Available online at www.sciencedirect.com
ScienceDirect
Gene xx (2007) xxx–xxx
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism

Mikhail Pachkov [a,b,*], Thomas Dandekar [c], Jan Korbel [d,e], Peer Bork [d,f], Stefan Schuster [a]

[a] *Department of Bioinformatics, Faculty of Biology and Pharmaceutics, Friedrich-Schiller University Jena, Ernst-Abbe-Platz 2, D-07743 Jena, Germany*
[b] *Division of Bioinformatics, Biozentrum, University of Basel, Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland*
[c] *Department of Bioinformatics, Biozentrum, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany*
[d] *European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany*
[e] *Yale University School of Medicine, Molecular Biophysics and Biochemistry Department, 266 Whitney Ave., PO Box 208114, New Haven, CT 06520, USA*
[f] *Max Delbrück Centre for Molecular Medicine, D-13092 Berlin, Germany*

Received 8 July 2005; received in revised form 26 November 2006; accepted 21 February 2007

Received by F. Salvatore

Abstract

Elementary modes analysis allows one to reveal whether a set of known enzymes is sufficient to sustain functionality of the cell. Moreover, it is helpful in detecting missing reactions and predicting which enzymes could fill these gaps. Here, we perform a comprehensive elementary modes analysis and a genomic context analysis of *Mycoplasma pneumoniae* nucleotide metabolism, and search for new enzyme activities. The purine and pyrimidine networks are reconstructed by assembling enzymes annotated in the genome or found experimentally. We show that these reaction sets are sufficient for enabling synthesis of DNA and RNA in *M. pneumoniae*. Special focus is on the key modes for growth. Moreover, we make an educated guess on the nutritional requirements of this micro-organism. For the case that *M. pneumoniae* does not require adenine as a substrate, we suggest adenylosuccinate synthetase (EC 6.3.4.4), adenylosuccinate lyase (EC 4.3.2.2) and GMP reductase (EC 1.7.1.7) to be operative. GMP reductase activity is putatively assigned to the NRDI_MYCPN gene on the basis of the genomic context analysis. For the pyrimidine network, we suggest CTP synthase (EC 6.3.4.2) to be active. Further experiments on the nutritional requirements are needed to make a decision. Pyrimidine metabolism appears to be more appropriate as a drug target than purine metabolism since it shows lower plasticity.

Keywords: Elementary modes; Functional genomics; Genome context methods; *Mycoplasma pneumoniae*; Nucleotide metabolism; STRING

1. Introduction

In the post-genomic era, a major challenge is to complete the annotation of genomes and to analyze the complex interactions of gene products. Such an analysis can improve the results of functional genomics (Dandekar et al., 1999; Schilling et al., 1999; Förster et al., 2002). In the case of metabolic genes, it is promising to study the interactions of enzymes within metabolic pathways. This can be done by metabolic pathway analysis (Mavrovouniotis et al., 1990; Schuster and Hilgetag, 1994; Schilling et al., 1999, 2000; Schuster et al., 2002a,b; Klamt and Stelling, 2003; Schuster and Kenanov, 2005). A central concept in metabolic pathway analysis is that of elementary flux modes, that is, minimal sets of enzymes that can operate together at steady state such that the directionality of all irreversible enzymes involved is complied with (Schuster and Hilgetag, 1994; Schuster et al., 1999, 2002a). The elementary modes approach has been applied to several subsystems of bacteria (Schuster et al., 2002b; Carlson and Srienc, 2004), yeast (Förster et al., 2002; Schwartz and Kanehisa, 2006) and higher organisms (Poolman et al., 2003; Schwender

Abbreviations: CTP, cytidine triphosphate; EC, Enzyme Catalogue; GMP, guanosine monophosphate; IMP, inosine monophosphate; ORF, open reading frame; Pi, inorganic phosphate; PEP, phosphoenolpyruvate; PPi, pyrophosphate; PRPP, 5-phospho-α-D-ribose-1-diphosphate; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; KEGG, Kyoto Encyclopedia of Genes and Genomes; R1P, α-D-ribose-1-phosphate; XMP, xanthosine monophosphate.

* Corresponding author. Division of Bioinformatics, Biozentrum, University of Basel, Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland. Tel.: +41 61 267 15 75; fax: +41 61 267 15 84.
E-mail address: mikhail.pachkov@unibas.ch (M. Pachkov).

0378-1119/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.02.033

et al., 2004; Schuster and Kenanov, 2005). A similar concept has been termed extreme pathways (Schilling et al., 2000; Schilling and Palsson, 2000; Price et al., 2002). However, the elementary modes approach has a number of advantages in comparison to extreme pathways (Klamt and Stelling, 2003).

Elementary modes analysis is useful in functional genomics for the following reason. As the set of identified enzyme genes is often incomplete, it is sensible to test whether at least one functional route is realizable by this set (Dandekar et al., 1999). If no pathway is found connecting some substrate to some product for which a conversion is found, for example, in experiment, then there must be a gap in the known metabolic network, that is, an enzyme gene that has not yet been annotated or detected. In mycoplasmas, several enzymatic activities found in biochemical assays could not be assigned to any primary gene sequence, while other enzymatic activities predicted on the basis of ORF sequences could not be detected experimentally (Pollack, 1997).

An approach to detecting pathway gaps based on databases and automated bioinformatics technology – the "pathway/genome database (PGDB)" and "PathoLogic" – has been developed and used (Green and Karp, 2004; Romero et al., 2004). Here, we choose to employ elementary-modes analysis because it is a very transparent method, is easy to implement and fully takes into account the mass balance of all co-substrates and by-products of a pathway (Schuster et al., 2002a).

In order to fill presumed pathway gaps, we also apply a set of computational methods that utilize evolutionary signals across various completely sequenced genomes for function prediction: the so-called genomic context methods. These analyze gene neighborhood (operon-like structures), the evolutionary conservation of predicted operons, gene fusion events, and co-occurrence of genes across genomes (correlating gene presence and absence) to predict functional associations between a query gene and other genes (Huynen et al., 2000; Osterman and Overbeek, 2003; Green and Karp, 2004). Predicted functional associations may indicate physical interactions between the encoded proteins, or the involvement in the same cellular process or biochemical pathway (Huynen et al., 2003). We use the STRING server (von Mering et al., 2003a,b, 2005) to identify functional relationships from the genomic context, using the option of including functional associations inferred from DNA microarray experiments, which have been performed in various species (von Mering et al., 2005).

Mycoplasma pneumoniae possesses one of the smallest genomes in self-replicating organisms and is a good candidate for modeling a minimal self-replicating cell (Morowitz and Tourtellotte, 1962; Hutchison et al., 1999). Several *Mycoplasma*

Table 1
Reactions included in the model of pyrimidine metabolism

EC number	Reaction	Reversibility	Gene[a]
2.7.1.11/40, 2.7.2.3	ATP+**dTDP**=ADP+**dTTP**	REV	K6PF/KPYK
	ATP+**dCDP**=ADP+**dCTP**	REV	PGK
	ATP+**UDP**=ADP+**UTP**	REV	
	ATP+**CDP**=ADP+**CTP**	REV	
2.7.4.14	ATP+**UMP**=ADP+**UDP**	IRREV	KCY
	ATP+**CMP**=ADP+**CDP**	IRREV	
	ATP+dCMP=ADP+**dCDP**	IRREV	
2.7.1.48	UTP+**Uridine**=UDP+**UMP**	REV	URK
	ATP+Cytidine=ADP+**CMP**	REV	
3.5.4.5	Cytidine+H_2O=**Uridine**+NH_3	IRREV	CDD
	Deoxycytidine+H_2O=**Deoxyuridine**+NH3	IRREV	
1.17.4.1	**dCDP**+Oxidized thioredoxin+H_2O=Thioredoxin+**CDP**	REV	RIR1/2
	dUDP+Oxidized thioredoxin+H_2O=Thioredoxin+**UDP**	REV	
2.4.2.4	**Deoxyuridine**+Pi=Uracil+2-Deoxy-R1P	REV	TYPH
	Thymidine+Pi=Thymine+2-Deoxy-R1P	REV	
2.4.2.1	**Deoxyuridine**+Pi=Uracil+2-Deoxy-R1P	REV	DEOD
2.7.1.21	ATP+**Deoxyuridine**=ADP+**dUMP**	REV	KITH
	ATP+**Thymidine**=ADP+**dTMP**	REV	
2.7.4.9	ATP+**dUMP**=ADP+**dUDP**	REV	KTHY
	ATP+**dTMP**=ADP+**dTDP**	REV	
	ATP+**Thymidine**=ADP+**dTMP**	REV	
2.7.7.7	**dCTP**+DNA=PPi+DNA	IRREV	DPO3A/B
	dTTP+DNA=PPi+DNA	IRREV	DPO3(X)
2.7.7.6	**UTP**+RNA=PPi+RNA	IRREV	RPOA RPOB
	CTP+RNA=PPi+RNA	IRREV	RPOC
2.1.1.45	**dUMP**+5,10-Methylenetetrahydrofolate=Dihydrofolate+**dTMP**	REV	TYSY
2.7.1.74	Deoxycytidine+ATP=**dCMP**+ADP	REV	[b]
2.4.2.3	**Uridine**+Pi=Uracil+R1P	REV	[b]
2.4.2.9	**UMP**+PPi<=>Uracil+PRPP	REV	UPP

Information on the presence of enzymes in *M. pneumoniae* was taken from KEGG database (http://www.genome.ad.jp/kegg/) and cross-checked with data from "The *M. pneumoniae* Genome Project" (http://www.zmbh.uni-heidelberg.de/M_pneumoniae/genome/Results.html). Information about reversibility/irreversibility was obtained from the EMP database (http://www.empproject.com/). Bold symbols indicate internal metabolites.

[a] Gene names are given according Swissprot classification. The suffices "_MYCPN" are omitted.

[b] Genes not yet found in the genome of *M. pneumoniae*.

Please cite this article as: Pachkov, M. et al. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. Gene (2007), doi:10.1016/j.gene.2007.02.033

species including *M. pneumoniae* have recently been analyzed with respect to the redundancy of their metabolism, by testing the effect of *in silico* deletion of enzymes (Mombach et al., 2006). Moreover, this parasitic bacterium is of medical importance because it is one of the causative agents of tracheobronchitis and atypical pneumonia (Cassell et al., 1997; Jacobs, 1997). Its genome was completely sequenced and includes 688 protein ORFs (Himmelreich et al., 1996; Dandekar et al., 2000). However, about 33% of its proteins remain unassigned. This is partly because experimental results on protein function in *M. pneumoniae* are relatively limited (Finch and Mitchell, 1992; Pollack et al., 1997; Razin et al., 1998). Nucleotide metabolism is of special interest because it is particularly amenable to therapeutic intervention and any manipulation directly affects cell reproduction. One of our aims is to make predictions about resistance effects and the ways *M. pneumoniae* can escape from drug treatment.

Here, we significantly extend the data set and analysis for nucleotide metabolism presented earlier (Pfeiffer et al., 2000; Schuster et al., 2002b) by including additional enzymes and metabolites, discussing elementary modes and nutritional requirements in detail, comparing the obtained pathways with experimental data and combining the analysis with genome-context methods (see above). A difficult problem in the pathway analysis of larger networks is that of combinatorial explosion of pathways (Mavrovouniotis et al., 1990; Dandekar et al., 2003). To cope with this problem, decomposing the metabolic network is appropriate. We use the method of decomposition presented earlier (Schuster et al., 2002b). In addition, for convenience and better correspondence with common biochemical classification, the network derived from the decomposition procedure is split into two parts — the subsystems of purine and pyrimidine metabolism.

2. Material and methods

We use the concept of elementary modes (Schuster and Hilgetag, 1994; Heinrich and Schuster, 1996; Schuster et al., 2000, 2002a,b). These are flux distributions that cannot be decomposed into two simpler flux distributions at steady state.

The starting point is provided by information stored in the freely accessible KEGG database (http://www.genome.ad.jp/kegg/). These are essentially two data sets: one with all enzyme genes identified in *M. pneumoniae* and one containing all EC numbers with the corresponding reactions. The list of annotated genes was cross-checked with data available from "The *M. pneumoniae* Genome Project" (http://www.zmbh.uni-heidelberg.de/M_pneumoniae/genome/Results.html). The data on the

Table 2
Reactions included in the model of purine metabolism

EC number	Reaction	Reversibility	Gene [a]
2.7.1.11/40, 2.7.2.3	**ATP**+Pyruvate [b]=**ADP**+PEP [b]	REV	K6PF/KPYK
	dATP+Pyruvate=**dADP**+PEP	REV	PGK
	dGTP+Pyruvate=**dGDP**+PEP	REV	
	GTP+Pyruvate=**GDP**+PEP	REV	ACKA
2.7.4.3	ATP+**dAMP**=ADP+**dADP**	REV	KAD
	ATP+**AMP**=ADP+**ADP**	REV	
1.17.4.1	Thioredoxin+**ADP**=**dADP**+Oxidized thioredoxin+H_2O	REV	RIR1/2
	GDP+Thioredoxin=**dGDP**+Oxidized thioredoxin+H_2O	REV	
2.4.2.8	**AMP**+PPi=**Adenine**+PRPP	REV	HPRT
	GMP+PPi=**Guanine**+PRPP	REV	
	IMP+PPi=**Hypoxanthine**+PRPP	REV	
	XMP+PPi=**Xanthine**+PRPP	REV	
2.4.2.7	**AMP**+PPi=**Adenine**+PRPP	REV	APT
	GMP+PPi=**Guanine**+PRPP	REV	
2.7.4.8	ATP+**GMP**=ADP+**GDP**	REV	KGUA
	ATP+**dGMP**=ADP+**dGDP**	REV	
2.7.7.7	**dATP**+DNA=PPi+DNA	IRREV	DPO3A/B
	dGTP+DNA=PPi+DNA	IRREV	DPO3(X)
2.7.7.6	**ATP**+RNA=PPi+RNA	IRREV	RPOA/B
	GTP+RNA=PPi+RNA	IRREV	RPOC
2.4.2.4	Deoxyinosine+Pi=**Hypoxanthine**+2-Deoxy-R1P	REV	TYPH
	Deoxyguanosine+Pi=**Guanine**+2-Deoxy-R1P	REV	
	Deoxyadenosine+Pi=**Adenine**+2-Deoxy-R1P	REV	
2.4.2.1	Deoxyinosine+Pi=**Hypoxanthine**+2-Deoxy-R1P	REV	DEOD
	Deoxyguanosine+Pi=**Guanine**+2-Deoxy-R1P	REV	
	Deoxyadenosine+Pi=**Adenine**+2-Deoxy-R1P	REV	
	Adenosine+Pi=**Adenine**+R1P	REV	
	Guanosine+Pi=**Guanine**+R1P	REV	
	Inosine+Pi=**Hypoxanthine**+R1P	REV	
	Xanthosine+Pi=**Xanthine**+R1P	REV	

Same source of information as in Table 1. Bold symbols indicate internal metabolites.

[a] Gene names are given according to Swissprot classification. The suffices "_MYCPN" are omitted.

[b] By way of example, we here indicate pyruvate and phosphoenolpyruvate (corresponding to pyruvate kinase, EC 2.7.1.40). The other glycolytic kinases involve other metabolites.

Please cite this article as: Pachkov, M. et al. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. Gene (2007), doi:10.1016/j.gene.2007.02.033

reversibility of enzymatic reactions were retrieved from the EMP database (http://www.empproject.com/). Following the information from the above-mentioned databases, several enzymes are taken to have broad substrate specificity. For example, in the case of hypoxanthine–guanine phosphoribosyltransferase (EC 2.4.2.8), four different reactions are considered.

All metabolites and metabolic reactions are shown in Tables 1 and 2. An important issue in Metabolic Pathway Analysis is how to define external and internal metabolites. The latter are balanced at steady state by the reactions included in the model, while external metabolites are assumed to be buffered by inexhaustible reservoirs (Heinrich and Schuster, 1996; Schuster et al., 2000). Note that the term "steady state" refers to the concentrations, that is, mole numbers per volume, so that such a state is (nearly) maintained even in the growth phase of bacteria. Metabolites known to be nutrients or excreted products, as well as substances available in large excess, were here defined to be external. There are not much experimental data on the composition of growth media of *M. pneumoniae* or *M. genitalium* due to the difficulty of their *in vitro* cultivation. In case information was not available, we assume that *M. pneumoniae* has similar properties as the majority of *Mycoplasma* species or consider two different hypothetical cases (see Section 3.2). The main sources of information are the reviews by Rodwell (1983), Pollack et al. (1997), Razin et al. (1998) and Finch and Mitchell (1992).

We then applied the decomposition method proposed in Schuster et al. (2002b). In this method, all internal metabolites participating in more then a threshold number of reactions are operationally set to external status. These metabolites form the boundaries of the obtained subnetworks. Here, we use a value of four for the threshold. This results in 19 subsystems for *M. pneumoniae* with one of them being nucleotide metabolism (Schuster et al., 2002a,b). The resulting nucleotide subnetwork can be decomposed into two parts, pyrimidine and purine metabolisms.

The pyrimidine network (Fig. 1) includes all enzyme activities that were predicted from the genome and are relevant for this network. In addition we have included uridine phosphorylase (EC 2.4.2.3), deoxycytidine kinase (EC 2.7.1.74) and nucleoside-diphosphate kinase activity (EC 2.7.4.6). The former two were detected biochemically (Pollack, 2002) but not yet in the genome sequence. The latter was detected neither way, but there is an evidence that its function can be replaced by the glycolytic kinases, 6-phosphofructokinase (EC 2.7.1.11), phosphoglycerate kinase (EC 2.7.2.3), and pyruvate kinase (EC 2.7.1.40) (Pollack et al., 2002). It was found by biochemical assays that these enzymes can, in *M. pneumoniae* and a number of other *Mycoplasma* species, operate not only with ATP and ADP but with a large number of NTPs and NDPs (Pollack et al., 2002). As 6-phosphofructokinase consumes NTP while phosphoglycerate kinase and pyruvate kinase produce NTP, an interconversion of all NTPs is feasible, thus effectively carrying out the function of nucleoside-diphosphate kinase. Therefore, in Table 1, that kinase activity is denoted by the EC numbers EC 2.7.1.11, 2.7.2.3, and 2.7.1.40 of glycolytic kinases. In the conversion

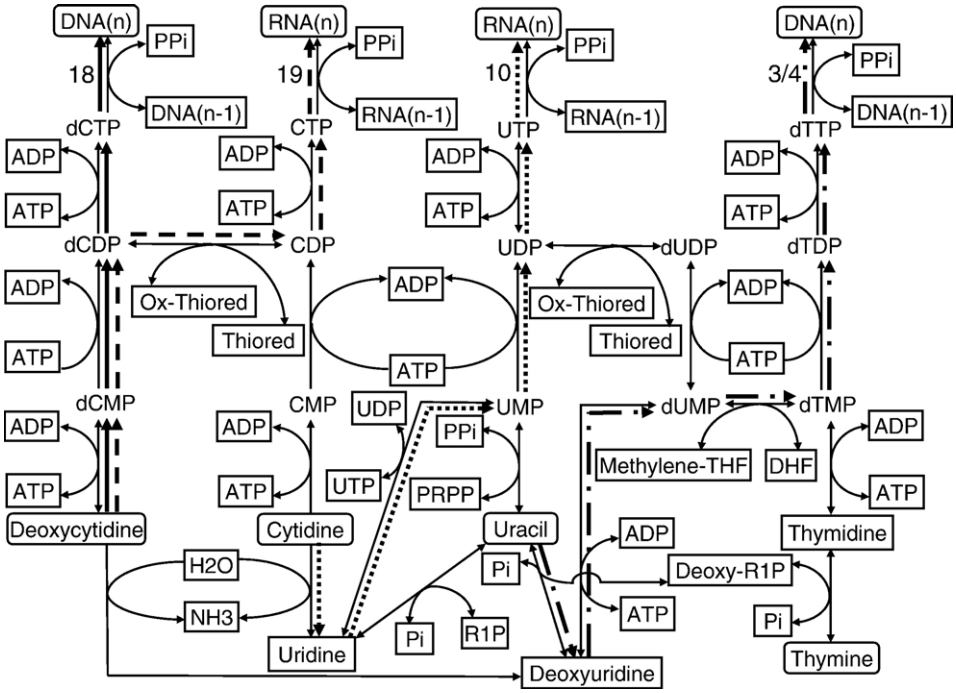

Fig. 1. Scheme of pyrimidine metabolism of *M. pneumoniae*. Metabolites in rectangular boxes are *a priori* external metabolites, that are, initial substrates and final products. Metabolites in rounded boxes were set to external status in the decomposition procedure. All other metabolites are internal. Bidirectional arrows indicate reversible reactions. Thick arrows indicate five exemplifying elementary modes, notably the modes 3 and 4 (dash–dotted arrows), 10 (dotted), 18 (solid) and 19 (dashed) given in Table 3. Choice and representation of these modes are arbitrary and are not related to their importance. For simplicity's sake, in the elementary modes, only one arrow in each bimolecular reaction is shown. Modes 3 and 4 differ in that the reaction from uracil to deoxyuridine is catalyzed by two different enzymes (see text and Tables 1 and 3). Abbreviations: Ox-Thiored, oxidized thioredoxin; Thiored, reduced thioredoxin. For further abbreviations, see list at the top of the manuscript.

Please cite this article as: Pachkov, M. et al. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. Gene (2007), doi:10.1016/j.gene.2007.02.033

between ribonucleotide diphosphates and desoxy-ribonucleotide diphosphates, the thioredoxin system is involved (Ben-Menachem et al., 1997). In total, the system contains 28 enzyme-catalyzed reactions.

In Fig. 2, the scheme of purine metabolism in *M. pneumoniae* is shown. Again, the metabolic system contains a nucleoside-diphosphate kinase activity (Pollack et al., 2002) in addition to the annotated enzymes. The network includes 30 reactions.

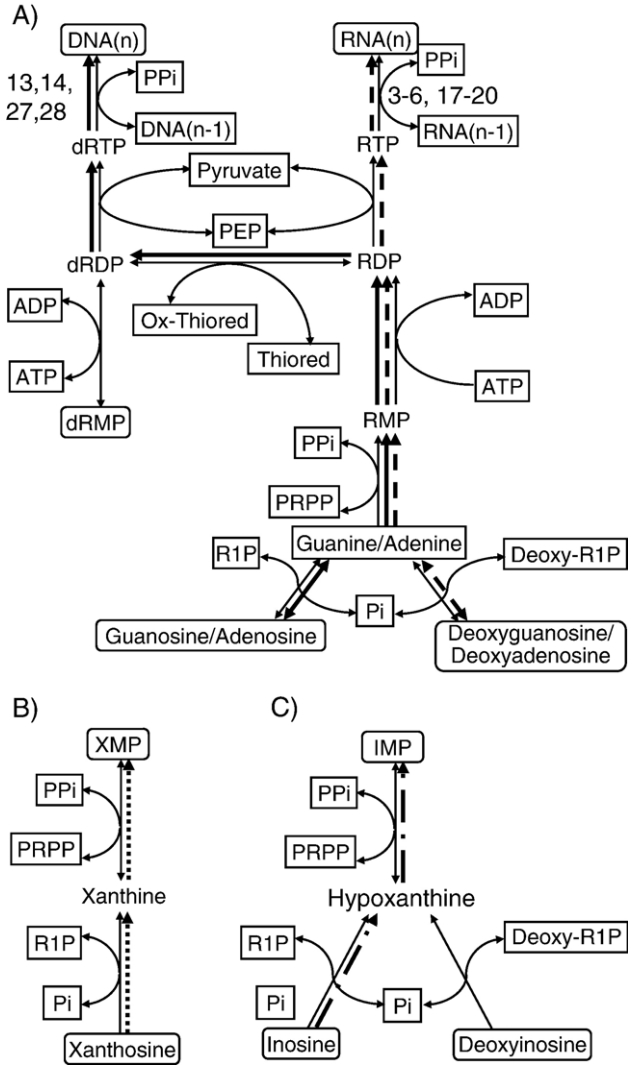

Fig. 2. Scheme of purine metabolism of *M. pneumoniae*, consisting of four subnetworks (A–C). The networks for adenine and guanine nucleotides are represented by one scheme (A), using "R" as a symbol for purines. However, in scenario a), conversions are only possible among adenine nucleotides and, separately, among guanine nucleotides. Metabolites in rectangular boxes are *a priori* external metabolites. Metabolites in rounded boxes were set to external status in the decomposition procedure. All other metabolites are internal. Bidirectional arrows indicate reversible reactions. Thick arrows indicate 14 exemplifying elementary modes, notably: A) the modes 3–6 and 17–20 (dashed arrows) 13,14, 27, 28 (solid); B) a mode leading from xanthosine to XMP (dotted); C) a mode leading from inosine to IMP (dash–dotted). Choice and representation of these modes are arbitrary. For simplicity's sake, in the elementary modes, only one arrow in each bimolecular reaction is shown. When more than one mode number is attached to a path of thick arrows, the modes differ in that some reactions are catalyzed by more than one enzyme (see text and Tables 2 and 4). Abbreviations: Ox-Thiored, oxidized thioredoxin; (d)RM(D,T)P, (deoxy)purine mono(di,tri)phosphate; Thiored, reduced thioredoxin. For further abbreviations, see list at the top of the manuscript.

The elementary modes were calculated for these two subsystems by the program METATOOL (Pfeiffer et al., 1999; von Kamp and Schuster, 2006) (for the current version see http://pinguin.biologie.uni-jena.de/bioinformatik). The obtained results were compared with experimental data to locate obvious gaps in the metabolic network. Then we would fill these with additional enzymes or additional reactions of multifunctional enzymes already present in the system. As for additional enzymes, we tried, in the first place, to use enzymes that had been found in other *Mycoplasma* species. In many situations, however, filling a pathway gap is possible in more than one way. We applied the criterion that the minimum possible number of new enzymes is added. This approach is based on the general principle of Ockham's razor, saying that the simplest solution is most likely to be the correct one (*cf.* Forster, 1999).

Resulting from the decomposition technique (Schuster et al., 2002b), most nucleoside phosphates such as ATP, ADP, dATP *etc.* are external metabolites. However, then it is impossible to trace, for example, the routes that go through ATP and dATP to RNA and DNA, respectively. Thus, we declared the nucleoside phosphates as external metabolites (that is, as sources or sinks with buffered concentrations) in all reactions in which they act as phosphate donors and as internal metabolites (which need to be balanced at steady state) whenever they are intermediates in the synthesis of RNA or DNA. That approach is similar to the treatment of "currency metabolites" by Ma and Zeng (2003). Currency metabolites are carriers for transferring certain functional groups (*e.g.* phosphate).

We also utilized comparative genome and gene expression analyses in an attempt at filling predicted gaps in nucleotide metabolism. In particular, functional associations between proteins were predicted by analyzing gene co-expression and the genomic context of genes across genomes, using the STRING server (von Mering et al., 2005) version 4. STRING integrates and combines evidence from several computational methods that analyze evolutionary signals across 110 completely sequenced genomes to predict functional associations between proteins — that is the involvement of proteins in the same biochemical pathway or cellular process. A combined list of 30 genes[1] encoding enzymes (or enzyme subunits) from both pathways was used as queries in STRING, using the 'protein mode'. We searched for genes that (i) are located in direct

[1] Query genes are: KPYK_MYCPN (2.7.1.40), KAD_MYCPN (2.7.4.3), RIR2_MYCPN (1.17.4.1), RIR1_MYCPN (1.17.4.1), HPRT_MYCPN (2.4.2.8), APT_MYCPN (2.4.2.7), KGUA_MYCPN (2.7.4.8), DP3B_MYCPN (2.7.7.7), Y007_MYCPN (2.7.7.7), DPO3_MYCPN (2.7.7.7), DP3A_MYCPN (2.7.7.7), Y315_MYCPN (2.7.7.7), Y360_MYCPN (2.7.7.7), DP3X_MYCPN (2.7.7.7), RPOE_MYCPN (2.7.7.6), RPOA_MYCPN (2.7.7.6), RPOC_MYCPN (2.7.7.6), RPOB_MYCPN (2.7.7.6), TYPH_MYCPN (2.4.2.4), DEOD_MYCPN (2.4.2.1), KCY_MYCPN (2.7.4.14), URK_MYCPN (2.7.1.48), CDD_MYCPN (3.5.4.5), KTHY_MYCPN (2.7.4.9), TYSY_MYCPN (2.1.1.74), PGK_MYCPN (2.7.2.3), PTA_MYCPN (2.7.2.1), K6PF_MYCPN (2.7.1.11), KITH_MYCPN (2.7.1.21), UPP_MYCPN (2.4.2.9).

Please cite this article as: Pachkov, M. et al. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. Gene (2007), doi:10.1016/j.gene.2007.02.033

genomic neighborhood of the queries, (ii) participate in conserved gene neighborhood clusters with these, (iii) are fused to these in genomes other than *M. pneumoniae*, (iv) show common presence and absence with these across genomes, (v) or have been shown to be co-expressed with these in DNA microarray experiments performed in other genera than *Mycoplasma* (the underlying methods are reviewed in, *e.g.*, Huynen et al., 2000; von Mering et al., 2005). Conservation of co-expression strengthens the inferred functional linkage for method (v). Methods (i)−(v) were recently benchmarked independently and combined using an integrated probabilistic approach (von Mering et al., 2005). We used genes retrieved with a combined STRING score of >400 in an additional search round in STRING, in order to obtain a protein functional association network of depth 2, which has 154 nodes and 4000 edges.

3. Results

3.1. Pyrimidine metabolism

The pyrimidine network (Fig. 1) gives rise to 45 elementary modes. There are 20 modes leading to DNA and RNA synthesis (Table 3), 7 cycles and 18 modes representing a transformation of bases into each other. The first group of elementary modes is most interesting since they reflect growth media requirements of *M. pneumoniae*. In Fig. 1, five arbitrarily chosen elementary modes are depicted. It can be seen that elementary modes may overlap each other (*e.g.* modes 18 and 19). The double arrows in Table 3 indicate reactions that are catalyzed by two enzymes each. For example, the reaction deoxyuridine + Pi = uracil + 2-deoxy-R1P is catalyzed both by thymidine phosphorylase (EC 2.4.2.4) and by purine-nucleoside phosphorylase (EC 2.4.2.1) (see Table 1). Accordingly, elementary modes 3 and 4, which only differ in these enzymes, are grouped together in Table 3 and Fig. 1.

Table 3
Elementary modes leading to DNA/RNA synthesis in the pyrimidine metabolic system

1. Uracil→UMP→UDP→UTP→RNA
2. Uracil→Uridine→UMP→UDP→UTP→RNA
3./4. Uracil⇒Deoxyuridine→dUMP→dUDP→UDP→UTP→RNA
5./6. Uracil⇒Deoxyuridine→dUMP→dTMP→dTDP→dTTP→DNA
7. Uracil→UMP→UDP→dUDP→dUMP→dTMP→dTDP→dTTP→DNA
8. Uracil→Uridine→UMP→UDP→dUDP→dUMP→dTMP→dTDP→ dTTP→DNA
9. Cytidine→CMP→CDP→CTP→RNA
10. Cytidine→Uridine→UMP→UDP→UTP→RNA
11. Cytidine→CMP→CDP→dCDP→dCTP→DNA
12. Cytidine→Uridine→UMP→UDP→dUDP→dUMP→dTMP→dTDP→ dTTP→DNA
13./14.Thymine→Thymidine⇒dTMP→dTMP→dTDP→dTTP→DNA
15./16.Thymine→Thymidine⇒dTMP→dTMP→dUMP→dUDP→UDP→ UTP→RNA
17. Deoxycytidine→Deoxyuridine→dUMP→dUDP→UDP→UTP→RNA
18. Deoxycytidine→dCMP→dCDP→dCTP→DNA
19. Deoxycytidine→dCMP→dCDP→CDP→CTP→RNA
20. Deoxycytidine→Deoxyuridine→dUMP→dTMP→dTDP→dTTP→DNA

Modes are grouped according to initial substrate. The double-arrow indicates reactions catalyzed by two enzymes.

Eight modes leading to RNA and DNA *via* triphosphates start with uracil. Four of these go *via* UTP and four, *via* dTDP. For reasons of clarity, only one example mode starting from uracil is depicted in Fig. 1. Four modes originate from cytidine and produce CTP, UTP, dCTP and dTTP, respectively. Deoxycytidine is analogous to cytidine in the number of modes and final products. Thymine gives four modes. Two generate dTTP and two others lead to dUTP. The production of UTP and dTTP (eight modes each with four of them starting with uracil) is more redundant than that of CTP and dCTP (two modes each).

The system is capable of using cytidine or deoxycytidine as a single source for production of all pyrimidine nucleotides and deoxynucleotides. Thus, for *M. pneumoniae*, cytidine is sufficient as a sole source of pyrimidines. This finding is in good agreement with empirical data (Finch and Mitchell, 1992; Pollack et al., 1997). In contrast, uracil would not have the capability of providing CTP and dCTP (despite the higher number of elementary modes starting from it) if the list of relevant enzymes was complete. However, uracil has been found experimentally to be a sufficient precursor for pyrimidines in other *Mycoplasma* species (Finch and Mitchell, 1992). There is a possible connection between uracil and (d)CTP *via* the reverse reaction of cytidine deaminase (EC 3.5.4.5). However, it has been shown that this reverse reaction is very unlikely to proceed *in vivo* (Cohen and Wolfenden, 1971). Thus, we declared cytidine deaminase as irreversible in our model, which makes (d)CTP synthesis inaccessible from uracil. Rather, the solution to this inconsistency appears to reside in CTP synthase, which is not included in our list because it has not yet been found in the genome (see below).

Thymine can be transformed to dTTP and UMP. Though we have not found any references about synthesizing UTP from thymine in mycoplasmas, the reaction of thymidylate synthase (2.1.1.45) is indicated as reversible in the EMP database, which opens a pathway from thymine to UTP.

Our results show that modes number 13 and 14 (Table 3) can be critical for dTTP production since a faster growth (at least in the host) is possible if dTTP is produced directly from thymine instead of producing it *via* UMP. This is only possible if there is a direct import of necessary precursors from the host cell — which is probably the case. Thus, dTTP production *via* modes 13/14 can limit the growth rate even under optimal conditions.

Another interesting point is that the concentration of dUTP has always to be kept low in the cell since otherwise dUTP instead of dTTP is incorporated into the DNA. This is usually done by dUTP diphosphohydrolase (EC 3.6.1.23) as it was observed for *Mycoplasma mycoides* (Neale et al., 1983). So far, there is no specific dUTP diphosphohydrolase known in *M. pneumoniae* (Williams and Pollack, 1988). Since some enzymes in mycoplasmas can have very broad specificities (Pollack et al., 2002), we speculate that thymidine phosphorylase (EC 2.4.2.4) may have a high affinity for dUTP and can, thus, keep its concentration low, taking into account that dCTP deaminase and dUTPase are both missing in *M. pneumoniae*. A proof of this suggestion requires experimental study of the corresponding enzyme activity.

Please cite this article as: Pachkov, M. et al. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. Gene (2007), doi:10.1016/j.gene.2007.02.033

The obtained network agrees in many points with a general model of mycoplasma pyrimidine metabolism (Finch and Mitchell, 1992) except the experimentally observed sufficiency of uracil as a source of pyrimidines. In order to obtain such functionality, we have included CTP synthase (EC 6.3.4.2) as a putative enzyme for *M. pneumoniae* because it is present in other mycoplasmas (Chambaud et al., 2001). This activity opens pathways from uracil to CTP *via* amination of UTP (Finch and Mitchell, 1992). In our simulation, including CTP synthase increases the number of elementary modes to 57 (not shown). Interestingly, the addition of that reaction opens routes to CTP from thymine along with uracil. Hence, thymine as well as uracil would be capable of serving as the single source of pyrimidines if CTP synthase (EC 6.3.4.2) was present.

3.2. Purine metabolism

The network of purine metabolism gives rise to 56 elementary modes (Fig. 2 and Table 4). 28 modes lead from various precursors to RNA and DNA. The remaining elementary modes represent the conversion of precursors into each other. The reaction set consists of four separate subsystems, which are dependent on different bases: adenine, guanine, xanthine and hypoxanthine. The former two lead to DNA and RNA synthesis. The latter two produce only the corresponding monophosphate nucleotides.

In Fig. 2, 14 arbitrarily chosen elementary modes are depicted. The elementary modes of the purine metabolism are redundant (see Table 4). The high redundancy arises because some reactions are "degenerate". The double arrows in Table 4 again indicate reactions that are catalyzed by two enzymes each. For example, the route connecting deoxyadenosine and RNA involves two reactions catalyzed by two enzymes each. Thus, there are four combinations of these reactions, which correspond to the four elementary modes 3–6. Moreover, there are four modes starting from dGMP and dAMP. However, it is unlikely that the latter modes are of physiological importance.

The guanine and adenine networks are not connected with each other. That would make them both necessary for growth.

Table 4
Elementary modes leading to DNA/RNA synthesis in the purine metabolic system

1. dAMP→dADP→dATP→DNA
2. dAMP→dADP→ADP→ATP→RNA
3.–6. Deoxyadenosine⇒Adenine⇒AMP→ADP→ATP→RNA
7.–10. Deoxyadenosine⇒Adenine⇒AMP→ADP→dADP→dATP→DNA
11./12. Adenosine→Adenine⇒AMP→ADP→ATP→RNA
13./14. Adenosine→Adenine⇒AMP→ADP→dADP→dATP→DNA
15. dGMP→dGDP→dGTP→DNA
16. dGMP→dGDP→GDP→GTP→RNA
17.–20. Deoxyguanosine⇒Guanine⇒GMP→GDP→GTP→RNA
21.–24. Deoxyguanosine⇒Guanine⇒GMP→GDP→dGDP→dGTP→DNA
25./26. Guanosine→Guanine⇒GMP→GDP→GTP→RNA
27./28. Guanosine→Guanine⇒GMP→GDP→dGDP→dGTP→DNA

Modes are grouped according to initial substrate. The double-arrow indicates reactions catalyzed by two enzymes.

From experiment, it is known that different *Mycoplasma* species have different purine requirements. While guanine is a sufficient precursor of all purine nucleotides in *M. mycoides*, adenine is required in addition in *M. capricolum* (Rodwell, 1983; Finch and Mitchell, 1992). For *M. pneumoniae* (and *M. genitalium*), the requirements are unknown. Here, we analyze two different scenarios: a) *M. pneumoniae* has the same requirements as *M. capricolum*, b) it shows the same requirements as *M. mycoides*. As the two species used for comparison are related very closely to each other, it is difficult to decide from reconstructed phylogenetic trees to which of the two species *M. pneumoniae* is related more closely (Pettersson et al., 1996; Rawadi et al., 1998). From the fact that *M. pneumoniae* has a very small genome, one may conclude that its metabolism resembles more closely to *M. capricolum* because the latter appears not to harbour any connection between the adenine and guanine networks.

In scenario a), we can leave the purine network as it is. In scenario b), we need to add at least three enzymes in order to open the necessary routes. We have chosen, according to the Ockham's razor principle, adenylosuccinate synthetase (EC 6.3.4.4), adenylosuccinate lyase (EC 4.3.2.2), and GMP reductase (EC 1.7.1.7) as candidates to fill the gaps. These enzymes had been found in *M. mycoides* (Westberg et al., 2004), *Acheloplasma* species and *Spiroplasma floricola* (Finch and Mitchell, 1992). Other possible sets of enzymes are larger or less likely in *Mycoplasma* species. For example, another possible but larger set consists of the following activities: GMP synthase (EC 6.3.5.2), IMP dehydrogenase (EC 1.1.1.205), adenylosuccinate synthetase (EC 6.3.4.4), adenylosuccinate lyase (EC 4.3.2.2), which were found in *M. penetrans* (Sasaki et al., 2002). The chosen set of activities would allow *M. pneumoniae* to use a pathway from GMP to AMP *via* IMP that is present in the above-mentioned species such as *M. mycoides*. The total number of elementary modes in that system is 65 (not shown). The chosen reactions not only provide the desired conversion of guanine to ATP and dATP but in addition join the inosine subnetwork to the rest of the purine network. A new pair of elementary modes to ATP and dATP would arise for each of the following precursors: inosine, deoxyinosine, guanosine and deoxyguanosine. That means that there would be two pathways from hypoxanthine to ATP and dATP, which is in accordance with the known nutritional requirements of mycoplasmas (Finch and Mitchell, 1992).

In scenario a), the two smaller subsystems may lead to accumulation of IMP and XMP (Fig. 2B,C). Such "isolated" reaction networks might explain the poor growth of mycoplasmas in rich growth media, when some nutrients turn out to be toxic (Razin et al., 1998). However, there may be hitherto unknown enzymatic activities, which utilize these metabolites, as in the second hypothesis put forward above for scenario b).

3.3. Search for hidden enzyme activities by genomic context methods

Our analysis suggested that some additional enzyme activities important for nucleotide metabolism might still be

Please cite this article as: Pachkov, M. et al. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. Gene (2007), doi:10.1016/j.gene.2007.02.033

hidden in the *M. pneumoniae* genome sequence. Context-based methods, for instance examining the genome context, can predict functional associations between the known members of a biochemical pathway and previously overlooked enzymes. Here we analyzed the *M. pneumoniae* nucleotide metabolism using these methods, *i.e.* by making use of the STRING server (von Mering et al., 2003, 2005).

Unexpectedly, the resulting functional association network (Fig. 3) revealed relatively few connections within or between proteins of the purine or pyrimidine metabolism — many enzymes are linked to large protein clusters involved in general cellular processes such as cell growth or protein biosynthesis. Altogether, eight proteins are singletons, in other words there is no genomic or expression evidence for a functional association. Usually, metabolic enzymes form extensive clusters that largely agree with the respective metabolic pathways (Van Noort et al., 2003; von Mering et al., 2003). We believe that the loose clustering observed here may reflect the central role of both purine or pyrimidine metabolism in cell growth and/or the existence of short, independent pathways that interconnect distinctly depending on the biochemical context. Nevertheless, when collapsing the largest clusters of associated proteins into single nodes (*e.g.* 'protein synthesis'; the resulting network has 56 nodes and 102 edges), we found that several of the remaining

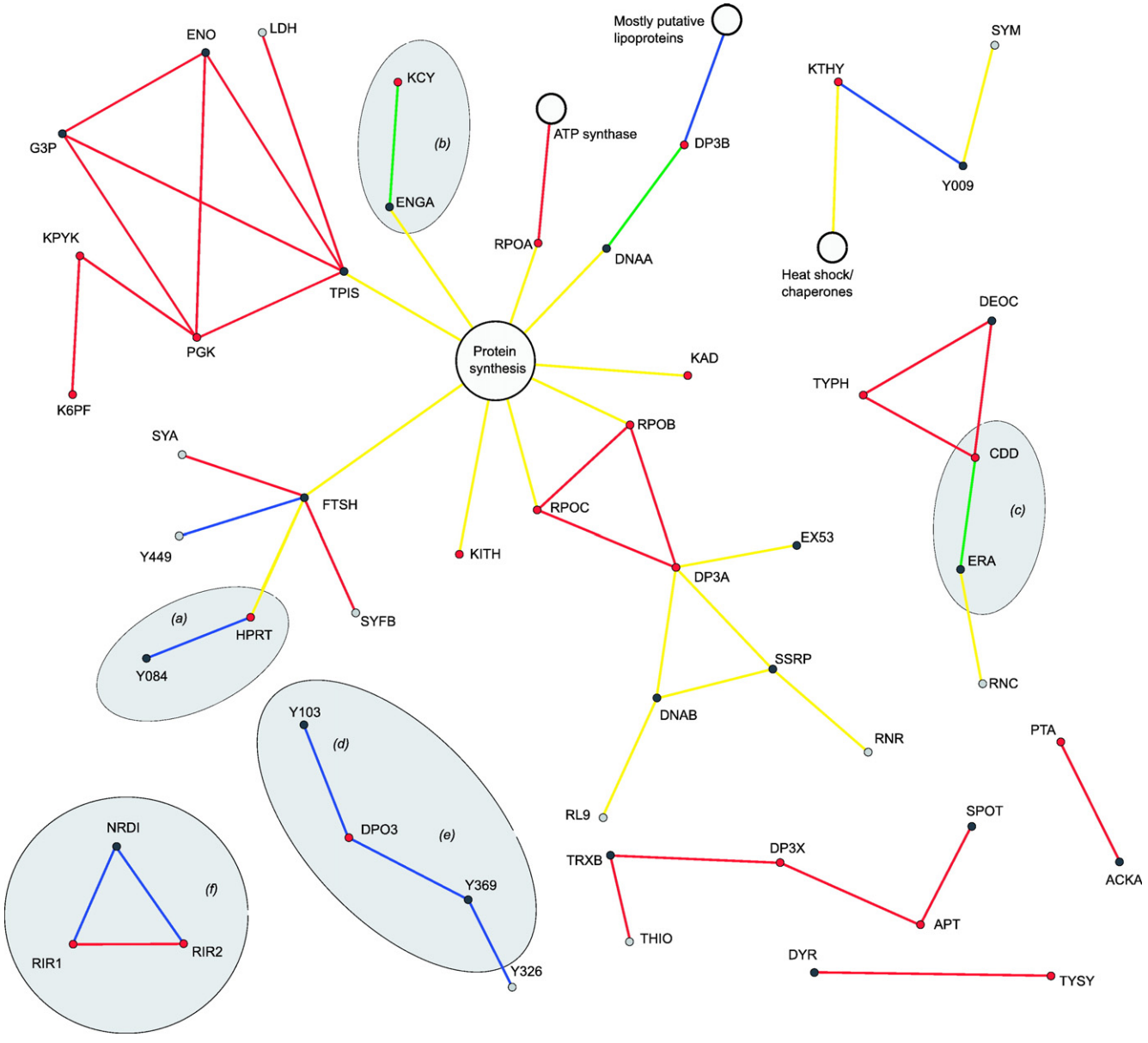

Fig. 3. Functional association network of purine and pyrimidine metabolism. The protein network was derived from the genomic context of genes, and from experimentally inferred co-expression. Poorly characterized proteins, which are predicted to be functionally linked to members of purine and pyrimidine metabolism, represent candidates for 'missing' enzymes. Edge colors indicate the presumable basis for functional association: red = functionally associated enzymes; green = regulatory associations; yellow = associations with general processes such as protein biosynthesis; blue = unknown. Node colors: red = query genes; dark grey = genes linked during first iteration; light grey = genes linked during second iteration; white = collapsed nodes. Swissprot identifiers are used as gene names, omitting the suffix "_MYCPN". Edge lengths are arbitrary. For a total number of eight genes, no functional relationships were observed; these 'singletons' were removed from the figure.

Please cite this article as: Pachkov, M. et al. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. Gene (2007), doi:10.1016/j.gene.2007.02.033

non-singleton proteins can indeed be functionally associated using STRING. Apart from known members of purine or pyrimidine metabolism, we observed several poorly characterized proteins in the remaining functional association network. Such poorly characterized proteins, which are associated with members of pyrimidine and purine metabolism, represent promising candidates for the missing enzymes. For instance, hypoxanthine guanine phosphoribosyltransferase (HPRT) is strongly associated to the poorly characterized Y084 (STRING score >800; prediction (a)), encoding a predicted ATPase of the PP-loop superfamily with a possible metabolic role. Furthermore, ENGA and ERA are putative GTPases assumed to be involved in cell growth and/or ribosome biogenesis (both are associated to known enzymes of purine and pyrimidine metabolism with scores >800 and >400, respectively). Nevertheless, their physiological roles are still tentative, and metabolic functions cannot be excluded (predictions (b)/(c)). Poorly characterized proteins linked to DPO3 are also candidates for the missing enzymes — for example, Y369 (score >650) was already predicted to be related to dihydroxyacetone kinase (predictions (d)/(e)). Last but not least, NRDI which is predicted to be tightly associated with RIR1 and RIR2 (scores >900), may be involved in ribonucleotide reduction or in another enzymatic activity of purine metabolism, putatively GMP reductase activity (prediction (f)).

4. Discussion

Here, we have established a network model of the nucleotide metabolism of *M. pneumoniae*. This has been obtained in an automated way as a subsystem of a larger part of *M. pneumoniae* metabolism, by a decomposition method suggested earlier (Schuster et al., 2002b). It is an interesting result that a decomposition method based on purely theoretical arguments gives rise to subsystems that are in agreement with biological experience, such as nucleotide metabolism. For this network, we computed the elementary modes (pathways) and conducted a genomic context search on the STRING server for finding missing enzymes.

To fill pathway gaps, we used the principle of Ockham's razor, that is, we used the simplest solution in that we inserted the smallest number of enzymes possible. Of course, a comprehensive analysis should include an extended set of hypotheses and evaluate these in parallel. However, this increases the number of possibilities in a combinatorial manner so that we have tested only two possible extensions in the case of the purine network and one extension of the pyrimidine network, for reasons of space.

Our elementary modes analysis shows that cytidine is a sufficient precursor for the production of all pyrimidine nucleotides while less clear assertions could be made regarding the other bases. For the likely case that also uracil is a sufficient pyrimidine precursor, we suggest CTP synthase (EC 6.3.4.2) to be active. We have analyzed two different scenarios depending on whether adenine is required as a substrate by *M. pneumoniae*. If it is required (as it is in *M. capricolum*), then the known enzymes are sufficient for reconstructing the purine metabolism network. Cytidine (or uracil), adenine and guanine would in that case be absolute requirements for cell growth. *M. pneumoniae* must take them up from the host and utilize in nucleotide metabolism. The synthesis of nucleotide triphosphates from these precursors is similar to the salvage pathways in higher organisms (*cf.* Schuster and Kenanov, 2005).

For the case that adenine is not required in the medium (like in *M. mycoides*), we have suggested a putative minimal set of enzymes to fill gaps in the network. These are adenylosuccinate synthetase (6.3.4.4), adenylosuccinate lyase (4.3.2.2) and GMP reductase (1.7.1.7). While the corresponding genes have not yet been found in the genome of *M. pneumoniae*, they have been found by biochemical assays in other *Mycoplasma* species. Also, it could be a case of non-orthologous gene displacement (Koonin et al., 1996), which is difficult to resolve. From the genomic context analysis carried out with STRING, it follows that a candidate for GMP reductase is the protein NRDI. However, as no indication for the other two enzymes could be found by the STRING analysis and *M. pneumoniae* has a very small genome, the more likely case is that it does not harbour these enzymes and does require adenine.

To further examine our conclusions, experiments on growth requirements for *M. pneumoniae* are desirable. The experimental data used in our work are not *M. pneumoniae* specific throughout and a final validation of our conclusions requires further research on *M. pneumoniae* metabolism.

As mentioned in the Introduction, other methods for reconstructing metabolisms have been proposed, such as "PathoLogic" (Green and Karp, 2004; Romero et al., 2004). Those methods include an automated comparison with sequence and pathway data for other organisms and a Bayesian approach. Our method based on elementary modes (Schuster et al., 2002a) and genome context (Von Mering et al., 2005) has strengths and weaknesses in relation to the other methods. Advantages are that the stoichiometric and thermodynamic feasibility of pathways is tested and that the method is very transparent and easy to implement. A weakness is that information from other species is included "manually" rather than in an automated way. This, however, has the advantage that expert knowledge about, for example, taxonomy can be taken into account more easily. For *M. pneumoniae*, it is sensible to preferably use information from other *Mycoplasma* species rather than from more distant species. Genome context analysis, on the other hand, also incorporates information from more distant taxa for function prediction. A strength of the Bayesian method is that probabilities for the candidate enzymes to show the desired functions can be calculated. In future work, it would be interesting and worthwile combining various methods.

The correct description of the adenine nucleotide phosphates in the model meets with difficulties. ATP, for example, is both a donor of the phosphate moiety and should, in this function, be treated as external. On the other hand, it is also a precursor of DNA and RNA and, hence, needs to be treated as internal. Following Ma and Zeng (2003), we chose to consider both internal and an external ATP pools. A theoretical justification for this way of description is, however, outstanding.

As mentioned in the Introduction, one of the goals of our analysis is the prediction of resistance effects in drug therapy. In this context, it is worth noting that *M. pneumoniae* does not

Please cite this article as: Pachkov, M. et al. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. Gene (2007), doi:10.1016/j.gene.2007.02.033

involve any "exotic" enzymes that would allow it to escape from treatment. Standard antibiotics against *M. pneumoniae* infections (mostly pneumonias) are tetracycline and erythromycin, both interfering with ribosomal protein synthesis (Clyde, 1993). On the basis of our model we can conclude that pyrimidine metabolism shows lower plasticity than purine metabolism and, thus, appear to be more appropriate as a drug target. The synthesis of CTP and dCTP possesses the lowest redundancy among the investigated pathways, which makes it most vulnerable for drugs.

The study confirms the known fact that nucleotide metabolism in *M. pneumoniae* is limited, and has in fact fewer enzymes than the human host. On the other hand there are no clear *M. pneumoniae* specific enzymes evident in this part of the metabolic map which are not present in the human host. Nevertheless, *M. pneumoniae* specific enzyme domain differences would allow targeting drugs towards *M. pneumoniae* nucleotide metabolism as already shown for its protein translation by standard therapies. In severe cases (immunocompromised patients) a *M. pneumoniae* infection could in principle be treated by nucleotide analogs, as is already used for antiviral treatment (Keam et al., 2004). Further, other enzymes directly interfering with nucleotide metabolism may be utilized, as *M. pneumoniae* has less enzymatic abilities to counterbalance their impairment than the human host. Furthermore, the interconversion of purines and the salvage pathways seems to be weak points of the *M. pneumoniae* metabolism that can be targeted pharmacologically as has been shown to work in *Cryptosporidium parvum* (Umejiego et al., 2004).

This study exemplifies for *M. pneumoniae* that the combination of genome context and network analyses allows one to examine and delineate nucleotide metabolism in pathogens including growth requirements and drug targets.

References

Ben-Menachem, G., Himmelreich, R., Herrmann, R., Aharonowitz, Y., Rottem, S., 1997. The thioredoxin reductase system of mycoplasmas. Microbiology 143, 1933–1940.

Carlson, R., Srienc, F., 2004. Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: creation of overall flux states. Biotechnol. Bioeng. 86, 149–162.

Cassell, G.H., Gray, G.G., Waites, K.B., 1997. Mycoplasma infections. In: Fauci, A.S., et al. (Ed.), Harrison's Principles of Internal Medicine. McGraw-Hill, New York.

Chambaud, I., et al., 2001. The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. Nucl. Acids Res. 29, 2145–2153.

Clyde Jr., W.A., 1993. Clinical overview of typical *Mycoplasma pneumoniae* infections. Clin. Infect. Dis. 17, S32–S36.

Cohen, R.M., Wolfenden, R., 1971. The equilibrium of hydrolytic deamination of cytidine and N14-mathylcytidine. J. Biol. Chem. 246, 7566–7568.

Dandekar, T., Schuster, S., Snel, B., Huynen, M., Bork, P., 1999. Pathway alignment: application to the comparative analysis of glycolytic enzymes. Biochem. J. 343, 115–124.

Dandekar, T., et al., 2000. Re-annotating the *Mycoplasma pneumoniae* genome sequence: adding value, function and reading frames. Nucleic Acids Res. 28, 3278–3288.

Dandekar, T., Moldenhauer, F., Bulik, S., Bertram, H., Schuster, S., 2003. A method for classifying metabolites in topological pathway analyses based on minimization of pathway number. BioSystems 70, 255–270.

Finch, L.R., Mitchell, A., 1992. Sources of nucleotides. In: Maniloff, J., McElhany, R., Fynch, L., Baseman, J. (Eds.), Mycoplasmas. Molecular Biology and Pathogenesis. American Society for Microbiology, Washington, D.C. ch. 13.

Förster, J., Gombert, A.K., Nielsen, J., 2002. A functional genomics approach using metabolomics and *in silico* pathway analysis. Biotechnol. Bioeng. 79, 703–712.

Forster, M.A., 1999. Model selection in science: the problem of language variance. Brit. J. Phil. Sci. 50, 83–102.

Green, M.L., Karp, P.D., 2004. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. BMC Bioinformatics 5, 76–91.

Heinrich, R., Schuster, S., 1996. The Regulation of Cellular Systems. Chapman & Hall, New York, NY.

Himmelreich, R., Hilbert, H., Plagens, H., Elsbeth Pirkl, E., Liw, B.-C., Herrmann, R., 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucl. Acid Res. 24, 4430–4449.

Hutchison III, C.A., et al., 1999. Global transposon mutagenesis and a minimal mycoplasma genome. Science 286, 2165–2169.

Huynen, M., Snel, B., Lathe III, W., Bork, P., 2000. Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome Res. 10, 1204–1210.

Huynen, M.A., Snel, B., von Mering, C., Bork, P., 2003. Functional prediction and protein networks. Curr. Opin. Cell Biol. 15, 191–198.

Jacobs, E., 1997. Mycoplasma infections of the human respiratory tract. Wien. Klin. Wochenschr. 109, 574–577.

Keam, S.J., Chapman, T.M., Figgitt, D.P., 2004. Brivudin (bromovinyl deoxyuridine). Drugs 64, 2091–2097.

Klamt, S., Stelling, J., 2003. Two approaches for metabolic pathway analysis? Trends Biotechnol. 21, 64–69.

Koonin, E.V., Mushegian, A.R., Bork, P., 1996. Non-orthologous gene displacement. Trends Genet. 12, 334–336.

Ma, H., Zeng, A.P., 2003. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 19, 270–277.

Mavrovouniotis, M.L., Stephanopoulos, G., Stephanopoulos, G., 1990. Computer-aided synthesis of biochemical pathways. Biotechnol. Bioeng. 36, 1119–1132.

Mombach, J.C.M., Lemke, N., da Silva, N.M., Ferreira, R.A., Isaia, E., Barcellos, C.K., 2006. Bioinformatics analysis of mycoplasma metabolism: important enzymes, metabolic similarities, and redundancy. Comp. Biol. Med. 36, 542–552.

Morowitz, H.J., Tourtellotte, M.E., 1962. The smallest living cells. Sci. Am. 206, 117–126.

Neale, G.A.M., Mitchell, A., Finch, L.R., 1983. Enzymes of pyrimidine deoxyribonucleotide metabolism in *Mycoplasma mycoides subsp. mycoides*. J. Bacteriol. 156, 1001–1005.

Osterman, A., Overbeek, R., 2003. Missing genes in metabolic pathways: a comparative genomics approach. Curr. Opin. Chem. Biol. 7, 238–251.

Pettersson, B., Leitner, T., Ronaghi, M., Bolske, G., Uhlen, M., Johansson, K.E., 1996. Phylogeny of the *Mycoplasma mycoides* cluster as determined by sequence analysis of the 16S rRNA genes from the two rRNA operons. J. Bacteriol. 178, 4131–4142.

Pfeiffer, T., Sanchez-Valdenebro, I., Nuño, J.C., Montero, F., Schuster, S., 1999. METATOOL: for studying metabolic networks. Bioinformatics 15, 251–257.

Pfeiffer, T., Dandekar, T., Moldenhauer, F., Schuster, S., 2000. Topological analysis of metabolic networks. Application to the metabolism of *Mycoplasma pneumoniae*. In: Hofmeyr, J.-H.S., Rohwer, J.M., Snoep, J.L. (Eds.), BioThermoKinetics 2000. Animating the Cellular Map. Stellenbosch University Press, Stellenbosch, pp. 229–234.

Pollack, J.D., 1997. Mycoplasma genes: a case for reflective annotation. Trends Microbiol. 5, 413–418.

Pollack, J.D., 2002. The necessity of combining genomic and enzymatic data to infer metabolic function and pathways in the smallest bacteria: amino acid, purine and pyrimidine metabolism in mollicutes. Front. Biosci. 7, d1762–d1781.

Pollack, J.D., Williams, M.V., McElhaney, R.N., 1997. The comparative metabolism of the mollicutes (mycoplasmas): the utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cell. Crit. Rev. Microbiol. 23, 269–354.

Please cite this article as: Pachkov, M. et al. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. Gene (2007), doi:10.1016/j.gene.2007.02.033

Pollack, J.D., Myers, M.A., Dandekar, T., Herrmann, R., 2002. Suspected utility of enzymes with multiple activities in the small genome Mycoplasma species: the replacement of the missing "household" nucleoside diphosphate kinase gene and activity by glycolytic kinases. OMICS 6, 247–258.

Poolman, M.G., Fell, D.A., Raines, C.A., 2003. Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. Eur. J. Biochem. 270, 430–439.

Price, N.D., Papin, J.A., Palsson, B.O., 2002. Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. Genome Res. 12, 760–769.

Rawadi, G., Dujeancourt-Henry, A., Lemercier, B., Roulland-Dussoix, D., 1998. Phylogenetic position of rare human mycoplasmas, *Mycoplasma faucium*, *M. buccale*, *M. primatum* and *M. spermatophilum*, based on 16S rRNA gene sequences. Int. J. Syst. Bacteriol. 48, 305–309.

Razin, S., Yogev, D., Naot, Y., 1998. Molecular biology and pathogenicity of mycoplasmas. Microbiol. Mol. Biol. Rev. 62, 1094–1156.

Rodwell, A.W., 1983. Defined and partly defined media. In: Razin, S., Tully, J.G. (Eds.), Methods in Mycoplasmology, vol. 1. Academic Press, New York, pp. 413–449.

Romero, P., Wagg, J., Green, M.L., Kaiser, D., Krummenacker, M., Karp, P.D., 2004. Computational prediction of human metabolic pathways from the complete human genome. Genome Biology 6, R2.

Sasaki, Y., et al., 2002. The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. Nucl. Acids Res. 30, 5293–5300.

Schilling, C.H., Palsson, B.O., 2000. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. J. Theor. Biol. 203, 249–283.

Schilling, C.H., Schuster, S., Palsson, B.O., Heinrich, R., 1999. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnol. Prog. 15, 296–303.

Schilling, C.H., Letscher, D., Palsson, B.O., 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J. Theor. Biol. 203, 229–248.

Schuster, S., Hilgetag, C., 1994. On elementary flux modes in biochemical reaction systems at steady state. J. Biol. Syst. 2, 165–182.

Schuster, S., Kenanov, D., 2005. Adenine and adenosine salvage pathways in erythrocytes and the role of S-adenosylhomocysteine hydrolase — a theoretical study using elementary flux modes. FEBS J. 272, 5278–5290.

Schuster, S., Dandekar, T., Fell, D.A., 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol. 17, 53–60.

Schuster, S., Fell, D.A., Dandekar, T., 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat. Biotechnol. 18, 326–332.

Schuster, S., Hilgetag, C., Woods, J.H., Fell, D.A., 2002a. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. J. Math. Biol. 45, 153–181.

Schuster, S., Pfeiffer, T., Moldenhauer, F., Koch, I., Dandekar, T., 2002b. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*. Bioinformatics 18, 351–361.

Schwartz, J.M., Kanehisa, M., 2006. Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Bioinformatics 7, 186.

Schwender, J., Goffman, F., Ohlrogge, J.B., Shachar-Hill, Y., 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. Nature 432, 779–782.

Umejiego, N.N., Li, C., Riera, T., Hedstrom, L., Striepen, B., 2004. *Cryptosporidium parvum* IMP dehydrogenase: identification of functional, structural and dynamic properties that can be exploited for drug design. J. Biol. Chem. 279, 40320–40327.

Van Noort, V., Snel, B., Huynen, M.A., 2003. Predicting gene function by conserved co-expression. Trends Genet 19, 238–242.

von Kamp, A., Schuster, S., 2006. Metatool 5.0: fast and flexible elementary modes analysis. Bioinformatics 22, 1930–1931.

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., Snel, B., 2003a. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 31, 258–261.

von Mering, C., et al., 2003b. Genome evolution reveals biochemical networks and functional modules. Proc. Natl. Acad. Sci. USA 100, 15428–15433.

von Mering, C., et al., 2005. STRING: known and predicted protein protein associations, integrated and transferred across organism. Nucleic Acids Res. 33, 433–437.

Westberg, J., et al., 2004. The genome sequence of *Mycoplasma mycoides* subsp. *mycoides* SC type strain PG1T, the causative agent of contagious bovine pleuropneumonia (CBPP). Genome Res. 14, 221–227.

Williams, M.V., Pollack, J.D., 1988. Uracil-DNA glycosylase activity. Relationship to proposed biased mutation pressure in the class mollicutes. In: Moses, R.E., Summers, W.C. (Eds.), DNA Replication and Mutagenesis. American Society for Microbiology, Washington, D.C., pp. 440–444.

Please cite this article as: Pachkov, M. et al. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. Gene (2007), doi:10.1016/j.gene.2007.02.033