

## Improving Template Selection for Homology Modeling

M.I. Sadowski & D.T. Jones

Abstract:

Homology modeling, the most accurate method currently available for predicting the structure of a protein sequence, generates a model using a homologous template of known structure.

Although there has been some recent progress the structural similarity between the template and the target is still the limiting factor in the accuracy of the model. Selecting the proper template is therefore paramount to generating models at the highest resolution. It is well-known that sequence similarity is a good predictor of structural similarity, however it is not perfect and the most similar sequence is not always the best template to use.

We have recently compared alternative template selection methods on a benchmark set of 732 proteins from 24 CATH families (Sadowski & Jones, *Proteins*, *in press*) and found that for highly similar proteins sequence identity is still the best predictor. More sophisticated profile-based methods improve template similarity only at < 40% sequence identity. Even at high levels of sequence-similarity (50-80%) choosing the most similar sequence leads to errors of 0.5 Angstroms or more for ~25% of our benchmark set.

In this poster we show three examples of incorrect template selections based on sequence similarity and describe how an identity measure based on a subset of sequence positions can be derived using a simple genetic algorithm and used to improve selection.