Genome Analyses of Spirochetes: A Study of the Protein Structures, Functions and Metabolic Pathways in Treponema pallidum and Borrelia burgdorferi

Rajdeep Das, Hedi Hegyi, and Mark Gerstein*

Department of Molecular Biophysics & Biochemistry
266 Whitney Avenue, Yale University
PO Box 208114, New Haven, CT 06520
(203) 432-6105, FAX (203) 432-5175
Mark.Gerstein@yale.edu

ABSTRACT

We perform a comprehensive genome analysis on two spirochetes, *T. pallidum* and *B. burgdorferi*. First, we focus on the occurrence of protein structures in these organisms. We find that there are only a few spirochete-specific folds, relative to those in other types of bacteria. The most common fold, by far, in the spirochetes is the P-loop NTP hydrolase, followed by the TIM barrel. These folds also happen to be amongst the most multifunctional of the known folds. We also survey the membrane-protein structures in *T. pallidum* and find a notable large family with twelve transmembrane (TM) helices, reflecting the prevalence of 12-TM transporters in bacteria. Then we move to analysis of the metabolic pathways and overall metabolism in the spirochetes, using the metabolic-flux-balancing method. We find that the lipid biosynthesis pathway is absent from the spirochetes. This strongly limits the degree to which these organisms can metabolize NADPH. In turn, we find that the spirochetes distribute flux disproportionately through the glycolytic pathway instead of the NADPH-providing pentose phosphate pathway.

INTRODUCTION

In the last five years there have been a number of organisms, spanning all three forms of life, whose genomes have been sequenced. Among these organisms, Borrelia burgdorferi (Fraser *et al*., 1997) and Treponema pallidum (Fraser *et al*., 1998), two pathogenic prokaryotic spirochetes, have generated a lot of interest among the scientific community. In this paper we have analyzed the occurrence of various protein-folds in these two genomes by comparing them to the structural domains in SCOP (Lo Conte *et al*., 2000). Such analyses of structure and fold families are important, as structural knowledge is essential to understand the precise molecular mechanism of the proteins. Comparing two or more complete, evolutionarily related genomes in terms of domains also allows one to define the essential building blocks of a functioning organism, while the differences can highlight the unique, often pathogenesis-related features of a pathogen microbe.

In addition to analyses of patterns in protein-fold occurrence, we have also studied metabolism in the two spirochetes. With the availability of genomic data, it is now possible to draw the metabolic maps of these two organisms, and analyze their metabolic characteristics. Glucose metabolism in *T. pallidum* has been shown to proceed through the Embden-Meyerhoff-Parnas and hexose monophosphate pathways. Lactate, acetate, and pyruvate are produced during glucose metabolism in the organism. Studies have shown that the amount of lactate and acetate produced in *T. pallidum* depends on the amount of dissolved oxygen in its growth medium. Although glycolytic and pentose-cycle enzymes are found in the spirochetes, TCA cycle enzymes are missing from their genomes (Fraser *et al*. 1997; 1998). In the absence of this pathway it is not clear how these organisms synthesize many complex macromolecules for biomass production requiring precursors produced in TCA cycle. In addition to glucose metabolism, lipid metabolism in these organisms is also intriguing. Although spirochetes, when compared

to bacteria, are relatively rich in lipids (22% of their dry weight), they lack all the enzymes of fatty acid biosynthesis. Studies have shown that they do not synthesize long chain fatty acids and need an external supply of fatty acids for their survival (Johnson, 1976). In this paper we discuss our analysis of glucose and lipid metabolism in the spirochetes. We have used the flux-balancing method to analyze their metabolic characteristics.

FOLD USAGE IN SPIROCHETES

We performed a careful structural analysis of the two genomes, using a combination of FastA and PSIblast searches with the domain sequences of SCOP 1.39, a structural database, as queries, against the two genomic sequences (Altschul *et al.*, 1997; Pearson, 2000). We found structural matches for less than one-fourth of the ORFs, more precisely 289 of the total of 1638 ORFs in *B. burgdorferi* and 252 of the total of 1031 ORFs in *T. pallidum.*

The most common folds are largely the same as in other prokaryotes, as shown in Table 1. The most common fold by far in both organisms is the P-loop containing NTP hydrolase, while the second most common fold, the TIM-barrel, has only about one-third as many matches. Both of these folds belong to the alpha/beta fold class, the most common structural class in bacteria. However, in contrast to most bacteria, including E. coli, there are also two abundant alpha-helical folds in the list, the alpha-alpha superhelix and the long helix oligomers. As shown in the table, in *E. coli* these two folds rank very low (62 and 70, respectively). Interestingly, both folds rank also quite high in yeast and in mycoplasmas (data not shown). Two of the folds, the class II aaRS synthetases and the Ferredoxin-like fold, belong to the alpha+beta class, while the rest of the folds in the table are also members of the alpha/beta class. None of the ten most abundant folds in either genome belong to the all-beta class, a highly abundant class in higher organisms.

Table 1: Top 10 folds in spirochetes

A more general functional distribution of the most common folds is shown in Figure **1**. It shows the 5 most commonly occurring folds in the Swissprot database and the number of different - mostly enzymatic - functions they are associated with (Hegyi & Gerstein, 1999). The most versatile fold, the TIM-barrel has as many as 16 different functions associated with it (it is commonly believed that this number was achieved as a result of convergent evolution, i.e. the fold evolved independently several times). Interestingly, the large majority of the folds, about 70% have only one enzymatic function associated with them (note however, that this analysis was based on single-domain proteins with enzymatic functions).

Figure 1: Top 5 folds in swissprot

In *B. burgdorferi* and *T. pallidum* we identified 123 and 126 folds, respectively. The two organisms share more than a hundred of their folds, 106 altogether. As shown in Figure **2**, there are only a few spirochete-specific folds: *T. pallidum* has no specific folds compared with *E. coli* and yeast, and only two compared with *E. coli* and *B. subtilis*. These two *T. pallidum*-specific folds are the 6-bladed beta-propeller (2sil) and a specific Cysteine-proteinase (1gcb). (The codes in parentheses are PDB identifiers with the corresponding chains (Sussman *et al.*,1998).) The Borrelia also has two specific folds, one of them an endonuclease (1smn, chain A), the other an outer-surface protein (1osp, chain O), occurring exclusively in this organism, in as many as 9 copies, which probably indicates an important pathogenic role for this fold.

Figure 2: Venn diagram

Figure **3** shows the phylogenetic tree of 10 completely sequenced organisms, including the two spirochetes (*T. pallidum* and *B. burgdorferi*). The tree was constructed based on the fold content of the organisms, in good agreement with the traditional phylogenetic trees derived from pairwise sequence similarity of their ribosomal genes. It shows that *T. pallidum* and *B. burgdorferi* are the most closely related to each other among the 10 organisms.

Figure 3: Phylogenetic tree

We also analyzed the two genomes for transmembrane helix content. The distribution of the ORFs with transmembrane helices with respect to the number of helices per ORF in *T. pallidum* is shown in Figure **4**. Interestingly, the distribution shows local maximums at 7 and 12. The significance of the 7-transmembrane proteins in higher organisms is well documented, and the local maximum at 12 indicates the important role of the 12-TM transporter proteins in bacteria.

Figure 4: TM distribution

In addition to protein structures and functions, we have also studied glucose and lipid metabolism in the two spirochetes.

GLUCOSE AND LIPID METABOLISM IN SPIROCHETES

Constraint On NADPH Production

Using genomic data, we have constructed a simplified metabolic map of glucose and lipid metabolism in spirochetes. Analyses of genomic sequences of spirochetes have shown that several metabolic enzymes are absent in *T. pallidum* and *B. burgdorferi*. Even though most of the glycolytic enzymes are present, a few are missing in their pentose phosphate pathway. A list of enzymes involved in these two pathways are shown in Table **2**, which shows the occurrence of these enzymes in the two organisms based on genome annotation. Note that two enzymes, 6-phosphoglucono lactonase and transaldolase, are missing in both organisms. In contrast to glucose metabolism, most enzymes for lipid metabolism are absent in spirochetes. Based on presence and absence of enzymes, we have drawn a simple metabolic model for glucose and lipid metabolism in spirochetes as shown in Figure **5**. In order to minimize complexity, we have not considered all the metabolic intermediates involved in a pathway. Consequently, many of the consecutive reactions were lumped into a single overall reaction. In mixed acid fermentative metabolism, glucose is degraded to lactate and acetate. The pentose phosphate pathway is present in the two spirochetes where glucose-6-phospahte is converted into fructose-6-phosphate and glyceraldehyde-3-phosphate. In the process, two molecules of NADPH are produced for every molecule of glucose-6-phosphate that is being siphoned into pentose pathway. In contrast to the pentose phosphate pathway, the fatty acid biosynthesis pathway requires a large amount of NADPH for condensation of Acetyl-CoA, as shown by the following reaction.

$$8\text{Acetyl-CoA} + 7\text{ATP} + 14\text{NADPH} = \text{Fatty acid } (C_8) + 7\text{ADP} + 7P_i + 7\text{NADP} + 8\text{CoA}$$

In the overall metabolism in a cell, the reducing equivalent generated needs to be consumed so that cellular equilibrium is maintained. Since spirochetes lack the ability to synthesize fatty acids de novo, the NADPH requirement will be limited for them. Therefore, if the metabolic flux through the pentose phosphate pathway is high, there will be an overproduction of NADPH in spirochetes, which will remain unused. This constraint will force the organisms to have a higher flux in the glycolytic route, compared to the pentose phosphate pathway. In this paper, we have performed a preliminary study on this problem.

Table 2: List of enzymes

Figure 5: Metabolic map

Pathway Analyses By Flux-Balancing Method

We have analyzed our model metabolic pathway by the flux-balancing method (Fell & Small 1986; Savinell & Palsson, 1992). There are several studies that used this method and addressed various aspects of metabolism in different organisms. In flux-balancing method a metabolic quasi-steady state is assumed in a cell. This assumption requires that all the metabolic fluxes involved in formation and degradation of any metabolite must balance which leads to following equation (Fell & Small 1986; Savinell & Palsson, 1992):

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b}$$

where **S** is the stiochiometric matrix of the catabolic reactions, **v** is a vector of all the metabolic reaction rates, and **b** is a vector containing the net metabolic uptake by the cell. We have formulated our model with a target to minimize the use of glucose for maximum synthesis of triglyceride (TG) and solved the above equation for the reaction fluxes. Note that here we have assumed that triglyceride is the only form of lipid present in the cell. Triglyceride is synthesized by condensing three molecules of acyl-CoA of long chain fatty acids with glycerol-3-phosphate. In the problem formulation, we have also assumed that there is an overall ATP requirement for cellular maintenance, as specified by reaction 23. There are a total of 24 reactions and 22 metabolites involved in our model. One should also note here that the model we have used in our study does not include other related pathways, and therefore it doesn't represent actual metabolic condition that is present in the cell. Although this is a simplified model, it provides a basic framework for flux analyses of lipid metabolism in spirochete. Similar studies have been performed where a part of the network is used for flux analyses (Fell & Small, 1986).

The results of our flux analyses are shown in Figure **6**. The results indicate that the metabolic flux in spirochetes doesn't pass through the pentose phosphate pathway but goes along the glycolytic route, whereby spirochetes avoid an overproduction of NADPH. This is a direct result of absence of fatty acid biosynthesis pathway in the organisms discussed earlier. Analyses also show that there is a net synthesis of lactate and acetic acid in the process with a flux split ratio of 12: 1. This result conforms to the previous observation that lactate and acetate are two predominant end products in T pallidum's anaerobic glucose metabolism (Barbieri & Cox, 1981). Our observation that the metabolic flux doesn't go through the pentose phosphate route implies that the enzymes involved in this pathway could either be greatly modified so that they became less efficient or they could be lowly expressed in the spirochetes.

Figure 6: Flux map

CONCLUSION

We have surveyed the protein-folds in *T. pallidum* and *B. burgdorferi*. We have found that they have few unique folds, and the NTP hydrolase fold is the most common fold. We have also studied the glucose and lipid metabolism in spirochetes. We have performed flux analyses on these two pathways. Our results show that in absence of fatty acid biosynthesis pathway, metabolic flux in *T. pallidum* and *B. burgdorferi* is constrained and is distributed disproportionately in favor of glycolytic pathway.

REFERENCES

Altschul, S.F., Madden, T. L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17), 3389-402.

Barbieri, J. T., Cox, C.D. (1981). Influence of oxygen on respiration and glucose catabolism by Treponema pallidum. *Infect Immun* **31(3)**, 992-7.

Fell, D. A., Small, J. A. (1986). Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem. J.* **238**, 781-786.

Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J., Venter, J.C., 1997. Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. *Nature* 390(6660), 580-6.

Fraser, C.M., Norris, S. J., Weinstock, G.M., White, O., Sutton, G.G., Dodson, R., Gwinn, M., Hickey, E.K., Clayton, R., Ketchum, K.A., Sodergren, E., Hardham, J.M., McLeod, M.P., Salzberg, S., Peterson, J., Khalak, H., Richardson, D., Howell, J.K., Chidambaram, M., Utterback, McDonald, L., Artiach P., Bowman, C., Cotton, M.D., Venter, J.C., 1998. Complete genome sequence of Treponema pallidum, the syphilis spirochete. *Science* 281(5375), 375-88.

Hegyi, H., Gerstein, M. 1999. The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J Mol Biol* 228(1), 147-64.

Johnson, C.R. 1976. Comparative spirochete physiology and cellular composition. In: The Biology of Parasitic Spirochetes. R.C. Johnson, eds. Academic Press, New York. p. 39-48.

Lo Conte L, Ailey, B., Hubbard, T.J, Brenner, S.E., Murzin, A.G., Chotia, C. 2000. SCOP: a structural classification of proteins database. *Nucleic Acids Res*(28(1)), 257-9.

Pearson, W.R. 2000. Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol Biol* 132, 185-219.

Savinell J.M., P. B. O. (1992). Network analysis of intermediary metabolism using linear optimization. II. Interpretation of hybridoma cell metabolism. *J Theor Biol* **154(4)**, 455-73.

Sussman J.L., Lin, D., Jiang J., Manning, N.O., Prilusky, J., Ritter O., Abola, E.E. 1998. Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr D Biol Crystallogr* 54(1 (Pt 6)), 1078-84.

CAPTION

Table 1. The 10 most frequently occurring folds in T. pallidum and B. burgdorferi. The numbers indicate how many times the folds were found in these two organisms, and also in E.coli. The numbers in parentheses for E.coli indicate the ranking of each fold in this organism.

Figure 1. The functional distribution of the most diverse folds occurring in Swissprot. The X-axis indicates the number of functions, and the Y-axis denotes the number of folds that have been associated with that many functions. A schematic is shown for each of the 5 functionally most diverse folds.

Figure 2. Two Venn-diagrams, showing the number of shared and unique folds among *T.pallidum, E.coli* and yeast (on the left-hand side), and among *T.pallidum, E.coli* and *Bacillus subtilis*.

Figure 3. A phylogenetic tree of 10 completely sequenced organisms, including *T.pallidum* and *B. burgdorferi*. The tree was constructed based on the pairwise similarity between the fold contents of the genomes.

Figure 4. Distribution of the ORFs with transmembrane helices in *T.pallidum* with respect to the number of helices per ORF. The highest number of TM-helices was found to be 19, detected only in one ORF.

Table 2. List of enzymes that are involved in our metabolic model used in the flux analyses. Column 4 and 5 show their occurrences in *T. pallidum* and *B. burgdorferi*. If an enzyme is present in the organism, it is shown by a "+" sign, whereas its absence is shown by "—".

Fugure 5. Simplified metabolic model for glucose and lipid metabolism in *T. pallidum* and *B. burgdorferi*. Reaction-numbers are numbered shown beside the arrows. Abbreviations used for the metabolites are as follows: Glu-6P (Glucose-6-phosphate), Fru-6P (Fructose-6-phosphate), Fru-1,6P2 (Fructose-1,6-biphosphate), GAP (Glyceraldehydes phosphate), DHAP (Di-hydroxy acetone phosphate), PEP (Phosphoenol pyruvate), Gro-3P (Glycerol-3-phosphate), Oxac (Oxaloacetate). Glucose and fatty acids are transported into cell from the extracellular medium. Since water forming NADH oxidase is present in both the organism, NADH is oxidized to NAD+ in the spirochetes as shown in reaction 24.

Figure 6. Flux values for the reactions as obtained by Flux-Balancing Method. Those reactions, for which flux values are zero, are not shown in the figure. In the flux analyses we have optimized the values of reaction rates subject to the condition of minimum glucose uptake. Examples of equations, for the reaction intermediates, solved by the optimization procedure:

$$v_5 - v_6 - v_7 + v_8 - v_9 + v_{18} - v_{19} + v_{20} = b_2 = 0 \text{ ; for glucose-6-phosphate}$$

$$v_2 - v_3 - v_4 + v_{19} + v_{20} = b_3 = 0 \text{; for fructose-6-phosphate}$$

where, v_i represents the individual reaction rate of the *i*th reaction as shown in figure 5.

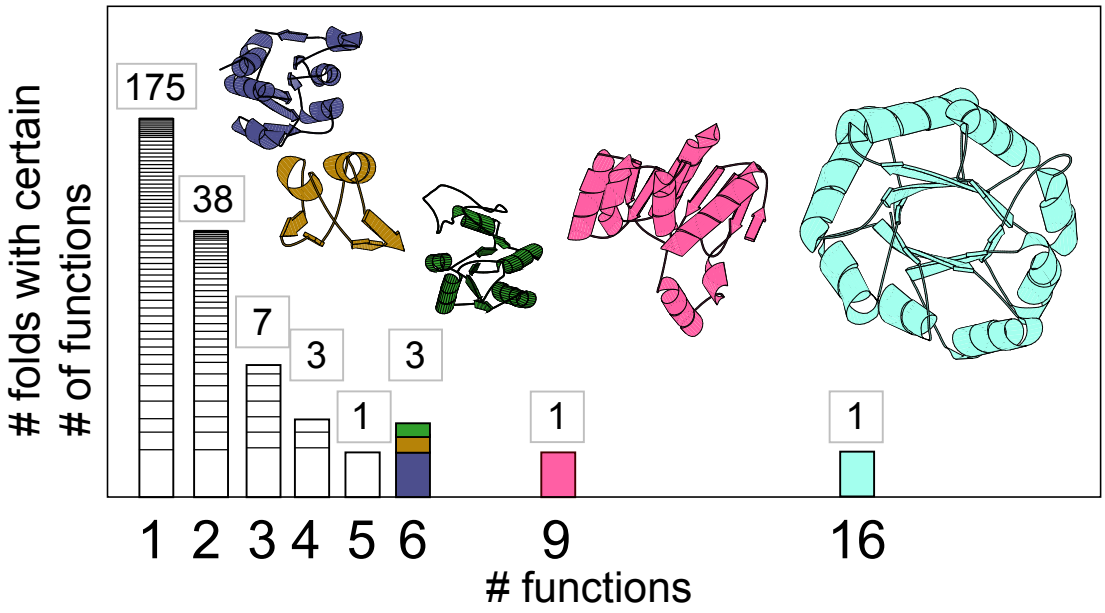
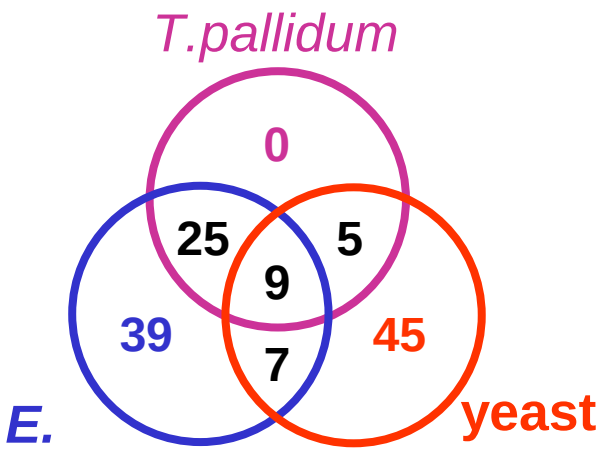

175
38
7
3
3
1
1
1
folds with certain
of functions
1 2 3 4 5 6 9 16
functions

Figure 1.


T.pallidum
0
25
5
9
39
45
7
E.
yeast

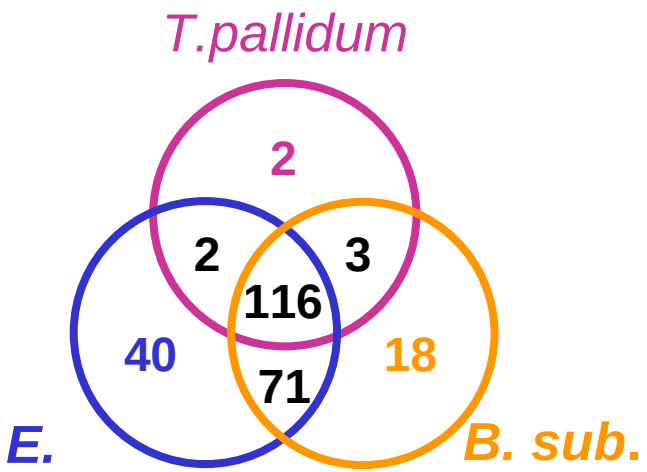

T.pallidum
2
2
3
116
40
18
71
E.
B. sub.

Figure 2.

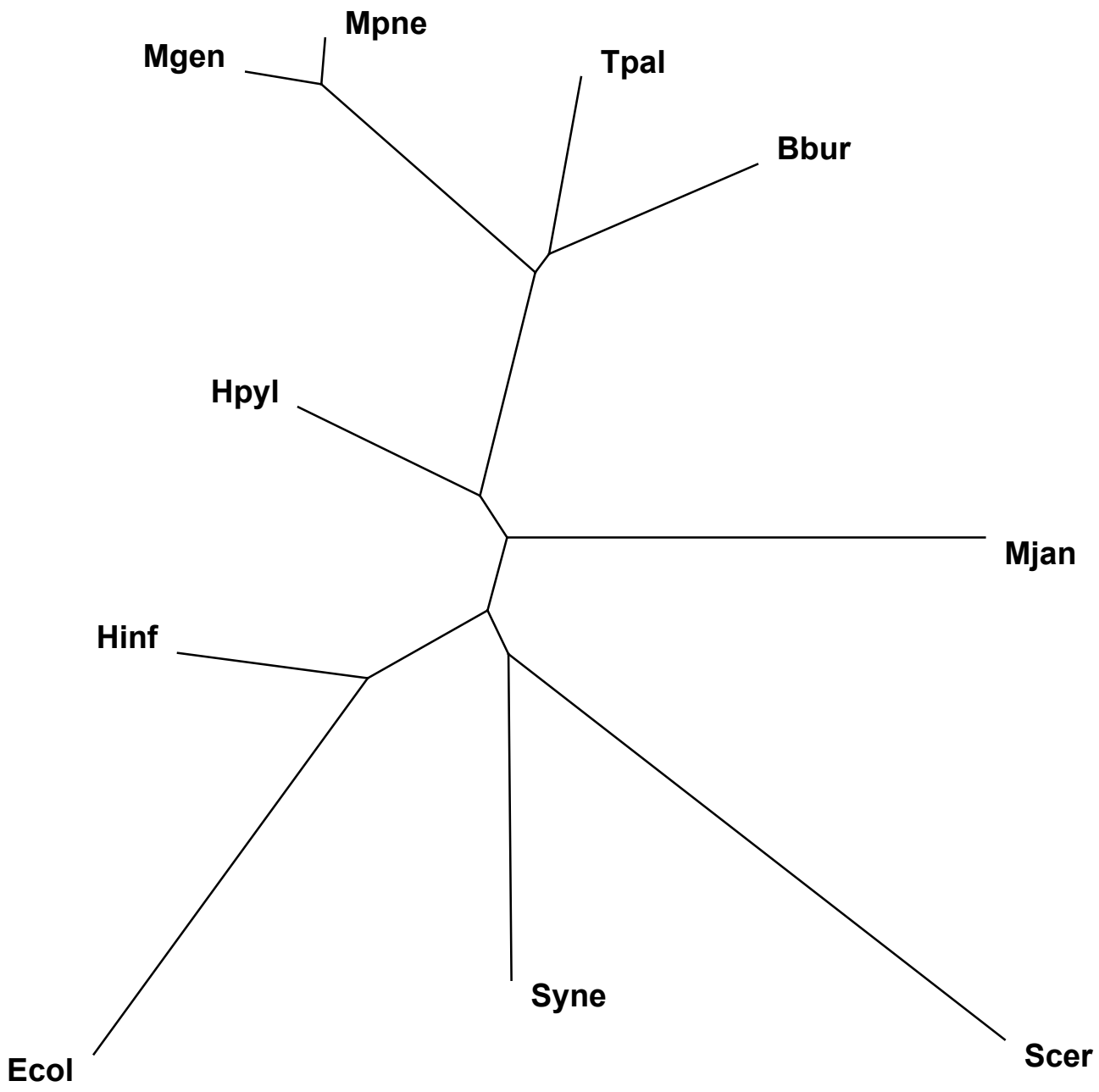


Figure 3

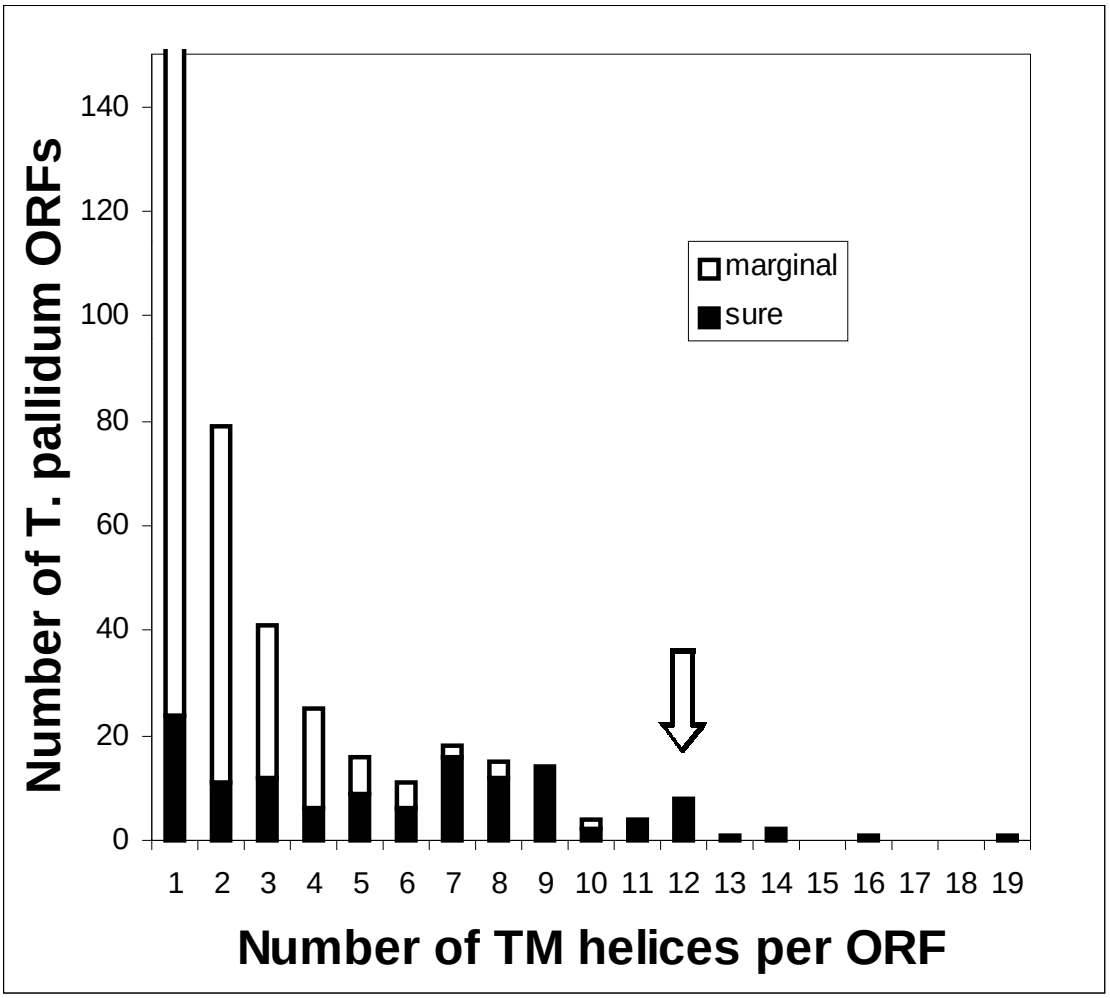
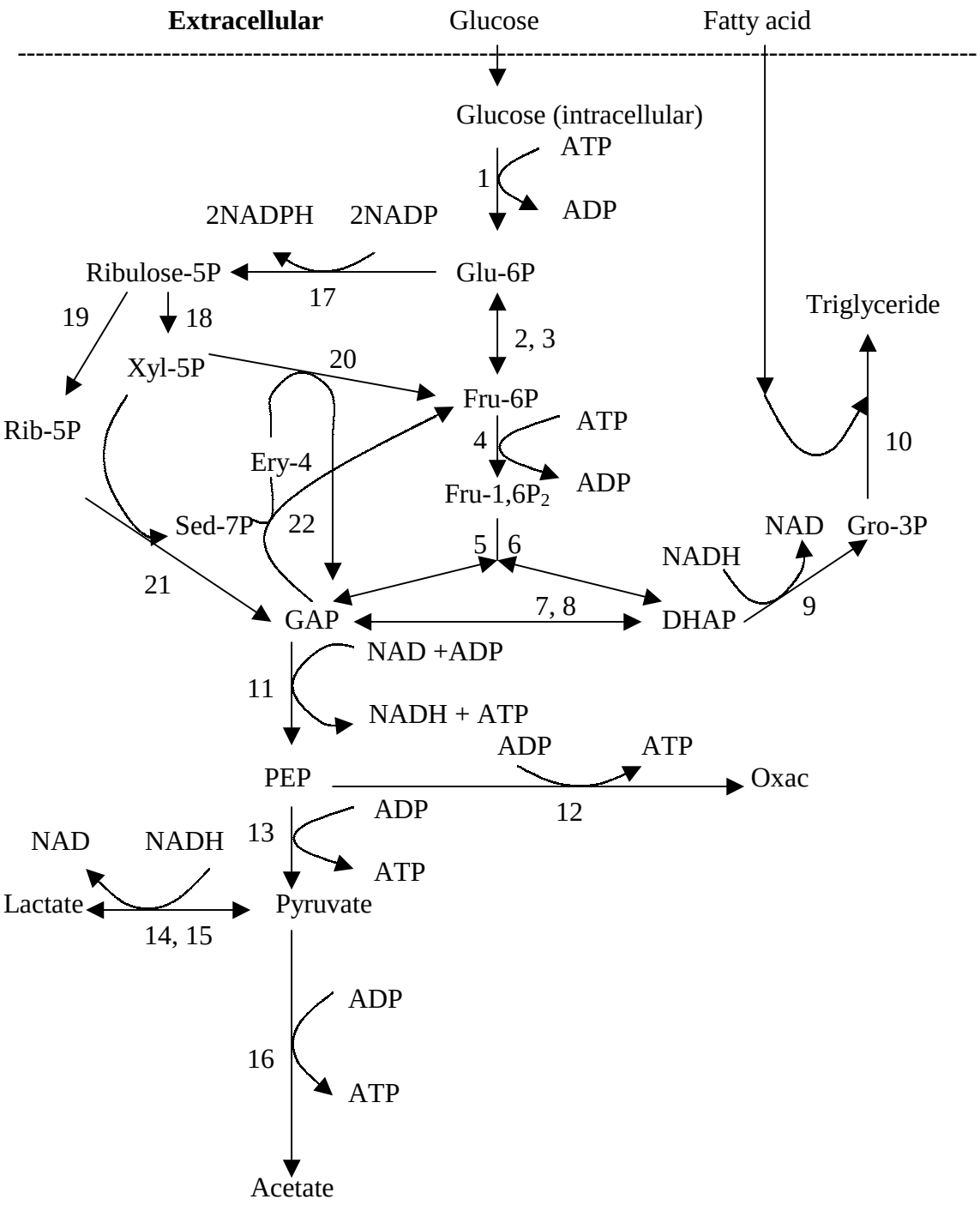

Number of T. pallidum ORFs
140
120
100
80
60
40
20
0
marginal
sure
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Number of TM helices per ORF

Figure 4.



$$ATP \xrightarrow{23} ADP$$

$$NAD(P)H + Ox. \xrightarrow{24} NAD$$

Figure 5.

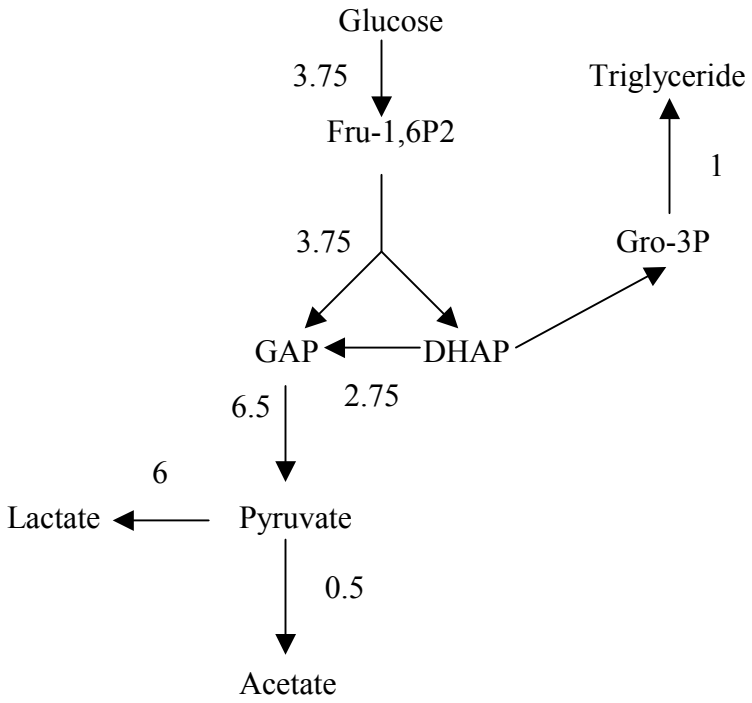
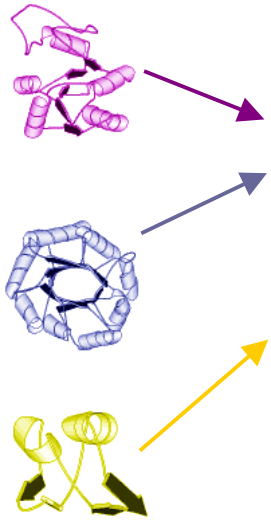

Glucose
3.75
Fru-1,6P2
3.75
GAP
DHAP
2.75
6.5
6
Lactate
Pyruvate
0.5
Acetate
Triglyceride
1
Gro-3P

Figure 6.



Fold		Number in T. pallidum	Number in B. burgdorferi	Number in E. coli
Name	ID			
P-loop NTP hydrolase	3.24	43	67	72(4)
TIM Barrel	3.1	14	20	93(1)
like Ribonuclease H	3.47	13	12	50(6)
class II aaRS synthetases	4.61	13	10	15(28)
alpha-alpha superhelix	1.91	11	6	6(62)
like Ferredoxin	4.34	11	6	82(2)
adenine NT alpha hydrolase	3.17	7	10	16(27)
FAD/NAD(P)-binding domain	3.4	6	4	38(7)
Periplasmic binding proteins - like II	3.82	7	9	36(9)
long helix oligomers	1.105	6	6	5(70)

Table 1.

Pathways	Enzyme	EC number	*T pallidum*	*B. burgdorferi*
Glycolysis	hexoinase	2.7.1.1	+	--
	glucose-6-phosphate isomerase	5.3.1.9	+	+
	aldolase	4.1.2.13	+	+
	triosphosphate isomerase	5.3.1.1	+	+
	glyceraldehyde-phosphate dehydrogenase	1.2.1.12	+	+
	phosphoglycerate kinase	2.7.2.3	+	+
	phosphoglycerate mutase	5.4.2.1	+	+
	enolase	4.2.1.11	+	+
	pyruvate kinase	2.7.1.40	--	+
Pentose Phosphate Pathway	glucose-6-phosphate dehydrogenase	1.1.1.49	+	+
	6-phosphoglucono lactonase	3.1.1.31	--	--
	phosphogluconate dehydrogenase	1.1.1.44	+	+
	ribulose-phosphate 3-epimerase	5.1.3.1	+	--
	ribose-5-phosphate isomerase	5.3.1.6	+	+
	transketolase	2.2.1.1	+	--
	transaldolase	2.2.1.2	--	--
Others	lactate dehydrogenase	1.1.1.28 1.1.1.27	+(28)	+(27)
	phosphoenolpyruvate carboxykinase	4.1.1.32	+	--
	glycerol-3-phosphate dehydrogenase	1.1.1.94	+	+
	pyruvate synthase	1.2.7.1	+	--
	phosphate acetyltransferase	2.3.1.8	+	+
	acetate kinase	2.7.2.1	+	+
	NADH oxidase	1.6.-.-	+	+

Table 2.