456b Ipro: Iterative Protein Redesign and Optimization Procedure: Application to Three Case-Studies

Manish C. Saraf, Brian A. Canada, and Costas D. Maranas

The iterative cycle of combinatorial library generation followed by selection, known as the directed evolution method, has emerged as a powerful and commonly used strategy for protein engineering. In directed evolution experiments, sequence diversity is generated by (i) recombination of related genes and/or (ii) random/directed mutagenesis. One of the key challenges in the use of such directed evolution techniques (e.g., DNA shuffling, StEP, ITCHY/SCRATCHY, SHIPREC) is that most of the reassembled sequences code for proteins that are either non-functional or have partially retained activity. Therefore, it appears that exploring protein sequence space freely comes at the expense of severely degrading the average activity of the combinatorial library.

By analyzing experimental data we find that energy-based metrics such as total energy, substrate or product binding energy can be used as surrogates of performance objectives such as overall stability, specific activity or specificity. For example, we find that the calculated binding energy between the hybrid proteins and their substrate for the *E.coli/B.subtilis* and *L.casei/B.subtilis* dihydrofolate reductases is linearly correlated with the logarithm of their specific activities. Drawing from the presence of such correlations we have developed the computational procedure IPRO (Iterative Protein Redesign and Optimization) for identifying optimal mutations for two different design paradigms: (1) downstream in the hybrid sequences, thereby enhancing the quality of each hybrid separately, and (2) upstream in the parental sequences, so that these mutations propagate into the combinatorial library, thereby improving the overall quality of the library in one step.

The redesign of hybrid and/or parental sequences will be discussed for three different protein engineering studies:

(i) Improving specific activities of hybrid sequences (Dihydrofolate Reductase). The objective here is to identify point mutations in the binding pocket of (a) a hybrid sequence or (b) the parental sequences for improving the specific activity of the hybrid and the average activity of the combinatorial library constructed from the parental sequences respectively.

(ii) Changing the shape of the binding site for influencing the binding orientation of the substrate (Nitroarene Dioxygenase (NDO)). It has been reported (Keenan et al., J. Bacteriol., 2005) that two different products (2,3-dinitrobenzyl alcohol, 4-methyl-3-nitrocatechol) are formed in different proportions when 2,3-dinitrobenzene (2,3-DNT) binds with NDO depending on the orientation with which 2,3-DNT binds with NDO. We find that the calculated binding energy of 2,3-DNT in the orientation corresponding to 2,3-dinitrobenzyl alcohol (98% yield) is significantly lower than its calculated binding energy in the orientation corresponding to 4-methyl-3-nitrocatechol (2% yield). This result is in agreement with the relative abundance of the two products. In this project we actively attempt to redesign the binding pocket to reverse the order of the binding energies of 2,3-DNT for the two orientations corresponding to the two different products and thus presumably reverse their relative selectivities. Computational and experimental results will be highlighted.

(iii) Substituting the residues in the binding site of a hybrid enzyme retained from one of the parental sequence to that of the other parental sequence (Glutathione S- Transferase (GST)) while retaining the same level of activity. The theta-class rat GST isozyme rGSTT2 has been shown to have high activity with respect to its human theta-class homolog, hGSTT1, specifically when bound to the conjugate complex of glutathione (GSH) and 7-amino-4-chloromethyl coumarin (CMAC). Certain hybrids of the human and rat sequences were found (Kawarasaki et al., Nucl. Acids Res., 2003) to retain the

characteristic high activity of the parental rat isozyme. However, if a gene encoding one of these hybrids were introduced into a human subject, there is a possibility that the rat-derived binding pocket of the resulting protein might contain epitopes that would be recognized by human antibodies, thereby triggering an immune response. Our objective here is to “humanize” the maximum number of residues in the rat-derived segment of the hybrid enzyme in order to eliminate any antigenic epitopes while retaining as much of the rat-like high activity as possible.