

Using Ancestral Sequence Reconstruction to Create Robust, Highly Expressed Proteins for Crystallography

Elizabeth M.J. Gillam¹, Raine E.S. Thomson¹, Stephina A. D'Cunha¹, Kurt L. Harris¹, Gabriel Foley¹, Mikael Boden¹, Aaron G. Bart², Emily E. Scott²

¹*The University of Queensland, Brisbane, Australia;*

²*University of Michigan, Ann Arbor, MI, Unites States*

e.gillam@uq.edu.au

Producing sufficient amounts of stable, properly folded protein is an essential prerequisite for protein crystallization, especially for challenging targets such as membrane proteins. However many proteins in their native form show limited stability and are refractory to expression in recombinant hosts. Over the last several years, ancestral sequence reconstruction (ASR) emerged as a useful tool by which protein engineers and crystallographers can obtain highly robust forms of proteins of interest, which are often also expressed at relatively high levels in *Escherichia coli*. ASR involves inferring the ancestral state from an alignment of the sequences of extant forms of a given protein family, an evolutionary tree that represents their phylogeny, and an amino acid substitution model. We have developed a suite of software (GRASP: Graphical Representation of Ancestral Sequence Predictions) which enables the inference of ancestral protein sequences from alignments of up to ~ 10000 sequences using maximum likelihood, joint or marginal reconstruction methods. GRASP has been exemplified using several families of eukaryotic cytochrome P450 enzymes, membrane-bound, haemoprotein monooxygenases that have typically been challenging to express and crystallise. To date ASR has enabled expression of ancestral eukaryotic P450s at levels up to ~ 7 μ M in *E. coli* cultures (~350 mg/L culture) leading to the successful crystallization of representative enzymes from several P450 subfamilies. While the resurrected ancestral proteins may not be identical to the extant proteins of principal interest, crystallization of ancestral homologues can provide insights into the structure of a protein family, including how to stabilize the protein fold. In addition, ancestors provide a robust and relevant template for structure function studies as well as protein engineering. When applied to previously uncharacterized sequences, ASR could enable the discovery of new folds and accelerate the functional and structural annotation of sequence-structure-function relationships in novel protein families.

Keywords: Ancestral Sequence Reconstruction, Thermostability, Cytochrome P450, Directed Evolution, Biocatalysis