
Preface

Proteins have evolved through selective pressure to accomplish specific functions. The functional properties of proteins depend upon their three-dimensional structures, which result from particular amino acid sequences folding into tightly packed domains. Thus, to understand and modulate protein function rationally, one definitely needs methods and algorithms to predict and decipher how amino acid sequences shape three-dimensional structures. Protein design aims precisely at providing the tools to achieve this goal.

The predictive power of rational protein design methods has dramatically increased over the past five years. A broad range of studies now illustrate how the sequence of proteins and peptides can be tuned to engineer biological tools with intended properties (1–3). The extensive characterization of peptides and protein mutants has enormously benefited the understanding of protein sequence-to-structure relationships. Synergies between computational and experimental approaches have also added momentum to the advancing limits of design methods. The potential applications in fundamental biochemistry and in biotechnology justify the considerable excitement that this progress has generated within the research community. The field is probably mature enough so that expert knowledge can assist researchers of diverse disciplines to rationally create or modify their favorite protein. Thus, the aim of *Protein Design: Methods and Protocols* is to account for the most up-to-date protein design and engineering strategies so that readers can undertake their own projects with maximum confidence in a successful return.

The basic concepts underlying rational design of proteins are intimately related to their three-dimensional structures. The stability of a given structure results from a complex combination of interactions that favor a specified conformation at the expense of any alternative one. Researchers have devised different strategies to extract the general principles on which protein structure is based. Proteins have been systematically mutated to address the question of how specific residues affect the stability of a given protein (4,5). Proteins have been also “redesigned,” starting with a protein of known structure and dramatically modifying features of its construction (6). For the sake of simplicity, initial works in the field of design were dedicated to the elucidation of the factors contributing to the stability of elementary building blocks. Peptide model systems have been shown to be very suitable to this end. They have served to dissect the relative energetic contributions of short- and long-range interactions to a given folding motif. They have provided key insights

into the relationship between sequence, folded structure, and stability (7). Major accomplishments have been achieved in the design and structural characterization of helical peptides and proteins (8,9). The main factors underlying α -helix stability have been largely identified, leading to advances in the rational design of helical proteins. Protein stability has been enhanced by maximizing helical propensities at specific sites, and protein structures have even been redesigned to adopt different folded topologies. The design of α -helices is surveyed in Chapter 1.

The rational understanding of β -sheet structure and stability has remained, however, more elusive, and it is only during the past five years that similar success has been achieved (10). In contrast to α -helices, β -sheets are propagated by residues remote in the polypeptide backbone. As a consequence, whereas in model helical peptide structure stabilization is largely a result of interactions between neighboring residues (local interactions), nonlocal interactions make important contributions to the stability of even minimal β -sheet peptides and proteins. This fact together with the intrinsic tendency of β -sheets to aggregate can be recognized as the main impediment to a comprehensive understanding of β -sheet structures. These studies are reviewed in Chapter 2. The basic rules derived from the analysis of these model systems can be used directly by the reader to increase the stability of a given protein through the local optimization of their constitutive secondary structure blocks. The predictive power of these rules has been tested already by the design, completely *de novo*, of several peptides and miniproteins. The conjunction of rational design principles and combinatorial approaches has been very successful also at finding sequences that highly populate a desired folding motif (11,12) and their applicability is demonstrated in Chapter 3.

These fundamentals can be exploited also to design, modify, or improve the interaction between peptide ligands and their receptor targets. This interaction commonly involves the formation of beta structures. Yet poor bioavailability and unfavorable pharmacokinetics significantly compromise the use of peptides as drugs. An additional problem is their conformational flexibility, which results in poor binding to the target. Thus, there is a great deal of interest in designing peptidomimetics with improved structural properties as therapeutic agents by mimicking β -turn and β structures. To this end, D-amino acids have been strategically introduced in polypeptide backbones to decrease the conformational flexibility of β -hairpin and β -sheet peptides designed *de novo* (13). β -peptides constitute one of the most important families of nonnatural polymers with the propensity to form well-defined secondary structures. They are attracting more and more attention because they have been found to have various applications in medicinal chemistry and biochemistry (14). These topics are covered in Chapters 4 and 5.

Procedures and strategies for engineering helices or β -sheets, solvent-exposed positions, or buried ones, common folds, or rare ones differ substantially. It is often difficult to account for all these factors using simple rules or relationships. Besides, one has rapidly to face a huge combinatorial complexity while increasing the number of positions in a sequence that are to be engineered simultaneously. For that purpose, integrated computational approaches have been developed based on different strategies (15–18). *Protein Design: Methods and Protocols* presents several of these algorithms, which require various degrees of computational complexity (Chapters 7–9). In addition to the basic philosophy underlying their work, the practical comments of the authors on the use of their tools will be of major interest to experimentalists selecting the strategy most adapted for their design problem.

Since protein binding is fundamentally ruled by the same laws as protein folding, the lessons learned by designing stable proteins have paved the way for important progress in the engineering of protein complex interfaces. This issue has a tremendous impact in many biological fields because it allows one to modulate the way protein-interaction networks in cells are organized. The specificity of protein–protein complex engineering is discussed through the success of three different applications (Chapters 6, 10, and 11).

A frequent pitfall hindering successful designs is the tendency of the engineered molecule to aggregate. Unspecific aggregation processes can trap most of the designed protein into amorphous aggregates. In other cases, proteins can aggregate in an organized fashion and lead to the formation of fibrillar aggregates, known as amyloid fibrils. Amyloid fibrils are also associated with a range of human disorders, such as spongiform encephalopathies, Alzheimer's disease, type II diabetes, and so forth (19). Recent progress in understanding the relationship between protein sequence and protein aggregation processes have provided clues on how to escape from these conformational traps (20). This knowledge may help to negatively design sequences that, while maintaining the compatibility with the template fold, either decrease or fully prevent self-association processes. Knowledge-based tools might also be applied to predict protein fragments responsible for the amyloidogenic behavior of a given pathogenic protein, and, as a further application, to design or screen for inhibitor molecules that specifically interact with these key aggregating regions, preventing aggregation or increasing clearance of the misfolded protein. Design approaches, validation methods, and application to predicting such behavior are discussed in Chapter 12. Therapeutic approaches that are currently under scrutiny for preventing or curing amyloidoses or protein misfolding diseases in general are discussed in Chapter 13.

How can I handle the design of my protein? How can I improve the binding of this peptide to my target protein? Might I avoid protein aggregation while retaining fold and stability? Which structural features should be considered with acute attention? How good are we at translating angstroms into calories? These are central questions addressed throughout *Protein Design: Methods and Protocols* with the expectation that researchers can find their way toward achieving successful designs.

Raphael Guerois
Manuela López de la Paz

REFERENCES

1. Kortemme, T. and Baker, D. (2004) Computational design of protein-protein interactions. *Curr. Opin. Chem. Biol.* **8**, 91–97.
2. Bolon, D. N., Voigt, C. A., and Mayo, S. L. (2002) De novo design of biocatalysts. *Curr. Opin. Chem. Biol.* **6**, 125–129.
3. Brannigan, J. A. and Wilkinson, A. J. (2002) Protein engineering 20 years on. *Nat. Rev. Mol. Cell Biol.* **3**, 964–970.
4. Mendes, J., Guerois, R., and Serrano, L. (2002) Energy estimation in protein design. *Curr. Opin. Struct. Biol.* **12**, 441–446.
5. Bava, K. A., Gromiha, M. M., Uedaira, H., Kitajima, K., and Sarai, A. (2004) ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.* **32** (Database issue), D120–121.
6. Regan, L. (1999) Protein redesign. *Curr. Opin. Struct. Biol.* **9**, 494–499.
7. Serrano, L. (2000) The relationship between sequence and structure in elementary folding units. *Adv. Protein Chem.* **53**, 49–85.
8. Munoz, V. and Serrano, L. (1995) Helix design, prediction and stability. *Curr. Opin. Biotechnol.* **6**, 382–386.
9. Doig, A. J. (2002) Recent advances in helix-coil theory. *Biophys. Chem.* **101–102**, 281–293.
10. Lacroix, E., Kortemme, T., Lopez de la Paz, M., and Serrano, L. (1999) The design of linear peptides that fold as monomeric beta-sheet structures. *Curr. Opin. Struct. Biol.*, **9**, 487–493.
11. Pastor, M. T., Lopez de la Paz, M., Lacroix, E., Serrano, L., and Perez-Paya, E. (2002) Combinatorial approaches: a new tool to search for highly structured beta-hairpin peptides. *Proc. Natl. Acad. Sci. USA* **99**, 614–619.
12. Wei, Y., Liu, T., Sazinsky, S. L., Moffet, D. A., Pelczar, I., and Hecht, M. H. (2003) Stably folded de novo proteins from a designed combinatorial library. *Protein Sci.* **12**, 92–102.
13. Venkatraman, J., Shankaramma, S. C., and Balaram, P. (2001) Design of folded peptides. *Chem. Rev.* **101**, 3131–3152.
14. Cheng, R. P., Gellman, S. H., and DeGrado, W. F. (2001) β -Peptides: from structure to function. *Chem. Rev.* **101**, 3219–3232.

15. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368.
16. Dahiyat, B. I. and Mayo, S. L. (1997) De novo protein design: fully automated sequence selection. *Science* **278**, 82–87.
17. Wernisch, L., Hery, S., and Wodak, S. J. (2000) Automatic protein design with all atom force-fields by exact and heuristic optimization. *J. Mol. Biol.* **301**, 713–736.
18. Reina, J., Lacroix, E., Hobson, S. D., Fernandez-Ballester, G., Rybin, V., Schwab, M. S., Serrano, L., and Gonzalez, C. (2002) Computer-aided design of a PDZ domain to recognize new target sequences. *Nat. Struct. Biol.* **9**, 621–627.
19. Dobson, C. M. (2003) Protein folding and misfolding. *Nature* **426**, 884–890.
20. Lopez de la Paz, M. and Serrano, L. (2004) Sequence determinants of amyloid fibril formation. *Proc. Natl. Acad. Sci. USA* **101**, 87–92.