Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Structural heterogeneity and dynamics in protein evolution and design

Kristoffer E Johansson and Kresten Lindorff-Larsen

Recent years have witnessed substantial progress in our ability to design proteins with specific structures and to introduce new functionalities into existing protein scaffolds. Such protein design efforts test our understanding of the biophysical and functional mechanisms of naturally evolved proteins. At the same time, we also know that proteins are dynamical entities, and that many proteins rely on detailed dynamical mechanisms for regulation and function. Thus, the success of design methods, especially in relation to functional proteins, might benefit from explicit considerations of conformational heterogeneity and dynamics. In this review, we compare results from the field of protein design with laboratory protein evolution with a focus on dynamics. Recent studies show that structural dynamics is altered during evolutionary trajectories, and that allosteric effects are pronounced. Interaction networks and the resulting coupling of structure and dynamics are suggested to facilitate these effects.

Address
The Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Corresponding authors: Johansson, Kristoffer E (kristoffer.johansson@bio.ku.dk), Lindorff-Larsen, Kresten (lindorff@bio.ku.dk)

Current Opinion in Structural Biology 2018, **48**:157–163

This review comes from a themed issue on **Proteins**

Edited by **Birte Hocker** and **Jakob Winther**

https://doi.org/10.1016/j.sbi.2018.01.010

Introduction

The ability to design proteins *de novo* holds great perspectives not only for biotechnological and medical applications, but also as a test of our basic understanding of protein evolution, folding and stability. In particular, the field of computational protein design is in rapid development. Design of entire amino acid sequences up to medium sized globular alpha/beta proteins of both known [1–3] and novel folds [4–6] has been achieved, and has helped elucidate general principles of protein structure [6]. Until recently, designed proteins were relatively rare because of the low success rate of computational design methods in combination with the laborious experiments needed to synthesize and characterize a defined sequence. Recent efforts to improve the success rate and the development of high-throughput expression and stability screens of many thousands of defined sequences hold great promise [7••].

Protein dynamics and the ability to change shape is known to play a widespread role in protein function, for example as seen in effector regulation, and is thus an evolved property and a potential cause of disease [8]. Furthermore, a recent study has linked structural heterogeneity to the 'evolvability' of enzyme active sites [9]. At the same time, advances in biophysical experiments and molecular simulations techniques provide us with new opportunities to characterize protein dynamics at the atomic level. Progress within these areas is reviewed elsewhere [10–12]. Computational protein design methods, on the other hand, generally rely on the assumption of static conformations in order to handle the vast complexity of sequence space, and sequence and structure is typically optimized iteratively in independent steps. As a consequence, the accuracy of these methods has been observed to correlate with structural rigidity [13] and to be hypersensitive towards minor conformational changes in the backbone coordinates [3]. Methods that simultaneously consider the variability of both the sequence and structure are less common, and are typically not employed in full sequence design [14]. Machine learning methods may be used to learn the relationships between structural environment and sequence, and thereby to propose amino acids substitutions without a priori assumptions about side-chain conformations [15].

The need for a rational approach to deal with conformational heterogeneity in computational protein design has been highlighted in recent reviews [16,17]. A promising approach to computational design of structural variability is the multistate approach [18], in which the amino acid sequence is simultaneously optimized on multiple backbone conformations [19–22]. Joh *et al.* used such a principle to design a flexible transmembrane four-helix bundle with selective ion transport function. The protein facilitates diffusion of certain transition metal ions using two binding sites with designed negative cooperativity, such that only one site may be occupied at a time [23•].

There has also been substantial progress in utilizing computational methods for the design of proteins with enzymatic function [24–31]. In particular, recent work has

shown improvements in the precision of positioning the catalytic side chains as necessary for enzymatic catalysis [32]. One current paradigm in computational enzyme design is to position a handful of residues in space to stabilize a theoretical transition state complex, a theozyme, and subsequently to search for a native protein to host this active site. The particular step of sequence optimization is generally based on static conformations, and only in subsequent steps of screening or optimization have methods like molecular dynamics simulations been applied [29,33]. Thus, any structural dynamics or heterogeneity of relevance for the efficiency of the overall catalytic cycle originates from the native host protein or subsequent optimization of catalysis by either computational or experimental methods.

Directed evolution is a powerful strategy to improve protein function, but generally requires that the starting point has a measurable amount of the target activity. The combination of computational design to generate new activities with subsequent optimization by directed evolution has thus proven a particular successful approach [13,18,34,35,36••]. The lower activity of the initial designed enzymes has been suggested to originate from the population of non-functional conformational states [37,38••,39•]. In light of this, strikingly few biophysical studies of the dynamical aspects of computationally designed enzymes have been reported. It has been suggested that computational design might open up a catalogue of structural variants that enable directed evolution for improved function [9,39•]. The simultaneous evolution of protein sequence, structure, dynamics and function may thus beneficially be studied in trajectories from either natural or directed evolution. We thus here review recent studies of protein dynamics in directed evolution trajectories to provide inspiration for future studies aimed towards incorporating protein dynamics into rational *de novo* protein design.

Protein dynamics and directed evolution

Proteins display dynamical properties on a wide range of time and length scales, ranging from small picosecond fluctuations of atomic bonds to larger conformational changes or folding processes on timescales on seconds and beyond. Indeed, it has been found that protein dynamics may change during optimization by directed evolution [38••,39•,40••,41•]. While it is often assumed that larger conformational changes are generally more relevant for function, chemistry occurs on short length scales, and it has long been recognized that 'an exceedingly subtle allosteric rearrangement' [42] of just a few Ångström can be sufficient to switch activity. Research into design of enzymes has confirmed this need for structural precision and controlled dynamics [18,32]. In the context of protein design and analysis of evolutionary trajectories this also means that one cannot focus solely on large structural differences or the regions that are most dynamic.

A recent example shows that fine-tuning of structure and dynamics may be one of the major consequences of directed evolution. Campbell *et al.* studied the simultaneous evolution of sequence, structure, dynamics and function in a directed evolution trajectory of a phosphotriesterase (PTE) that was evolved to gain a high level of arylesterase (AE) function [40••]. Crystal structures along the trajectory showed a maximum pairwise Cα RMSD of 0.4 Å except for minor changes in the flexible loops that surround the active site, which were all expected to be involved in either PTE or AE function (Figure 1). The first round of directed evolution introduced an active site mutation, H254R, and mutations from several subsequent rounds were found to fine tune the side-chain conformation of this arginine residue by introducing substitutions at an increasing distance from the active site. At the same time, the conformational heterogeneity of some loops was found to change consistently along the trajectory resulting from substitutions that perturbed interactions with these loops. A further discussion of this work in relation to structural dynamics and catalysis can be found in [43].

Another example shows that directed evolution may alter the timescales of the dynamics. Gonzalez *et al.* used NMR spectroscopy to characterize the dynamics of proteins from a trajectory of a metallo-β-lactamase evolved towards an expanded substrate profile [41•]. Here, the mutation N70S alone lowered the amount of slow timescale dynamics and impaired activity while in the background of another mutation, G262S, the mutation enhanced the slow time-scale dynamics and expanded the substrate profile. This study shows an epistatic effect between the two mutations in which the second mutation appears to facilitate the first. Only by introducing G262S does N70S show the desired effect.

Strategies for directed evolution include both site-directed mutagenesis at selected sites, and sequence-wide random mutations. The first requires some analysis and understanding of the properties that are to be optimized, but on the other hand allows for saturation mutagenesis or even exhaustive screening of many combinations at a few selected positions. Podgornaia and Laub did the latter on four positions and screened all 160 000 possible quadruple mutations for effects on protein–protein interactions using fluorescence activated cell sorting [44•]. They found widespread epistasis, suggesting that the possibility of introducing multiple substitutions between fitness evaluations might be a particularly powerful approach to protein evolution.

Sequence-wide random mutagenesis samples a close neighbourhood in sequence space of typically one or a few nucleotide substitutions at random positions, and is