VANDERBILT
School of Engineering

Game Theoretic Antibody Design

Swetasudha Panda, Alexander M. Sevy, James E. Crowe Jr,
Jens Meiler and Yevgeniy Vorobeychik

Vanderbilt University
Nashville, TN, USA

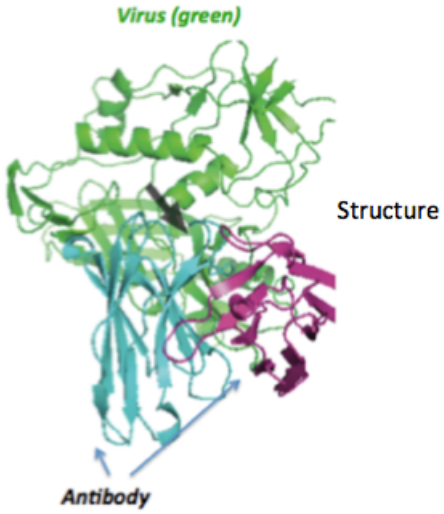
Introduction

- Infectious diseases: major threat to public health
- Vaccines: major prevention tools
  - Vaccine evokes immune response
  - Immune system generates antibodies that kill the vaccine
- Common computational approaches involve two steps:
  - Finding an antibody with the desired characteristics
  - Finding a vaccine which binds to the desired antibody, thereby invoking the target immune response

Antibodies and Viruses

- Antibodies and viruses: proteins
  - Chain of 20 possible amino acids
  - Primary structure

    ANTGRPSGQQNAAN



- These proteins bind at specific binding sites
  - By binding, an antibody neutralizes the virus
- Major challenge is virus escape through mutations e.g., HIV has extremely high mutation rates

Goals in Antibody Design

- Antibody should bind (and thereby neutralize)
  - A diverse array of target viruses
  - Potential mutations from the target viruses

Our Contributions

- A game theoretic (Stackelberg game) model for antibody and virus interaction
  - Bi-level optimization problem: intractable with integer variables

- To address intractability, we
  - Learn a linear approximation of the antibody-virus energy score
  - Formulate the optimal virus escape problem as an integer linear program (ILP)
  - Relax the integrality constraint in the ILP and take its dual to formulate the antibody design bi-level problem as a mixed-integer linear program (MILP)

Related Work

- Game-theoretic (Stackelberg Games) for Vaccine Design

  Panda, Vorobeychik, AAMAS 2015

  – Relies on local search approaches whereas our compact formulation computes the optimal solution

- Computational protein design: Multi-specificity design with respect to more than one targets

  Sevy et al., PLOS Computational Biology 2018

  – Breadth Optimization in Antibody Design (BROAD): machine learning and sequence optimization for efficient sampling in the sequence space

  – BROAD maximizes breadth over an existing virus panel whereas our approach additionally considers potential mutations

Game Theoretic Model

- We model the interaction between an antibody and a virus as a Stackelberg game
  - Antibody acts first (leader), choosing a sequence **a**
  - Virus (follower) observes the sequence **a**, and responds to it by choosing a sequence **v** so as to escape binding to **a**

Virus Objective

- Evade the antibody *a* through a series of mutations (l_0 distance) from native (initial) virus **c** (T sequences)

Over the set of feasible sequences

$$\underset{\mathbf{v}^t \in \mathcal{V}}{\text{maximize}} \sum_{t=1}^{T} \mathcal{Z}(\mathbf{a}, \mathbf{v}^t)$$

Z-score: binding and stability scores combined

$$\text{subject to } \|\mathbf{v}^t - \mathbf{c}^t\|_0 = \alpha, \forall t.$$

Number of mutations

- We constrain the feasible mutations to those that are observed in nature, using mutation frequencies from an exhaustive database

Antibody Objective

- Minimize the energy score to strengthen binding (and stability), accounting for potential mutations

$$\min_{\mathbf{a}\in\mathcal{A}} \max_{\mathbf{v}^t\in\mathcal{V}} \sum_{t=1}^{T} \mathcal{Z}(\mathbf{a}, \mathbf{v}^t)$$

$$\text{subject to } \|\mathbf{v}^t - \mathbf{c}^t\|_0 = \alpha, \forall t$$

- We restrict the antibody and virus design space to the binding sites in the native sequences
- We use Rosetta protein structure modeling software to determine the energy score

Bi-Linear Energy Score Model

- We approximate the complex black-box Rosetta energy function by a bi-linear function of the antibody and virus sequences
- The assumption is that the binding and stability of an antibody-virus complex is primarily determined by
  - The individual amino acids at each binding position of the antibody and the virus respectively, and
  - The effects of the pairwise amino acid interactions between the antibody and the virus.

Bi-Linear Energy Score Model

$$\mathcal{Z}(\mathbf{a},\mathbf{v}) = \sum_{i=1}^{N_a}\sum_{j=1}^{M} x_{ij}a_{ij} + \sum_{i=1}^{N_v}\sum_{j=1}^{M} y_{ij}v_{ij}$$

$$+ \sum_{k=1}^{N_a}\sum_{l=1}^{N_v}\sum_{u=1}^{M}\sum_{m=1}^{M} a_{ku}q_{kl}^{um}v_{lm} + I$$

- N_a and N_v : number of binding sites on **a** and **v**, M = 20
- a_{ij}, v_{ij} : binary variables, 1 if position i has amino acid j
  0 otherwise
- x_{ij}, y_{ij}, q_{kl}: associated weights for each position/amino acid on the antibody and virus side respectively, and every pairwise interacting positions
- I : intercept

ILP for Virus Objective

- Evade the antibody *a* through a series of mutations (l_0 distance) from native (initial) virus **c** (T sequences)

$$\underset{\mathbf{v}^t \in \mathcal{V}}{\text{maximize}} \sum_{t=1}^{T} \mathcal{Z}(\mathbf{a}, \mathbf{v}^t)$$

Z-score: bi-linear

$$\text{subject to } \|\mathbf{v}^t - \mathbf{c}^t\|_0 = \alpha, \forall t.$$

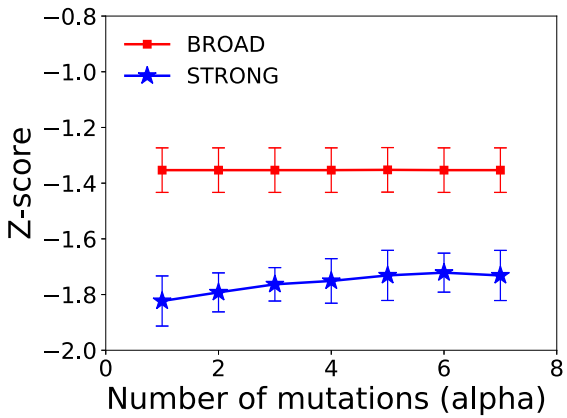
MILP for Antibody Objective

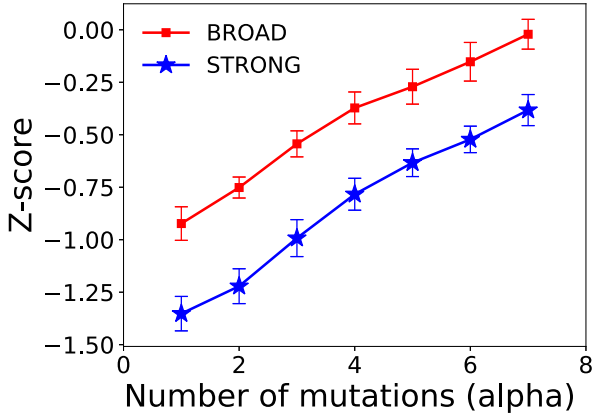
- We relax the integrality constraint of the inner (virus escape) problem
  - The LP relaxation of the virus escape ILP has integer optimal solutions
  - The primal relaxed LP is feasible and bounded; therefore, the dual is also feasible and bounded
  - By strong duality, the dual has the same solution as the primal
- We integrate the dual LP into the antibody optimization problem to formulate the MILP

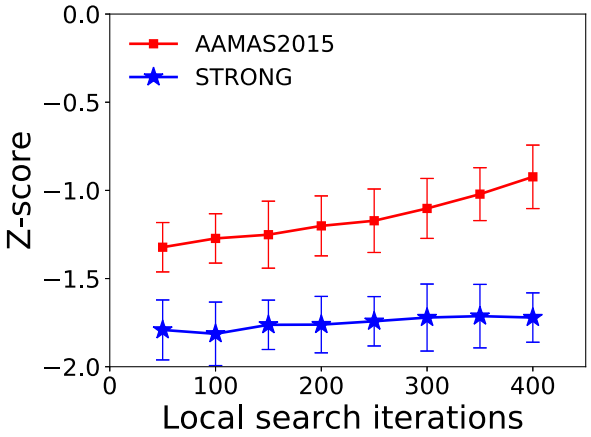
Data

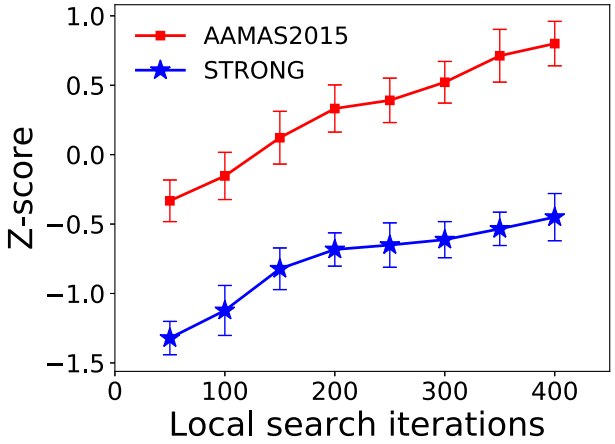
- Antibody sequences
  - Native sequence: VRC23 (anti-HIV broadly neutralizing antibody)
  - 27 binding sites
  - Variants: random substitutions at the binding sites
- Virus sequences
  - Native virus panel: 180 diverse HIV sequences
  - 32 binding sites
  - Variants: random substitutions at the binding sites
- Rosetta structure modeling and energy minimization to compute the Z-scores on 7360 antibody and virus pairs
- Mutation frequencies from the Los Alamos HIV sequence dataset
- We denote the proposed approach as STRONG and the prior approaches as BROAD and AAMAS2015

Simulation Experiments









STRONG vs. BROAD (top row) and STRONG vs. AAMAS2015 (bottom row) Z-score objective (lower is better) on the 180 native virus panel (left) and the 180 escaping virus set (right)

Conclusions

- We proposed an efficient approach for computational antibody design
  - Using a Stackelberg game model for the interaction between the antibody and the virus
- Our simulation experiments demonstrate that our approach significantly outperforms the prior approaches

Thank you