisms addressed in this work. The Pfam and COG assignments showed a good agreement. Indeed, the sequences in each of the starting ensembles of zinc proteins had the same composition in terms of Pfam domains as well as, when available, the same COG annotation. This provides significant evidence that the 204 groups of sequences defined in the original prediction of the zinc proteome [1] with which we started our analysis were indeed groups of homologs. In addition, we analyzed all the 540 individual sequences that could not be assigned to any ensemble (i.e., they did not have homologs in the organisms subjected to analysis in [1]).

Of the 1,472 sequences lacking functional information in [1], only for about 26% could we obtain structural models of any kind. These can be further separated into homology models (150), threading models encompassing the entire length of the target protein (100), and threading models encompassing only a part of the target protein (134). It is to be noted that we imposed a relatively restrictive threshold on the degree of sequence similarity between the target and template sequences (40% sequence

identity over the entire target length), to ensure that only reliable models were produced [39]. On the other hand, we used a relatively loose threshold (*E* value lower than 10.0) to select results from threading to ensure that the maximum amount of structural information could be obtained. We were thus left with 1,086 proteins for which no structural information could be gathered at all. We obtained useful functional information from the analysis of sequence features alone for 704 of these proteins (Table S1).

In total, we therefore assigned, with variable degrees of confidence, a possible function to 1,090 proteins, of which 721 belonged to 132 ensembles of homologs. This assignment is entirely based on the application of bioinformatic methods and therefore is typically not supported by specific experimental evidence (although there may be experimental evidence available for other, relatively close systems). The assignment results are shown in the pie graph in Fig. 1a. After our analysis, the functional assignment of the zinc proteomes has almost been completed (Figs. 1b, 2; the percentage distribution of the functional categories assigned is given in Table S2). In the reference work, and subsequently in this work, release number 36 of the human proteome was analyzed, counting about 40,000 proteins, of which 9.2% constituted the zinc proteome. The present functional annotation shows that 44% of human zinc proteins are involved in the regulation

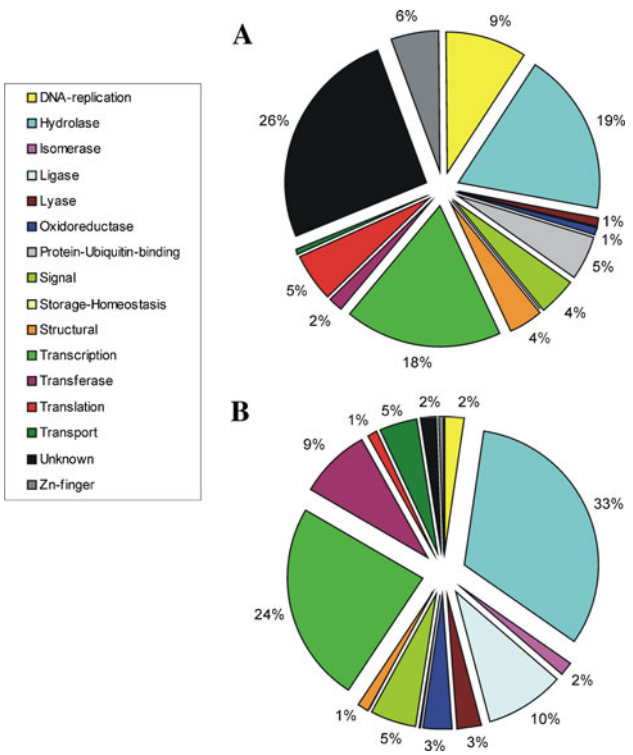


Fig. 1 The functional assignment obtained for **a** the 1,472 sequences analyzed in this work that were previously unassigned, and **b** all the 18,336 proteins in the complete 57 zinc proteomes

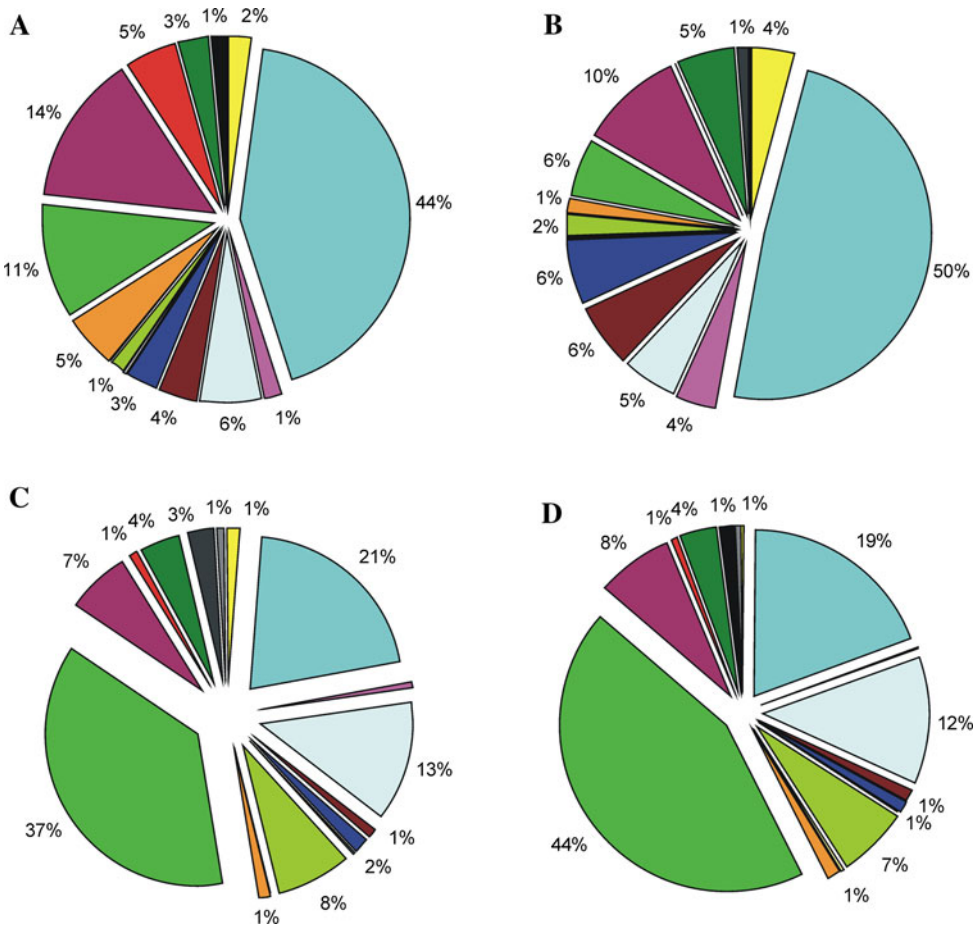
of gene expression, followed by 12% hydrolases (Fig. 2d). These figures compare well with the average distribution in Eukarya. The present work also provided some new hints on the cellular role of various families of zinc proteins, as discussed in more detail below. The portfolio of functional categories to which we could assign zinc-binding proteins was larger and finer-grained than that described in [1], thanks to a more detailed comparison of the various databases used in this work, using functional categories from GO as the reference.

It must be pointed out that about 75% of the proteins still lacking a functional assignment are hypothetical, putative, or predicted eukaryotic proteins. The average size of unassigned ensembles is 2.9, indicating that a significant fraction of them in fact contain only two/three proteins. Nearly two thirds of the proteins that we could not assign had no hits against the Pfam database. To a large extent, these sequences may actually be the result of noncodifying regions of the genome sequence that were erroneously interpreted, e.g., due to wrong positioning of introns [40].

In the following, we analyze three selected case studies. These cases were chosen on the basis of the different content of information available for each of them, to exemplify the degree of insight that can be reached in each case. In the first test case, the amount of available information is extensive also at the functional level. In the second case, at the time of preparation of this manuscript, there was good structural information but hardly any even indirectly relevant functional information; nevertheless, the information obtained from the analysis of gene organization features allowed us to obtain a detailed functional prediction. Similarly, for the third case study, we could obtain a detailed functional prediction, on the basis of the analysis of potential physiological protein partners.

ZIP proteins

Four archeal sequences contained the ZIP domain, which is characteristic of various transmembrane zinc transporters found in all domains of life [41–43]. ZIP proteins move zinc to the cytoplasm from the extracellular medium or from vacuoles, in contrast with the action of cation diffusion facilitator proteins, which mediate the reverse process. Through structure threading methods, we could model part of the NP_147044.1 sequence onto the PDB structure of a Cl^- transporter with 11 transmembrane regions (PDB code 1KPL [44]) (Fig. 3). In particular, although the *E* value (see "Materials and methods") is quite poor (4.0), the model can be used to interpret the common sequence properties of this ensemble of proteins. On the basis of the high conservation in sequence of the residues involved (Fig. 3b), we propose that Hx(3)Ex(29)H is the putative ZBP; note that the Glu ligand is either conserved or



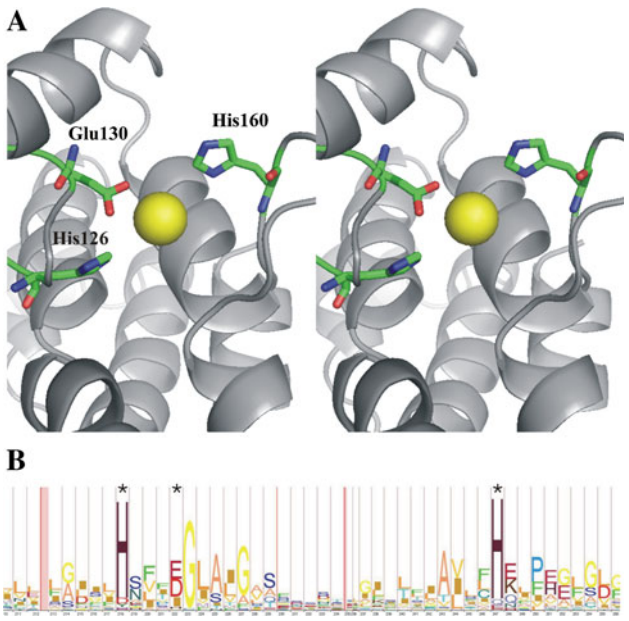


Fig. 3 **a** Protein structure threading onto 1KPL for the representative sequence NP_147044.1. The side chains of the amino acids in the proposed zinc-binding pattern Hx(3)Ex(29)H are shown. **b** The hidden Markov model (HMM) logo of the ZIP domain; the three residues are highlighted with *asterisks*

conservatively substituted by Asp. The present proposition is reinforced by the fact that these residues are close in space in the structural model (Fig. 3a). Note that although zinc binding by ZIP proteins has been established, the mode by which this is accomplished is still not fully supported by structural evidence; the present data therefore provide novel insight into the atomic-level features of the archaeal system.

A putative regulator of the metabolism of fatty acid

When we collected our data, for an ensemble of 27 archeal sequences we identified a homolog of known structure with PDB code 2GNR (from *Sulfolobus solfataricus*). This was the structure of a protein dimer solved at the Joint Center of Structural Genomics. The protein was described as having unknown function. Each subunit binds a zinc ion with the known ZBP Cx(2)Cx(10)Cx(2)C. Using STRING, we found a functional correlation between the target protein and acetyl-CoA acetyltransferase, the first enzyme in the fatty acid biosynthetic pathway. Other putative functional partners were 3-hydroxy-3-methylglutaryl-CoA synthetase, 3-hydroxyacyl-CoA dehydrogenase, 3-ketoacid-CoA

transferase, hydroxymethylglutaryl-CoA reductase, and pyruvate/ferredoxin oxidoreductase, which are all involved in the fatty acid anabolism. These enzymes are coded by genes in well-defined operons, upstream of which is located the codifying sequence of the protein under analysis here. Finally, the zinc-binding knuckle contained in the structure corresponds to a known zinc-finger fold, potentially able to bind DNA. Figure 4a shows the structure of the protein, with the two subunits in light blue and green. As the sequence logo shows, the four Cys in each subunit are either completely or very highly conserved (Fig. 4b). CastP [45] recognized a putative binding pocket; the residues involved are shown in red in Fig. 4a. The size and shape of the pocket were compatible with acetyl-CoA, the starting molecule for biosynthesis of fatty acids. All these hints allowed us to propose that these proteins are putative transcriptional factors regulating fatty acid metabolism, possibly responding to acetyl-CoA concentration. It has to be noted that, after the completion of this work, the reference PDB structure 2GNR was superseded by 3IRB (doi:10.2210/pdb3irb/pdb), classified as acyl-CoA binding protein. We considered the recent data as a validation of the proposed functional predictions.

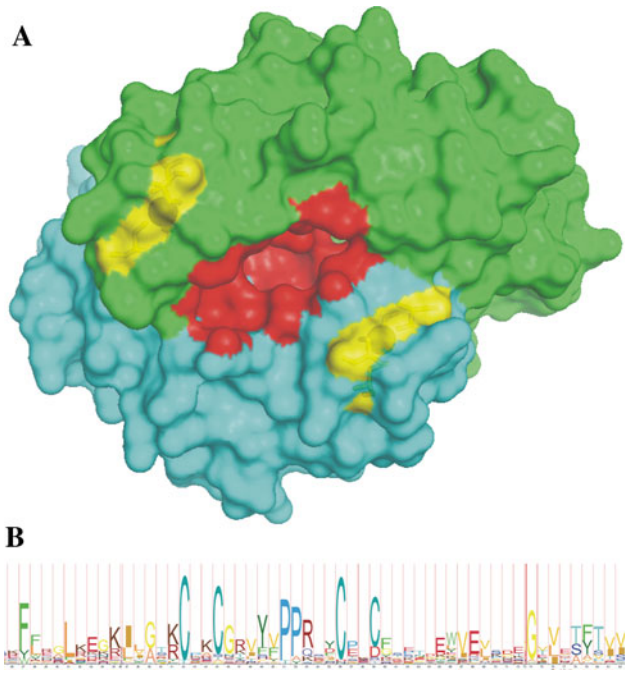


Fig. 4 **a** The homodimeric Protein Data Bank structure of 3IRB. The two monomers are in *light blue* and *green*, the eight Cys constituting the zinc-binding pattern are in *yellow*. The residues constituting the putative acetyl-CoA binding-pocket are reported in *red*. **b** HMM logo of the corresponding ensemble of proteins (only one monomer is shown)

A putative transcriptional factor regulating pilin biosynthesis

Twelve bacterial sequences had a homolog of known structure corresponding to PDB entry 1LV3 [46], which is a monomeric zinc-binding protein (YacG) with unknown function whose structure was solved by the Northeast Structural Genomics Consortium. In these sequences, we identified a DNA-binding zinc-finger domain that is present in known transcriptional factors. We now additionally propose that they are involved in the type II secretion system. Using the STRING tool, we identified various putative functional partners belonging to this system, such as PilA (pre-pilin), a precursor of type IV fimbrial pilin, PilB, and PilC, ATPases for PilA maturation and assembly, and PilD, a peptidase processing the N-terminal region of

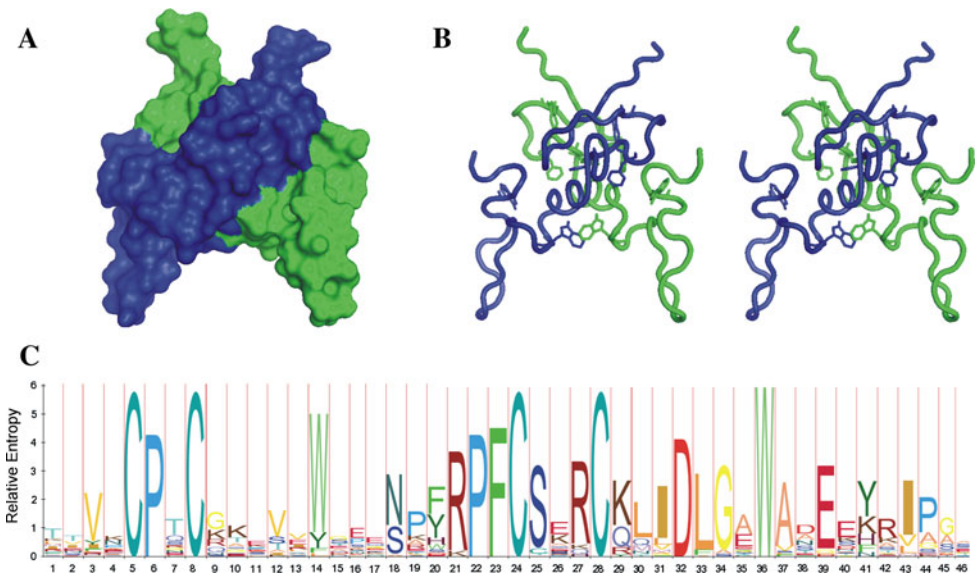


Fig. 5 Proposed model for dimerization for YacG based on the 1LV3 structure. **a** Protein surface, **b** protein backbone representation, showing the side chains of the highly conserved residues (stereoview), and **c** HMM logo of the corresponding ensemble of proteins

PilA. In archea and bacteria, polymers of type IV fimbrian pilin form flagella, for twitching motility, and f-pilus, for DNA transfer in processes such as conjugation, infection, and transformation. All the reported codifying sequences are contained in well-characterized operons, having the target sequence downstream. A model for the possibly functionally active YacG homodimer, which is commonly the oligomerization state for transcriptional factors, could be successfully built (Fig. 5). The HMM-logo in Fig. 5c shows that the zinc-binding residues are completely conserved.

Conclusions

Genome sequencing projects are continuously making new DNA sequences and potential protein sequences available. Several of these are not experimentally characterized, and thus a hypothetical function can be proposed only by functional prediction [47, 48]. Functional prediction can be a powerful and relatively high confidence method to this end, exploiting sequence and/or structural features [49, 50]. The composition of the analyzed sequences in terms of functional domain(s), their cellular localization, hints about functional partners, and conservation of residues among homologs are all important information allowing computational biologists to figure out hypothetical functions. The analysis of protein structures and protein surfaces can provide even more reliable and detailed hints [51].

There are computational approaches reported in the literature that can provide the metal proteome for each completely sequenced genome. We showed that this information can be complemented for the zinc proteome by an essentially complete functional annotation, again as the result of the systematic application of computational prediction methods. An experimental verification of these predictions is thus generally warranted, at least for selected cases of particular interest, where the purpose would be beyond the validation of the functional predictions proposed in this work. The annotations of the zinc proteomes of the 57 organisms analyzed are available at http://www.cerm.unifi.it/home/research/genomebrowsing.html.

References

1. Andreini C, Banci L, Bertini I, Rosato A (2006) J Proteome Res 5:3173–3178
2. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Nat Genet 25:25–29
3. Dobson PD, Cai YD, Stapley BJ, Doig AJ (2004) Curr Med Chem 11:2135–2142
4. Baker EN, Arcus VL, Lott JS (2003) Appl Bioinformatics 2:S3–10
5. Lee D, Redfern O, Orengo C (2007) Nat Rev Mol Cell Biol 8:995–1005
6. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR (2004) Nucleic Acids Res 32(Database issue):D138–D141
7. Sonnhammer EL, Eddy SR, Durbin R (1997) Proteins 28:405–420
8. Andreini C, Bertini I, Rosato A (2009) Acc Chem Res 42:1471–1479
9. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
10. Eddy SR (1998) Bioinformatics 14:755–763
11. Andreini C, Bertini I, Rosato A (2004) Bioinformatics 20:1373–1380
12. Andreini C, Banci L, Bertini I, Rosato A (2006) J Proteome Res 5:196–201
13. Castagnetto JM, Hennessy SW, Roberts VA, Getzoff ED, Tainer JA, Piquet ME (2002) Nucleic Acids Res 30:379–382
14. Coggill P, Finn RD, Bateman A (2008) Curr Protoc Bioinformatics 23:2.5.1–2.5.17
15. Zhang Y, Gladyshev VN (2009) Chem Rev 109:4828–4861
16. Tatusov RL, Koonin EV, Lipman DJ (1997) Science 278:631–637
17. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova RD, Koonin EV (2001) Nucleic Acids Res 29:22–28
18. Chen Y, Yu P, Luo J, Jiang Y (2003) Mamm Genome 14:859–865
19. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) J Mol Biol 305:567–580
20. Moller S, Croning MD, Apweiler R (2001) Bioinformatics 17:646–653
21. Sonnhammer EL, von Heijne G, Krogh A (1998) Proc Int Conf Intell Syst Mol Biol 6:175–182
22. Sprenger J, Fink JL, Teasdale RD (2006) BMC Bioinformatics 7(Suppl 5):S3
23. Liu J, Kang S, Tang C, Ellis LB, Li T (2007) Nucleic Acids Res 35:e96
24. Snel B, Lehmann G, Bork P, Huynen MA (2000) Nucleic Acids Res 28:3442–3444
25. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B (2003) Nucleic Acids Res 31:258–261
26. von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Kruger B, Snel B, Bork P (2007) Nucleic Acids Res 35:D358–D362
27. van Bakel H, Huynen M, Wijmenga C (2004) Bioinformatics 20:2644–2655
28. Galperin MY, Koonin EV (2000) Nat Biotechnol 18:609–613
29. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A (2007) Curr Protoc Protein Sci 50:2.9.1–2.9.31
30. Sali A, Potterton L, Yuan F, Van Vlijmen H, Karplus M (1995) Proteins Struct Funct Genet 23:318–326
31. Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008) Methods Mol Biol 426:145–159
32. Laskowski RA, Watson JD, Thornton JM (2005) Nucleic Acids Res 33:W89–W93
33. de Vries SJ, van Dijk AD, Bonvin AM (2006) Proteins 63:479–489
34. Neuvirth H, Raz R, Schreiber G (2004) J Mol Biol 338:181–199
35. Consortium The Uniprot (2007) Nucleic Acids Res 35:D193–D197

36. Leinonen R, Diez FG, Binns D, Fleischmann W, Lopez R, Apweiler R (2004) Bioinformatics 20:3236–3237
37. Bennett-Lovsey RM, Herbert AD, Sternberg MJ, Kelley LA (2008) Proteins 70:611–625
38. Kelley LA, Sternberg MJ (2009) Nat Protoc 4:363–371
39. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) Annu Rev Biophys Biomol Struct 29:291–325
40. Brent MR, Guigo R (2004) Curr Opin Struct Biol 14:264–272
41. Grotz N, Fox T, Connolly E, Park W, Guerinot ML, Eide D (1998) Proc Natl Acad Sci USA 95:7220–7224
42. Eide DJ (2006) Biochim Biophys Acta 1763:711–722
43. Gaither LA, Eide DJ (2001) Biometals 14:251–270
44. Dutzler R, Campbell EB, Cadene M, Chait BT, MacKinnon R (2002) Nature 415:287–294
45. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) Nucleic Acids Res 34:W116–W118
46. Ramelot TA, Cort JR, Yee AA, Semesi A, Edwards AM, Arrowsmith CH, Kennedy MA (2002) Proteins 49:289–293
47. Godzik A, Jambon M, Friedberg I (2007) Cell Mol Life Sci 64:2505–2511
48. Baker D, Sali A (2001) Science 294:93–96
49. Pandit SB, Bhadra R, Gowri VS, Balaji S, Anand B, Srinivasan N (2004) BMC Bioinformatics 5:28
50. Madera M (2008) Bioinformatics 24:2630–2631
51. Serres MH, Riley M (2004) OMICS 8:306–321