Five Computational Developability Guidelines for Therapeutic Antibody Profiling

Matthew Raybould
Oxford Protein Informatics Group
University of Oxford

Next-Generation Investigator Session, HOS2021
Common Antibody Developability Issues

- Many different *in vitro* assays to test for each of these issues
- However, the time/quantity of monoclonal antibody (mAb) needed to experimentally test for each of these is often prohibitive in early-stage development
- Therefore, desire to generate *in silico* assays that can rapidly filter out mAb drug candidates with poor developability

**in silico developability assessment tools (2018)**

1. Various algorithms for “humanness” assessment via comparison to natural antibody sequences
2. Statistically-fit predictors of *in vitro* assay values (e.g. CamSol, Developability Index, FvCSP) or sites of post-translational modification

No publicly-available method that captured general developability
The Therapeutic Antibody Profiler:
A structure-based, *in silico* method for rapidly detecting mAbs with poor developability

**Assumptions**
- Many instances of poor developability are caused by the chemical properties of a region of the antibody surface.
- The most variable region between antibodies is the Fv region, so we analyse this region alone.
- The best way to measure Fv surface properties is *via* a structural representation.
- A set of these properties may offer some predictive power to identify more “drug-like” antibodies, *cf.* Lipinski rules.
- We assume that therapeutics that have reached Phase-II of clinical trials have acceptable developability.

**Requirements**
- We must be able to identify poor developability mAbs in a high-throughput manner.
- This necessitates using homology models over *ab initio* models or crystal structures.
Five properties:

1. CDRH3 or Total CDR length [aggregation, flexibility, topology]
2. Patches of Surface Hydrophobicity (PSH) across the CDR Vicinity [aggregation, viscosity, polyspecificity]
3. Patches of Surface Positive Charge (PPC) across the CDR Vicinity [poor expression, aggregation, viscosity, polyspecificity]
4. Patches of Surface Negative Charge (PNC) across the CDR Vicinity [poor expression, aggregation, viscosity, polyspecificity]
5. Structural Fv Charge Symmetry Parameter [aggregation, viscosity]

Datasets:

- 137 Post-Phase I Therapeutic Models
- 14k Representative Human Antibody Models
- 2 Datasets of MedImmune Developability Failures

Sets the acceptable bounds of the five properties Provides a “natural antibody comparison” Used to validate that we can selectively highlight mAbs with developability issues

---

Comparisons: Therapeutics vs. Human Antibodies

- Therapeutics tend to have shorter CDRH3s and smaller patches of surface hydrophobicity than human antibodies
Comparisons: Therapeutics vs. Human Antibodies

Patches of Surface Positive Charge (PPC)

Patches of Surface Negative Charge (PNC)

\[ \sum_{R1 \neq R2} \left| Q(R1) \right| \left| Q(R2) \right| / \gamma_{12}^2 \]

Blue: Therapeutic Antibody Models
Red: Human Antibody Models

- Therapeutics and human Abs have similar sizes of positive charge and negative charge patches
Comparisons: Therapeutics vs. Human Antibodies

Structural Fv Charge Symmetry Parameter (SFvCSP)

\[
\begin{bmatrix}
\sum_{R_H} Q(R_H) \\
\sum_{R_L} Q(R_L)
\end{bmatrix}
\]

Blue: Therapeutic Antibody Models
Red: Human Antibody Models

- Both therapeutic and human antibodies have an aversion to strongly oppositely-charged VH and VL chains
Validation

- Found a further 105 post-Phase I therapeutic sequences, as “developable antibodies”
- Only 8/105 were assigned by TAP to have a property outside the existing distributions. Most (except PPC) were minorly adjusted:

<table>
<thead>
<tr>
<th>Property</th>
<th>Red Threshold (137 Phase-II+ therapeutics)</th>
<th>Red Threshold (242 Phase-II+ therapeutics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total CDR Length (Lower)</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>Total CDR Length (Upper)</td>
<td>59</td>
<td>60</td>
</tr>
<tr>
<td>PSH (Lower)</td>
<td>85.64</td>
<td>83.34</td>
</tr>
<tr>
<td>PSH (Upper)</td>
<td>168.30</td>
<td>173.85</td>
</tr>
<tr>
<td>PPC</td>
<td>1.51</td>
<td>3.16</td>
</tr>
<tr>
<td>PNC</td>
<td>3.50</td>
<td>3.50</td>
</tr>
<tr>
<td>SFvCSP</td>
<td>-19.50</td>
<td>-20.40</td>
</tr>
</tbody>
</table>
**Validation**

**M-1912** aggregated uncontrollably during development, and exhibited extremely high values in our CDR Vicinity PSH metric. **M-1912STT** resolved the issue.

**A001** had prohibitively poor expression levels, and exhibited extremely high values in our CDR Vicinity PNC metric. **A-DDEN** fixed the issue (backbone engineering).
TAP Developability Guidelines

Values based on 242 clinical-stage therapeutic antibodies as of Feb’ 2019

<table>
<thead>
<tr>
<th>Metric</th>
<th>(Bottom 5%/Top 5%)</th>
<th>(Below/Above)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total CDR Length</td>
<td>Amber Flag Region</td>
<td>Red Flag Region</td>
</tr>
<tr>
<td></td>
<td>39 ≤ L ≤ 43</td>
<td>L &lt; 39</td>
</tr>
<tr>
<td></td>
<td>54 ≤ L ≤ 60</td>
<td>L &gt; 60</td>
</tr>
<tr>
<td>PSH, CDR Vicinity</td>
<td>83.84 ≤ PSH ≤ 100.71</td>
<td>PSH &lt; 83.84</td>
</tr>
<tr>
<td></td>
<td>156.200 ≤ PSH ≤ 173.850</td>
<td>PSH &gt; 173.850</td>
</tr>
<tr>
<td>PPC, CDR Vicinity</td>
<td>1.25 ≤ PPC ≤ 3.16</td>
<td>PPC &gt; 3.16</td>
</tr>
<tr>
<td>PNC, CDR Vicinity</td>
<td>1.84 ≤ PNC ≤ 3.50</td>
<td>PNC &gt; 3.50</td>
</tr>
<tr>
<td>SFvCSP</td>
<td>-20.40 ≤ SFvCSP ≤ -6.30</td>
<td>SFvCSP &lt; -20.40</td>
</tr>
</tbody>
</table>

NB: Metric values for therapeutics can change as model quality improves

These metrics could be rapidly calculated:

- During early-stage discovery
- During in silico affinity maturation

to help select mAbs more amenable to therapeutic development
Notes

• The TAP thresholds are now set by c. 400 CSTs in Phase-II+ development. We actively track these in Thera-SAbDab (http://opig.stats.ox.ac.uk/webapps/therasabdab). Thresholds have proven robust to the addition of more data.

• Typical runtime for TAP is < 30s/antibody on a single core (if all loops are homology-modellable)

• The TAP metrics were chosen to be developability-linked and interpretable. With sufficient “negative” data, they could be more systematically derived. As could the amber/red threshold percentile values

• The TAP metrics are guidelines, not strict rules. They could change over time with advances in process development

• These principles could be extended to other classes of protein therapeutics

The Therapeutic Antibody Profiler is described in our paper in PNAS¹

Software Availability

- Free OPIG Webserver
  (http://www.opig.stats.ox.ac.uk/webapps/tap)

If data is IP-sensitive...

- Vagrant VirtualBox
- Coming Soon: Singularity Container

enquiries to: opig@stats.ox.ac.uk
Acknowledgements

With special thanks to my supervisors:
Dr Claire Marks (Oxford), Dr Bruck Taddese (AZ), Dr Alan Lewis (GSK), Dr Alex Bujotzek (Roche), Dr Jiye Shi (UCB), Prof Charlotte Deane (Oxford)

And to my DPhil funders: EPSRC, MRC, the Systems Approaches to Biomedical Sciences CDT (Oxford) & partner companies

And to the organisers of CASSS HOS2021 for inviting me to speak as a “Next-Generation Investigator”
Supplementary Slides
Making a set of “representative human antibody” models

Designed to capture as much sequence & structural diversity as possible within the “modellable space”

Protocol used in TAP metric comparison described in PNAS 116(10):4025-4030

Splitting Therapeutics by Kappa/Lambda LCs

Models containing Lambda light chains seemed inherently less ‘developable’ than those containing kappa light chains (90% of CSTs involve kappa light chains).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>TAP Metric</th>
<th>Kappa Subset ($\mu \pm \sigma$)</th>
<th>Lambda Subset ($\mu \pm \sigma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>242 CST Models</td>
<td>PSH</td>
<td>120.89 ± 15.10</td>
<td>142.03 ± 19.09</td>
</tr>
<tr>
<td></td>
<td>PPC</td>
<td>0.21 ± 0.47</td>
<td>0.53 ± 0.56</td>
</tr>
<tr>
<td></td>
<td>PNC</td>
<td>0.38 ± 0.64</td>
<td>0.60 ± 0.77</td>
</tr>
<tr>
<td></td>
<td>SFvCSP</td>
<td>3.82 ± 7.38</td>
<td>1.67 ± 7.87</td>
</tr>
<tr>
<td>14,072 VdH Ig-seq Models</td>
<td>PSH</td>
<td>131.27 ± 21.41</td>
<td>141.68 ± 17.82</td>
</tr>
<tr>
<td></td>
<td>PPC</td>
<td>0.17 ± 0.40</td>
<td>0.52 ± 0.73</td>
</tr>
<tr>
<td></td>
<td>PNC</td>
<td>0.27 ± 0.48</td>
<td>0.74 ± 0.83</td>
</tr>
<tr>
<td></td>
<td>SFvCSP</td>
<td>4.56 ± 7.44</td>
<td>0.84 ± 6.48</td>
</tr>
<tr>
<td>19,019 UCB Ig-seq Models</td>
<td>PSH</td>
<td>125.40 ± 18.56</td>
<td>139.66 ± 17.88</td>
</tr>
<tr>
<td></td>
<td>PPC</td>
<td>0.11 ± 0.31</td>
<td>0.31 ± 0.53</td>
</tr>
<tr>
<td></td>
<td>PNC</td>
<td>0.22 ± 0.40</td>
<td>0.65 ± 0.88</td>
</tr>
<tr>
<td></td>
<td>SFvCSP</td>
<td>3.67 ± 5.30</td>
<td>0.12 ± 5.24</td>
</tr>
</tbody>
</table>

• Consistent with DeKosky et al. (Lambda L3’s much more hydrophobic than Kappa L3’s)

Splitting Therapeutics by Species Origin

Table S8. 242 CST TAP values split by species origin.

<table>
<thead>
<tr>
<th>TAP Metric</th>
<th>101 Human ($\mu \pm \sigma$)</th>
<th>108 Humanized ($\mu \pm \sigma$)</th>
<th>30 Chimeric ($\mu \pm \sigma$)</th>
<th>3 Mouse ($\mu \pm \sigma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total CDR Length</td>
<td>48.68 ± 4.09</td>
<td>47.80 ± 3.42</td>
<td>46.77 ± 3.55</td>
<td>46.33 ± 1.25</td>
</tr>
<tr>
<td>PSH</td>
<td>127.76 ± 18.56</td>
<td>120.90 ± 14.20</td>
<td>115.73 ± 15.58</td>
<td>117.26 ± 9.44</td>
</tr>
<tr>
<td>PPC</td>
<td>0.29 ± 0.58</td>
<td>0.20 ± 0.36</td>
<td>0.26 ± 0.55</td>
<td>0.05 ± 0.06</td>
</tr>
<tr>
<td>PNC</td>
<td>0.34 ± 0.56</td>
<td>0.50 ± 0.75</td>
<td>0.30 ± 0.63</td>
<td>0.50 ± 0.50</td>
</tr>
<tr>
<td>SFvCSP</td>
<td>4.06 ± 7.44</td>
<td>3.13 ± 7.80</td>
<td>3.29 ± 5.99</td>
<td>7.58 ± 6.75</td>
</tr>
</tbody>
</table>

- Appears that the more human mAbs have larger patches of hydrophobicity than mouse mAbs.
- We also split by clinical progression (P2, P3, Approved) and drug campaign status (active/discontinued) but found no significant differences in TAP metric values.