OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE

Fun GCAT (Functional Genomic and Computational Assessment of Threats)

Proposers' Day Brief – John Julias

28 June 2016

INTELLIGENCE ADVANCED RESEARCH PROJECTS ACTIVITY (IARPA)

Agenda

Time	Topic	Speaker
9:00am – 9:30 am	Logistics, Program Introduction	Dr. John Julias Program Manager
9:30am – 10:00 am	IARPA Overview	Mr. Tarek Abboushi Chief Acquisitions Officer
10:00am – 10:30am	Break	
10:30am – 11:00am	Fun GCAT Technical Overview	Dr. John Julias Program Manager
11:00am – 11:30am	Fun GCAT BAA Overview	Dr. John Julias Program Manager
11:30am-12:00pm	Doing Business with IARPA	Mrs. Katie Cole IARPA Acquisitions
12:00pm-12:30pm	Q&A Session	IARPA Program Team
12:30pm – 1:30pm	Lunch – on your own	
1:30pm – 3:00pm	Poster Session and Teaming Discussions	Attendees (**No Government**)

Disclaimer

- This Proposers' Day Conference is provided solely for information and planning purposes
- The Proposers' Day Conference does not constitute a formal solicitation for proposals or proposal abstracts
- Nothing said at Proposers' Day changes the requirements set forth in a Broad Agency Announcement (BAA)

Proposers' Day Goals

- Familiarize participants with IARPA's interest in research in DNA sequence screening, biological threat characterization, and bioinformatics
- Familiarize participants with IARPA's mission and how to do business with IARPA
- Provide answers to participants' questions
  - This is your chance to alter the course of events
- Foster discussion of synergistic capabilities among potential program participants, i.e., facilitate teaming
  - Take a chance – someone might have a missing piece of your puzzle

Important Points

- Proposers' Day slides will be posted on iarpa.gov
- Please save questions for the end, write on notecards
- Posters are available for browsing after lunch
- Government will not be present during the poster/teaming session
- Discussions with the PM are allowed until the BAA release
  - Once BAA is published, questions can only be submitted and answered in writing that will be provided via the BAA guidance
- Name/email list of Proposers' Day participants provided to the group **with permission**

Fun GCAT Program Introduction

Current Practice:
Pathogens are curated on DNA sequence similarity

- Organization of pathogens and genes based on genetic relatedness is impractical for biodefense purposes
  - No detection of novel genes and reassortments
  - Infrastructure for pathogen analysis is academic and static
  - Contributions from synthetic biology and the evolution and emergence of new pathogens complicate traditional approaches
  - New approaches and technologies are needed to be applied towards biodefense
  - Need to focus on phenotype of pathogens
  - Need better tools to screen, predict and understand the function and risk of unknown nucleic acid sequences

Limitations in Current Practices

- Databases curated on sequence with no capacity for gene function
- Redundant (hundreds of thousands to millions of accession #s for many genes)
- Identify novel sequences but cannot extend to knowledge
- Poor integration of informatics tools and modelling for unknowns
- Challenging for users to make confident assessments of the threat potential for novel genes

Current Practices in DNA Sequence Screening

DNA order screening
Voluntary: No requirements for screening
Costly: Time to follow up on potential hits costly
Weak: Screen genes > 200 bp not oligos
Guidance not policy

Bioinformatics in Biodefense
Based solely on genetic relatedness
Databases poorly organized • Emphasize genomic sequences not function • Many organisms that are harmless • Genes associated with house-keeping functions • Partial genomes- clutter the analysis • Lack metadata/context • Redundancies lead to inefficiency
No capacity or framework to understand unknowns

Fun GCAT Overview

Improve our analytical capabilities to reduce biological threats that arise either accidentally or intentionally by curating genes based on function.

We need to develop and provide effective approaches for the assessment of DNA sequences and improve current capabilities in the characterization and analysis of novel nucleic acid sequences to assess threat potential of unknowns, whether they are manmade or the products of evolution.

IARPA Introduction

Tarek Abboushi

Chief, IARPA Acquisitions

Fun GCAT Technical Overview

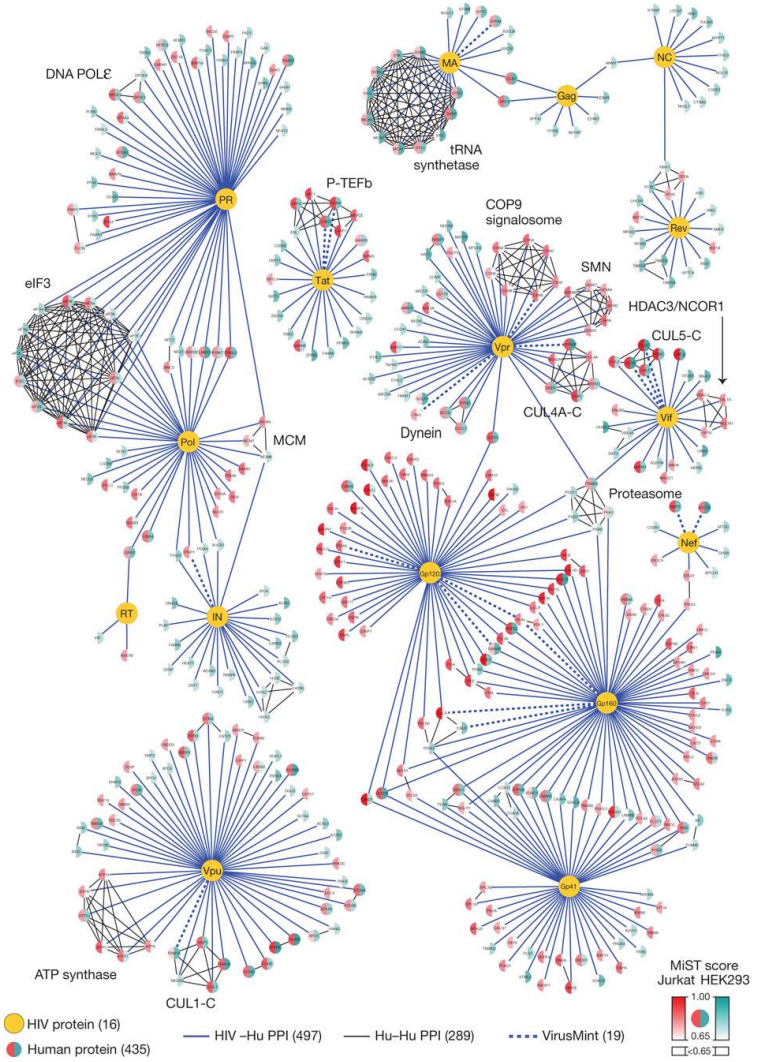
Curate Bio-Threats and Understand Novel Gene Sequences

- Problem: Pathogens are currently curated solely on DNA sequence similarity. This organization based on genetic relatedness is impractical for biodefense purposes.
  - Infrastructure for pathogen analysis is academic and static
  - New approaches and technologies need to be applied for biodefense
  - Need to reduce the risk of the intentional or accidental creation of a biological threat

- Solution:
  - Improve our analytical capabilities to reduce manmade biological threats by developing experimental approaches and bioinformatic tools to better understand the properties of novel DNA sequences. Demonstrate the benefits of curating genes based on function to provide the IC with effective approaches for risk assessment of unknown DNA sequences.

Functional Annotation

- Functional attributes to annotate
- Biochemical functions: What do the proteins do?
- Interactions/Network: What other proteins does a protein interact with and form a network
  - Cellular processes
  - Pathogen- host interactions
- Systems: How do networks relate?
- Biological functions: How pathogens interact with host and affect physiology
- Gene expression: Characteristics of promoters/enhancers in regard to how well they work and in what context
- What functionalities do novel organisms/ genes possess?
- Use data to inform models of new or emerging potential threats
- Signatures change over time

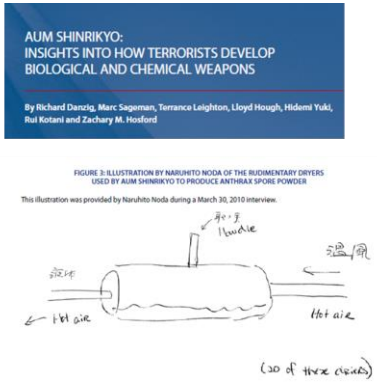


Impact of a Functional Database and Computational Modeling Tools

- Direct benefit of a functionally curated database is understanding phenotype of interest
- Enable longer-term, predictive applications from aggregation of biomolecular interactome
- Applications in forensics/attribution/prediction-prevention of future bio-crimes and bio-accidents
- Maintain and gain capability to understand rapid advances in synthetic biology, genomic engineering, and applications/implications of technologies

Why Should We Care?

- Advances in biology and engineering have transformed bio research
  - Biomedical applications: vaccines, drugs, host-pathogen relationships
  - Crops: Bioengineering and enhanced traits- global stability factors
  - Commodities: Biofuels, industrial enzymes, crops, flavor and fragrance industry
  - Micro fermentation
  - Gene synthesis and genetic modification
  - Drug delivery
- Legitimate applications with “dual-use” implications enable nefarious acts
- Expansion/democratization of biotechnology increases likelihood of accidents


AUM SHINRIKYO:
INSIGHTS INTO HOW TERRORISTS DEVELOP BIOLOGICAL AND CHEMICAL WEAPONS

By Richard Danzig, Marc Sageman, Terrance Leighton, Lloyd Hough, Hidemi Yuki, Rui Kotani and Zachary M. Hosford

FIGURE 3: ILLUSTRATION BY NARUHITO NODA OF THE RUDIMENTARY DRYERS USED BY AUM SHINRIKYO TO PRODUCE ANTHRAX SPORE POWDER

This illustration was provided by Naruhito Noda during a March 30, 2010 interview.


ENERGY & ENVIRONMENT

Modified Wheat Is Discovered in Oregon

By ANDREW POLLACK MAY 29, 2013

Unapproved genetically engineered wheat has been found growing on a farm in Oregon, federal officials said Wednesday, a development that could disrupt American exports of the grain.

The Agriculture Department said the wheat was of the type developed by Monsanto to be resistant to the herbicide Roundup, also known as glyphosate. Such wheat was field-tested in 16 states, including Oregon, from 1998 through 2005, but Monsanto dropped the project before the wheat was ever approved for commercial planting.

The department said it was not known yet whether any of the wheat got into the food supply or into grain shipments. Even if it did, officials said, it would pose no threat to health. The Food and Drug Administration reviewed the wheat and found no safety problems with it in 2004.

Still, the mere presence of the genetically modified plant could cause some countries to turn away exports of American wheat, especially if any traces of the unapproved grain were found in shipments. About $8.1 billion in American wheat was exported in 2012, representing nearly half the total $17.9 billion crop, according to U.S. Wheat Associates, which promotes American wheat abroad. About 90 percent of Oregon's wheat crop is exported.


750 sickened in Oregon restaurants as cult known as the Rajneeshees spread salmonella in town of The Dalles

Salad-bar attack by followers of Bhagwan Shree Rajneesh was the largest act of bioterrorism on U.S. soil

BY MARA BOVSUN / NEW YORK DAILY NEWS / Saturday, June 15, 2013, 7:20 PM

Share 683 Tweet

JACK SMITH/ASSOCIATED PRESS

Bhagwan Shree Rajneesh (far r.) had hundreds of followers and made millions from various schemes before being embroiled in plot to spread salmonella and take over a community in Oregon.

Bhagwan Shree Rajneesh breezed into rural Oregon in the early 1980s to spread love, enlightenment and, for those who did not believe, a little bit of salmonella.

On Sept. 17, 1984, the Wasco County health department fielded what seemed

Why Should We Care?

New and emerging technologies redefine the threat space- examples in gene synthesis

De novo synthesis of a virus

REPORTS

Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template

Jeronimo Cello, Aniko V. Paul, Eckard Wimmer*

Full-length poliovirus complementary DNA (cDNA) was synthesized by assembling oligonucleotides of plus and minus strand polarity. The synthetic poliovirus cDNA was transcribed by RNA polymerase into viral RNA, which translated and replicated in a cell-free extract, resulting in the de novo synthesis of infectious poliovirus. Experiments in tissue culture using neutralizing antibodies and CD155 receptor–specific antibodies and neurovirulence tests in *CD155* transgenic mice confirmed that the synthetic virus had biochemical and pathogenic characteristics of poliovirus. Our results show that it is possible to synthesize an infectious agent by in vitro chemical-biochemical means solely by following instructions from a written sequence.

9 AUGUST 2002 VOL 297 SCIENCE www.sciencemag.org

Potential recreation of eradicated viruses

Mouth lesions are a sign of rinderpest, which has long decimated cattle throughout the world.

INFECTIOUS DISEASE

Officials act to secure cattle-plague virus

Risk of accidental reintroduction shadows rinderpest eradication effort.

Baron's home lab contains more than 100 different rinderpest virus isolates, which he says represent "basically the history of the disease". He intends to sequence them all in the next few years — so that they can be recreated if ever needed — and then destroy them. ■

2 AUGUST 2012 | VOL 488 | NATURE | 15

Current Practices

- We are good at collecting DNA sequences from a variety of sources
  - Sequencing environmental samples
  - Clinical samples
- Technology drivers are commercial
  - Biotech
  - Pharma
  - Fermentation
- Data exist for the analysis of well-characterized pathogens
  - High impact infectious disease- HIV, HepB, HepC, Herpes, Influenza
  - Extensive traditionally annotated genetic databases
  - **Not robust or practical for biodefense threats and unknowns**

Limitations of Current Practices

Threats are evolving and dynamic

A static, threat-based list is obsolete; natural or manmade threats are dynamic

- Emerging and re-emerging threats
- Multivariable problem
- Future threats manmade/naturally occurring
- Human factors accelerating emergence in areas of great ecological diversity

OPEN ACCESS Freely available online

PLOS | PATHOGENS

Pearls

Emerging Infectious Diseases: Threats to Human Health and Global Stability

David M. Morens*, Anthony S. Fauci

National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, United States of America

Table 1. Some major factors that underlie disease emergence and reemergence [2,5].

The Microbial Agent	The Human Host	The Human Environment
Genetic adaptation and change	Human susceptibility to infection	Climate and weather
Polymicrobial diseases	Human demographics and behavior	Changing ecosystems
	International trade and travel	Economic development and land use
	Intent to harm (bioterrorism)	Technology and industry
	Occupational exposures	Poverty and social inequality
	Inappropriate use of antibiotics	Lack of public health services
		Animal populations
		War and famine
		Lack of political will

doi:10.1371/journal.ppat.1003467.t001

Limitations of Current Practices

Acquisition of virulence and novel organisms

Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax

Alex R. Hoffmaster*†, Jacques Ravel†‡, David A. Rasko†‡, Gail D. Chapman§, Michael D. Chute§, Chung K. Marston*, Barun K. De*, Claudio T. Sacchi*, Collette Fitzgerald*, Leonard W. Mayer*, Martin C. J. Maiden¶, Fergus G. Priest‖, Margaret Barker‖, Lingxia Jiang‡, Regina Z. Cer‡, Jennifer Rilstone‡, Scott N. Peterson‡, Robbin S. Weyant*, Darrell R. Galloway§, Timothy D. Read†§, Tanja Popovic*†, and Claire M. Fraser†‡**

*Epidemiologic Investigations Laboratory, Meningitis and Special Pathogens Branch, Centers for Disease Control and Prevention, 1600 Clifton Road, MS G34, Atlanta, GA 30333; †Microbial Genomics and Pathogen Functional Genomic Resource Center, Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850; ‡The Peter Medawar Building for Pathogen Research and Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3SY, United Kingdom; §School of Life Sciences, Heriot Watt University, Edinburgh EH14 4AS, United Kingdom; and ¶Biological Defense Research Directorate, Naval Medical Research Center, 503 Robert Grant Avenue, Silver Spring, MD 20910

Communicated by John B. Robbins, National Institutes of Health, Bethesda, MD, April 5, 2004 (received for review March 24, 2004)

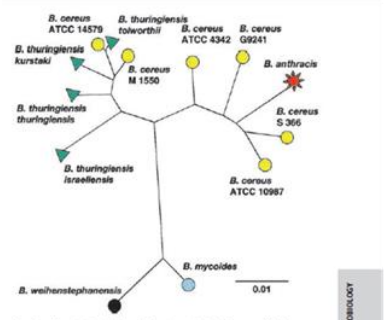


Fig. 2. Unrooted, neighbor-joining tree derived from multiple-locus sequence typing of *B. cereus* G9241 and other *Bacillus* spp. (http://pubmlst.org/bcereus).

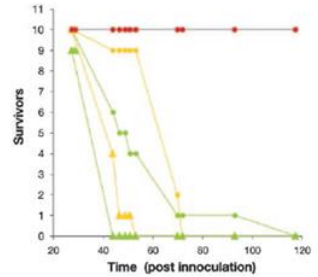


Fig. 4. Survival of A/J mice i.p. challenged with *B. cereus* G9241 (green), *B. anthracis* Sterne (yellow), and *B. cereus* ATCC10987 (red). High-spore dose (1×10^8) and low-spore dose (1×10^6) are represented by triangles and circles, respectively. The experiment was monitored for 14 days, after which the mice inoculated with *B. cereus* ATCC10987 were still alive.

SYNTHETIC BIOLOGY

Complete biosynthesis of opioids in yeast

Stephanie Galanie,[1] Kate Thodey,[2] Isis J. Trenchard,[2] Maria Filsinger Interrante,[2] Christina D. Smolke[2]*

Opioids are the primary drugs used in Western medicine for pain management and palliative care. Farming of opium poppies remains the sole source of these essential medicines, despite diverse market demands and uncertainty in crop yields due to weather, climate change, and pests. We engineered yeast to produce the selected opioid compounds thebaine and hydrocodone starting from sugar. All work was conducted in a laboratory that is permitted and secured for work with controlled substances. We combined enzyme discovery, enzyme engineering, and pathway and strain optimization to realize full opiate biosynthesis in yeast. The resulting opioid biosynthesis strains required the expression of 21 (thebaine) and 23 (hydrocodone) enzyme activities from plants, mammals, bacteria, and yeast itself. This is a proof of principle, and major hurdles remain before optimization and scale-up could be achieved. Open discussions of options for governing this technology are also needed in order to responsibly realize alternative supplies for these medically relevant compounds.

A decade ago, when we began work to realize total biosynthesis of opioids in yeast, we were motivated by the many foreseeable benefits yet mindful of potential negative impacts. Specifically, we were and remain concerned that a yeast-based opioid supply might contribute to opioid abuse (*21*, *22*). Thus, before starting this project, we sought and received permission to carry it out via Stanford University's institutional research registration with the U.S. Drug Enforcement Agency (DEA). Gaining permission required (i) background screening for researchers handling Schedule II compounds or yeast strains capable of making such compounds; (ii) detailed protocols limiting fermentation volumes and compound concentrations and including provisions for culture and product destruction and disposal immediately after experiments; (iii) increased physical containment for the strains and controlled compounds; (iv) increased laboratory security; and (v) explicit management and reporting. Taken together, these requirements reduce the chance that any compounds or strains generated in our research would directly enable individuals to abuse opioids.

4 SEPTEMBER 2015 • VOL 349 ISSUE 6252 1095

SCIENCE sciencemag.org

Limitations

- Organisms are curated in static lists
  - But organisms evolve, they are not genetically static
  - We have little or no capacity to deal with novel, emerging, or unknown sequences
  - Naturally emerging threats: Ebola, ZIKA, etc.
  - Technical advances broaden the “threat” space
    - Generation of novel or “remake” a natural threat
    - Approaches that redefine what is a biothreat
- No commercial drivers for biodefense
  - Diseases endemic to low income countries
  - No commercial viability
- Shifting landscape
  - Synthetic biology transforming lens of how to classify things

Approaches to Address Limitations

- Advances in biology converge to provide better capacity to understand novel sequences and their functions
  - Allow the curation of genes based on function
    - Models informed by ‘OMICS: proteomics, transcriptomics, etc
    - Applicable to emerging threats or enhanced and advanced agents
    - Inform predictive modelling for protein structure and function
  - Apply new computational tools and machine learning approaches to genes of interest for biodefense
    - Leverage advances in data analysis and computation for biodefense

Gene Selection Criteria

Organisms or sequences	Excluded
Model systems to validate approaches	Housekeeping functions/metabolism
Select Agents	Well known functions that pose no risk
Genes associated with pathogenesis	Gain-of-function DURC
Critical role in host response to infection	
Genes associated with synthetic pathways for substances of interest	
Contextual dependence for genes of concern	

Informatics tools

Be creative, novel and forward looking- where should we be in 3-5 years and beyond?

- Experimental approaches to enable understanding risks of genes
  - Evidence based comparison of risk
  - Measure and understand protein-protein interactions important for host response to pathogens

- Suite of tools that allow user to understand risk potential of genes and DNA sequences
  - Methods to predict structure and protein-protein interactions for unknowns
  - Advances in computational prediction of structure and functions
  - Accelerate DNA sequence data analysis and searches

Assign relative risk for unknown DNA sequences

Host-Pathogen Interactions Inform Prediction

Empirically Ranks Proteins Responsible for Phenotype

Identification of conserved regulatory proteins using
A network biology approach by measuring RNA and proteins.

Demonstrates approaches and analytics can be used
to understand conservation and regulation of gene function.

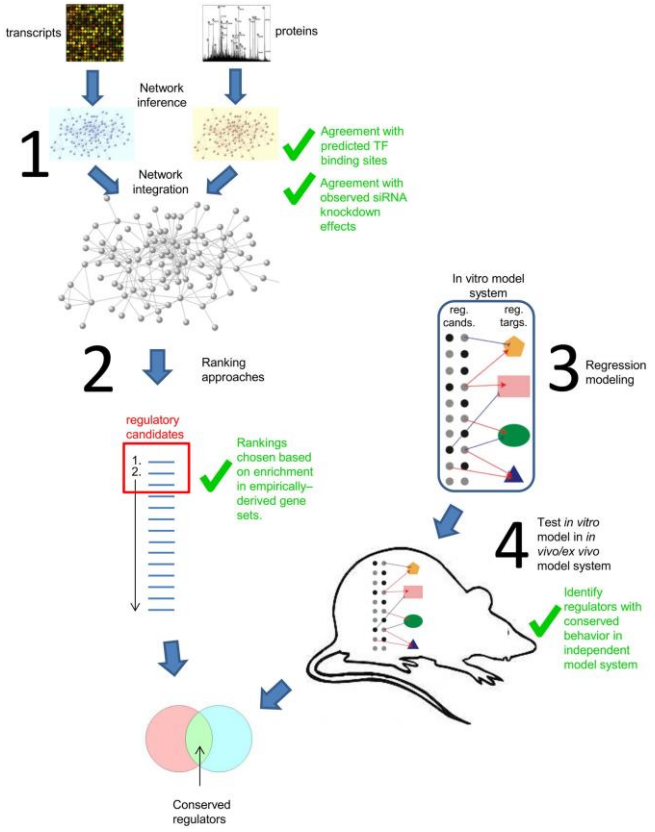
OPEN ACCESS Freely available online

PLOS ONE

A Network Integration Approach to Predict Conserved Regulators Related to Pathogenicity of Influenza and SARS-CoV Respiratory Viruses

Hugh D. Mitchell[1]*, Amie J. Eisfeld[2], Amy C. Sims[3], Jason E. McDermott[1], Melissa M. Matzke[1], Bobbi-Jo M. Webb-Robertson[1], Susan C. Tilton[1], Nicolas Tchitchek[4], Laurence Josset[4], Chengjun Li[2], Amy L. Ellis[2], Jean H. Chang[4], Robert A. Heegel[6], Maria L. Luna[6], Athena A. Schepmoes[6], Anil K. Shukla[6], Thomas O. Metz[6], Gabriele Neumann[2], Arndt G. Benecke[4,5], Richard D. Smith[6], Ralph S. Baric[3], Yoshihiro Kawaoka[2,7,8,9], Michael G. Katze[4,10], Katrina M. Waters[1]

1 Computational Sciences and Mathematics Division, Pacific Northwest National Laboratory, Richland, Washington, United States of America, **2** Department of Pathobiological Sciences, Influenza Research Institute, University of Wisconsin-Madison, Madison, Wisconsin, United States of America, **3** Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **4** Department of Microbiology, University of Washington, Seattle, Washington, United States of America, **5** Université Pierre et Marie Curie, Centre National de la Recherche Scientifique UMR7224, Paris, France, **6** Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington, United States of America, **7** Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Tokyo, Japan, **8** Department of Special Pathogens, International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Tokyo, Japan, **9** ERATO Infection-Induced Host Responses Project, Saitama, Japan, **10** Washington National Primate Research Center, University of Washington, Seattle, Washington, United States of America



Approaches for Modeling

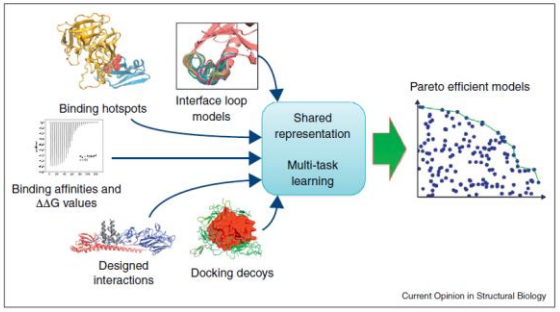
Leverage the incorporation of multi-task learning into protein modeling and the benefits of different approaches to examine energy changes in local environments.

Available online at www.sciencedirect.com
SciVerse ScienceDirect
Current Opinion in Structural Biology
ELSEVIER

Scoring functions for protein–protein interactions

Iain H Moal[1], Rocco Moretti[2], David Baker[2,3] and Juan Fernández-Recio[1]

Figure 2



Multi-task learning for simultaneously considering multiple modelling probelms. Multiple modelling tasks can be given a shared representation by using common features, such as desolvation models or pair potentials. These tasks can then be considered simultaneously using a shared learning environment, using a machine learning method capable of transferring information between tasks regarding the importance of, and interactions between, the descriptors. Models which are capable of performing well on two tasks can be selected by plotting the performances of the models against one another, as shown on the right hand side. The Pareto optimal models are those for which no other models exist with a greater performance at both tasks. Collectively, the Pareto optimal models define the Pareto frontier (green). This concept can be expanded to higher dimensions.

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL
Volume No: 5, Issue: 6, February 2013, e201302006, http://dx.doi.org/10.5936/csbj.201302006
CSBJ

SMALL WORLD NETWORK STRATEGIES FOR STUDYING PROTEIN STRUCTURES AND BINDING

Neil R. Taylor [a,*]

Abstract: Small world network concepts provide many new opportunities to investigate the complex three dimensional structures of protein molecules. This mini-review explores the published literature on using small-world network approaches to study protein structure, with emphasis on the different combinations of descriptors that have been tested, on studies involving ligand binding in protein-ligand complexes, and on protein-protein complexes. The benefits and success of small world network approaches, which change the focus from specific interactions to the local environment, even to non-local phenomenon, are described. The purpose is to show the different ways that small world network concepts have been used for building new computational models for studying protein structure and function, and for extending and improving existing modelling approaches.

MINI REVIEW ARTICLE

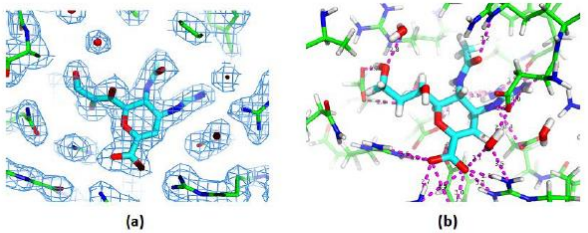


Figure 1. Different representations of 3D protein structure provide different types of understanding. (a) The electron density map for ligand and surrounding binding site residues shows a good quality, rigid model. (b) Typical computational chemistry view, showing that protein-ligand binding involves many short hydrogen bonds. (c) A network view of all favourable polar interactions in the binding site region shows how the protein-ligand hydrogen bonds are parts of highly connected local environments, which also involve numerous bound water molecules. (d) Secondary structure view, suited for structural bioinformatics analysis. Protein is Neuraminidase in complex with Zanamivir. Images generated using PyMol (www.pymol.org), PDB ids 2cml and 1nnc.

Focus on local environment not specific interactions

Current Practices in DNA Sequence Screening

DNA order screening
Voluntary: No requirements for screening
Costly: Time to follow up on potential hits costly
Weak: Screen genes > 200 bp not oligos
Guidance not policy

Bioinformatics in Biodefense
Based solely on genetic relatedness
Databases poorly organized • Emphasize genomic sequences not function • Many organisms that are harmless • Genes associated with house-keeping functions • Partial genomes- clutter the analysis • Lack metadata/context • Redundancies lead to inefficiency
No capacity or framework to understand unknowns

Current Practices: Inefficient

The International Gene Synthesis Consortium (IGSC) is comprised of seven companies that as a group provide what they estimate to be about 80% of commercial gene-length synthetic DNA. The IGSC member companies have developed their own biosecurity practices that are in line with the HHS Guidance.

Box B: Estimate of Time Spent and Costs for Bioinformatic Screening

The numbers below are estimates based on data collected from IGSC member companies but do not represent any single company's orders or costs. Green, yellow, and red sequences are described in the text.

Time for Screening

Type	% of orders of this type	Bioinformatics review time	Customer follow-up	Cost, assuming labor @ $150/hour
"Green"	95%	0.5 min	0 min	$1.25
"Yellow"	4.3%	4.5 min	79 min	$209
"Red"	0.7%	7.5 min	232 min	$598

For any given order, a company can expect to spend, on average, $14.35 on bioinformatic screening and the necessary follow-up with the customer (based on an average of costs weighted by the percentage likelihood of green, yellow, and red sequences). Because genes often cost on the order of $500–$1,000, this screening plus follow-up represents approximately 1.5-3% of total costs. As the price of gene synthesis goes down, this percentage will increase.

Of the total time spent screening orders, approximately 13% is devoted to bioinformatics-review time to determine whether the sequence is red or green or the more ambiguous yellow. The remaining 87% of time is devoted to customer follow-up for the red and yellow sequences. Close to 60% of the total screening time is devoted to customer follow-up for orders that are unlikely to be able to cause harm (i.e. yellow sequences). A more selective definition of "sequences of concern" might lower screening costs by half or more. (See discussion on a database of sequences of concern on page 10.)

5 These companies are DNA2.0, Genscript, Gen9, Integrated DNA Technologies (IDT), Origene, SGI-DNA, and Thermo Fisher Scientific. Disclaimer: the authors of this report are with the J. Craig Venter Institute, which holds stock in SGI-DNA.

Hit	Description
"Green"	No concern
"Yellow"	Similar to pathogen
"Red"	Clear link to pathogen

Deliverables

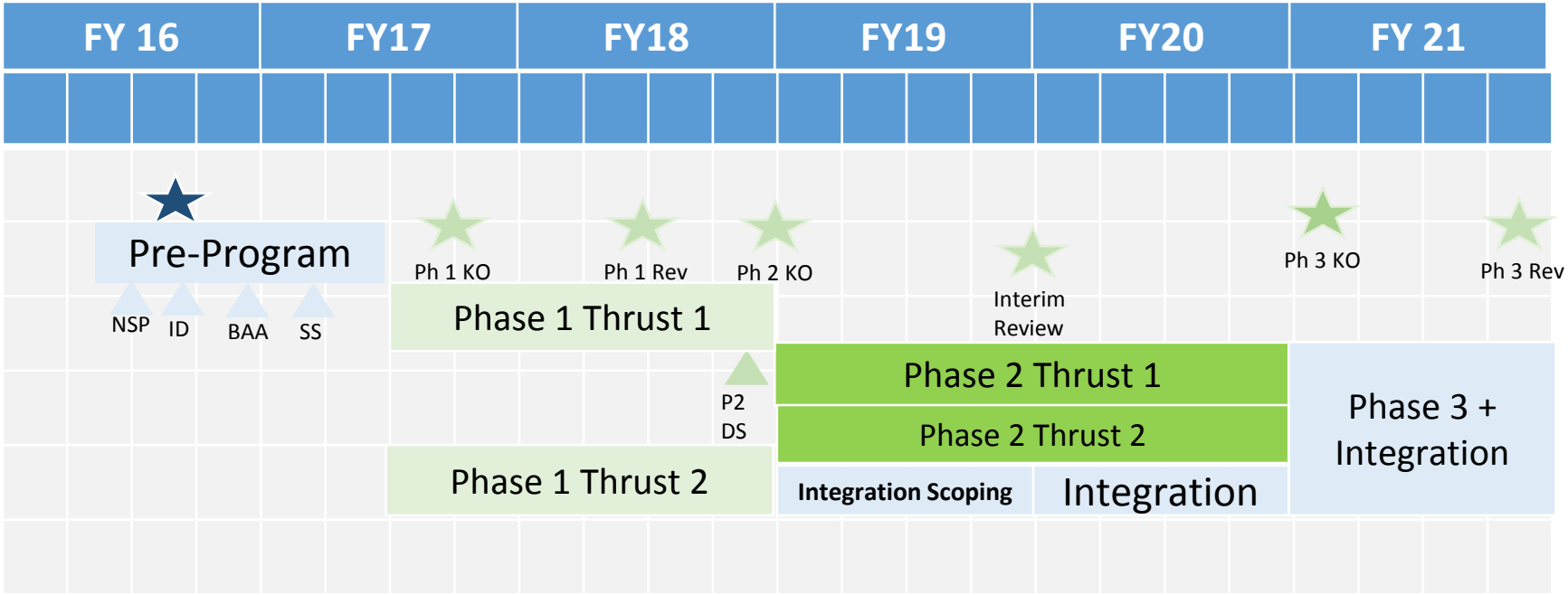
1. Functional screening methodology adaptable to any data source
   - Approaches to analyze unknowns
   - Applicable to various sources of sequence data
   - Improved modelling of structure/function of unknown genes
2. Improved capability to address unknown threats
   - Computational models that inform new/emerging threats
     - Targets: bacterial, toxin, viral, hybrid, etc.
     - Test data provided by T&E performer
3. Framework for a functionally annotated db derived from integration of data allowing DNA sequence screening and improved prediction of unknowns

Fun GCAT BAA Overview

Program Details

Item	Phase I	Phase 2	Phase 3
Structure	Competitive source selection through unclassified BAA	Possible down selection from Phase I performers; primes can re-team	Integration of data
Data	Performers choose organisms/genes	Select Agents and USG furnished data	Select Agents and USG furnished data
Security	Unclassified	Unclassified	Unclassified/FOUO
Duration	18 months	24 months	12 months
T&E	JHU-APL/LANL/PNNL/LLNL	JHU-APL/LANL/PNNL/LLNL	JHUAPL/LANL/PNNL/LLNL

Program Schedule



Milestones and Waypoints

- **Milestones** are Government-defined progress metrics that must be met by the end of each phase
- **Waypoints** are offeror-defined, task-driven intermediate steps towards a milestone
  - Depending on an offeror's specific approach, progress towards a milestone is not expected to be linear in all areas
  - Waypoints are how the offeror clearly explains to the Government the quantitative and timely progress that must be made for their overall concept to meet the end-of-phase Milestones – performance against these waypoints is reviewed throughout program
- **Technical reviews** held at months **throughout the program** will quantify progress against the waypoints & assess whether course corrections are needed for success

Gene Characterization

- Tests will be designed to characterize functional interactions accurately and rapidly
- Issue test genes with an allotted time for functional characterization
- Performance attributes:
  - Ascribing function/biological role to test genes (protein, NOT organism)
  - Scored on: Accuracy of protein changes detected, determination of gene function, length of time to generate answer
  - T&E team developing statistically robust plan

- Phase One: Derived from naturally occurring sequences only
- Phase Two: Interim evaluation sequence and Phase 2 down selection from Expansion of test to a variety of targets and contextual dependencies

Informatics and Modelling

- Dependent on types of tools being developed
- Integrate data/sequences of concern and show progress to low percentages of false positives and false negatives
  - Critical to approach from perspective of key information and contextual data
- Phase 2 will need demonstrated progress at binning threat level of genes with accuracy; low rate of false positive/negative to ensure utility in sequence screening

BAA Highlights

- Single BAA for Phases 1 and 2 and Thrusts 1 and 2
- May bid to 1st and/or 2nd thrust areas
- Will encourage use of existing data sets- provide justification for appropriateness and comprehensiveness of data
- Encourage use of cell-based and organotypic models, animal validation as needed; justify approach, use of what models, and why appropriate
- The Government anticipates that proposals submitted under this BAA will be unclassified
- Multiple awards are expected
- Foreign participants and/or individuals may participate to the extent that such participants comply with any necessary Non-Disclosure Agreements, Security Regulations, Export Control Laws and other governing statutes applicable under the circumstances
- Publications and presentations at conferences will require review by government prior to submission

Notional/Target Schedule