March 15th, 2003

Supplementary Materials

for “Integrative Study of *H. influenzae*-Host Interactions” proposal, P01 AI056298-01, submitted to NIAID, October 2002 and assigned to Special Panel ZAI1 RB-M (M1); PI: E. Kolker, PhD.

Contents

I. Building HIC
II. Recent work
III. Recent publications

This report describes the status of the various research projects at the five *H. influenzae* Consortium (HIC) institutions since the P01 submission in October 2002. Supporting information can be found at the BIATECH site: http://www.biatech.org/.

I. Building HIC

Since the P01 submission, the five HIC institutions have concentrated on establishing procedures for sample and data exchange, standardizing sample preparation and purification protocols, analyzing different types of data previously produced, and generating new data. Specifically: ***First,*** BIATECH has developed a secure web directory shared by the HIC investigators for data exchange purposes; this secure web has been updated daily, and currently consists of various subdirectories containing publications in preparation, raw and processed data, protocols, and other information specific to our shared research goals; ***Second,*** a new web interface up-/down-loading service has been launched to support the secured HIC web directory; ***Third,*** the HIC public web page was initiated, and it currently consists of an updated overview of the work done by the HIC labs during last half year; ***Fourth,*** Kolker and Munson labs have adopted RNA sample preparation protocols developed by Smith lab, and additionally Kolker lab is currently reproducing a protein purification protocol developed by Akerley lab; ***Fifth,*** exchange among the five HIC institutions is not limited to protocols, procedures, and samples, but also includes ideas and unpublished data, along with regular conference calls, which already significantly improved the quality of our research; ***Sixth,*** the efforts outlined above as well as new data generated and its analyses have resulted in ***three*** joint HIC publications: the first is an overview of the integrated studies of our model microorganism *H. influenzae*, proposed by HIC consortium [A1]; the second focuses on proteome analysis of *H. influenzae* [B1]; and the third compares an *in silico* metabolic model to proteome and mutation data [B2].

II. Recent work

1. Kolker lab

The Kolker lab has been primarily focusing on building BIATECH information technology (IT) and computational infrastructure along with infrastructure for its two facilities: the Sample Preparation and Array facility and the Proteomics and Metabolomics facility. These facilities are utilizing new approaches for sample preparation, gene characterization, and protein and metabolite expressions. Together with the IT and computational core, these two facilities are prototypes of the proposed HIC’s Research and Development Centralized Cores. Additionally, BIATECH’s research (for more information on this and other projects, please visit the HIC web page) resulted in papers on: statistical models for proteomics [A2, B3] and arrays [B4]; high-throughput proteomics [A3]; large-scale analysis of hypothetical proteins [B5]; and more.

As stated in the P01 proposal, all the papers and data (including raw data), will be available on BIATECH’s web site upon publishing. For example, gene expression data from *E. coli* produced by analysis of whole genome arrays and described in two 2002 publications can be downloaded at no cost and have been requested from over 50 labs. Further, the protein control data set for proteomics research, described in the 2002 paper, was recently made available on our web site, and has been already used by over 20 research institutions and firms, including the Swiss Bioinformatics Institute, ETH (Switzerland), UCLA, UCSF, UCSD, Agilent Technologies, GeneProt, Lion Bioscience, Merck, and Oxford GlycoSciences.

BIATECH recently established collaborative relationships relevant to the proposed project in the following areas: proteomics, with groups lead by R. Aebersold (ISB), R. Smith (PNNL; BIATECH is analyzing data produced by FTICR at PNNL, and PNNL is implementing our statistical models), and J. Yates (Scripps); trancsriptomics – Affymetrix, Agilent, MWGBiotech, and Operon; metabolomics – Genomatica and Phenomenome Discoveries; and data analysis and integration – J. Crowley (CRAB, institute-level collaboration), groups lead by P. Bernstein (Microsoft Research) and B. Mishra (NYU).

Clarification on current BIATECH support. The 2nd year of BIATECH's grant from DOE, which started 11/01/02, does ***not*** provide any support for work on *H. influenzae*. As an exception, DOE generously agreed to allow BIATECH to finish the work on *H. influenzae* within the first three months of the grant's 2nd year (11/01/02 – 02/01/03). Due to the absence of alternative funding, this work is currently ***suspended***.

2. A. Smith lab

On going projects in the Smith lab consist of investigations of mechanisms of invasion by nontypeable *H. influenzae* as well as quorum sensing and biofilm formation in *H. influenzae* isolated from the middle ear of patients with otitis media. The value of *H. influenzae* Rd strain KW20 in understanding pathogenesis has been limited as it has been used as the avirulent control in animal experiments. We have found that this strain has certain virulence properties which are cell-system and model dependent [A4]. Thus, certain aspects of virulence can be examined in this strain. We developed a chemically defined media which will permit studies of metabolism, including quorum sensing [B6]. The molecules used by *H. influenzae* in the invasion of epithelial surfaces are unknown. To gain insight into this process we have begun experiments examining the role of the *vap* system in this process [B7]. We have determined that the serum resistance of the nontypeable invasive *H. influenzae* is not due to evasion of the classical complement cascade through the mitigating action of the mannose binding lectin. Preliminary experiments indicate that serum-resistant organisms bind the C4B binding protein, which in turn interferes with C3 convertase effectively stopping the deposition of the membrane attack complex.

We have found that a *H. influenzae* homolog of the quorum sensing autoinducer (AI-2) synthetase (*luxS) is* a borinic acid ester of furanone and the bacterium secretes a molecule which is recognized by *Vibrio harveyi*. Mutation of $luxS_{Hi}$ in otitis media isolates yielded a strain which has a shorter mean generation time in defined media, and is more virulent in a chinchilla model of otitis media. Experiments are in progress to better assign a more tractable phenotype to the $luxS_{Hi}$ mutant.

3. Akerley lab

Since the submission of our initial component of this application, the Akerley lab has made progress in a several areas relevant to the proposed studies. Some of this work was recently published [A5, A6], and other work described in this report has been submitted for publication [B8]. Both of these papers are directly relevant to the proposal as described in these Materials. In addition, we have contributed experimental data to the HIC collaborative manuscript that integrates proteomics, metabolic modeling, and physiological information from mutational studies [B2]. Our recent progress can be divided into three main areas described in the update for project 2: comparative genomics; genetic systems; and integrative studies. Comparative genomics will allow the extension of our results to virulent clinical isolates of *H. influenzae*. Genetic systems, which were initially described in our preliminary data, have now been applied in the context for which they are needed for the proposed work, providing validation that they will be effective. Initial integrative studies mark the transition to utilizing the expertise of the HIC to answer questions that are difficult to address using a gene-by-gene approach.

4. Munson lab

In aim 1 of project 3, we proposed to describe, at the transcriptome level, the role of cAMP and the cAMP binding protein (*crp*) in *H. influenzae* gene regulation. The first goal was the construction and characterization of a strain with a mutation in the adenylate cyclase gene (*cya*). Previously, Redfield and

coworkers characterized a *cya* mutant constructed by transposon mutagenesis. Their mutant was characterized with respect to the role of cAMP in carbohydrate transport and competence for transformation. The data clearly indicate that there are cAMP-dependent and cAMP-independent responses. We have constructed a new strain that contains a nonpolar mutation in the *cya* gene. We have further verified this result by Southern blot analysis. This mutant has been used in a number of microarray experiments, several of which are described in the update [A7].

With respect to Core B, we proposed to use the Operon 70-mer ORF set for *H. influenzae* Rd strain KW20. This is a unique set of 70-mer oligos, one for each gene, and it became available in early 2003. We have performed a number of experiments using this strain and the *cya* mutant to optimize our microarray conditions. We have optimized our RNA isolation procedure, determined that it is not necessary to remove the rRNAs when working with the Operon oligo set, and have increased the data reproducibility by optimizing our hybridizations on a Genomic Systems GENETAC Hybridization Station. The short-term goals of refining our methodologies to produce the best and most consistent RNA possible and then defining conditions that will minimize the slide-to-slide and within slide variation in hybridization have largely been completed. We are currently working to optimize data normalization protocols.

Thus, the Munson lab has made considerable progress in establishing the 70-mer slide based arrays in our facility. We have demonstrated up-regulation of known competence genes during the development of competence and identified new genes that were not previously associated with competence development. We have also demonstrated that up-regulation of competence genes is partially dependent on cAMP levels. These data confirm and extend previously described data in the literature based largely on *lacZ* transcriptional fusions, and will provide significant new information that is more global in nature. These experiments will all be repeated, confidence levels determined, and slide-to-slide variation factored in. All microarray data will be processed according to MIAME standards.

5. Palsson lab

The Palsson lab continues to be active in building genome-scale models of microorganisms and validating them experimentally as the enclosed publication list shows. A brief summary of the most relevant achievements with respect to the P01 proposal are:

1. We have implemented the dHPLC method to find mutations that occur during adaptive evolution and applied it to our *E. coli* evolved strains [A10]. A total of four mutations in a transcription factor gene were identified. This study shows that the proposed re-sequencing method works.
2. We have installed a pulse-field electrophoresis gel apparatus and begun to examine genome-scale rearrangements that take place during adaptive evolution in all the evolved strains of *E.coli* K12. We could extend this method to detect mobilization of genetic elements in *Haemophilus* once we have evolved HI strains in different conditions.
3. We have nearly completed the construction of the *H. influenzae* model published in JBC in 1999 in a new software package SimPheny by Genomatica. A number of updates over and above the published model are being incorporated.
4. We have performed a total of approximately 80 adaptive evolutions for *E. coli* and the procedure for these evolutions is now routinely implemented in our laboratory. The first data set was published in *Nature* [A9]. This study demonstrated for the first time that the outcome of adaptive evolution could be predicted based upon genome-scale *in silico* modeling. We have now evolved three dozen knock-out strains of *E. coli* simulating the loss of gene-product activity due to inactivation by a pharmacological agent [B9]. This same procedure can clearly be applied to *H. influenzae*.
5. We have now completed all paperwork required to obtain the permission for growing *H. influenzae* in our laboratory and culturing will begin soon, unless no further funding can be obtained.

III. Recent Publications (all the papers referenced in the text above are located in the subdirectory "Papers")

A. Selected papers from the 21 that have been published or in press since the P01 submission:

1 (Joint). Kolker, E., Purvine, S., Picone, A. F., Cherny, T., Akerley, B. J., Munson, R. S. Jr., Palsson, B. O., Daines, D. A., and Smith, A. L. 2002. *H. influenzae* consortium: Integrative study of *H. influenzae* - Human interactions. *OMICS A Journal of Integrative Biology*, 6(4), 341-48

2. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical Chemistry*, 74, 5383-5392.

3. Guina, T., Purvine, S., Yi, E., Eng, J. K., Goodlett, D., Aebersold, R., and Miller, S. 2003. Quantitative proteomic analysis indicates increased synthesis of a quinolone by *Pseudomonas aeruginosa* isolates from cystic fibrosis airways. *Proceedings of the National Academy of Sciences*, 100(5), 2771-2776.

4. Daines, D. A., Cohn, L. A., Coleman, H. N., Kim, K. S., and Smith, A. L. 2003. *Haemophilus influenzae* Rd has virulence properties. *Journal of Medical Microbiology*, 52, 1-6.

5. Akerley, B. J., and Lampe, D. J. 2002. Analysis of gene function in bacterial pathogens by GAMBIT. *In* V. L. Clark and P. M. Bavoil (eds.) Methods in bacterial pathogenesis. Part C. *Methods in Enzymology,* 358, 100-108.

6. Bergman, N. H., and Akerley, B. J. 2003. Position-based scanning for comparative genomics and identification of genetic islands in *Haemophilus influenzae* type b. *Infect. Immun.,* 71, 1098-1108.

7. Mason, K. M., Munson, R. S. Jr., and Bakaletz, L. O. Nontypeable *Haemophilus influenzae* gene expression induced *in vivo* in a chinchilla model of otitis media. *Infect. Immun.*, in press.

8. Sun, S., Scheffler, N. K., Gibson, B. W., Wang, J., and Munson, R. S. Jr. 2002 Identification and characterization of the N-acetylglucosamine glycosyltransferase gene of *Haemophilus ducreyi. Infect. Immun.,* 70, 5887-5892.

9. Ibarra, R.U., Edwards, J.S., and Palsson, B.O. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420, 186-189.

10. Raghunathan, A., and Palsson, B.O. 2003. Scalable method to determine mutations that occur during adaptive evolution of *Escherichia coli. Biotechnology Letters*, 25, 435-441.

B. Selected papers from the 20 that have been submitted or to be submitted since the P01 submission:

1 (Joint). Kolker, E., Purvine, S., Galperin, M. Y., Stolyar, S., Goodlett, D. R., Yi, E., Nesvizhskii, A. I., Xie, T., Keller, A., Eng, J. K., Picone, A. F., Tjaden, B. C., Siegel, A. F., Hood, L., Reilly, T. J., Makarova, K. S., Palsson, B. O., and Smith, A. L. Initial proteome analysis of model microorganism *Haemophilus influenzae* Rd strain KW20, *Journal of Bacteriology*, subm.

2 (Joint). Kolker, E., Price, N., Raghunathan, A., Galperin, M. Y., Makarova, K. S., Xie, T., Purvine, S., Picone, A. F., Cherny, T., Tyler, R., Akerley, B. J., Munson, R. S. Jr., Smith, A. L., and Palsson, B. O. *In silico* metabolic model and protein expression of *Haemophilus influenzae* Rd strain KW20 in rich medium.

3. Nesvizhskii, A. I., Keller, A., Aebersold, R., and Kolker, E. Estimating the accuracy of protein identifications made by MS/MS and database search, *Analytical Chemistry*, subm.

4. Pauler, D. K., Jiang, C., Tjaden, B. C., Picone, A. F., Cherny, T., Jacobson, J., Crowley, J. J., and Kolker, E. Comparing apples and oranges: cDNA versus oligonucleotide arrays.

5. Kolker, E., Xie, T., Shabalina, S., Purvine, S., Picone, A. F., Cherny, T., and Galperin, M. Y. Large-scale analysis of conserved hypothetical proteins from *H. influenzae* and *E. coli*.

6. Coleman, H. N., Daines, D. A., Jarisch, J., and Smith, A. L. Chemically defined media for the growth of *Haemophilus influenzae* strains.

7. Daines, D. D., Jarisch, J. and Smith, A. L. Characterization of $vapD_{Hi}$ in nontypeable *Haemophilus influenzae*.

8. Wong, S. M., and Akerley, B. J. A Tightly-regulated conditional expression system and marker-linked mutagenesis approach for functional genomics of *Haemophilus influenzae.*

9. Fong, S. S., Marciniak, J. Y., and Palsson, B. O. Description and interpretation of adaptive evolution of *Escherichia coli* K12 MG1655 using a genome-scale *in silico* metabolic model, *Journal of Bacteriology*, subm.

10. Allen, T. E., and Palsson, B. O. The wobbling proteome: the systemic effect of codon usage, *Science*, subm.