H. influenzae Consortium: Integrative Study of H. influenzae–Human Interactions

EUGENE KOLKER,1 SAMUEL PURVINE,1 ALEX PICONE,1 TIM CHERNY,1 BRIAN J. AKERLEY,2 ROBERT S. MUNSON, JR.,3 BERNHARD O. PALSSON,4 DAYLE A. DAINES,5 and ARNOLD L. SMITH5

ABSTRACT

Developments in high-throughput analysis tools coupled with integrative computational techniques have enabled biological studies to reach new levels. The ability to correlate large volumes of diverse data types into cohesive models of organism function has spawned a new systematic approach to biological investigation. The creation of a new consortium has been proposed to investigate a single organism utilizing these comprehensive approaches. The Haemophilus influenzae Consortium (HIC) would be comprised of five laboratories, each providing separate and complementary areas of expertise in the study of Haemophilus influenzae (HI). The 5-year study proposes to develop coherent models of HI, both as a stand-alone organism, and more importantly, as a human pathogen. Studies in growth condition specificity followed by genomic, metabolic, and proteomic experimentation will be combined and integrated through computational and experimental analyses to form dynamic and predictive models of HI and its responses. Data from the HIC will allow greater understanding of cellular behavior, pathogen–host interactions, bacterial infection, and provide future scientific endeavors with a template for studies of other pathogens.

INTRODUCTION

Understanding how microbial cells sense and respond to their environment is one of the ultimate goals of microbiology. A model that accurately accounts for bacterial responses to external environmental changes would enable scientists to predict how microorganisms in general and pathogens in particular would respond to changes in a given chemical milieu. This model could then be extended to predict bacterial response in a host. Knowledge of this highly complex, interactive pathogen-host paradigm will permit new insights into the treatment of infectious disease and suggest new approaches to antibiotic development. The proposed H. influenzae Consortium (HIC) will address the development of these predictive

1BIATECH, Bothell, Washington.
2University of Michigan Medical School, Ann Arbor, Michigan.
3Children’s Research Institute, Columbus, Ohio.
4University of California San Diego, La Jolla, California.
5Seattle Biomedical Research Institute, Seattle, Washington.
models through the integrated efforts of a wide strata of complementary teams incorporating interdisciplinary perspectives and principles from diverse academic backgrounds. The HIC will combine whole genome experimental approaches, including gene expression arrays and global protein expression studies (proteomics), with comprehensive data analysis and modeling. Hypotheses from these studies will then be tested in biochemical, physiological, and genetic experiments to yield a thorough quantitative understanding of pathogenic microbial cells. Data from these experiments will then be subjected to computational and statistical analysis and mathematical modeling, which will result in new quantitative hypotheses that will be subsequently tested by further experimentation. This iterative process will build on information obtained from in vitro and in vivo studies. Each experimental approach will be designed in accordance with, and complemented by, other multidisciplinary experimental methods. Because this work will be accomplished by the five collaborating institutions involved in HIC, we will develop a common suite of analysis and visualization tools to enable all HIC investigators to have access to the multitudinous volumes of data. We believe that a unique feature of the proposed project is the integration of diverse interdisciplinary, quantitative, experimental, and computational approaches for better understanding of behavior and responses of a pathogenic microorganism.

HIC will focus its research on one human-restricted organism, *Haemophilus influenzae* (*HI*), to learn its metabolism as it colonizes and/or invades human respiratory epithelial cells including those from patients with cystic fibrosis, and its response to carbohydrate, carbon, and/or phosphate limitation as well as to the environment inside infectious compartments. The HIC’s research will serve as a model for a more complete understanding of bacterial cells in general, with specialized emphasis on pathogenesis. We will obtain a better understanding of what *HI* does while growing in animal models (infant rats), confirming insights on human tissue as a step towards understanding the process of pathogenesis in the complex environments of the normal upper airway, the diseased lower airway of the cystic fibrosis lung and the environment encountered by *HI* after it has invaded the bloodstream. This project will bring new knowledge, novel methods, and innovative technologies that bear relevance to important problems in infectious disease. The formation of this multi-institutional Consortium will serve an important pioneering role, creating an innovative technological and biological basis for further studies of pathogenic microorganisms, their interrelationships with their hosts and with antibiotics.

**WHY *H. INFLUENZAE***?

Bergey’s *Manual of Determinative Bacteriology* (Holt et al., 1994) defines the genus *Haemophilus* as being comprised of facultative anaerobes requiring preformed growth factors, particularly protoporphyrin IX and/or nicotinamide adenine dinucleotide. It also states that the members of the genus are chemo-organotrophic, having both a respiratory and a fermentative type of metabolism. This text, which is the primary reference by which bacteria are defined, indicates that >90% of *HI* ferment galactose, glucose, maltose, ribose, and xylose, but less than 10% ferment fructose, mannose, or glycerol (Holt et al., 1994). Bergey’s *Manual* does not indicate the source for the *HI* speciation, nor the number of strains contributing to the phenotypic definition. It is not clear whether the *HI* strain Rd KW20, for which the complete genome sequence is known (Fleischmann et al., 1995), has the metabolism of the majority of the species or the minority. For example, it has been recently shown that strain Rd could ferment glycerol, and a PTS outer membrane transporter for glucose is absent from the Rd genome (Macfadyen and Redfield, 1996; Macfadyen et al., 1996). Research conducted by the HIC will clarify this issue.

We wish to study *HI* strain Rd KW20. Its genome is only 1.83 Mb (Fleischmann et al., 1995), and it is the first genome of a free-living organism to have been fully sequenced. Recent annotations of this genome show over 1,700 predicted protein coding regions, of which approximately 1,470 (>85%) have been assigned a tentative function based upon sequence homology comparisons. *HI* is a human-restricted bacterium and is acquired as a commensal in the first few months of life (Sell et al., 1973). By the time that infants are 10 months old, 50% carry *HI* in their nasopharynx; by the time they are 24 months old, 80% will have harbored *HI* at some time (Sell et al., 1973). In the first 5 years of life, the point prevalence of infected individuals was 26%, of which 75% were unencapsulated (and therefore nontypeable); the cumulative annual
carriage was 100%. In contrast to this study, which did not use selective media, a recent study (Fontanals et al., 2000) found that the cross-section carriage rate was 55% in children younger than 6 years old (Fontanals et al., 2000); of these, 93% were nontypeable. In young U.S. infants, the frequency of colonization by \textit{HI} is higher at the time they are diagnosed with otitis media (Faden et al., 1997). In the United States, 90% of children have at least one episode of acute otitis media by the time they are 2 years old (Daly and Giebink, 2000; Paradise et al., 1997). Of these cases, approximately one-third are caused by \textit{HI} (Giebank, 1989).

Common physician practice is to administer antibiotics to children with otitis media, and in the late 1980s, office visits and prescriptions for otitis media cost $1 billion annually (Stool and Field, 1989).

In virtually all reviews, nontypeable \textit{HI} is the most common bacterium isolated from the lower respiratory tract of patients with chronic obstructive pulmonary disease (COPD; chronic bronchitis) (Ball, 1996; Read, 1999; Miravitlles et al., 1999; Sethi and Murphy, 2001), which is the fourth most common cause of death in the United States (Sethi and Murphy, 2001). Loss of pulmonary function correlates with the presence of \textit{HI} in patients with mild to moderate disease (Miravitlles et al., 1998). Available data suggest that the evaluation and treatment of patients with an exacerbation of their chronic bronchitis costs $200 million annually in the United States (Grossman, 2000). \textit{HI} pneumonia is a well-recognized clinical entity (Smith, 1988). In infants and children in the United States, virtually all cases are preventable by vaccination with a type b capsular polysaccharide conjugate vaccine. In contrast, over half of \textit{HI} pneumonia in adults is caused by nontypeable strains, as is the disease in developing countries (Plouffe and McNally, 1998). Adults who acquire nontypeable \textit{HI} pneumonia often have a serious underlying disease (Shana, 1999). In “normal” children with \textit{HI} pneumonia in Pakistan, Papua New Guinea, Gambia, and the Philippines, more than 75% of the isolates are nontypeable (Shana, 1999). Disease due to nontypeable strains is not preventable by immunization with type b conjugate vaccines.

\textit{HI} infections are a serious health problem. The near universal administration of \textit{HI} capsular polysaccharide-conjugate vaccines has eliminated virtually all invasive disease (such as sepsis, meningitis, and septic arthritis) in the United States. Disease due to nontypeable strains, such as otitis media, sinusitis, and bronchitis, continues to be a health burden, and should not be overlooked in the strategic planning for the entire health care system. Further, serotype b disease remains a serious problem in much of the developing world where conjugate vaccines are not readily available (Peltoa, 2000). Understanding the metabolism and responses of \textit{HI} will permit the design of novel strategies that will prevent or interrupt the infectious processes. We are seeking to develop experimental approaches to analyze and model the metabolism of bacterial cells \textit{in vitro} and \textit{in vivo}. We have chosen \textit{HI} for a number of key reasons. Foremost, it is one of the most relevant organisms to utilize when addressing the question “How do bacterial cells exist in the human host?” Nontypeable \textit{HI} can cause otitis media, sinusitis, pneumonia, chronic bronchitis, and occasional invasive infections, while serotype b organisms are responsible for serious life-threatening invasive infections in non-immunized populations. \textit{HI} has been extensively studied, grows on defined media, and is amenable to genetic manipulation. \textit{HI} strain Rd KW20 has been extensively studied \textit{in vitro}, its genome has been sequenced and it retains many virulence determinants associated with adherence and invasion of cultured human cells. When transformed with the \textit{HI} serotype b capsule gene cluster, Rd KW20 produces the type b capsular polysaccharide and is virulent in infant rats, further indicating that this highly studied model organism retains important virulence determinants. Ongoing projects (www.microbial-pathogenesis.org/site/) together with the studies proposed here will, when completed, help to define the relationships between serotype b organisms, nontypeable organisms, and strain Rd KW20. The differences between these strains and the environmental differences at the various anatomical sites they infect will be layered onto the model that is proposed herein as these data become available.

More specifically:

1. \textit{HI} has one of the smallest known genomes, 1.83 Mb. It thus represents an optimal model of a free-living organism with a small, “not-so-complex” genome (e.g., compared to \textit{Escherichia coli}).
2. \textit{HI} Rd KW20 was the first free-living organism whose genome was completely sequenced (Fleischmann et al., 1995) by TIGR. The genome has been extensively re-annotated (Tatusov et al., 1996). Although this strain is reportedly avirulent in the infant rat model, we have recently found that it can replicate on
and invade human respiratory epithelial cells in vitro and survive in a human bronchiolar xenograft for at least 3 weeks (Daines et al., 2002).

3. In contrast to E. coli, in which there are marked strain-to-strain differences in genome organization, HI strain Rd KW20 contains a genetic background in which genes unique to HI clinical isolates can be oriented with respect to chromosomal location. Sequencing of several other important HI strains is under way in two labs participating in this proposal.

4. HI has only one natural host—humans—can exist either as a commensal or pathogen, and strain Rd KW20 is amenable to genetic manipulation.

5. HI is a facultative anaerobe that can be grown under different conditions, which is crucial for high-throughput quantitative experimental approaches, such as proteomics (protein expression), transcriptomics (RNA and gene expression), and metabolomics (metabolite expression).

6. Methods are available for generating gene deletions, insertional mutations (both polar and nonpolar), gene inactivation and transcomplementation, which are all techniques used routinely by the participating laboratories (Akerley et al., 2002; Georgellis et al., 2001; Munson, 2002; Daines and Smith, 2001).

7. The methods and technologies developed for the analysis of this microbe can be then transferred and implemented for analysis of other human pathogens.

Despite its comparative simplicity, we still do not have a detailed understanding of how exactly HI exists in the human host. Unanswered questions include the substrates it uses for growth, how its metabolic and regulatory systems operate, how it interacts with various human cell types and the innate immune system, and if changes in gene expression are necessary for colonization and persistence versus disease. For example, let us illustrate the uniqueness of the multi-disciplinary integrated study proposed here by its ability to address one more fundamental biological question, “When is a bacterium dead?”

The traditional answer is “When it can no longer grow, or is lysed.” β-lactam antibiotics kill bacteria by interfering with cell wall biosynthesis leading to a failure to maintain ionic gradients. With loss of the cell barrier, essential cytoplasmic contents leak from the cell and replication is no longer possible. Less is known about nonlytic cell death. What happens within the bacterium to cause an irreversible failure of replication? Is there a “final common pathway” similar to eukaryotic apoptotic cell death? Is individual bacterial cell death beneficial for the population? What is the role of carbon metabolism in nonlytic cell death? Answers to questions such as these will be revealed as a result of cooperative studies conducted by the five labs constituting this Consortium.

HIC APPROACH AND STRUCTURE

Five overall approaches of the entire Program Project, to be accomplished by the HIC’s three Research Cores and four Research Projects, include the following:

1. Cataloging and quantitative proteomics, using LC-MS/MS and isotope-coded affinity tags (Gygi and Aebersold, 2000) and methyl esterification (Goodlett et al., 2001) methodologies, to identify cell proteins and key metabolites, their relative and absolute patterns of expression (Keller et al., 2002; Kolker et al., 2002)

2. Whole genome mRNA and cDNA microarrays to determine expressed genes, translated and non-translated transcripts, their expression levels, and cellular regulation under different environmental conditions (Tjaden et al., 2002a,b)

3. Mutational and genetic analyses to identify essential genes (Akerley et al., 2002; Georgellis et al., 2001) and characterize phenotypes focusing primarily on H. influenzae-host interactions

4. Biochemical and physiological analyses to generate hypotheses obtained from the above analyses to guide further experiments toward building, validating, and refining models

5. Data processing and integration, computational and statistical analyses, and mathematical modeling to study correlations between expressed genes and proteins, and to reconstruct metabolic and regulatory networks (Edwards et al., 1999, 2002; Covert et al., 2001), including specific focus on pathogen–host and pathogen–antibiotic related processes
FIG. 1. Model of bacterial cellular processes. This model represents the main species of biomolecules and interactions among them that constitute bacterial cellular processes (top). The proposed study will be focused on key experimental approaches (in parentheses) and result in integral parameters (bottom). GE, gene expression; CPE, cataloging protein expression; ICAT, isotope-coded affinity tags; ME, methyl esterification.

FIG. 2. Integration of experimental and computational approaches. The proposed project tightly integrates experimental approaches, including BPE (biochemical and physiological experiments), GE (gene expression), CPE (cataloging protein expression), ICAT (isotope-coded affinity tags), ME (methyl esterification), and MA (mutational analysis), through DINAMICS (data integration, analysis, and modeling framework).
Applying approaches 1–5 to address biological questions, described above, will complement four Research Projects (Table 1):

1. Cidal action of chloramphenicol (the Smith lab)
2. Iron homeostasis and oxidative stress in infection (the Akerley lab)
3. Analysis of the cya and pho regulons (the Munson lab)
4. Building and validating the in silico genome-scale model (the Palsson lab)

We propose combining leading-edge experimental and computational technologies and methodologies to develop complementary Research Cores within the HIC framework (Table 1):

A. Global analysis of proteome and metabolome (the Kolker lab)
B. Global study of transcriptome (the Kolker and Munson labs)
C. Data analysis and integration (the Kolker lab)

In addition, the HIC has enlisted the support of six external experts, Drs. Richard Moxon (Oxford University, U.K.), Philip Bourne (University of California San Diego), Eugene Koonin (National Center for Biotechnology Information, NIH), Stephen Lory (Harvard University), Evgeni Selkov (Argonne National Laboratory), and Barry Wanner (Purdue University), to act as members of the HIC’s Advisory Board.

HIC STUDIES H. INFLUENZAE–HUMAN INTERACTIONS

The proposed integrative study necessitates efficient interdisciplinary collaboration between the researchers at the five collaborating institutions constituting the HIC. It will also be essential to develop software tools to enable the facile analysis and sharing of data and models. Linking the results of all researchers in this way will enable the construction of a hierarchy of biological systems, as shown in Figure 1. As information is obtained on multiple global processes in the cell, the hierarchy can be used to assemble a higher order understanding of complex cellular outcomes such as growth, death, and pathogenesis. This integrated approach will provide a new framework for understanding HI and will serve as a model for the study of more complex microorganisms. To understand its metabolism, we will first examine HI in vitro with different carbon and other nutrient sources and in different phases of growth, based on chemically defined cell culture medium.

Specific aims for the entire Program Project, corresponding to three Research Cores and four Research Projects, for each year are described below. The underlying principle behind the entire experimental design is that biochemical and physiological experiments, and gene expression studies and mutational analysis will be performed in advance to lead the proteomics cataloging protein expression and quantitative labeling experiments. Each year biochemical, physiological, gene expression, and mutant studies will cover experimental space to be pursued by proteomics experiments during current and subsequent years. Additionally, biochemical, physiological, and gene expression studies will explore wider ranges of environmental condi-
tions to complement development of knockouts and mutational analysis (Fig. 2). All the obtained experimental data will be processed, stored, and integrated for numerous computational analyses, including genome analysis, statistical models, correlation analysis, and mathematical modeling.

This will provide us a working and testable mathematical model for predicting the growth properties of wild-type and mutant strains grown under various conditions, including their behavior in the hosts and their drug resistance, a guidebook for engineering *HI* to any desired specifications.

**INTEGRATION OF EXPERIMENTAL AND COMPUTATIONAL APPROACHES**

A schematic view of the integration of experimental and computational approaches of the proposed project is shown in Figure 2. From the project’s onset, cutting-edge experimental, technological, and computational approaches will be integrated such that experiments are conducted which most effectively test and refine current models, and new models will be constructed which best incorporate available experimental data. This interrelationship will help ensure the development of accurate working models that are consistent with the data obtained, and which can be used as predictive tools. For example, putative essential genes and DNA regulatory sequences earlier identified by the participating HIC labs (Projects 1–4) can be verified through proteomic (Core A), array (Core B), mutational and genetic (Project 2), biochemical and physiological experiments (Projects 1 and 3), evolutionary experiments (Project 4), and computational and statistical analyses (Core C).

In addition, mutants developed by Project 2 will be evaluated for the effect of the mutation on antibiotic susceptibility in collaboration with Project 1. This will allow increasing understanding of the regulons studied by Project 2, which will impact the construction of the mutants for the *cya* and *pho* regulons to be studied in Project 3. Analysis of data acquired via the whole genome mRNA and cDNA array (Core B and Projects 1–3) and proteomics and metabolomics (Core A and Projects 1–4) will provide insights into post-transcriptional and post-translational mechanisms of protein regulation, and supply data for estimating parameters of our models of metabolic and regulatory networks (Core C and Project 4). This newly acquired evidence for sets of co-regulated essential genes can be used as additional input for subsequent genomic analysis. Model predictions can be tested through subsequent physiological and biochemical experimental studies on both wild type and newly generated mutants (Projects 1–4). This tight integration between experimental and computational approaches will enable us to achieve our goals of defining the genes and proteins involved in growth, in modeling the metabolic and regulatory networks of *HI*, and in developing an integrated understanding of *HI* as a complex dynamic system. Finally, these data will aid us in understanding *HI* interactions with the human host; the ultimate objective of the proposed HIC project.

**ACKNOWLEDGMENTS**

This work was supported in part by the DOE’s *Genomes to Life* grant DE-FG03-03ER63527 to E.K. We highly appreciated critical comments from Tatiana Gill, Brian Tjaden, Jason Papin, Anu Raghunathan, and Nathan Price.

**REFERENCES**


Munson, R.S. Jr., personal communication.


Address reprint requests to:
Dr. Eugene Kolker
BIATECH
19310 North Creek Parkway, Suite 115
Bothell, WA 98011
E-mail: ekolker@biatech.org.
This article has been cited by: