Genome-wide prediction of vaccine targets for human herpes simplex viruses using Vaxign reverse vaccinology

Zuoshuang “Allen” Xiang, Yongqun “Oliver” He

Unit for Laboratory Animal Medicine
Department of Microbiology and Immunology
Center for Computational Medicine and Bioinformatics
Comprehensive Cancer Center
University of Michigan Medical School
Ann Arbor, MI 48109
Human Herpes Simplex (HSV) Viruses

- Herpesviruses are a family of DNA viruses that cause diseases in humans and various animals.
- All herpesviruses are species-specific.
- Human herpesviruses (HHVs) have eight members, including HSV-1 and HSV-2, the most common infectious agents of humans.
- Infectious virions are spherical.

Virion structure: a linear, double-stranded DNA molecule densely packaged into a protein cage called **capsid**. The capsid is surrounded by an amorphous protein layer, called the **tegument**, consisting of viral proteins and viral mRNAs and a lipid bilayer membrane (envelope).
Human Herpes Simplex (HSV) Viruses

- HSV-1 and -2 are the most common infectious agents of humans.
- USA: Seroprevalence of HSV-1 and HSV-2 in adults is 68% and 21%, respectively.
- USA: ~700-2000 cases of neonatal HSV infections per year occur in the US.
- No safe and effective HSV vaccines are available.

Herpesvirus genomes available:
  - 52 herpesvirus genomes
  - 3 HHV-1 genomes: HHV-1 genome has 77 proteins
  - 12 HHV genomes

**Question:** Can we predict vaccine candidates using these genome data?
**Reverse Vaccinology (RV)**

Definition: RV is a vaccine development strategy that starts with bioinformatics analysis to find potential vaccine candidates from pathogenic genomes. These candidate genes can then be tested in normal wet lab for protective immune responses.

- Johri et al. *Nat Rev Microbiol.* 2006 Dec; 4(12):932–42.
- Rappuoli R. *Curr Opin Microbiol.* 2000 Oct;3(5):445-50.
The First RV Success: MenB

MenB: Serogroup B meningococcus

1) Genome sequence of *Neisseria meningitidis* serogroup B strain MC58 was obtained and used.
2) 570 genes predicted to code for surface-exposed or exported proteins.
3) 350 were successfully cloned to *E. coli*, expressed, and purified.
4) Mice were immunized.
5) 25 proteins induced bactericidal antibodies, which correlate with vaccine efficacy in humans.

Reference: Pizza M, *et al.* *Science.* 2000 Mar 10;287(5459):1816-20.
Reverse Vaccinology Criteria

Original Criteria:
1. Subcellular localization
   ✓ Outer membrane proteins
   ✓ Secreted proteins

New Criteria:
1. Transmembrane domains
   ✓ >2 α-helix domains → difficult to isolate
2. Adhesin probability
   ✓ Adhesin is important for pathogen invasion
3. MHC-Epitope binding
   ✓ MHC class I epitope → cell-mediated immunity
   ✓ MHC class II epitope → antibody response
4. Sequence conservation and exclusion
   ✓ Shared genes in pathogens but not in avirulent strains
5. Similarity to host proteins
   ✓ Avoid autoimmunity or immune tolerance
It is challenging to apply reverse vaccinology without a comprehensive pipeline.

To address this challenge, we developed **Vaxign**: [http://www.violinet.org/vaxign](http://www.violinet.org/vaxign)

- **Vaxign vaccine design for *Brucella***:

  **Citations:** -- Xiang Z, He Y. 2009. Vaxign: a web-based vaccine target design program for reverse vaccinology. *Procedia in Vaccinology*. Volume 1, Issue 1, Pages 23-29.

  -- He Y, Xiang Z. Bioinformatics analysis of *Brucella* vaccines and vaccine targets using VIOLIN. *Immunome Res*. 2010 Sep 27;6 Suppl 1:S5.

- **Vaxign vaccine design for uropathogenic *E. coli***:

  **Citation:** He Y, Xiang Z, Mobley HLT. Vaxign: the first web-based vaccine design program for reverse vaccinology and an application for vaccine development. *Journal of Biomedicine and Biotechnology*. Volume 2010 (2010), Article ID 297505, 15 pages.

*Many have cited Vaxign, esp. this year!*
Vaxign: Vaccine Design System

- Aim: Vaccine target prediction for reverse vaccinology

- The 1st web-based reverse vaccinology system
- Freely available: http://www.violinet.org/vaxign

Reference: He Y, Xiang Z, Mobley HLT. Vaxign: the first web-based vaccine design program for reverse vaccinology and an application for vaccine development. *Journal of Biomedicine and Biotechnology*. Volume 2010 (2010), Article ID 297505, 15 pages. [PMID: 20671958]
Two Forms of Vaxign Usage

Pre-computed data query

- Pre-computed results
- > 200 genomes in database
- Easy to query

Dynamic analysis

- Runtime execution
- Can analyze up to 500 proteins at one time
Human Herpesvirus Vaccine Design

- Downloaded from NCBI RefSeq database:
  - Three HHV-1 genomes:
    - 77 proteins in human herpesvirus 1
  - 12 HHV genomes
  - 52 herpesvirus genomes

- Vaxign pre-computation and results saved in Vaxign database

- User-friendly web interface for result query
Workflow and Result Summary

**Human herpesvirus 1 (HHV1) as seed genome**

- Absence in 9 other HHV genomes
- Absence in 40 non-human herpesviruses
  - Adhesin probability (>0.51)
  - No human protein similarity
    - Transmembrane helix (<2)
  - 7

- Conserved in 2 other HHV1 genomes
- Conserved in 9 other HHV genomes
  - 77
  - Transmembrane helix (<2)
  - Adhesin probability (>0.51)
  - No human protein similarity
  - 19
  - 17
  - 1

- Total HHV1 proteins
  - 77

- 5
Seven HSV-1 proteins having No orthologs in all 40 non-human herpesviruses

So these proteins are human-specific. Not good for animal test.
19 HSV-1 proteins conserved in human herpesviruses

| #  | Protein Accession | Protein GI | Protein Note                          | Adhesin Probability | Transmembrane helices |
|----|-------------------|------------|---------------------------------------|---------------------|-----------------------|
| 1  | NP_044603.1       | 9629382    | uracil-DNA glycosylase                | 0.262               | 0                     |
| 2  | NP_044606.1       | 9629385    | helicase-primase helicase subunit     | 0.115               | 0                     |
| 3  | NP_044655.1       | 9629434    | helicase-primase primase subunit      | 0.163               | 0                     |
| 4  | NP_044607.1       | 9629386    | capsid portal protein                 | 0.241               | 0                     |
| 5  | NP_044620.1       | 9629399    | major capsid protein                  | 0.113               | 0                     |
| 6  | NP_044627.1       | 9629406    | capsid maturation protease (UL26)     | 0.386               | 0                     |
| 7  | NP_044628.1       | 9629407    | capsid scaffold protein UL26.5        | 0.675               | 0                     |
| 8  | NP_044611.1       | 9629390    | envelope glycoprotein gM              | 0.244               | 8                     |
| 9  | NP_044629.1       | 9629408    | envelope glycoprotein gB              | 0.229               | 3                     |
| 10 | NP_044613.1       | 9629392    | deoxyribonuclease                     | 0.203               | 0                     |
| 11 | NP_044616.1       | 9629397    | DNA packaging terminase subunit 1     | 0.165               | 0                     |
| 12 | NP_044630.1       | 9629409    | DNA packaging terminase subunit 2     | 0.188               | 0                     |
| 13 | NP_044626.1       | 9629405    | DNA packaging tegument protein UL25   | 0.210               | 0                     |
| 14 | NP_044634.1       | 9629413    | DNA packaging protein UL32            | 0.185               | 0                     |
| 15 | NP_044635.1       | 9629414    | DNA packaging protein UL33            | 0.264               | 0                     |
| 16 | NP_044625.1       | 9629404    | nuclear protein UL24                  | 0.195               | 0                     |
| 17 | NP_044631.1       | 9629410    | single-stranded DNA-binding protein   | 0.168               | 0                     |
| 18 | NP_044632.1       | 9629411    | DNA polymerase catalytic subunit      | 0.101               | 0                     |
| 19 | NP_044641.1       | 9629420    | ribonucleotide reductase subunit 1    | 0.193               | 0                     |
UL26.5 for HHV-1 Vaccine Development?

- U26.5 capsid scaffold protein is important for virus capsid formation
- Has not been reported for vaccine development
- U26.5 capsid scaffold protein has adhesin-like characteristics? Why?

http://microbiologybook.org/mhunt/dna1.htm

Scaffold protein lost in mature virus

http://www.ncbi.nlm.nih.gov/pubmed/21927635
MHC Class I Epitope Prediction Example

Run MHC I epitope prediction using IEDB consensus method and compare with Vaxitope.

| Index | Epitope     | Epitope Length | MHC Allele  | P value | Matching from | Matching to | Location |
|-------|-------------|----------------|-------------|---------|---------------|------------|----------|
| 1     | GLSGHYPPPHY | 10             | HLA-A*02:01 | 0.0212  | 63            | 72         | outside  |
| 2     | HQYPGWFLFG  | 10             | HLA-A*02:01 | 0.0267  | 74            | 83         | outside  |
| 3     | DLFYSGNMGA  | 10             | HLA-A*02:01 | 0.0495  | 319           | 326        | outside  |

1 unique MHC I alleles.
Vaxign-Vaxitop MHC Class I and II Epitope Predictions

- Internally developed project
- Based on position specific scoring matrices (PSSM)
- Unique: Calculate statistical P-value for each prediction

![ROC Curve](image)

AUC=0.929 (2009 data)  
AUC=0.971 (2012 data)
Vaxign and IEDB prediction comparison on MHC Class I Epitope Prediction

### Results overlap
- Vaxitop is more conservative in predicting positive results.

| Index | Epitope       | Epitope Length | MHC allele | Matching from | Matching to | IC50 (IEDB consensus) | Vaxitop P-value |
|-------|---------------|----------------|------------|---------------|-------------|-----------------------|-----------------|
| 1     | GLSQHYPHPHV   | 10             | HLA-A*02:01| 63            | 72          | 1.06                  | 0.0212          |
| 2     | DLFYSGMMGA    | 10             | HLA-A*02:01| 319           | 329         | 4.45                  | 0.0495          |
| 3     | HQYPGVLFSG    | 10             | HLA-A*02:01| 74            | 83          | 5.4                   | 0.0267          |
| 4     | ALMGAVTSLLQ   | 10             | HLA-A*02:01| 174           | 183         | 5.4                   | >0.1            |
| 5     | FGPAFAAGSV    | 10             | HLA-A*02:01| 46            | 56          | 6.95                  | >0.1            |
| 6     | TALMGAVTSLQ   | 10             | HLA-A*02:01| 173           | 182         | 7.05                  | 0.007           |
| 7     | YLWIPASHYN    | 10             | HLA-A*02:01| 20            | 29          | 7.2                   | >0.1            |
| 8     | SAPYGMYTPV    | 10             | HLA-A*02:01| 194           | 203         | 7.4                   | 0.0641          |
| 9     | VLFSGPSPLE    | 10             | HLA-A*02:01| 79            | 88          | 8.96                  | >0.1            |
| 10    | GMYTPVAHYR    | 10             | HLA-A*02:01| 198           | 207         | 9.45                  | >0.1            |
| 11    | DTARADLFV     | 10             | HLA-A*02:01| 313           | 322         | 9.45                  | >0.1            |
| 12    | GVLFSGPSPL    | 10             | HLA-A*02:01| 78            | 87          | 9.9                   | >0.1            |

| Index | Epitope       | Epitope Length | MHC Allele | Matching from | Matching to | P-value | IC50 (IEDB consensus) |
|-------|---------------|----------------|------------|---------------|-------------|---------|-----------------------|
| 1     | GLSQHYPHPHV   | 10             | HLA-A*02:01| 63            | 72          | 0.0212  | 1.05                  |
| 2     | DLFYSGMMGA    | 10             | HLA-A*02:01| 319           | 329         | 0.0267  | 5.4                   |
| 3     | HQYPGVLFSG    | 10             | HLA-A*02:01| 74            | 83          | 0.0495  | 4.45                  |
| 4     | ALMGAVTSLLQ   | 10             | HLA-A*02:01| 174           | 183         | 0.0841  | 7.4                   |
| 5     | FGPAFAAGSV    | 10             | HLA-A*02:01| 46            | 56          | 0.0762  | >5.0                  |
| 6     | TALMGAVTSLQ   | 10             | HLA-A*02:01| 173           | 182         | 0.037   | 7.05                  |
Conclusion

• Vaxign is a specific and sensitive predictor of vaccine targets.

• It is web-based, user-friendly, and free.

• More information in next demo and hands-on training

http://www.violinet.org/vaxign/
Discussion

Challenges:
How to better rank predicted vaccine targets?
What’s unique about protective antigens?

Directions: integrated with microarray, proteomics, literature mining, 3D structure, and VIOLIN components.
Acknowledgements

He Laboratory / U of Michigan (UM)

** Zuoshuang “Allen” Xiang

** Funding: NIH-NIAID Grant 1R01AI081062