Abstract

Protein design is a rapidly maturing field, with the goal of designing proteins with customized structures and functions. A large number of protein structures are experimentally determined, which have facilitated an understanding of sequence-structure relationships. However, very few studies have explored the full sequence space around each structural fold and further in understanding which of these sequences will lead to functional proteins. Moreover, there are not enough studies to date, reporting the design of entirely novel folds, especially with a designed functionality. There has been significant advance in the computational design methods, which enable us to explore de novo protein design. The challenges in protein design stem from the very large number of potential conformations that need to be explored, as well as scoring and ranking them. Designing proteins with predetermined functions is even more difficult since protein interactions with ligand molecules add a further layer of complexity. To date, reports of functional de novo protein designs remain rare.

In this thesis, different algorithms have been successfully employed to solve four diverse protein and peptide design challenges. The algorithms were selected according to the design problem posed. For large proteins and binding site design, atomic resolution models and corresponding energetics were used. For antimicrobial peptides with non-specific targets, machine learning algorithms were used to understand sequence and structural features of existing proteins to incorporate into the new design.

I will describe computational and experimental approaches for the following design challenges:

- The design and implementation of algorithms for understanding ligand binding-site (pocket) substructures. Three approaches for analysing and classifying pockets are described.

- The computational design and experimental characterization of symmetric TIM barrel proteins using the Rosetta software suite, followed by structural characterization through NMR studies.

- The design, biophysical characterization and NMR structure determination of a heme-binding peptide adopting a novel twin beta hairpin topology.

- The sequence-based design of antimicrobial peptides using a new long short-term memory network based algorithm, along with *in vitro* antimicrobial activity characterization.

- The structure-based design of a second set of antimicrobial peptides based on a maximum common sub-graph detection algorithm optimized using simulated annealing methods, followed by *in vitro* and *in vivo* antimicrobial activity characterization, and NMR structural determination of the best performing antimicrobial peptide . The applications of the algorithms presented to the broader field will also be discussed.