OMICS A Journal of Integrative Biology
Volume 8, Number 1, 2004

In Silico Metabolic Model and Protein Expression of *Haemophilus influenzae* Strain Rd KW20 in Rich Medium

ANU RAGHUNATHAN,[1,*] NATHAN D. PRICE,[1,*] MICHAEL Y. GALPERIN,[2] KIRA S. MAKAROVA,[2] SAMUEL PURVINE,[3] ALEX F. PICONE,[3] TIM CHERNY,[3] TAO XIE,[4] THOMAS J. REILLY,[5] ROBERT MUNSON, JR.,[6] RYAN E. TYLER,[7] BRIAN J. AKERLEY,[8] ARNOLD L. SMITH,[9] BERNHARD O. PALSSON,[1] and EUGENE KOLKER[3]

ABSTRACT

The intermediary metabolism of *Haemophilus influenzae* strain Rd KW20 was studied by a combination of protein expression analysis using a recently developed direct proteomics approach, mutational analysis, and mathematical modeling. Special emphasis was placed on carbon utilization, sugar fermentation, TCA cycle, and electron transport of *H. influenzae* cells grown microaerobically and anaerobically in a rich medium. The data indicate that several *H. influenzae* metabolic proteins similar to *Escherichia coli* proteins, known to be regulated by low concentrations of oxygen, were well expressed in both growth conditions in *H. influenzae*. An *in silico* model of the *H. influenzae* metabolic network was used to study the effects of selective deletion of certain enzymatic steps. This allowed us to define proteins predicted to be essential or non-essential for cell growth and to address numerous unresolved questions about intermediary metabolism of *H. influenzae*. Comparison of data from *in vivo* protein expression with the protein list associated with a genome-scale metabolic model showed significant coverage of the known metabolic proteome. This study demonstrates the significance of an integrated approach to the characterization of *H. influenzae* metabolism.

[1]Department of Bioengineering, University of California at San Diego, La Jolla, California.
[2]National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, Maryland.
[3]BIATECH, Bothell, Washington.
[4]St. Jude Children's Research Hospital, Memphis, Tennessee.
[5]Department of Molecular Microbiology and Immunology, University of Missouri–Columbia, Columbia, Missouri.
[6]Children's Research Institute and The Ohio State University, Columbus, Ohio.
[7]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan.
[8]Department of Molecular Genetics and Microbiology, University of Massachusetts Medical School, Worcester, Massachusetts.
[9]Seattle Biomedical Research Institute, Seattle, Washington.
*These authors contributed equally to this work.

INTRODUCTION

Haemophilus influenzae is a facultative anaerobe which requires β-NAD for growth under any condition and heme for aerobic growth. In nature, the organism is largely found in the human nasopharynx (Smith, 1988). *H. influenzae* is an important cause of otitis media in children as well as serious infectious diseases in adults, such as chronic bronchitis, which is the fourth most common cause of death in the United States (Sethi and Murphy, 2001; Sethi et al., 2002; St Geme, 2002; Kolker et al., 2002). Serotype b organisms are a common cause of meningitis and other serious invasive diseases in non-immunized children. In addition to the medical importance of this organism, it is a valuable model microorganism to study pathogen–host interactions (Kolker et al., 2002). However, despite its relative simplicity (e.g., compared to *Escherichia coli*, which has 2.5-fold larger genome) and being the first cellular life form to have a completely sequenced genome (Fleischmann et al., 1995), little is known about the basic metabolic capabilities of *H. influenzae*, including sugar fermentation. According to Bergey's manual, over 90% of *H. influenzae* isolates ferment glucose, galactose, maltose, ribose, or xylose while the remainder ferment fructose, mannose, or glycerol (Holt et al., 1994). Direct fermentation experiments showed that the *H. influenzae* Rd strains can ferment glucose (Hollander, 1976; Macfadyen and Redfield, 1996; Macfadyen et al., 1996). However, the *H. influenzae* strain Rd KW20 glucose transporter remains unidentified; it is not annotated in the genome database (Fleischmann et al., 1995; Karlin et al., 1996, 2001; Tatusov et al., 1996) despite the fact that strain Rd KW20 is considered the prototypic strain representing the minimal (or close to the minimal) genome shared by a majority of *H. influenzae* strains (Morlin et al., 2004). The only protein annotated as a component of the glucose transport machinery, HI1711, a homolog of the *E. coli crr* gene product, is apparently a part of the fructose-specific phosphoenolpyruvate (PEP)–dependent phosphotransferase system and likely plays a role in catabolite repression (Macfadyen and Redfield, 1996; Macfadyen et al., 1996). Identification and characterization of the glucose transporter as well as characterization of over 650 other uncharacterized "hypothetical" ORFs will require further significant experimental effort. To address these problems at the whole-genome scale, diverse analyses, including *in silico* modelling (Edwards and Palsson, 1999; Schilling and Palsson, 2000; Papin et al., 2002), bioinformatics (Fleischmann et al., 1995; Smith et al., 1995; Karlin et al., 1996, 2001; Tatusov et al., 1996), mutagenesis (Akerley et al., 1998, 2002; Herbert et al., 2002), physiology (Hollander, 1976; Smith, 1988; Macfadyen and Redfield, 1996; Macfadyen et al., 1996; Reidl and Mekalanos, 1996), and proteomic analyses (Cash et al., 1997; Link et al., 1997; Langen et al., 2000; Kolker et al., 2003), as well as integrative studies (Kolker et al., 2002) have to be employed.

Here we report the results of an integrated study of the intermediary metabolism of *H. influenzae* employing a combination of proteome analysis, mutation data, and mathematical modeling. The results suggest interesting differences from the global model of the *E. coli* metabolic network in response to levels of oxygen. During anaerobiosis *H. influenzae* appears to maintain expression of numerous proteins traditionally considered being specific to aerobic growth and, conversely, during microaerobic growth it expresses several proteins restricted to anaerobic growth conditions in other bacteria. Overall, this suggests a markedly different pattern of global metabolic control in *H. influenzae* compared with *E. coli* and other model organisms, and these differences are predicted to influence the set of genes required for growth with different oxygen levels.

MATERIALS AND METHODS

Growth conditions

The *H. influenzae* strain Rd KW20 cells were analyzed after growth in two different conditions: anaerobic and microaerobic. The reasons for this choice, as well as a detailed description of the experimental design, have been presented earlier (Kolker et al., 2003). These conditions were chosen as they were expected to be close to the natural state of *H. influenzae* in its human host, as it is more frequently isolated from clinical material after anaerobic incubation (A.L. Smith, unpublished data).

H. influenzae strain Rd KW20 was initially obtained from Dr. Hamilton Smith in 1980, having been described in 1975 (Wilcox and Smith, 1975). It was subcultured on supplemented brain-heart infusion (sBHI)

agar containing 10 mg/L each of β-NAD, equine hemin hydrochloride, and L-histidine (all from Sigma, St. Louis, MO), and 15 g of agar (Difco) per liter (Macfayden et al., 1996), and frozen at -70°C in sterile skim milk. A fresh subculture was used for each experiment. A sterile loop of organisms was used to transfer an inoculum from a fresh BHI agar plate into 50 mL of sBHI broth in a 250-mL Erlenmeyer flask. The initial inoculum of strain Rd KW20 was 1.2–4.6 $\times$ 10^5 cfu/mL and was incubated at 37°C in sBHI in room air shaking at 200 rpm. During incubation in room air, in three experiments, serial 1-mL samples of the media were obtained, and the partial pressures of oxygen, carbon dioxide, and pH were measured by an Irma SL Blood Gas Analysis system (Diametrics Medical, St. Paul, MN). We found the partial pressure of oxygen ranged from 145 to 128 Torr ($n = 3$) during logarithmic growth to stationary phase, except at mid-logarithmic phase ($A650_{nm} = 0.64$) when the mean pO_2 was 53 Torr ($n = 3$). In stationary phase the pressure of oxygen returned to 138–145 Torr ($n = 3$). Thus, the conditions were microaerobic during late logarithmic growth.

The cultures were also incubated anaerobically in the same flasks fitted with butyl rubber stoppers in which air in the headspace was purged with nitrogen and the flasks placed in a Difco anaerobic chamber. Nitrogen was bubbled through the sBHI broth prior to inoculation. Anaerobiosis was indicated by a disposable GasPak® anaerobic indicator (Becton Dickenson, Franklin Lakes, NJ) and verified by growth of the obligate anaerobe *Clostridium butyricum* spores on chocolate agar plates placed into the incubation chamber.

H. influenzae strain Rd KW20 cells were grown overnight under anaerobic and microaerobic conditions and harvested by centrifugation. The initial pH of the sBHI medium was 7.34 $\pm$ 0.12 ($n = 9$) while at the conclusion of anaerobic incubation the pH was 4.69 $\pm$ 0.59 ($n = 6$); after incubation in room air pH was 5.97 $\pm$ 0.16 ($n = 4$; for more details on the two growth conditions, see Kolker et al., 2003).

Simulated sBHI medium

The proteome data were obtained for *H. influenzae* growing in sBHI broth (Kolker et al., 2003). The exact composition of sBHI is unknown. However, to define allowable inputs and outputs to the metabolic model, sBHI was simulated by a defined rich medium. This simulated rich medium contained all metabolites for which a transport reaction existed in the metabolic model as we assumed that these metabolites were present in the medium. Thus, these metabolites were allowed as inputs to the metabolic model (Tables 1 and 1S, supplementary materials). Transport reactions associated with proteins that were not identified in the proteome data were constrained to have zero flux in order to simulate the consequences of these transporters being absent (Table 1S; for details, see *www.Biatech.org/publications/HI-metabolome*). The allowable inputs to the network present in the simulated sBHI medium and those metabolites eliminated due to lack of proteomics evidence are listed in Table 1. In addition to protein expression data (Kolker et al., 2003) and model information for both growth conditions, Table 1S also includes mutation data on essential and non-essential gene assessments for microaerobic condition from the earlier study by one of the authors (Akerley et al., 2002).

Enzymatic activities and organic acids

Along with *in silico* modelling, activities of the following enzymes in sBHI under two growth conditions were determined as described in the corresponding references (Table 2): acetate kinase (Aceti and Ferry, 1988); citrate lyase (Wright et al., 1985); enolase (Spring and Wold, 1975); malate dehydrogenase (Gregory et al., 1971); malic enzyme (Geer et al., 1979); phosphoenolpyruvate carboxykinase (Hunt and Kohler, 1995); pyruvate kinase (Turner and Plaxton, 2000); and pyruvate dehydrogenase (Millar et al., 1998).

Organic acids were measured in stationary phase cultures; 15 mL of ice-cold 0.8 M perchloric acid was added to 25 mL of liquid culture in a 125-mL Erlenmeyer flask, mixed, and held on ice for 30 min before centrifuging at 20K $\times$ g at 4°C for 15 min. The supernatant was removed and adjusted to pH 6.7–7.0 with 1.6 M KOH. An additional centrifugation at 20K $\times$ g at 4° for 10 min pelleted the $KClO_4$. All organic acids were measured enzymatically on the neutralized perchlorate supernantant: acetate by the method of (Guynn and Veech, 1974) and confirmed by assay with the method of (Schulman and Wood, 1975); formate by the method of (Quayle, 1966). All chemicals, enzymes, and coenzymes were purchased from Sigma.

TABLE 1. SIMULATED RICH MEDIUM FOR THE *IN SILICO* MODEL[a]

Acetate	Deoxyguanosine	Histidine	Pantothenate
Acetaldehyde	Deoxyinosine	Hypoxanthine	Proline
Adenine	Dipeptide	Inosine	Putrescine
Adenosine	Thymidine	Potassium	Pyruvate
α-Ketoglutarate	Deoxyuridine	D-Lactate	Ribose
Arginine	Ethanol	L-Lactate	Serine
Asparagine	Formate	L-Lysine	Sialic acid
AspartateThreonine	Fructose	Malate	Sulfate
Branched chain amino acids	Fucose	Meso-diaminopimelate	Spermidine
Myristic acid	Fumarate	Methionine	Succinate
Palmitic acid	4-Aminobutanoate	Sodium	Threonine
Stearic acid	Glycerol	Ammonia	Tyrosine
Carbon dioxide	Glycerol 3-phosphate	Oligopeptide	Uracil
Cysteine	Galactose	Ornithine	Urea
Cytidine	D-Glyceraldehyde	Peptide	Uridine
Cytosine	α-D-Glucose	Phenylalanine	Xanthine
Deoxyadenosine	Guanine	Phosphate (inorganic)	D-Xylose
Deoxycytidine	Guanosine		

[a]In microaerobic *in silico* modeling, oxygen is also an allowed input.

In silico *metabolic modeling for microaerobic and anaerobic conditions*

A previously published flux balance model of *H. influenzae* strain Rd KW20 was utilized for all model calculations herein (Edwards and Palsson, 1999; Schilling and Palsson, 2000; Papin et al., 2002). Genes and proteins were associated with 353 reactions in the metabolic model of *H. influenzae.* These gene-protein-reaction associations were based on data regarding which genes encode for which proteins, and which reactions are catalyzed by each protein. Reactions can be catalyzed by multiple proteins and some proteins catalyze multiple reactions, as shown in Table 1S. Using the protein-reaction relationships, the effect of the absence of a protein can be evaluated in the context of the *in silico* metabolic model. Reactions catalyzed by proteins for which no evidence was found in the proteome data (Kolker et al., 2003) were eliminated from the model one at a time. In this manner, a core set of essential proteins whose functions were always required for simulated growth by the *in silico* model was calculated. The reaction(s) associated with each protein were individually "deleted" from the model to recalculate the optimal growth solution in the absence of this protein. If biomass was not able to be produced without the reaction(s) associated with that protein, then the protein was deemed individually essential for the *in silico* metabolic model.

TABLE 2. ENZYMATIC ACTIVITIES FOR KEY PROTEINS INVOLVED IN THE CENTRAL METABOLISM (μMOL/MIN/MG)[a]

Enzyme (ID)	*MAE*	*ANA*
1. Pyruvate kinase (HI1573)	0.131 ± 0.026	0.019 ± 0.004
2. Pyruvate dehydrogenase (HI1231-HI1233)	0.06 ± 0.002	0.02 ± 0.001
3. Malic enzyme (HI1245)	0.025 ± 0.005	0.023 ± 0.005
4. Malate dehydrogenase (HI1210)	0.53 ± 0.011	1.36 ± 0.272
5. PEP carboxykinase (HI0809)	0.37 ± 0.074	0.47 ± 0.094
6. Acetate kinase (HI1204)	1.42 ± 0.284	0.64 ± 0.128
7. Citrate lyase (HI0022-HI0024)	0.02 ± 0.001	0.04 ± 0.002
8. Enolase (HI0932)	0.15 ± 0.03	0.2 ± 0.4

[a]Average values with corresponding errors, based on triplicate experimental measurements.
MAE, microaerobic conditions; ANA, anaerobic conditions.

RESULTS

General properties of the H. influenzae *proteome*

Using different statistical analyses (Keller et al., 2002a,b; Nesvizhskii et al., 2003; Kolker et al., 2003), 414 proteins were identified with high confidence in the soluble and membrane fractions of *H. influenzae* strain Rd KW20 cells grown under both conditions (for more details, see Kolker et al., 2003). Of those, 221 proteins were detected in both conditions (53% of total high confidence identifications), whereas 146 and 65 proteins were detected exclusively in microaerobically and anaerobically grown cells, respectively. The entire set of 414 proteins accounted for approximately a quarter of all predicted Rd KW20 proteins based on analysis of the genome sequence (Fleischmann et al., 1995; Karlin et al., 1996; Tatusov et al., 1996). Alternatively, searching and filtering using traditional parameters, coupled with incomplete trypsin digestion permitted identification of 1284 "candidate" proteins. This value representing more than 75% coverage of the theoretical protein set encoded in the strain Rd KW20 genome (Kolker et al., 2003). Based on the above protein identifications, we attempted to reconstruct the central metabolic pathways and evaluate the carbon flow in *H. influenzae* strain Rd KW20 cells under both growth conditions.

Given that the cells were grown in a rich medium, comprised of BHI broth supplemented with heme, β-NAD, numerous sugars, peptides, vitamins, and nucleosides, it was not surprising that enzymes involved in amino acid, purine, pyrimidine, and cofactor biosyntheses were generally poorly expressed and therefore typically identified as "candidate" proteins. For example, although all purine biosynthesis enzymes were identified in microaerobically grown cells in our experiments (Table 1S), only the enzymes catalyzing the last steps of the pathway and the interconversion of AMP and GMP (PurA, PurB, GuaA, and GuaB gene products; hereafter we refer to *H. influenzae* enzymes by the gene names of their *E. coli* orthologs) were represented by numerous non-overlapping peptides and were therefore identified with high confidence. Further, phosphoribosylpyrophosphate synthetase (HI1609, *prsA*), which is not only shared by purine and histidine biosynthesis, but which can also participate in the accommodation of exogenous nitrogen bases, was similarly well detected.

The same trend was seen in pyrimidine biosynthesis enzymes, most of which (PyrD, PyrE, PyrF, Ndk, and PyrG gene products) were either undetectable or only detected as "candidate" proteins. In contrast, adenylate kinase (HI0349, *adk*) and uridylate kinase (HI1065, *pyrH*), catalyzing the interconversion of nucleoside mono-, di-, and tri-phosphates, were confidently identified in all samples.

Sugar fermentation

In contrast to the enzymes of nucleotide and amino acid biosynthesis, enzymes participating in sugar fermentation were typically highly expressed and identified with high confidence. Peptides derived from fructose 1,6-bisphosphate aldolase (HI0524, *fba*), transketolase (HI1023, *tktA*), transaldolase (HI1125, *talB*), glyceraldehyde 3-phosphate dehydrogenase (HI0001, *gapdH*), 3-phosphoglycerate kinase (HI0525, *pgk*), phosphoglycerate mutase (HI0757, *gpmA*), enolase (HI0932, *eno*), and pyruvate kinase (HI1573, *pykA*) were the most commonly detected, indicating that these proteins were highly expressed. These data suggest that, consistent with earlier predictions (Tatusov et al., 1996), *H. influenzae* cells used glycolysis as the primary pathway of sugar catabolism during either microaerobic or anaerobic growth (Figs. 1 and 2) with a possible contribution of the pentose-phosphate shunt.

However, direct measurements of enzyme activities in cells grown in both environments revealed a more complex picture with dramatic differences between enzymes. While enolase was detected with high confidence and the specific activity of the enzyme in microaerobic and anaerobic cell extracts was essentially the same (Table 2), pyruvate kinase activity in microaerobically grown cells was six-fold higher than in anaerobically grown cells (131 vs. 19 nmol/min/mg protein, Table 2). This suggests that carbon flow from PEP to pyruvate was relatively suppressed in anaerobic conditions. Accordingly, while pyruvate dehydrogenase components (HI1231-HI1233, *lpdA, aceEF*) were detected under both growth conditions, anaerobic cells had three-fold lower pyruvate dehydrogenase activity (60 vs. 20 nmol/min/mg protein, Table 2), indicating reduced carbon flow from pyruvate to acetyl-CoA in anaerobic conditions.

It should be also noted that the late logarithmic phase cells had a glycolytic flux that used a stored poly-

saccharide, such as glycogen. Glycogen synthase (HI1360, *glgA*) was detected as a "candidate" protein in both microaerobic and anaerobic cells, while glycogen phosphorylase (HI1361, *glgP*) was identified in both conditions with high confidence. Glucose-6-phosphate isomerase (HI1576, *pgi*) was also found, albeit without high confidence.

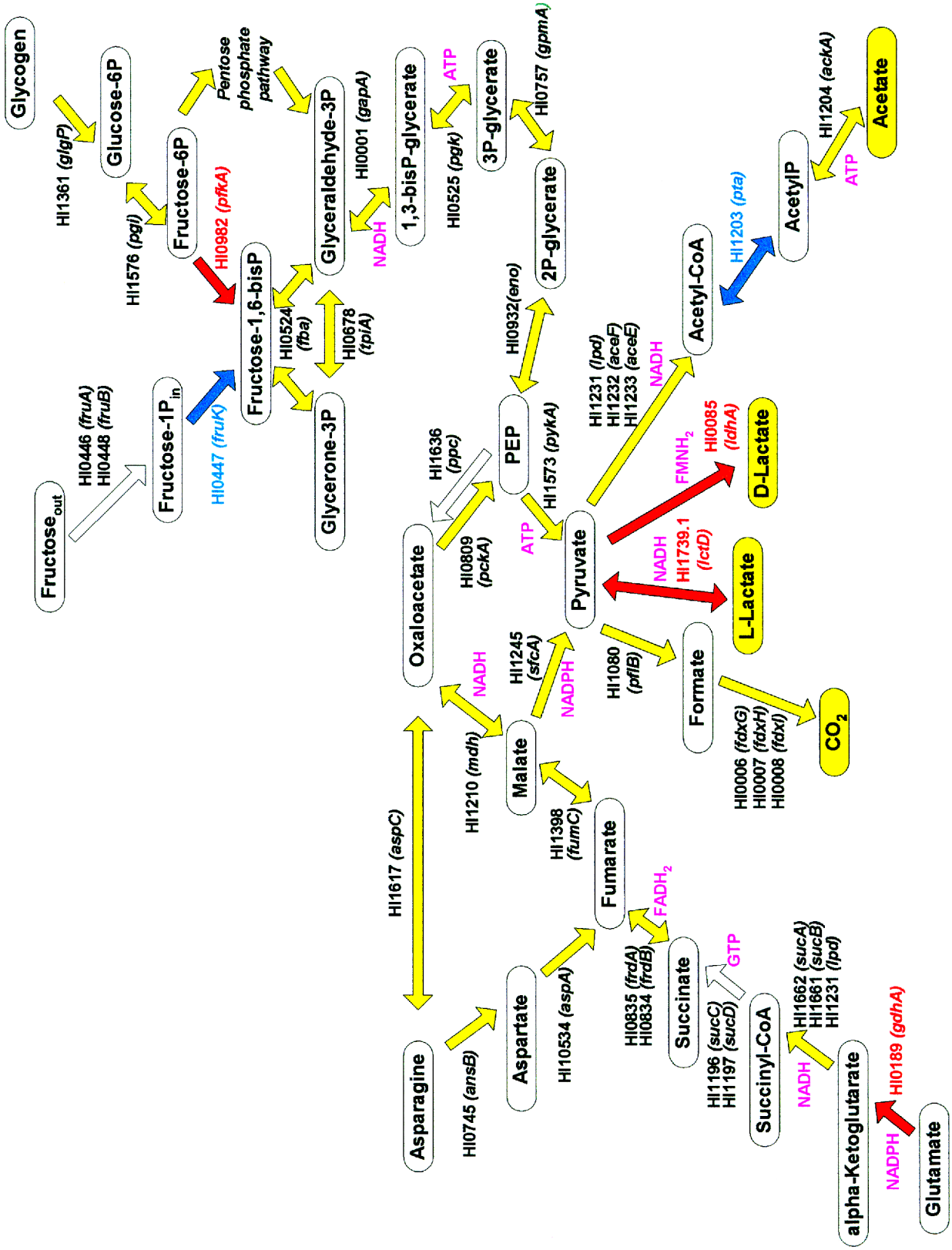
An outstanding example of the simultaneous presence of an enzyme for respiratory and fermentative catabolism under the conditions examined is pyruvate-formate lyase (HI0180, *pflB*). Distinct peptides for pyruvate-formate lyase were detected with high confidence multiple times, suggesting that this enzyme is abundant in the *H. influenzae* cells and that it might be active under both growth conditions. While it is known that pyruvate-formate lyase from *E. coli* is oxygen-sensitive (de Graef et al., 1999), activity of this enzyme has been detected in *E. coli* under microaerobic conditions (Alexeeva et al., 2000). It is quite probable that the respiratory activity in the *H. influenzae* cells results in maintaining the reducing cytoplasmic environment that allows for pyruvate-formate lyase activity. It has been also proposed that cytochrome *bd*-type quinol oxidase CydAB plays a crucial role in creating such an environment in *E. coli* (Alexeeva et al., 2000). Most likely, the terminal oxidase, encoded by the HI1076 (*cydA*) and HI1075 (*cydB*) genes, plays a similar role in *H. influenzae* strain Rd KW20. Indeed, CydA was detected under microaerobic conditions with high confidence, although CydB was not (Table 1S), probably due to the well-known problem with the identification of hydrophobic proteins (Kolker et al., 2003). These data are consistent with our previous identification of the three genes (HI1075, *cydB*; HI1076, *cydA*; and HI0180, *pfl*) as essential for the microaerobic growth of *H. influenzae* (Akerley et al., 2002). Furthermore, in contrast to the situation in *E. coli* (Alexeeva et al., 2000) and in a full agreement with an earlier observation on *H. influenzae* (Cash et al., 1997), superoxide dismutase (HI1088, *sodA*) was detected with high confidence in both microaerobic and anaerobic conditions (Table 1S).

TCA cycle and related reactions

Based on genomic analysis, that *H. influenzae* strain Rd KW20 is reported to have an incomplete TCA cycle (Fleischmann et al., 1995). The results of the proteome analysis and *in silico* modeling support this idea. We additionally confirmed this by Southern and PCR analysis of the Rd KW20 genome using *E.* coli K12 genomic DNA as the reference standard (A.L. Smith, data not shown). The reductive portion of the TCA cycle from oxaloacetate to succinyl-CoA through malate and fumarate (Fig. 1) was active in cells grown in both microaerobic and anaerobic conditions: malate dehydrogenase (HI1210, *mdh*), fumarate hydratase (HI1398, *fumC*), and fumarate reductase (HI0834-HI0835, *frdAB*), were all detected with high confidence, while succinyl-CoA synthetase (HI1196-HI1197, *sucCD*) subunits were detected as "candidate" proteins. In addition, all four components of the α-ketoglutarate dehydrogenase complex: dihydrolipoamide acyltransferase HI1232 (*aceF*), dihydrolipoamide dehydrogenase (HI1231, *lpdA*), and E1 and E2 components, (HI1662, s*ucA*) and (HI1661, *sucB*), were detected with high confidence (*sucA* only under microaerobic condition). These data also confirmed the absence of a frameshift in the HI1398 (*fumC*) gene, included in the original genome description of *H. influenzae* strain Rd KW20 (Fleischmann et al., 1995), but later reported to be an error (Manukhov et al., 2000). Several distinct peptides covering most of the FumC polypeptide chain were detected with high confidence, indicating that the strain Rd KW20 genome encodes a full-size functional enzyme (Kolker et al., 2003).

Oxaloacetate can be formed through the activity of citrate lyase. This enzyme complex (HI0022-HI0024, *citDEF*) was detected in cells grown in both conditions (except CidD) and exhibited similar enzyme activ-

FIG. 1. Summary of catabolic carbon flow by *H. influenzae* strain Rd KW20. Yellow arrows and black gene names indicate the enzymes detected under both conditions with high confidence, the empty arrows correspond to "candidate" proteins detected under both conditions. The blue arrows and gene names indicate the enzymes detected only under anaerobic conditions with high confidence, while the red arrows and gene names indicate the enzymes detected only under microaerobic conditions. The green NADH and ATP labels are placed at the positions where they are generated. Double-headed arrows indicate that the reaction may proceed in either direction, depending upon the pathway used. The substances in rounded boxes have been included in the metabolic model. The final fermentation products are shown on yellow background.


Glycogen
HI1361 (glgP)
Glucose-6P
HI1576 (pgi)
Fructose-6P
Pentose phosphate pathway
HI0982 (pfkA)
Fructose-1P$_{in}$
HI0447 (fruK)
Fructose-1,6-bisP
HI0524 (fba)
HI0678 (tpiA)
Glyceraldehyde-3P
HI0001 (gapA)
NADH
1,3-bisP-glycerate
HI0525 (pgk)
ATP
3P-glycerate
HI0757 (gpmA)
Fructose$_{out}$
HI0446 (fruA)
HI0448 (fruB)
Glycerone-3P
2P-glycerate
HI0932(eno)
HI1636 (ppc)
PEP
HI0809 (pckA)
ATP
HI1573 (pykA)
Oxaloacetate
Pyruvate
HI1231 (lpd)
HI1232 (aceF)
HI1233 (aceE)
NADH
Acetyl-CoA
HI1203 (pta)
AcetylP
HI1204 (ackA)
ATP
Acetate
FMNH$_2$
HI0085 (ldhA)
D-Lactate
NADH
HI1739.1 (lctD)
L-Lactate
HI1617 (aspC)
HI1210 (mdh)
NADH
Malate
HI1245 (sfcA)
NADPH
HI1080 (pflB)
Formate
HI0006 (fdxG)
HI0007 (fdxH)
HI0008 (fdxI)
CO_2
HI1398 (fumC)
Fumarate
HI0745 (ansB)
Asparagine
HI10534 (aspA)
Aspartate
HI0835 (frdA)
HI0834 (frdB)
FADH$_2$
Succinate
HI1196 (sucC)
HI1197 (sucD)
GTP
Succinyl-CoA
NADH
HI1662 (sucA)
HI1661 (sucB)
HI1231 (lpd)
alpha-Ketoglutarate
NADPH
HI0189 (gdhA)
Glutamate

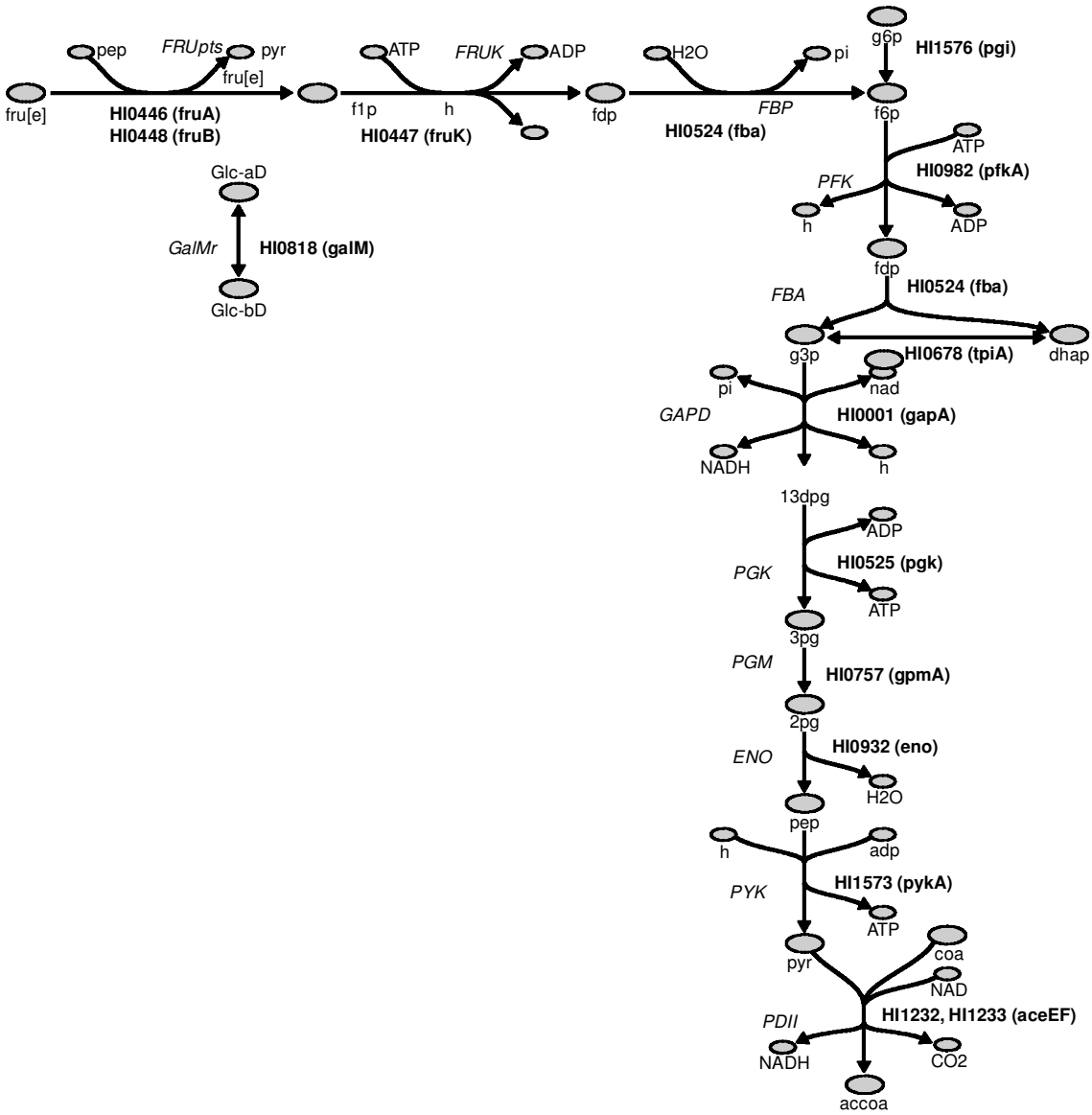


FIG. 2. Snapshot of the glycolysis pathway as generated using Sim Pheny. Images of the reactions are generated automatically and then can be linked together to form network diagrams.

ity (Table 2). An alternative route for the formation of oxaloacetate, PEP carboxylase (HI1636, *ppc*), was not detected with confidence under either growth condition. However, the reverse reaction, decarboxylation of oxaloacetate by PEP carboxykinase (HI0809, *pckA*) in an ATP-utilizing reaction that yields CO_2, was found to be active under both microaerobic and anaerobic conditions: the peptides were detected with high confidence and the enzyme activity was similar under both conditions (Table 2). Formate was produced from acetyl-CoA by formate dehydrogenase (HI0006-HI0008, *fdxGHI*) under both conditions, an observation confirmed by the nearly identical concentrations of formate in each culture: 0.30 ($\pm$0.06) mM in the microaerobic and 0.29 ($\pm$0.06) mM in the anaerobic environment.

Utilization of acetate appeared to be the same under each condition with the detection of phosphate acetyl transferase (HI1203, *pta*) and acetate kinase (HI1204, a*ck*) although the specific activity of the latter was higher under microaerobic conditions (1.42 vs. 0.64 mol/min/mg protein, Table 2). The concentration of

acetate in stationary phase cultures was nearly identical with growth in both environments: 0.46 (±0.09) mM microaerobic and 0.60 (±0.12) mM anaerobic. Although malate dehydrogenase (HI1210, *mdh*) activity was greater in anaerobically grown cells (1.36 vs. 0.53 mol/min/mg protein, Table 2), malic enzyme (HI1245, *maeB*) activity (Fig. 1) was the same in both cell extracts (Table 2).

The above experimental data indicate that, during growth in an environment where the partial pressure of oxygen varies from 53 to 145 Torr, classic anaerobic respiration is used. This is additionally confirmed by finding of FrdA and FrdB subunits of fumarate reductase (HI0835-HI0832, *frdABCD*) with high confidence in cells from both growth environments. The observation that these two subunits of fumarate reductase were detected under both growth conditions, while the two subunits that are specific for the succinate dehydrogenase were not detected in both conditions, suggested that this arm of the TCA cycle in *H. influenzae* operates differently from that in *E. coli* and other model organisms. To further investigate this hypothesis, an insertion mutation that disrupts the *H. influenzae frdA* gene (HI0835) predicted to encode fumarate reductase flavoprotein subunit was isolated from an ordered collection of *mariner* transposon mutants generated previously and mapped by PCR (Akerley et al., 2002). Growth of the *frdA* mutant versus wild-type was compared under conditions of low and high oxygen supply (Fig. 3). Surprisingly, the mutant was attenuated for growth under both conditions indicating a requirement for fumarate reductase both anaerobically and microaerobically. The growth defect could not be reversed by addition of 10 mM succinate and 10 mM nitrate to the media. It is possible that fumarate serves as a primary electron acceptor under both conditions and that its role in conversion of fumarate to succinate does not account for the growth defect of the *frdA* mutant. In summary, these results demonstrate a role in growth for *frdA* under diverse redox conditions consistent with the protein expression data.

Of the enzymes that couple the TCA cycle with amino acid metabolism, aspartate ammonia lyase (HI0534, *aspA*), catalyzing the conversion of aspartate to fumarate, and aspartate aminotransferase (HI1617, *aspC*) were detected with high confidence under both microaerobic and anaerobic conditions. This data suggests that *H. influenzae* cells grown on sBHI broth, lacking added carbohydrates, can generate oxaloacetate (via HI1617, *aspC*) and fumarate (via HI0534, *aspA*).

Electron transport

Energy generating electron transport pathways also showed utilization of different terminal electron acceptors under both growth conditions. In *E. coli,* the *cydAB* operon is expressed under conditions of limited aeration, and will reduce oxygen to water. We detected the *H. influenzae* homologs (HI1075-HI1076, *cydAB*) only under microaerobic conditions. None of the subunits of the predicted putative sodium-translocating quinone reductase (HI1683-HI1688, *ydgLMNOPQ*; Hase et al., 2001) were detected in cell extracts from either condition. However, most of the sodium translocating NADH dehydrogenase (HI0164-HI0171, *nqrABCDEF*) subunits were identified under both conditions. Chain A (HI1047, *dmsA*) of the anaerobic dimethyl sulfoxide reductase complex was detected with high confidence under both conditions. Of the other subunits in this operon, chain B of dimethyl sulfoxide reductase (HI1046, *dmsB*) was found only in aerobically grown cells, while chain C (HI1045, *dmsC*) was not detected in cells grown in either environment.

Periplasmic nitrate reductase (HI0347, *napB*) was found with high confidence in cells grown in both conditions (Kolker et al., 2003). This enzyme appears to be part of an operon in which nitrate is the final electron acceptor as the (HI0344, *napA*) gene belongs to the family of selenocysteine-containing anaerobic dehydrogenases. Since genes from HI0342 through HI0348 appear to form an operon, we suspect that all these genes are expressed under both growth conditions. Further evidence of utilization of anaerobic pathways under both conditions is that the *nrfA* (HI1069) gene for the cytochrome *c* subunit of the nitrite reductase complex (HI1069-HI1066, *nrfABCD*) was detected under both growth environments; other components of that complex were either not detected or detected as "candidate" proteins (Kolker et al., 2003). None of the Na^+/H^+ antiporters (HI0225, *nhaA*; HI0427, *nhaB*; HI1107, *nhaC* [Hase et al., 2001; Kolker et al., 2003]) was detected with high confidence.

Comparison of proteome data with genome-scale metabolic reconstruction

Protein identifications from our recent proteome study (Kolker et al., 2003) were compared to the proteins in a previously published genome-scale model of *H. influenzae* metabolic network (Edwards and Pals-

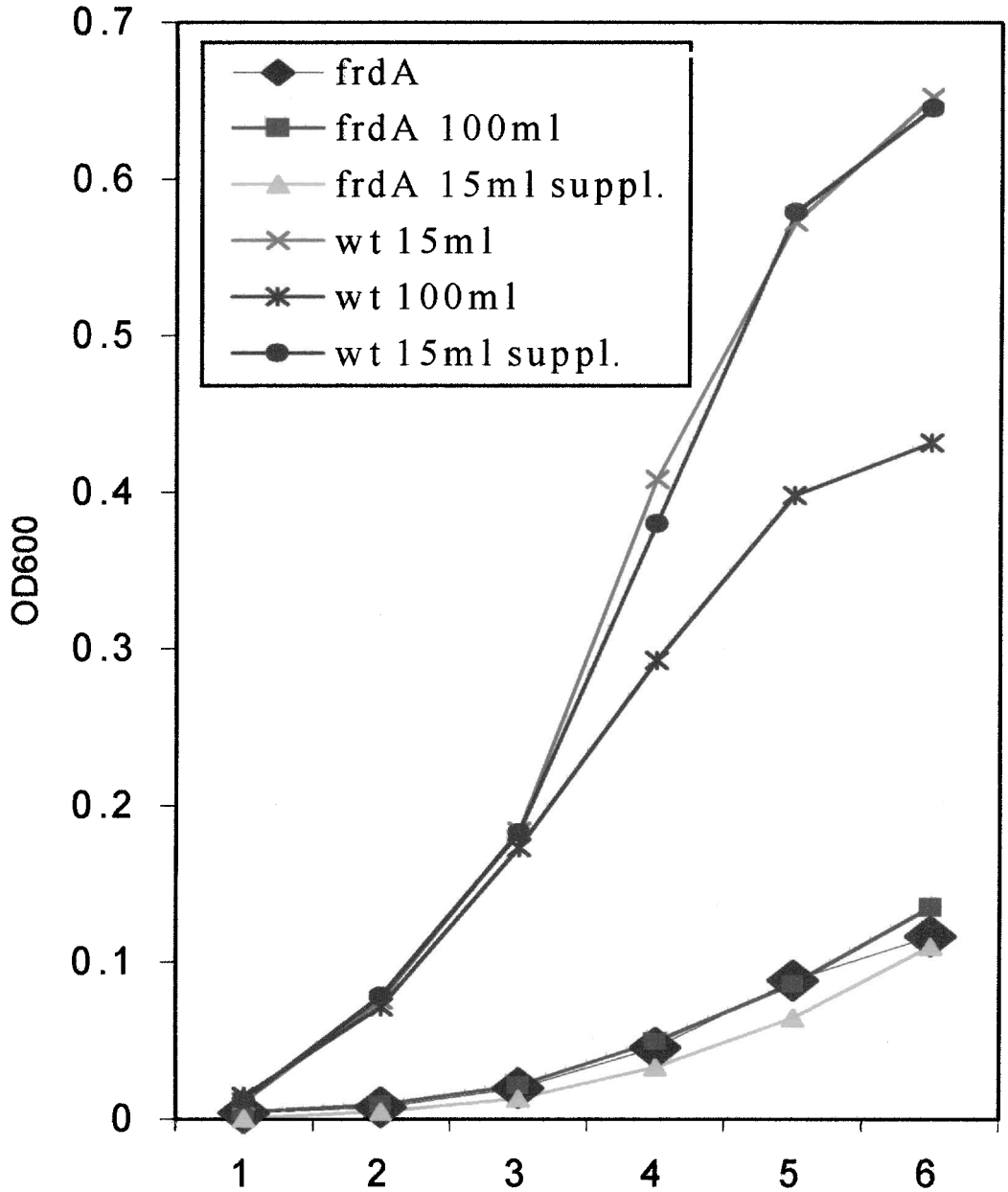


FIG. 3. sBHI growth of the wild-type strain Rd KW20 and its *frdA* mutant. The *frdA* mutant strain and parental wild-type strain Rd KW20 were inoculated from overnight cultures to a starting OD600 of 0.003 and compared for growth in sBHI under low aeration (100-mL cultures in 125-mL shake flasks at 220 rpm) versus high aeration (15-mL cultures in 125-mL shake flasks at 220 rpm). The 15-mL cultures were compared both in the absence and presence (supplement) of 10 *mM* succinate and 10 *mM* $NaNO_3$. Each point represents the mean of three experiments, and standard deviations were below 10% of the means for all determinations.

son, 1999). Proteins in the model identified under each condition either with high confidence or as "candidate" proteins are shown in Table 1S. Approximately 38% of the 353 proteins associated with the genome-scale metabolic model were identified with high confidence in the proteome study (Kolker et al., 2003), and approximately 90% of the proteins in the model were experimentally identified as "candidate" proteins under either growth condition. Interestingly, the proportion of metabolic proteins (identified with either high

confidence or as "candidate" proteins) detected in both conditions is higher than total amount of all detected proteins. That is, 59% of high confidence (80 out of 120, Table 1S) and 64% of "candidate" (206 out of 320, Table 1S) metabolic proteins were detected in both conditions, compared respectively to 53% and 45% of all high confidence and "candidate" proteins identified in both conditions (Kolker et al. 2003). Summary of the catabolic carbon flow, depicted in Figure 1, also confirms the above observation: 68% (19 out of 28) of the enzymatic reactions are performed by proteins detected with high confidence.

We specifically decided to focus on the above estimates only, rather than pursue a more detailed comparative analysis of all three types of available information on protein expression, mutation, and *in silico* metabolic data (Table 1S). Two main obstacles motivated this decision: the first complication stems from the fact that the precise compositional contents of the growth media used in the earlier studies could not be determined (Akerley et al., 2002; Kolker et al., 2003); the second is concerned with the variability of the batch cultures implemented in the proteome (Kolker et al., 2003) and mutation (Akerley et al., 2002) studies was not strictly controlled. Further experimentations using standardized experimental designs will be performed with a recently developed defined medium (Coleman et al., 2003) in order to overcome these obstacles.

In silico *metabolic modeling for microaerobic condition*

The *H. influenzae* strain Rd KW20 was subjected to *in silico* deletions in the gene products found to be absent in the proteome study (Kolker et al., 2003). The reconstructed metabolic model of *H. influenzae* (Edwards and Palsson, 1999) was used to calculate which of the unidentified proteins were essential for the defined metabolic network to produce biomass. Theoretical optimal growth performance was evaluated, and each of the reactions catalyzed by the gene products not detected in the proteome was removed from the system one at a time. Gene products were simultaneously removed when their genes coded for the components of the same enzyme complex (i.e., AceEF and SucCD).

A set of 41 enzymes not detected by the proteome analysis during microaerobic growth was considered in this analysis. The ability of the altered metabolic network to compensate for the loss of enzymatic function due to each unidentified protein was evaluated *in silico* during simulated growth in the simulated rich medium (Table 3). Proteins were grouped into two categories based upon the effect of the loss of their enzymatic function: essential or non-essential for biomass production. It was observed that *H. influenzae* could not "grow" *in silico* when all the 41 proteins were excluded from the model. This *in silico* single deletion analysis resulted in classifying 16 proteins as individually essential and 25 as individually non-essential to growth in the simulated rich medium under microaerobic conditions (Table 3). Proteins that have enzymatic functions predicted to be essential to the production of biomass provide hypotheses for proteins that may need to be included in the proteome but were not detected due to sensitivity of the method.

When the 25 so-called "non-essential" proteins were deleted simultaneously from the genome set, the resulting strain did not allow growth on simulated sBHI media. Although these 25 proteins were found to be individually non-essential for the production of biomass, growth was not possible when the entire redundant or non-essential protein set was excluded from the model. Thus, while no one of these proteins was necessarily essential, as a group they perform essential functions, so that some of these proteins are required for the defined metabolic network to produce biomass. Various combinations of these proteins could be deemed essential depending on the conditions and which of the other gene product within the redundant set are not present. For example, the *in silico* model predicts 322 pairwise combinations of protein double deletions from within the non-essential protein set that inhibit biomass production under the microaerobic growth condition.

In silico *metabolic modeling for anaerobic condition*

The reconstructed metabolic network was also used to evaluate the proteomics data under anaerobic conditions in the context of the reconstructed metabolic network. Similarly, to the microaerobic situation, the removal of all the non-identified proteins under anaerobic conditions again resulted in a loss of the capacity of the *in silico* strain to produce biomass. Thus, based upon the reconstructed metabolic network, the identified protein complement under anaerobic conditions on simulated rich medium was not sufficient to ensure cellular growth.

Table 3. Proteins Found Essential or Non-Essential During *In Silico* Growth of *H. influenzae* Under Microaerobic Condition in Simulated sBHI Medium[a]

Essential proteins		*Non-essential proteins*	
ID	*Protein name*	*ID*	*Protein name*
HI0972	AccC	HI1230	ApT
HI0185	AdhC	HI1025	AraD
HI0970	AroD	HI0226	BrnQ
HI1308	DapB	HI1350	CdD
HI0750	DapF	HI0447	FruK
HI1140	DdlB	HI0819	GalK
HI1062	FabZ	HI0818	GalM
HI0899	FolA	HI0686	GlpT
HI1447	FolE	HI1079m	GlnQ
HI0812	GalU	HI1153	HpT
HI1743	GmK	HI1607	HemM
HI0058	KdsB	HI0471	HisB
HI0876	NdK	HI0475	HisI
HI0123	PgsA	HI0738	IlvD
HI1077	PyrG	HI0881	IspB
HI0456	TmK	HI0988	LeuC
		HI0989	LeuD
		HI0969	MenC
		HI0122	MetC
		HI1104	NanT
		HI0272	PyrE
		HI0529	TdK
		HI0416	ThiD
		HI1305	ThiL
		HI0905	ThyA

[a]These proteins were classified as essential or non-essential based upon whether or not they were necessary for biomass production within the defined metabolic network under microaerobic conditions in simulated rich medium.

The impact of removing each of the non-identified proteins individually from the metabolic model was assessed. The proteins from Table 4 were calculated to be *always* required for the production of biomass under simulated anaerobic conditions on a simulated sBHI medium. Since the model predicts that all of these proteins would be identified under conditions in which *H. influenzae* is growing, we presume that these proteins are likely to be present in the cell under anaerobic conditions, but are not being detected with the current approach. Alternatively, verifying the absence of these proteins from the proteome through further experimentation would suggest the existence of alternate synthesis routes not currently included in the reconstructed metabolic network.

DISCUSSION

The initial goal of this work was to investigate the effects of low oxygen pressure and anaerobiosis on the expression of the key metabolic enzymes in the *H. influenzae* strain Rd KW20 cells. In these experiments, cells of *H. influenzae* were harvested after 18–20 h of growth in sBHI broth under conditions designed to be microaerobic and anaerobic. During exponential growth (5 h after the incubation began in mid-log phase), we detected a transient decrease in the partial pressure of oxygen in the media. As stationary

Table 4. Proteins Found Essential or Non-Essential During *In Silico* Growth of *H. Influenzae* Under Anaerobic Condition in Simulated sBHI Medium[a]

Essential proteins		*Non-essential proteins*					
ID	*Protein name*	*ID*	*Protein name*	*ID*	*Protein name*	*ID*	*Protein name*
HI1589	AroA	HI0457	PabC	HI1527	HtrB	HI1633	PurA
HI0208	AroB	HI0123	PgsA	HI1438	IspA	HI1726	PurC
HI0970	AroD	HI0764	RibB	HI0085	LdhA	HI0888	PurD
HI0655	AroE	HI1025	AraD	HI0727	LysA	HI1207	PurF
HI1547	AroG	HI1022	BioB	HI1027	Lyx	HI0887	PurH
HI0631	CoaA	HI1219	CmkA	HI0968	MenB	HI1616	PurK
HI1308	DapB	HI1646	CmkB	HI0283	MenD	HI1428	PurN
HI0750	DapF	HI0133	Dcd	HI0194	MenE	HI0272	PyrE
HI1140	DdlB	HI1645	Fbp	HI0122	MetC	HI1224	PyrF
HI0064	FolK	HI0611	FucA	HI1234	MgsA	HI0505	RbsK
HI0812	GalU	HI0351	GalE	HI0199	MsbB	HI0288	SdaA
HI1743	Gmk	HI0819	GalK	HI0140	NagA	HI1033	SerB
HI0058	KdsB	HI0818	GalM	HI0747	Ndh	HI0529	Tdk
HI0988	LeuC	HI0820	GalT	HI1066	NrfD	HI0088	ThrB
HI0989	LeuD	HI0605	GpsA	HI1306	PgpA	HI1290	TyrA
HI1144	LpxC	HI1607	HemM	HI0211	Pgpb	HI112	XylA
HI0915	LpxD	HI0468	HisG	HI0212	PgpB	HI0558	Zwf
HI1136	MurD	HI0475	HisIE	HI0591	PotE		

[a]These proteins were classified as essential or non-essential based upon whether or not they were necessary for biomass production within the defined metabolic network under anaerobic conditions in simulated rich medium.

phase was approached (6 h after inoculation), the oxygen tension returned to nearly initial levels. Thus, these data are representative of microaerobic and anaerobic growth at stationary phase. It is not clear what effect the transient period (approximately 1.5 h) of reduced oxygen tension had on the final cellular composition. From this data, it appears that *H. influenzae* strain Rd KW20 uses classic anaerobic respiration, microaerobic respiration, and fermentative pathways. Data from the proteome analysis was confirmed by assay of selected enzymes and metabolites.

Fermentation in *H. influenzae* strain Rd KW20 results in the production of D-lactate, acetate, CO_2, and hydrogen (the latter two through the catabolism of formate). Ethanol is an end product of acetyl-CoA catabolism in *E. coli*, a reaction that reoxidizes two NADH molecules. However, we did not detect the alcohol dehydrogenase (HI0185, *adhC*) with high confidence under either growth condition. On the other hand, *adhC* gene was found to be expressed in recent cDNA array experiments using *H. influenzae* strain Rd KW20 cells grown on sBHI medium (R. Munson Jr., unpublished data). Both Pta and AckA can also function in the reverse direction facilitating the growth on acetate. Thus, depending upon the pathway used, acetate can be either an end product or a substrate. The detection of pyruvate-formate lyase (HI0180, *pfl* or *pflB*) under both conditions is to be expected: Pfl is constitutively produced and requires activation to function anaerobically. The Pfl-activating enzyme, Act or PflA (HI0179), was not detected in either growth environment, nor was the formate transporter (HI0181, *focA*). Pfl is also under transcriptional control of the anaerobic global regulator Fnr (HI1425). However, we did not detect Fnr in cell extracts from either growth condition.

Anaerobic respiration typically utilizes nitrate as an alternative electron acceptor. The nitrate concentration in sBHI broth was not measured, but we have found that it would not support the anaerobic growth of *Pseudomonas aeruginosa* strain PAO1 (Kolker et al., 2003). The genome of *H. influenzae* strain Rd KW20 contains an operon for nitrate utilization: HI0342-HI0348 (*napFDAGHBC*). However, the only protein we were able to detect with high confidence was NapB, the periplasmic nitrate reductase. Similarly, nitrite uti-

lization is facilitated by the nitrite reductase operon (HI1066-1069, *nrfABCD*). We detected HI1069 (*NrfA*) with high confidence under both growth conditions, suggesting that anaerobic respiration is used in both environments. The products of the gene cluster (HI0006-HI0008, *fdxGHI*) permit formate oxidation. Proteins from all genes in this cluster were detected with high confidence under both growth conditions.

Microaerobic respiration in *H. influenzae* strain Rd KW20 presumably involves the transfer of H^+ to the quinone pool by fumarate reductase (HI0835-HI0832, *frdABCD*) catalyzed formation of fumarate from succinate. However, the finding of FrdB and FrdA proteins under both conditions suggests that reductive fermentation may also be operative. Of the cytochromes of the *bd* complex (HI1076-HI1075, *cydAB*), CydA was found only under microaerobic conditions. The presence on an alternative glucose transporter, expressed during aerobic growth, is supported by the findings of others (Dandekar et al., 1999), who reported that strain Rd KW20 fermented glucose after overnight incubation in sBHI in "loosely capped tubes." We have found that an OMP-P4 mutant, which would only grow anaerobically on solid media (Reidl and Mekalanos, 1996), grows in 2 mL of sBHI broth in a 16 × 125 mm loosely capped tube incubated in air at 37°C (Fig. 3). Similarly, Tuyau et al. (1984), using reference capsulated strains and a complex media, found that succinate and acetate were produced after 48 h of aerobic growth. In addition, Hollander (1976) reported that *H. influenzae* strain NCTC 4560 demonstrated glucose-dependent oxygen consumption, with loss of glucose from the media and an inability of that strain to utilize fructose. Finally, further protein expression analysis suggests that one of galactose transporters (*mglB)* might fulfil role of "mysterious" glucose transporter (A.L. Smith, unpublished data). Thus, these observations, as well as the data presented herein, indicate that *H. influenzae* utilizes glucose both microaerobically and anaerobically, which has not been evident from the genome sequence analysis (Fleischmann et al., 1995; Tatusov et al., 2001). This example shows the importance of incorporating different whole-genome approaches such as proteome analysis in any global modeling of organism's metabolism. Alternatively, there is also the possibility that the complement of enzymes for glucose transport and utilization varies from strain to strain.

There are noteworthy differences between the response to oxygen of two related proteobacteria, *E. coli* and *H. influenzae.* Whereas exposure to oxygen affects expression of a large number of *E. coli* proteins (Smith and Neidhardt, 1983; Sawers et al., 1988; Lynch et al., 1996; VanBogelen et al., 1996), *H. influenzae* cells show less pronounced changes in their protein expression patterns. While one could think of a number of possible explanations for this schism, it obviously correlates with the greatly diminished variety of the signalling systems in *H. influenzae*, as compared to *E. coli*, even taking into account 2.5-fold smaller genome size of the former one. Comparing only sets of homologous genes known to be involved in the signalling processes in these two proteobacteria and, for simplicity sake, ignore possible multiplicities of their regulatory networks. Indeed, *H. influenzae* encodes a set of only four sensor histidine kinases and six response regulators, as opposed to 28 histidine kinases (seven-fold higher compared to *H. influenzae*) and 32 response regulators (over five-fold higher compared to *H. influenzae*) in *E. coli* (Galperin et al., 2001). Furthermore, *H. influenzae* encodes just one protein, HI1378, with the PAS domain that is involved in reception of the redox and energy state of the cell (Taylor and Zhulin, 1999) versus 14 PAS domain-containing proteins in *E. coli* (Galperin et al., 2001). However, the lack of a PAS domain in the sensor histidine kinase ArcB of *H. influenzae* does not prevent it from sensing the redox state of the cell (Georgellis et al., 2001). This example clearly points towards future studies of cellular regulatory networks and multiple regulations, including post-translational, transcriptional, translational, and allosteric ones. *H. influenzae* does not encode other signalling domains, such as Cache, GAF, GGDEF, and EAL, which are all well-represented in the *E. coli* genome (Galperin et al., 2001). This reduction in the signalling systems is probably due to the relatively simple lifestyle of *H. influenzae*, which does not seem to be found in nature outside its human host (Smith, 1988). Accordingly, its responses to oxygen need not be as sophisticated as the ones of *E. coli.* Another manifestation of the simpler lifestyle is the presence in *H. influenzae* of a single membrane component of the PEP-dependent sugar/phosphotransferase system with apparent specificity for fructose (Macfadyen and Redfield, 1996; Macfadyen et al., 1996), which contrasts with multiple systems with various specificities in *E. coli.*

Potential reasons for the inability of the identified complement of metabolic proteins to ensure the production of biomass are worth noting. They include three main factors: incomplete protein identification, inaccuracies in the *in silico* model, and utility of unidentified components of the complex medium utilized in

the proteome study. Most likely, all these factors play a role: no protein identification method has yet been shown to perfectly identify all expressed proteins, and current knowledge of metabolic networks is not sufficiently complete to fully match the working of a cell's metabolic network. Nonetheless, the integration of a large-scale *in silico* metabolic model with a large-scale coverage of the protein complement and directed mutation experiments is undoubtedly a worthy exercise, as demonstrated by this study. Such integrative approaches will direct the design of future studies from which deeper knowledge of *H. influenzae* metabolism, behaviour, and responses will be obtained (Kolker et al., 2002).

ACKNOWLEDGMENTS

We highly appreciate critical comments from Ludmila Chistoserdova, Dan Drell, Jim Fredrickson, Lee Hood, Eugene Koonin, Mary Lidstrom, Ken Nealson, Monica Riley, Joe St. Geme, and Judy Wall. Derren Barken, Andy Keller, Alex Nesvizhskii, Serg Stoliar, and Brian Tjaden are very much obliged for the technical assistance and fruitful discussions. This work was supported by the National Institutes of Health grants DC005980 to R.M., DC03915 to L. Bakaletz, subcontract to R.M., and AI44002 to A.L.S., a grant from the American Heart Association to B.J.A., and the Department of Energy's *Genomics: GTL* grant DE-FG03-01ER63527 to E.K.

REFERENCES

ACETI, D.J., and J.G. FERRY. (1988). Purification and characterization of acetate kinase from acetate-grown *Methanoscarcina thermophila.* J Biol Chem **263,** 15444–15448.

AKERLEY, B.J., E.J. RUBIN, A. CAMILLI, et al. (1998). Systematic identification of essential genes by *in vitro* mariner mutagenesis. Proc Natl Acad Sci USA **95,** 8927–8932.

AKERLEY, B.J., E.J. RUBIN, V.L. NOVICK, et al. (2002). A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae.* Proc Natl Acad Sci USA **99,** 966–971.

ALEXEEVA, S., B. DE KORT, G. SAWERS, et al. (2000). Effects of limited aeration and of the ArcAB system on intermediary pyruvate catabolism in *Escherichia coli.* J Bacteriol **182,** 4934–4940.

CASH, P., E. ARGO, P.R. LANGFORD, et al. (1997). Development of a *Haemophilus* two-dimensional protein database. Electrophoresis **18,** 1472–1482.

COLEMAN, H., D.A. DAYNES, J. JARISCH, et al. (2003). Chemically defined media for growth of *Haemophilus influenzae* strains. J Clin Microbiol **41,** 4408–4410.

DANDEKAR, T., S. SCHUSTER, B. SNEL, et al. (1999). Pathway alignment: application to the comparative analysis of glycolytic enzymes. Biochem J **343,** 115–124.

DE GRAEF, M.R., S. ALEXEEVA, J.L. SNOEP, et al. (1999). The steady-state internal redox state (NADH/NAD) reflects the external redox state and is correlated with catabolic adaptation in *Escherichia coli.* J Bacteriol **181,** 2351–2357.

EDWARDS, J.S., and B.O. PALSSON. (1999). Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. J Biol Chem **274,** 17410–17416.

FLEISCHMANN, R.D., M.D. ADAMS, O. WHITE, et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269,** 496–512.

GALPERIN, M.Y., A.N. NIKOLSKAYA, and E.V. KOONIN. (2001). Novel domains of the prokaryotic two-component signal transduction system. FEMS Microbiol Lett **203,** 11–21.

GEER, B.W., D. KROCHKO, M.J. OLIVER, et al. (1979). A comparative study of the NADP-malic enzymes form *Drosophila* and chick liver. Comp Biochem Physiol **65,** 25–34.

GEORGELLIS, D., O. KWON, E.C. LIN, et al. (2001). Redox signal transduction by the ArcB sensor kinase of *Haemophilus influenzae* lacking the PAS domain. J Bacteriol **183,** 7206–7212.

GREGORY, E.M., F.J. YOST, M.S. ROHRBACH, et al. (1971). Selective chemical modification of malate dehydrogenase. J Biol Chem **246,** 5491–5497.

GUYNN, R.W., and R.L. VEECH. (1974). Direct enzymic determination of acetate in tissue extracts in the presence of labile acetate esters. Anal Biochem **61,** 6–15.

HASE, C.C., N.D. FEDOROVA, M.Y. GALPERIN, et al. (2001). Sodium ion cycle in bacterial pathogens: evidence from cross-genome comparisons. Microbiol Mol Biol Rev **65,** 353–370.

HERBERT, M.A., S. HAYES, M.E. DEADMAN, et al. (2002). Signature tagged mutagenesis of *Haemophilus influenzae* identifies genes required for *in vivo* survival. Microb Pathog **33,** 211–223.

HOLLANDER, R. (1976). Energy metabolism of some representatives of the *Haemophilus* group. Antonie Van Leeuwenhoek **42,** 429–444.

HOLT, J.G., N. KRIEG, P. SNEATH, et al. (1994). *Bergey's Manual of Determinative Bacteriology* (Williams & Wilkins, Baltimore, MD).

HUNT, M., and KOHLER, P. (1995). Purification and characterization of phopshoenol pyruvate carboxykinase from *Trypanosoma brucei.* Biochem Biophys Acta **249,** 15–22.

HUYNEN, M.A., T. DANDEKAR, and P. BORK. (1999). Variation and evolution of the citric-acid cycle: a genomic perspective. Trends Microbiol **7,** 281–291.

KARLIN, S., J. MRAZEK, and A.M. CAMPBELL. (1996). Frequent oligonucleotides and peptides of the *Haemophilus influenzae* genome. Nucleic Acids Res **24,** 4263–4272.

KARLIN, S., J. MRAZEK, A.M. CAMPBELL, et al. (2001). Characterizations of highly expressed genes of four fast-growing bacteria. J Bacteriol **183,** 5025–5040.

KELLER, A., S. PURVINE, A.I. NESVIZHSKII, et al. (2002a). Experimental protein mixture for validating tandem mass spectral analysis. OMICS **6,** 207–212.

KELLER, A., A.I. NESVIZHSKII, E. KOLKER, et al. (2002b). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem **74,** 5383–5392.

KOLKER, E., S. PURVINE, A. PICONE, et al. (2002). *H. influenzae* consortium: Integrative study of *H. influenzae*–human interactions. OMICS **6,** 341–348.

KOLKER, E., S. PURVINE, M.Y. GALPERIN, et al. (2003). Initial proteome analysis of model microorganism *Haemophilus influenzae* Rd strain KW20. J Bacteriol **185,** 4593–4602.

KOONIN, E.V., and M.Y. GALPERIN. (2002). *Sequence—Evolution—Function. Computational Approaches in Comparative Genomics* (Kluwer Academic Publishers, Boston).

LANGEN, H., B. TAKACS, S. EVERS, et al. (2000). Two-dimensional map of the proteome of *Haemophilus influenzae.* Electrophoresis **21,** 411–429.

LYNCH, A.S., and E.C.C. LIN. (1996). Responses to molecular oxygen. *In* F. C. Neidhardt, R. Curtiss III, J.L. Ingraham, et al., eds. *Escherichia coli and Salmonella: Cellular and Molecular Biology,* 2nd ed. (ASM Press, Washington, DC).

LINK, A.J., L.G. HAYS, E.B. CARMACK, et al. (1997). Identifying the major proteome components of *Haemophilus influenzae* type-strain NCTC 8143. Electrophoresis **18,** 1314–1334.

MACFADYEN, L.P., and R.J. REDFIELD. (1996). Life in mucus: sugar metabolism in *Haemophilus influenzae.* Res Microbiol **147,** 541–551.

MACFADYEN, L.P., I.R. DOROCICZ, J. REIZER, et al. (1996). Regulation of competence development and sugar utilization in *Haemophilus influenzae* Rd by a phosphoenolpyruvate/fructose phosphotransferase system. Mol Microbiol **21,** 941–952.

MANUKHOV, I.V., Y.V. BERTSOVA, D.Y. TROFIMOV, et al. (2000). Analysis of HI0220 protein from *Haemophilus influenzae*, a novel structural and functional analog of ArcB protein from *Escherichia coli.* Biochem (Mosc) **65,** 1321–1326.

MILLAR, H.A., C. KNORPP, C.J. LEAVER, et al. (1998). Plant mitochondrial pyruvate dehydrogenase complex: purification and identification of catalytic components in potato. Biochem J **334,** 571–576.

MORLIN, G., V. KUWAJIMA, K. BRADLEY, et al. (2004). Genome size as an adaptation in *Haemophilus influenzae*: genomic comparison of the sequenced Rd KW20 strain with its archived ancestors and other natural isolates (submitted).

NESVIZHSKII, A.I., A. KELLER, E. KOLKER, et al. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem **75,** 4646–4658.

PAPIN, J.A., N.D. PRICE, J.S. EDWARDS, et al. (2002). The genome-scale extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. J Theor Biol **215,** 67–82.

PELTOLA, H. (2000). Worldwide *Haemophilus influenzae* type b disease at the beginning of the 21st century: global analysis. Clin Microbiol Rev **13,** 302–317.

QUAYLE, J.R. (1966). Formate dehydrogenase. Methods Enzymol **9,** 360–368.

REIDL, J., and J.J. MEKALANOS. (1996). Lipoprotein e(P4) is essential for hemin uptake by *Haemophilus influenzae*. J Exp Med **183,** 621–629.

SAWERS, R.G., E. ZEHELEIN, and A. BOCK. (1988). Two-dimensional gel electrophoretic analysis of *Escherichia coli* proteins: influence of various anaerobic growth conditions and the fnr gene product on cellular protein composition. Arch Microbiol **149,** 240–244.

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

SCHULMAN, M., and H.G. WOOD. (1975). Enzymatic determination of microquantities of acetate. Methods Enzymol **35,** 298–299.

SETHI, S., and T.F. MURPHY. (2001). Bacterial infection in obstructive pulmonary disease in 2000: a state-of-the-art review. Clin Microbiol Rev **14,** 336–363.

SETHI, S., N. EVANS, B.J. GRANT, et al. (2002). New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. N Engl J Med **347,** 465–471.

SMITH, A.L. (1988). *Haemophilus influenzae* pneumonia. In J. E. Pennington, ed. *Respiratory Infections: Diagnosis and Management* (Raven Press, New York), pp. 364–380.

SMITH, H.O., J.F. TOMB, B.A. DOUGHERTY, et al. (1995). Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome. Science **269,** 538–540.

SMITH, M.W., and F.C. NEIDHARDT. (1983). Proteins induced by anaerobiosis in *Escherichia coli.* J Bacteriol **154,** 336–343.

SMITH, M.W., and F.C. NEIDHARDT. (1983). Proteins induced by aerobiosis in *Escherichia coli.* J Bacteriol **154,** 344–350.

SPRING, T.G., and F. WOLD. (1975). Enolase from *E. coli.* Methods Enzymol **42,** 323–329.

ST. GEME, J.W., III. (2002). Molecular and cellular determinants of non-typeable *Haemophilus influenzae* adherence and invasion. Cell Microbiol **4,** 191–200.

TATUSOV, R.L., A.R. MUSHEGIAN, P. BORK, et al. (1996). Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli.* Curr Biol **6,** 279–291.

TAYLOR, B.L., and I.B. ZHULIN. (1999). PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev **63,** 479–506.

TURNER, W.L., and W.C. PLAXTON. (2000). Purification and characterization of cytosolic pyruvate kinase from banana fruit. Biochem J **352,** 875–882.

TUYAU, J.E., W. SIMS, and R.A. WILLIAMS. (1984). The acid end-products of glucose metabolism of oral and other *haemophili.* J Gen Microbiol **130,** 1787–1793.

VANBOGELEN, R.A., K.Z. ABSHIRE, A. PERTSEMLIDIS, et al. (1996). Gene-protein database of *Escherichia coli* K-12, edition 6, p. 2067-2117. *In* F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, et al., eds. *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd ed. (ASM Press, Washington, DC).

WILCOX, K.W., and H.O. SMITH. (1975). Isolation and characterization of mutants of *Haemophilus influenzae* deficient in an adenosine 5′-triphosphate-dependent deoxyribonuclease activity. J Bacteriol **122,** 443–453.

WRIGHT, C., A. DAY, N. HOOGENRAAD, et al. (1985). A two-step purification of ATP-citrate lyase from rat liver and its use in a fluorometric assay for *N*-acetylglutamate synthetase. Anal Biochem **144,** 604–609.

Address reprint requests to:
Eugene Kolker
BIATECH
19310 North Creek Pkwy., Ste. 115
Bothell, WA 98011

E-mail: ekolker@biatech.org