Biochem. J. (1999) 343, 115–124 (Printed in Great Britain)

Pathway alignment: application to the comparative analysis of glycolytic enzymes

Thomas DANDEKAR*†[1], Stefan SCHUSTER†, Berend SNEL*†, Martijn HUYNEN*†‡ and Peer BORK*†

*EMBL, P.O. Box 102209, D-69012 Heidelberg, Germany, †Max Delbrück Centre for Molecular Medicine, Robert-Rössle-Strasse 10, 13092 Berlin-Buch, Germany, and ‡Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Suggests that both PYK gene from cyanobacteria are from FBP alg. fusion.

Comparative analysis of metabolic pathways in different genomes yields important information on their evolution, on pharmacological targets and on biotechnological applications. In this study on glycolysis, three alternative ways of comparing biochemical pathways are combined: (1) analysis and comparison of biochemical data, (2) pathway analysis based on the concept of elementary modes, and (3) a comparative genome analysis of 17 completely sequenced genomes. The analysis reveals a surprising plasticity of the glycolytic pathway. Isoenzymes in different species are identified and compared; deviations from the textbook standard are detailed. Several potential pharmacological targets and by-passes (such as the Entner–Doudoroff pathway) to glycolysis are examined and compared in the different species. Archaean, bacterial and parasite specific adaptations are identified and described.

Key words: elementary modes, glycolysis, genome comparison, isoenzymes, sequence analysis.

INTRODUCTION

Sequence alignment is a well-established tool for investigating and comparing nucleic acids from different species and for identifying characteristic gaps, insertions and dissimilarities. The availability of full genomic sequences and the increasing body of biochemical data open up higher-order possibilities for comparative analysis [1,2]. The alignment of biochemical pathways from different species is an important step towards a more global comparison from a physiological viewpoint.

Here we present a detailed comparison of the glycolytic pathway, including potential by-passes and species-specific enzyme sets, in different species. The recognition of such differences is interesting for biotechnology (the identification of alternative enzymes) and pharmacology (differences in drug targets). We extend and combine biochemical data with elementary-mode analysis of substrate fluxes, comparative genome analysis and pathway alignment to reveal a much greater variety in glycolysis than assumed in textbook standard knowledge, including implications for evolution and biotechnology.

Some other methods have been suggested to permit large-scale comparisons of genomic data and pathways, notably clusters of orthologous sequences [3] and the application of related enzyme clusters [4]; recent analyses have examined different genomic aspects of glycolysis [2,5].

Whereas these approaches stress the genomic nature of the data exploited for the comparison, our pathway alignment focuses more on the biochemical capacity of the organisms compared. Attempts in this biochemical direction have already been made [6,7]. However, we show here how the use of complementary data and tools extends the biochemical approach and improves the predictive power of such efforts. The availability (reviewed in [1]) of full genomic sequences is exploited; pathway fluxes and by-passes that use the identified enzymes are tested for consistency by algebraic methods developed for this purpose [8]. Pathway alignment is greatly facilitated by the availability of databases on the World Wide Web (e.g. TIGR, http://www.tigr.org; ExPasy, http://www.expasy.ch; KEGG, http://www.tokyo-center.genome.ad.jp/kegg/kegg4.html; WIT, http://wit.mcs.anl.gov/WIT2/). However, as databases are based on automated methods that include pre-formed pathway charts, these automated predictions are here further refined for an accurate analysis.

We combine our strategies for a systematic comparison of glycolysis, alternative pathways such as the Entner–Doudoroff pathway, and several functionally related enzymes such as pyruvate dehydrogenase. We compare this part of the metabolic chart in 16 completely sequenced prokaryotes as well as yeast.

MATERIALS AND METHODS

For the analysis of the glycolytic enzymes, species-specific biochemical evidence from known experimental data and literature was collected.

Available completely sequenced genomes were extensively cross-compared for a better identification of all encoded glycolytic enzymes in these organisms. Gene duplications, replacement by unrelated sequences (non-orthologous displacement; see [1]) and gene neighbourhood were taken into account. We used detailed sequence analysis protocols as described and reviewed previously [1].

Phylogenetic analysis was applied for a further analysis of gene duplication events and to clarify substrate specificities of the encoded enzymes.

In the pathway analysis of glycolysis, we applied methods to calculate elementary flux modes as described in [8,9]. For these calculations we used the programs EMPATH (J. Woods, personal communication) and METATOOL [10]; both programs are

Abbreviations used: PEP, phosphoenolpyruvate; PPP, pentose phosphate pathway.

[1] To whom correspondence should be addressed at the EMBL address (e-mail dandekar@embl-heidelberg.de).

available from ftp://bmsdarwin.brookes.ac.uk/pub/mca/software/ibmpc).

RESULTS AND DISCUSSION

Biochemical analysis of the pathway plasticity of glycolysis

After collection of relevant literature and databases, and extension of this by detailed sequence analysis, the biochemical analysis examines the following questions: Besides the standard pathway, which alternatives are available and how do these alternatives compare? Which enzyme sets are involved? Where are organism-specific, in particular parasite-specific, adaptations?

In a number of bacteria (including *Escherichia coli*), glucose is anaerobically converted into pyruvate by the 'standard' glycolytic pathway as known and determined in higher organisms including humans; at least some experimental evidence for this is available. An initial indication for this is the ability to grow on glucose as sole carbon source; a detailed study for different sugars in *Haemophilus influenzae* has been done [11], and for glucose in *Treponema hyodysenteriae* [12].

However, this does not mean that the presence of each glycolytic enzyme was directly determined biochemically. Table 1 gives an overview of glycolytic enzymes that were characterized directly biochemically, either in the species compared or in close relatives. Additional biochemical evidence is available from metabolite-tracing experiments and less direct data (some of this is summarized on our Web site http://www.bork.embl-heidelberg.de/Genome/glycolytic_enzymes). Nevertheless, most of the specific glycolytic enzymes from completely sequenced organisms have not been characterized in biochemical detail. For example, little is known about the biochemical capabilities of parasites such as *Treponema pallidum* and the obligate intracellular *Chlamydia trachomatis*, respectively because of difficulties in cultivating them continuously *in vitro* and because their metabolism has to be studied against the overwhelming background of the host cell metabolism [13]. *Chlamydia* species are considered to be energy parasites, depending on their hosts to generate ATP, because no metabolic pathway that results in the generation of energy could be detected [14]. Some experiments indicate that glucose can be degraded to CO_2 via the pentose phosphate pathway (PPP) and by either glycolysis or the Entner–Doudoroff pathway, although with a very low, hardly detectable rate [14].

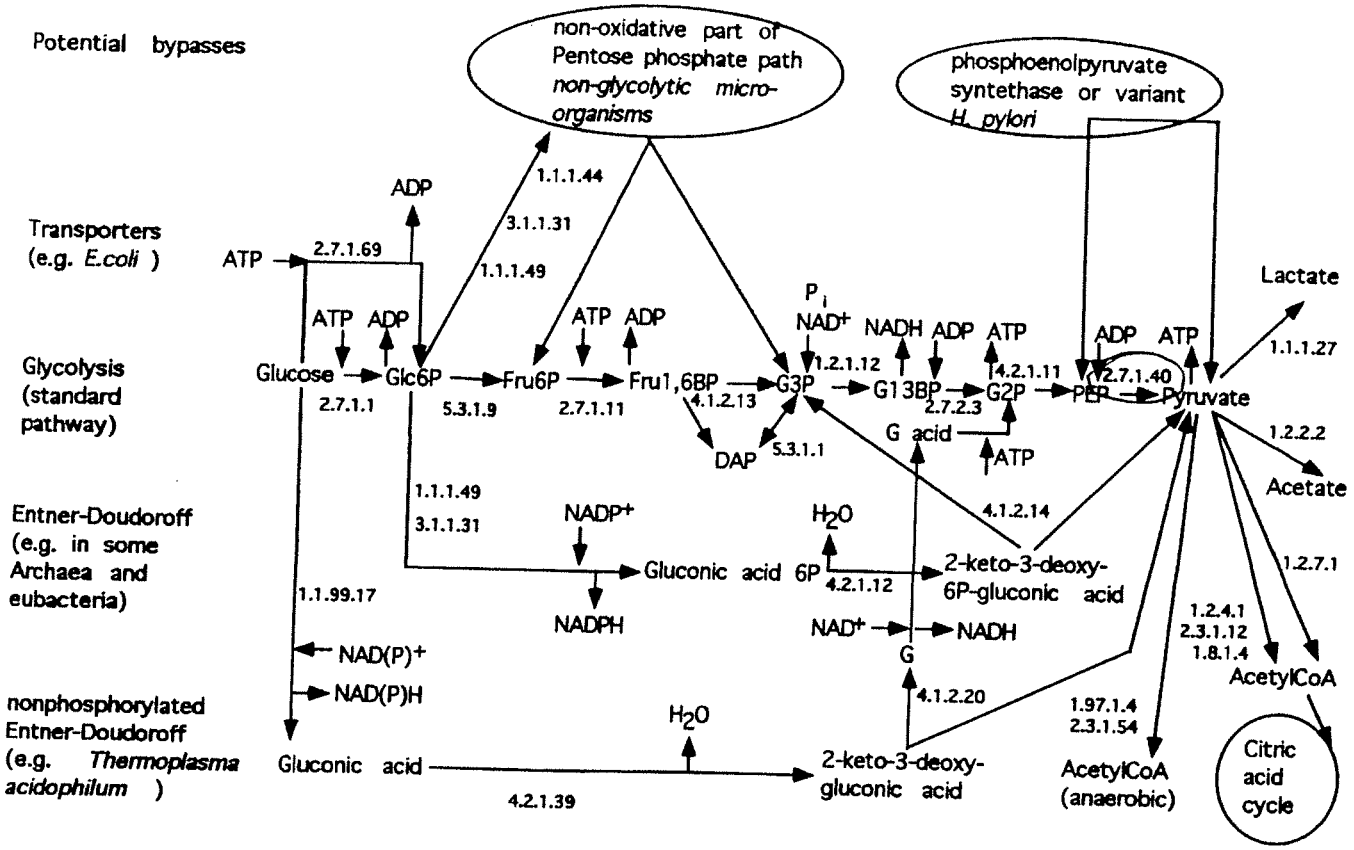
To obtain a better overview of the glycolytic enzymes present in the different organisms, the biochemical analysis is extended by an analysis of the different compounds, the use of databases (e.g. TIGR, ExPasy, KEGG, and WIT) and sequence searches (e.g. http://www.ncbi.nlm.nih.gov/BLAST/, http://dove.embl-heidelberg.de/Blast2). Biochemical function, sequence-based enzyme comparisons and detailed enzyme specificities (see also details available on our Web site) are made more accurate by a careful comparison of domain architecture, detection of duplication and gene displacement events, operon organization and gene families. This is more exact than in database overviews that use copious genome data but rely strongly on an automated assignment of enzyme functions by reciprocal gene similarities.

Summarizing these analyses, the flow from glucose 6-phosphate onwards presents a relatively well-conserved region of the glycolytic pathway (Scheme 1). However, for a number of other organisms several glycolytic enzymes are missing, according to comparative sequence analysis and/or biochemical data. As an example, we mention the absence of phosphofructokinase and aldolase activity in *Mycoplasma hominis* [15] or the apparent lack of genes encoding phosphofructokinase and pyruvate kinase in *Helicobacter pylori*, a causative agent of stomach ulcers and chronic ulcerative gastritis [16].

Alternative methods of processing glucose to pyruvate can be compared (Scheme 1). In glycolysis (shown as the central pathway), two molecules of triose are derived from one of hexose; the energy yield is 2 mol of ATP/mol of glucose. The complete glycolytic pathway is present in *E. coli*. This is also true in most eukaryotic (including human) cells. However, a first variation concerns the route of transfer of glucose into the cell.

Table 1 Biochemical characterization of glycolytic enzymes in various species (data from the literature)

Species	Enzyme
Borrelia burgdorferi	Triose-phosphate isomerase, phosphoglycerate kinase [34]; glyceraldehyde-3-phosphate dehydrogenase [35]
Bacillus subtilis	Phosphopyruvate hydratase, fructose bisphosphate aldolase A, phosphoglycerate kinase, glyceraldehyde-3-phosphate dehydrogenase, pyruvate dehydrogenase (B, C and D subunits) [36]; phosphofructokinase B [37]; glucokinase [27] glucokinase *yajF* (measured for the homologous sequence from *Streptomyces coelicor* [26a])
Chlamydia trachomatis	Phosphoglucoisomerase (measured in *Chlamydia psittaci* [14])
Escherichia coli	Standard glycolysis reviewed in [25]; class I aldolase (EC2097) [30]
Haemophilus influenzae	PEP phosphotransferase system [38]; phosphoglucoisomerase [39]
Helicobacter pylori	Glucose-6-phosphate dehydrogenase, 6-phosphogluconolactonase, 6-phosphogluconate dehydratase, 2-keto-3-deoxy-6-phosphogluconic acid aldolase, triose-phosphate isomerase glucokinase [40]; pyruvate synthase [32]
Mycoplasma genitalium, *M. pneumoniae*	PEP phosphotransferase system, phosphoglucoisomerase, phosphofructokinase, fructose bisphosphate aldolase, triose-phosphate isomerase, glyceraldehyde 3-phosphate dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, phosphopyruvate hydratase, pyruvate kinase, pyruvate dehydrogenase [15]
Methanococcus jannaschii	Phosphofructokinase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate mutase, phosphoglycerate kinase, triose-phosphate isomerase, fructose bisphosphate aldolase, phosphoglucoisomerase (all measured in the related *Methanococcus maripaludis* [41]); class I aldolase activity has been measured in Halobacteria [42]; pyruvate synthase [43]; dihydrolipoamide dehydrogenase A (MJ0636) (measured in the *Haloferax volcanii* homologue [44])
Mycobacterium tuberculosis	Polyphosphate glucokinase, using ATP or polyphosphate as phospho donor [45]; pyruvate dikinase measured in the related (high-GC, Gram-positive) *Propionibacterium shermanii* [46]
Pyrococcus horikoschii	Glucokinase, phosphoglucoisomerase, fructose bisphosphate aldolase (all measured in *Pyrococcus furiosus*; the glucokinase is ADP dependent [23]); glyceraldehyde-3-phosphate oxidoreductase (measured in *P. furiosus* [26])
Synechocystis	Glyceraldehyde-3-phosphate dehydrogenase [47]
Saccharomyces cerevisiae	Standard glycolysis (reviewed in [48])



Scheme 1 Pathway plasticity incorporating genomic and biochemical data for variants of the glycolytic pathway

Alternative routes to the glycolytic pathway (centre) found in various species as shorter by-passes (top) or variant pathways (bottom) are shown. Paths and species examples are described on the left. Arrows symbolize the substrate flux of the enzymic reactions. The release or consumption of ATP, ADP, NAD(P)(H), P_i and water is indicated. Enzyme numbers and names are as given and summarized in Table 2. Hexokinase, EC 2.7.1.1 (shown, typical for the standard pathway in higher organisms) is replaced by glucokinase (EC 2.7.1.2) in prokaryotes. Phosphoglycerate mutase (EC 5.4.2.1) is located after phosphoglycerate kinase (EC 2.7.2.3), but not shown to save space. To illustrate pathway plasticity further, the following additional enzymes are shown: L-lactate dehydrogenase (EC 1.1.1.27); pentose phosphate pathway start: glucose 6-phosphate 1-dehydrogenase (EC 1.1.1.49), 6-phosphogluconolactonase (EC 3.1.1.31), phosphogluconate dehydrogenase (decarboxylating) (EC 1.1.1.44); non-phosphorylated Entner–Doudoroff pathway: glucose dehydrogenase (EC 1.1.99.17). Gluconate dehydratase (EC 4.2.1.39), 2-dehydro-3-deoxyglucarate aldolase (EC 4.1.2.20). Abbreviations: DAP, dihydroxyacetone phosphate; Fru6P, fructose 6-phosphate; Fru1,6BP, fructose 1,6-bisphosphate; G, glyceraldehyde; G acid, glyceric acid; Glc6P, glucose 6-phosphate; gluconic acid 6P, 6-phosphogluconic acid; G3P, 3-phosphoglycerate; G13BP, 1,3-bisphosphoglycerate; G23BP, 2,3-bisphosphoglycerate; G2P, 2-phosphoglycerate; Rib5P, ribulose 5-phosphate.

In contrast with higher organisms, the major *E. coli* pathway for glucose transport uses the phosphotransferase system glucose transporter that uses phosphoenolpyruvate (PEP) as an energy-rich compound to phosphorylate glucose directly to glucose 6-phosphate (second line in Scheme 1; this transporter is present in several other organisms).

If well-known glycolytic enzymes are not found in a species, the question arises of which alternative enzymes, if any, might compensate for them. Pyruvate kinase is missing from *H. pylori*. Sequence analysis shows that a homologue of PEP synthase is present in *H. pylori* (indicated in Scheme 1). If properly adapted, it could compensate for the lack of pyruvate kinase. Note that pyruvate kinase is not as resistant to low pH as PEP synthase [17]. *H. pylori* has to adapt to the highly acidic environment of the stomach and several possible species-specific adaptations of *H. pylori* have been described recently [18].

The connections of the glycolytic pathway to other parts of metabolism such as the PPP allow further plasticity (shown in the centre of the top line of Scheme 1).

Besides detours in transport or by-passes to the standard glycolytic pathway, more dissimilar alternatives for converting glucose to pyruvate exist in other organisms (Scheme 1, lower part). The Entner–Doudoroff pathway is used instead of glycolysis in some bacteria [19]. The ATP yield is only 1 mol/mol of glucose. The advantage of this route is a direct yield of NADPH for biosynthesis. From biochemical data the non-phosphorylated Entner–Doudoroff pathway (Scheme 1, bottom) is yet another alternative route for the conversion of glucose to pyruvate. This is considered to be an example of palaeometabolism [6]. It is present, for instance, in some Archaea such as *Thermoplasma acidophilum* [20]. The direct conversion of glucose to gluconic acid is not optimized in this ancient pathway and there is no ATP yield from glucose.

In summary, after the establishment of a solid knowledge basis from literature and experimental data (Table 1) and from the extensive use of sequence and genome analysis, in the present paper the full repertoire of glycolytic enzymes present in the analysed organisms is better characterized, including organisms that are difficult to analyse biochemically such as intracellular parasites. Furthermore, pathway plasticity is examined and shows alternative pathways and detours within standard glycolysis for various organisms (Scheme 1).

Table 2 Enzyme families and isoenzymes encoded in the different genomes

Enzyme families from the standard glycolytic pathway, hexose phosphorylation, Entner–Doudoroff pathway, pyruvate dehydrogenase and three alternatives in pyruvate processing are compared. Closely related enzyme activities including isoenzymes are grouped together to allow easy comparison of their respective species distributions. Enzymes are labelled by their EC number and enzyme names are given for all enzyme families compared. Seventeen species are compared with the following species codes. SY, *Synechocystis* sp. PCC6803 (blue-green algae); CT, *Chlamydia trachomatis*. Spirochaetes: TP, *Treponema pallidum*; BB, *Borrelia burgdorferi*. Gram-positive bacteria: BS, *Bacillus subtilis*; TB, *Mycobacterium tuberculosis*; MG, *Mycoplasma genitalium*; MP, *Mycoplasma pneumoniae*. Proteobacteria (Gram-negative bacteria): EC, *Escherichia coli*; HI, *Haemophilus influenzae*; HP, *Helicobacter pylori*. Further bacteria: AQ, *Aquifex aeolicus*. Archaea: AF, *Archaeoglobus fulgidus*; MJ, *Methanococcus jannaschii*; MT, *Methanobacterium thermoautotrophicum*; PH, *Pyrococcus horikoshii*. Eukaryotic species: Y, yeast. Completeness of the upper (hexoses) and lower (trioses) part of glycolysis as well as the pyruvate dehydrogenase (lipoamide) subunits or the pyruvate formate-lyase path is indicated by 'co'; important missing enzyme activities are either phosphofructokinase (pfk) or pyruvate kinase (pyk). CoT indicates that the upper part of glycolysis is complete, assuming glucose phosphorylation by a suitable transport system. To save space, the specific enzyme sequences are not listed; however, genome numbers extended by their respective species code allow the retrieval of each enzyme sequence. Several places where corrections have been made to the assignment of enzyme families are indicated: *, changes with respect to KEGG; †, changes with respect to published data from genome sequencing efforts; ‡, changes with respect to WIT. Our Web page on glycolytic enzymes can be consulted for additional corrections and justification details. Abbreviation: PTS, phosphotransferase system.

	Genome number																
Species code...	SY	CT	BS	TP	TB	BB	EC	HI	HP	MG	MP	AQ	AF	MJ	MT	PH	Y
Upper part of glycolysis																	
2.7.1.2 Glucokinase (glk family)	sll0593						2388		1103								
2.7.1.2 Glucokinase (yajF family)	slr0329		glck			0831	0394†	0182				1496					
2.7.1.2 Glucokinase† (yhc1 family)					0650		3222	0144									
2.7.1.2 Yeast type glucokinase				0505													YCL040W
5.3.1.9 Glucose-6-phosphate isomerase	slr1349	378	pgi	0475	0946	0730	4025	1576	1166	111	0583	–†		1605			YBR196C
2.7.1.11 6-Phosphofructo-kinase (family A)	sll1196	205*		0108	3010	0727*	3916	0982		215	0535	1708					YGR240C
2.7.1.11 6-Phosphofructokinase (second copy of family A)	sll0745	207		0542		0020*											
2.7.1.11 6-Phosphofructo-kinase (family B)			frub		2029	0630	1723										
2.7.1.56 1-Phosphofructo-kinase							2168	0447		063	0076						
4.1.2.13 Fructose-bisphos-phate aldolase					0363	0445	2925	0524									YKL060C
4.1.2.13 Fructose-bisphos-phate aldolase (second family)	sll0018		fbaa	0662					0176	023	0129	1390					
4.1.2.13 Fructose-bisphos-phate aldolase (class I type family)†‡		215					2097						0108	400	0579	0082	
Completeness	co	coT	co	co	co	co	co	co	pfk	coT	coT	co					co
Lower part of glycolysis																	
5.3.1.1 Triose-phosphate isomerase	slr0783	328	tpi	0537	1438	0055	3919	0678	0194	431	0213	0360	1304	1528	1041	1884	YDR050C
1.2.1.12 Glyceraldehyde -3-phosphate DH (phosphorylating)	slr0884	505	gapa	0844	1436	0057	1779	0001	0921	301	0410	1065	1732	1146	1009	1830	YJL052W
1.2.1.12 Second family	sll1342		gapb				1417		1346								
1.2.7.– Oxidoreductase family														1185		0457	
2.7.2.3 Phosphoglycerate kinase	slr0394	693	pgk	0538	1437	0056	2926	0525	1345	300	0411	0118	1146	0641	1042	1218	YCR012W*
5.4.2.1 Phosphoglycerate mutase (family A)		722		0168	0489	0658	0755	0757									YKL152C
5.4.2.1 Phosphoglycerate mutase (family B)	slr1124		yhfr		2228†		4395					1744					YKR043C
5.4.2.1 Phosphoglycerate mutase (2,3-bisphospho-glycerate independent form) *†‡	slr1945		pgm				3612		0974	430	0214	0542	1751	1612	1591	0037	
4.2.1.11 Phosphopyruvate hydrolase (enolase)	slr0752	587	eno	0817	1023	0337	2779	0932	0154	407	0236	0484	1132	0232	0043	1942	YGR254W
2.7.1.40 Pyruvate kinase family A)	sll0587	332			1617	0348	1854	1573		216	0534			0108		0570	
2.7.1.40 Pyruvate kinase (family F)			pyka				1676										YAL038W
2.7.9.2 PEP synthase	slr0301		ptsi	0746	1127		1702		0121			2142	0710	0542	1118	0092	
Completeness	co	co	co	pyk	co	co	co	co	pyk	co	co	pyk	pyk	co	pyk	co	co

Table 2 (*cont.*)

Species code...	Genome number SY	CT	BS	TP	TB	BB	EC	HI	HP	MG	MP	AQ	AF	MJ	MT	PH	Y
Glucose phosphorylation																	
2.7.1.69 PTSII IIA component			ybfs			0559	2417	1711		069	0625						
2.7.1.69 PTSII IIBC component			ybfs			0645	1101			069	0625						
2.7.1.69 *malX* PTS						0645	1621			069	0625						
2.7.3.9 PEP protein phosphatase (enzyme 1 of the phosphotransferase system)		336				0558	2416	1712		429	0215			0542			
Fructose																	
2.7.1.4 Fructokinase-related‡	slr1448		rbsk		2436		3752	0505					0356‡	0406	0404	1845	YCR036W
Entner–Doudoroff pathway																	
4.2.1.12 6-Phosphogluconate dehydratase							1851		1100								
4.1.2.14 2-dehydro 3-deoxy phosphogluconate aldolase	sll0107		kdga	0568			1850	0047	1099								
Pyruvate dehydrogenase																	
1.2.4.1 E1 component (lipoamide)					2241		0114	1233									
1.2.4.1 E1 α subunit‡	slr1934	340	pdha		2497					274	0446						YER178W
1.2.4.1 E1 β subunit ‡	sll1721	246	bfubab		2496					273	0447						Y8R221C
2.3.1.12 Dihydrolipoamide S-acetyltransferase (E2 component)			pdhc		2215		0115	1232		272	0448						YNL071W
2.3.1.12 Second copy of E2	sll1841	400	bfmbb				0727										YDR148C
1.8.1.4 Dihydrolipoamide dehydrogenase	slr1096	557	pdhb		0462		0116	1231		271	0449						YFL018C
Alternative pyruvate processing																	
1.2.2.2 Pyruvate oxidase (cytochrome)			ydap				0871										
1.97.1.4 Formate acetyltransferase activating enzyme															0476		
1.97.1.4 Second copy					3138		0902	0179				1861	1450	0808	0345	1391	
2.3.1.54 Formate C-acetyltransferase (pyruvate formate-lyase)							0823	0180									
2.3.1.54 Second copy							0903	0180					1449		0346		
1.1.1.1 Alcohol dehydrogenase														1227	1586		
1.2.7.1 Pyruvate synthase (subunit A)‡	sll0741			0939			1241						0024			0743	YGL256W
1.2.7.1 Pyruvate synthase (subunit) B	sll0741			0939			1378		1110			1167	1701	0267	1739	0680	
1.2.7.1 Pyruvate synthase (subunit C)	sll0741			0939			1378		1111			1168	1702	0266	1738	0681	
1.2.7.1 Pyruvate synthase (subunit D)	sll0741			0939			1378		1108			1169	1699	0269	1740	0678	
									1109			1171a	1700	0268	1740	0679	

Algebraic pathway analysis

For a further assessment of the different paths detected or suggested for glucose in the different organisms (Scheme 1), we calculate substrate fluxes depending on the enzymic activities present ('algebraic pathway analysis'). This checks the consistency of the different enzyme sets deduced from the biochemical and genomic analyses. Moreover, it reveals which glycolytic bypasses are viable, where futile cycles are (often indicating allosteric regulation of key enzymes involved) and which paths might efficiently be blocked pharmacologically in different organisms.

We have developed elementary-mode analysis as a tool for this analysis [8,9], including specific software (see the Materials and methods section) [10]. A related approach has been developed by Mavrovouniotis et al. [21]. Starting from a minimum amount of necessary information, including knowledge of the stoichiometry and thermodynamic directionality of reactions, all non-decomposable flux distributions ('elementary flux modes') at stationary states can be computed. These can be interpreted as biochemical pathways or futile cycles in the system.

In particular, alternative routes for glycolysis (Scheme 1, first line) can be re-examined by applying the method. We give three examples.

In *Mycoplasma hominis*, phosphofructokinase, aldolase and glucose-6-phosphate dehydrogenase seem to be absent on the basis of biochemical data [15]. It has been argued [15] that these species instead channel glucose by the PPP, yielding glyceraldehyde 3-phosphate plus ribose for nucleotide biosynthesis. However, on the assumption that the enzymes of the non-oxidative PPP are operative, computing the elementary modes of the combined glycolysis and PPP scheme shows that this is not sufficient to by-pass the above-mentioned enzymes. This is indicative of a missing link in the metabolism of *Mycoplasma hominis*. One possibility would be the presence of fructose-6-phosphate phosphoketolase (as found, for example, in *Acetobacter xylinum*; see the BRENDA database via http://srs.ebi.ac.uk/srs5bin) and of phosphate acetyltransferase (as found, for example in *E. coli* and several *Mycoplasma* species; see the ENZYME database, http://www.expasy.ch/enzyme/enzyme_details.html). It can be shown that, with these activities, glucose can indeed be converted into intermediates of the lower glycolytic pathway (elementary mode 1 in the Appendix). The presence of these enzymes in *Mycoplasma hominis* can now be tested experimentally. Another possibility would be the presence of enzymic activities phosphorylating sedoheptulose 7-phosphate and splitting the resulting seduheptulose bisphosphate into dihydroxyacetone phosphate and erythrose 4-phosphate [10]. In addition, elementary-mode analysis indicates that for non-glycolytic species possessing glucose-6-phosphate dehydrogenase activity, an elementary mode by-passing phosphofructokinase via the PPP is available (elementary mode 3 in the Appendix).

Another interesting example is *Methanococcus jannaschii*, in which new connections and pathway routes are apparent. All enzymes of the non-oxidative PPP have been identified in its genome [22]. In contrast, the two dehydrogenases required for the oxidative PPP could not be identified, either in the genome or by biochemical assays. Glycolysis seems to be operative, although hexokinase, 6-phosphofructokinase, fructose bisphosphate aldolase and phosphoglycerate mutase could not be identified in the genome [22]. Assuming that these enzymes or alternative enzymes are present (for instance, there are some indications of ADP-dependent sugar kinases in Archaea; [22,23]), we have computed the elementary modes for this system. One of these (elementary mode 2 in the Appendix) corresponds to the transformation of glucose 6-phosphate into ribose 5-phosphate, which is required for nucleotide biosynthesis. Thus the oxidative PPP is not needed for converting hexoses into pentoses under steady-state conditions. This is in agreement with results on glucose-6-phosphate dehydrogenase deficiencies [24].

In summary, algebraic pathway analysis examines admissible routes through metabolism (Appendix). We applied it here to inprove our understanding of the function of different organism-specific sets of glycolytic enzymes. In addition, required enzyme actitivities can be identified better (the example shown is *Mycoplasma hominis*), as well as additional viable flux modes, for instance involving parts of glycolysis and PPP (the example shown is *M. jannaschi*).

Comparative genome analysis of glycolysis

To characterize further the species-specific differences in glycolytic enzymes, we now compare in detail the biochemical data and results and the algebraic pathway analysis with the sets of enzymes encoded in completely sequenced genomes. Enzyme families and isoenzymes (often differently allosterically regulated) become apparent; subsequently we compare each reaction step and the whole pathway.

Identifying enzyme and isoenzyme families

The result of a systematic genome sequence search to identify glycolytic enzymes is shown in Table 2. The following species and clades are compared (from left to right in Table 2): Blue-green algae (*Synecchocystis* sp.), *C. trachomatis*, *Bacillus subtilis*, *Treponema pallidum*, *Mycobacterium tuberculosis*, *Borrelia burgdorferi*, three proteobacteria (including *E. coli*), *Mycoplasma genitalium* and *Mycoplasma pneumoniae*, the bacterium *Aquifex aeolicus*, four Archaea and finally yeast (the only eukaryote). The resulting enzyme patterns of completeness of glycolysis and by-passes in the different species are summarized in Figure 1.

However, the different enzyme and isoenzyme families with their differences in regulation and specificities are shown in Table 2, which also describes their respective variations in sequence and corresponding structure; genome accession numbers in Table 2 permit the retrieval of the exact amino acid sequence for each enzyme family member. This also permits a comparison of our data with published genome sequencing efforts and other genome-based analyses. Furthermore, with this information, sequences from newly sequenced genomes can be added much more easily to the correct glycolytic enzyme family. The classification of the encoded enzyme activities shown in Table 2 has been derived by a careful comparison of several approaches as mentioned in the Materials and methods section and includes corrections compared with standard databases (KEGG, WIT and genome sequencing efforts). Nevertheless, our analysis includes many genome-based predictions. Biochemical analyses will have to confirm these; related as well as completely new enzymic activities might become apparent as the biochemical analysis of the various species proceeds.

Further considerations and comparisons that we made to evaluate encoded enzyme specificities and family relations as listed in Table 2 are given, including additional corrections and justification details, on our glycolytic enzyme Web page (http://www.bork.embl-heidelberg.de/Genome/glycolytic_enzymes). The analyses included a further, less related glucokinase-like family as well as details on the polyphosphate glucose phosphotransferase (EC 2.7.1.63) in *Mycobacterium tuberculosis*, and related glycolytic and non-glycolytic enzymic activities such as 2-keto-3-deoxygluconate kinases (EC 2.7.1.45, involved in pectin degradation) and gluconate kinases (EC 2.7.1.12), PEP synthases (EC 2.7.9.2) and pyruvate dikinases (EC 2.7.9.1), dihydroxyacid dehydratase (EC 4.2.1.9) and 6-phosphogluconate dehydratase (EC 4.2.1.12) (see our Web page on glycolytic enzymes for details).

Enzyme and isoenzyme families

Many different families of glycolytic enzymes can be distinguished (Table 2). *E. coli*, with its large genome and high adaptability, is particularly rich in glycolytic enzymes, including several by-passes. The closely related *Haemophilus influenzae* shares a large set of enzymes with *E. coli* [7]. Alternatives to the *E. coli* set of enzyme families are seen in the more distantly related species (Table 2), such as the use of a class I type aldolase in *C. trachomatis* or a yeast-type glucokinase not only in yeast but also in *T. pallidum*.

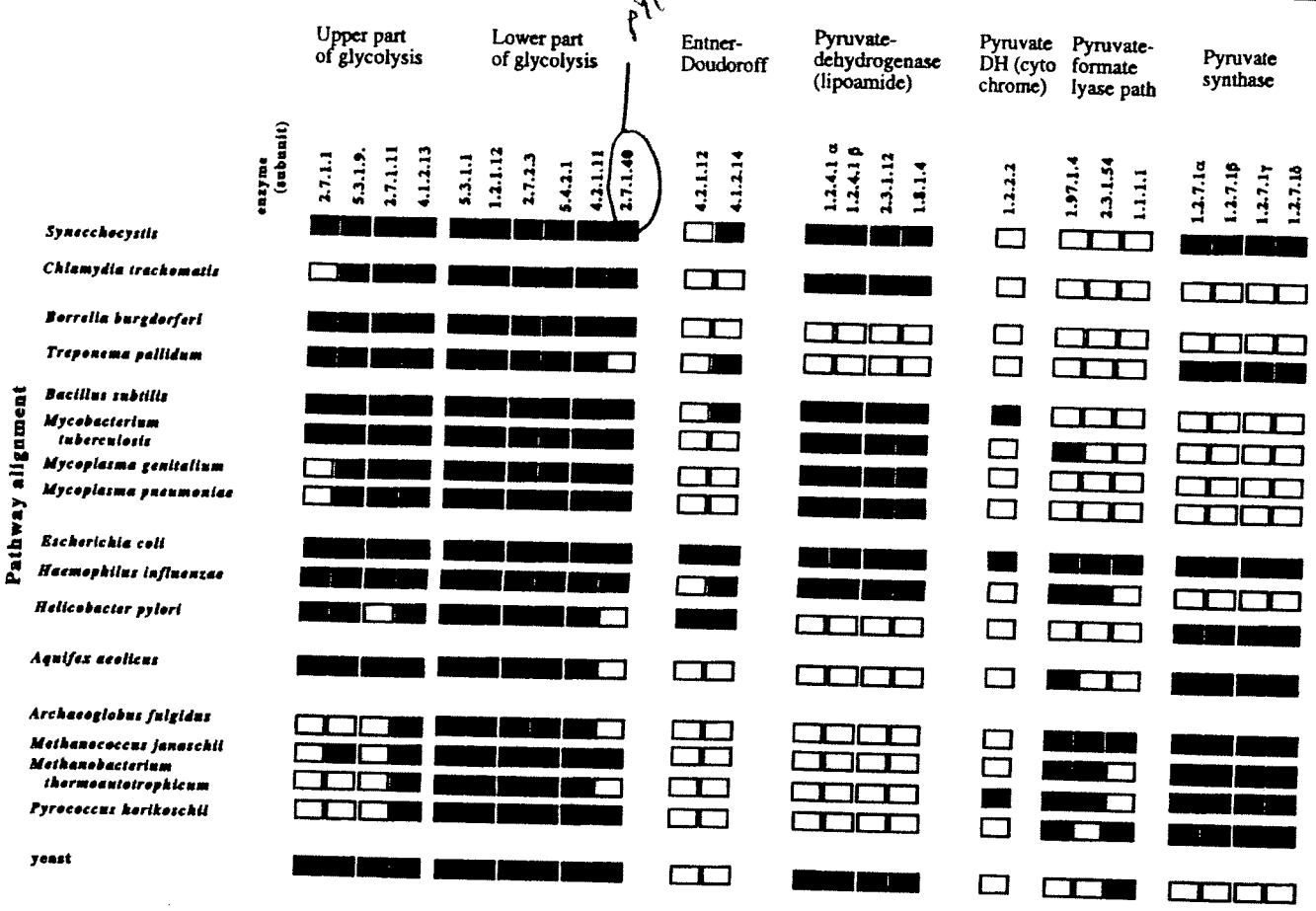


Figure 1 Pathway alignment for glycolysis, Entner–Doudoroff pathway and pyruvate processing

Enzymes for each pathway part (top; EC numbers and enzyme subunits are given below these) are compared in 17 organisms and represented as small rectangles. Filled and empty rectangles indicate the presence and absence respectively of enzyme-encoding genes in the different species listed at the left. Further details are given in the text; different isoenzymes and enzyme families are listed in Table 2.

The isoenzymes identified for different organisms and listed in Table 2 are interesting because they are often regulated differently. For example, different allosteric activators are known for the pair 6-phosphofructokinase, family A (6-phosphofructokinase number 1, main 6-phosphofructokinase, activated by any nucleotide diphosphate, inhibited by PEP) and 6-phosphofructokinase, family B (6-phosphofructokinase number 2, which supplies only 10% of the activity of 6-phosphofructokinase, family A, and is inhibited by ATP) or the pair pyruvate kinase, family F (pyruvate kinase 1, activated by fructose 1,6-bisphosphate) and pyruvate kinase, family A (pyruvate kinase II, allosterically activated by AMP and other metabolites; different structure, expression and kinetics from those of pyruvate kinase, family F) on the basis of data from *E. coli* [25]. A larger adaptive capacity to changes in pH or oxygen content might also be implied in such duplications for several species (Table 2). Furthermore, we show that isoenzymes in glycolysis are widespread among prokaryotes and are not restricted to eukaryotes. Moreover, some new cases are suggested in which different allosteric regulation might be expected (e.g. AQ0542 and AQ1744). The recently discovered [26] induction of the glyceraldehyde-3-phosphate ferredoxin oxidoreductase (gapor) during glycolysis in hyperthermophilic *Pyrococcus furiosus*, while repressing the standard glyceraldehyde-3-phosphate dehydrogenase, is an intriguing example. Our comparative genome analysis suggests a similar induction for PH0457 and MJ1185 (glyceraldehyde-3-phosphate oxidoreductase) where as PH1830 and MJ1146 (glyceraldehyde-3-phosphate dehydrogenase) are repressed during glycolysis.

Glucokinase

Following the direction of glycolytic flow, we shall next discuss the enzymes compared in Table 2 from top to bottom, including clade-specific or species-specific peculiarities. For the different glucokinases we shall illustrate in more detail the assignment of the genome-encoded sequences to enzyme families (similar considerations for the other enzyme families listed in Table 2 can be found on our Web page). The *glk* glucokinase is the major glucokinase in *E. coli*. If the phosphotransferase system is genetically defective, the presence of this glucokinase gene is essential for residual growth on glucose [25]. Regarding the *yajF* glucokinase family, Angell et al. [26a] have biochemically measured glucokinase activity for its sequence homologue from *Streptomyces coelicolor*. Skarlatos and Dahl [27] did the same for the *B. subtilis* glucokinase from the glk family (*BSglck* in Table 2) and showed by insertional mutagenesis that, in contrast with the situation in *E. coli*, this glucokinase is the major one for *B. subtilis*. However, some of the proteins of this glucokinase family

(for example, the *E. coli* gene product) might function as repressors in carbohydrate utilization (as they are for example related to the *E. coli* xylulose repressor and the *E. coli yjcT* gene), with the glucokinase activity being complemented by a regulatory repressor function. For the third family shown in Table 2, genetic experiments and sequence analyses suggest that in *E. coli*, the preferred substrate for the encoded glucokinase from the yhcI family is *N*-acetylmannosamine [28]. Another enzyme (number 3437 in *E. coli* and annotated as glucokinase) is in fact a gluconokinase and uses gluconate. An ADP-dependent glucokinase has been characterized in *Pyrococcus furiosus* [23] (Table 1). Relatives of this might be present in other Archaea. In the pathogenic *T. pallidum* a gene encodes an enzyme homologous with the eukaryotic yeast glucokinase (Table 2). This is likely to be a horizontal gene transfer from eukaryotes, because prokaryotic homologues are far more divergent (see [29] for further comparisons within this large family of sugar kinases). In Mycobacteria a polyphosphate glucokinase (EC 2.7.1.63) is also present.

Upper part of glycolysis (processing of hexoses)

Enzymes for this part of glycolysis seem to be missing from several Archaean species. In contrast, the bacterial species are in most cases relatively complete in this part; the absence of 6-phosphofructokinase in the pathogenic *Helicobacter pylori* is an exception. 6-Phosphofructokinase appears in two different isoenzyme families, A and B. The fructose bisphosphate aldolase gene found in *C. trachomatis* (CT215) encodes a class I fructose bisphosphate aldolase that has recently been characterized biochemically [30]. Homologues, including Archaean genes, are given in Table 2. Moreover, note that in Table 1 there are some additional glycolytic activities measured in *Methanococcus maripaludis* and *P. furiosus* for which genome counterparts have not yet been identified.

Lower part of glycolysis

This route is shown to be operative in many species (Table 2; Figure 1). Four enzymes in the lower part of glycolysis are conserved in all species (Table 2); phosphoglyceromutase is also present in all species but uses different variants. No gene has been found for pyruvate kinase in 5 of the 17 species considered. Perhaps this indicates the presence of yet another undetected enzyme that has displaced the 'classical' form. Note that PEP-driven transport systems for glucose import and phosphorylation also allow PEP to be turned into pyruvate without requiring pyruvate kinase [31]. A third alternative is the use of pyruvate dikinase (present in *Mycobacterium tuberculosis* and *T. pallidum*). In the remaining species, completion of the lower part of glycolysis is possible because several versions of phosphoglyceromutase and pyruvate kinase are identified.

Glucose transport with phosphorylation

A range of specific transporters is able to phosphorylate glucose, alleviating the need for a hexokinase or glucokinase for instance in compact genomes (Mycoplasmae). Cross-species comparison confirms that the *E. coli*-like protein-N(PI)-phosphohistidine-sugar phosphotransferases as well as the *E. coli*-type phosphoenolpyruvate protein phosphatase have a limited species range (Table 2). A range of other sugar transporters are present. Fructokinase is one example (note that in *E. coli* the specific fructokinase is plasmid-borne; the chromosomal enzyme is a ribokinase with broader specificity). Another entry point for fructose into glycolysis is the formation of fructose 1,6-bisphosphate by the action of the *fruK* gene product (i.e. by the 1-phosphofructokinase instead of the closely related 6-phosphofructokinase; Table 2).

Entner–Doudoroff pathway

This is functional in *E. coli* and *H. pylori*. In several other species, genes encoding 2-dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) are present (Table 2) (some potential phosphogluconate dehydratase genes are listed on our Web site).

Pyruvate processing

The specific pyruvate dehydrogenase repressor used in *E. coli* is not found in the other species (Table 2). Pyruvate dehydrogenase (lipoamide) subunits are missing from *Borrelia*, *Helicobacter*, *T. pallidum*, *Aquifex* and Archaea. Instead, these species, except *Borrelia*, use a pyruvate synthase (EC 1.2.7.1) for this reaction [32]. Furthermore, some species harbour pyruvate dehydrogenase (cytochrome), EC 1.2.2.2, which decarboxylates pyruvate to acetate and carbon dioxide with a lower energetic efficiency than pyruvate dehydrogenase (lipoamide), EC 1.2.4.1, or the pyruvate formate-lyase path, which converts pyruvate to acetyl-CoA under strict anaerobic conditions [composed of the pyruvate formate-lyase-activating enzyme (EC 1.97.1.4)] and formate acetyltransferase (EC 2.3.1.54); the further path uses for instance alcohol dehydrogenase/acetaldehyde dehydrogenase).

Comparing glycolytic pathways between species

Glycolysis is present to some extent across all species (Figure 1), which confirms that it is one of the most basic pathways in life. The lower part of glycolysis (trioses) is best conserved. Different enzymes of the upper part of glycolysis are missing from Archaea. The first phosphorylation step provided by glucokinase or hexokinase in many organisms seems to be replaced by transport phosphorylation in some species. The lower part of glycolysis also serves anabolic and conversion functions in amino acid and carbohydrate metabolism. It has more stringent requirements for being maintained. It seems to be the older part of this metabolic pathway. Its widespread distribution suggests that it was present even in the last common ancestor. The Entner–Doudoroff pathway (Scheme 1), although generally believed to be a more ancient catabolic route for glucose than glycolysis [6,33], is not as well conserved. It is present only in the Gram-negative bacteria in our comparison. Both the phosphorylated and the non-phosphorylated Entner–Doudoroff pathways have been described in several other Archaea and Eubacteria (Scheme 1) [33] and additional non-homologous sequences might have escaped our detection. These observed as well as potential variations in the presence of different Entner–Doudoroff enzymes again indicate that the pathway need not be well conserved. The secondary loss of an inefficient old pathway is another possibility to explain its lack of conservation, but it seems not to be that inefficient as it successfully co-exists with glycolysis in several species.

Pyruvate dehydrogenase (lipoamide) is often used in Eubacteria. Its anaerobic counterpart in Archaea is pyruvate synthase; pyruvate formate-lyase is also often present in Archaea. *E. coli*, an adaptation generalist with its large genome, harbours all three alternatives; as for the parasites, *B. burgdorferi* has none and *T. pallidum* (probably) has only the pyruvate synthase (EC 1.2.7.1).

Correlations between the different metabolic steps analysed exist: for instance, glycolysis is one of the main generators of

pyruvate in the cell; for all of the organisms analysed, if pyruvate dehydrogenase is present, glycolysis is also present. Similarly a lack of pyruvate kinase is correlated with a lack of pyruvate dehydrogenase (Figure 1). However, alternative by-passes (Scheme 1) for each of the steps as well as the patterns for Archaea (Figure 1) or the different pyruvate oxidoreductases [counterpart in Archaea is pyruvate synthase (EC 1.2.7.1)] (Figure 1) show that such metabolic correlations are not strict.

In summary, comparative genome analysis identifies glycolytic isoenzymes such as different versions of phosphoglycerate mutase preferentially used in Archaea and bacteria or different types of glucokinase. Species-specific enzymes become apparent, such as class I aldolases in *Chlamydia* and Archaea, the absence of enzymes (e.g. the lack of 6-phosphofructokinase and pyruvate kinase in *H. pylori* and several Archaea) and by-passes such as Entner–Doudoroff enzymes in some species (Table 2). Pyruvate dehydrogenase (oxygen might be present in the cell; many Eubacteria) and pyruvate synthase (anaerobic lifestyle; Archaea) are good examples of observed phenotypic adaptations. The variance in glycolysis is highlighted by the fact that none of the species compared show the same pattern of enzymes present, with the exception of the closely related *Mycoplasma* species (Figure 1).

Conclusion

We show that even the simple and basic pathway of glycolysis has a high plasticity and versatility both in a single species (e.g. *E. coli*) and in a multi-genome comparison (Scheme 1).

Currently the adaptability of organisms seems to be underestimated if only the textbook pathway is considered. Except for the two *Mycoplasma* species, no two species encode the same pattern of glycolytic enzymes (Figure 1). We find by elementary-mode analysis that even the blocking of central enzymes such as aldolase as an antibiotic strategy might not be effective in some pathogens because alternative substrate fluxes have to be considered (Appendix). We reveal different species-specific sets of isoenzymes (e.g. three types of phosphoglyceromutase), alternative enzymes (such as pyruvate dehydrogenase and pyruvate synthase) and by-passes (e.g. Entner–Doudoroff) by combining biochemical data (Table 1) with comparative genome analysis (Table 2).

The differences identified can now be examined further biochemically. Several are interesting for medicine and drug design, such as parasite-specific adaptations in *Mycoplasma* and *H. pylori*, the different sugar transporters, or the polyphosphate glucokinase in *Mycobacterium tuberculosis*. Starting points for metabolic engineering are also apparent, e.g. a higher yield of NADPH by expressing enzymes of the Entner–Doudoroff pathway or for the improvement of other product-to-substrate ratios in a micro-organism.

We thank Warren C. Lathe III for stylistic corrections. Research was supported by Deutsche Forschungsgemeinschaft (DFG) and Bundesministerium für Bildung und Forschung (BMBF).

REFERENCES

1 Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M. and Yuan, Y. (1998) J. Mol. Biol. **283**, 707–725
2 Koonin, E. V., Tatusov, R. L. and Galperin, M. Y. (1998) Curr. Opin. Struct. Biol. **8**, 355–363
3 Tatusov, R. L., Koonin, E. V. and Lipman, D. J. (1997) Science **278**, 631–637
4 Goto, S., Bono, H., Ogata, H., Fujibuchi, W., Nishioka, T., Sato, K. and Kanehisa, M. (1997) in Proceedings of the Pacific Symposium on Biocomputing '97 (Altman, R., Dunker, K., Hunter, L and Klein, T., eds.), pp. 175–186, World Scientific Publishers, Singapore
5 Galperin, M. Y. and Brenner, S. E. (1998) Trends Genet. **14**, 332–333
6 Meléndez-Hevia, E., Waddell, T. G., Heinrich, R. and Montero, F. (1997) Eur. J. Biochem. **244**, 527–543
7 Tatusov, R. L., Mushegian, A. R., Bork, P., Brown, N. P., Hayes, W. S., Borodovsky, M., Rudd, K. E. and Koonin, E. V. (1996) Curr. Biol. **6**, 279–291
8 Schuster, S. and Hilgetag, C. (1994) J. Biol. Syst. **2**, 165–182
9 Schuster, S., Dandekar, T. and Fell, D. (1999) Trends Biotechnol. **17**, 53–60
10 Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J. C., Montero, F. and Schuster, S. (1999) Bioinformatics **15**, 251–257
11 Hollander, R. (1976) Antonie van Leeuwenhoek **42**, 429–444
12 Stanton, T. B. (1989) Appl. Environ Microbiol. **55**, 2365–2371
13 Ojcius, D. M., Degani, H., Mispelter, J. and Dautry-Varsat, A. (1998) J. Biol. Chem. **273**, 7052–7058
14 Moulder, J. W. (1966) Annu. Rev. Microbiol. **20**, 107–130
15 Pollack, J. D., Williams, M. V. and McElhaney, R. N. (1997) Crit. Rev. Microbiol. **23**, 269–354
16 Tomb, J.-F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A. et al. (1997) Nature (London) **388**, 539–547
17 Podesta, F. E. and Plaxton, W. C. (1992) Biochim. Biophys. Acta **1160**, 213–220
18 Huynen, M. A., Dandekar, T. and Bork, P. (1998) FEBS Lett. **426**, 1–5
19 Danson, M. J. and Hough, D. W. (1992) in The Archaebacteria: Biochemistry and Biotechnology (Danson, M. J., Hough, D. W. and Lunt, G. G., eds.), Biochemical Society Symposium Series No, 58, pp. 7–21, Portland Press, London
20 Fields, J. H. A. (1987) Can. J. Zool. **66**, 1036–1040
21 Mavrovouniotis, M. L., Stephanopoulos, G. and Stephanopoulos, G. (1990) Biotechnol. Bioeng. **36**, 1119–1132
22 Selkov, E., Maltsev, N., Olsen, G. J., Overbeek, R. and Whitman, W. B. (1997) Gene **197**, GC11–GC26
23 Kengen, S. W., Tuininga, J. E., de Bok, F. A., Stams, A. J. and de Vos, W. M. (1995) J. Biol. Chem. **270**, 30453–30457
24 Pandolfi, P., Sonati, F., Rivi, R., Mason, P., Grosveld, F. and Luzzatto, L. (1995) EMBO J. **14**, 5209–5215
25 Fraenkel, D. G. (1996) in *E. coli* and *Salmonella*: Cellular and Molecular Biology (Neidhardt, F. C., ed.), pp. 189–198, FC American Society for Microbiology (ASM) Press, Washington, DC
26 Van der Oost, J., Schut, G., Kengen, S. W., Hagen, W. R., Thomm, M. and de Vos, W. M. (1998) J. Biol. Chem. **273**, 28149–29154
26a Angell, S., Schwarz, E. and Bibb, M. J. (1992) Microbiol. **6**, 2833–2844
27 Skarlatos, P. and Dahl, M. K. (1998) J. Bacteriol. **180**, 3222–3226
28 Plumbridge, J. and Vimr, E. (1999) J. Bacteriol. **181**, 47–54
29 Bork, P., Sander, C. and Valencia, A. (1993) Protein Sci. **2**, 31–40
30 Thomson, G. J., Howlett, G. J., Ashcroft, A. E. and Berry, A. (1998) Biochem. J. **331**, 437–445
31 Postma, P. W., Lengeler, J. W. and Jacobson, G. R. (1996) in *E. coli* and *Salmonella* Cellular and Molecular Biology (Neidhardt, F. C., ed.), pp. 1149–1174, FC American Society for Microbiology (ASM) Press, Washington, DC
32 Hughes, N. J., Clayton, C. L., Chalk, P. A. and Kelly, D. J. (1998) J. Bacteriol. **180**, 1119–1128
33 Romano, A. H. and Conway, T. (1996) Res. Microbiol. **147**, 448–455
34 Gebbia, J. A., Backenson, P. B., Coleman, J. L., Anda, P. and Benach, J. L. (1997) Gene **188**, 221–228
35 Anda, P., Gebbia, J. A., Backenson, P. B., Coleman, J. L. and Benach, J. L. (1996) Infect. Immun. **64**, 262–268
36 Schmid, R., Bernhardt, J., Antelmann, H., Volker, A., Mach, H., Volker, U. and Hecker, M. (1997) Microbiology **143**, 991–998
37 Gay, P. and Delobbe, A. (1977) Eur. J. Biochem. **79**, 363–373
38 Macfadyen, L. P., Dorocicz, I. R., Reizer, J., Saier, Jr., M. H. and Redfield, R. J. (1996) Mol. Microbiol. **21**, 941–952
39 Porras, O., Caugant, D. A., Lagergard, T. and Svanborg-Eden, C. (1986) Infect. Immun. **53**, 71–78
40 Hazell, S. L. and Mendz, G. L. (1997) Helicobacter **2**, 1–12
41 Yu, J. P., Ladapo, J. and Whitman, W. B. (1994) J. Bacteriol. **176**, 325–332
42 Krishnan, G. and Altekar, W. (1991) Eur. J. Biochem. **195**, 343–350
43 Shieh, J. S. and Whitman, W. B. (1987) J. Bacteriol. **169**, 5327–5329
44 Vettakkorumakankav, N., Danson, M. J., Hough, D. W., Stevenson, K. J., Davison, M. and Young, J. (1992) Biochem. Cell Biol. **70**, 70–75
45 Hsieh, P. C., Shenoy, B. C., Samols, D. and Phillips, N. F. (1996) J. Biol. Chem. **271**, 4909–4915
46 O'Brien, W. E., Bowien, S. and Wood, H. G. (1975) J. Biol. Chem. **250**, 8690–8695
47 Valverde, F., Losada, M. and Serrano, A. (1997) J. Bacteriol. **179**, 4513–4522
48 Strathern, J. N., Jones, E. W. and Broach, J. R. (1982) The Molecular Biology of the Yeast *Saccharomyces*, part B (Metabolism and Gene Expression), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

APPENDIX

Elementary-mode analysis of by-passes in glycolysis

Looking at the different sets of glycolytic enzymes present in various organisms, it is interesting to identify substrate flux modes that work despite the lack of textbook glycolytic enzymes. In algebraic pathway analysis, an algorithm systematically calculates all elementary substrate flux modes for a given set of enzymes. Three examples are listed with the use of the algorithm and programs given in the Materials and methods section. We limit our analysis to glycolysis and the PPP, starting at glucose 6-phosphate (Glc6P) and ending at 1,3-bisphosphoglycerate (G13BP) for reasons of simplicity. Note that the conversion of G13BP into pyruvate yields 2 mol of ATP. The elementary flux modes are idealized situations; any observed substrate flux distribution in a living cell is a linear combination of the few elementary flux modes calculated from the set of enzymes present. In what follows, a minus sign indicates that a reaction proceeds in the opposite direction; numbers correspond to the relative fluxes carried by the different reactions catalysed.

Elementary mode 1

How could *Mycoplasma hominis* convert glucose despite the lack of glucose-6-phosphate dehydrogenase and 6-phosphofructokinase? One elementary flux mode calculated for the species-specific (see the text) modified set of enzymes yields the following succession of enzymes as a possible solution: 2 phosphoglucoisomerase, −ribose-5-phosphate isomerase, xylulose-5-phosphate epimerase, −transketolase reaction I, 2 transketolase reaction II, −transaldolase, 3 fructose-6-phosphate phosphoketolase, −3 phosphate acetyltransferase, 2 glyceraldehyde-3-phosphate dehydrogenase.

Elementary mode 2

An elementary flux mode probably operative in *Methanococcus jannaschii* converts glucose 6-phosphate into ribose 5-phosphate without the requirement of the oxidative PPP by using the following enzyme reactions in succession: 5 phosphoglucoisomerase, fructose-1,6-bisphosphate aldolase, triose-phosphate isomerase, 4 ribose-5-phosphate isomerase, −4 xylulose-5-phosphate epimerase, −2 transketolase reaction I, −2 transketolase reaction II, −2 transaldolase, 6-phosphofructokinase.

Elementary mode 3

The following flux mode can process glucose if phosphofructokinase is absent; it uses PPP enzymes for the by-pass (as depicted in Scheme 1): −2 phosphoglucoisomerase, glyceraldehyde-3-phosphate dehydrogenase, 3 glucose-6-phosphate dehydrogenase, 3 phosphogluconolactonase, 3 gluconate-6-phosphate dehydrogenase, ribose-5-phosphate isomerase, 2 xylulose-5-phosphate epimerase, transketolase reaction I, transaldolase, transketolase reaction II.

Reaction yields of the modes

The overall reaction yields calculated for these elementary modes test the results of the algorithm and prove that each set of enzymes achieves what was claimed (see parenthetical statements).

For elementary mode 1 (glucose 6-phosphate is channelled into the lower part of glycolysis):

$$3CoA + 2NAD^+ + 2Glc6P \rightarrow 3Ac\text{-}CoA + 2NADH + 2G13BP$$

For elementary mode 2 (glucose 6-phosphate is efficiently converted into ribose 5-phosphate):

$$ATP + 5Glc6P \rightarrow ADP + 6R5P$$

where R5P is ribose 5-phosphate.

For elementary mode 3 (glucose 6-phosphate is channelled into the lower part of glycolysis by a different route):

$$Glc6P + NAD^+ + 6NADP \rightarrow NADH + 6NADPH + 3CO_2 + G13BP$$

Received 22 March 1999/1 June 1999; accepted 15 July 1999