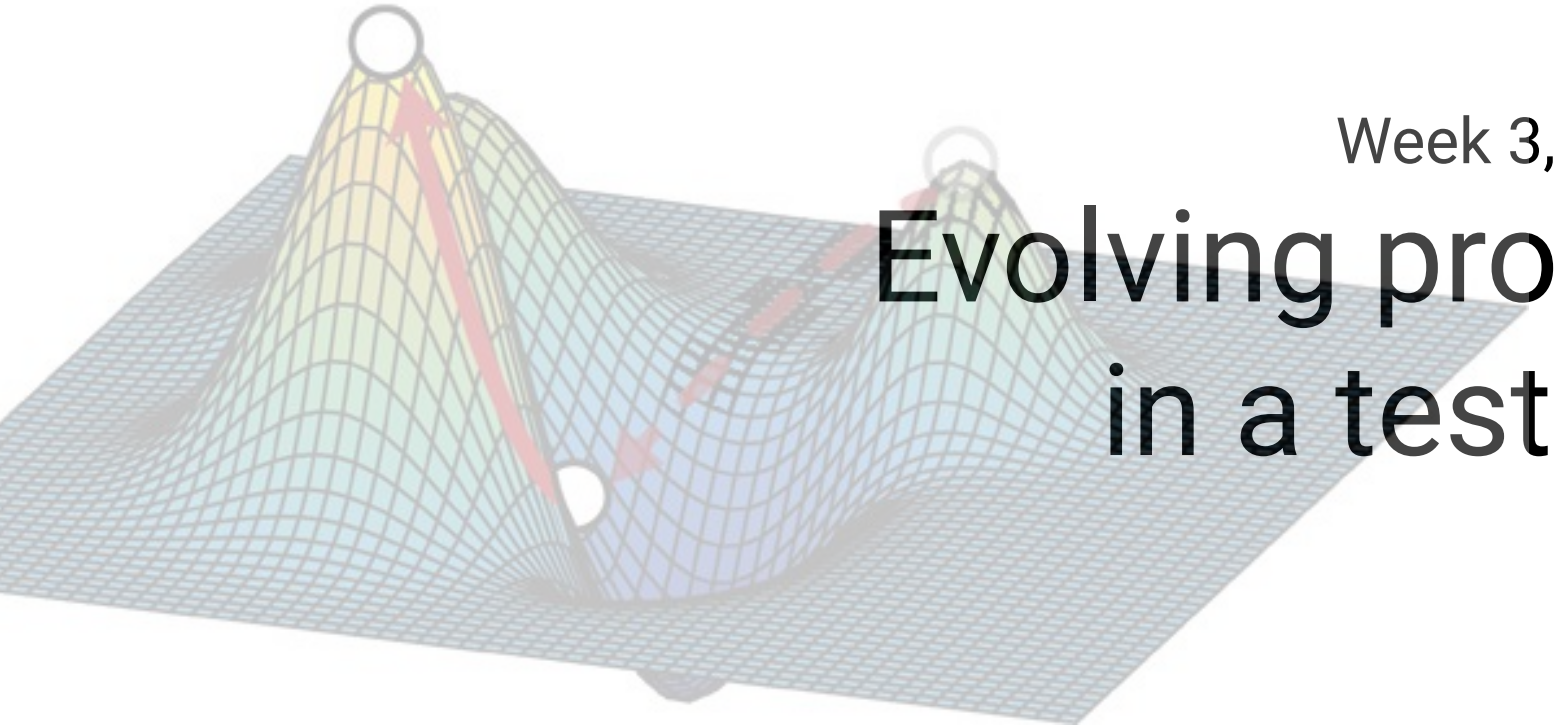
Class core values

1. Be **respect**ful to yourself and others
2. Be **confident** and believe in yourself
3. Always do your **best**
4. Be **cooperative**
5. Be **creative**
6. Have **fun**
7. Be **patient** with yourself while you learn
8. Don’t be shy to **ask “stupid” questions**



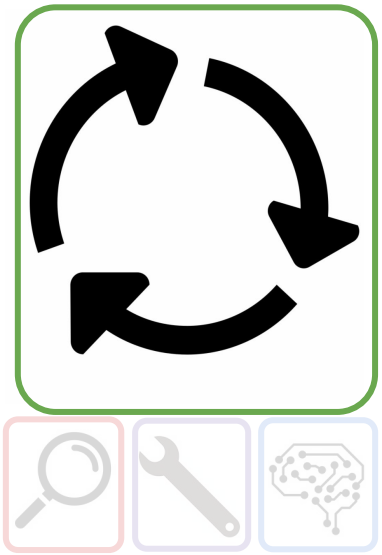
Week 3, Lecture 2

Evolving proteins in a test tube

Learning Objectives

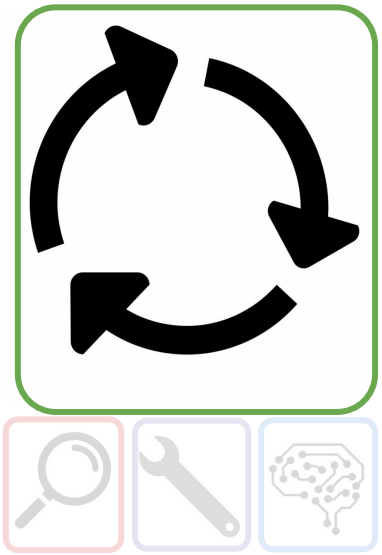
1. Describe the concept of high dimensionality in directed evolution
2. Identify the advantages and disadvantages of different methods for creating diversity in populations
3. Describe the problem of local maxima and potential ways to overcome it
4. Critically evaluate the choice of mutagenesis methods

Directed evolution



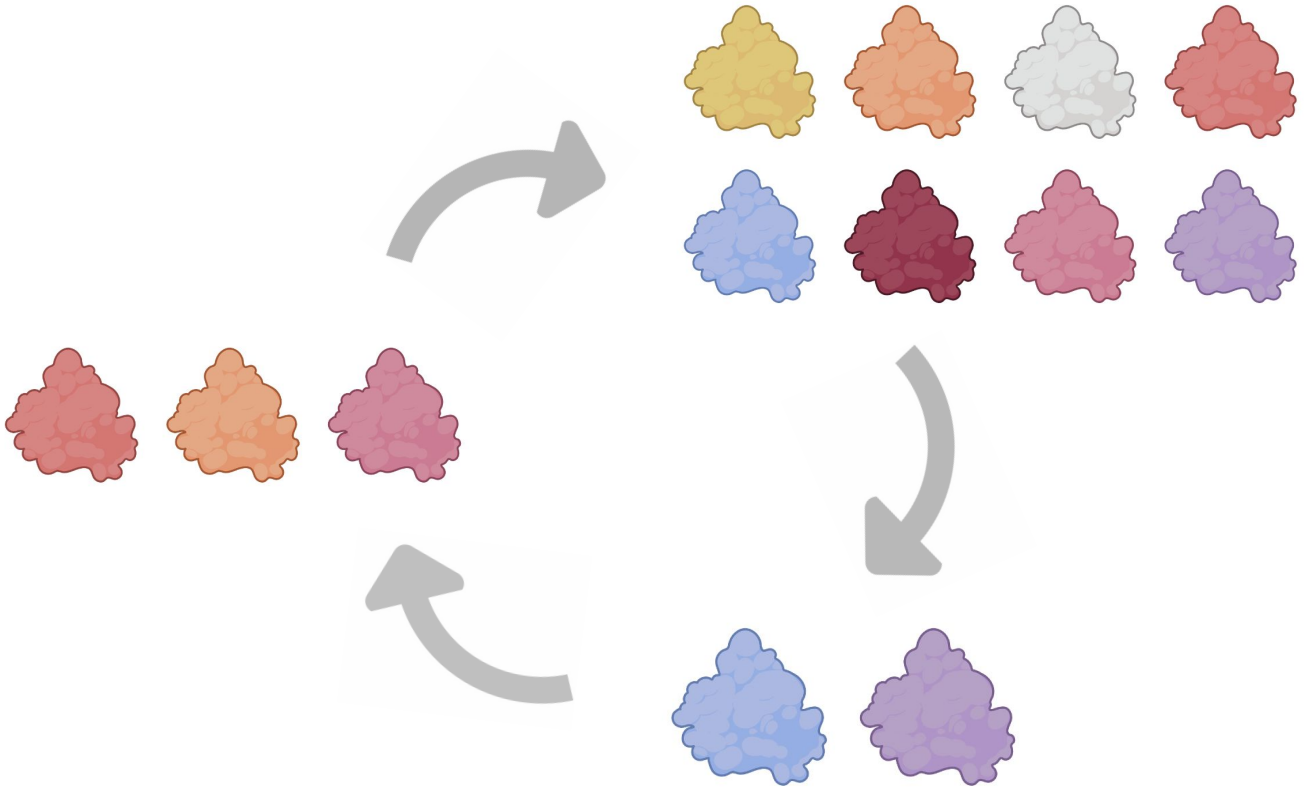
Directed evolution is a powerful method that is inspired by nature

Uses a process of mutation and selection, inspired by evolution

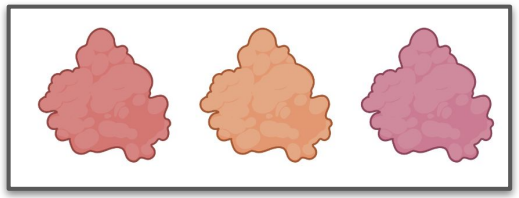


Advantage	Disadvantage
High throughput and can test many things at once	Sequence space is very large and not everything can be sampled
Don’t need prior knowledge	Is not interpretable (we don’t know why things worked)
Can screen multiple features at once	We’re not sure if we’re getting the best result
Random mutations = casting a wider net to make sure things work	You need to have a selection/screening method and you have to have some minimal activity

The overall process of directed evolution



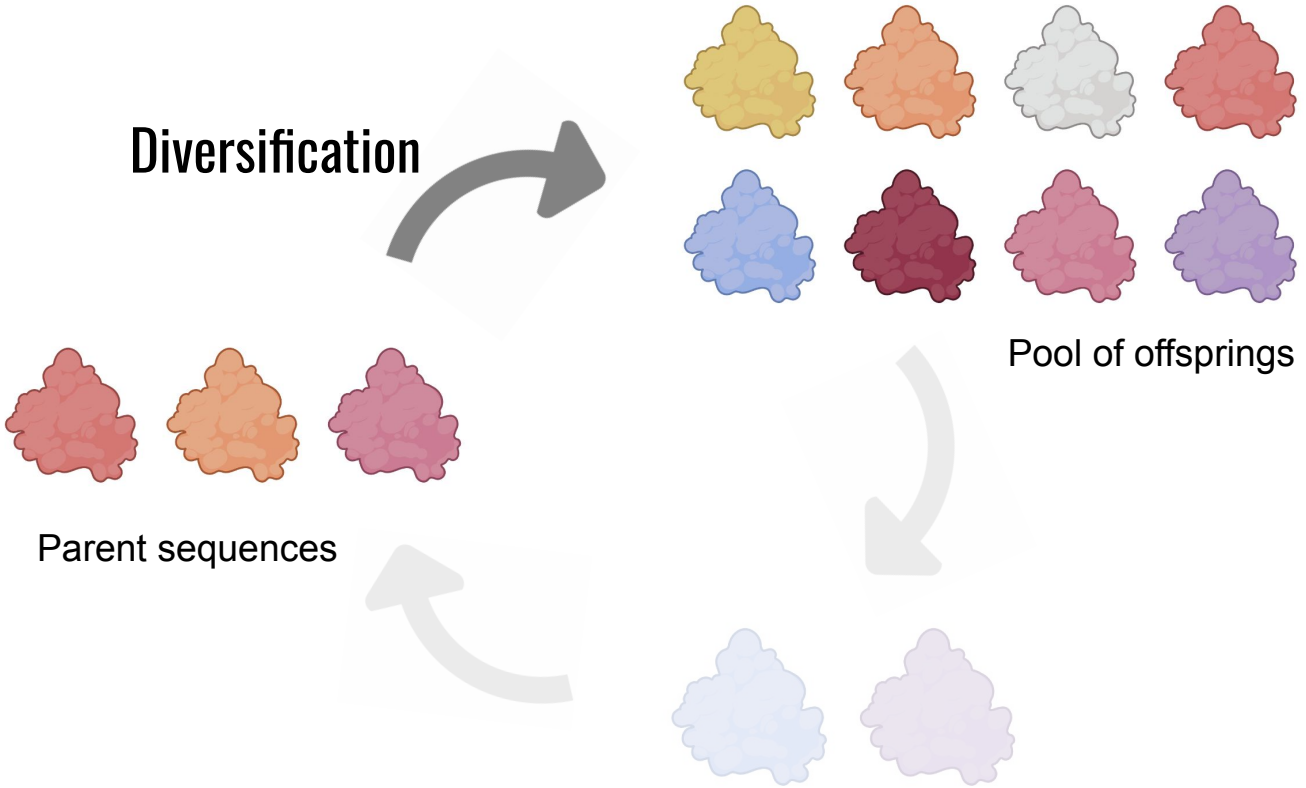
The starting parent sequences should …



Parent sequences

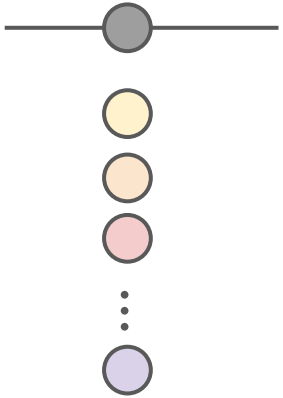
1. Has some initial activity
2. If lots of homologs, starting from ancestral ones can be better as they're more flexible for being mutated
3. Using sequences of proteins that are similar to each other
4. An activity that can be measured

The overall process of directed evolution



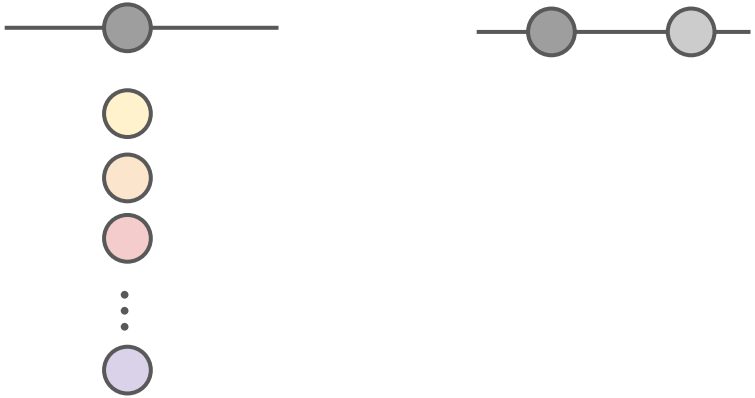
The curse of dimensionality

The curse of dimensionality



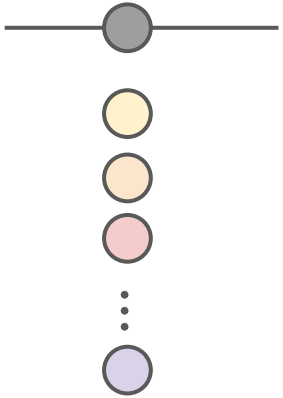
20 possibilities

The curse of dimensionality

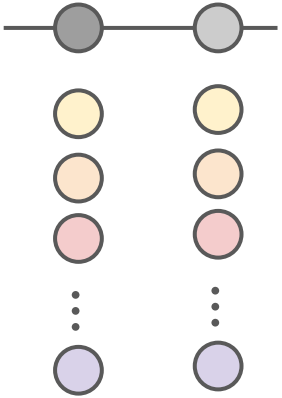


20 possibilities

The curse of dimensionality

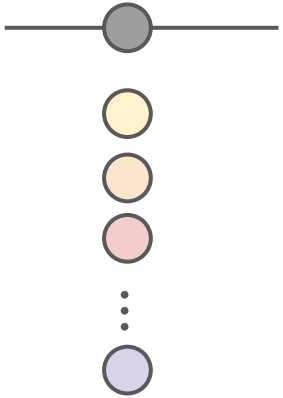


20 possibilities

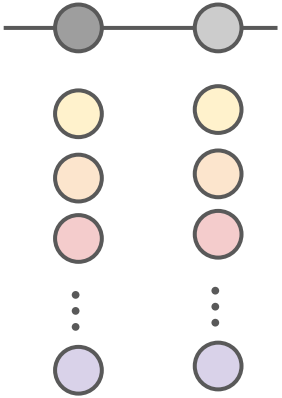


20^2 = 400 possibilities

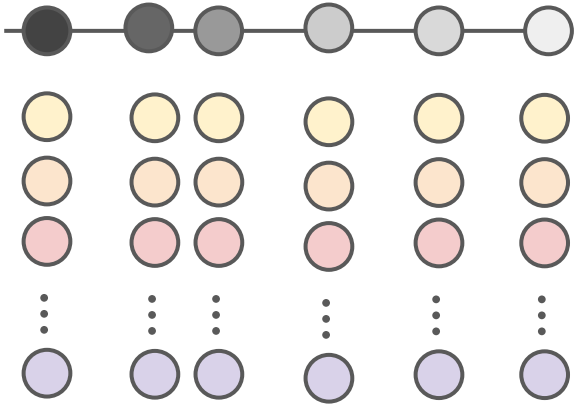
The curse of dimensionality



20 possibilities

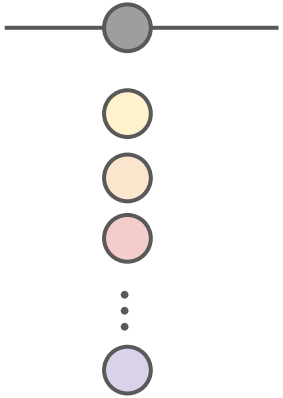


$20^2 = 400$ possibilities

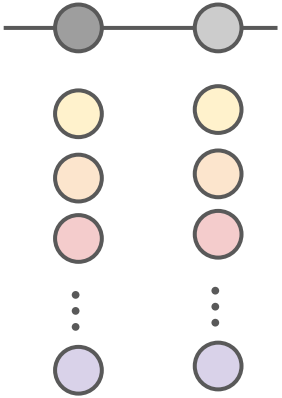


$20^6 \sim 10^7$ possibilities

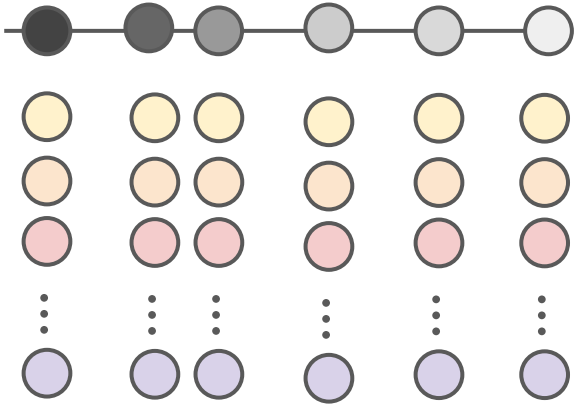
The curse of dimensionality



20 possibilities



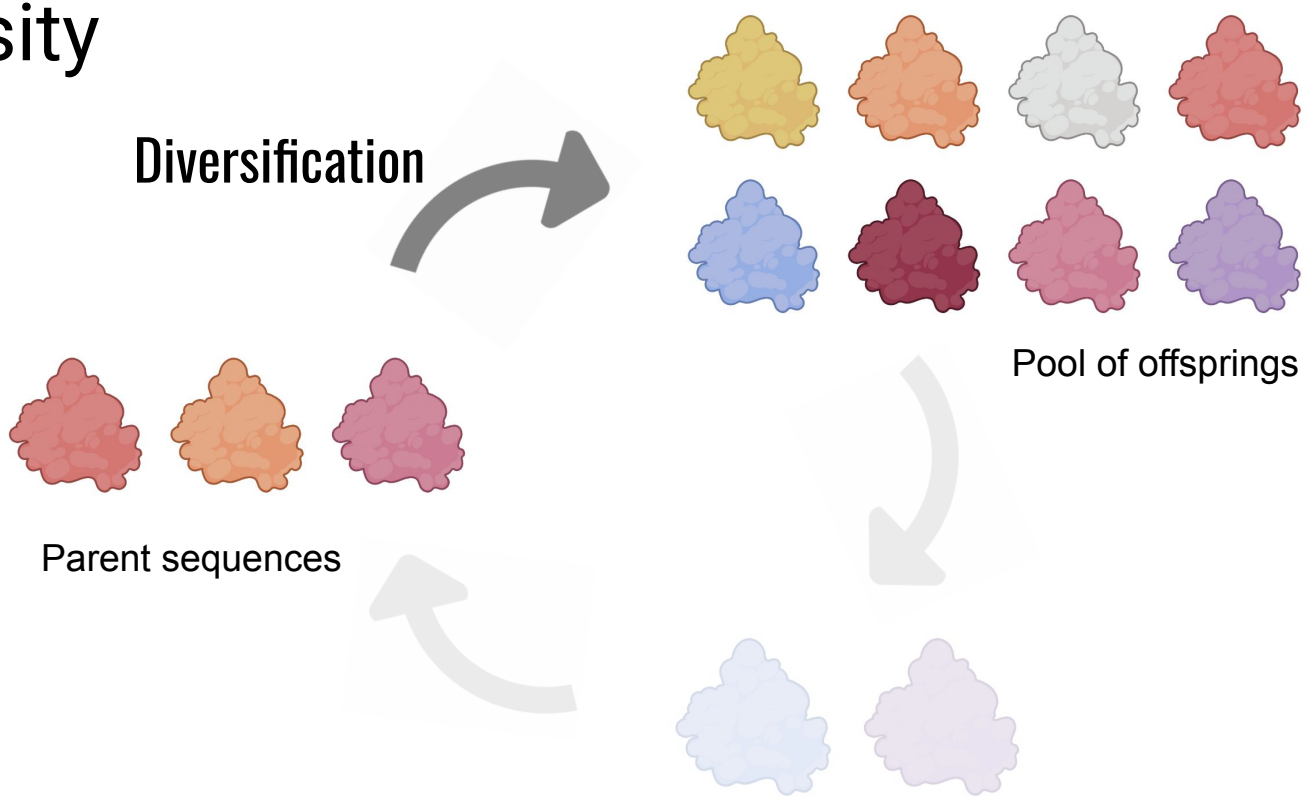
20^2 = 400 possibilities



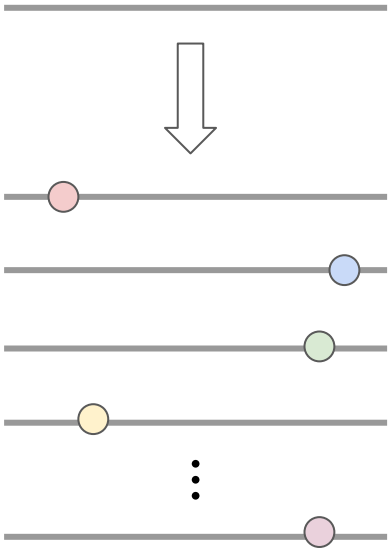
$20^6 \sim 10^7$ possibilities

Limit of yeast surface display!

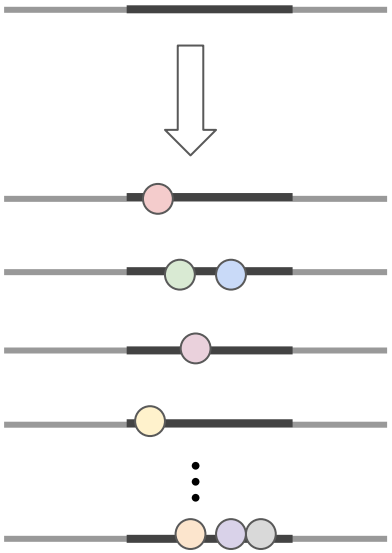
The goal of diversification is to create meaningful diversity



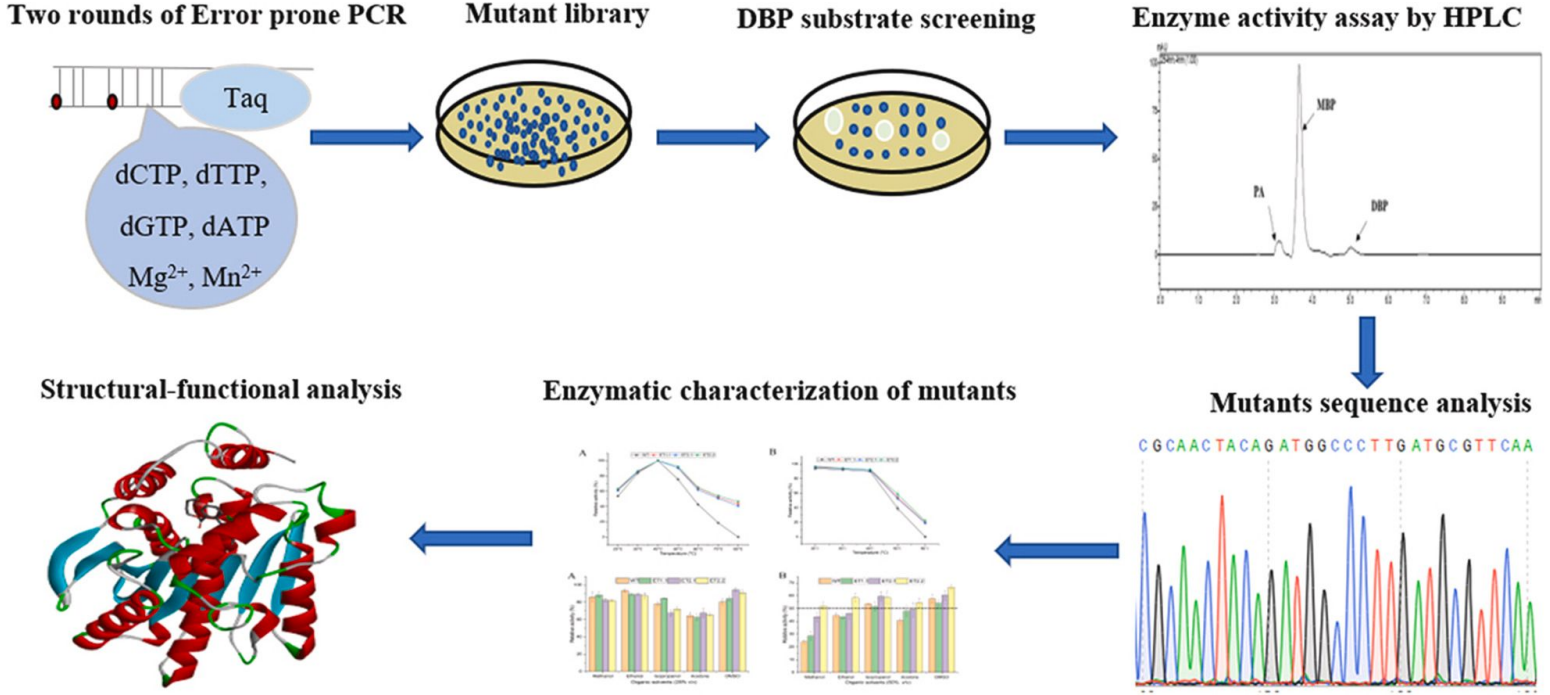
Random mutagenesis is one of the most commonly used methods to generate diversity



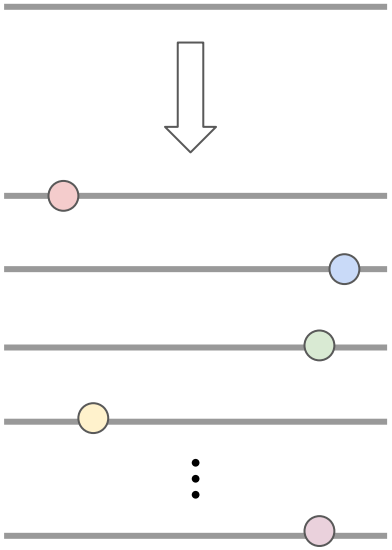
MORPHING is a local version of random mutagenesis



Random mutagenesis enhanced activity and thermal stability of phthalate degrading enzymes

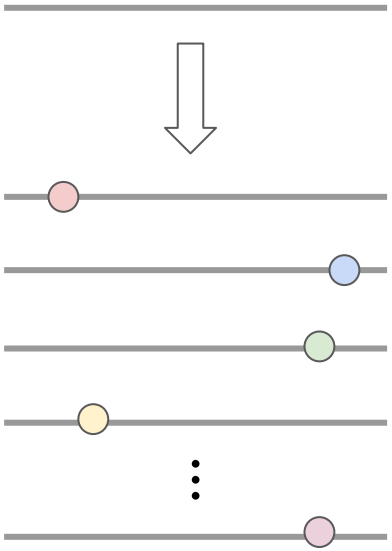


Random mutagenesis is one of the most commonly used methods to generate diversity



Advantage	Disadvantage
No need for prior knowledge	Most mutations are non-functional
You get mutations that you couldn't predict (pleasant surprises)	Not great for a new activity
You can get enhanced stability for free	Can be tricky for the larger proteins
Easy to run and iterate through	Results are not always interpretable

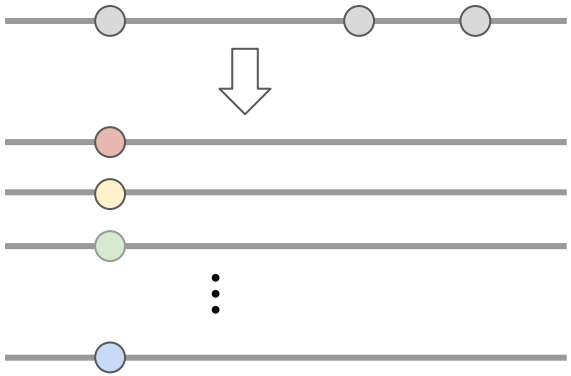
In class activity:



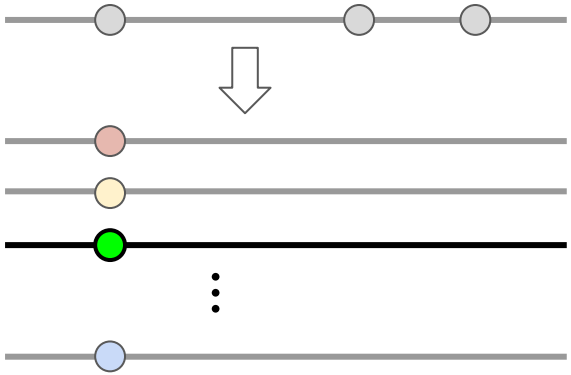
Random mutagenesis for evolving function

1. There's a sweat spot on # of iterations
2. Increased in rate of mutation increased the negative results as well
3. Starting with the fittest parent doesn't always get you to the global maximum

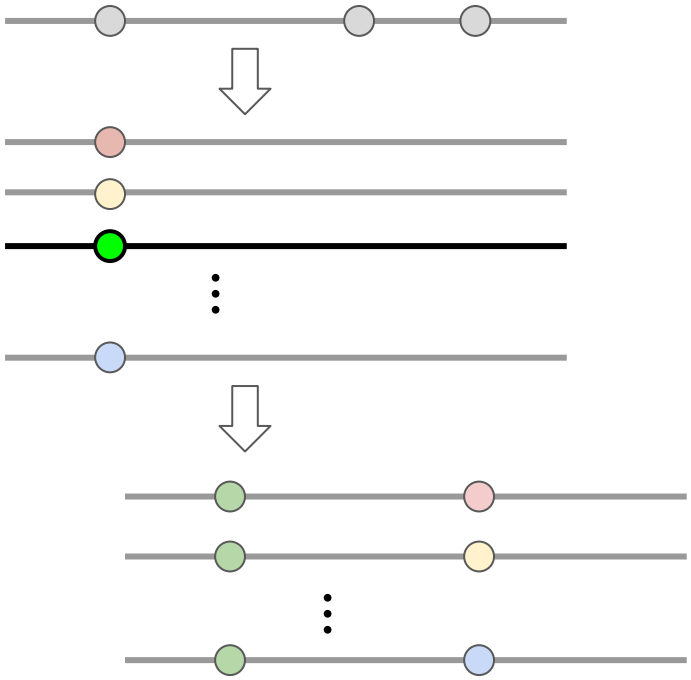
If structural knowledge is present, iterative site saturation mutagenesis (ISM) can be used



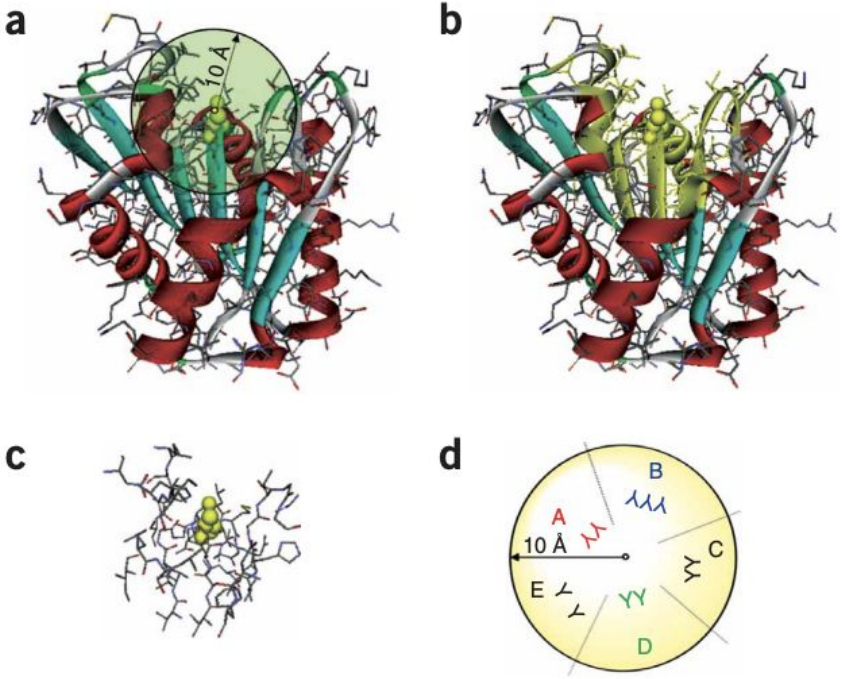
If structural knowledge is present, iterative site saturation mutagenesis (ISM) can be used



If structural knowledge is present, iterative site saturation mutagenesis (ISM) can be used

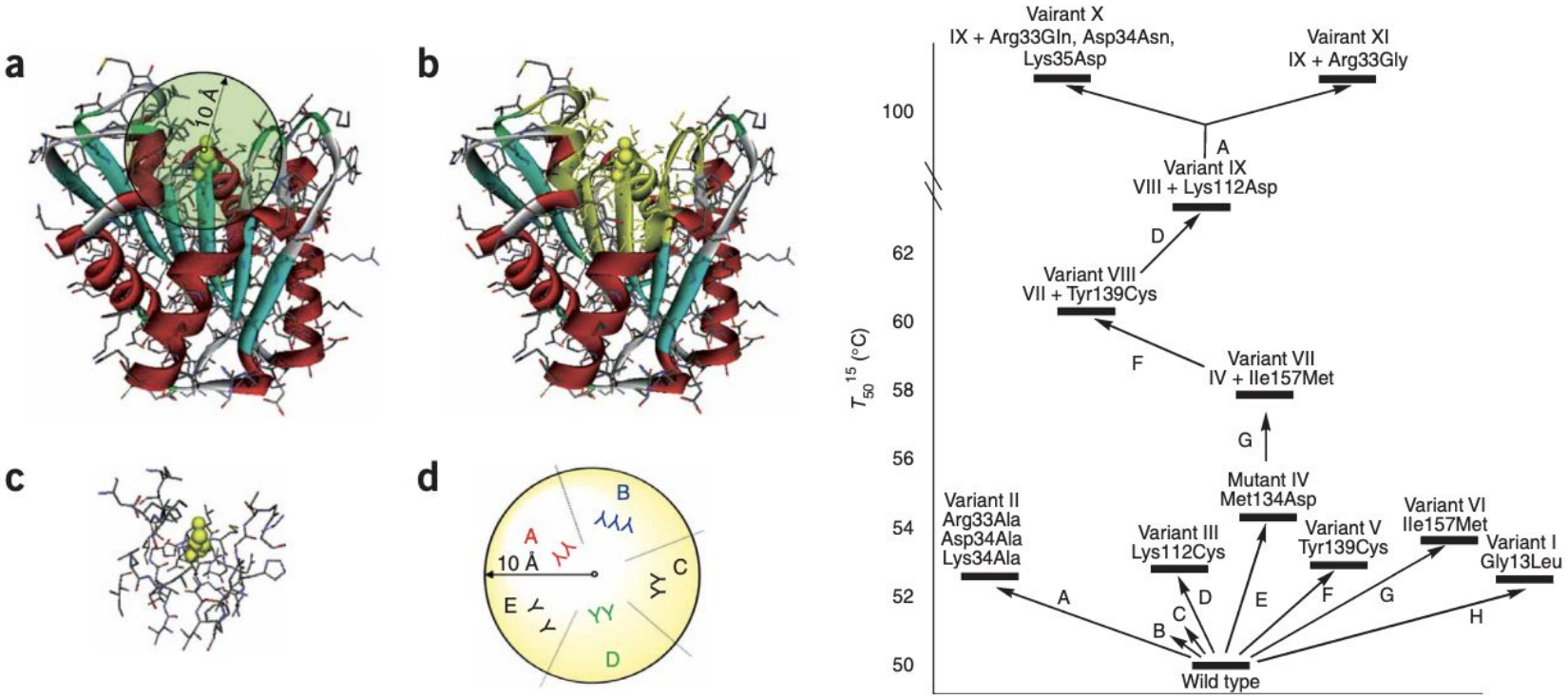


Enhancing thermal stability of LipA using ISM

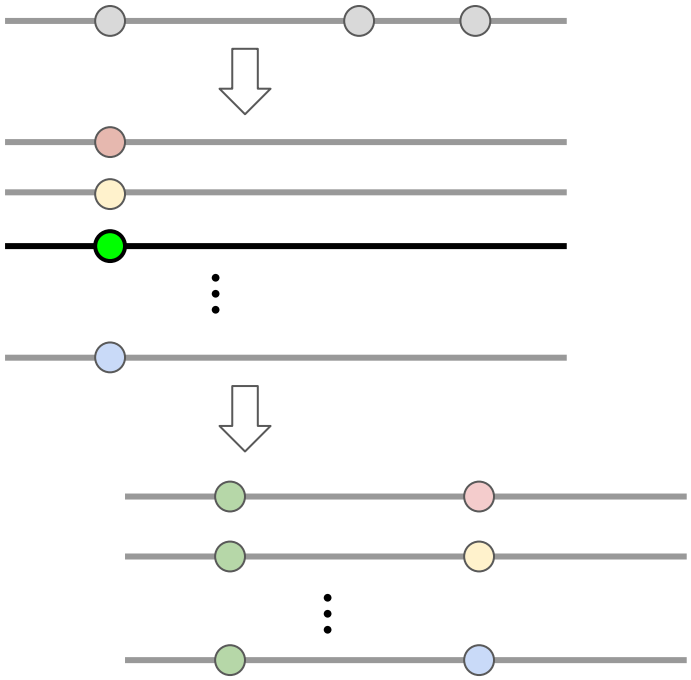


Nature Protocols 2007; 2(4)

Enhancing thermal stability of LipA using ISM

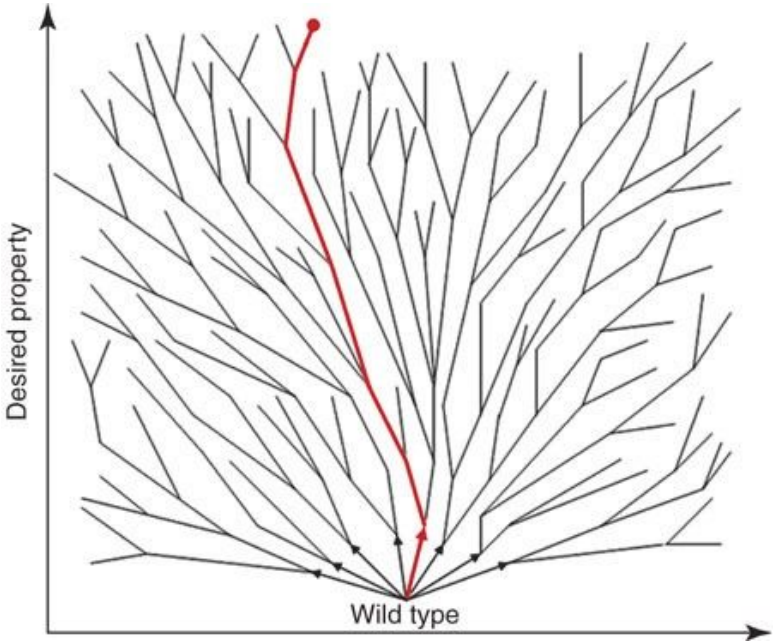


If structural knowledge is present, iterative site saturation mutagenesis (ISM) can be used



Advantage	Disadvantage
A faster approach to get to a maxima compared to random	Doesn't capture distal mutations that can affect activity
You can get synergistic effects or mutations in proximal regions	More likely to get stuck in a local minima
Results are more instructive	Often require structural knowledge

In class activity:

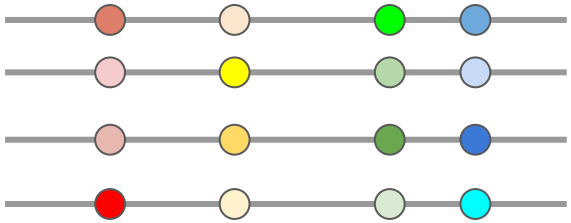


Try ISM to evolve your protein

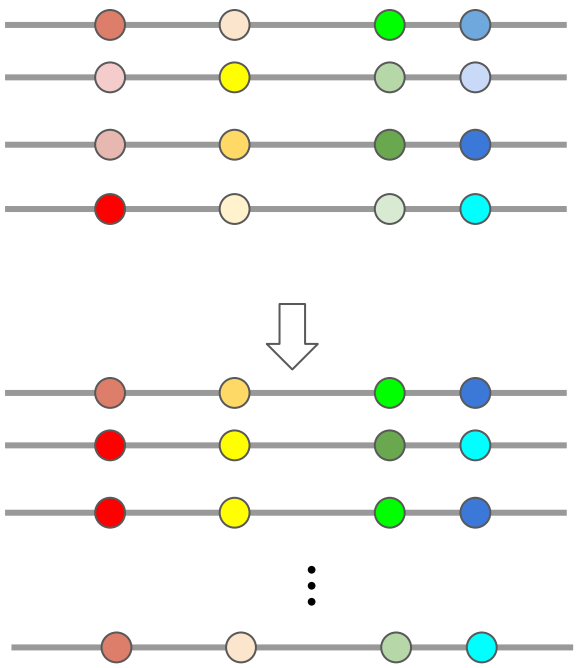
1. Not all mutations are additive
2. Which location you pick matters in getting to the maximum
3. Moves faster than random (in that fewer variants need to be tested)

Image from *Nature Protocols*, **2**, 891-903 (2007)

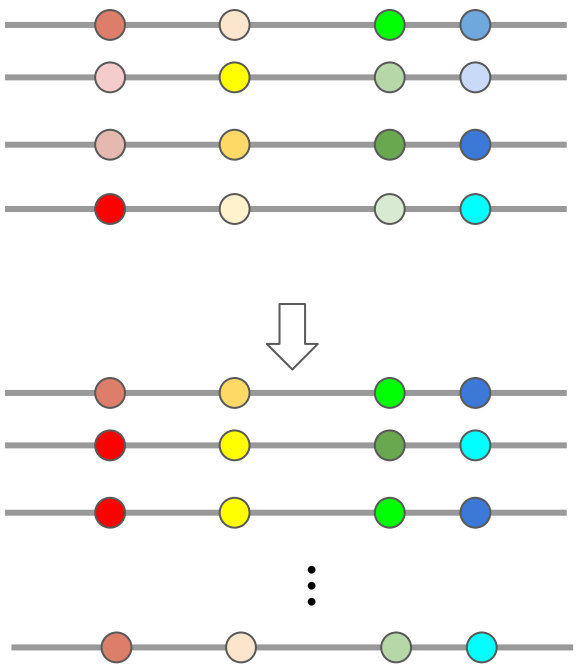
Generating chimeric sequences can be a powerful method for diversifying the sequences



Generating chimeric sequences can be a powerful method for diversifying the sequences

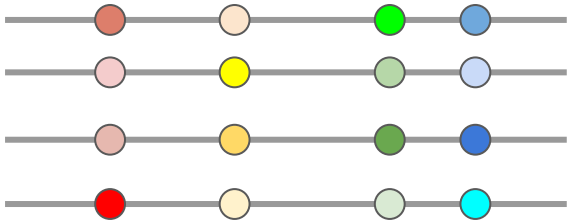


Multiple methods exist for generating these chimeras

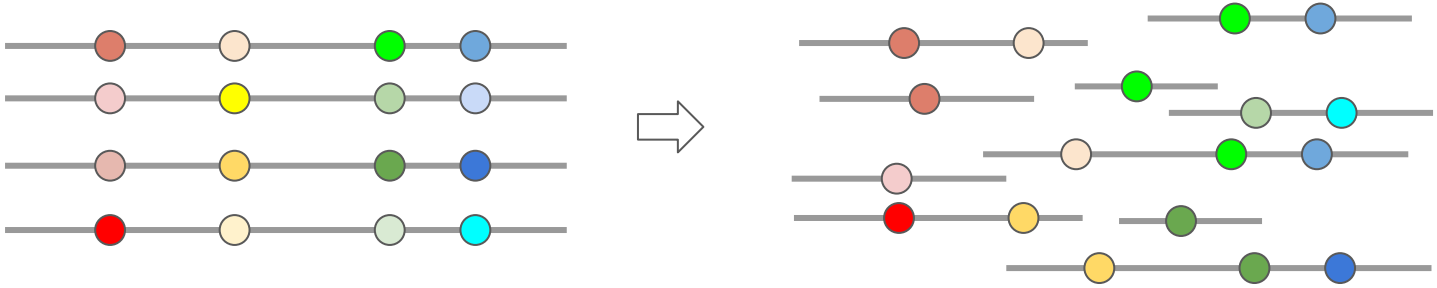


- StEP
- ITCHY
- RACHIT
- Gene shuffling
- SHIPREC
- SCHEMA
- …
- Synthetic libraries

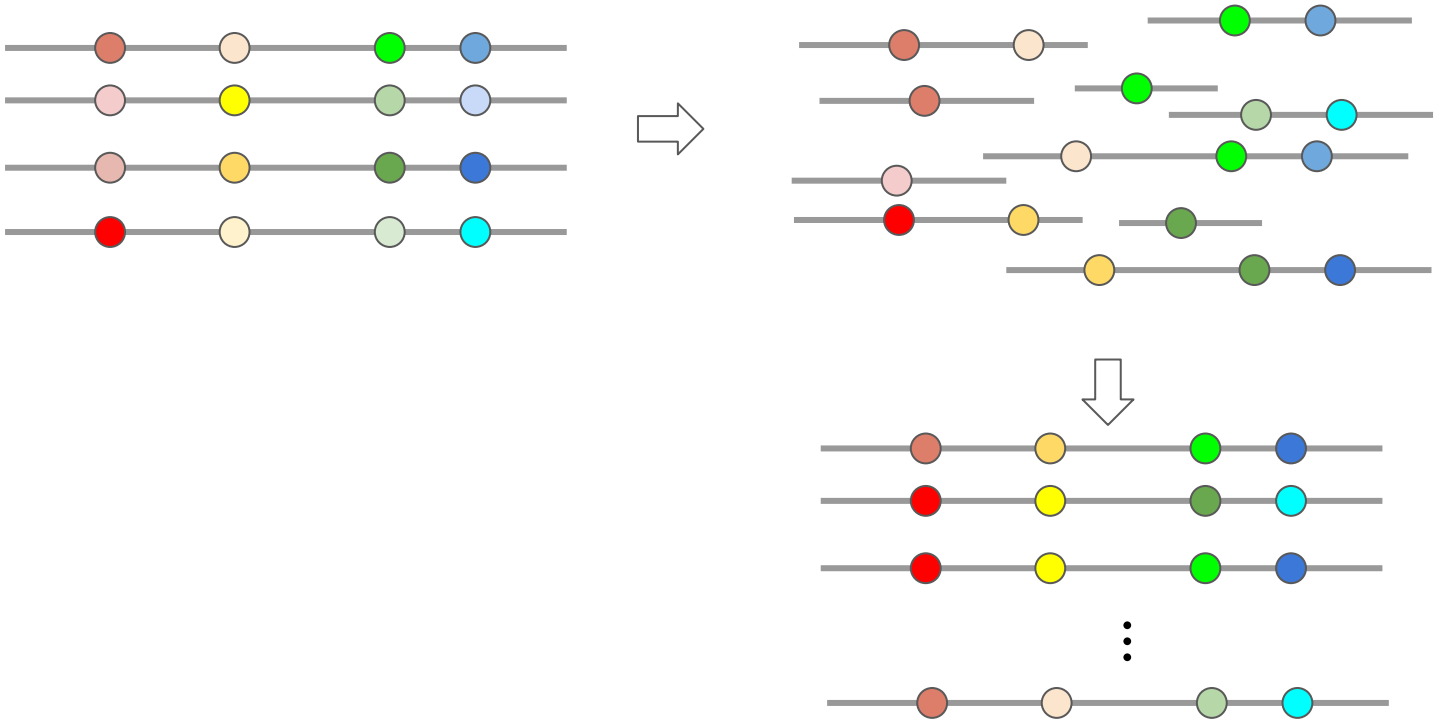
Gene shuffling is one method for generating chimeras



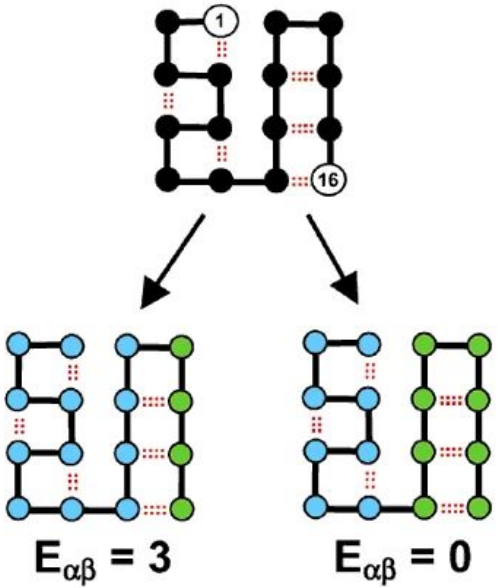
Gene shuffling is one method for generating chimeras



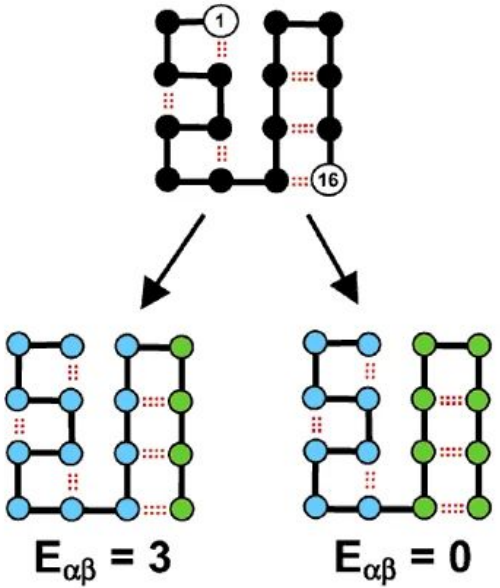
Gene shuffling is one method for generating chimeras

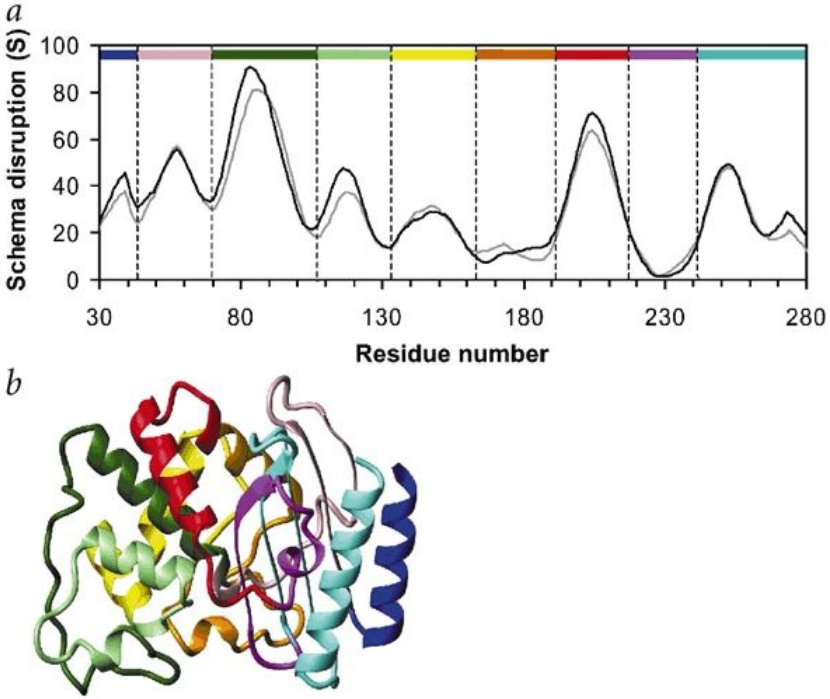


SCHEMA is a computational-powered method to generate chimeric proteins



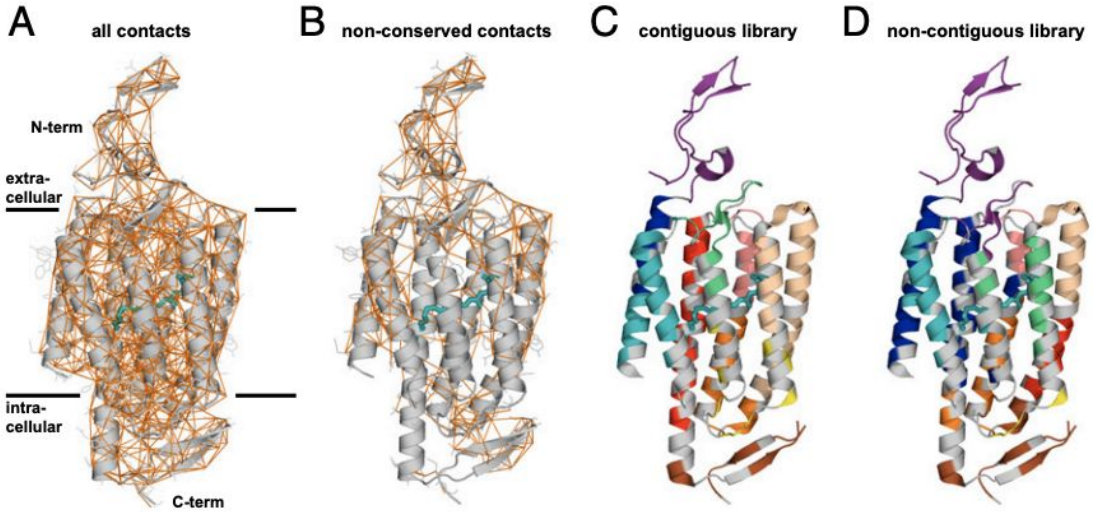
SCHEMA is a computational-powered method to generate chimeric proteins



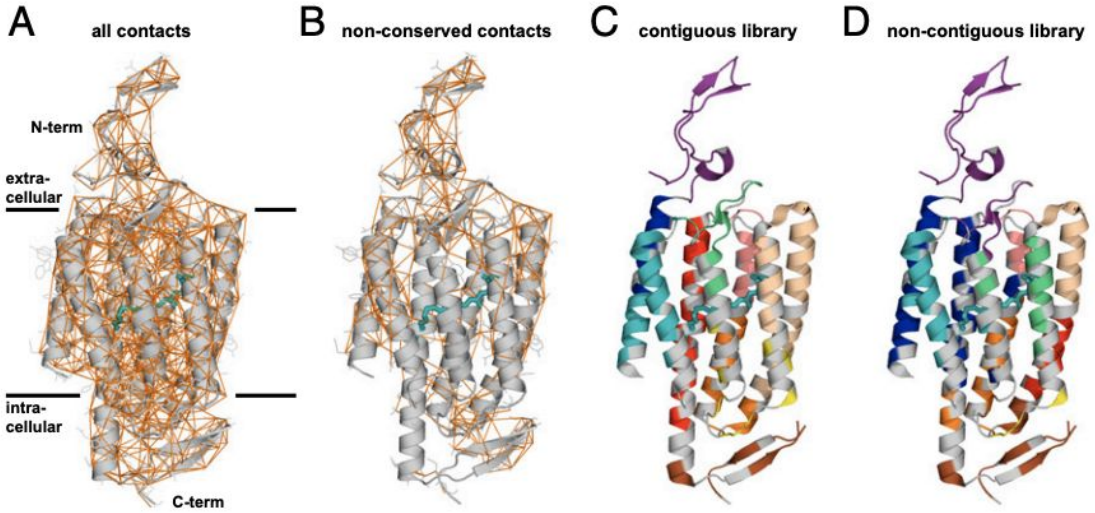


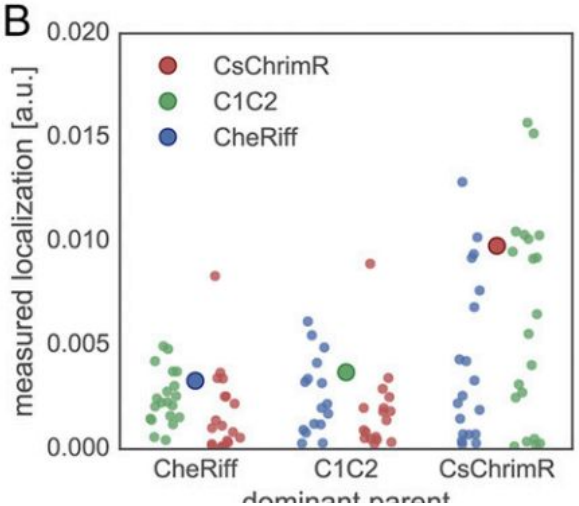
Nat Struct Biol 2002, **9,** 553-558

SCHEMA can be used to generate chimeric channelrhodopsins

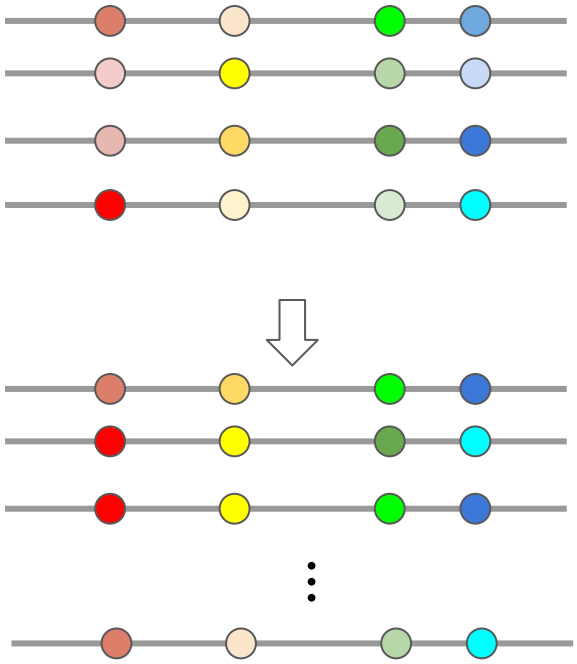


SCHEMA can be used to generate chimeric channelrhodopsins



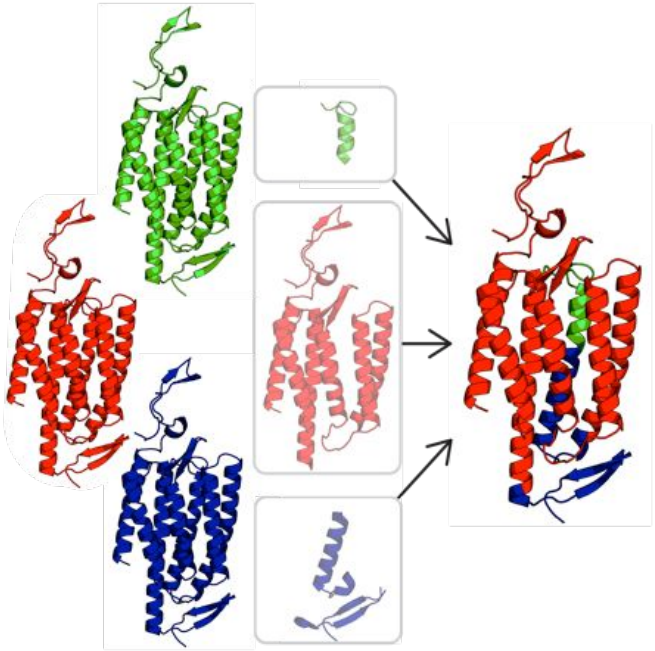


Shuffling is another powerful method for large families of proteins



Advantage	Disadvantage
Can help generate new functions/properties in a known fold	Difficult/tricky to run
Border scope than ISM	Require lots of homologous sequences
You can get different modifications that are hard to sample by ISM or random methods	There is a chance they don't fold
You get lots of functional random mutations for free (nature already selected)	

In class activity:



Shuffle your sequences to evolve their function

1. You may not get to the global maximum no matter how much you souffle depending on your starting sequences

Image from Frances Arnold's group webpage

Why can’t we predict these mutations?

Why can’t we predict these mutations?

1. Missing links in our knowledge of structure-function relationship
2. Epistatic effects of mutations
3. Effects of stability on overall performance
4. Intractable large sequence space

For the next lecture:

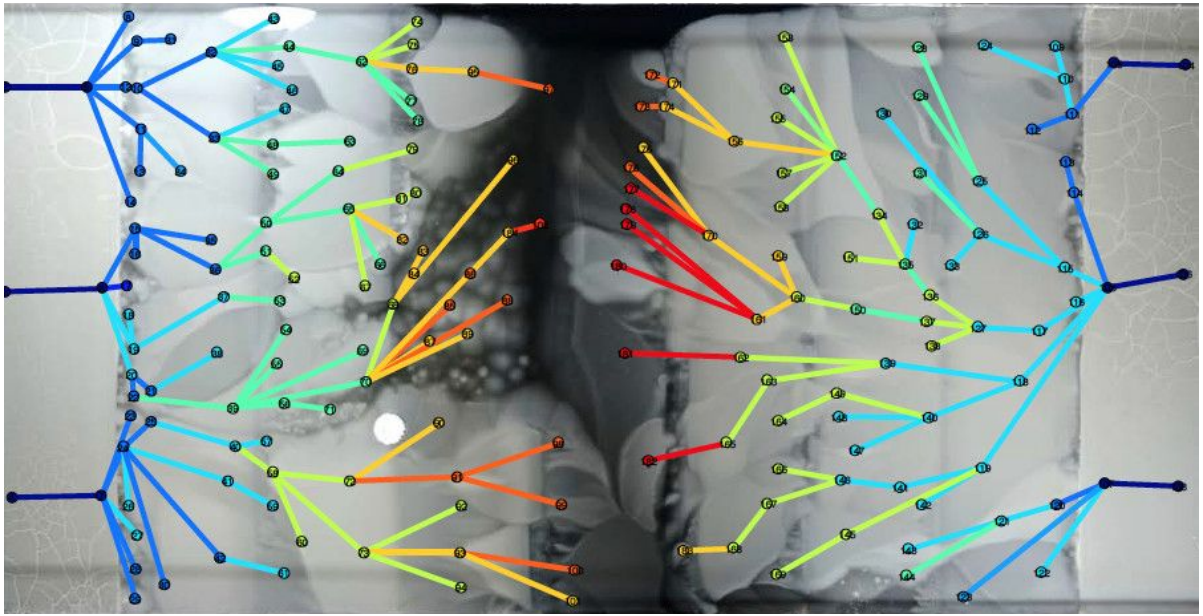
1. Pre-class assessment for the next lecture
   None

2. Post-class assignment
   The one from W2L2 due next lecture

3. Second journal: Will be discussed next week

Next lecture:

You get what you select for …



The molecular ecologist: “Visualizing the evolution of bacterial resistance”