Research at the Command’s Bioinformatics Cell

Jaques Reifman, PhD†
Thomas McKenna, PhD††

Bioinformatics is just one area the Telemedicine and Advanced Technology Research Center (TATRC) explores in its quest to find leading-edge medical technologies to help the warfighter. The TATRC is a subordinate element of the U.S. Army Medical Research and Materiel Command (USAMRMC) and is charged with managing core research development test and evaluation and congressionally mandated projects in telemedicine and advanced medical technologies. To support its research and development efforts, TATRC maintains a productive mix of partnerships with federal, academic, and commercial organizations. The TATRC also provides short duration, technical support (as directed) to federal and defense agencies; develops, evaluates, and demonstrates new technologies and concepts; and conducts market surveillance with a focus on leveraging emerging technologies in health care and health care support. Ultimately, TATRC’s activities strive to make medical care and services more accessible to Soldiers, Sailors, Marines, and Airmen, reduce costs, and enhance the overall quality of military health care.

Introduction

Recognizing the need to develop in-house expertise in the quickly growing and rapidly changing fields of bio and medical informatics, the USAMRMC charged TATRC to establish a bioinformatics cell (BIC). The BIC consists primarily of physical scientists with backgrounds in statistics, computer science, mathematics, and engineering, and has both internal and external missions. Its external mission is to monitor new and emerging technology developments in bioinformatics to enable the identification of new opportunities, publicize Army requirements, influence the course of biotechnology developments, establish strategic alliances and partnerships with industry, academia, and other government agencies, including other Services within the Department of Defense, and to serve as the Command’s focal point for coordination of bioinformatics-related Congressional Special Interest (CSI) projects.

The BIC’s internal mission is to serve as the USAMRMC advisor for bioinformatics to help align the Command’s portfolio and investment strategy with new developments and emerging technologies and to support research efforts that cut across the four major research focus areas of the Command: Military Infectious Diseases, Combat Casualty Care, Military Operational Medicine, and Medical Chemical and Biological Defense. The BIC staff is involved in various activities ranging from organizing bioinformatics workshops to hosting tutorials to performing joint research with the Command’s life scientists in both bio and medical informatics. The following sections summarize our research activities in these two research areas.

Bioinformatics Research

Bioinformatics research at the BIC involves the development of software systems to warehouse, manage, and analyze genomic and proteomic data. These activities will help us gain insight into gene function and protein function/structure in support of the Command’s missions to develop improved assays for threat detection and diagnostics, characterize health effects from exposure to military relevant toxic hazards, and develop medical countermeasures in the form of drugs and prophylactics.

High Throughput Gene Functional Analysis. High throughput deoxyribonucleic acid (DNA) microarray technology is being widely exploited by USAMRMC investigators across the four major research focus areas of the Command. In support of these activities, we are collaborating with Wayne State University to extend functional analysis tools, such as Onto-Tools, to support animal models, such as Caenorhabditis elegans and rats, being used by our investigators at the Center for Environmental Health Research and the U.S. Army Institute for Surgical Research (USAINSIR), respectively. These tools allow the automatic translation of lists of genes found to be differentially regulated under given conditions into functional profiles, which permit the characterization of the impact of the condition studied upon various biological processes and pathways.

In collaboration with the U.S. Army Research Institute of Environmental Medicine (USARIEL), we are developing analysis of variance (ANOVA)-based algorithms that permit statistical evaluation of the contribution and significance of different factors to changes in gene expression. Unlike most off-the-shelf statistical analysis packages, which limit the number of repeated measures multivariate ANOVA analyses that can be performed at a time, our algorithms have no such restrictions. This allows us to perform time course analyses of all 30,000 genes of prevailing gene arrays within a single computation, where time is the repeated measured factor.
Management and Analysis of DNA Microarrays. The BIC is also initiating a new project involving the management and analysis of high throughput DNA microarrays. Our goal is to employ mostly open-source software to warehouse, filter, and analyze DNA microarray studies and associated animal-model physiologic data. Open-source relational database systems will be modified to accommodate specific data requirements and will be integrated with sophisticated open-source and open-development analysis packages, such as Bioconductor. This open-source strategy leverages the work of a large number of contributors world-wide and accommodates the necessity of frequent software updates for this emerging and rapidly changing technology.

Design of Biosensors for Threat Diagnosis. We are developing bioinformatics-based tools to help guide the design of gene chips (assays) to detect and identify human exposure to infectious agents (Figure 1) in collaboration with life-science colleagues from the Command’s Research Institute of Infectious Diseases. Our current approach uses existing software that searches for sequence similarity to identify fragments of a pathogen DNA or “fingerprints” that are unique to that pathogen when compared across all known sequenced genomes. Due to the large dimensionality of the search space ($10^6$–$10^9$ base-pairs/pathogen and ~140,000 organisms) this computation for a single pathogen may take several days in a serial 2.4 GHz machine. Hence, future efforts involve the scaling of these algorithms to a high-performance computing platform and the development of new algorithms that are designed from their onset to identify sequence dissimilarity (which is our interest) rather than similarity.

![Image](image.png)

*Fig 1. Bioinformatics-based software to help guide the design of diagnostic assays ("gene chips") to detect and identify human exposure to infectious agents.*

Characterization of Protein Structure and Function. Another challenging bioinformatics project involves the development of algorithms for the identification of protein markers that can be used to characterize protein function and structure. These algorithms will form the basis for methods to identify infectious threat agents that have been genetically engineered to foil conventional sequence-based detection approaches, to identify markers of molecules that render microorganisms resistant to currently available antibiotics, and to characterize protein function. This last functionality will help analyze the large percent of amino acid sequences identified by present-day microbial genome projects that have no identifiable structures, as assessed by standard methods, thus requiring prolonged and expensive laboratory efforts to characterize their functions.

In collaboration with Argonne National Laboratory, a multi-disciplinary Department of Energy laboratory, we have developed machine-learning support vector machine algorithms that combine protein sequence information, and more recently, protein 3-dimensional structure information to classify proteins based on their function. One of the key features of these algorithms is the ability to sift through various amino acid substitutions (in this study, sequences with up to 40% of amino acid substitutions were observed) to identify which site substitutions are critical in changes in protein function.

We are also initiating a new project focused on characterizing the structure and function of proteins that exhibit low sequence similarity (<25%) when compared with sequences of known structure and function in existing databases. In this project, new algorithmic search strategies will be developed and implemented with an objective of characterizing 50% of the present-day microbial genome sequences that consist of “hypothetical” proteins of unknown function, as they have insufficient identity to well-characterized proteins. A case in point of a microorganism of military relevance is the recently sequenced malaria parasite, *Plasmodium falciparum*, consisting of ~3000 regions that encode into proteins (open reading frames) of which 60% have unknown – or perhaps, unrecognizable – function.

Medical Informatics Research

The BIC’s research efforts in medical informatics involve the development of computer systems for management and analysis of physiologic data and the development of predictive and decision support algorithms for the prevention of disease and the management of nonbattle and combat casualty injuries.

Data Management and Analysis System. Our physiology analysis system (PAS) is built on a combination of proven computational platforms and is easily accessible through the web. The PAS is designed to provide a flexible solution to warehouse, manage, and mine large volumes of time-series
Fig 2. The PAS provides a web-based, flexible platform to warehouse, manage, and mine multiple studies involving large amounts of time-series physiologic data in a central location.
The innovative concept of PAS relates to its architecture where both the data and the analysis tools reside at the server while offering — through a web browser — a feature-rich, workstation-like environment that researchers require to perform sophisticated data mining. This architecture eliminates the need to download tools and data from the server to the user’s computer and avoids the often-painstaking process of installing the downloaded tools into the desktop. It allows rapid incorporation of user-provided analytical tools into PAS’s library of functions and automatically keeps track of the sequence of analysis steps performed by the user and generates derivative data only when requested, reducing the amount of data the system has to manage.

Casualty Triage Algorithms. We are mining pre-hospital trauma data to identify physiologic parameters that are diagnostic and prognostic for casualty status and to develop triage algorithms for the Warfighter Physiologic Status Monitoring (WPSM) system.7 Our pre-hospital civilian trauma data is collected during life flight helicopter transport to the University of Texas Health Science Center at Houston, a CSI partner. The data is collected with a Propaq system and associated hardware/software interface configured by the USAISR. The chief objective of this project is the identification of key physiologic parameters that are diagnostic or predictive of some clinical outcome, such as internal hemorrhage or a need for a lifesaving intervention, which will then be employed to develop triage algorithms.

A second project, in collaboration with a small business, centers on developing a statistics-based decision-support algorithm to interpret streaming physiologic data (for example, heart rate, respiratory rate, skin temperature) from a suite of wearable biosensors to remotely and automatically assess the physiologic status (life sign) of a wounded Soldier in the battlefield.8 The algorithm incorporates, as part of the decision-making process, both clinical uncertainty (for example, potential contradictory evidence, the number of available physiologic sensors, and temporal evolution of the symptoms from injury onset) and data imprecision (for example, missing, degraded, or corrupted data due to sensor faults, data transmission failures, and sensor dislodgment). The algorithm, based on Bayesian belief networks and related Decision Theory technologies, has a probabilistic foundation and provides the mathematical formalism whereby medical judgment may be expressed as the degree of belief in an outcome given a set of observations.

Physiology-Based Predictive Algorithms. In collaboration with USARIEM and in support of the WPSM, we are developing hybrid computer-based algorithms to predict the physiologic state of individual Soldiers.9 For instance, as illustrated in Figure 3, these algorithms may take as inputs a multitude of anthropometric, environmental, and on-line physiology measurements to infer, in real time, an individual’s level of fatigue or propensity to become a heat casualty. The hybrid algorithms combine mechanistic first-principles-based models, such as the macroscopic conservation of mass and energy, with data-driven algorithms in the form of artificial neural networks. This combined modeling approach yields Soldier-specific predictive models that maximize the use of prior physiology knowledge and, through information extracted from on-line measurements, complements our partial understanding of physiologic phenomena and accounts for inter-person variability.

Automatic Control of Resuscitation Fluids. In collab-

![Fig 3. Hybrid models combining first-principles-based physiologic models and data-driven models for predicting the physiologic state of an individual Soldier.](image)
oration with researchers at the Walter Reed Army Institute of Research (WRAIR), we are modeling blood pressure responses in pigs to infusion of various fluids during resuscitation after severe blood loss. Our objective is to use the models to evaluate candidate algorithms to automate resuscitation of the pigs, and ultimately humans, via control of an infusion pump. Our pressure/volume models are based on autoregressive moving average and autoregressive integrating moving average algorithms. Pressure/volume relationships during resuscitation are erratic because of numerous sources of variability, such as differences in pig-to-pig physiology, the blood pressure against which the infusion is occurring, the rate of infusion, the total interval of infusion, and the viscosity and distribution characteristics of the different infusion fluids. Because of this variability, we are investigating a range of potential infusion pump control algorithms, including proportional-integrating-differential and its variations, fuzzy logic, and self-tuning. The ultimate goal of this project is to integrate the selected controller into a computer assisted resuscitation algorithm (CARA), which is a critical component of WRAIR's Critical Care System for Trauma and Transport platform (CSTAT). The CSTAT is an enhanced litter that incorporates significant patient monitoring and, as exemplified by CARA, computer assisted patient care capabilities.

Conclusion

The BIC is a USAMRMC resource that supports bio and medical informatics activities that cut across the four major research focus areas of the Command: Military Infectious Diseases, Combat Casualty Care, Military Operational Medicine, and Medical Chemical and Biological Defense. The BIC performs inter-disciplinary research that covers a wide range of technology areas in collaboration with multiple life-science investigators throughout the Command’s laboratories. As we move forward, we plan to leverage, to the maximum extent possible, the BIC’s bioinformatics-related CSI projects with these internal activities.

Disclaimer

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U. S. Army or of the U. S. Department of Defense. "This paper has been approved for public release; distribution is unlimited."

References


AUTHORS:

†Doctor Reifman is a Department of the Army Senior Research Scientist and serves as the Director of the U.S. Army Medical Research and Materiel Command Bioinformatics Cell.

††Doctor McKenna is an IPA with the Henry M. Jackson Foundation and serves as the Deputy Director of the U.S. Army Medical Research and Materiel Command Bioinformatics Cell.
AMEDDC&S Welcomes MG George W. Weightman  
From the Commander, USAMRMC  
MG Lester Martinez-Lopez  
From the USAMRMC Command Sergeant Major  
CSM Domingo Costa  
USAMRIID: The Cornerstone for Medical Biodefense  
COL Erik A. Henchal, MS, USA/Caree L. Vander-Linden  
United States Army Medical Research Unit-Kenya  
COL Samuel Martin, MC, USA/COL Kip Hartman, MC, USA  
Research at the Command’s Bioinformatics Cell  
Jaques Reifman, PhD/Thomas McKenna, PhD  
Hu BChe: A Bioscavenger for Protection Against Organophosphate Chemical Warfare Agents  
Ashima Saxena, PhD, et al  
Five Keys to Deploying Activity-Based Costing  
Ken Whittaker  
USARIEM: Physiological Research for the Warfighter  
COL Karl E. Friedl, MS, USA/Jeffrey H. Allan  

Also in this issue...  
Dietary Supplement Intake in the Active Duty Enlisted Population  
MAJ Karen Brasfield, MS, USA