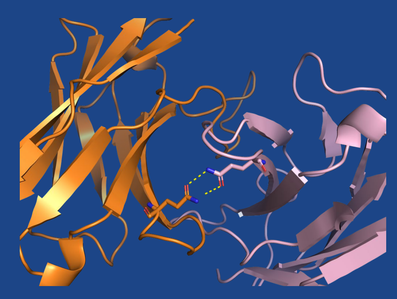
Domesticating natural proteins for use in synthetic protein switches

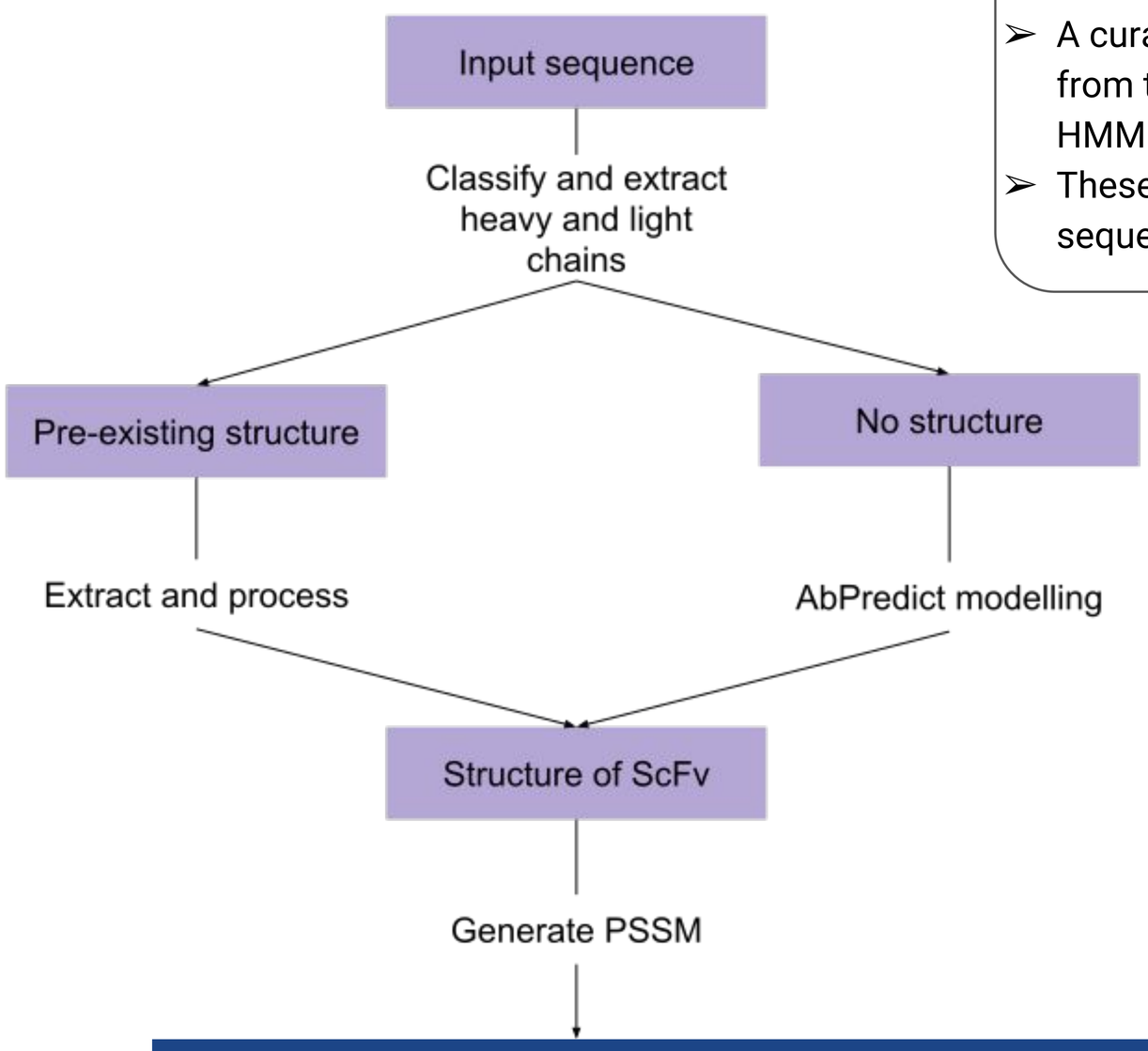
Jake Parker[†§], Kirill Alexandrov[†], and Sarel Fleishman[§]
(†) Institute for Molecular Bioscience, University of Queensland
(§) Weizmann Institute of Science



Background

- Antibodies, and even antibody fragments, are not suited to many in vitro applications due to their poor biophysical properties, often leading to misfolding and aggregation.
- An automated pipeline for the re-engineering of antibody fragments from structure and sequence was developed.
- The general design strategy combines both evolutionary analysis and ROSETTA atomistic design.
- Antibody sequence diversity is complex; both the heavy and light chains arise from the recombination of multiple genes to form a large number of functional proteins (V(D)J recombination).
- This necessitated special treatment when collecting “evolutionary information” of antibodies for use in design.
- In total 28 antibodies– 12 with a predetermined structure and 16 without– were subjected to design.

Pipeline overview

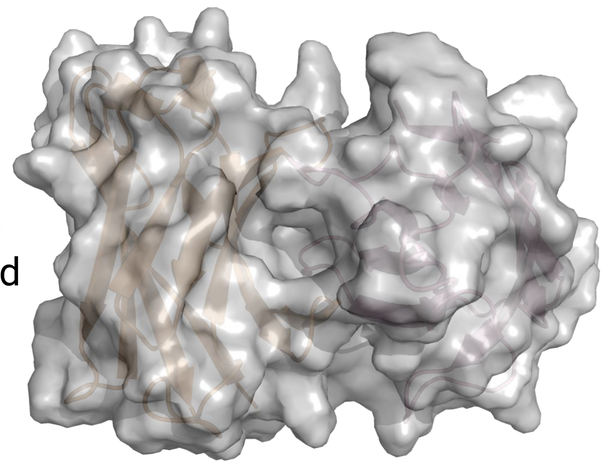


Chain classification and collection

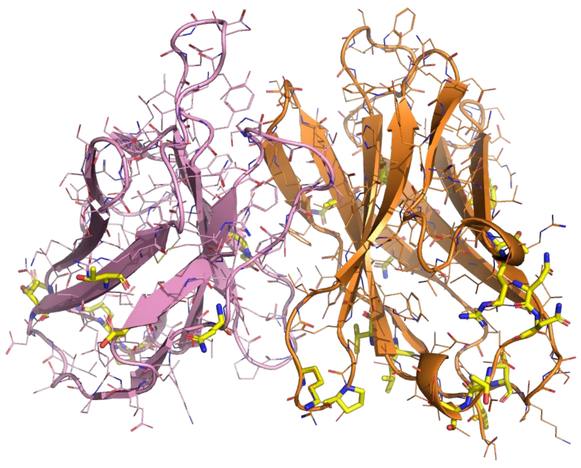
- A curated set of heavy-, light kappa-, and light lambda-chain sequences from the SAbDab database were collected, aligned, and used to build an HMM model of each chain-type.
- These models were then used to collect and classify thousands of sequences from the abYsis database for use in PSSM generation.

Homology model generation

- Homology models were built using AbPredict.
- AbPredict produces models with reduced backbone strain which is important for design.

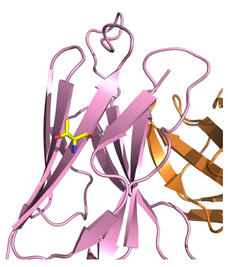


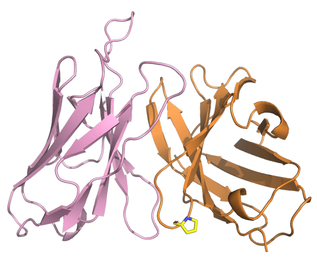
ROSETTA atomistic design

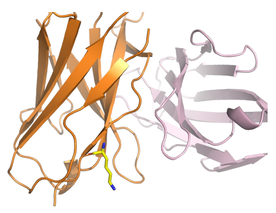


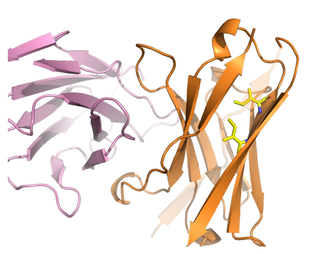
- Chain-specific PSSMs are generated by classifying the input sequence into chain-types and then aligning them against the chain-specific collections described above.
- The PROSS stabilisation algorithm was run as per Goldenzweig et al. 2016:
  - first, “non-permissible” mutations with a PSSM less than 0 score are excluded from the sequence space;
  - second, Rosetta’s FilterScan algorithm is used to find mutations which increase $\Delta\Delta G_{CALC}$ by more than -0.45 Rosetta energy units;
  - finally, Rosetta combinatorial sequence design is used to find an optimal combination of mutations within the space of potentially stabilising mutations.
- On average the designs contained >15% mutations from wild-type (excluding the CDRs).
- Mutations were driven by a combination of the ROSETTA energy function and the PSSM scores.

Four classes of common mutations were observed: increase of surface charge, better core-packing (void filling), proline loop rigidification, and alpha- to beta-branched amino acids for backbone rigidification.


Backbone rigidification


Loop rigidification


Increased surface charge


Void filling