Prediction of Protein-Protein Binding Sites and Epitope Mapping
Epitope Mapping

- Antibodies interact with antigens at epitopes
  - Epitope is collection residues on antigen
  - Continuous (sequence) or non-continuous
  - Patentable druggable sites

- Methods for mapping epitopes
  - X-ray crystallography (gold standard)
  - Peptide scanning
  - Display technologies (e.g. phage display)
  - Hydrogen-Deuterium exchange (HDX)

- In silico methods for mapping epitopes
  - Patch Analysis (antigen surface properties)
  - Protein Docking (pose prediction)
  - Very challenging: large surface area / conformational flexibility

- Key question: are in silico methods accurate?
  - Suggest mutagenesis experiments early in development
  - Reduce crystallographic efforts / expense
Hydrophobic Patches

- Identify key hydrophobic regions on the molecular surface
  - Hydrophobic patches implicated in protein binding
  - Desolvation of hydrophobic patches is a driver of protein-protein interactions (e.g. binding affinity, HIC, aggregation, etc.)

**Protein Patch Analyzer Methodology**
- Derive patches from local lipophilic potential\(^1\)
- Map atomic logP (octanol/water partition coefficient) onto surface
- Remove portions of surface with local logP < 0.1 (≈ methyl C)
- Retain contiguous patches of at least 50 Å\(^2\) (≈ methane)

---

% Inclusion Body Predictions for 31 Adnectins

- **Sequences and experimental data provided in publication**¹
  - % IB: a measure of insoluble protein aggregate formation
  - 31 triple mutants in series

- **Mutating key residues of hydrophobic patches improves solution behavior**
  - Mutated residues correspond to hydrophobic patches
  - Polar variants reduce % IB
  - Rational design using homology models effective in this case

<table>
<thead>
<tr>
<th>Adnectin</th>
<th>% IB</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>76</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
</tr>
</tbody>
</table>

¹ Trainor et al., JMB 2016, 428(6), p1365-1374
Hydrophobic Patches Predict Adnectin Solubility

- Homology model of all 31 mutant adnectins
  - Measure hydrophobic patch surface area
  - Correlate with experimental data – compare with SAP score

**MOE Hydrophobic Patch Descriptor**
- Sum of hydrophobic patch surface areas
- Averaged over 25 LowModeMD conformations

**SAP Spatial Aggregation Propensity (Trout)**
- Score of exposed residues
- Averaged over MD simulation

Hydrophobic Patch Statistics

- ~3000 crystal structures of antibodies in PDB, 1582 used
  - Removed structures with resolution > 3 Å or missing segments in Fv domain

- Small hydrophobic patches are common
  - Unusually large patches present concern for developability
  - Multiple hydrophobic patches together can reduce solubility
HIC Predictions for 137 Clinical Candidates

- Recent publication providing sequences and experimental data\(^1\)
  - Diverse set of antibody Fv regions expressed on common IgG1 scaffold
  - Use data as benchmark to measure broad applicability of HIC predictions

- Generate homology models and predictions with descriptors / QSPR
  - Automatic antibody modeling and patch surface area calculation

\(^1\): Jain et al. PNAS 2017;114:944-949
HIC Predictions for 137 Clinical Candidates

- **3D hydrophobicity descriptors outperform sequence-based index**
  - Signals are generally modest
  - Likely due to clinical candidates having been solubility optimized
  - Difficult to differentiate narrow range of low HIC RT values

- **CDR region important**
  - Candidates -> framework mutations
  - LowModeMD sampling improvement

- **Prediction quality is case dependent**
  - Adnectin predictions more accurate
  - Triple mutants easier than multi-class

- **QSPR model more predictive**
  - 4-point QSPR can enhance signal
HIC Predictions for 137 Clinical Candidates

- Sum of hydrophobic patch surface areas of CDRs calculated

![Graph showing frequency distribution of cdr_hyd SA (Å²)](image)

- Many candidates hydrophobic vs PDB antibodies
  - Trastuzumab has a low HIC RT (9.7 min)
  - Cixutumumab has higher HIC RT (11.8 min) but longer H3 loop dilutes the hydrophobicity
  - Cituxumumab removed from pipeline after phase II

- Predictions across classes are difficult
  - Relative predictions within a series is more reliable
  - Ab:Ag co-crystal structure useful for optimization
Hydrophobic Patches Prevalent at PPI Sites

- 750 antibody:antigen complexes from antibody structural database
  - Antigens larger than 2000 residues removed
  - Count hydrophobic patch frequency at PPI
  - Order by patch surface area (blue is largest)

- Useful for epitope mapping and identifying residues for PPI optimization
  - To reduce undesired PPIs, mutate residues forming hydrophobic patches
Patch Analysis of Interleukin-1β

- Known complexes with Canakinumab and Gevokizumab
  - Protein-protein interfaces consist of clusters of multiple patches
  - Patches are present in other exposed regions

- Patch analysis alone is insufficient to predict epitopes
  - Ranking patches by size does not correlate with interface
Protein-Protein Docking

- **Predict macromolecular complexes with protein docking**
  - More comprehensive approach for identifying binding sites
  - Computationally more expensive than Protein Patch Analysis

- **First principles physical approximations**
  - No protein structure training set used in parametrization

- **Site restraints incorporated using generic energy term**
  - Applied automatically to antibody CDR regions

![Antibody PDB: 3G6D](image)

Antigen Pose Predictions (100-200)
Antibody Docking Results

- **ZDOCK 4.0 Benchmark test set**
  - Success defined as a structure with ligand RMSD < 10 Å
  - Prior knowledge (site restraint) needed to achieve high accuracy with state-of-the-art dockers

Performance of Protein Dockers on Antibodies

![Performance Graph](chart.png)

- **Repacking/refinement improves ranking**
  - Backbone and sidechain rearrangement in complex
  - Scoring the single correct pose is not realistic
  - Ranking of top 100 shows good enrichment
  - Use ensemble of poses to refine prediction

Epitope Mapping Uses Multiple Docked Poses

- An approach for analyzing the protein poses is required
  - 100-200 poses needed for reasonable performance
  - Use protein-protein interaction fingerprints to describe poses

- **Goal: Identify sets of interacting residues**
  - Dock antibody Fv restricted to CDRs to the entire antigen
  - For each pose, identify antigen residues that are implicated
  - Generate statistical summaries of docked poses
  - Is epitope identification more robust than predicting a single pose?

- **Validation: Similarity to native complex structure**
  - Measure overlap between sets of interacting residues in different poses
  - Use ZDOCK5 benchmark complex structures
Identify Interactions using Surface Patch Contacts

- Identify residue-residue interactions from antibody poses
  - Fingerprints: surface patch contacts on the antigen in each pose
  - Count residue on the antigen if it is interacting with the antibody

- Classify patch contacts into four categories
  - Both HYD, One HYD, Any POS/NEG, None

- Example: Colicin E7 nuclease (7CEI), key residue N26

Native complex crystal contacts  Unbound structure docked pose (5Å)
Surface Contacts Describe Native Complexes

- **Benchmark data set of protein-protein complexes**
  - 150 rigid-body targets from ZDOCK test set
  - 32 antibody-antigen complexes, 118 others

- **Protein contacts calculated using distance and energy**
  - Cut-offs applied to select most significant interactions
  - Average number of contacts: ~6 per complex

- **Surface interactions calculated from patch fingerprints**
  - Most contacts correspond to surface interactions
  - Energetic calculations tend to over-emphasize charged interactions
  - Surface contacts better represent hydrophobic contacts

![Protein Contacts Graph](image1)

![Residues in Common Graph](image2)
Protein-Protein Interaction Fingerprints

- Populations are calculated for each residue in all poses
  - The top-200 refined poses are used in the analysis
  - The most likely residues to interact with the antibody are identified

- Highest peaks do not necessarily represent epitopes
  - Cluster poses to identify residues that are spatially close

**Interleukin-1β/canakinumab Pose Interaction Histogram**
Using Fingerprints to Determine Epitopes

- **Cluster poses by fingerprint similarity**
  - Connected poses are defined by a similarity cutoff (0.4)

- **Rank clusters by ensemble free energy**
  - Effective temperature is an empirical parameter fit to score values

- **For each cluster generate the epitope**
  - Each cluster is a potential epitope (spatially close)
  - Select highest-population residues from each cluster
Epitope Benchmark Results

- Predictions carried out using ZDOCK test set
  - Selected targets with representative docking performance
  - Divided into antibody (20) and non-antibody (30) classes

- Native contacts identified in predicted epitopes

- At least 3 native contacts in top-ranked epitope
  - Antibody: 60%, non-antibody: 53%
Predicted Epitopes of IL-1β Complexes

- **Results of docking with different antibody structures**
  - Correct ordering impossible from patch analysis alone
  - Top-ranked docking poses are not in top-ranked clusters

- **PPI fingerprint analysis correctly predicts crystal epitopes**
Hydrogen-Deuterium Exchange (HDX)

- **Soak the protein with deuterium**
  - Deuterium will replace H over time; buried atoms no replacement
  - Solvent accessibility, H-bonding, sidechain neighbors affect uptake

- **Digest with pepsin and use mass spectrometry**
  - Collection of partially deuterated sequence fragments

- **Measure deuteration of backbone amides in solution**

**HDX Useful to Probe Binding Interfaces**

- **Compare bound and unbound deuterated structures**
  - Reduction in deuteration indicates interface residues

![Diagram showing IL-23 and p40 domains with deuterated residues indicated]

- **Provides low-resolution information – sequence segments**
  - Qualitative effect, divided into weak and strong categories
  - Effects may not be due to contact, e.g. flexibility changes
  - Not all interactions appear

- **Idea: use HDX data to bias docking to find IL-23:adnectin epitope residues**

---

Characterization of Binding Interface

- HDX data alone is not enough to characterize epitope
  - Resolution does not label single residues
  - Energy-based methods assume accurate atomic structure
  - Buried surface alone does not distinguish different types of interactions (e.g., hydrophobic, H-bonds)
  - Surface Patch Fingerprints can be effective

- Patch contacts reflect key interactions
  - Close agreement with atomic-level interactions in native structure
  - More selective than contacts based on energy alone
  - Surface patches provide a more robust characterization of docked conformations
Interleukin-23 : Adnectin Calculation

- **IL-23:Adnectin show conformational changes in complex**
  - Docking the **bound** conformations works very well (RMSD < 0.5 Å, #1)
  - Docking **unbound** conformations fails (too much loop movement)

- **Docking procedure**
  1. 20 Adnectin models
  2. 20 IL-23 models
  3. HDX IL-23 constraints
  4. FG-loop constraints
  5. Dock Pairs – 400 runs

- **Epitope Analysis**
  - Keep top 1,000 poses
  - Calculate patch FPs and cluster
  - Rank clusters by docking score, contact area, HDX overlap
  - Output top rank

*Image of molecular models with annotations:*
- 20 Adnectin Homology Models (from PDB:1FNA - unbound)
- 20 IL-23 Homology Models (from PDB:3D87 - unbound)
- IL-23 loop not well resolved in unbound crystal structure
- Adnectin FG-loop
Interleukin-23 : Adnectin Epitope Prediction

- Ensemble modeling preserves key features of complex
  - Consensus fingerprints show good correspondence (4/8) to crystal interactions
  - Homology models outperform docking of unbound structures when there is significant conformational flexibility

- HDX biased protein docking can elucidate epitope
Conclusions

- **Protein Patch Analyzer for detecting protein hot spots**
  - Fast method for predicting protein binding sites on a single body
  - Identify hydrophobic patches implicated in PPIs

- **Predict protein-protein complexes with docking**
  - Predicts protein interactions between two bodies
  - Site restraints derived from automatic sequence annotation are effective for docking antibody-antigen complexes

- **Analysis of docked poses can be used for epitope mapping**
  - Similar poses determined by clustering PPI fingerprints
  - First-ranked epitope correct in 56% of antibody-antigen targets

- **Epitope analysis can be used to interpret HDX data**
  - Case study: Adnectin complex with IL-23
  - HDX data can be incorporated into docking calculations
  - Consensus fingerprints show good agreement with crystal structure

- **Epitope mapping available in MOE 2018**