# CIBRED: Engineering Education on Cyberinfrastructure with a Multidisciplinary Approach for Non-Engineering Students

Arun K. Datta,[1,2,*] Jacqueline Caesar,[1,3] Daphne Rainey,[4] Stephen Cammer,[4] Julie Schuman[4], Oswald Crasta[4]

[1] National University of Community Research Institute (NUCRI), [2] School of Engineering and Technology, [3] College of Letters and Science (COLS), National University, San Diego, CA; [4] Virginia Bioinformatics Institute, Virginia Tech University, Blacksburg, VA.

*For Correspondence

ABSTRACT

Although the role of cyberinfrastructure in engineering education and research is advancing, the use of the concept and infrastructure are quite limited in the courses for the allied health professionals. CIBRED* (CI-TEAM Implemenrtation for Biological Researchers, Educators, and Developers), an NSF funded project, provides a unique opportunity for these authors to introduce the concept of cyberinfrastructure to non-engineering educators and students by designing, developing, and deploying course materials with a interdisciplinary approach. At present, two courses are being developed for deployment and assessment during Fall of 2009. These interdisciplinary courses are being developed in a modular format integrating scientific and technology information from a variety of disciplines. These modules can be incorporated into existing or newly developed courses. One module is for allied health professionals to learn about cyberinfrastructure for healthcare management. The other module focuses on human migration, which introduces engineering education to the undergraduate students from humanities and social sciences. Project-based learning concepts have been implemented in developing these courses to teach various relevant disciplines. The focus is to teach students from diverse disciplines some essential concepts on computer technology in the context of applying cyberinfrastructure. These courses developed for K13 & K14 levels will be offered in an innovative classroom setting for hands-on experimental learning with a focus on solving scientific problems as a team. These courses will also be deployed for online learning in a virtual classroom. The effectiveness of such an approach, introducing concepts from engineering education to the non-engineering students, will be assessed through formative and summative methods for further development and dissemination.

---

*CIBRED is funded by NSF award OCI-0753375 to O. Crasta

## INTRODUCTION

Advances in various technologies in biology, medicine and computation have enabled researchers to generate more experimental data for understanding medical science at the molecular level. This trend is now common in most of the scientific fields that can be exploited for data utilization, data analysis and more useful data mining and visualization of data. Moreover, the advances in computational technology are changing the way research is conducted in all aspects of science and have led to the generation of seemingly limitless possibilities of national and international collaboration and sharing of data for research, education and training. The Office of Cyberinfrastructure (OCI) was created at the National Science Foundation in June 2005[1] following the recommendation of a Blue Ribbon Committee chaired by Daniel Atkins of the University of Michigan to formulate the scientific policies leading to more useful collaboration using computer technology and coordinating these activities building the infrastructure[2,3] The vision of the OCI is making it possible to bridge the gap between the research data infrastructure and the scientific community through federated database systems, collaboratories, and a powerful portal with high performance computing capability for tracking, analyzing, visualizing and interpreting the experimental data and results[4]. Through competitive, merit-reviewed awards for leading-edge, IT-based infrastructure, which is increasingly essential to science and engineering leadership in the 21st century, OCI is developing an infrastructure, termed Cyberinfrastructure (CI) to facilitate rapid progress in education and research.

## Cyberinfrastructure and teragrid

“Cyberinfrastructure” describes integrated information and communication technologies for distributed information processing and coordinated knowledge discovery, which promises to revolutionize the way that science and engineering are done in the 21st century and beyond[3]. It is the coordinated aggregate of software, hardware and other technologies, as well as human expertise, required to support current and future discoveries in science and engineering. Thus, CI includes supercomputers, data management systems, high capacity networks, digitally-enabled observatories and scientific instruments, and an interoperable suite of software and middleware services and tools for computation, visualization, and collaboration[1]. The challenge is to integrate relevant and often disparate resources to provide a useful, usable, and enabling framework for research and discovery characterized by broad access and “end-to-end” coordination. NSF has sponsored projects to develop principles for the “design and evaluation of IT-enabled scientific "collaboratories," or "centers without walls" in which researchers can perform their research without regard to physical location, interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information and data in digital libraries and repositories" (details at www.nsf.gov). One such project, TeraGrid, is a national-scale high performance computing facility to provide large scale resources and services in support of advancing scientific research and education[5,6]. Presently, TeraGrid (TG) includes 11 Resource Providers (RPs) and is a leader and major component of the emerging national and international e-Science (details at www.teragrid.org). TeraGrid is an open scientific discovery infrastructure combining leadership class resources at partner sites to create an integrated, persistent computational resource. Using high-performance network connections, the TeraGrid integrates high-performance computers, data resources and tools, and high-end experimental facilities around the country. Currently, TeraGrid resources include more than a petaflop of computing capability and more than 30 petabytes of online and archival data storage, with rapid access and retrieval over high-performance networks. Researchers can also access more than 100 discipline-specific databases in TeraGrid. With this combination of resources, TeraGrid is the world's largest, most comprehensive distributed cyberinfrastructure for open scientific research. Thus, CI supported by Teragrid is becoming the foundation of 21st Century's discovery through innovative scientific research.

Over the years, the role of CI has evolved. Initially, CI was defined by its role – a research environment to support advanced data acquisition, data storage, data management, data integration, data mining, data visualization and other computing and information processing services over the Internet. Soon the definition incorporated the advances enabled by CI and CI became the technological solution to the problem of efficiently connecting data, computers, and people with the goal of enabling derivation of novel scientific theories and knowledge. In its current stage, CI is taking its place in the advancement of education as a platform for delivery of learning content and continues to advance into a platform to expedite the transition from research data to pedagogical practice[7]. The latest role of CI advances in the face of the realization that a new workforce needs to be trained in the CI environments in which it will function.

Earlier, an awareness program on CI was launched under NSF funded 'inFormation Year 06-07' by the HASTAC organization (www.hastac.org). Among the events organized by nine participating institutions, 'InCommunity' was organized by the NU Community Research Institute (NUCRI) of the National University to promote the concept of CI to the community (archives can be viewed at http://nucri.nu.edu/incommunity). Nevertheless, integration of the engineering concept of CI is very limited, especially to the non-engineering students due to the lack of appropriate course curriculum for schools and colleges.

**Skilled workforce needed to sustain CI/Teragrid Program**

Central to this CI environment is the most important element, the skilled workforce that is required to maintain and sustain this program (Figure 1). Two components toward the development of the new workforce are needed: educators with an understanding of the demands of CI and materials for training the students. In answer to some of these needs, OCI has now established a rich source of educational and research materials through TeraGrid to meet the 21st Century's demand for scientific talent (Materials are freely available through CI/TeraGrid – www.teragrid.org). Additionally, OCI put forth the CI-TEAM (Cyberinfrastructure, Training, Education, Advancement, and Mentoring for Our 21st Century Workforce) program to aid education initiatives directed toward this new workforce. Training of a 21$^{st}$ Century workforce demands students have a firm grounding in interdisciplinarity, especially in the sciences.

Today's young generation born in the technological landscape of the digital age is already familiar with participatory media[8] and the practice of integrating knowledge with the aid of digital tools, such as, iPhone or Google phone, and even interacting with unseen peers in the digital metaverse space, such as, in Second Life (http://secondlife.com/). The goal now is to harness and develop these skills. However, the problem is to find instructors who can meet the challenge of training these students and serve as a bridge between these techno-savvy students and the scientific discoveries happening in the CI supported techno-sphere. Many educators today are not familiar with the requirements of the 21$^{st}$ Century digital age where information literacy[9,10] and multimedia savvy are required skills.

**The role of CI in education and training and CIBRED**

The National Research Council and the US Department of Education conducted a study[11] to evaluate teaching and learning and the use of technology. This study recognized that the proper use of computer technology helps the learning process. Computational technologies have proven to have profound impacts on the practice of science and engineering[3]. TeraGrid has developed many resources of research and educational materials exploiting the computational technologies (TeraGid07 & TeraGrid08). CI's and its role in adhering systems, tools, and services emerging

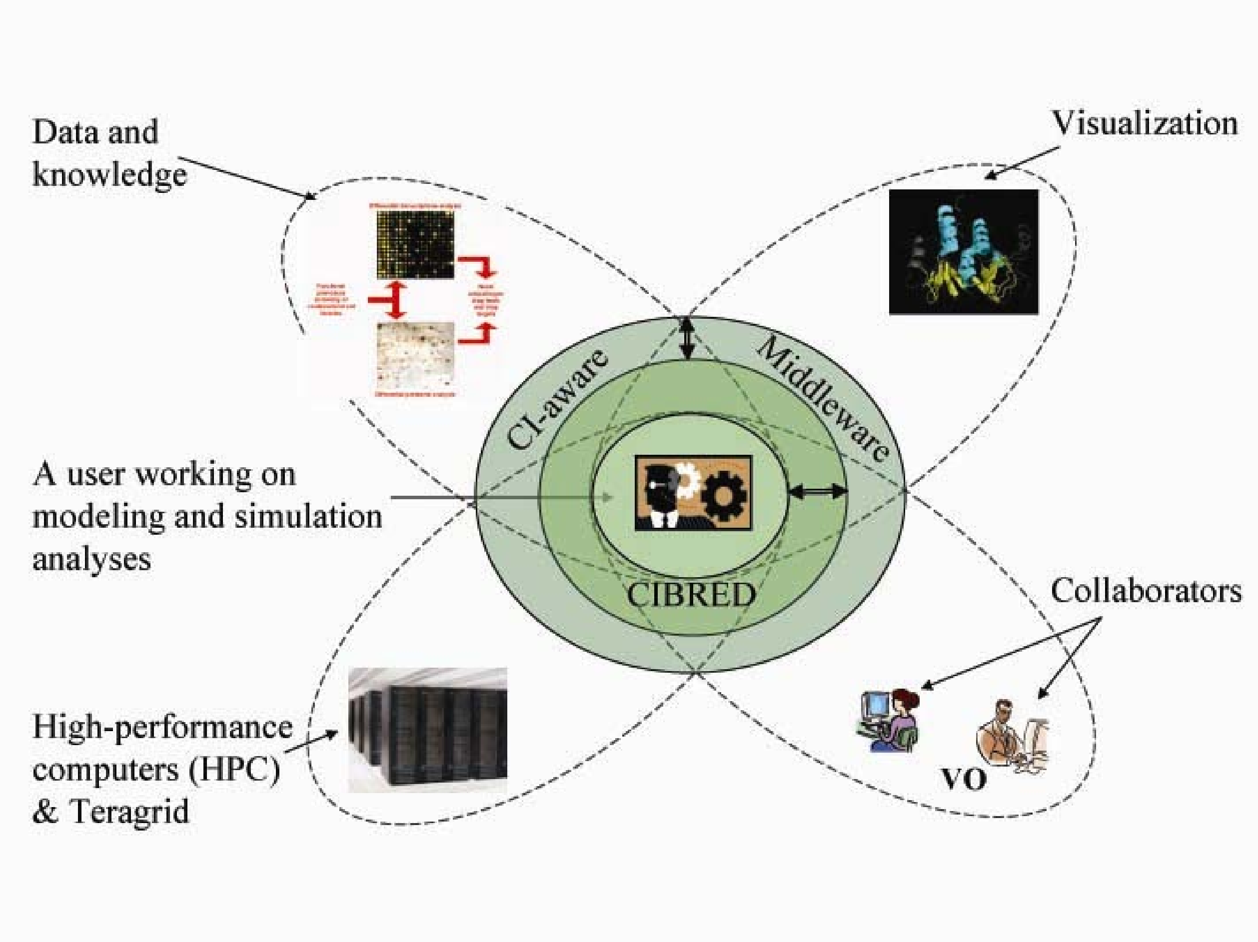


**Figure 1**. Relationships between the knowledge worker at the center of the CI with high-performance computers and Teragrid, middleware, VO (virtual organizations), data management and knowledge discovery, and visualization services. The objective of CIBRED is to educate and bring awareness of CI (*Courtesy of S. Wang*[12]; *courtesy of Stan Watowich*[13]).

from computational technologies, enables individuals, groups, and organizations to advance research and education in ways that revolutionizes the practice of participation. Once again, a new workforce empowered with the knowledge and skills to design, deploy, adapt and apply CI, are needed to sustain this revolution across all areas of science and engineering. The OCI CI-TEAM program supports educational training projects that position the national science and engineering community to engage in integrated research and education activities promoting, leveraging and utilizing cyberinfrastructure systems, tools and services (more info at, http://www.nsf.gov/crssprgm/ci-team/).

One such program, CIBRED (CI-TEAM Implementation for Biological Researchers, Educators, and Developers), brings together faculty, researchers, educators, and academic administrators from multiple institutions to design, develop and deploy course materials in a collaborative way with the objective of preparing future scientists, engineers, and educators to adopt and deploy, cyber-based tools and environments for research and learning, both formal and informal (for more detail, see: cibred.vbi.vt.edu). The activities involve a diverse group of people and organizations, with particular emphasis on the inclusion of traditionally underrepresented individuals, institutions, and communities as both creators and users of cyberinfrastructure. Like all CI-TEAM funded projects, CIBRED seeks to broaden and diversify the population of individuals and institutions leveraging existing or current development efforts in cyberinfrastructure technologies, open software standards and open educational resources. The vision of this program is to realize the potential of cyberinfrastructure and high performance computing to empower a larger and more diverse set of individuals and institutions to participate in science and engineering education, research and innovation. The mission is to develop and evaluate an extensible, scalable, and comprehensive learning program for the students at academic institutions to effectively utilize CI and TeraGrid resources and services, to advance scientific discovery in all fields.

Six fundamental tenets underlie the CIBRED project:

1. New scientific discovery will increasingly require transdisciplinary research within a cyberinfrastructure (CI) environment.
2. Scientists require increased knowledge and proficiency in team science to conduct transdisciplinary research within a CI environment.
3. There is a decrease in the number of students entering the STEM disciplines.
4. Curricula for existing students require new courses that are built on the principles of team science and methodology for conducting transdisciplinary research within a CI environment.
5. Learning environments must increasingly engage students in STEM disciplines, build awareness of team science and transdisciplinary research within a CI environment, and inspire students to further investigate or pursue a career in these areas.
6. To prepare future scientists, existing researchers and developers must be linked to educators and learners to create data-rich learning environments.

CIBRED, will help bridge this gap by training the trainers through workshops. Although TeraGrid and SC Community (http://sc08.supercomputing.org/) periodically conduct courses through workshops to train people for CI/TeraGrid, CIBRED workshops will specifically serve the school teachers and college instructors to enable teaching CI embedded courses in the classroom.

**CIBRED for interdisciplinary courses**

This project was not proposed to design courses for a specific scientific discipline, but rather to utilize the information from multiple disciplines while developing some courses for CIBRED. Participating institutes are: Virginia Tech University (VT), National University, Howard University, Hampton University, Denbigh High School, Pheobus High School, Blacksburg High School, Auburn High School, and Galileo Magnet High School. The project established the objectives and goals for team collaborations via a kick-off meeting on April 10, 2008 (The agenda and the copy of all presentations of this meeting can be found at: http://staff.vbi.vt.edu/bsharp/CIBRED_Workshop/Minutes.htm). A variety of CIBRED-course materials for both high school (K8 -K10) and college levels (K13 – K14) are under development by the participating institutions. These course materials are being developed in a modular way, the concept of CI as the common module, for disseminating institutions to choose from according to their standards of learning. The courses are developed by integrating scientific and technology information from a variety of interdisciplinary fields ranging from basic concept of CI, Molecular Biology, Bioinformatics, Genomics and Proteomics to Health Informatics into modules. These modules can be incorporated into existing or newly developed courses. To accommodate specific needs of various institutions and to facilitate adaptation by additional institutions in the future, we have standardized the framework in the following areas[14]:

a. Introduction to CI
b. Inclusion of current research (context, data, and tools)
c. Role-based involvement of the students for transdisciplinary learning; and
d. Forward-looking conclusion of the project with review of the CI systems.

Project based learning concepts[15-18] have been implemented in developing these courses to teach various relevant disciplines. The focus is to teach students from diverse disciplines essential concepts on computer technology in the context of applying CI. The type and number of role-based modules to be incorporated into a CI course will be solely at the discretion of the deploying institution. Each course will finish with a CI reflection component that reiterates the basic CI components learned, relates them to the role-based modules, and addresses the potential applications of both the specific scientific domains utilized and CI.

**CI for allied health professionals**

Concerns for the seemingly exponential rate of increase in US healthcare costs have had an impact on the national economy. This concern was reflected in President Obama's first State of the Union Address on February 24, 2009. A component of these rising costs has been identified as the lack of coordination of patient medical records due to the continued and widespread use of paper records. The use of paper records inhibits the ability to coordinate care, measure quality, and reduce medical errors. Processing paper claims also costs twice as much as processing electronic claims. The Obama administration has adeptly targeted a strategy to increase the use of healthcare informatics as a means to update the healthcare system and simultaneously reduce costs associated with uncoordinated patient information. A thorough analysis of the present state of healthcare has revealed that the lack of standards for sharing health related information contributes to increased health care costs through a redundancy of services[19]. Developing the state of health informatics has the potential to eliminate many of these concerns and increase the efficiency of the system. The present US administration's plan is to invest approximately $50 billion over the next five years to move towards the adoption of standards-based electronic health information systems, including electronic health records in the nation's health system. In February 2009, President Obama provided $19.2 billion for health IT in the American Recovery and Reinvestment Act. This legislature is expected to create a demand for skilled healthcare professionals with knowledge of CI. With this in mind, a CIBRED course, 'Cyberinfrastructure in Healthcare Management' has been created for the students at undergraduate level (K13-14).

Medical informatics is a transdisciplinary science that requires domain knowledge of various disciplines, such as, medical science (anatomy, physiology, etc.), molecular biology and bioinformatics, computer technology, among other fields. In the field of modern molecular medicine, advanced computational technologies will certainly play an important role in managing and analyzing massive quantities of medical data once knowledgeable workers are available to serve this growing field. In the treatment of cancer, for example, it is already accepted that it is more beneficial if a patient is treated on the basis of his/her personal health profile by using information obtained from various data including data from microarray analysis, toxicogenomic analysis, pharmacogenomic analysis[20,21]. Such analysis supported by CI will enable a medical specialist to treat a cancer patient in future by prescribing right medicine at the right dose(s) based on the individual's profile. Over time, it has been realized that individuals respond differently to drugs and sometimes the effects are unpredictable. Safer treatment will utilize the knowledge synthesized from the intersection of genomics and medicine supported by CI, which has the potential to yield a new set of molecular diagnostic tools that can be used to individualize and optimize drug therapy[22].

As discussed during BIO2008, it is becoming clear that a new understanding of the dynamic interplay between genes and environment, made possible by technologies arising from the Human Genome Project, helps support the individualization of medicine[23,24]. Currently, available software tools and focus on data analysis do not provide a platform for the management of patient information[25]. A shift in technology fueled by computational thinking[26] is needed to speed the coupling cyberinfrastructure and medical informatics rapidly and economically[25,27]. Computational modeling and simulation based on cyberinfrastructure-enhanced medical informatics allows researchers to tackle large and complex medical problems[28].

Although the author (AKD) has gained knowledge in both the molecular biology and computer science by educational training and working several years in medical institutes, it is virtually impossible for any one educator/domain expert to develop a course focusing on transdisciplinary concepts and problem solving needed to train the health-informaticians and others needed to serve the 21st Century workforce. A collaborative framework has been designed involving domain experts from the Virginia Bioinformatics Institute at Virginia Tech and the National University in an attempt to develop such a course. Modules will be used so that subjects can be flexible enough for adapting the course. The current modules focus on helping students to develop and understanding of the importance of cyberinfrastructure framework to address critical issues at the societal level, thus introducing the transdisciplinary nature of the project. The course has four main modules: 1) Introduction to CI, 2) Molecular Biology and Bioinformatics, 3) Health Informatics, 4) Vaccine Development and Drug Discovery. The second module is being developed with the help of domain experts available from the Virginia Bioinformatics Institute at Virginia Tech. One author (AKD) is developing the third module 'health informatics'. The Health

Informatics module stems from a course developed and taught for undergraduate allied health professional students at the School of Health and Human Services at the National University. The feedback obtained from the students has become useful in developing this current module. The other two modules are being developed by these authors with the feedback from TeraGrid Resource Providers, feedback from the Campus Champion Program[6,29,30], University of California-San Diego, and The Scripps Research Institute.

One case study used in the project is dengue management. Nevertheless, for the project, an instructor can consider any other disease management using this case study as a model. Dengue, transmitted to humans by the domestic, daytime-biting mosquito, *Aedes aegypti* (Figure 2), is caused by four closely related virus serotypes of the genus Flavivirus, a member of the family Flaviviridae. Each serotype is sufficiently different that there is no cross-protection and epidemics caused by multiple serotypes (hyperendemicity) can occur. According to WHO, 50 million cases

**Figure 2**. This infected female mosquito, *Aedes aegypti* (and rarely *Aedes albopictus*) transmits dengue virus to humans through the bites. Mosquitoes generally acquire the virus while feeding on the blood of an infected person. After an incubation period of eight to 10 days, an infected mosquito is capable, of transmitting the virus during probing and feeding for the rest of its life[31]. Public awareness to control this daytime-biting domestic mosquito needs to improve through education and community outreach.

are reported annually across the globe (Figure 3). Often the victims suffer from more serious symptom, Dengue Hemorrhagic Fever (DHF/DSS), which may turn fatal. In 2002 alone, 91 deaths occurred in Brazil. This disease previously unknown in US is becoming a concern of the Center for Disease Control (CDC). National Institute of Allergy and Infectious Diseases (NIAID) experts now see dengue as potential threat to U.S. health[32]. Unfortunately, there is no drug or vaccine, at present, to cure dengue. Public health education to control the mosquito or the disease is not good enough even in the countries where it is prevalent. Outbreak of this disease is also common. Moreover, global warming apparently is changing the pattern of infection: more people residing at higher altitude in countries like Costa Rica are becoming infected, causing a grave concern. There are various elements to learn by studying this disease, and thus, may be used as a case study for CI – supported disease management.


**Figure 3**. Global distribution of dengue infection. The vector mosquitoes are prevalent in tropical and sub-tropical areas. However, these are now causing concern not only in Central and South Americas but also in North Americas. Apparently climate change has a profound effect on these vectors. People residing in higher altitude in Costa Rica, as for example, are becoming infected.

**Method of teaching/learning -** Project-centric hands-on experimental learning

Learning is a complex phenomenon that depends on various factors including the student's state of mind[11]. The same topic delivered by the instructor in the classroom is 'learnt' differently by different students. There are various techniques to improve the teaching techniques for facilitating learning[33]. Among the teaching techniques we tested, we realized that agile techniques supports learning techniques better[34]. In general, many strategies can be employed, including problem-based learning[15-18], technology-based learning[35], game-based learning[36,37], work-based learning[38,39], inquiry-based learning[40,41], project-based learning[42,43], team-based learning[44], web-based learning[45,46] and participatory learning[8]. Our approach for teaching this course centered on project-centric learning[14] with support of agile teaching technique[34]. The course has been designed with background information (cyberinfrastructure) introduced at the beginning and summarized at the end. However, how to balance the load (in terms of both time and knowledge) among students from multiple disciplines remains a challenge.

A disease management includes a variety of disciplines including Clinical study, Public health education and training, Epidemiology, Environmental issues, local and national resources supported by Geospatial Information Systems (GIS) in case of an outbreak/epidemic, Electronic

Health Record and other Health Informatics –tools and Services, Disease outbreak predictive tools, Decision Support Systems, Drug Discovery Research, Vaccine development - to name a few. Moreover, a student needs to know various public policies including HIPAA compliance (see: http://www.hhs.gov/ocr/privacy/index.html) and FERPA (see: http://www.ed.gov/policy/gen/guid/fpco/ferpa/index.html) for school-aged children. Patient's data sharing even among the healthcare service providers are restricted by these compliances. The law also dictates how the permission should be obtained from parents while sharing the health related data on minors.

In a typical classroom setting, an instructor will initially provide all the relevant information on this disease management supported by this course through lecture and demonstration. In this phase, the CIBRED course instructors will provide background information necessary for the students to understand the scenario for the project and to develop strategies for investigations including identification of the virus, accessing the TeraGrid and National network of bioinformatics resources, and suggesting control of preventive measures in case of an outbreak and the role of Geospatial Information Systems (GIS) in coordinating the resources. In the second phase, following the agile problem driven teaching techniques[34], an instructor will survey the students' aptitude & interest and will facilitate forming a group/team working on a given project, such as, dengue case management. The students will then work through the scenarios assuming different roles in the team. Each team may utilize different tools and address the problem from a different perspective. Some students can assume a role as clinicians, as for example, to 'diagnose' the disease by conducting a series of pathological tests including blood tests and checking the physical symptoms. In doing so, the students need to learn what the symptoms are for dengue and what sort of laboratory tests are required to diagnose such a case. This dengue management project will specifically require students to use various computational and web based tools to discover the identity of the virus, given some preliminary biological data under the supervision of the instructor and in consultation with the researchers and sequences to test a hypothesis regarding the origin of the sample sequence (Figure 4). Students will practice what they would learn by becoming familiar with large, CIBRED team. They should also need to learn about the DNA sequencing that generates a code for a portion of a pathogenic genome derived from a biological or environmental sample. The project will require learning about how gene sequences are stored and how to access these comprehensive repositories such as the NCBI database Gen-Bank, and focused bioinformatics resources, such as, Patho-Systems Resource Integration Center (PATRIC),

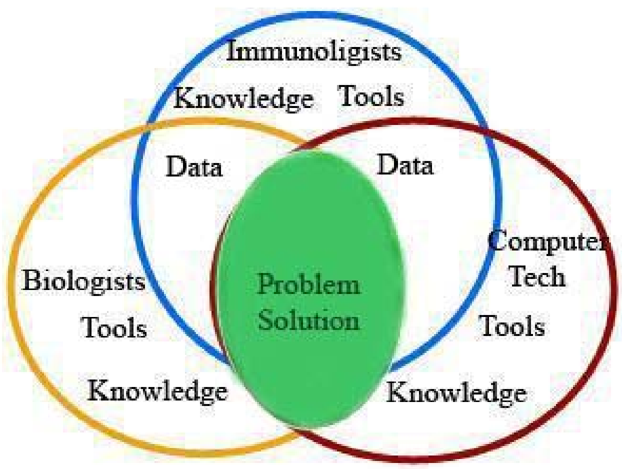


**Figure 4**. The key to solve a scientific problem is to gather vital information from a variety of disciplines, synthesize those to acquire knowledge. These information are gathered through experiments, literature search, enquiry and communication.

Pathogen Information (PathInfo), and Molecular Interactions Network (MINet) documents that are maintained at Virginia Bioinformatics Institute (VBI), TeraGrid, Biological Workbench http://workbench.sdsc.edu/) or elsewhere on the Internet. Techniques for comparing and manipulating sequences will be addressed to insure students could use similarity to a known genetic sequence to predict possible function for a sequence of interest. In the process, putative genes needed to have their sequences translated to protein sequences for functional features to be identified. Predictions of structural featutures of the protein, such as secondary structure, signal sequence and transmembrane segments, will provide them the first clues as to what types of proteins are present. They will learn the basic information on proteins, how a protein is synthesized in a living cell and folded into a 3D-structure, what factors primarily are responsible for folding into that structure, how a protein can be selected as a drug target. Protease is regarded as a potential drug target for dengue[48,49], and the references therein). Some students can use molecular modeling techniques for this protein[50], during the investigation. Target selection can, however, be made after several proteins are investigated using multiple data resources, including expression data where available. Experimental laboratory analysis could provide students with new information about the target that may be analyzed to generate possible leads in the search for drug candidates. In the process, students will also learn about small molecule drug discovery supported by World Community Grid (http://www.worldcommunitygrid.org/) of IBM and how that is being used by a scientist, Stan Watowitch, at The University of Texas Medical Branch (Galveston, TX) for small molecule potential drug candidates against dengue virus[51] (also, see the story at http://www.tacc.utexas.edu/research/users/features/worldgrid.php). Searching his publications in PubMed followed by reviewing the information, and may even interacting with his team at UTMB, the students will learn how computational screening is done for identifying small molecules as potential drug candidates and later conducting cell biology experiments to verify their computational screening results. They will also learn how a small molecule candidate drug can be 'discovered' through structure-based drug design technique when a 3D- structure of the protein molecule is known following the experiments using X-ray crystallography or NMR technique. Some students can assume a role as researchers for vaccine development. They will learn what is vaccine, what is immunity, what is antibody and how do they take part in conferring immunity against a disease, such as, dengue. They will also learn how the environment plays an important role in host-pathogen interaction (Figure 5). In the process, they will need to know the genus (Flavivirus) and family (Flaviviridae) of a dengue virus. They will also learn how a single family can be responsible for causing multiple major diseases including hepatitis C, dengue, Yellow fever, and West Nile fever. They will learn that these viruses are positive strand RNA viruses and their mode of pathogenecity is similar (see: http://www.bioinformatics.org/dengueDTDB/Pages/main.htm). It will be necessary to characterize


Figure 5. Environmental factors play important roles in the host-pathogen interaction[47].

the family and use the sequences to generate a multiple sequence alignment to profile the family for subsequent investigation. They will also learn how a 'protein motif' can be selected using computational technique and bioinformatics tools before selecting the right monoclonal antibody(ies) for clinical trial. Such a study is being done in the laboratory of Eva Harris at UC-Berkeley[52,53]. Her collaboration with researchers in Managua (Nicaragua) has established informatics system to identify and monitor dengue infected cases[54]. Most of these information and experimental results are available in PubMed. In addition, this approach will provide the students unique opportunity for hands-on experimental learning with a focus on solving scientific problems as a team. At the end, it is expected that the members of each sub-team will generate a report on their analyses and present those to the other team members. These reports should contain enough information and need to be presented in a clear fashion for other members to understand the information and for the entire group to agree upon a conclusion. Such exercise will also teach them working in a team through social networking & associated tools, because it will be necessary for these students to cooperate with each other to achieve the best results. Finally, the students will assemble the information, synthesized interpretations and conclusions, and may communicate the results to the live and online community. Earlier, during the CIBRED demonstration phase, one of the participating students described her involvement to the local media, and the interview was posted for the government bio-defense agencies such as DoD, NIAID, and the scientific experts in this process.

**CI for Humanities and Social Scientists**

A module for teaching CI impact on studying 'Human Migration' will be designed and deployed at the college level (K13-14). Advances in technology have had an impact not only on biological and medical sciences, but also on social sciences. Anthropologists have begun to look at the relationships between human migration and cultural diversity with a different lens. Influenced by discoveries in genetics as a result of the Human Genome Project and other scientific milestones, new data has opened the door to meaningful research and provided insights into the connections between human evolution, migration, cultural diversity and diversity[55] (also see, https://genographic.nationalgeographic.com/genographic/index.html for more information). Old concepts have been revisited in light of new information such as the role of natural selection and mutations in genetic variations and the more recent emphasis on the concept of genetic drift. Projects are underway to create an International databank for DNA samples in an effort to obtain larger and more comprehensive data sources. Needless to say, an interdisciplinary and transdisciplinary academic approach that includes cyberinfrastructure, science and the social science is a relevant and timely prospect for helping students become scientifically literate.

Students will be introduced to the extensive information on the evolution of mankind. Research by key theorists and practitioners, especially the most recent works of Luigi Luca, Sforza-Cavalli and Stephen Wells[56-58] will be explored with the idea of tracing ancestry and cultural groups and tracking scientific methodologies to the present. They will also have an opportunity to review the work of Allan Wilson and Rebecca Cann's research on mitochondrial diversity and mtDNA sequencing[59], which supported the notion of a "single ancestry" prompting the sentiment "Eve is the connecting link for humankind." Information from the National Institutes of Health's National Human Genome Research Institute offers examples of human migration and genetics (see: http://www.genome.gov/25019968). Other scientific search data sources will be included that will focus on understanding cells, genes, DNA, chromosomes (mtDNA and Y - chromosome) and the impact of the environment on diseases. Assignments will be directed toward investigating and gathering data from the major sources identified earlier especially the NCBI Gnome database (the Genome sequence, gMap and Map Viewer and Salmonella SNP), the analytical tools in BLAST, the modeling of the Virtual Cell by CMC-Nature Gateway (Figure 6).

Based on the scientific inquiry model, groups of students will actively engage in experiential projects that are designed to trace patterns of migration by learning about DNA sequencing, DNA hybridization and the Hardy-Weinberg gene frequency and variation theorem. The objective is to connect the sources to disease patterns within cultural groups. They will discuss and review diseases such as sickle cell anemia and Tay Sachs disease, cystic fibrosis as case studies. Students will then engage in problem solving by conducting research based on a case study on diseases impacting more diverse populations such as multiple sclerosis and cancer. Teams of students with diverse skill levels and experiences will engage in investigative data search and analysis.

One of the most important aspects of the course is that expert resources will be available to confer with students on their projects. Finally, students are encouraged to question and explore.

They are introduced to such engaging resources as *Case It*, a National Science Foundation-sponsored case-based learning project involving molecular biology, computer simulations and

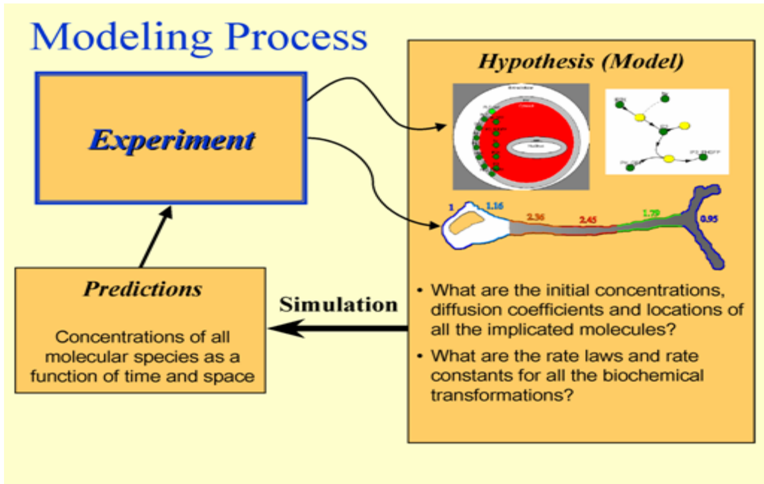


**Figure 6.** The Virtual Cell was designed specifically as a tool for a wide range of scientists. It was developed at the National Resource for Cell Analysis and Modeling at the University of Connecticut Health Center (http://www.nrcam.uchc.edu) as a unique computational tool that provides a software environment for analyzing, modeling, and simulating cellular function that can range from metabolic pathways to signaling networks to membrane transport and electrophysiology. These models can be based on both experimental data (biochemistry, molecular biology, imaging, etc.) and purely theoretical assumptions by using a very general conceptual structure of a model as a collection of arbitrary processes.

Internet conferencing[60]. The students also are exposed to the Environmental Genome Project sponsored by the National Institute of Environmental Health Sciences, which consider how specific human genetic variations contribute to environmentally-induced disease susceptibility[61,62].

NSF supported CI/Teragrid usage is free for such academic exercise. Teragrid has two important features. It provides the computation power (supported by training) necessary for conducting drug discovery and vaccine developmental research. It also provides a vast amount of scientific information through portals developed for specific type of scientific projects. One can obtain all these information by visiting www.teragrid.org. For computational biology experiments (for the definition, see: http://www.bisti.nih.gov/docs/CompuBioDef.pdf), a student can get access to the computing power of TeraGrid through the instructor serving as PI/Principal Investigator. NSF has facilitated this process through Campus Champion Program of TeraGrid (http://www.teragrid.org/eot/campuschamps.html). The Campus Champions program supports campus representatives as the local source of knowledge about high-performance computing opportunities and resources. This knowledge and assistance will empower campus researchers, educators, and students to advance scientific discovery[30,63]. An instructor for teaching CIBRED developed courses will be encouraged to apply for the teragrid computing resources through Campus Champion Program. Being a Campus Champion Program of the National University, AKD will be helping the students for such computing analysis.

Availability of publications in most of the peer-reviewed journals was not free at the individual level, until recently. The student or the instructor needed to visit public or University/institute library to get the access to any printed peer-reviewed publications. Online access was also limited and subscription based. However, recently US Federal Government law passed in December 2008 has empowered the National Institute of Health (NIH) to execute the requirement by all of its grantees who publish articles springing from their govt. funded research to file a copy in PubMed Central (http://www.pubmedcentral.nih.gov), a free open access website run by the National Library of Medicine, within one year of its initial publication date. This has made very easy to obtain most of the publications free of charge.

**Assessment of the Outcome of CIBRED**

The CIBRED assessment for these and other courses will address four research questions derived from a comprehensive list of possible questions. The criteria for selection of those questions included a determination of which are most important to meeting project goals, relevant to the project deliverables and are within the scope of available time and resources. Additional criteria were derived from NSF 02-057 (*The 2002 User-Friendly Handbook for Project Evaluation*). Following those criteria, a document has been generated that identifies the questions and links them to generalized project goals.

An assessment of course readiness will be conducted based on:

1. Accessibility of the network to the collaborators, other institutions and participating student populations
2. Outreach efforts
    a. What other institutions have expressed an interest
    b. What type of reaction have other institutions had

c. CIBRED Web site usage statistics: To disseminate the CIBRED course model, CI courses, and related materials to other participating high schools and colleges via a web-based CIBRED network.

3. The comparative results between the two groups of students completing the courses
   a. Perceptions and experiences of the course developers, course instructors/educators and participating students
   b. Comparison of students interests in and perceptions of STEM disciples before and after completing the courses
   c. Comparison of students knowledge of team science, transdisciplinary research and CI before and after taking the course
   d. Comparison of student's ability to participate in team science, transdisciplinary research and CI activities before and after the course taken.

**Importance or relevance to other institutions**:

The CIBRED courses will be offered to two groups of students at each institution. The first group will complete the courses in the Spring, 2009, and the second group in the Fall, 2009. A comparison of results will be conducted to determine course readiness for deployment beyond the participating institutions. Those results will be obtained by surveying the participants before and after taking the course, interviews and feedback from all other stakeholders. The CIBRED Course Model may prove to be an effective and sustainable means for developing CIBRED courses through such collaborative efforts. Later, these courses will be made available for use in other academic institutions evaluating the method of deployment on a case-by-case basis.

**Acknowledgement**

The authors acknowledge contribution through stimulating discussions by various participants of CIBRED, particularly Bruce Sharp of the Virginia Bioinformatics Institute, Tom Klagholz of Klagholz and Associates, Thomas Maccalla of the National University Community Research Institute, Pradip Dey of the School of Engineering and Technology, and Michael Maxwell of the College of Letters and Science, National University. This study is conducted with NSF funding to O. Crasta (OCI-0753375).

## References

[1]. National Science Foundation (2006). About Office of Cyberinfrastructure (OCI). Retrieved February 26, 2009 from www.nsf.gov/od/oci/about.jsp

[2]. National Science Foundation (2006). NSF Names Daniel Atkins to Head New Office of Cyberinfrastructure. February 8, 2006, from http://www.nsf.gov/news/news_summ.jsp?cntn_id=105820

[3]. Atkins DE, et al. (2003). Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyber - infrastructure. (http://www.nsf.gov/od/oci/reports/atkins.pdf)

[4]. National Science Foundation (2006). NSF'S Cyberinfrastructure vision for 21st century discovery. January, 2006, from http://www.nsf.gov/od/oci/ci_v5.pdf

[5]. McLean, S.A. (2007). Gateway to the TeraGrid for Bioscientists. TeraGrid '07, the second annual TeraGrid conference, June 4 - 8, Madison, Wisconsin. Retrieved Feb 24, 2009 from http://www.teragrid.org/events/teragrid07/agenda.html

[6]. Lathrop, S. (2007). *Education, Outreach and Training (EOT).* TeraGrid '07, the second annual TeraGrid conference, June 4 - 8, Madison, Wisconsin. Retrieved Feb 24, 2009 from http://www.teragrid.org/events/teragrid07/agenda.html; Advancing Scientific Discovery through TeraGrid. TeraGrid '08, the 3rd annual TeraGrid Conference, TeraGrid '08 www.teragridforum.org/mediawiki/images/8/8f/TeraGrid_Overview.ppt

[7]. Madhavan KPC, Goasguen S. (2007). *Integrating Cutting-edge research into learning through Web 2.0 and virtual environments*. Retrieved online February 25, 2009 from http://library.rit.edu/oajournals/index.php/gce/article/viewFile/94/55.

[8]. Barab, S. A., Hay, K. E., Barnett, M., Squire, K. (2001). Constructing virtual worlds: Tracing the historical development of learner practices. *Cognition and Instruction, 19*(1), 47–94; also see: http://en.wikipedia.org/wiki/Participatory_Media for definition

[9]. Shapiro, J.J., and SK Hughes, S.K. (1996). Information literacy as a liberal art. Educom Review, 31(2): 31; Association of College and Research Libraries (1989). *Presidential Committee on Information Literacy: Final Report.* American Library Association, January 10, 1989, in Washington, D.C. Retrieved February 24, 2009 from http://www.ala.org/ala/mgrps/divs/acrl/publications/whitepapers/presidential.cfm

[10]. Breivik. P. S. & Senn, J. A. (1998). *Information literacy: Educating children for the 21st century*. (2nd ed.). Washington, DC: National Education Association.

[11]. Bransford, J. D., Brown, A. L., Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

[12]. Wang, S., and Zhu, X-G. (2008). Coupling Cyberinfrastructure and Geographic Information Systems to Empower Ecological and Environmental Research. *BioScience*, 58 (2): 94-95.

[13]. Watowitch, S. (2009). Personal communication; also a part of the Figure 1 was retrieved Feb 24, 2009 from http://www.bmb.utmb.edu/faculty/watowich/

[14]. Rainey, D., Faulkner, S., Craddock, L., Crammer, L., Tretola, B. (2007). *A project-centric approach to cyberinfrastructure education*. TeraGrid '07, the second annual TeraGrid conference, June 4 - 8, Madison, Wisconsin. Retrieved Feb 24, 2009 from http://www.teragrid.org/events/teragrid07/agenda.html

[15]. Barell, J. (2007). Problem-based learning: An inquiry approach (2nd Edition). Thousand Oaks: Corwin Press.

[16]. Duch, B. (2008). *Problem-based learning*. Retrieved, February 10, 2008 from http://www.udel.edu/pbl/

[17]. Kaminski, J. A., Sloutsky, V.M., Heckler, A. F. (2008). Learning theory: The advantage of abstract examples in learning math. *Science*, 320: 454 – 455.

[18]. Savin-Baden, M. (2003). *Facilitating problem-based learning: Illuminating perspectives*. Buckingham: SRHE/Open University Press.

[19]. Devine, K., O'Clock, P. and Lyonsc, D. (2000). Health-Care Financial Management in a Changing Environment. Journal of Business Research, 48 (3): 183-191.

[20]. Furic L, Livingstone M, Dowling RJ, Sonenberg N.(2009) Targeting mtor-dependent tumours with specific inhibitors: a model for personalized medicine based on molecular diagnoses. *Curr Oncol.*16(1):59-61

[21]. Whitfield, K.E., McClearn, G. (2005). Genes, Environment, and Race: Quantitative Genetic Approaches. *By Whitfield, Keith E.; McClearn, Gerald American Psychologist.*, Vol 60(1):, Jan 2005, 104-114.

[22]. Evans, W.E, Relling, M.V. (2004). Moving towards individualized medicine with pharmacogenomics. *Nature*, 429 (6990): 464-8.

[23]. Grimaldi D, Claessens YE, Mira JP, Chiche JD. (2009) Beyond clinical phenotype: the biologic integratome. *Crit Care Med*. 37(1 Suppl):S38-49.

[24]. Individualized Medicine Emerging From Gene-Environment Studies. *ScienceDaily*. Retrieved February 22, 2009, from http://www.sciencedaily.com- /releases/2005/01/050123213425.htm

[25]. Cao Y, Wang N, Ying X, Li A, Wang H, Zhang X, Li W. (2009) BioSunMS: a plug-in-based software for the management of patients information and the analysis of peptide profiles from mass spectrometry. *BMC Med Inform Decis Mak*. 9(1):13.

[26]. Wing, J.M. (2006). Computational thinking. *Communications of the ACM* 49: 33–35.

[27]. Sharpe RE Jr, Sharpe M, Siegel E, Siddiqui K. (2009) Utilization of a Radiology-Centric Search Engine. J Digit Imaging. 2009 Feb 12 [Epub ahead of publication PubMed: 19214633]

[28]. Hoffman MA. (2007) The genome-enabled electronic medical record. *J Biomed Inform*. 40(1):44-6.

[29]. Hunt, K., and Lathrop, S. (2008). *Campus Champions – bringing HPC to your campus* (BoF Topic), Nov 19, in SC08: International Conference on high performance computing, networking, storage and analysis. Austin, TX. Nov 15-21, 2008. Retrieved, November 19, 2008 from http://teragridforum.org/mediawiki/index.php?title=SC08_BoF&printable=yes

[30]. Hunt, H., Lathrop, S., Kucera, B., Barr von Oehsen, J., Moye, R., Datta, A. (2009). *Campus Champions – Bringing High Performance Computing to your Campus*. EDUCAUSE 2009 Annual Conference (Panel discussion), Denver, Colorado, November 3–6, 2009.

[31]. Schneider BS, Higgs S. (2008) The enhancement of arbovirus transmission and disease by mosquito saliva is associated with modulation of the host immune response. *Trans R Soc Trop Med Hyg*. 102(5):400-8; also see: [WHO] World Health Organization. *Dengue*. Retrieved Feb 24, 2009, from http://www.who.int/topics/dengue/en/

[32]. Morens, D.M., Fauci, A.S. (2008). Dengue and Hemorrhagic Fever: A Potential Threat to Public Health in the United States. *JAMA*, 299: 214 - 216.

[33]. Dey, P., Uhlig, R., Amin, M., Datta, A., Romney, G., Gatton, T., Wyne, M., Cruz, A. (2009a) Teaching Mathematical Reasoning in Science, Engineering, and Technology. *Journal of Research in Innovative Teaching* , Volume 2, *in press*.

[34]. Dey, P., Gatton, T., Amin, M., Wyne, M., Romney, G., Farahani, A., Datta, A., Badkoobehi, H., Belcher, R., Tigli, O., and Cruz.A. (2009b). Agile Problem Driven Teaching in Engineering, Science and Technology. ASEE/PSW to be held in San Diego, March 19-20, *accepted*.

[35] Trondsen, E. (1998). *Technology-based learning*. Retrieved February 10, 2008 from http://www.sric-bi.com/LoD/summaries/LoDRpt01.shtml

[36]. Prensky, M. (2004). *Digital game-based learning*. New York: McGraw-Hill.

[37]. Van, R. E. (2008). *Game-based learning*. EDUCAUSE Review, 41(2), 16–30.

[38]. Bailey, T. R. (2003). Working knowledge: Work-based learning and education reform. New York: RoutledgeFalmer.

[39]. Cunningham, I., Dawes, G., & Bennett, B. (2004). *The handbook of work-based learning*. Adershot: Gower Publishing.

[40]. Eick, C. J., & Reed, C. J. (2002). What makes an inquiry oriented science teacher? The influence of learning histories on student teacher role identity and practice. *Science Teacher Education*, 86: 401–416.

[41]. Educational Broadcasting Corporation (2008). *Inquiry-based learning*. Retrieved, February 10, 2008 from https://www/thirteen.org/edonline/concept2class/inquiry/inded.html

[42]. Helic, D., Maurer, H., & Scerbakov, N. (2004). Knowledge transfer processes in a modern WBT system. *Journal of Network and Computer Applications*, 27(3):163–190.

[43]. George Lucas Educational Foundation (2008). *The Project based learning*. Retrieved, February 10, 2008 from http://www/edutopia.org/projectbasedlearning

[44]. Michaelsen, L. K., Knight, A. B., & Fink, D. L. (2008). *Team-based learning: A transformative use of small groups in college teaching*. Sterling, VA: Stylus Publishing.

[45]. Lee, M. & Baylor, A. L. (2006). Designing Metacognitive Maps for Web-Based Learning. *Educational Technology & Society*, 9 (1), 344-348.

[46] O'Neil, H. F., Perez, R. S. (2006). *Web-based Learning: Theory, research, and practice*. Hillsdale, NJ: Erlbaum.

[47]. Sadikot, R.T., Blackwell, T.S., Christman, J.W., and Prince, A.S. (2005). Pathogen–Host Interactions in Pseudomonas aeruginosa Pneumonia. American Journal of Respiratory and Critical Care Medicine, 171: 1209-1223.

[48]. Chambers, T. J., Hahn, C.S., Galler, R., Rice, C. (1990). Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol*, 44: 649–688

[49]. Rosaura, P., Valle, C., and Falgout, B. (1998). Mutagenesis of the NS3 Protease of Dengue Virus Type 2. *J Virol.*, 72(1): 624–632.

[50]. Lee, Y.K., Tan, S.K., Wahab, H.A., Yusof, R., Rahman, N.A. (2007). Nonsubstrate Based Inhibitors of Dengue Virus Serine Protease: A Molecular Docking Approach to Study Binding Interactions between Protease and Inhibitors. *Asia Pacific Journal of Molecular Biology & Biotechnology*, 15 (2): 53-59.

[51]. Wang, E., Ni, H., Xu, R., Barrett, A.D., Watowich, S.J., Gubler, D.L., and Weaver, S.C. (2000). Phylogenetic Analysis of Endemic/Epidemic Dengue Emergence from Sylvatic Progenitors, J. Virology 74, 3227-3234; also see, http://www.bmb.utmb.edu/faculty/watowich/)

[52]. Kyle, J.L., Harris, E. (2008). Global spread and persistence of dengue. *Annu Rev Microbiol.*, 62:71-92. Review

[53]. Kyle, J.L., Balsitis, S.J., Zhang, L., Beatty, P.R., Harris, E. (2008). Antibodies play a greater role than immune cells in heterologous protection against secondary dengue virus infection in a mouse model. *Virology*, 380(2): 296-303

[54]. Avilés, W., Ortega, O., Kuan, G., Coloma, J., Harris, E. (2007). Integration of information technologies in clinical studies in Nicaragua. *PLoS Med*. 4(10):1578-83.

[55]. Behar, D.M., Rosset, S., Blue-Smith, J., Balanovsky, O., Tzur, S., et al. (2007) The Genographic Project public participation mitochondrial DNA database. PLoS Genet 3(6):1083- 1095

[56]. Stone, L., Lurquin, P. F., and Cavalli-SforzaGenes, L.L. (2007). Culture, and Human Evolution: A Synthesis. Blackwell Publishing: MA.

[57]. Sforza-Cavalli, L.L. (2001). *Genes, Peoples, and Languages*. University of California Press: Berkeley.

[58]. Well, S. (2002). *The Journey of Man: A Genetic Odyssey*. Random House: New York 2002

[59]. Cann, R.L, Stoneking, M., Wilson, A.C. (1987). Mitochondrial DNA and Human Evolution, Nature, 325: 31-6.

[60]. Bergland, M., Klyczek, K., Mogen, K., and Johnson, D., Lundeberg, M., Nelson, M.A. (2001). Case It! - a collaborative BioQUEST project to enhance case-based learning in university and high school biology education worldwide via molecular biology computer simulations and Internet conferencing. CAL-laborate Volume 6 June 2001 (Published by: UniServe Science, URL: http://science.uniserve.edu.au/pubs/callab/vol6/index.html ).

[61]. Levin, S.A., Grenfell, B., Hastings, A., Perelson, A.S. (1997). Mathematical and computational challenges in population biology and ecosystems science. *Science*, 275: 334–343 (also see: http://www.niehs.nih.gov/research/supported/programs/egp/)

[62]. Molnar, S. (2006). *Human Variation* (Sixth Edition). Pearson Prentice Hall.

[63]. Datta, A., Crasta, O., Rainey, D., and Maccalla, T. (2009). *CIBRED- educating Cyberinfrastructure concept and teragrid resources to the 21st Century's biological researchers*. EDUCAUSE 2009 Annual Conference (Panel discussion), Denver, Colorado, November 3–6, 2009.

BIOGRAPHICAL INFORMATION

**Arun K. Datta, Ph.D**., has double majors in Biochemistry and Software Engineering. As a Director (Technology) of the National University Community Research Institute, he supervises various national and international projects (see: www.arundatta.info) particularly on healthinformatics. He developed a course on 'Informatics for Health Professionals' and now teaches for the School of Health and Human Services (SOHHS). As a Senior Faculty Associate of the National University, he is currently teaching computer science and related subjects in various schools of the University, particularly in the School of Engineering and Technology (SOET). His research interests are: educational technology, health informatics, and protein motif analysis, especially those associated with pathological conditions.