Multi-State Design of Antibody-Antigen Interactions Confers Conformational Flexibility Hypothesis

Jordan Willis
Crowe Lab
Meiler Lab
RosettaCon 2011
Antibody Structure is Constructed through 3 Genes

Heavy chain and light chain joins 3 and 2 gene segments respectively to form combinatorial diversity. Junctions form complimentary determining regions.
Antibody Diversity

Combinatorial Diversity

V<sub>H</sub> (40-50) D (25) J<sub>H</sub> (6) C<sub>μ</sub>

Junctional Diversity

(40-50) (approx. 10<sup>4</sup>)

Somatic Hypermutation

(25) (approx. 10<sup>11</sup>)

Segment Count: *Immunobiology* (Janeway)
Junctional Diversity
V-Gene codes for a majority of antibody variable region

Heavy chain variable region

Blue - V Gene
Red - N-additon
Pink - D Gene
Cyan - n addition
Orange - J Gene

Red - Framework I
Green - CDR1
Yellow - Framework II
Pink - CDR2
Cyan - Framework III
Orange - CDR3
Wheat - Framework IV
Motivation - HT sequencing reveals progenitor genes

• Crowe lab uses 454 pyro-sequencing to access antibody repertoire of healthy and viral infected patients.
• Antibody repertoire is the same for all healthy patients

<table>
<thead>
<tr>
<th></th>
<th>Reads</th>
<th>High-Quality Reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peripheral Blood</td>
<td>149896</td>
<td>132248</td>
</tr>
<tr>
<td>Bone Marrow</td>
<td>171111</td>
<td>156177</td>
</tr>
<tr>
<td>Small Intestine</td>
<td>118044</td>
<td>108279</td>
</tr>
<tr>
<td>Lung</td>
<td>198660</td>
<td>181299</td>
</tr>
<tr>
<td>Lymph Node</td>
<td>165091</td>
<td>152175</td>
</tr>
<tr>
<td>Tonsil</td>
<td>197846</td>
<td>180319</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1000648</td>
<td>910497</td>
</tr>
</tbody>
</table>

95.1% of all reads are high quality antibody sequences

Briney, Willis, Crowe *Blood* 2011
Antibody Repertoire - VH3-23 dominates
Antibody Repertoire - VH3-23 dominates

IGHV3 Super Family

N=47323

IGHV3-23 24%
IGHV3-21 11%
IGHV3-15 5%
IGHV3-11 4%
IGHV3-3 3%
IGHV3-20 1%
IGHV3-22 0%
IGHV3-21 11%
IGHV3-15 5%
IGHV3-11 4%
IGHV3-3 3%
IGHV3-20 1%
IGHV3-22 0%
IGHV3-23 24%
IGHV3-21 11%
IGHV3-15 5%
IGHV3-11 4%
IGHV3-3 3%
IGHV3-20 1%
IGHV3-22 0%
IGHV3-23 24%
IGHV3-21 11%
IGHV3-15 5%
IGHV3-11 4%
IGHV3-3 3%
IGHV3-20 1%
IGHV3-22 0%
IGHV3-23 24%
IGHV3-21 11%
IGHV3-15 5%
IGHV3-11 4%
IGHV3-3 3%
IGHV3-20 1%
IGHV3-22 0%
IGHV3-23 24%
PDB Antibody Repertoire - Recapitulates Sequencing Repertoire

Search by antibody-antigen protein complexes
There exists conformational flexibility on commonly used germline genes that accommodates a variety of antigenic structures. Using multi-state design we can test if germline sequences are optimal to bind a set of native complexes.

**Promiscuous antigen binding**

- **VH169 Germline Antibody**
  Affinity = $10^{-4} / 10^{-5}$ M

- **Intermediate Progenator**
  Affinity = $10^{-6}$ M

- **Mature Antibodies**
  Affinity = $10^{-9}$ M

**Mutations Affinity**
Multi-constraint computational design suggests that native sequences of germline antibody H3 loops are nearly optimal for conformational flexibility

Mariana Babor¹,² and Tanja Kortemme¹,²*
mAB X, Y, and Z all use the same germline, but bind separate, and structurally unique antigens. The germline sequence must be flexible to accommodate all these positions.

Conformational Flexibility Hypothesis

mpi_msd.linuxrelease - design in mutations that give a lower energy structure for each state
mAB X, Y, and Z all use the same germline, but bind separate, and structurally unique antigens. The germline sequence must be flexible to accommodate all these positions.

Conformational Flexibility Hypothesis
## VH1-69 Mature Antibody Complexes

<table>
<thead>
<tr>
<th>Entry (PDB ID)</th>
<th>Antibody Name</th>
<th>Type</th>
<th>Ligand</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1g9m</td>
<td>17b</td>
<td>FAB Kappa</td>
<td>Envelope Glycoprotein gp120 (HXBC2)</td>
<td>2.20</td>
</tr>
<tr>
<td>2b4c</td>
<td>X5</td>
<td>FAB Kappa</td>
<td>Envelope Glycoprotein gp120 (JRFL)</td>
<td>3.30</td>
</tr>
<tr>
<td>2cmr</td>
<td>D5</td>
<td>FAB Kappa</td>
<td>Gp41 Fusion Intermediate</td>
<td>2.0</td>
</tr>
<tr>
<td>2dd8</td>
<td>m396</td>
<td>FAB Lambda</td>
<td>SARS Spike</td>
<td>2.30</td>
</tr>
<tr>
<td>2xra</td>
<td>HK20</td>
<td>FAB Kappa</td>
<td>Transmembrane protein (synthetic)</td>
<td>2.30</td>
</tr>
<tr>
<td>2xtj</td>
<td>ID05</td>
<td>FAB Kappa</td>
<td>Proprotein convertase substilin</td>
<td>2.70</td>
</tr>
<tr>
<td>3fku</td>
<td>F10</td>
<td>ScFv Kappa</td>
<td>Hemmaglutanin</td>
<td>3.20</td>
</tr>
<tr>
<td>3gbn</td>
<td>CR6261</td>
<td>FAB Lambda</td>
<td>Hemmaglutanin Peptide</td>
<td>2.20</td>
</tr>
<tr>
<td>3ma9</td>
<td>8066</td>
<td>FAB Lambda</td>
<td>Transmembrane Glycoprotein</td>
<td>2.05</td>
</tr>
<tr>
<td>3mac</td>
<td>8062</td>
<td>FAB Lambda</td>
<td>Transmembrane Glycoprotein</td>
<td>2.50</td>
</tr>
<tr>
<td>3nps</td>
<td>S4</td>
<td>FAB Kappa</td>
<td>Suppressor of tumorigenicity protein</td>
<td>1.50</td>
</tr>
<tr>
<td>3p30</td>
<td>1281</td>
<td>FAB Lambda</td>
<td>Gp41 Fusion Intermediate</td>
<td>3.30</td>
</tr>
</tbody>
</table>

12 candidate test complexes using VH1-69
VH1-69 Mature Antibody Complexes

Divergent from germline

Conservation

Quality Consensus

Wednesday, August 24, 11
VH1-69 Multi-State Design

VH1-69 MSD (11 States)

19/28 recovered to VH1-69
Antibody 17b designs towards germline in MSD with correctly designed amino acids shown in dark blue. Incorrect designed are shown in orange.
VH1-69 Single-State Design (3GBN)

16/28 recovered to VH1-69
24/28 Native
## VH1-69 MSD/SSD Design

<table>
<thead>
<tr>
<th>Design</th>
<th>Percentage recovered to native</th>
<th>Percentage to VH1-69</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSD of 11 VH1-69 States</td>
<td>-</td>
<td>68</td>
</tr>
<tr>
<td>1g9m</td>
<td>68</td>
<td>36</td>
</tr>
<tr>
<td>2cmr</td>
<td>71</td>
<td>57</td>
</tr>
<tr>
<td>2dd8</td>
<td>71</td>
<td>61</td>
</tr>
<tr>
<td><strong>2xra</strong></td>
<td><strong>79</strong></td>
<td><strong>35</strong></td>
</tr>
<tr>
<td>2xtj</td>
<td>64</td>
<td>54</td>
</tr>
<tr>
<td>3fku</td>
<td>57</td>
<td>36</td>
</tr>
<tr>
<td>3gbn</td>
<td>85</td>
<td>50</td>
</tr>
<tr>
<td>3ma9</td>
<td>50</td>
<td>42</td>
</tr>
<tr>
<td>3mac</td>
<td>64</td>
<td>39</td>
</tr>
<tr>
<td>3nps</td>
<td>71</td>
<td>64</td>
</tr>
<tr>
<td>3p30</td>
<td>46</td>
<td>39</td>
</tr>
</tbody>
</table>
VH1-69 MSD/SSD Design

VH1-69 Sequence Recovery

- Native Sequence Recovery Average = 66%
- VH1-69 Recovery Average = 46%
## VH3-23 Mature Antibody Complexes

<table>
<thead>
<tr>
<th>Entry (PDB ID)</th>
<th>Antibody Name</th>
<th>Type</th>
<th>Ligand</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1s78</td>
<td>Pertuzumab</td>
<td>FAB Kappa</td>
<td>ErbB-2</td>
<td>3.25</td>
</tr>
<tr>
<td>2fjg</td>
<td>G6</td>
<td>FAB Kappa</td>
<td>Vascular endothelial growth factor 1</td>
<td>2.80</td>
</tr>
<tr>
<td>2qqn</td>
<td>Semaphorin Blocking</td>
<td>FAB Lamda</td>
<td>Neurophilin-1</td>
<td>2.20</td>
</tr>
<tr>
<td>2r56</td>
<td>IgE Fab Fragment</td>
<td>FAB Kappa</td>
<td>Beta-lactoglobulin allergen</td>
<td>2.80</td>
</tr>
<tr>
<td>2vxs</td>
<td>Unnamed</td>
<td>FAB Lamda</td>
<td>Interleukin-17A</td>
<td>2.63</td>
</tr>
<tr>
<td>2vyr</td>
<td>Unnamed Single VH chain</td>
<td>Single Chain</td>
<td>MDM4 Protein</td>
<td>2.00</td>
</tr>
<tr>
<td>3bn9</td>
<td>E2</td>
<td>FAB Kappa</td>
<td>Suppressor of tumorigenicity protein 14</td>
<td>2.17</td>
</tr>
<tr>
<td>3dvn</td>
<td>Apu2.16</td>
<td>FAB Kappa</td>
<td>Ubiquitin</td>
<td>2.70</td>
</tr>
<tr>
<td>3kr3</td>
<td>DX-2647</td>
<td>FAB Kappa</td>
<td>Insulin-like growth factor II</td>
<td>2.20</td>
</tr>
</tbody>
</table>

9 candidate test complexes using VH3-23
VH3-23 Mature Antibody Complexes

9 candidate test complexes using VH3-23
VH3-23 Multi-State Design

9 States - VH3-23 fixed backbone

16/27

17/33 Amino Acids Recovered
VH3-23 Multi-State Design

Orange - Correct, Red - Incorrect
VH3-23 Single-State Design

2FJG SSD

11/34 recovered to VH3-23
22/34 recovered to Native
11 more recovered to single state

Wednesday, August 24, 11
<table>
<thead>
<tr>
<th>Design</th>
<th>Percentage recovered to native</th>
<th>Percentage recovered to VH3-23</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSD of VH3-23 States</td>
<td>-</td>
<td>60</td>
</tr>
<tr>
<td>1S78</td>
<td>46</td>
<td>28</td>
</tr>
<tr>
<td>2FJG</td>
<td>64</td>
<td>32</td>
</tr>
<tr>
<td>2QQN</td>
<td>47</td>
<td>44</td>
</tr>
<tr>
<td>2R56</td>
<td>56</td>
<td>47</td>
</tr>
<tr>
<td>2VXS</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>2VYR</td>
<td>47</td>
<td>35</td>
</tr>
<tr>
<td>3DVN</td>
<td>26</td>
<td>20</td>
</tr>
<tr>
<td>3BN9</td>
<td>50</td>
<td>38</td>
</tr>
<tr>
<td>3KR3</td>
<td>47</td>
<td>32</td>
</tr>
</tbody>
</table>
VH3-23 MSD/SSD Design

VH3-23 Sequence Recovery

Native Sequence Recovery Average = 48%
VH1-69 Recovery Average = 36%
# VH5-51 Mature Antibody Complexes

<table>
<thead>
<tr>
<th>Entry (PDB ID)</th>
<th>Antibody Name</th>
<th>Antibody Description</th>
<th>Ligand</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2b1a</td>
<td>2219</td>
<td>FAB Lamda</td>
<td>UG1033 Peptide</td>
<td>2.35</td>
</tr>
<tr>
<td>2xwt</td>
<td>K1-70</td>
<td>FAB Lamda</td>
<td>TSH-R</td>
<td>1.90</td>
</tr>
<tr>
<td>3hmx</td>
<td>ustekinumab Fab</td>
<td>FAB Lamda</td>
<td>IL-12</td>
<td>3.00</td>
</tr>
<tr>
<td>2dd8</td>
<td>m396</td>
<td>FAB Lambda</td>
<td>SARS Spike</td>
<td>2.30</td>
</tr>
</tbody>
</table>

4 candidate test complexes using VH5-51
VH5-51 Mature Antibody Complexes
VH5-51 Multi-State Design

VH5-51 MSD

11/17 recovered to VH5-51
VH5-51 Single-State Design

2B1A MSD

6/17 recovered to VH5-51
14/17 Native

VH5-51 2B1A MSD

G F T S Y M I T Q I K S T A Y A M
T F S D Y M F S E M R N T A H P L

Wednesday, August 24, 11
<table>
<thead>
<tr>
<th>Design</th>
<th>Percentage recovered to native</th>
<th>Percentage to VH5-51</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSD of VH5-51 States</td>
<td>-</td>
<td>65</td>
</tr>
<tr>
<td>2b1a</td>
<td>82</td>
<td>35</td>
</tr>
<tr>
<td>2xwt</td>
<td>82</td>
<td>47</td>
</tr>
<tr>
<td>3hmx</td>
<td>82</td>
<td>70</td>
</tr>
</tbody>
</table>
VH5-51 MSD/SSD Design

VH5-51 Sequence Recovery

Native Sequence Recovery Average = 82%
VH1-69 Recovery Average = 50.6%
Conclusions

• Multi-state design recovers sequences closer to germline progenitor.

• Single state design recovers sequences closer to native (mature) antibody sequences, showing an *in silico* maturation.

• Germline sequences are optimally flexible in frequently used germline genes to accommodate binding of many antigens.
Future Directions

• Combine states of frequent and infrequently used germline genes to see which sequences are recovered.

• Full quantitative workup to find frequently used amino acids (PSSM)

• Iterative relax and MSD to accommodate clashing rotamers and improve sequence recovery.

• Apply MSD to HIV antibodies to bind a diverse panel of antigens
Acknowledgments.

Crowe Lab
• Natalie Thornburg PhD
• Bryan Briney
• Gopal Sapparapu
• Mohammed Aiyegbo
• Frances House, MS
• Fyza Shaikh
• John Bates, PhD
• Mark Hicar, MD PhD
• Lidiya Pundor
• Valentine Chukwuma
• Scott Smith MD PhD
• Chelsey Huffman

Meiler Lab
• Will Lowe, PhD
• Brian Weiner PhD
• Carrie Fortenberry MS
• Jeff Mendenhall
• Caitlin Heuberger
• Mert Karakas
• Nils Woetzal
• Mariusz Butkiewicz
• Steffen Lindert
• Julia Koehler
• Sten Heinze

Meiler Lab
• Elizabeth Dong
• Nathan Alexander
• Gordon Lemmon
• David Nannemann
• Steven Combs
• Sam DeLuca
• Stephanie Hirst
• Gregory Sliwoski
• Britney Allison

Spearman Lab, Emory
• Paul Spearman MD
• Xuemin Chen

Kuhlman Lab
• Colin Smith
• Mariana Babor
• Tanja Kortemme

Kuhlman Lab
• Brian Kuhlman
• Andrew Leaver-Fay

Funding
• NIH U01 - Clonal Analysis of the Human B-Cell Response
• NIH U01 - Broadly Neutralizing Monoclonal Antibodies to HIV-1
• HIV Training Grant - M&IM (Chris Aiken)
Antibody Diversity

Pejchal et al., PNAS 2010
Tian et al. reported on healthy and diseased repertoire using Sanger sequencing.

Gene Usage is Driven by Structure

Healthy Donor

RSV Infected

Tian et al Immunology 2007