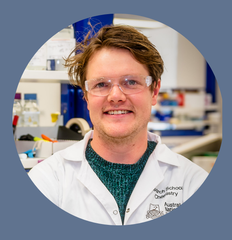
Evolution-based approaches for protein engineers



Joe A Kaczmarski & Colin J Jackson

Research School of Chemistry, 137 Sullivans Creek Road, The Australian National University, Canberra.

joe_kaczmarski | joe-kaczmarski | kaczmarski.joe@gmail.com | jkaczmarski.com

Australian National University

Protein Engineering

= developing useful or valuable proteins

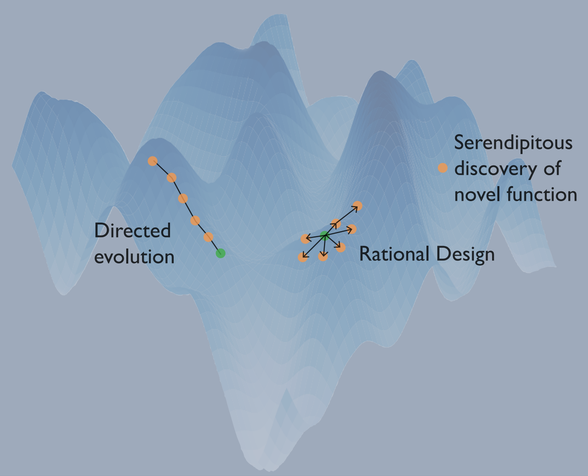
Challenge = need to optimise several properties at once

primary function/activity
stability
ligand/substrate specificity
interactions with other proteins
bioorthogonal interactions
folding kinetics

How to effectively explore sequence space?

(i) DISCOVERY
finding sequences that confer novel/useful functions or properties
(tends to involve a more broad search of seq. space)

(ii) TWEAKING/OPTIMISATION
identify/generate variants with improved function or properties
(tends to be a more localised search of seq. space)



Evolution-based approaches

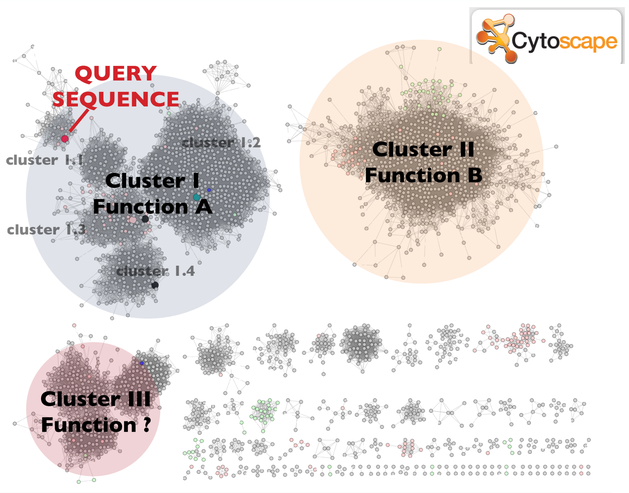
(that we are using)

Exploring sequence space using SSNs

How?

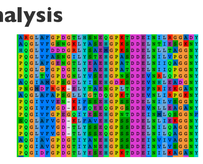
EFI-EST or in-house workflow + scripts

Identify target sequence, enzyme class or protein family
↓
Collect sequences & perform all-by-all BLAST
↓
Annotate network (PDB, Swiss-Prot, Phylum etc)
↓
Adjust edge threshold to separate clusters of isofunctional groups
or separate further into sub-clusters
↓
Identify uncharacterised regions of seq. space
↓
Cluster analysis (EFI-EST)
↓
Characterise representative or consensus sequences



Per cluster analysis

Cluster HMM profiles can be used to identify representative sequences or consensus sequences to characterise



Alignments from individual clusters can be used to perform profile-profile alignment (e.g. for ASR)

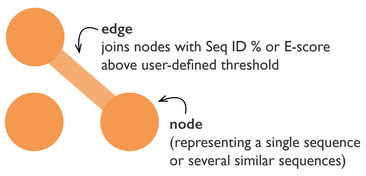
Why?

Visual overview of sequence-function space

Reveals unexplored regions of sequence-function space
- a potential source of novel functions/properties

Guides the selection of representative sequences from distinct clusters
- for broader characterisation of sequence space to better understand drivers of functional diversification
- to use as starting points for engineering (e.g. by directed evolution)

Complements ASR & consensus design workflows
- guides the selection of input sequences
  - e.g. consensus sequences based on a single cluster may be more useful than consensus of whole family.
- per-cluster MSAs and HMMs can aid refinement of multiple sequence alignments



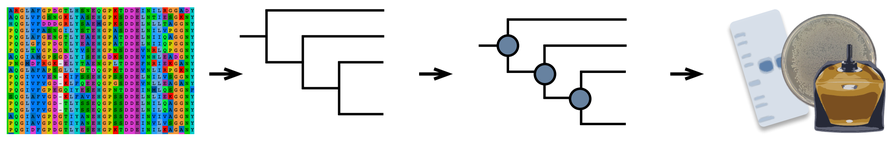
Considerations & Challenges

- large networks can be demanding on (computational) memory
- choosing edge cut-off is not always trivial
- output alignments from cluster analysis will require further refinement
- functionally distinct proteins can share high sequence similarity and may be missed

Useful References:
Copp et al. (2018) *Revealing Unexplored Sequence-Function Space Using Sequence Similarity Networks.*
Gerlt et al. (2015) *Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks.*

Ancestral sequence reconstruction

How?



Collect sequences & align → Infer trees → Reconstruct sequences at ancestral nodes → Characterise

Why?

Good option for exploring *functionally-dense* sequence space to find useful protein sequences or starting points for future optimisation.

Ancestral proteins reconstructed using maximum liklihood approaches ...
- are often thermostable
- may have unique functions/properties useful for engineering (e.g. novel or promiscuous activities)
- thermostable + promiscuous ancestors can make suitable starting points for subsequent engineering

Studying historical evolution highlights drivers of functional diversification within target protein families. More generally, ASR studies can highlight the fundamental drivers of functional innovation.

Considerations & Challenges

- How many sequences to collect?
- Monofunctional or functionally diverse?
- Which nodes to reconstruct?
- How important is it that reconstructed sequences reflects historical ancestors?
- Large trees can be computationally demanding
- Deletions & insertions can make reconstruction more difficult

See:
Spence, Kaczmarski, Saunders & Jackson (2021). *Sequence reconstruction for protein engineers.*

Let's chat!

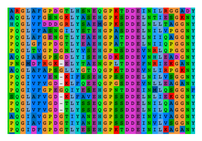
How could evolution-based approaches help your protein engineering & synthetic biology projects?

How could you help us to improve our evolution-based protein engineering workflows?

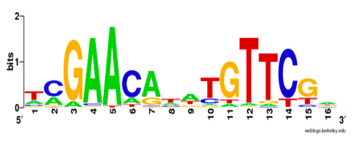
we are particularly interested in leveraging high-throughput screening (e.g. display technologies) & applying machine learning to map protein sequence-function space

Consensus design

How?



Collect sequences (e.g. clusters from SSN) & refine alignment



Construct HMM profile

Full consensus seq. reconstruction

OR

Introduce consensus residues into WT

Why?

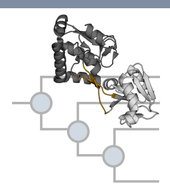
- Reasonably straight-forward (once alignment is refined)
- Enhanced thermostability
- Reduced immunogenicity

Considerations & Challenges

- Relies on good alignment
- Different outcomes depending on diversity of sequences used
- May not capture epistatic interactions/ dynamics

Reference:
Porebski & Buckle (2016) *Consensus protein design*

Or ask me about some of my other work!



Evolution of an enzyme from a non-catalytic binding protein
Using ASR, we traced the emergence and optimisation of cyclohexadienyl dehydratase activity from an ancestral solute-binding protein. A shift in conformational sampling led to an increase in catalytic activity. ***Potential implications for enzyme design.***

Understanding the interactions between neutralising monoclonal antibodies and the *Plasmodium falciparum* circumsporozoite surface protein (CSP)
with Ian Cockburn, JCSMR, ANU
Using a combination of biophysical measurements (ITC/SPR), structural biology, immunological tests & evolution-based approaches, we are determing the factors that confer long-term protection against malaria. ***Implications for rational-vaccine design*** + a better understanding of the antibody affinity maturation process (especially against complex antigens).

Structural basis for the allosteric regulation of the SbtA bicarbonate transporter by the PII-like protein, SbtB, from *Cyanobium* sp. PCC7001
with Dean Price, RSB, ANU
A better understanding of how the cyanobacterial carbon-concentrating mechanism (cCCM) is regulated is supporting efforts to improve crop yields through the incorporation of cCCM components into the chloroplasts of crop plants.