Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

# Protein function machinery: from basic structural units to modulation of activity

Igor N Berezovsky[1,2], Enrico Guarnera[1], Zejun Zheng[1], Birgit Eisenhaber[1] and Frank Eisenhaber[1,3]

Contemporary protein structure is a result of the trade off between the laws of physics and the evolutionary selection. The polymer nature of proteins played a decisive role in establishing the basic structural and functional units of soluble proteins. We discuss how these elementary building blocks work in the hierarchy of protein domain structure, co-translational folding, as well as in enzymatic activity and molecular interactions. Next, we consider modulators of the protein function, such as intermolecular interactions, disorder-to-order transitions, and allosteric signaling, acting via interference with the protein's structural dynamics. We also discuss the post-translational modifications, which is a complementary intricate mechanism evolved for regulation of protein functions and interactions. In conclusion, we assess an anticipated contribution of discussed topics to the future advancements in the field.

**Addresses**
[1] Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, #07-01, Matrix, Singapore 138671, Singapore
[2] Department of Biological Sciences (DBS), National University of Singapore (NUS), 8 Medical Drive, Singapore 117579, Singapore
[3] School of Computer Engineering (SCE), Nanyang Technological University (NTU), 50 Nanyang Drive, Singapore 637553, Singapore

Corresponding author: Berezovsky, Igor N (igorb@bii.a-star.edu.sg)

**Current Opinion in Structural Biology** 2017, **42**:67–74

This review comes from a themed issue on **Proteins: bridging theory and experiment**

Edited by **Igor N Berezovsky** and **Ugo Bastolla**

For a complete overview see the Issue and the Editorial

Available online 16th November 2016

http://dx.doi.org/10.1016/j.sbi.2016.10.021



## Introduction

The views on protein function have undergone significant transformation since Emil Fisher's 'lock-and-key' model of 1894 [1], which provided the foundation for the classical structure–function paradigm. In this review, we first present the current views on basic structural and functional units of soluble proteins and their involvement in organization of the tertiary structure, co-translational folding, and function. Second, we turn to consideration of the most common mechanisms of protein function regulation that act via affecting the structural dynamics: intermolecular interactions, intrinsic disorder/flexibility, allostery, and post-translational modifications. We conclude with a critical evaluation of discussed topics from the perspective of future research tasks and challenges.

## Basic units of protein structure and function

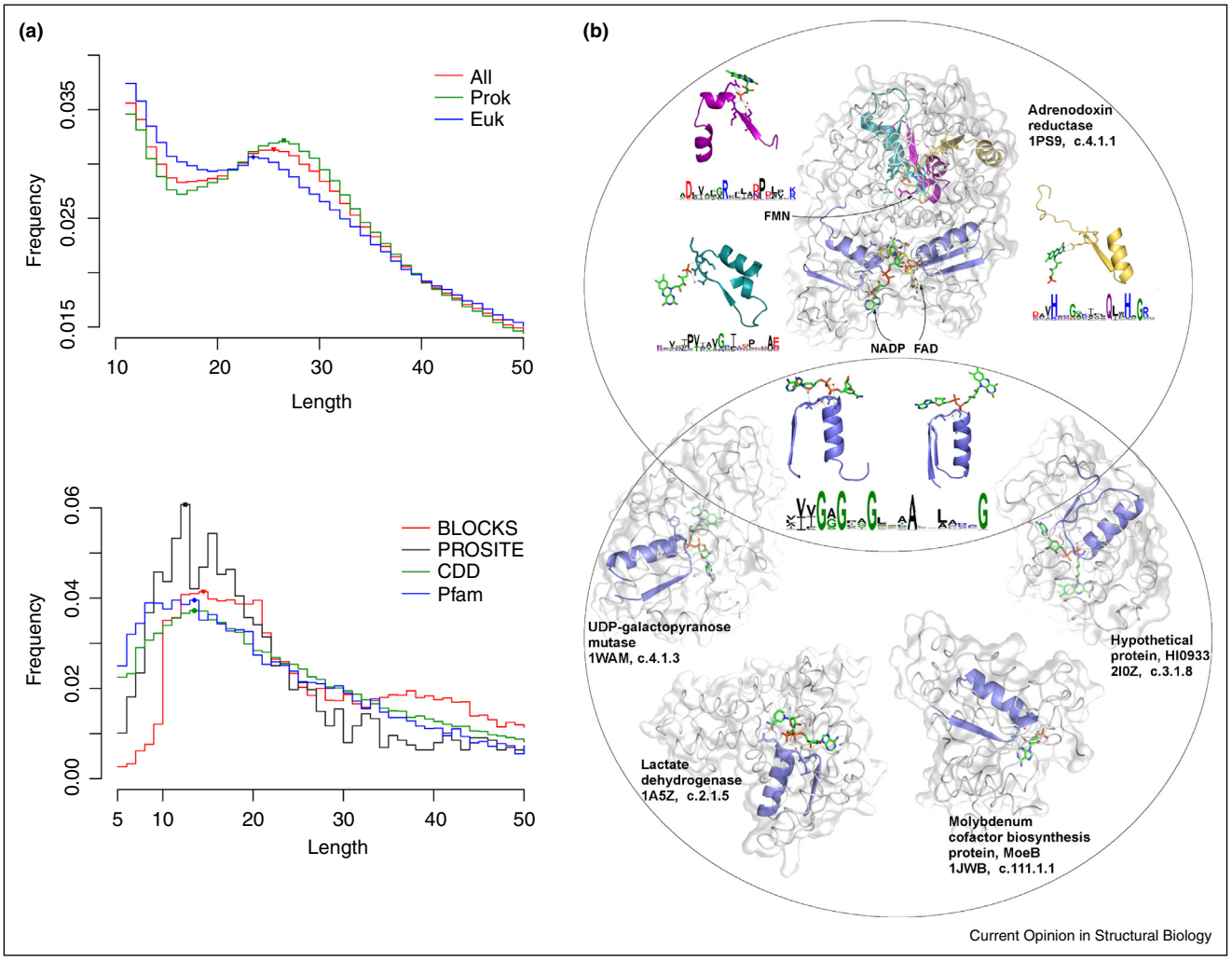
### Closed loops

Protein modularity and architecture is a topic of intense discussion since the very beginning of the protein structure studies [2]. How to define the basic structural unit is one of the major questions, which reappeared in the study of the hierarchy of protein domain structure [3,4]. It was hypothesized that the protein backbone is the major determinant of the protein architecture and, therefore, it can be instrumental in detecting the protein partitioning [5]. Exhaustive enumeration of the polypeptide chain's sections with short spatial distances between their ends revealed the common basic unit of globular proteins — closed loop or return of the polypeptide backbone with preferential contour length of 25–30 amino acid residues [5]. Figure 1a (top chart) shows the universality and omnipresence of closed loops in proteins of prokaryotes and eukaryotes. The polymer nature of protein chains was found to be the origin of the shape and size of this universal unit [5], with implications in the emergence of protein folds/domains [6] as combination of prebiotic ring-like peptides [7••], in the architecture of multidomain structures and hierarchy of protein domain structure [8,9], as well as in the scenario of the co-translational protein folding [10]. The hierarchy of domain structure [3,4] determined by the protein's loop-n-lock structure [6,9] reconciles different methods of detecting protein domains [8].

### Elementary functional loops

Size distributions of the non-gapped functional signatures Figure 1a (bottom chart) prompt to the hypothesis that closed loop can serve as a structural basis of the elementary unit of protein enzymatic function, elementary functional loop (EFL). The correspondence between the elementary functions goes beyond the similarity between functional superfamilies and even folds, as it originated in the very emergence of the first folds from prebiotic functional peptides [11]. Therefore, computational approach for derivation of the evolutionary prototypes of elementary functions is based on sequence analysis [12,13]. Hidden evolutionary relations between the enzymatic functions

**Figure 1**




Closed loops and elementary functional loops are elementary building blocks of the soluble proteins' structures and functions. **(a)** Closed loops of nearly standard size are omnipresent in proteins of both prokaryotes and eukaryotes (top chart). Complete set of structures was downloaded from Protein Data Bank, and CD-HIT was used to eliminate redundancy at 50% level. Protein chains longer than 600 residues were also excluded, as they can be dominated by non-globular structures. In total, 14 501 and 10 732 protein chains of eukaryotes and prokaryotes, respectively, were analyzed. Closed loops are defined as subtrajectoires of polypeptide backbone with end-to-end (Cα–Cα) distance within 10 Å. Distributions are plotted for loops longer than 10 residues. Length distributions of functional signatures shows that they can be carried by closed loops (bottom chart). Preferential size of signature is about 15–20 residues, which together with segments of van der Waals locks (3–5 residues on each loop terminus) results in the closed loop's typical contour length. Data on non-gapped functional signatures are obtained from CDD (52241 entities), PFAM (16295), BLOCKS (32125), and PROSITE (2416) databases. Distributions of non-gapped BLOCKS's lengths are plotted directly. Multiple sequence alignments from CDD, PFAM, and PROSITE databases are split into non-gapped blocks (single-residues gaps are allowed; in each block total number of small gaps in individual sequences should not exceed 5% of number of residues in the block), then distributions of their lengths are plotted. **(b)** The combination of elementary functional loops in the protein (upper oval) is shown for adrenodoxin reductase (PDB ID: 1ps9, nucleotide-binding domain fold). The glycine-rich motif with a characteristic signature GxGxxG binds phosphates in dinucleotide-containing flavin adenine dinucleotide (FAD) and nicotine adenine dinucleotide phosphate (NADP). Glycine-rich motif (characteristic signature GxGxxG) provides binding of the phosphate moiety in dinucleotide-containing ligands in many different proteins (bottom oval). Set of structures in the bottom oval shows that this signature is reutilized in different functional superfamilies and folds: c.111.1.1 is activating enzymes of the ubiquitin-like proteins fold; c.2.1.5 — NAD(P)-binding Rossmann-fold; c.4.1.3 — nucleotide-binding domain; c.3.1.8 — FAD/NAD(P)-binding domain.

in archaeal kingdom were unraveled using prototypes of EFLs [14]. In particular, the reutilization of EFLs in building new functional folds along with repurposing of functional domains was explored in methanogenesis pathway archetypal for Archaea [14]. The evolution of protein function from its emergence to the contemporary realm of proteins was comprehensively surveyed in [7••]. Specifically, the mutual work of physics and biology in shaping structures of modern proteins and achieving diversity of their enzymatic functions was analyzed.