Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Internal symmetry in protein structures: prevalence, functional relevance and evolution☆

Santhanam Balaji

Symmetry has been found at various levels of biological organization in the protein structural universe. Numerous evolutionary studies have proposed connections between internal symmetry within protein tertiary structures, quaternary associations and protein functions. Recent computational methods, such as SymD and CE-Symm, facilitate a large-scale detection of internal symmetry in protein structures. Based on the results from these methods, about 20% of SCOP folds, superfamilies and families are estimated to have structures with internal symmetry (Figure 1d). All-β and membrane proteins fold classes contain a relatively high number of unique instances of internal symmetry. In addition to the axis of symmetry, anecdotal evidence suggests that, the region of connection or contact between symmetric units could coincide with functionally relevant sites within a fold. General principles that underlie protein internal symmetry and their connections to protein structural integrity and functions remain to be elucidated.

Address
MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, United Kingdom

Corresponding author: Balaji, Santhanam (bsanthan@mrc-lmb.cam.ac.uk)

Current Opinion in Structural Biology 2015, **32**:156–166

This review comes from a themed issue on **Sequences and topology**

Edited by **M Madan Babu** and **Anna R Panchenko**

http://dx.doi.org/10.1016/j.sbi.2015.05.004

Introduction

Symmetry was one of the first recognized features of protein structures that were determined at atomic detail in three-dimensions (3-D) [1,2]. With the accumulation of such structural data, it has become clear that symmetry in protein structures is not uncommon, and has been suggested to be relevant for folding, function and evolution of proteins [3–6,7•,8–10]. Symmetry in protein structures is found at various levels of biological organization: whole biological assemblies, the quaternary association of proteins, within individual proteins and even within protein domains (Figure 1a). Analyses of quaternary associations in protein structures have led to the elucidation of various principles that underlie the symmetry of biologically relevant oligomeric states [11–16].

Presence of symmetry in proteins at the level of tertiary structures or domains, termed internal symmetry, was first recognized through the examination of homologous amino acid sequence repeat stretches in proteins [17,18]. Subsequently it became evident with the determination of numerous crystal structures of proteins that internal symmetry also exists in 3D, independent of sequence identity [19,20]. The hierarchical classification of protein structures organized in databases such as SCOP and CATH [21,22] has allowed the recognition of structural symmetry being widespread, from individual domains to domain folds (e.g. Ferredoxin-like, all forms of β-propellers, β-barrels, alpha toroids). Computational methods that utilize both sequence and structural information to systematically identify symmetry in protein structures have been developed over many years, which include COSEC2, DAVROS, OPAAS, Swelfe, RQA, GANGSTA+, SymD and CE-Symm (see Myers-Turnbull *et al.* [23••] for further details). Attempts to elucidate evolutionary models on a selected set of symmetric protein domains such as β-trefoils and β-propellers have been successful [7•,24]. This inspired the design of symmetric protein structures with a potential for various applications [25–27]. Particularly in the last five years, there has been significant progress in identifying protein structural symmetry and describing its relevance to functions and evolution of proteins, which will be the primary focus below.

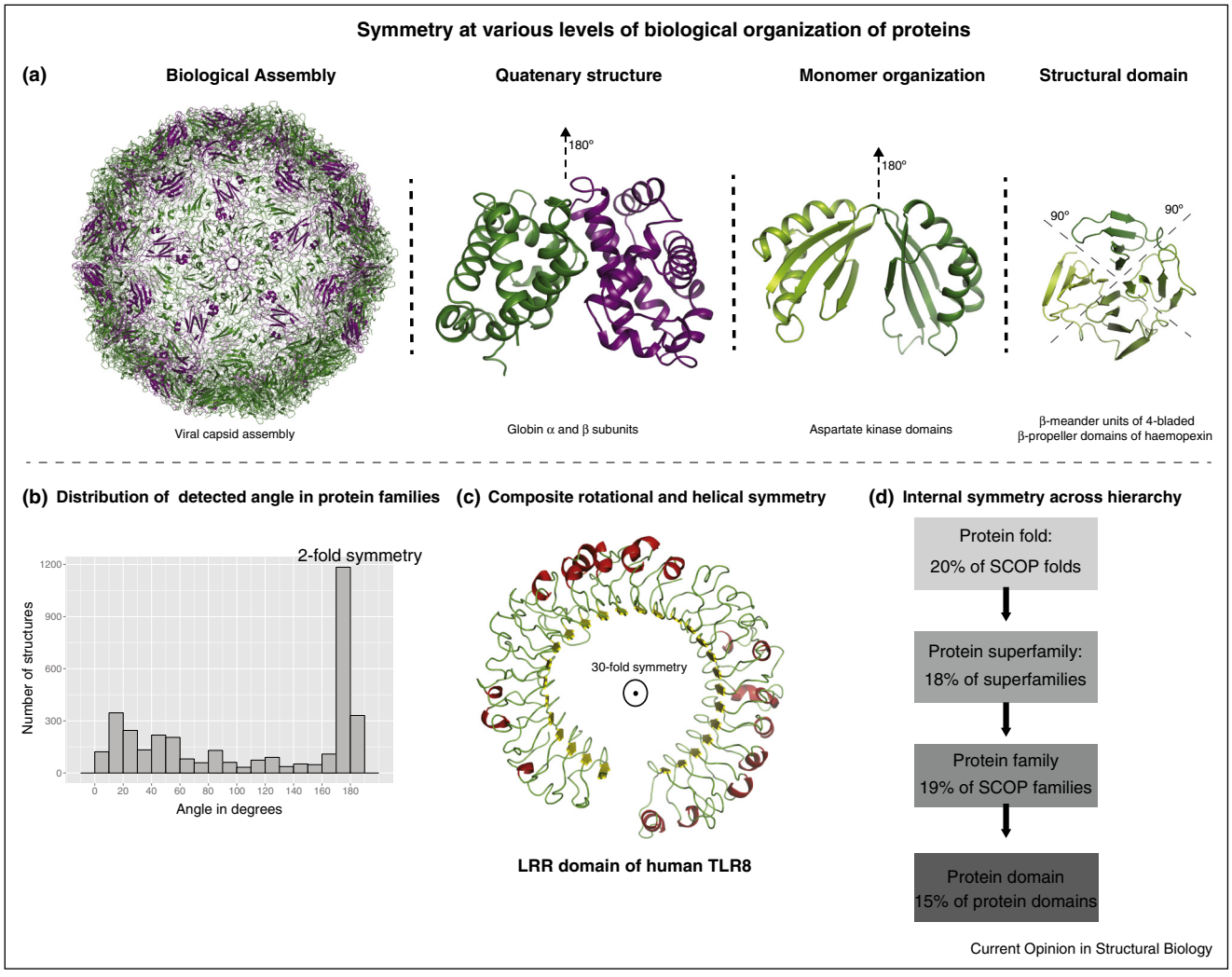
Internal symmetry and its prevalence in protein structures

Recent computational methods to systematically identify protein internal symmetry

Recently developed computational methods to identify symmetry in protein structures include SymD and CE-Symm, which are amenable to large-scale applications [23••,28••]. These encode complementary procedures that align a given protein structure to itself through systematic circular permutations while excluding diagonal matches,

☆ *Glossary*: This review focuses on internal symmetry at the level of tertiary structures and domains in proteins. Internal symmetry as defined here is independent of sequence identity, accommodates minor deviations from perfect symmetry and could also be termed as 'pseudo symmetry'. The terms 'symmetry' and 'pseudo symmetry' in the context of protein tertiary structure or domain refer to internal symmetry in protein structures.

Figure 1



Symmetry in biological systems and prevalence of internal symmetry in evolutionarily related proteins. **(a)** Symmetry is present at various levels of biological organization from: (i) over all biological assembly virus — picornavirus capsid; PDB: 4CTF, (ii) functional quaternary association — alpha and beta subunits of haemoglobin; PDB: 1HHO, (iii) tertiary protein structure — aspartate kinase domains; PDB: 2DTJ and (iv) within domain fold — hemopexin; PDB: 1QJS. **(b)** Distribution of fold of symmetry angle on the 3559 representative protein structures. The angle around 180 degrees is maximally enriched with more than one-third of structural representatives. This suggests that two-fold symmetry is the most common type of rotational symmetry in protein families. **(c)** Composite 30-fold rotational and helical symmetries in Leucine Rich Repeat extracellular domain of human Toll-like receptor 8 (PDB: 4R0A). The axis of symmetry is shown as circle with central dot, which denotes that the axis is going into the plane of paper. **(d)** Percentage of folds, superfamilies, families and domains with internal symmetry across SCOP classification hierarchy. The percentage values have been calculated based on additional file 1 of Kim *et al*. [28••] and adopted for superfamilies from Myers-Turnbull *et al*. [23••].

that is, self-matches. SymD relies on the residue level alignment of protein structures and employs an alignment scan procedure for each circular permutation of a protein structure and considers only the best optimal solution from the procedure. On the contrary, CE-Symm uses a dynamic programming method that identifies optimal off-diagonal matches between a protein structure and its circularly permuted versions. Both SymD and CE-Symm report protein structural symmetry order, angle and axis, and yield largely congruent results (see below) [23••]. The results from both of these methods provide a comprehensive view of the prevalence, nature and distribution of internal symmetry in the protein structural universe.

Various types of rotational, translational and dihedral internal symmetry detected by these methods are listed in Table 1 [23••]. Based on the current literature evidence, pure rotational symmetry beyond the order of 10 is not found in any of the determined protein structures (Table 1). The most common form of symmetry is 2-fold