

# Stackelberg Games for Robust Vaccine Design

## (Doctoral Consortium)

Swetasudha Panda  
Supervisor: Yevgeniy Vorobeychik  
Electrical Engineering and Computer Science  
Vanderbilt University  
Nashville, TN  
swetasudha.panda@vanderbilt.edu

### ABSTRACT

Drug and vaccination therapies are important tools in the battle against infectious diseases such as HIV and influenza. However, many viruses, including HIV, can rapidly escape the therapeutic effect through a sequence of mutations. We propose to design vaccines, or, equivalently, antibody sequences that make such evasion difficult. We frame this as a bilevel combinatorial optimization problem of maximizing the escape cost, defined as the minimum number of virus mutations to evade binding an antibody. Binding strength can be evaluated by a protein modeling software, Rosetta, that serves as an oracle and computes a binding score for an input virus-antibody pair. However, score calculation for each possible such pair is intractable. We propose a three-pronged approach to address this: first, application of local search, using a native antibody sequence as leverage, second, machine learning to predict binding for antibody-virus pairs, and third, a poisson regression to predict escape costs as a function of antibody sequence assignment. We demonstrate the effectiveness of the proposed methods, and exhibit an antibody with a far higher escape cost (7) than the native (1).

### Categories and Subject Descriptors

G.1.6 [Numerical Analysis]: Optimization-Stochastic programming; I.2.1 [Artificial Intelligence]: Applications and Expert Systems-Medicine and science; I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search-Heuristic methods

### Keywords

Stackelberg games; Heuristic search; Optimization; Machine learning

## 1. INTRODUCTION

We formulate antibody design as a Stackelberg game between the vaccine designer who stimulates an antibody with particular binding characteristics (this is the binding site in the antibody sequence), and the virus subsequently responds

**Appears in:** *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015)*, Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.

Copyright © 2015, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

to the antibody by attempting to evade it (evade binding to it) through a series of local mutations. So, the “designer” chooses an antibody, and the virus responds through the shortest sequence of mutations leading to escape. In nature, evasion models natural selection where fitness criterion principally includes not binding to the antibody. Conceptually, our work follows on the steps of Stackelberg game modeling efforts in security [10]. However, the specific models developed for security are completely inadequate for our domain. There is superficial similarity to game theoretic models of vaccination decisions [2, 3, 9]. Other related work include [7], [1],[8], [11] and [5]. However, our work, to our knowledge, is the first game theoretic model of molecular-level interaction between infectious disease treatment and disease.

## 2. ANTIBODY DESIGN AS A STACKELBERG GAME

Let  $v^0$  denote the native virus, which we treat simply as a sequence (vector) of amino acids, and  $v$  and  $a$  arbitrary virus and antibody sequences, respectively. Let  $O(a, v)$  represent binding energy for the antibody-virus pair  $(a, v)$ . The “dilemma” faced by the virus is the following constrained optimization problem:

$$\min_{v \in V} \|v^0 - v\|_0 \quad (1a)$$

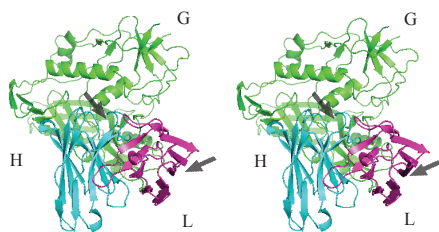
$$\text{s.t. : } O(a, v) \geq \theta, \quad (1b)$$

where  $V$  is the space of virus sequences under consideration, and  $\theta$  is a threshold on binding energy which designates escape. The  $l_0$  norm computes the number of sequence positions in  $v$  that are different from  $v^0$ . The optimization problem 1 can be viewed as a *best response* of the virus to a fixed antibody  $a$ . Let  $v(a)$  be the solution to this problem, a function of the antibody choice  $a$ . The designer’s decision problem is then

$$\max_{a \in A} \|v^0 - v(a)\|_0, \quad (2)$$

where  $A$  is the antibody design space. Sequence permutations are restricted to the binding site.

This game poses two challenges: 1) enormous search space for both the designer and the virus ( $\geq 10^{50}$  in each case), and 2) determination whether an arbitrary antibody-virus pair bind. We propose, and compare the performance of, several stochastic local search heuristics [6], using the native antibody as a “springboard”. Even for computing virus escape alone, this approach scales poorly. We use Rosetta, a premier computational protein modeling tool [4] to address



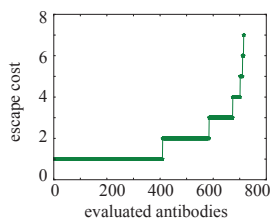
**Figure 1: The native antibody, H and L, with the native virus, G (left) and antibody with escape cost=7 (right). The arrows point at some significant differences.**

the second challenge. Rosetta, however, can be extremely time consuming even for a single evaluation. To significantly speed up the search, we use classification learning to predict whether or not an antibody-virus pair bind, limiting Rosetta evaluations only to cases in which the classifier predicts that they do not. The bi-level nature of the problem means that antibody design is still quite time consuming. To address this, we make use of Poisson regression to predict virus escape cost. In summary, we make the following contributions:

1. A bi-level optimization (Stackelberg game) model of antibody design and virus escape interaction,
2. stochastic local search techniques to determine optimal virus escape, with classifier-in-the-loop used to speed up the evaluations, and
3. stochastic local search techniques for optimal antibody design, making use of Poisson regression to predict minimal virus escape time.

### 3. EVALUATION

We used the native co-crystal structure of the antibody VRC01 complexed with the HIV envelope protein GP120. The binding site on the virus is chain G with 45 residues, while the binding site on the antibody includes chains H and L with a total of 52 residues. The visual representation of the native binding structure is shown in Figure 1 (left).



**Figure 2: Evaluated antibodies for  $\theta = 0$ , ranked by escape cost. The native antibody escape cost is 1.**

The actual set of antibodies we generated as a part of our search process, ranked in terms of evaluated escape cost (Figure 2). We found many antibodies which are much more robust to escape than the native when  $\theta = 0$ . Our best has escape cost of 7, and the resulting antibody complexed with the native virus is shown in Figure 1 (right). Visually, the

differences appear quite small, but make a significant difference in the ultimate breadth of binding, emphasizing the importance of a computational micro-level design approach.

### 4. CONCLUSION

We have, for the first time, formulated the virus evading antibodies problem as a Stackelberg game. We were able to exploit the problem structure to develop effective classification algorithms to significantly speed up the search. Finally, we exhibited an antibody that is far more robust to virus escape than the native.

### REFERENCES

- [1] M. Archetti. Evolutionarily stable anti-cancer therapies by autologous cell defection. *Evolution, Medicine, and Public Health*, pages 161–172, 2013.
- [2] C. T. Bauch and D. J. Earn. Vaccination and the theory of games. *Proceedings of the National Academy of Sciences*, 101(36):13391–13394, 2004.
- [3] C. GB, L. M, V. J, I. Y, T. D, Y. H, and G. AP. Using game theory to examine incentives in influenza vaccination behavior. *Psychological Science*, 23(9):1008–1015, 2012.
- [4] J. J. Gray, S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C. A. Rohl, and D. Baker. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of molecular biology*, 331(1):281–299, 2003.
- [5] P. Hernandez-Leal, L. Fiedler-Cameras, A. Rios-Flores, J. A. González, L. E. Sucar, and S. M. Tonantzintla. Contrasting temporal bayesian network models for analyzing hiv mutations.
- [6] H. H. Hoos and T. Stutzle. *Stochastic Local Search: Foundations & Applications*. Morgan Kaufmann, 2004.
- [7] S.-H. Huang, W. Zhou, A. Jong, and H. Qi. Game theory models for infectious diseases. In *Frontiers in the Convergence of Bioscience and Information Technologies*, pages 265–269, 2007.
- [8] R. H. Lathrop and M. J. Pazzani. Combinatorial optimization in rapidly mutating drug-resistant viruses. *Journal of Combinatorial Optimization*, 3(2-3):301–320, 1999.
- [9] J. Liu, B. F. Kochin, Y. I. Tekle, and A. P. Galvani. Epidemiological game-theory dynamics of chickenpox vaccination in the usa and israel. *Journal of the Royal Society Interface*, 9(66):68–76, 2012.
- [10] P. Paruchuri, J. P. Pearce, J. Marecki, M. Tambe, F. Ordóñez, and S. Kraus. Playing games with security: An efficient exact algorithm for Bayesian Stackelberg games. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems*, pages 895–902, 2008.
- [11] L. Richter, R. Augustin, and S. Kramer. Finding relational associations in hiv resistance mutation data. In *Inductive Logic Programming*, pages 202–208. Springer, 2010.