Reconstructing ancestral sequences through a combined bioinformatics and molecular modelling approach

Subha Kalyaanamoorthy | OCE Post-Doctoral Fellow

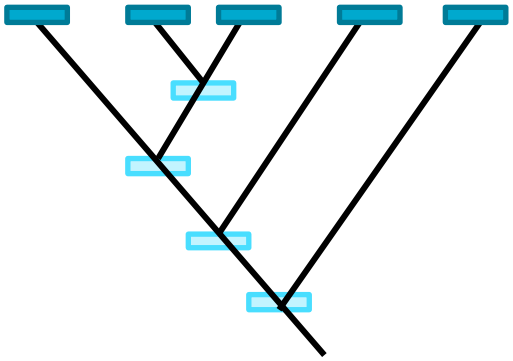
07 November 2013

CSIRO ECOSYSTEM SCIENCES
www.csiro.au

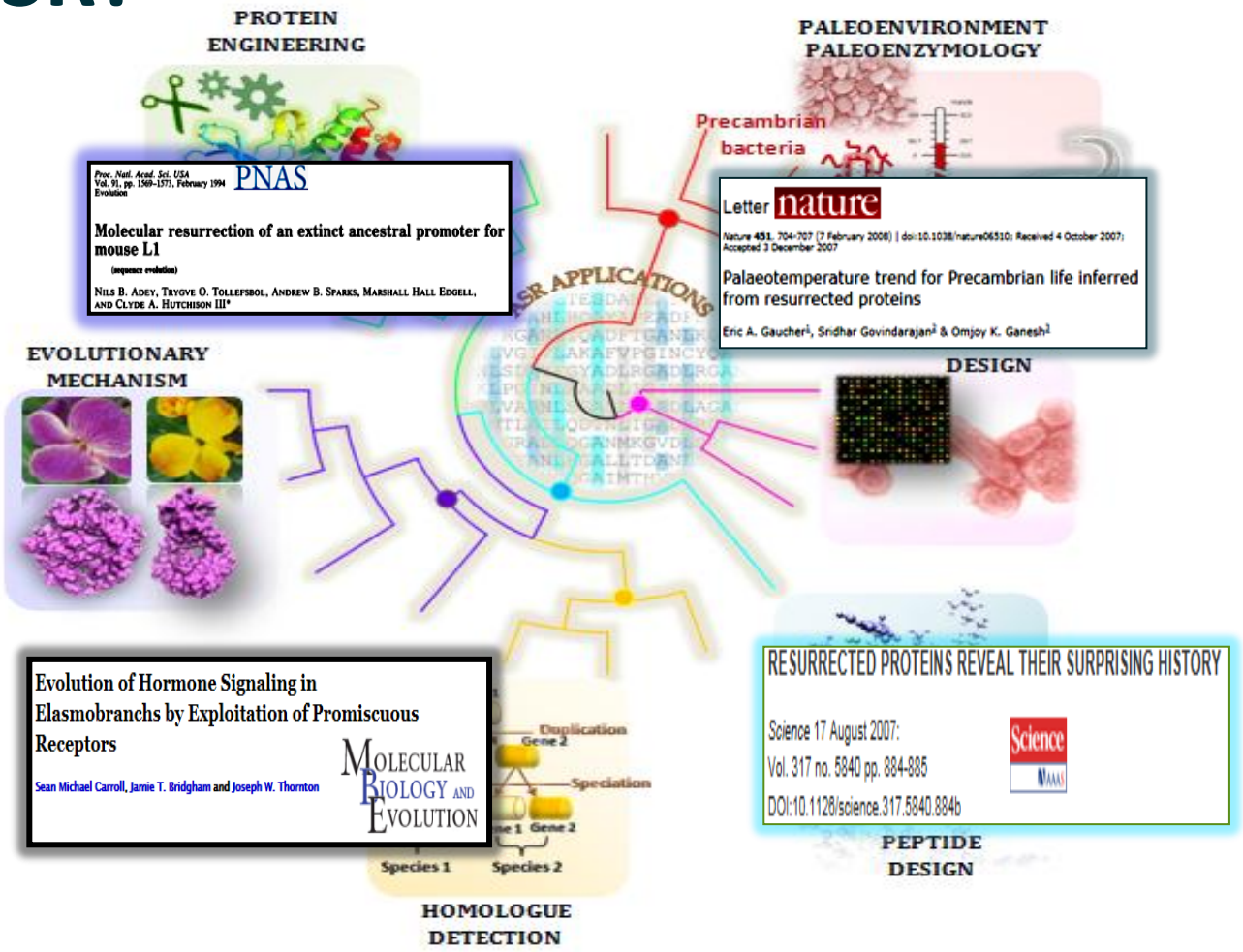
Ancestral Sequence Reconstruction

- A rearward approach that uses the evolutionary relationships between the extant sequences to infer their common ancestral state, which no longer exists in earth

- Replaying the evolutionary trajectory to trace the molecular trials through which the present day sequences have evolved has now become possible through phylogenomic approaches



Why ASR?



Applications of ASR

Image Source: Manuscript in preparation

Current ASR projects

- We are reconstructing the ancestral states of biologically and/or industrially useful enzymes, including **hydrolases** and **oxidoreductases**, in order to understand their evolutionary process and engineer them for various 'present-day' applications..

- Phylogenetic methods and advanced molecular modelling approaches, in corroboration with experiments, are employed in the reconstruction

Protocol

- Approximately 60 extant sequences from different insect species were collected from various sources and aligned
- The alignments are visually assessed and their completeness were analysed using Alistat
- The extant sequences were verified if they have evolved under stationary conditions using Homo.
- The best-fitting evolutionary model was selected and the phylogenetic tree was constructed.
- Given the tree and the evolutionary model, the ancestral states were estimated using maximum likelihood methods

Maximum Likelihood

- The ML method calculates the ancestral character state based on the likelihood scores (i.e., the probability that the extant sequences would evolve from the inferred ancestral state), given a phylogenetic tree and an explicit stochastic model to describe the evolution.

- Two types of methods:

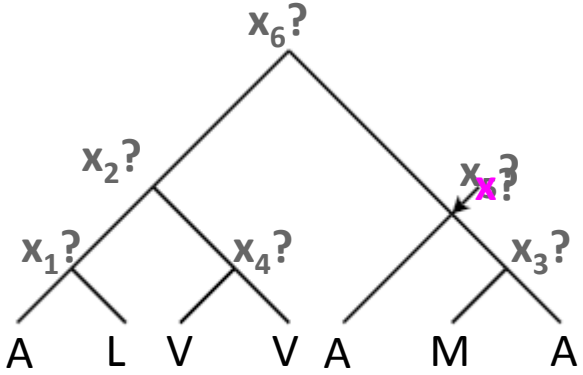
  Joint and Marginal reconstruction

Joint and Marginal

Joint:

Identifies the most probable set of character states for all internal nodes at any given site, resulting in the maximum joint likelihood of the tree

$$\Pr(x_{1\text{-}6} \mid D,\theta) = \frac{\Pr(x_{1\text{-}6} \mid D, \theta)}{\Pr(D, \theta)}$$

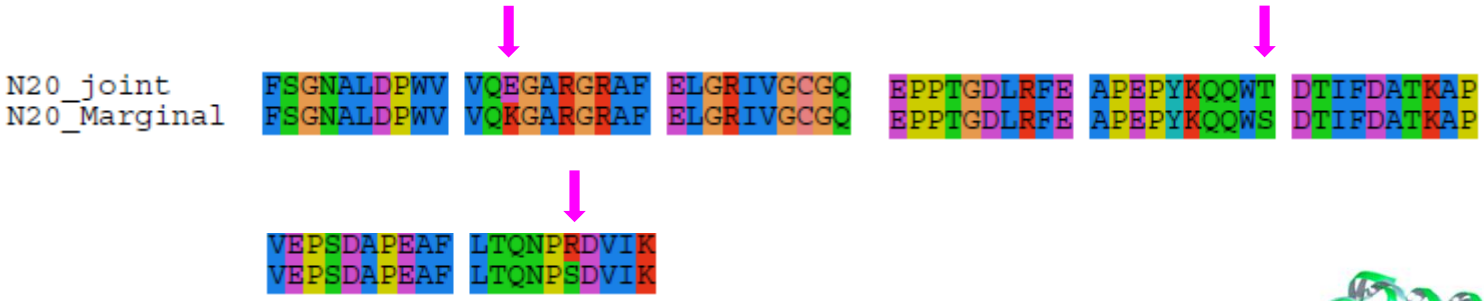


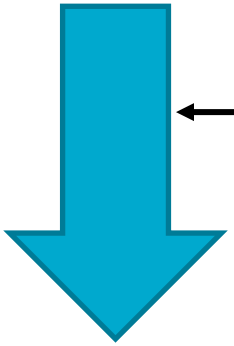
Marginal:

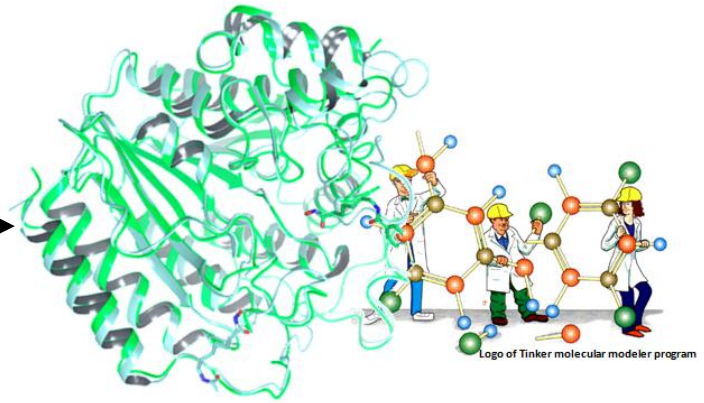
Identifies the most probable character state for a particular site of the tree, by comparing the likelihoods of different character states at the given site

$$\Pr(x \mid D,\theta) = \frac{\Pr(x \mid D, \theta)}{\Pr(D, \theta)}$$

Reconstructed ancestral states

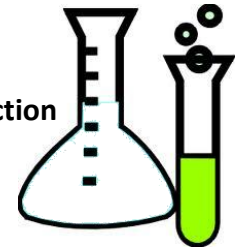






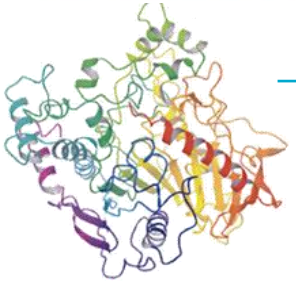
Molecular modelling and dynamics-based screening!!!

In-vitro resurrection



Ancestral sequence reconstructed using the joint method

..... FSGNALDPWV VQEGARGRAF ELGRIVGCGQ ..... EPPTGDLRFE APEPYKQQWT DTIFDATKAP VEPSDAPEAF LTQNPRDVIK .....



The template X-ray crystal structure with 28% identity with ancestral sequence.

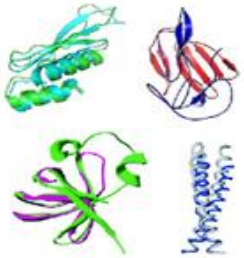
Homology modeling

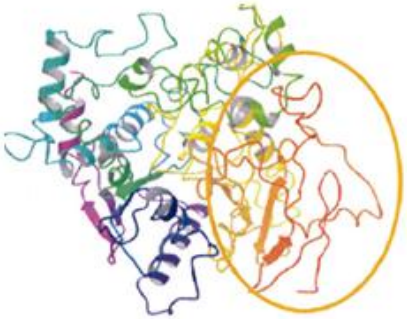
(Evolutionarily related proteins may have similar structures)

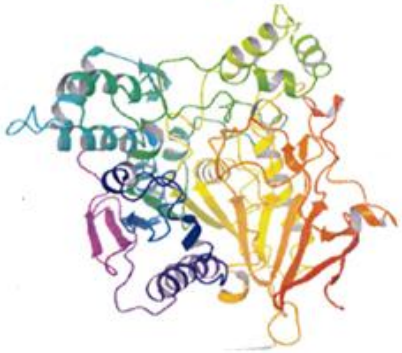
<30% identity; not evolutionarily related

Threading

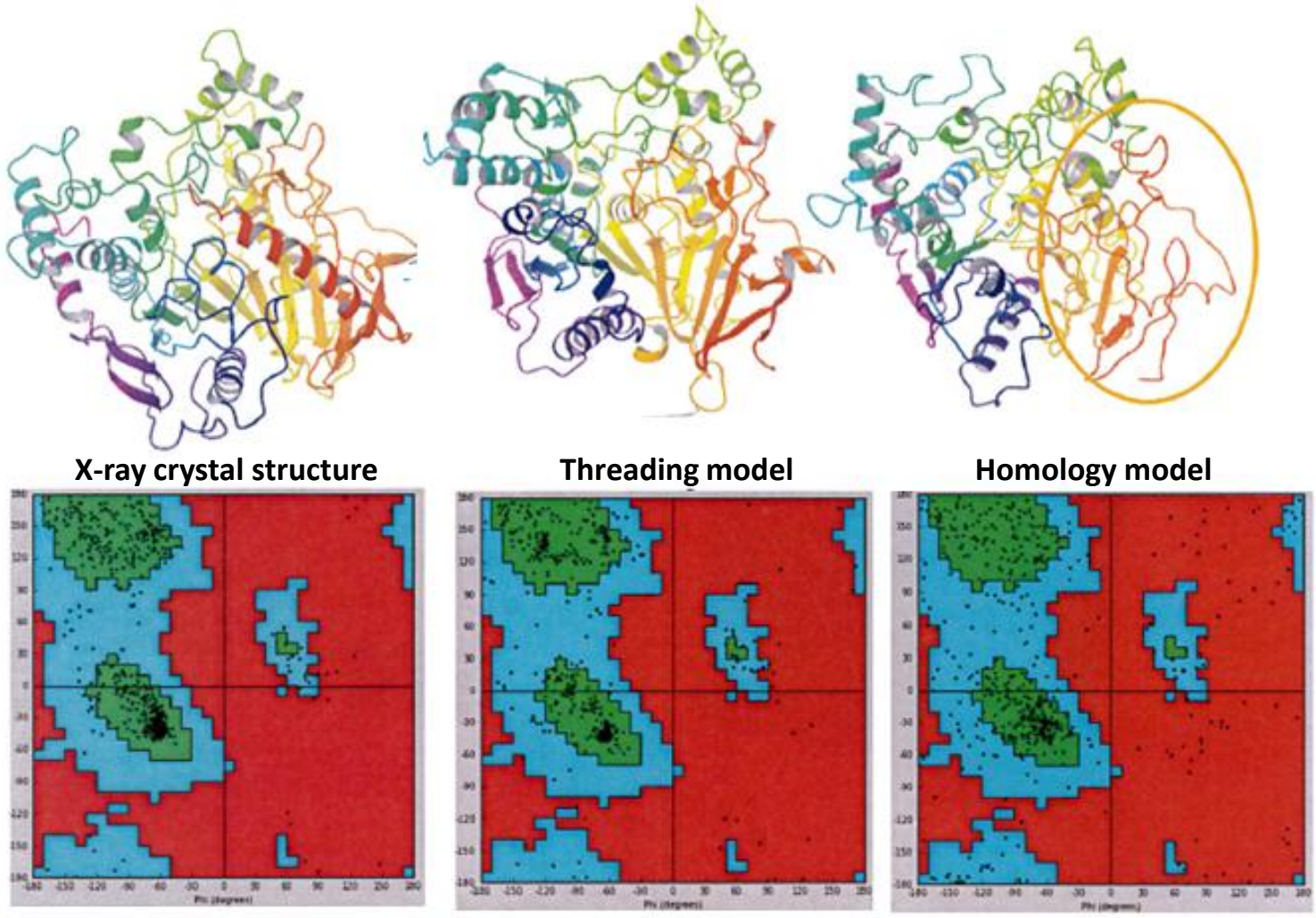
(Most proteins share similar folding patterns irrespective of their evolutionary relationships)



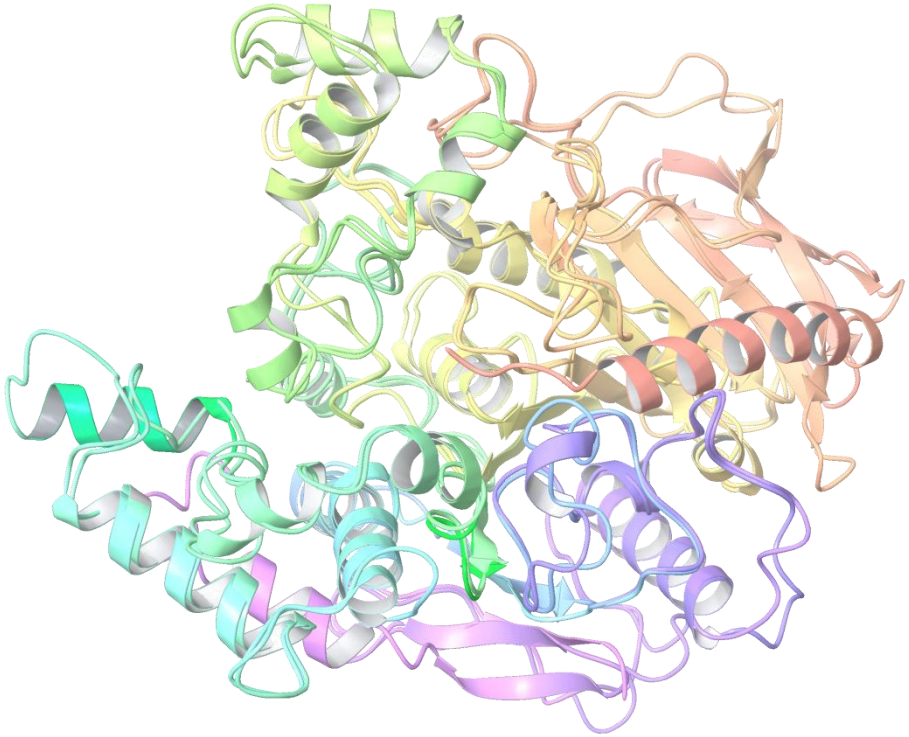




Model quality assessment



Superimposed ancestral model and extant homologue (crystal structure)



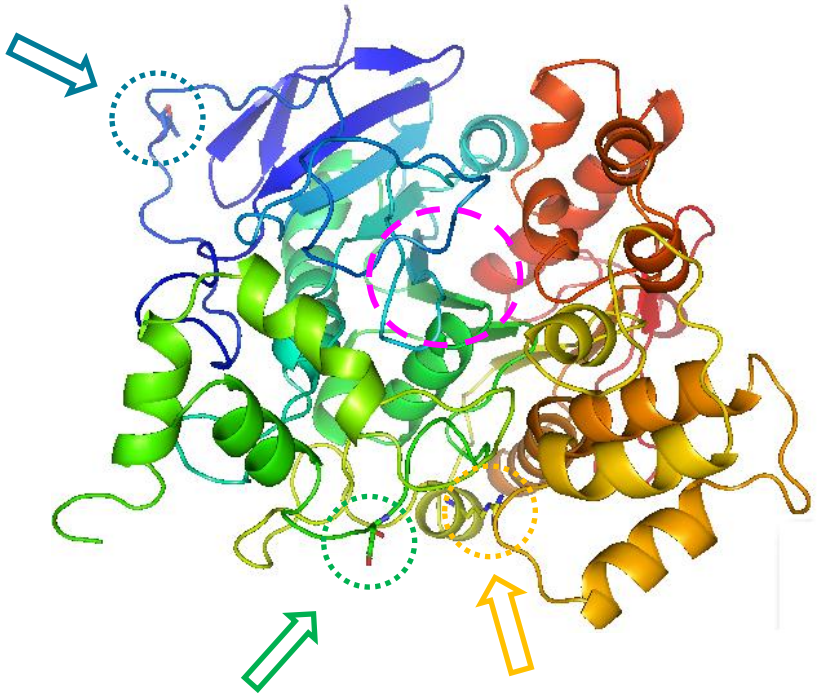
Molecular dynamics

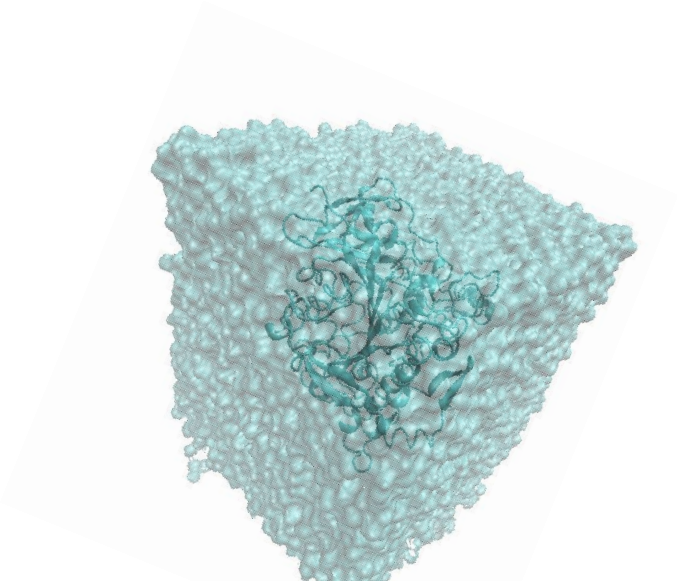
- Molecular dynamics (MD) is a computational biophysics approach to study the equilibrium structure, stability, dynamic and transport properties of bio-molecules, such as proteins

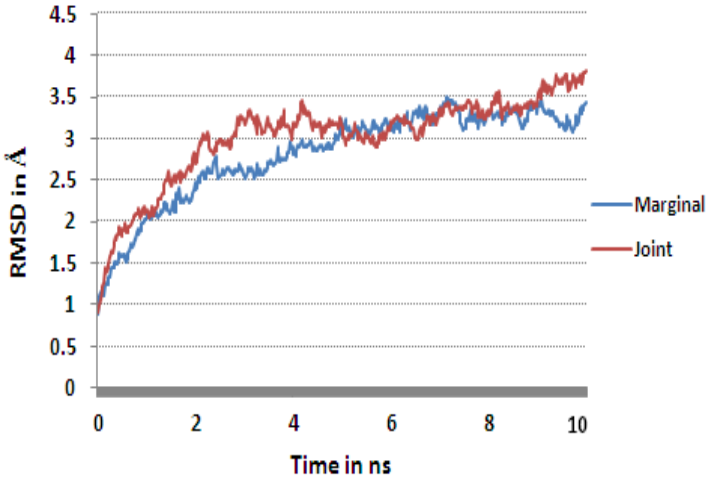
- MD offers significant insights into the time- and temperature-dependent fluctuations and conformational changes in biological systems, which are useful to understand their physical and functional features

- We employ classical (molecular mechanics based) approaches, as implemented in NAMD, to study the effects of amino acid substitution in the dynamic behavior of the ancestral enzymes at physiological temperature and pressure

Root Mean Square Deviation

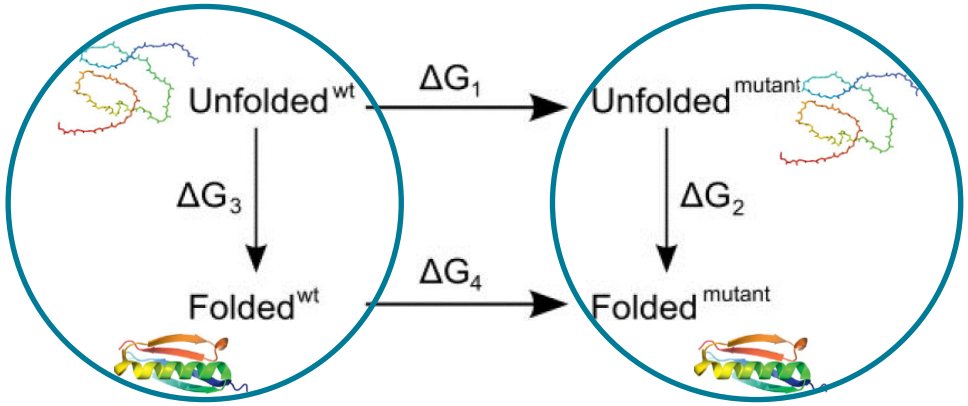






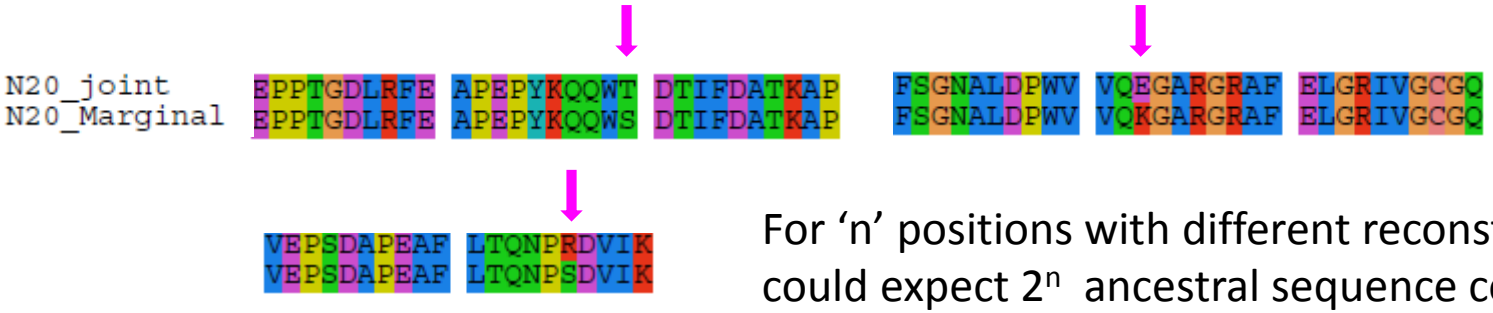
Free Energy Perturbation

Free energy perturbation (FEP) is a computational chemistry method for calculating the relative free energy difference for two states



From the thermodynamic cycle, the folding free energy difference $\Delta\Delta G = \Delta G_3 - \Delta G_2$ between the wild-type protein and the mutant can be calculated via $\Delta G_1 - \Delta G_4$

The more positive the $\Delta\Delta G$, the more the mutant is destabilized in comparison to the wild type



For ‘n’ positions with different reconstructions -- we could expect 2^n ancestral sequence combinations

Residue	Original	Mutated	ΔΔ Stability (gas)	ΔΔ Stability (solvated)	ΔΔ Prime energy
A:51	THR	SER	2.05	0.84	4.2
A:51	SER	THR	-6.01	-1.23	-4.59
A:229	GLU	LYS	69.83	-0.55	-2.47
A:229	LYS	GLU	125.83	3.61	14.31
A:306	ARG	SER	25.56	11.59	41.37
A:306	SER	ARG	192.22	-2.65	-32.42

Preference :
(1) T51-K229-R306
(2) T51-E229-R306

Similar procedures were repeated for the ancestral sequences constructed for the other node, which produced different reconstruction in 7 positions

Our preliminary one-to-one substitution analysis resulted in three preferred sequences out of the 7, which has significant impact to subsequent experimental costs and time

Summary

- Studies on the molecular trails and the characteristics of the desired ancestral character states provide enormous insights into the evolutionary relationships between the extant and extinct

- Preliminary results presented here serve as a good example of how different computational methods for bioinformatics, phylogenomics and molecular modeling and dynamics can combine to play an increasingly relevant role in resurrecting ancestral character states with minimum cost and time.

- Last but not the least, it should be acknowledged that ***Discovering the past becomes a valuable key for Inventions of 'The Future'***

Thank you

Bioinformatics & Phylogenomics Team
Subha Kalyaanamoorthy
OCE Post-Doctoral Fellow

t +61 2 6246 4522
e subha.kalyaanamoorthy@csiro.au