Exploitation of gene context

Martijn Huynen*†‡, Berend Snel*, Warren Lathe III*‡ and Peer Bork*‡

Recently, a number of techniques have been proposed that use completely sequenced genomes for the function prediction of individual proteins encoded therein. They use the fusion of genes, their conserved location in operons or merely their co-occurrence in genomes to predict the existence of functional interactions between the proteins they encode. This type of information complements functional features that are predicted by classical homology-based search techniques.

Addresses
*European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany
†e-mail: huynen@embl-heidelberg.de
‡Max-Delbrück-Centrum for Molecular Medicine, 13122 Berlin-Buch, Germany

Current Opinion in Structural Biology 2000, **10**:366–370

0959-440X/00/$ – see front matter

Introduction

The sequencing of complete genomes has created the possibility to not only analyze a protein's role within the context in which it occurs, but also predict protein function from the genomic context of its genes. Here, we review three formalizations of the usage of genomic context in the prediction of gene function. They all use co-localization of genes on the genome, but vary in the physical proximity that is required, using either the fusion of genes [1•,2•], the local neighborhood of genes [3•,4•] or merely the presence of genes in the same genome [5•,6•]. All methods depend on the comparison of multiple genomes to increase the selectivity of their predictions; the number of genomes that are required increases with a decrease in the constraints on the physical proximity of the genes. In contrast to classical homology-based function prediction, methods based on genomic context do not directly predict the function of gene products, but rather the existence of functional interactions between gene products.

Types of genomic context

Gene fusion

During evolution, genes can fuse with each other into a larger, 'composite' gene or fall apart (gene fission) into smaller, 'component' genes. The observation that two genes also occur in parallel as a larger composite gene (also called 'Rosetta stone sequence') is an indication that they are functionally related and can thus be used to predict a functional relation between them [1•,2•]. The single occurrence of a composite gene in one genome is regarded as sufficient to infer a functional interaction between its components in other genomes. As one might expect, the large majority of composite genes result from gene fusion [7]. The process of gene fission occurs mainly in thermophiles [7].

Local gene neighborhood

In bacteria and archaea and, incidentally, in eukaryotes [8], genes are transcribed into polycistronic mRNAs encoding co-regulated and functionally related genes. This concept has been used in numerous cases in which gaps in metabolic pathways could be filled by hypothetical genes encoded in the operon for that pathway (e.g. [9]). In computational genomics, operon structure is inferred from the conserved physical proximity of genes in phylogenetically distant genomes (small subunit ribosomal RNA identity <88% [10]). Such conservation has been measured in two ways: via the conservation of pairs of genes as neighbors in all the genomes compared [3•,11,12] and via the co-localization of genes in potential operons in at least two of the genomes compared [13]. The latter does not specifically measure conservation; it thereby dramatically increases the number of genes for which functional relations can be predicted [13].

Note that it is not a requirement that the genes for which a functional relation is predicted are located in an operon in the specific genome that is being studied. As is the case in gene fusion, one can infer a functional relation from their physical proximity in other genomes. This provides an opportunity to predict also functional relations among genes from species with little or no operon structure, such as eukaryotes, at least to the extent that such species have orthologs in prokaryotes.

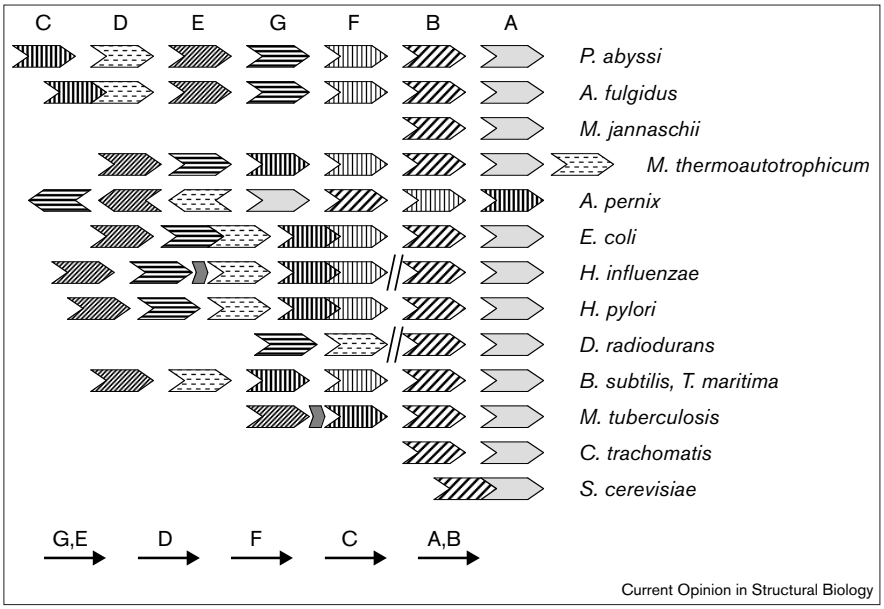
Co-occurrence of genes in genomes

The co-occurrence of genes in genomes [5•,6•] (also called 'phylogenetic profiles' or 'phylogenetic signatures' [14]) is the most general form of genomic context. The idea is that genes that are functionally related tend to occur together in the same genome, irrespective of their relative location in that genome, or be absent together. Conceptually, this method is the most powerful, as it is not obvious that all proteins that functionally interact have genes that are either fused or located in operons with each other. Methodologically, however, it appears to be the most cumbersome, as the signal-to-noise ratio per genome is very small and accurate comparisons of large numbers of genomes are required.

Types of functional interaction

As the function of a protein is not a well-defined concept that can be described at many levels [15], the issue of what constitutes a functional interaction between proteins is probably even harder to tackle. In analyses of gene context, the types of functional interaction are generally divided into physical interaction, metabolic and regulatory pathways, and biological process, which includes functional interactions not caught by the first

Figure 1



Operon organization and fusion of genes (*trpA–G*) involved in tryptophan synthesis in published genomes. The small open reading frames (ORFs) (dark gray) are not conserved with tryptophan genes in one operon. Fusion is depicted by the absence of an intergenic region, vertical lines indicate separate operons. (Gene names are based on *Pyrococcus abyssi,* in which all the genes are separate and in one operon.) The arrows at the bottom give a schematic representation of the tryptophan synthesis pathway, with the genes coding for the enzymes that catalyze the steps.

two types. The types of interactions that have been reported depend on the database that is used to measure them, like WIT (metabolic pathways; http://wit.mcs.anl.gov/) [16] or DIP (physically interacting proteins; http://dip.doe-mbi.ucla.edu) [17]. Systematic 'manual' analysis of the functional relations among proteins has only been done for limited sets of proteins [2•,3•]. A quantification of which type of genomic context correlates with which type of functional interaction has not been reported.

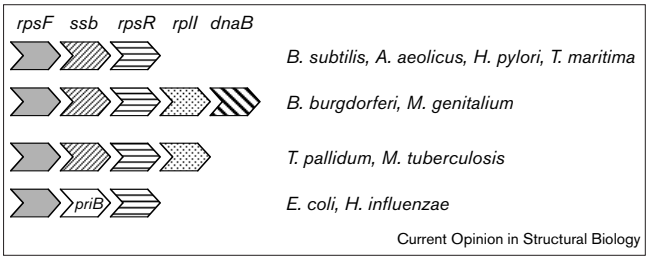
Physical interaction

Physical interaction between proteins appears to be especially dominant if the gene order is conserved among all the species compared [3•,12]. Also, within a set of functionally interacting genes, such as those encoding the proteins in the tryptophan operon, the strongest conserved aspect is that *trpA* and *trpB*, which encode physically interacting proteins, are neighbors on the genome. The arrangement of the other genes within this set, even if they are together in one operon, varies (Figure 1). The conservation of the relative order of the transcription of pairs of genes (e.g. *trpA* and *trpB*) has led to speculation that the physical interaction might start during or immediately after translation [3•]. This can, however, only be part of the explanation; in the relatively well-conserved ribosomal protein operons, for example, the conserved gene pairs do not separate the small and large subunits.

Metabolic pathways

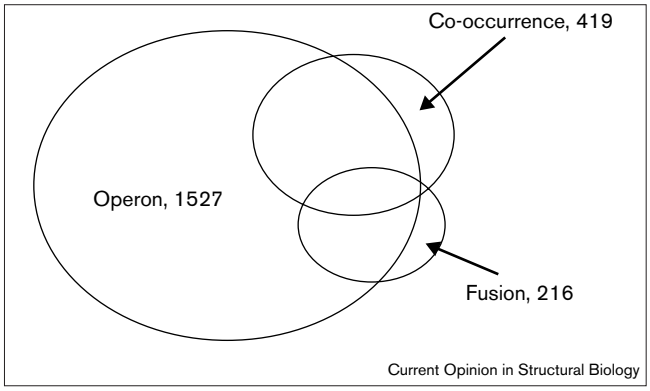
In order to predict functional interactions among a set of genes, for example, in predicting a metabolic pathway, one has to show that all the genes in the set are coupled to each other, either directly or indirectly, but not to any other genes. One example of how gene neighborhood predicts the complete set of genes involved in a pathway is illustrated by the tryptophan synthesis pathway (Figure 1). Similarly, the genes in the purine synthesis pathway could be detected by correlations in gene neighborhood [4•], albeit with one 'false positive', a regulatory gene. At the moment, using gene fusion alone does not reveal the complete tryptophan synthesis pathway, but instead reveals three subpathways. Similarly, the purine and shikimate biosynthesis pathways have been shown to be reflected in gene fusion events [1•], but not to the extent that all the genes could be linked to each other. An increase in the number of sequenced genomes might solve this, but it is not inconceivable that the specific constraints on the function of proteins (e.g. the interaction of TrpA with TrpB) might prevent their genes from fusing with any other genes. The co-occurrence of genes cannot, as of yet, distinguish between various amino acid synthesis pathways [6•], but again this might be a matter of increasing the number of sequenced genomes. Notice, however, that the distribution of the tryptophan biosynthesis genes shows variation among published genomes; *trpD* and *trpE* are absent from *Chlamydia trachomatis* [18], even though the other genes are present. The degree to which pathways can be detected by analysis of genomic context depends on their variation among species, which, in turn, appears to be a function of their linearity and interconnectedness with other pathways. A pathway such as glycolysis is 'well-connected' and can, so far, only partly be detected by gene neighborhood [4•]. Finally, of the genes coding for the eight enzymes of the citric acid cycle, which displays a large amount of variation among the sequenced prokaryotes, only two occur repeatedly in a single operon [19•].

Figure 2



The unexpected conservation of proteins involved in translation and proteins involved in DNA replication in one operon. The gene for single stranded DNA binding protein (*ssb*) and *dnaB* occur in conjunction with ribosomal proteins *rpsF*, *rpsR* and *rplI*. *Prima facie*, there is no functional relation between *ssb* and *dnaB,* which are mainly involved in DNA replication, and ribosomal proteins. However, *rpsF* has been shown to be a supressor of a heat-sensitive mutation in *priB*, which codes for RNA primase [36], with which the ribosomal genes occur in *E. coli* and *H. influenzae*, indicating that *rpsF* is involved in (regulating) DNA replication.

Figure 3



Coverage and overlap of the three types of genomic context for the genes of the *E. coli* genome and its orthologs in other genomes. For gene 'fusion', a single occurrence of a 'composite protein' or 'Rosetta stone sequence' was regarded significant. For 'operon', the co-localization of two genes within one potential operon in at least two genomes was considered significant, including the situation in which they were fused in one of the genomes, but not the other. For 'co-occurrence', only proteins with a nonphylogenetic pattern of occurrence were considered: that is, they were not specific for either *E. coli*, or for *E. coli* and *H. influenzae*, or for *E. coli*, *H. influenzae* and *R. prowazekii*, or for *E. coli*, *H. influenzae*, *R. prowazekii* and *H. pylori*, or for the bacteria among the set of genomes published up to January 1st 2000. Furthermore, genes were required to minimally occur five times and maximally 20 times (in 25 genomes) for their pattern of co-occurrence to be regarded significant. Orthology was determined using the pairwise, mutual best-hit approach, with an E-value cut-off of 0.01 [5•]. The numbers in the figure are the numbers of *E. coli* genes for which a specific type of genomic context could be detected.

Sensitivity and selectivity

Homology versus orthology

The transfer of protein function is more reliable when it is done between orthologs, rather than between sequences that are merely homologs of each other. However, although the presence of large databases of orthologous proteins, such as COG (http://www.ncbi.nlm.nih.gov/COG/) [20•], might suggest otherwise, to predict the presence of a pathway in a species, orthology prediction has to be done manually [19•]. Methods that rely on homology, rather than orthology [1•,6•] have the advantage that they cover a larger set of proteins, although the selectivity of the predictions is lower [2•]. The analysis of gene fusion of homologous proteins ('domain fusion' [1•]) borders on the analysis of the co-occurrence of domains in proteins, which can be done using databases like SMART (http://bork.embl-heidelberg.de) [21] and PFAM (http://www.sanger.ac.uk/Software/Pfam/) [22].

Counting false positives

In the absence of a quantitative model for the evolution of genomes and the functional relations among proteins, the quality of the prediction of functional interaction is generally established by counting the fraction of false positives, that is, gene pairs in which the function of each of the genes is known, but the genes do not have a described functional relation. Such an estimate is always a maximum; the relations might exist, but have not been discovered yet or they might not have been caught by the databases of functional relations used. For example, linking ribosomal proteins to proteins involved in DNA replication would, in general, be regarded as a false positive; there is, however, some experimental evidence that some ribosomal proteins play a (regulatory) role in DNA replication (Figure 2). Nevertheless, even though they can only be regarded as maxima, the reported fractions of false positives are very high, often reaching levels of 30% or higher [2•,23•]. Only the strict conservation of gene order in multiple genomes reported low fractions of false positives [3•,12]

Quantitative coverage

We compared the three types of context for their coverage of the *Escherichia coli* genome (Figure 3), including only orthologous relations. Quantitatively, the usage of local neighborhood appears to be the most powerful method, although the cut-off for what is considered to be a significant pattern of co-occurrence is arguably subjective.

Variation of functional interactions

The transfer of functional interactions between genes from one genome to another assumes that those interactions do not change. This assumption is not always correct. To give one example, the iron–sulfur subunit of fumarate reductase is, in *Methanobacterium thermoautotrophicum*, fused to a heterodisulfide reductase that oxidizes coenzyme M and coenzyme B, providing the electrons for the reduction of fumarate [24]. On the basis of this fusion, these proteins have been predicted to interact in *E. coli*, where they occur as separate genes [2•]. This is unlikely, however, as the electrons for the reduction of fumarate in *E. coli* are derived from quinol in the membrane, rather than from coenzyme M and coenzyme B, which appear to be absent from this species. In a species more closely related to the methanogens, *Sulfolobus acidocaldarius*, the heterodisulfide

reductase occurs in the same operon as an ortholog of fumarate reductase, with which it has the same functional interaction as in *M. thermoautotrophicum*, although the direction of the reaction is reversed [25]. Fumarate reductase occurs fused to or in an operon with heterodisulfide reductase in some archaea and with membrane-binding subunits in one operon in bacteria. In both cases, there is a functional interaction between the proteins in terms of electron transfer. This illustrates that gene context reflects the evolution of functional interactions and can thus be used to delineate the evolution of pathways [10].

Also, genes that occur in the same operon can be related in a very species-specific manner; for example, genes of the glycolysis pathway are co-regulated or occur in an operon with genes involved in translation [26,27]. Conservation of physical proximity is thus required, not only to find operons, but also to ensure that the functional relation is not species-specific. Even then, however, 'false positives' can appear, like the location of *dnaE* (DNA polymerase III) in an operon with *pfk* and *pyk*, which code for enzymes in the glycolysis pathway, in two genomes [4•]. The abundance of neighborhood information allows one to rate the functional relatedness of two genes on the basis of how often they occur in each other's neighborhood in phylogenetically distant genomes [4•].

The co-occurrence of genes exploits the conservation of functional relations among genes to the fullest and will not give predictions at all for functional relations among genes that change in evolution, unless it is only applied to a subset of the genomes. Indeed, for this method, genomic divergence appears to be the strongest requirement, as there is a strong phylogenetic signal in the number of genes genomes share [28,29]. Although such 'taxon-specific' genes might fall into separate functional classes or only one, there is no *a priori* expectation that they do.

Combining various types of context – information from homology

Although quantitative trends can be discriminated in which types of functional relation seem to correlate with types of genomic context (M Huynen, unpublished data), these are not so strong that one can, on the basis of context alone, predict with any reliability what type of functional interaction exists between two proteins, except in the case of conserved gene order among multiple genomes. In general, the combination of context analysis with homology-based function prediction appears to be very useful, as the latter predicts the molecular function of proteins. Thus, combining context analysis with homology searches one might be able to predict both the pathway in which an enzyme operates and its molecular function [19•,30]. Various types of context analysis can also be combined with each other to increase the reliability of the predictions and to test the validity of the assumptions of either method [23•]. For example, the extent to which adjacent genes are coexpressed in a species like *Saccharomyces cerevisiae* can be assessed by comparisons with expression data [31]. Information about coexpression can also be derived from a comparison of promotor sequences, leading to interesting new predictions of a possible functional interaction between heat-shock genes and histone genes in archaea [32•].

Conclusions

Methods that predict functional interactions between genes on the basis of their genomic context can not only lead to the discovery of new functional relations, such as physical interactions and being active in the same metabolic pathway, but can also lead to an objective definition and demarcation of cellular processes. In the absence of a statistical model for the evolution of genomes and functional relations, heuristic methods that incorporate reliability scores for the different types of genomic context will have to be developed to reduce the fraction of false positives. Quantitatively, local gene neighborhood appears to be the most powerful type of information, but its role is restricted to proteins that have orthologs in bacteria and archaea. The combination of various context-based predictions of functional relations with each other and with other types of information, like homology searches, provides the most valuable information. The total number of papers that review these methods (eight) — [10,15,33,34], two papers in this issue and two others to appear soon (S Teichmann, personal communication; E Koonin, personal communication) — already outnumbers the total number of predictions that were shown to be correct (one) [6•,35] (although it is hard to disprove a functional interaction), but it is the latter that will finally decide on the value of these methods to experimental genomics.

Update

We have introduced a web server to facilitate the search for the conservation of the local neighborhood of genes: STRING (Search Tool for Recurring Instances of Neighboring Genes). It detects for a given query gene whether it repeatedly occurs with certain other genes in potential operons (http://www.Bork.EMBL-Heidelberg.DE/STRING/).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Marcotte EM, Pellegrini M, Ng H, Rice WD, Yeates TO, Eisenberg D:
• **Detecting protein function and protein–protein interactions from genome sequences.** *Science* 1999, **285**:751-753.
The authors introduce the idea of inferring functional interactions among genes on the basis of their fused occurrence. The paper only considers homologous relations. The paper shows that metabolic pathways are reflected in gene fusion events.

2. Enright A, Ilipoulos I, Kyrpides N, Ouzounis C: **Protein interaction**
• **maps for complete genomes based on gene fusion events.** *Nature* 1999, **402**:86-90.
A smaller scale approach to that described in [1•], this method distinguishes between orthologs and paralogs, and uses a manual check of the functional interactions

3. Dandekar T, Snel B, Huynen M, Bork P: **Conservation of gene order:**
• **a fingerprint of proteins that physically interact.** *Trends Biochem Sci* 1998, **23**:324-328.
A comparison of three sets of three genomes shows that genes whose order is conserved within the genome tend to code for proteins that physically

interact with each other. The paper predicts interactions for some hypothetical proteins on the basis of conserved gene order.

4. • Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N: **The use of gene clusters to infer functional coupling.** *Proc Natl Acad Sci USA* 1999, **96**:2896-2901.
A method for the large-scale detection of genes that occur together in operons. The method focuses on predicting metabolic pathways and shows that, at the moment of publication, 35% of the genes coding for enzymes in pathways were co-localized in potential operons with other enzymes coding for that same pathway in sequenced genomes. The method is distinct from that described in [3•] in that it does not require conservation in multiple genomes (two is enough), which is responsible for the higher fraction of genes coding for proteins that are involved in the same pathway relative to proteins that physically interact than in [3•].

5. • Huynen MA, Bork P: **Measuring genome evolution.** *Proc Natl Acad Sci USA* 1998, **95**:5849-5856.
The first large-scale analysis comparing the rates of evolution of various aspects of the genome. The authors introduce the idea of detecting functional interactions among genes from their co-occurrence in genomes.

6. • Pellegrini M, Marcotte EM, Thompson JM, Eisenberg D, Yeats TO: **Assigning protein functions by comparative genome analysis: protein phylogenetic profiles.** *Proc Natl Acad Sci USA* 1999, **96**:4285-4288.
A large-scale implementation of the idea that genes that have the same phylogenetic distribution are functionally related. The paper includes some functional predictions, one of which was shown to be correct, and analyses of correlations in the phylogenetic distribution of proteins that functionally interact.

7. Snel B, Bork P, Huynen M: **Genome evolution: gene fusion versus gene fission.** *Trends Genet* 2000, **16**:9-11.

8. Blumenthal T: **Gene clusters and polycistronic transcription in eukaryotes.** *Bioessays* 1998, **20**:480-487.

9. Nassau P, Martin SL, Brown R, Weston A, Monsey D, McNeil M, Duncan K: **Galactofuranose biosynthesis in *Escherichia coli* K-12: identification and cloning of UDP-galactopyranose mutase.** *J Bacteriol* 1996, **178**:1047-1052.

10. Huynen MA, Snel B: **Gene and context: towards an integrative approach to genome analysis.** In *Analysis of Amino Acid Sequences*, vol 10. Edited by Bork P. San Diego: Academic Press; 2000:345-379.

11. Tamames J, Casari G, Ouzounis C, Valencia A: **Conserved clusters of functionally related genes in two bacterial genomes.** *J Mol Evol* 1997, **44**:66-73.

12. Mushegian AR, Koonin EV: **Gene order is not conserved in bacterial evolution.** *Trends Genet* 1996, **12**:289-290.

13. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N: **Use of contiguity on the chromosome to predict functional coupling.** *In Silico Biology* 1998, **1**:9. [URL: http://www.bioinfo.de/isb/1998/01/0009/]

14. Gaasterland T, Ragan MA: **Microbial genescapes: phyletic and functional patterns of ORF distribution among prokaryotes.** *Microb Comp Genomics* 1998, **3**:199-217.

15. Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M, Yuan Y: **Predicting function: from genes to genomes and back.** *J Mol Biol* 1998, **283**:707-725.

16. Selkov E, Grechkin Y, Mikhailova N, Selkov E: **MPW: The Metabolic Pathways Database.** *Nucleic Acids Res* 1998, **26**:43-45.

17. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D: **DIP: the database of interacting proteins.** *Nucleic Acids Res* 2000, **28**:289-291.

18. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q *et al.*: **Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis.*** *Science* 1998, **282**:754-759.

19. • Huynen MA, Dandekar T, Bork P: **Variation and evolution of the citric acid cycle: a genomic perspective.** *Trends Microbiol* 1999, **7**:281-291.
Comparative analysis of the citric acid cycle via the gene content of the published genomes. By including the gene context and other information about the species compared, the authors solve some of the pitfalls in orthology prediction to predict pathways. The authors reconstruct the evolution of the citric acid cycle in archaea.

20. • Tatusov R, Galperin MY, Natale D, Koonin EV: **The COG database: a tool for genome-scale analysis of protein functions and evolution.** *Nucleic Acids Res* 2000, **28**:33-36.
The latest article describing clusters of orthologous groups (COGs). This paper shows the uses of COGs not only for conventional genome annotation, but also points out how 'phylogenetic patterns of COG distribution' can be used for more advanced analysis. The COG web page (http://www.ncbi.nlm.nih.gov) allows the automatic extraction of groups of orthologous genes with a similar COG distribution pattern.

21. Schultz J, Copley R, Doerks T, Ponting CP, Bork P: **Smart: a web-based tool for the study of genetically mobile domains.** *Nucleic Acids Res* 2000, **28**:231-234.

22. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL: **The pfam protein families database.** *Nucleic Acids Res* 2000, **28**:263-266.

23. • Marcotte E, Pellegrini M, Thompson M, Yeates T, Eisenberg D: **A combined algorithm for genome-wide prediction of protein function.** *Nature* 1999, **402**:83-86.
The authors describe an approach that combines different methods to predict functional links among proteins. Using this combination, high coverage and increased confidence can be attained. All the computational predictions rely on homologs. Potential biologically relevant links for several proteins are described.

24. Heim S, Kunkel A, Thauer RK, Hedderich R: **Thiol:fumarate reductase (tfr) from *Methanobacterium thermoautotrophicum* – identification of the catalytic sites for fumarate reduction and thiol oxidation.** *Eur J Biochem* 1998, **253**:292-299.

25. Janssen S, Schafer G, Anemuller S, Moll R: **A succinate dehydrogenase with novel structure and properties from the hyperthermophilic archaeon *Sulfolobus acidocaldarius*: genetic and biophysical characterization.** *J Bacteriol* 1997, **179**:5560-5569.

26. Kromer W, Arndt E: **Halobacterial S9 operon. Three ribosomal protein genes are cotranscribed with genes encoding a tRNA(Leu), the enolase, and a putative membrane protein in the *Archaebacterium haloarcula* (halobacterium) *marismortui.*** *J Biol Chem* 1991, **266**:24573-24579.

27. Tornow J, Santangelo GM: **Efficient expression of the *Saccharomyces cerevisiae* glycolytic gene ADH1 is dependent upon a cis-acting regulatory element (UASRPG) found initially in genes encoding ribosomal proteins.** *Gene* 1990, **90**:79-85.

28. Snel B, Bork P, Huynen M: **Genome phylogeny based on gene content.** *Nat Genet* 1999, **21**:108-110.

29. Huynen MA, Snel B, Bork P: **Lateral gene transfer, genome surveys and the phylogeny of prokaryotes.** *Science* 1999, **286**:1441a.

30. Galperin M, Koonin EV: **Sequence analysis of an exceptionally conserved operon suggests enzymes for a new link between histidine and purine biosynthesis.** *Mol Microbiol* 1997, **24**:443-445.

31. Kruglyak S, Tang H: **Regulation of adjacent yeast genes.** *Trends Genet* 2000, **16**:109-111.

32. • Gelfand M, Koonin EV, Mironov AA: **Prediction of transcription regulatory sites in archaea by a comparative genomic approach.** *Nucleic Acids Res* 2000, **28**:695-705.
The authors attempt to detect regulatory sites in archaea using profiles of known sites and comparative genomics. Their method proves to be able to predict new members for several regulons.

33. Doolittle RF: **Do you dig my groove?** *Nat Genet* 1999, **23**:6-8.

34. Sali A: **Functional links between proteins.** *Nature* 1999, **402**:23-24.

35. Karzai A, Susskind MM, Sauer RT: **SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of Ssra (tmRNA).** *EMBO J* 1999, **18**:3793-3799.

36. Britton R, Lupski JR: **Isolation and characterization of suppressors of two *Escherichia coli* dnaG mutations, dnaG2903 and parB.** *Genetics* 1997, **145**:867-875.