Mission

Bring together scientists from the School of Medicine, the College of Arts & Science, and the School of Engineering who create a complex pipeline that integrates structural biology with big data for the discovery of novel antibodies that will spur the development of next-generation vaccines.

Aims

• Form pipeline between the Vanderbilt Vaccine Center (VVC) and the Center for Structural Biology (CSB) that characterizes the most promising novel antibodies in terms of the structural determinants that drive neutralization of pathogenic viruses

• Integrate big data and computational engineering with the VVC and CSB to engineer novel antibodies and vaccines

• Graduate and undergraduate student training in integrating structural biology with big data for the design of next generation vaccines
Faculty Participants

Jens Meiler (PI)
Professor, Chancellor Faculty Fellow
Departments of Chemistry, Pharmacology, and Biomedical Informatics
Center for Structural Biology, Institute for Chemical Biology
jens.meiler@vanderbilt.edu

Walter Chazin, Ph.D.
Chancellor’s Professor and Ingram Professor of Cancer Research
Department of Biochemistry and Chemistry
Director of Center for Structural Biology

Melanie Ohi, Ph.D.
Associate Professor
Department of Cell and Developmental Biology and Center for Structural Biology

Benjamin Spiller, Ph.D.
Associate Professor
Departments of Pharmacology and Pathology, Microbiology and Immunology

Kevin L Schey, Ph.D.
Professor
Department of Biochemistry
Director of Mass Spectrometry Core

James E Crowe Jr, M.D., (co-PI)
Professor and Ann Scott Carell Chair
Departments of Pediatrics and Pathology, Microbiology and Immunology
Director of Vanderbilt Vaccine Center
james.crowe@vanderbilt.edu

Ivelin Georgiev, Ph.D.
Assistant Professor
Department of Pathology, Microbiology and Immunology

Yevgeniy Vorobeychik, Ph.D.
Assistant Professor
Department of Electrical Engineering and Computer Science

James C. Slaughter, Ph.D.
Associate Professor
Department of Biostatistics

Daniel Fabbri, Ph.D.
Assistant Professor
Department of Biomedical Informatics
Overview

Integrating Structural Biology with Big Data

Naïve, previously infected, or vaccinated humans

Next generation sequencing of human antibody repertoire

Rapid structure-based in silico search for novel antibodies

Biological characterization of novel antibodies

Recombinant generation of human antibodies

Electrofusion to create antibodies

Computational design of novel antibodies

DXMS to determine epitope and paratope

X-ray, EM and computation to determine antibody/antigen complex structure
Stipends Available!

Opportunities exist for:
• Undergraduate summer research internships
• Graduate students
• Post-Doctoral fellows

Inquiries:
Dr. Jens Meiler – jens.meiler@vanderbilt.edu

Dr. James E. Crowe – james.crowe@vanderbilt.edu
Structural Biology of Poxvirus Protein and Human Antibody Interaction

- The initial focus – determining structures of neutralizing antibodies to orthopoxviruses bound to antigen by using X-ray crystallography and single particle electron microscopy.

- Computational approaches will be used to dock high resolution crystal structures of the complexed antibody and antigen into the lower resolution electron density map.

- The long term goal is to be able to computationally predict the structure of the antibody-antigen complex once the sequence of the antibody is known.

Amritraj Patra
Mentors: Melanie D. Ohi, Jens Meiler, and James E. Crowe Jr.
Human Neutralizing Antibodies for Norovirus

Goals:
- Provide a detailed understanding of the human humoral response to Norovirus infection
- Define the molecular and structural basis for inhibition of Norovirus by human antibodies.

Norovirus (NoV)
- Leading cause of sporadic and epidemic gastroenteritis in humans.
- No vaccines, therapeutics, or prophylactics available to prevent or treat infection
- Vaccine design difficult due to antigenic variation within & between NoV genogroups

We intend to characterize three-dimensionally where broadly binding & blocking human monoclonal antibodies bind to current strains of NoV.

Gabriela Alvarado
Mentors: James E. Crowe Jr., Ivelin Georgiev
Delineation of polyclonal antibody specificities in sera from donors infected by or vaccinated against viruses of biomedical significance

We are developing a novel algorithm for simulating antibody neutralization fingerprints. Preliminary analysis shows that the algorithm can successfully recapitulate known HIV-1 antibody fingerprints. Ultimately, the algorithm will be used to design and identify novel antibody fingerprints and epitopes.

- Computational neutralization fingerprinting uses large-scale antibody-antigen functional data to delineate the complex epitope specificities found in polyclonal responses to infection and vaccination.

- The goal is to develop next-generation neutralization fingerprinting technology for identifying novel, previously uncharacterized, antibody specificities in polyclonal responses from infection or vaccination.

- Application to HIV-1 and dengue

Nagarajan Raju
Mentors: Ivelin Georgiev, James E. Crowe Jr.
Conservation of dengue antibody epitopes on protein E

- Maximum antibody-interactive area for each antigen residue is plotted as a heatmap, capped at 100 Å²
- Darker color corresponds to antigen residues that are a substantial part of at least one antibody epitope

Nagarajan Raju
Mentors: Ivelin Georgiev, James E. Crowe Jr.
Computational Antibody Design

• Computational design in the combinatorial antibody and virus sequence space
• New paradigm for designing antibodies with increased breadth, high potency and high similarity to antibodies in the human repertoire
• Goal: to develop highly scalable methods for sequence-based design of antibodies for HIV that exhibit very broad recognition of diverse isolates
• Algorithms: combination of game theoretic modeling, computational structure design, scalable sequence search, and scalable bi-level combinatorial optimization

The native VRC01-GP120 complex (left) and game-theoretically designed antibody in complex (right). GP120 needs at least 7 mutations to escape binding to the designed antibody, as compared to just 1 mutation in case of VRC01.

Swetasudha Panda
Mentors: Yevgeniy Vorobeychik, Jens Meiler
Broadly Binding Antibody Design

Vaccination therapies fight infectious diseases e.g., HIV and influenza.

Goal: Design broadly binding antibodies, that bind to a high fraction of viruses, given a panel of virus sequences. With the proposed algorithm we exhibit stable antibodies with ~98% breadth compared to 60% in case of the native (in terms of Rosetta computations).

**Features for Linear SVM**
- N binding sites on each side
- M=20 possible amino acids
- N x N edges represent possible interactions
- Antibody and virus: vectors a and v of length N
- a_j and v_i = 1, if site i has amino acid j, 0 otherwise

\[ a \cdot v = \sum_i a_i v_i \]

- a binds v if binding score < \theta
- \[ \sum_i a_i v_i + \sum_j \sum_i A_i v_j + \sum_i \sum_j \sum m_n Q_m v_i v_j + I < \theta \]

\( A_i \) and \( V_j \) are coefficients for site i, amino acid j, \( Q_m \) is the coefficient matrix for interaction between sites i and j, I = intercept

\[ S(a) = \sum_i \sum_j A_i v_j + I = \text{stability score when antibody} \ a \]

Algorithm:
- Generate antibody and virus data (binding and stability scores) with 180 virus panel
- Learn linear SVM to predict binding and stability
- Evaluate using Rosetta
- Antibody design Integer program to maximize breadth and stability
- Designed antibody

Data:
- Antibody VRC23 and a panel of 180 virus sequences similar to GF120.
- Rosetta binding and stability scores generated by making 5 random mutations on the antibody and using the 180 panel.
- Binding classifier trained with 32% error, stability regression model with 0.85 correlation coefficient on test set (10 fold cross validation).

**Antibody Design Integer Program**

\[ \min \sum_i \sum_j (A_i v_j + A_m v_n) \]

subject to

\[ \sum_i (A_i v_j + A_m v_n) \leq \theta - \epsilon \]

\( \forall n \in 1, \ldots, \text{train size} \)

\[ \sum_m (A_m v_n) = 1/n \forall m \]

\[ A_m \in \{0,1\} \]

Results:
- Designed Antibody with Training Set = 30
- Predicted Breadth
- Predicted Stability
- Rosetta Evaluated Breadth

- Designed antibody binding score vs VRC23 binding score with 180 panel
- Designed antibody stability score vs VRC23 stability score with 180 panel

Swetasudha Panda
Mentors: Yevgeniy Vorobeychik, Jens Meiler
Game-Theoretic Antibody Design

Challenge: Many viruses can rapidly escape vaccination therapies by a sequence of mutations.

Goal: Design broadly binding antibodies such that the virus does not easily escape. We formulate antibody design as a designer-virus Stackelberg game, where the designer chooses an antibody to maximize the escape cost, i.e., the shortest sequence of mutations that lead the virus to escape. With the proposed method we exhibit an antibody sequence with a far higher escape cost (7) than the native (1) (in terms of Rosetta computations).

Antibody Design as a Stackelberg Game

- The bi-level optimization problem can be expressed as,

\[
\max_{x, \theta} \left\{ \sum_{i=1}^{n} c_i x_i \right\}
\]

s.t. \( C(x, \theta) \geq 0 \)

- We propose local search approaches starting with the native pair since evaluating binding for single point mutations is faster (~15 minutes) than evaluating multiple mutations (~40 minutes).

Algorithm 1: High-level algorithm for antibody design

function ADVSearch(a, \theta)
    s = initializeState(a, \theta)
    for K iterations do
        e = chooseNext(s)
        s = NextState(a, e)
        \( e^* = \arg\max E(s) \)
    return \( s^* \)

Computing Minimal Virus Escape

- Baseline approach is greedy local search algorithm (Algorithm 2)
- To speed this up, we train two linear SVM classifiers:

\[
\Psi(a, v) = \begin{cases} 
1 & \text{if } O(a, v) > M(a, v) \\
0.95 & \text{if } O(a, v) = M(a, v) \\
0 & \text{if } O(a, v) \leq M(a, v) 
\end{cases}
\]

\[
\Psi(a, v) = \begin{cases} 
1 & \text{if } O(a, v) > \theta \\
\frac{1}{2} & \text{if } O(a, v) = \theta \\
0 & \text{if } O(a, v) \leq \theta 
\end{cases}
\]

using \( \ell_2 \) loss, \( \ell_1 \) penalty and class weights inversely proportional to their frequencies in the training dataset.

- Feature vector for each binding site:
  a) binary indicator to denote if amino acid is different from the native
  b) 15 amino acid features

Results

- We evaluate on the HIV native antibody-virus pair VRC01 and HIV envelope protein GP120.
  52 and 45 binding sites respectively. \( O(a^n, v) = -49.5 \).

- Virus escape: We generate training data with 346 antibodies from a native-biased random distribution.

Comparison between baseline (A) and classifier-based greedy (B) algorithms for computing virus escape in terms of the number of evaluations (top) and computed escape time (bottom): (a) \( \theta = 0 \), 75\% of the data for training; (b) \( \theta = 0 \), 55\% of the data for training; (c) \( \theta = -15 \), 75\% of the data for training; (d) \( \theta = -15 \), 55\% of the data for training. Horizontal axes denote antibodies.

- Antibody design: Poisson regression model trained on the 346 antibodies (correlation of 0.66 between actual and predicted escape costs, 10 fold cross validation) is used to guide random search and simulated annealing search to maximize escape cost.

Example greedy search to compute escape cost. Binding scores against point mutations in each virus sequence position, relative to the sequence from previous iteration. C,P,A,W,R,D,N are the 7 major amino acid classes.

Swetasudha Panda
Mentors: Yevgeniy Vorobeychik, Jens Meiler
Structure Based Antibody and Antigen Design for Sialic Acid Binding Domains

- **Goal** - to characterize sialic acid (SIA) binding in proteins by looking at the electrochemical environment as well as the stereochemical conformation bias inherent in the SIA carbohydrate

- **Computational screening and model building done using EON and ROCS**
  - Results show high levels of homology between hemagglutinin and neuraminidase SIA binding pockets

- **Experimental testing of predicted trends to be run using ELISA and biolayer interferometry**

Clayton Anderson
Mentors: James E. Crowe Jr., Jens Meiler
Restricting the Sequence Space in Rosetta Antibody Design to Human-like Antibodies

1. Detection of sequence liabilities

Humanization, which is also referred to as reshaping, complementarity determining region (CDR)-grafting, veneering, resurfacing, specificity-determining residue (SDR)-transfer, or DelImmunization™, comprises strategies for reducing the immunogenicity of monoclonal antibodies (mAbs) from animal sources and for improving their activation of the human immune system.

A webserver was created to detect and score 13 common liabilities, which are likely to cause complications for elicitation as a vaccine (Figure above). The current experience based scoring model will be replaced by a statistical approach, the Bayes estimator, which enables harvest of the huge sequence pool of Next Generation Sequencing (NGS). For this purpose, the CroweLab antibody sequence database will be harvested to collect mutation information and frequencies of human antibodies extracted from real patients. The knowledge about in vivo mutations will allow better evaluation of possible sequence liabilities. The addition of data from ongoing studies, may be incorporated into the model using Bayesian inference.

2. Detection of structure liabilities

Residue pair couplings extracted from natural homologous sequences give insight into the evolution of proteins and encode functional and structural liabilities of a given protein family.

Incorporation of residue pair couplings has proven to be a successful strategy to enhance computational protein design.

Using actual evolutionary data may compensate for the limitations of common design techniques, attributable to a simplified molecular force field or lacking knowledge about protein functionality during the design process.

A physical model, the Ising model is commonly used to estimate residue pair couplings in sequence data gremlin.bakerlab.org. As part of my work, I will update the coupling prediction by replacing the Ising model by the Potts model.

For this purpose, I’ve begun collaborating with one of the developers of Gremlin, Sergey Ovchinnikov.

3. Guided computational antibody design

ROSETTA, a software suite for computational protein design, is capable to dock and design antibodies to specific targets. There are two major drawbacks which come with the current version of this procedure. First, it requires a significant amount of computational resources by naively sampling random sequence modifications from the huge variety of possible sequences. Second, ROSETTA is guided by structural stability, but omits functional and structural constraints. A system to incorporate knowledge on these constraints could circumvent antibody sequences which would cause immunogenicity effects or exhibit low expression rates in the model system.

To overcome these issues, sequence and structure liabilities (goal 1 and 2) can be incorporated into the ROSETTA Design process. This additional information will reduce the sequence space to be sampled dramatically and save computational resources. Furthermore, the designs are expected to be more human-like, which will reduce the post processing effort dramatically.

With the saved computational resources, design on multiple conformations (multistate design) will become even more feasible and allow multispecificity of antibodies for broadly neutralizing vaccines.

Samuel Schmitz
Mentors: Jens Meiler, James E. Crowe Jr.
Symposium on Modeling Immunity
April 27\textsuperscript{th}
1220 MRBIII

- 12:30pm lunch
  - Attendees must register by April 7\textsuperscript{th}

- 1:00-4:30pm
  - 5 confirmed speakers
    - Frank DiMaio – University of Washington
    - Sarel Fleishman – Weizmann Institute of Science
    - Steven Kleinstein – Yale Center for Medical Informatics
    - Jens Meiler – Vanderbilt University
    - Andrew Ward – The Scripps Institute
  - Coffee/tea break

- 4:30-5:30pm poster session and reception

- https://my.vanderbilt.edu/modelingimmunity/eventpage/