Preface

The aim of this first-ever book entitled *Computational Protein Design* (CPD) is to bring the latest know-how on the CPD methods in respect to the process, success, and pitfalls of the field. The book is organized so as to introduce and present the general methodology and main challenges followed by a description of specific software and applications. As seen in the description below, there is more than one way to cluster the different chapters, each highlighting a different aspect of the field.

While there has not been a book dedicated to CPD, books on protein design have often included chapters on CPD. Here, following a chapter on the framework of CPD (Chapter 1) and a summary of past achievements and future challenges (Chapter 2), a chapter on the experimental aspects of production of the designed protein is presented (Chapter 3). Beyond the need to understand the experimental aspects of the computational endeavor, this is to remind us that the final outcome of the computational process is the production of a real protein.

It is widely considered that a global minimum energy conformation (GMEC) reflects the actual native structure of the protein. The protein design process is intrinsically computationally intensive as sequence and structure space should be rigorously sampled in the search for the GMEC of the requested target. Deterministic search methods (Chapter 4) of which dead-end elimination (DEE) is among the first to be used, are guaranteed to find the GMEC while stochastic methods are not guaranteed to find it. Other methods, e.g., the A* search algorithm, were optimized to run in parallel taking advantage of the graphic processing unit (GPU) processor infrastructure (Chapter 13). Complementarily, the CPD effort should consider the solvating milieu, e.g., via a geometric potential (Chapter 5). In addition, the residue-level core building block focus of CPD should be analyzed and predicted in respect to phylogenetic, structural, and energetic properties. These should be treated according to the immediate and possibly changing microenvironment, e.g., as in protein–protein complexes (Chapter 6). The GMEC considers a single minimum conformation and can be applied for the redesign of a given scaffold (Chapter 10), for requested functional motifs (Chapter 11) or for emphasizing specific types of available data, e.g., evolutionary information (Chapter 12). Yet, proteins within their native physiological surrounding are dynamic ensembles intrinsically requiring conformational dynamics. As such, it is important to *a priori* design the protein as a multistate entity (Chapter 7), a characteristic that can be introduced via integrating to the design process methods that analyze dynamics such as molecular dynamics (Chapter 8) or normal mode analysis (Chapter 9).

The computational design scheme can be tailored to specific types of proteins or domains, which in turn should be assessed as to their resemblance to the requested domain or specific designated characteristic. Examples include protein–protein interaction interfaces (Chapter 14), drug-resistance mutations (Chapter 15), symmetric proteins of identical sequence repeats (Chapter 16), self-assemblies exploiting synthetic amino acids (Chapter 17), oligomerized conformations of the defensins (Chapter 18), ligand-binding proteins (Chapter 19), proteins with reduced immunogenicity (Chapter 20), antibodies (Chapter 21), membrane curvature-sensing peptides (Chapter 22), and allosteric drug-binding sites within proteins (Chapter 23). Taken together, these application focus areas

present the breadth of the CPD field along with the intrinsic achievements and challenges upon examining the "devil" in the details of key examples.

The general field of protein design, let alone the computational aspect of it, is expected to present an exponential increase in quality and quantity alike. Such change is fostered by the need to expand protein space for understanding biology, for applying biotechnology, and for expanding pharmaceuticals from the common small molecules to biologics – specific and side-effect-free proteins. Importantly, while scientific research of proteins is often focused towards pharmaceutical applications, CPD presents the possibility to expand the use of proteins in food-tech and white biotechnology, namely, the use of proteins for industrial applications. In addition, the field is nurtured by the exponential increase in raw sequence and structure data, and the increase in cost-effect computational hardware in general and hardware tailored to protein application, in particular. Not less important is the careful feedback loop of quantitative parameterization sequence and fold space followed by software design that will efficiently test our parameterization and produce novel protein design, which in turn can be materialized and characterized experimentally.

Karmiel, Israel   *Ilan Samish*