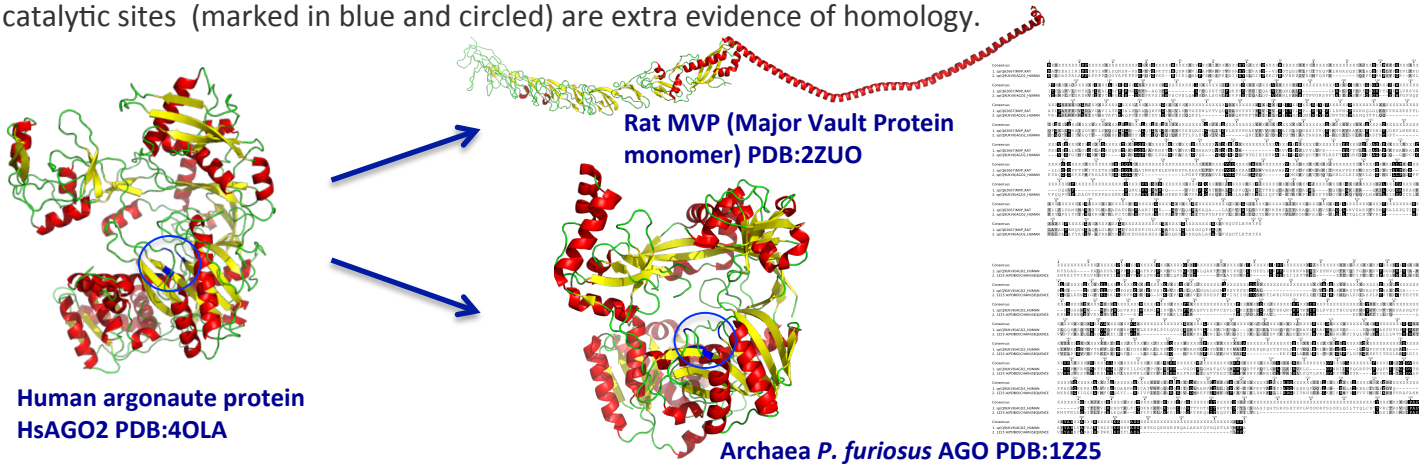
Beyond BLASTing
How 3D structural prediction can support or refute BLAST results

Toni Daly[1,2], David Penny[1] and Andrew Sutherland-Smith[1]

Biologists know that not all BLAST results are homologs and that trees lose information at deep times. We have therefore developed a pipeline to bring increased confidence to the BLAST result. We use structural prediction to support or refute the inclusion of protein sequences into multiple sequence alignments. Ancestral sequence reconstruction (ASR) together with structural prediction can be used to create seed sequences to help find remote homologs. The process can be used to help with annotation.

The problem……

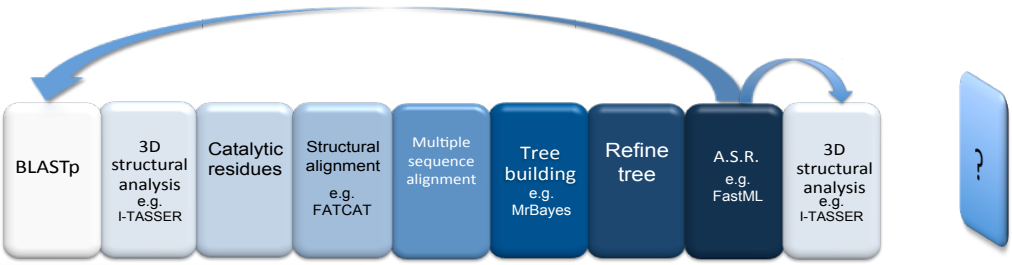
The proteins below are all solved crystal structures from the Protein Data Bank. The human argonaute (HsAGO2) has 14% sequence identity with each of the other two structures but HsAGO2 and rat MVP are clearly **not** homologs. The predicted structure of the archaea *Pyrococcus furiosus* and retention of the catalytic sites (marked in blue and circled) are extra evidence of homology.



Neither is found by BLAST alone, even if they were they would be ignored due to their unlikely 'Expect' value. By creating ancestors of known argonautes and checking their folds we can use those as a BLAST query to find more remote homologs. Structural homology is not always predictable from BLAST results.

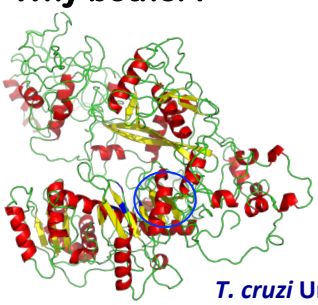
Our solution

We have developed a pipeline emphasising **structural homology** as evidence of evolutionary relationships between remote ancestors. It will also add certainty to the annotation of new sequences.



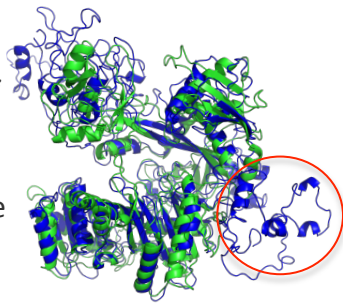
Additional modules can slot in depending on the protein e.g. we have used docking algorithms such as RosettaDock in the past for oligomeric proteins such as the vault monomers.

Why bother?



T. cruzi UniprotKB:D7RU30

Trypanosome cruzi was thought to lack argonautes. With only 14% sequence identity the catalytic sites are in place and the structure is feasible. Aligning the predicted *T. cruzi* model (blue) with the known HsAGO2 (green), we can see that that there is an insert peripheral to the catalytic sites which should not affect function, otherwise the models are structurally similar. This method can be used more generally to demonstrate homology in other proteins.



FATCAT server alignment

[1]Institute of Fundamental Science, Massey University, Palmerston North
[2]Department of Applied & Environmental Sciences, NorthTec, Whangarei