BP 605 T. Pharmaceutical Biotechnology (Theory)

Brief Introduction to Protein Engineering

Dr Chandresh Sharma

Assistant Professor

Department of Biotechnology

Chhatrapati Shahu Ji Maharaj University, Kanpur

Overview

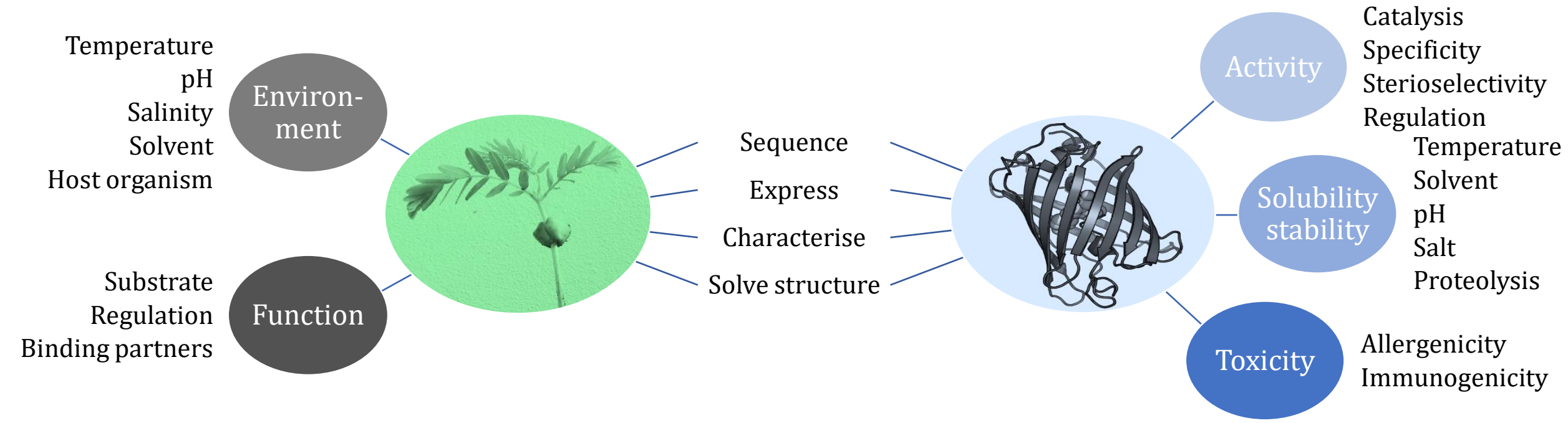
- What protein engineering is
- Protein Engineering Methods
- Protein Engineering and Applications

PROTEIN ENGINEERING

Protein engineering: Techniques which are used to manipulate the structure and function of a protein so that it acquires specific desired properties.

Genetic engineering: The alteration of the genome of an organism by laboratory techniques

AIMS OF PROTEIN ENGINEERING



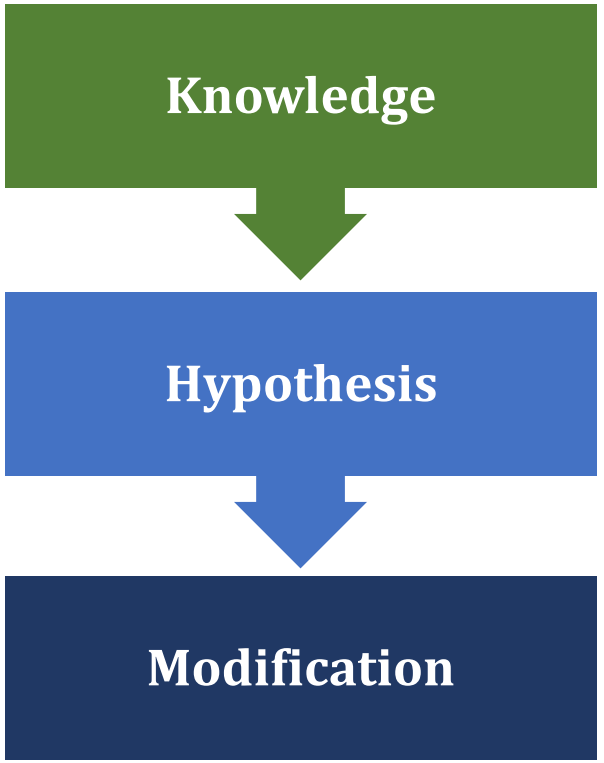
Natural evolution | Analysis | Protein engineering

APPLICATIONS OF PROTEIN ENGINEERING

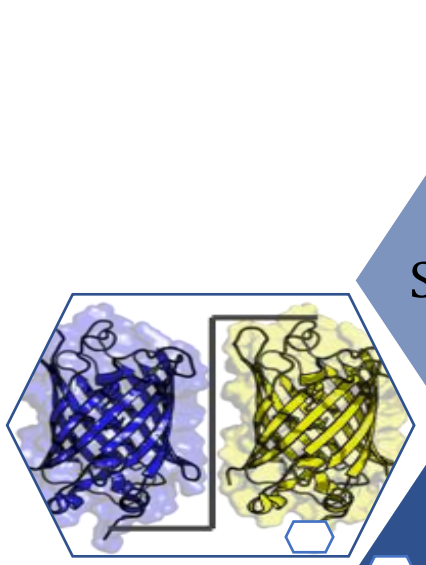
Images: Wikimedia commons

THE RATIONAL DESIGN PROCESS

- Based on protein knowledge
  - Structure
  - Mechanism
  - Dynamics
  - Natural variation
- Analogous to mechanical engineering

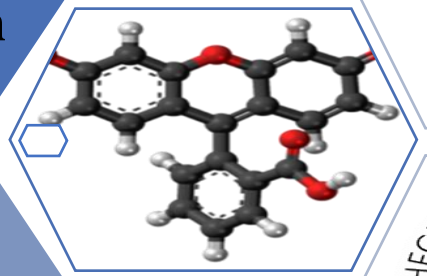


MODIFICATIONS

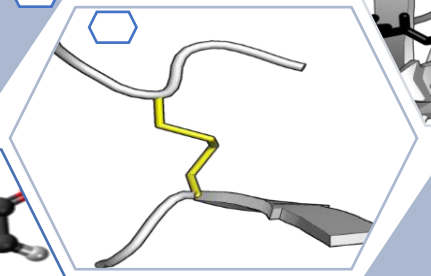


Split proteins

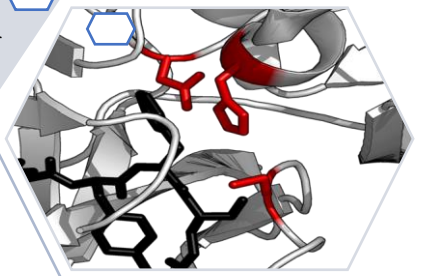
Chemical modification



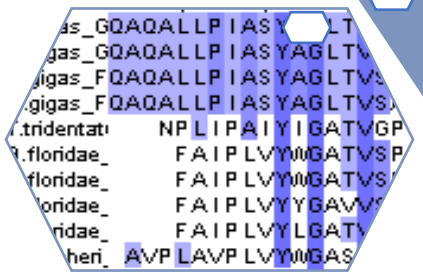
Disulphides



Active site modification

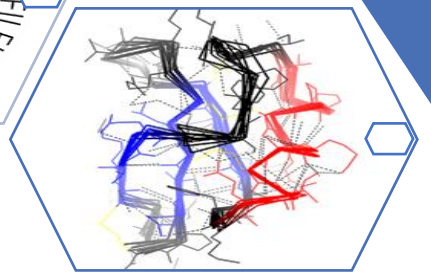


Fusion proteins



Bio-informatics

Cyclisation

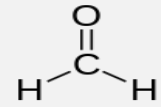


Computer modelling

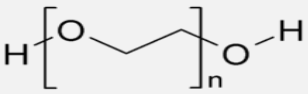
CHEMICAL MODIFICATION

- **Formaldahyde**
  - Extensive modification ⟶ Inactivated toxoid production
- **PEGylation**
  - Flexible hydrophilic coat ⟶ Solubility
  - Reduced accessibility ⟶ Protease resistance and non-antigenicity
  - Increased size ⟶ Serum half-life
- **Fluorophores**
  - Fluorescent labelling ⟶ Tracking location or dynamics
- **Prosthetic catalytic groups**
  - Modified reactivity ⟶ Altered or novel catalysis
- **Considerations**
  - Exposure of modified residues
  - Original function of modified residues

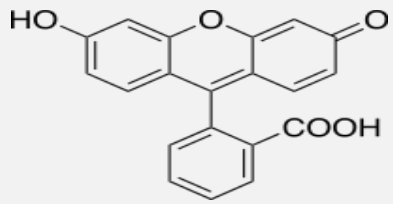
Formaldehyde



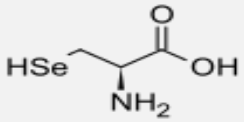
Poly-ethylene glycol



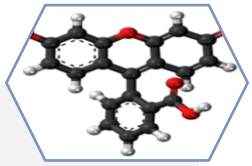
Fluorescein



Selenocysteine



Uricase immunogenicty



Aim

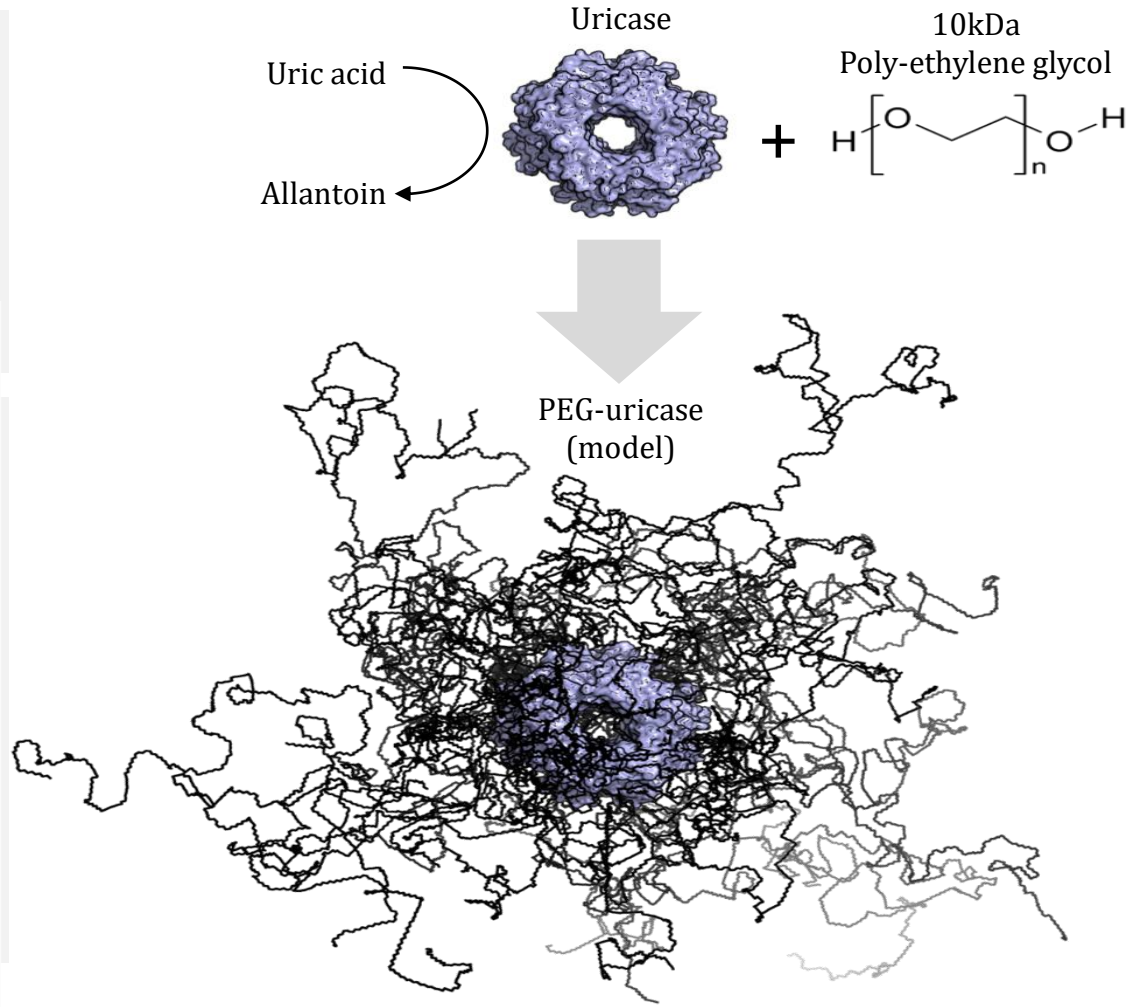
- Optimise uricase as gout treatment
  - Reduce immunogenicity
  - Increase serum half-life

Engineering

- Attached PEG polymers
  - Lysine coupling
- Optimised PEG number and length
  - Maximise improvements
  - Avoid destabilisation or activity reduction

Outcome

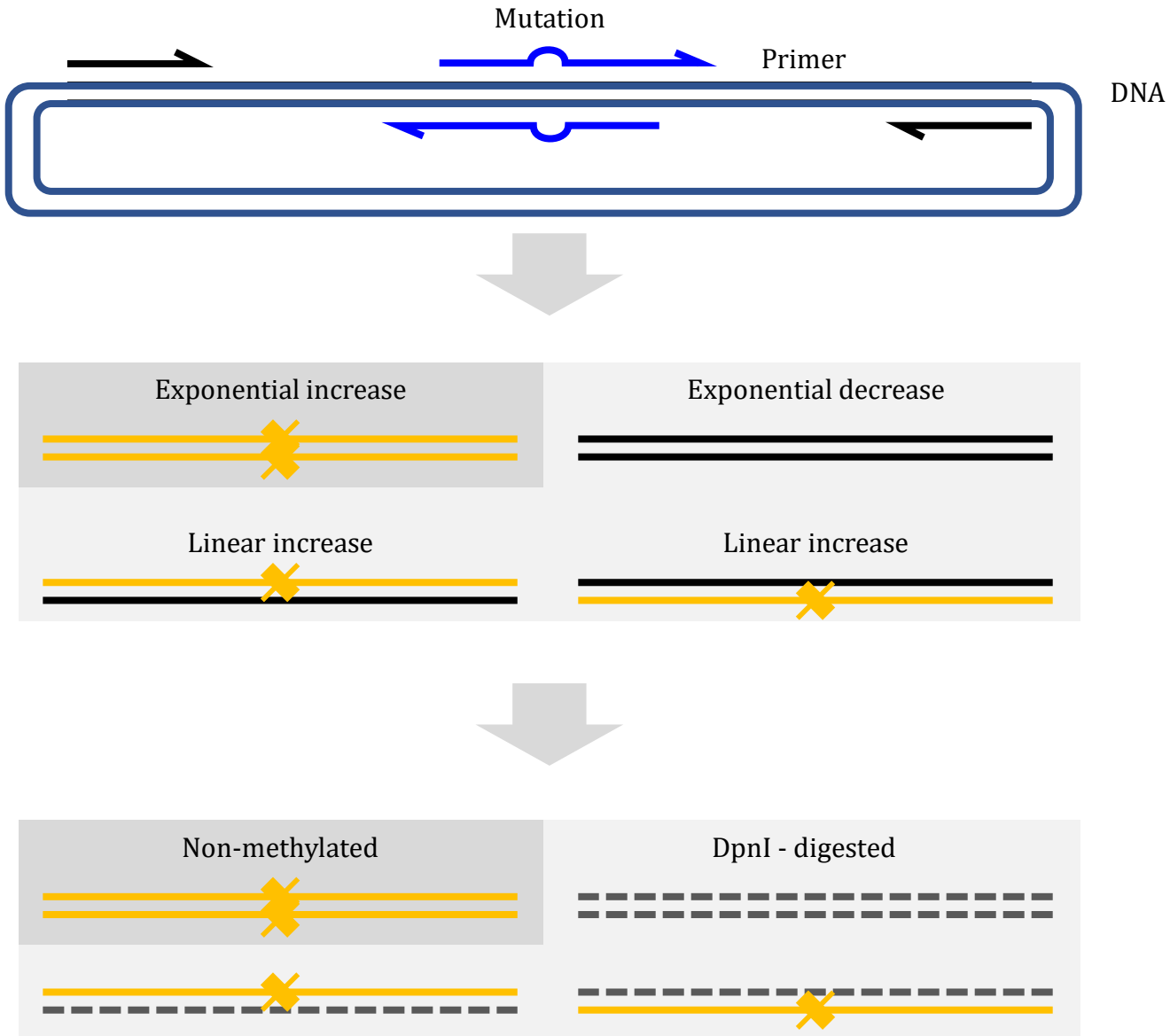
- Optimal PEG number and length
  - 10kDA polymers
  - 9 polymers per subunit of the tetramer
- 1000x reduced antigenicity
  - Also improved solubility at neutral pH
  - Also increased serum half-life
- Krystexxa (Crealta Pharmaceuticals)



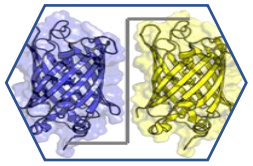
PEG-uricase in the management of treatment-resistant gout and hyperuricemia, (M Sherman 2008)
Images: Wikimedia commons

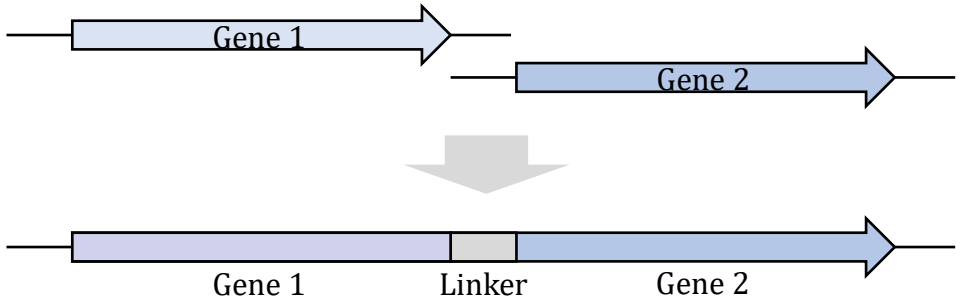
Site-directed mutagenesis

- Modified PCR
  - Whole plasmid
  - Overlap extension
- Introduce point mutations
- Introduce *short* insertions or deletions



Fusion proteins



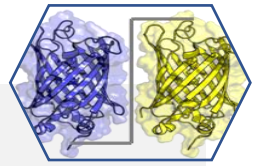


- Creation
  - Remove stop codon of first gene
  - Ligate genes together in frame
  - Include linker codons
- Aims
  - Combine the properties of the components
    - E.g. Addition of antibody Fc fragment to proteins increases their serum halflife
  - Co-localise the components
    - E.g. Set of enzymes that work in a reaction pathway
- Considerations
  - Linker length and flexibility
    - Ability for proteins rotate relative to each other
    - Distance between protein components
    - Protease resilience
    - Ability for domains to fold

Pfu polymerase Processivity

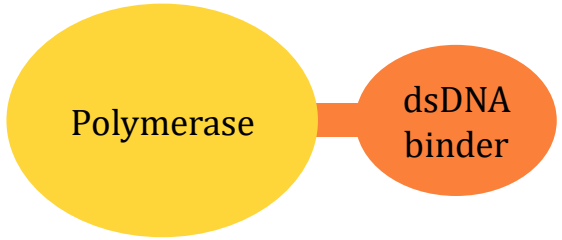
AIM

- Create a polymerase for long templates
  - Increase processivity
  - Retain fidelity and stability



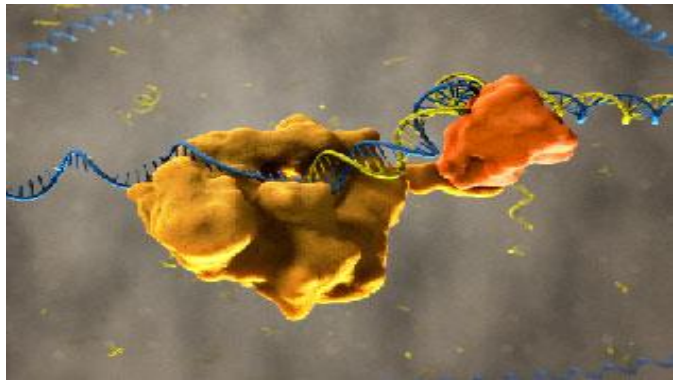
ENGINEERING

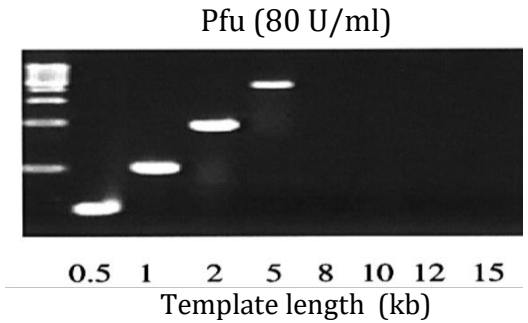
- Fusion
  - *Pyrococcus furiosus* DNA polymerase (Pfu)
  - *Sulfolobus solfataricus* dsDNA binding domain (Sso7d)
- Linker
  - Short tripeptide linker
- Generality
  - Also works with other polymerases

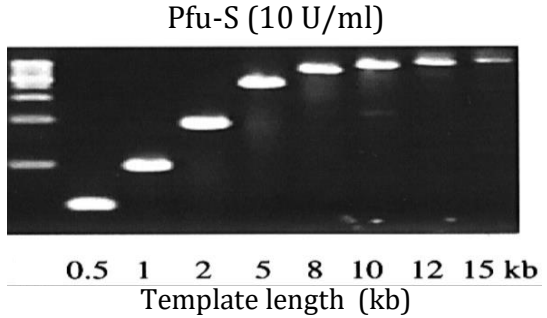


OUTCOME

- Improvements
  - 10x increase in processivity
  - Improved salt tolerance
  - Can amplify >15kb templates
- Phusion (New England Biolabs)



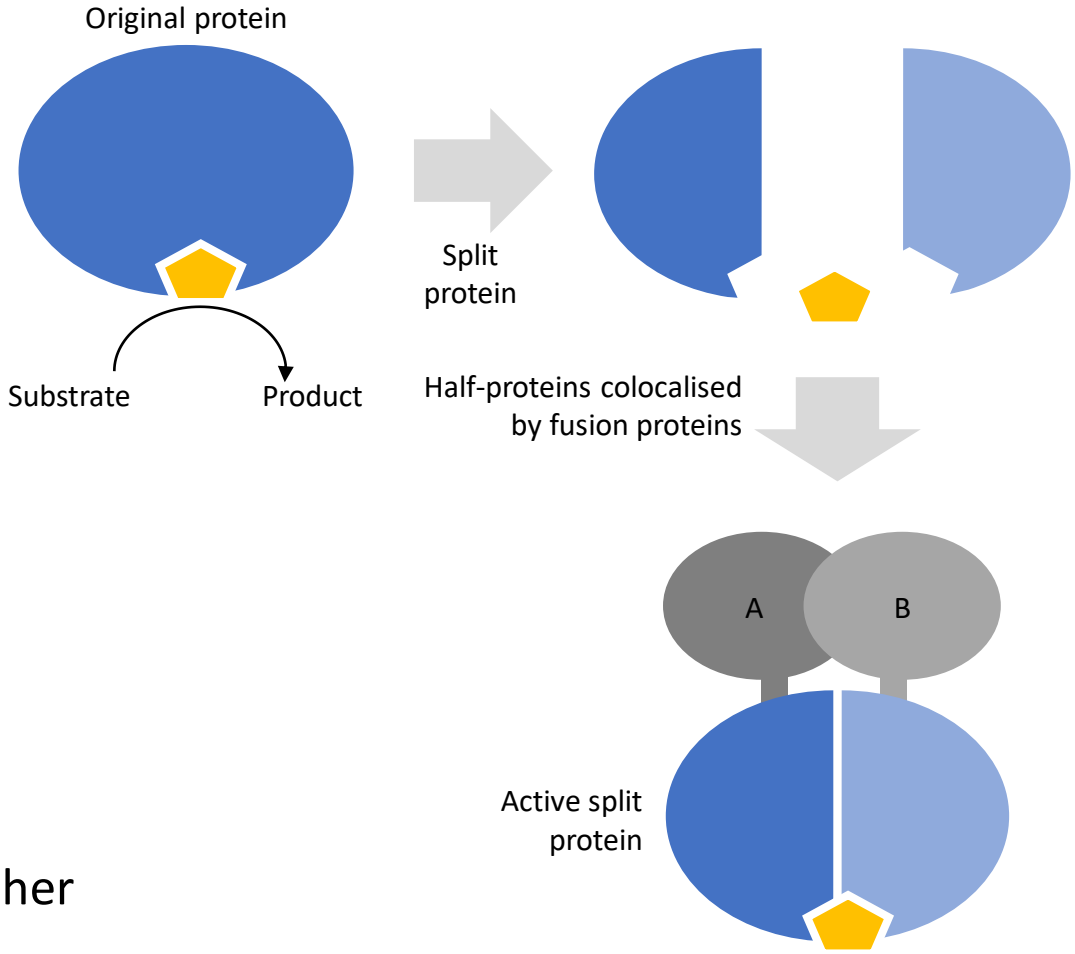




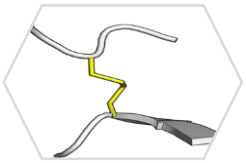
A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro, (Y Wang 2004)
New England Biolabs

Split proteins

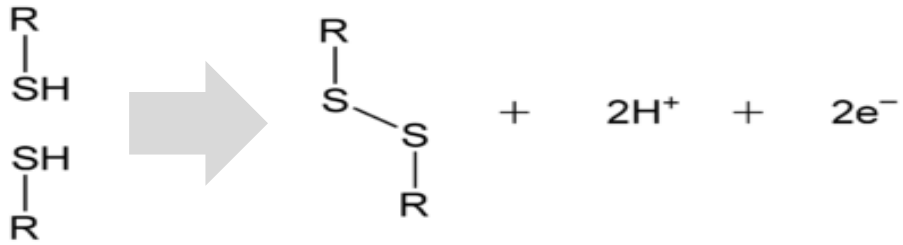
- Creation
  - Locate flexible, surface loops
  - Create two open reading frames
    - First half of protein with stop codon in loop
    - Second half of protein with start codon in loop
- Aims
  - Couple colocalisation to activity
  - Fuse half-proteins to other proteins
    - Measure protein binding
    - Biosensor
    - Logic gates
- Considerations
  - Half-proteins must: fold independently
  - not spontaneously ‘dimerise’
  - be inactive when apart
  - bind and be active when brought together



Disulphides

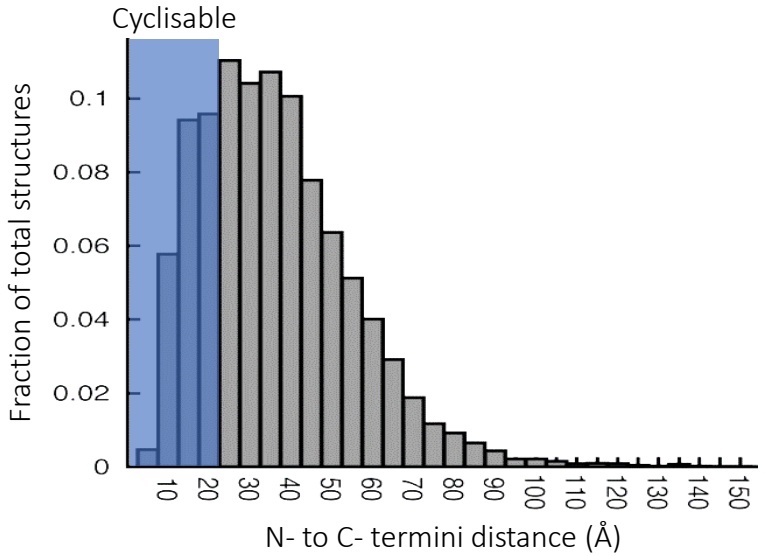


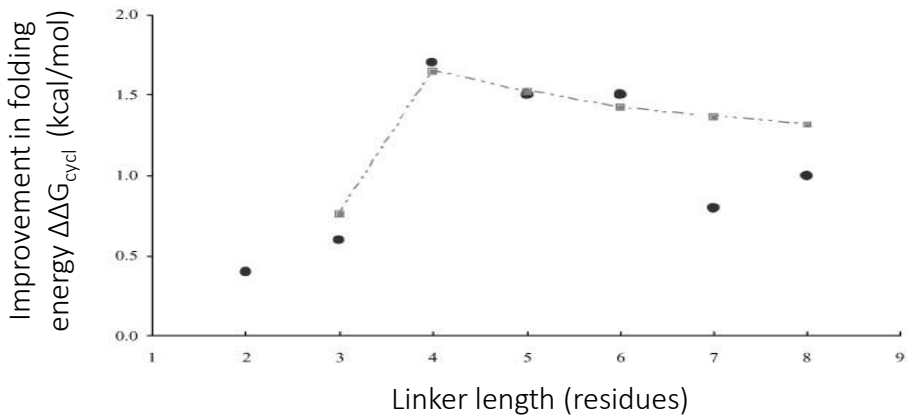
- Creation
  - Mutation of two codons to cysteine
  - Protein kept in oxidising environment
- Aims
  - Stability enhancement
    - Enthalpy increase ≈ 3.5 kcal/mol
    - Entropy decrease ≈ Logarithm of trapped loop length
- Considerations
  - Inter-cysteine distance
  - Inter-cysteine orientation
  - Trapped loop length and flexibility
  - Original function of mutated residues
  - Original function of flexibility
  - Folding pathway of protein (multistep)



Cyclisation

- Creation
  - Termini of most proteins happen to be close together
  - Express protein with extra linker to bridge gap
  - Ligate peptide ends
- Aims
  - Thermostability
    - Up to 1.7 kcal/mol
  - Protease resistance
    - Especially exopeptidase
- Considerations
  - Linker length
  - Ligation method
    - Chemically (e.g. by solid-phase synthesis)
    - Enzymatically (e.g. by sortase)





CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering, (C Wang 2008)
Effect of Backbone Cyclization on Protein Folding Stability: Chain Entropies of both the Unfolded and the Folded States are Restricted, (H Zhou 2003)

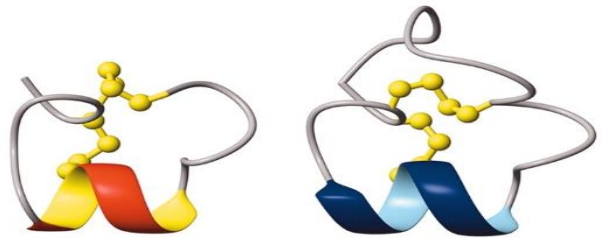
Conotoxin stability

Aim

- Increase conotoxin protease resistance
  - Pain killer activity by specific binding to ion channels
- Improve stability in human blood

Engineering

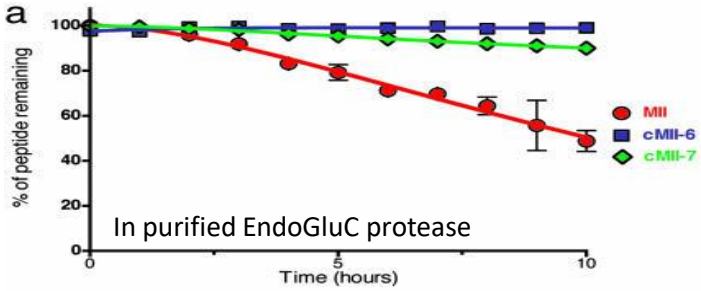
- Produced whole peptide by solid-phase synthesis
  - Linker length of 5, 6, or 7 residues
  - cMII-5, cMII-6, cMII-7

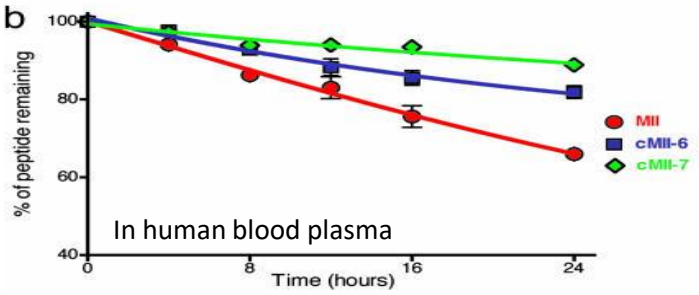


Original **MII** Cyclic c**MII-6**

Outcome

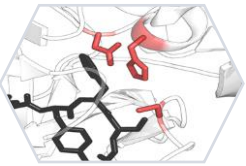
- cMII-5
  - No longer folded or functional
- cMII-6 and cMII-7 retained full activity
  - Specific ion channel blocking
  - Minimal structural difference
- Reduced protease susceptibility
  - With purified EndoGluC protease (a)
  - In human blood plasma (b)



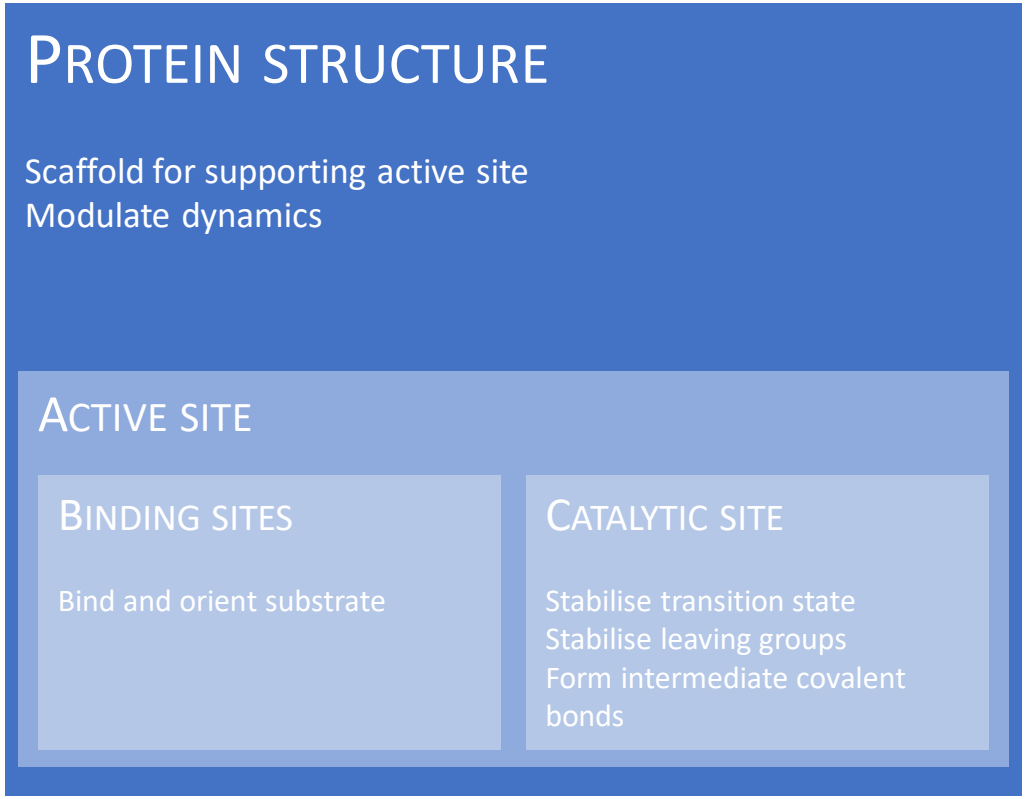


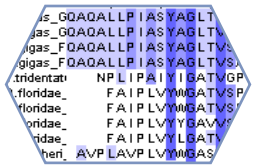
Engineering stable peptide toxins by means of backbone cyclization: stabilization of the alpha-conotoxin MII, (R Clark 2005)

Active site modification



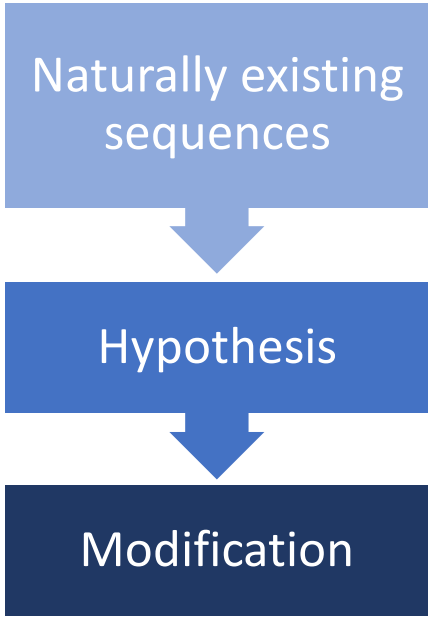
- Creation
  - Structural insight into function of active site residues
  - Site-directed mutagenesis to alter key functional groups
- Aims
  - Modify binding
    - Affinity
    - Specificity
    - Sterioselectivity
  - Modify catalysis
  - Modify regulation
- Considerations
  - Requires knowledge of protein structure and mechanism
  - Mutations may have additional, unpredicted effects



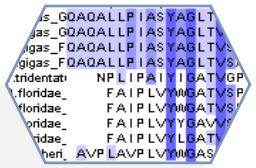


Bioinformatic approaches

- Codon optimisation
  - Different organisms have different tRNA ratios
  - Matching codon frequency to host increases expression
- Considerations
  - Altered codons can affect mRNA (stability, 2° structure, IRES)
  - Increased translation rates can cause misfolding
- Consensus sequence
  - Most mutations are mildly destabilising
  - Through genetic drift, homologues accumulate different mutations
  - Therefore consensus should be more stable than existing sequences
- Considerations
  - Availability of homologous sequences



Phytase stability



Aim

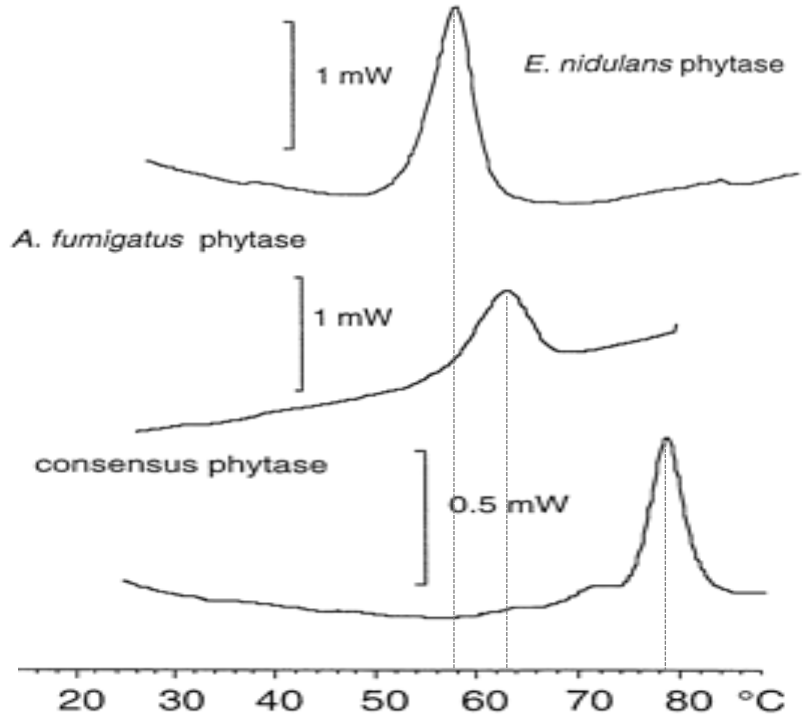
- Improve phytase thermostability
  - Improving phosphorous bioavailability in animal feed

Engineering

- Align 13 related fungal sequences
  - Sequences 50 - 70% identical to each other
  - If no consensus in column ⟶ most common residue (*)
  - ⟶ residue from most stable (^)
- Starting thermostabilities (T_M) 56 - 63 °C

Starting sequences

```
QDLGAQFYRR YDTLTRHINP FVRAADSSRV HESAEKFVEG
VNSGIKFYQR YESLTRNIVP FIRSSGSSRV IASGKKFIEG
VNSGIKFYQR YKALARSVVP FIRASGSDRV IASGEKFIEG
VDSGAKFYRR YKNLARKNTP FIRASGSDRV VASAEKFING
IQLGIKFYNH YKSLARNAVP FVRCSGSDRV IASGRLFIEG
VNSGIKFYRR YRALARKSIP FVRTAGQDRV VHSAENFTQG
                  *       ^
Con.  VNSGIKFYRR YKALARKIVP FIRASGSDRV IASAEKFIEG
```

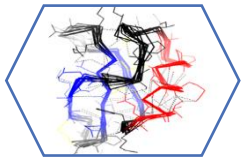


Outcome

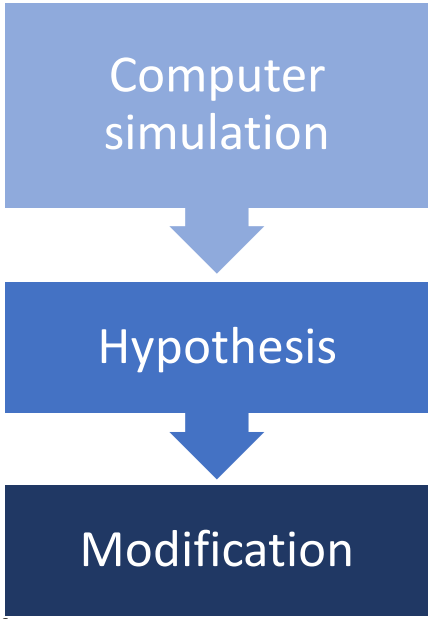
- Final T_M = 78 °C
  - Crystal structure resolves loops too flexible to be seen in natural phytases
  - Some residues form hydrogen bond network
- Later work further increased T_M to 90°C
  - Added 6 extra sequences to alignment
  - Changed consensus residues that weren't stabilising

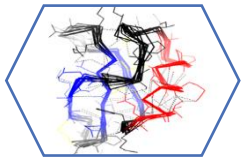
From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase, (M Lehmann 1999)

Computational modelling



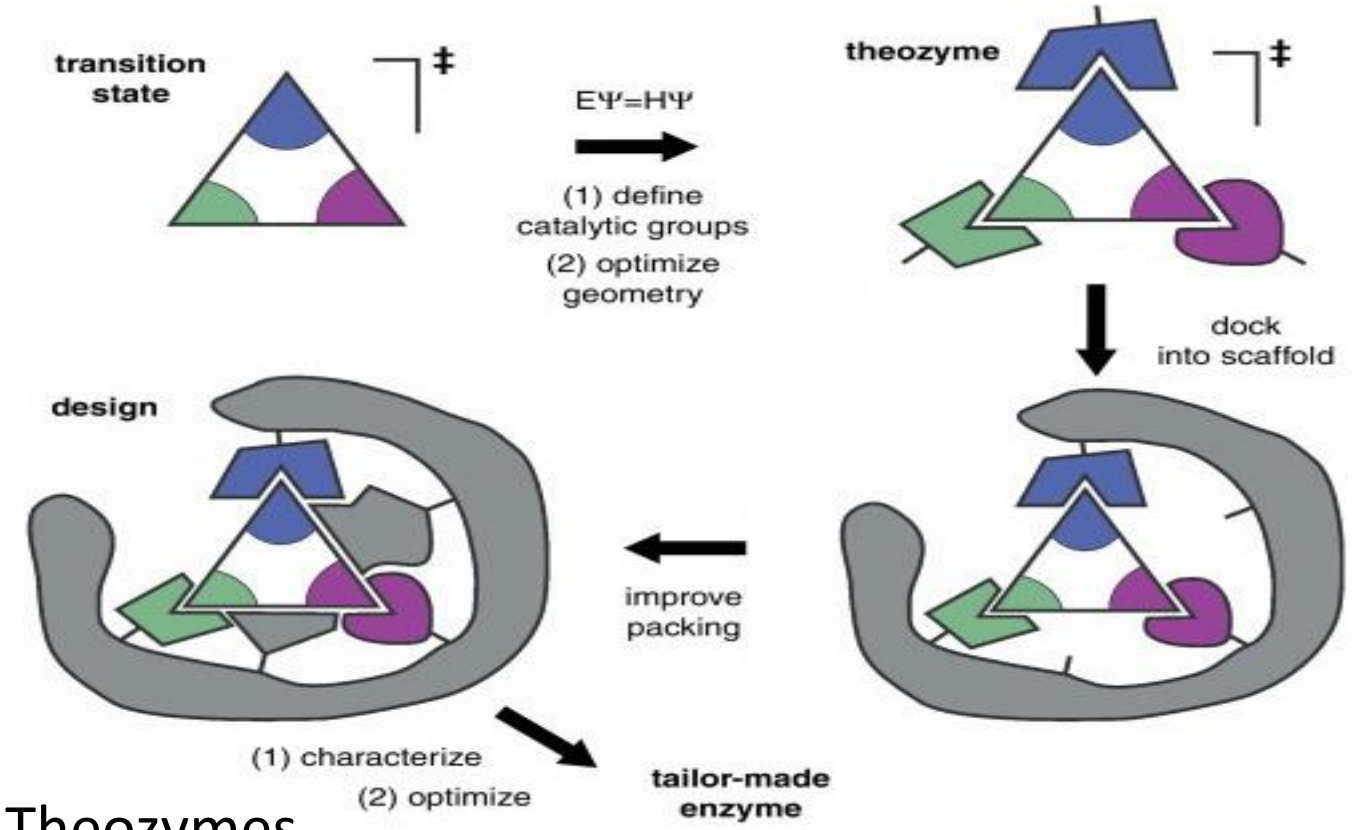
- Improving stability
  - Model energy of folded and unfolded protein variants
- Improving activity
  - Increase existing catalysis
  - Catalyse new reactions, never seen in nature
    - e.g. Kemp elimination or Retro-aldol
- Considerations
  - Requires deep knowledge of reaction mechanism
  - Requires extreme computational power
  - Simulation either ignores: quantum mechanism of active site
  - or structure and dynamics in rest of protein





De novo enzyme design

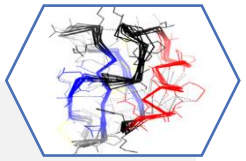
- Disembodied amino acids placed to stabilise reaction transition state
- Existing protein structures searched for backbones with correct orientations
- Other residues in active site optimised for packing



- Theozymes
  - Theoretical enzyme
  - Quantum mechanical modelling

De novo enzymes by computational design, (H Kries 2013)

Creating a retro-aldolase

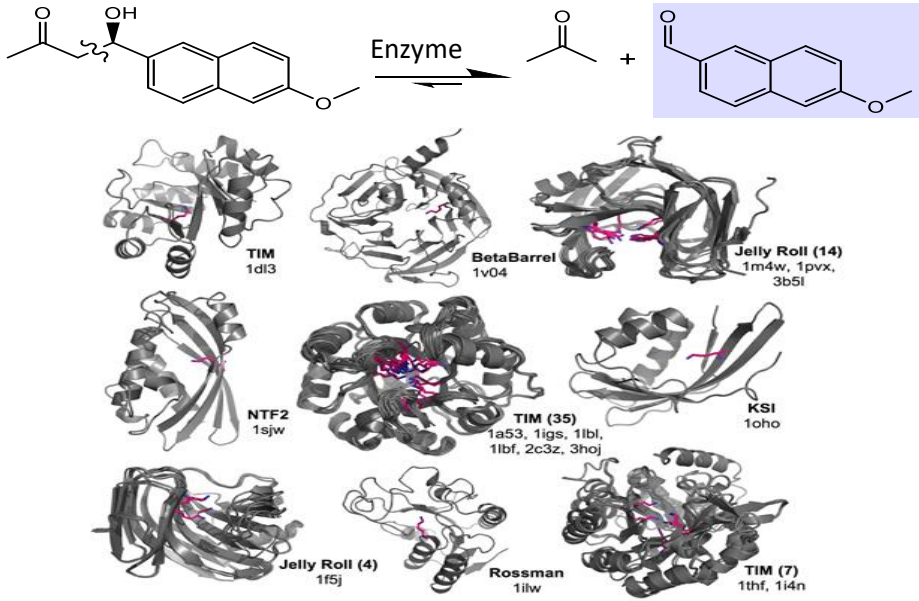


Aim

- Enzymatically catalyse unnatural reaction
  - Retro-aldol reaction not performed by any known enzyme

Engineering

- Theozyme
  - Amino acids positioned to increase reactivity of nucleophilic Lys, stabilise transition state, stabilise leaving group
  - Protein structures searched for backbones that could correctly position these residues
  - Surrounding residues optimised for packing
- 42 designs in 13 protein scaffolds
  - Active sites grafted onto backbone
  - Genes synthesised and expressed



Outcome

- 75% of variants showed rate enhancements 10^1-10^4 k_{cat}/k_{uncat}
  - Still many orders of magnitude worse than natural enzymes
- Crystal structure of most active complexed with covalent inhibitor
  - Confirmed mechanism proceeds as designed

Robust design and optimization of retroaldol enzymes, (E Althoff 2012)

Pros and cons of rational design

Benefits

- Intellectually satisfying
- Controlled outcome
- Range of available techniques
- Increasing computational power

Limitations

- Requires deep understanding
  - Natural variation
  - Structure
  - Dynamics
  - Mechanism
  - …for starting protein *and* changes
- High failure rate
  - Failures rarely reported

Pharmaceutical Biotechnology

Concepts and Applications

Gary Walsh

University of Limerick, Republic of Ireland

John Wiley & Sons, Ltd

For Query

chandreshsharma@csjmu.ac.in; sharmac3001@gmail.com