Active Participation in Current Faculty Research Inspires Student Achievement

Majid Masso, PhD
School of Systems Biology
George Mason University

Aspiring Scientists Summer Internship Program (ASSIP)

- Established Summer 2007 at GMU
- Students work one-on-one with GMU STEM faculty on cutting-edge research projects
- Win-Win: students learn hands-on how research is conducted; faculty receive assistance with advancing their research projects
- Additional student benefits: manuscript co-authors, conference presentations, patents

Applicants to ASSIP

- Highly competitive (< 9% out of ~1000 in 2018)
- Courses completed, GPA, volunteer/work experience, personal statements considered; interviews follow for highly-qualified candidates
- Mentor selected from 3-4 given in application
- Mainly high-school (16+), undergraduates, high school STEM teachers (professional development)
- Mainly local, national, international
- Responsible for own housing, transport, meals
- $25 application fee; otherwise free program

ASSIP Faculty Researchers

From GMU and collaborating institutes, including:

1. Center for Applied Proteomics and Molecular Medicine
2. Center for Secure Information Systems
3. Departments of Chemistry and Biochemistry; Atmospheric, Oceanic, and Earth Sciences; Computer Science; Physics and Astronomy; Psychology; Electrical and Computer Engineering; Mathematical Sciences; and Bioengineering
4. Microbiome Analysis Center
5. National Center for Biodefense and Infectious Diseases
6. School of Systems Biology
7. Virginia Serious Games Institute
8. VT Marion duPont Scott Equine Medical Center

ASSIP Curriculum

- Summer, 7–9 week session (early start option)
- First official day: orientation and required training
- Working hours: 9:00 am – 5:00 pm, Mon – Fri
- Supplementary weekly webinars and lectures: colleges, careers, networking, resumes, etc.
- Optional parallel participation in Young Inventors Club (popular)
- Concluding Poster Symposium and Reception

Projects Under My Mentorship

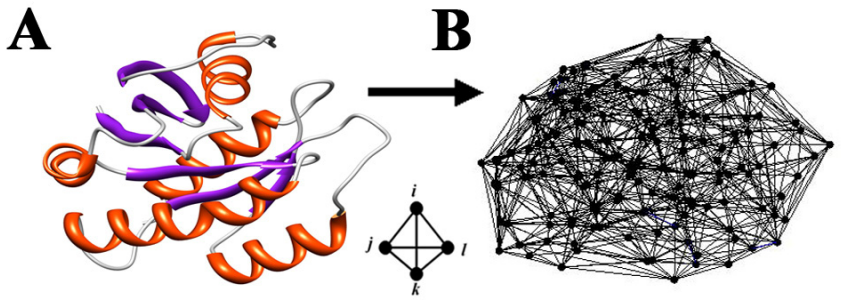
- Protein structure analysis: predicting relative changes to protein function due to mutations
- Functional changes to protein stability, activity, fitness, drug susceptibility or resistance, neutral mutation or association with a human disease, etc.
- Mutants represented as feature vectors using native protein structure and computational mutagenesis
- Machine learning used for training predictive models using mutants with known functional effects

Methods Implemented

- Steep initial learning curve: lectures and reading materials provided over the first 4-5 days
- Work integrates concepts drawn from computational geometry, probability theory, finite mathematics, statistical mechanics, biology, chemistry
- Programming skills (e.g., Perl, C++, or Python)
- Software (training on the fly): **Matlab** (Delaunay tessellation of protein structures and analysis), **Qhull** (tessellations—free), **Weka** (suite of machine learning tools—free), **Excel** (extensive use for data analysis and graphics), **UCSF Chimera** (protein structure visualization—free)

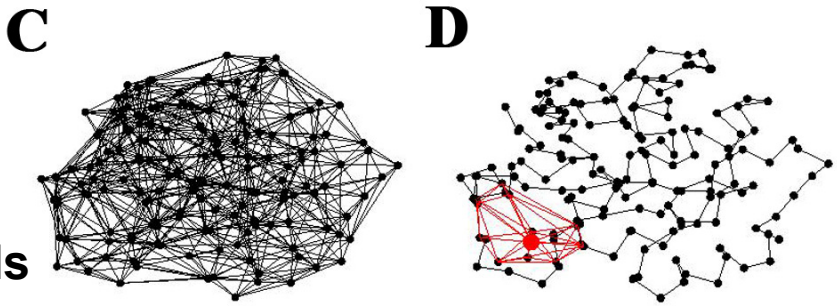
Delaunay Tessellation

UCSF Chimera software



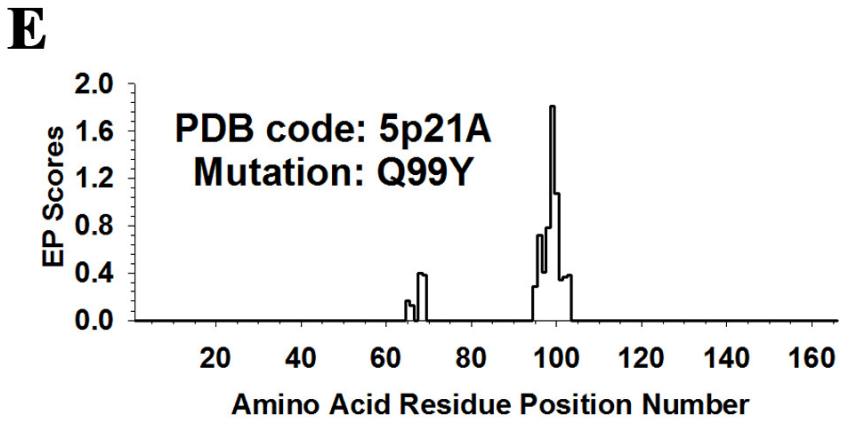
Points-amino acids
Qhull-tessellation
Matlab-visualization

Remove edges ≥ 12Å; each tetrahedral simplex identifies 4 interacting amino acids



Amino acid Q at position 99 (large red point), shared as a vertex by 18 tetrahedral simplices, and has 12 neighbors

Feature vector for mutant Q99Y (Q replaced with Y at position 99)



Counting Amino Acid Quadruplets

n **= size of amino acid alphabet = 20;** ***r*** **= size of the subsets = 4**

Repetitions Allowed?	Permutations Allowed?	Number of Quadruplets
yes	yes	$n^r = 20^4 = 160{,}000$
yes	no	$\binom{n+r-1}{r} = \binom{23}{4} = 8855$
no	yes	$\frac{n!}{(n-r)!} = \frac{20!}{16!} = 116{,}280$
no	no	$\frac{n!}{r!(n-r)!} = \binom{n}{r} = \binom{20}{4} = 4845$

only realistic choice for proteins

only realistic choice when identifying quadruplets of interacting amino acids based on the four unordered vertices of tetrahedra in a protein tessellation

Four-Body Statistical Potential

Amino Acid Quadruplet	"Pseudo-Energy" Log-likelihood s(i,j,k,l)
CCCC	3.29042538
CCCH	2.09542785
CCCS	1.96177162
CCCG	1.84022021
CCCI	1.79961166
CCCF	1.77139046
CCCT	1.76378293
CCCP	1.74840641
ACCC	1.74777711
CCCW	1.74711265
CCHH	1.70747111
CCCN	1.69741431
HHHH	1.61473339
.	.
.	.
.	.
HMNP	0.000221495
DGGY	0.000178988
DRSV	9.45855E-05
EHHV	4.979E-06
LRYY	-6.29797E-05
DGKP	-9.73563E-05
NPSS	-0.000100914
IPRW	-0.000136526
MMRT	-0.000168007
GLLP	-0.000294376
EKNT	-0.000312593
EKQR	-0.000343148
.	.
.	.
.	.
HKKW	-0.66398714
KKKP	-0.66875323
CDEQ	-0.67215257
CKKW	-0.75315166
CDDM	-0.76390474
HHKK	-0.85974
CKKR	-0.88002907
CIKR	-0.90372634
CHKW	-0.94458122
CEEE	-1.02439761
HKKM	-1.14234339

S = log (f / p)

f = observed relative frequency of occurrence in a diverse set of protein structure tessellations

p = rate expected by chance, based on relative frequencies of occurrence of the 20 types of amino acid letters in the protein set

***** p is calculated using the multinomial distribution *****

S ~ quadruplet interaction energy, by the inverted Boltzmann principle

Multinomial Reference Distribution

$n =$ number of independent trials of an experiment

$k =$ number of mutually exclusive and exhaustive outcomes for the experiment, say $A_1, A_2, \ldots, A_k$

$P(A_i) = p_i, i = 1, 2, \ldots, k$ on each trial with $\sum_{i=1}^{k} p_i = 1$

Let random variable X_i be the number of times A_i occurs in the n trials, $i = 1, 2, \ldots, k$.

If $x_1, x_2, \ldots, x_k$ are nonnegative integers such that $\sum_{i=1}^{k} x_i = n$, then the probability that A_i occurs x_i times, $i = 1, 2, \ldots, k$ is given by

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

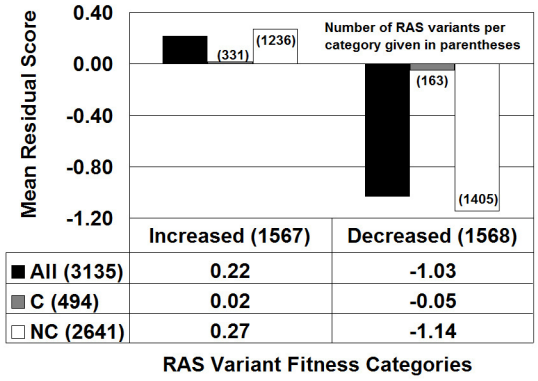
In our case, each experiment consists of selecting an amino acid ($k = 20$), and there are $n = 4$ trials..

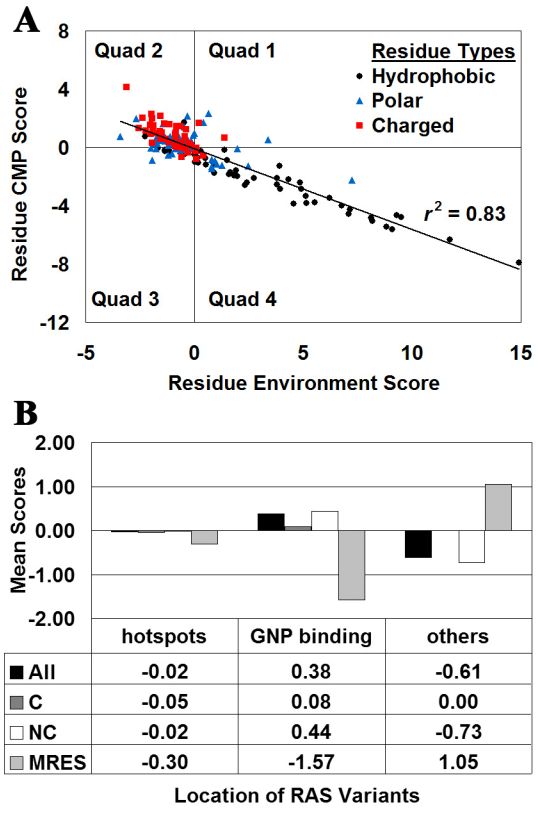
Each A_i represents a different amino acid type, where p_i is the proportion of all amino acids in the 1400 + proteins that are of type i, and x_i is the number of times that amino acid A_i occurs in the quadruplet. So,

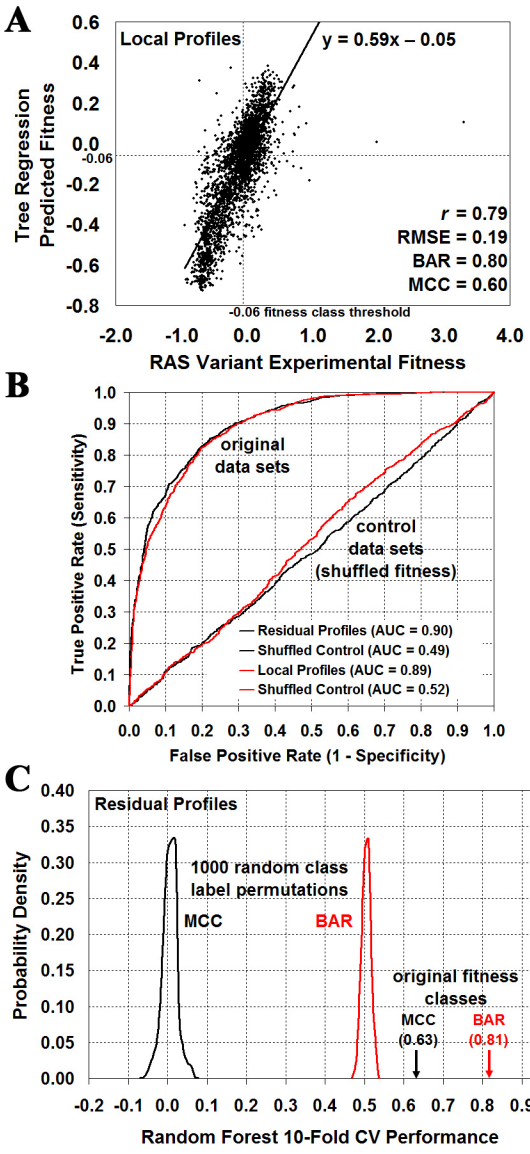
$$P(X_1 = x_1, X_2 = x_2, \ldots, X_{20} = x_{20}) = \frac{4!}{\prod_{i=1}^{20} x_i!} \prod_{i=1}^{20} p_i^{x_i}$$

is the random chance of occurence of any given quadruplet, where $\sum_{i=1}^{20} x_i = 4$.

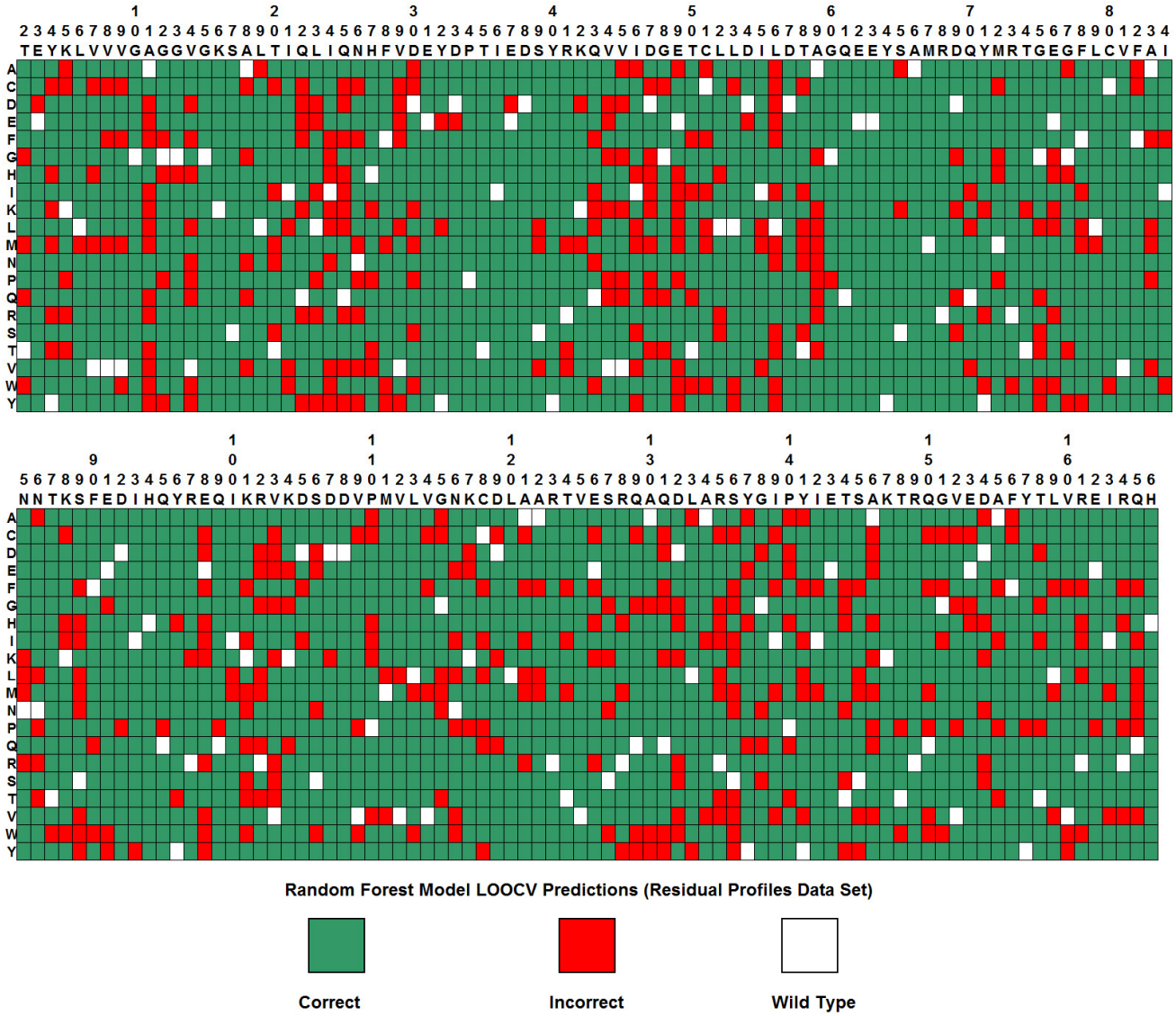
Graphical Views of Results







Graphical Views of Results



Publications

Antiviral Research 106 (2014) 5–12

Contents lists available at ScienceDirect

Antiviral Research

journal homepage: www.elsevier.com/locate/antiviral

Structure-based predictors of resistance to the HIV-1 integrase inhibitor Elvitegravir

Majid Masso *, Grace Chuang, Kate Hao, Shinar Jain, Iosif I. Vaisman

Laboratory for Structural Bioinformatics, School of Systems Biology, George Mason University, 10900 University Boulevard MS 5B3, Manassas, VA 20110, USA

ARTICLE INFO

Article history:
Received 14 January 2014
Revised 14 March 2014
Accepted 17 March 2014
Available online 25 March 2014

ABSTRACT

The enzyme integrase (IN) of human immunodeficiency virus type 1 (HIV-1) mediates integration of reverse transcribed viral DNA into the human genome, an essential step in the HIV-1 replication cycle. Elvitegravir (EVG) is an HIV-1 strand transfer inhibitor that binds IN and is the second drug in its class to be approved for clinical use in combination with other anti-HIV-1 medications. However, certain IN sequence mutational patterns have an effect on inhibitor binding, thereby altering the degree of IN mutant susceptibility to EVG. Employing a dataset of 115 translated IN sequences, each having a known

IEEE International Conference on Bioinformatics and Biomedicine

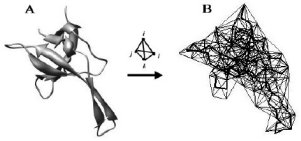
Structure Based Functional Analysis of Bacteriophage f1 Gene V Protein

Majid Masso, Ewy Mathe, Nida Parvez, Kahkeshan Hijazi, and Iosif I. Vaisman
Laboratory for Structural Bioinformatics, George Mason University, 10900 University Blvd. MS 5B3, Manassas, VA 20110, USA
{mmasso, ivaisman}@gmu.edu, mathee@mail.nih.gov, nidaparvez@hotmail.com, kahk2001@yahoo.com

Abstract

A computational mutagenesis methodology utilizing a four-body, knowledge-based, statistical contact potential is applied toward globally quantifying relative structural changes (residual scores) in bacteriophage f1 gene V protein (GVP) due to single amino acid residue substitutions. We show that these residual scores correlate well with experimentally measured relative changes in protein function caused by the mutations. For each mutant, the approach also yields local measures of environmental perturbation

model system for protein engineering experiments given its small size.



Modeling transcriptional activation changes to Gal4 variants via structure-based computational mutagenesis

Majid Masso, Nitin Rao and Purnima Pyarasani

Laboratory for Structural Bioinformatics, School of Systems Biology, George Mason University, Manassas, VA, United States of America

ABSTRACT

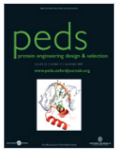
As a DNA binding transcriptional activator, Gal4 promotes the expression of genes responsible for galactose metabolism. The Gal4 protein from *Saccharomyces cerevisiae* (baker's yeast) has become a model for studying eukaryotic transcriptional activation in

CHAPTER 36
Bioinformatics Research and Applications

Computational Mutagenesis of E. coli Lac Repressor: Insight into Structure-Function Relationships and Accurate Prediction of Mutant Activity

Authors: Majid Masso · Kahkeshan Hijazi · Nida Parvez · Iosif I. Vaisman

A computational mutagenesis methodology that utilizes a four-body, knowledge-based, statistical contact potential is applied toward quantifying relative changes (



Volume 22, Issue 11
November 2009

Article Contents

Abstract
Introduction
Materials and methods
Results and discussion

Modeling the functional consequences of single residue replacements in bacteriophage f1 gene V protein

FREE

Majid Masso, Ewy Mathe, Nida Parvez, Kahkeshan Hijazi, Iosif I. Vaisman

Protein Engineering, Design and Selection, Volume 22, Issue 11, 1 November 2009, Pages 665–671, https://doi-org.mutex.gmu.edu/10.1093/protein/gzp050
Published: 18 August 2009 **Article history**

Split View | PDF | Cite | Permissions | Share

Abstract

A computational mutagenesis methodology utilizing a four-body, knowledge-based, statistical contact potential is applied toward globally quantifying relative environmental perturbations (*residual scores*) in bacteriophage f1 gene V protein (GVP) due to single amino acid substitutions. We show that residual

Where Are They Now?

| Department of Biology, Lahore University of Management Sciences (LUMS)

Computational Genomics and Systems Biology Lab

Home Research Publications People Software Contact

CO-PRINCIPAL INVESTIGATOR

Home / People /

Dr. kahkeshan Hijazi

Assistant Professor

Research Interests:

Transcriptomics

Genomics

Translational Bionformatics

Biography

Dr. Kahkeshan Hijazi received her Bachelor's degree in Bioinformatics from Mohammad Ali Jinnah University, Islamabad, Pakistan in 2006. In 2009, she was awarded the J. William Fulbright Doctoral Award from the United States Educational Foundation (USEFP), Pakistan. She received her Master's degree and a PhD in Bioinformatics from Boston University, Massachusetts, USA in 2014. Her research during her PhD was focused on developing predictors of tobacco-induced airway epithelial cell damage and the risk for having or developing tobacco-associated lung disease in humans at the Boston University Medical Center (BUMC) under the supervision of Dr. Avrum Spira. Prior to joining LUMS she served at the Research Center for Modeling and Simulation, National University of Sciences and Technology (NUST), Islamabad as Assistant Professor of Bioinformatics. Dr. Hijazi's expertise in Bioinformatics gives her great experience in the application of techniques from computer science and statistics to identify and understand patterns in the ever-more complex datasets produced by genome-wide

Enter Keywords...

CONTACT US

Principal investigator

Lab Manager

Site Administrator

Mailing address

Concluding Remarks

- ASSIP sessions are densely packed
- Students thrive at the opportunity to perform cutting-edge research
- Co-authoring manuscripts and presenting work at conferences are natural motivators
- Skills learned are put to good use by students in future endeavors
- ASSIP sessions are at least as equally satisfying for the faculty mentors!