FEBS Letters 584 (2010) 1059–1065

journal homepage: www.FEBSLetters.org

Generating artificial homologous proteins according to the representative family character in *molecular mechanics properties* – an attempt in validating an underlying rule of protein evolution

Xin Liu, Ya-Pu Zhao *

The State Key Laboratory of Nonlinear Mechanics, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 5 January 2010
Revised 1 February 2010
Accepted 2 February 2010
Available online 9 February 2010

Edited by Takashi Gojobori

Keywords:
Protein evolution
Protein design
Hydrophobic interaction
Hydration
Binding assay

ABSTRACT

The *molecular mechanics property* is the foundation of many characters of proteins. Based on intramolecular hydrophobic force network, the representative family character underlying a protein's *mechanics property* is described by a simple two-letter scheme. The tendency of a sequence to become a member of a protein family is scored according to this mathematical representation. *Remote homologs of the WW-domain family* could be easily designed using such a mechanistic signature of protein homology. Experimental *validation* showed that nearly *all* artificial homologs have the representative folding and bioactivity of their assigned family. Since the *molecular mechanics property* is the only consideration in this study, *the results indicate its possible role in the generation* of new members of a protein family during evolution.

1. Introduction

The amino acid sequence is believed to specify a protein's atomic structure and biological function [1]. Proteins are diversiform due to differences in residue sequence. Although their compositions are quite different, some proteins share common biological properties with one another. For instance, some remotely homologous proteins can have less than 30% identical residues. However, the reason for such functional uniformity, which arises from the diversity of intramolecular details, is still unknown.

Two levels of studies are related to protein homology research: investigations of a single physical system and those of the uniformity of multiple systems.

(i) The first type refers to studies focusing on the properties of a biomolecule-solution physical system, including the native fold, function, and conformational motion. Since only one system is investigated, the basic and universal physical principles, quantities, and methods are applicable in this type of studies. For example, the free energy of the physical system per protein is believed to play a vital role in protein folding [2,3].

(ii) The second type refers to studies focusing on the reasons behind the occurrence of homology. This includes the present study on why the folding that generates the representative family biological properties, but not other decoy folds, is specified as the native structure of a protein family member. This type of studies usually focuses on something common within a homolog set, and embodies the selection pressure during the process of choosing the eligible molecules from the outcome of the basic physical principle.

As each protein corresponds to a physical system, a set of systems must be jointly investigated so that some common mechanisms within these systems can be identified. Since multiple systems are simultaneously focused on, the methodology will be different from that of a single physical system. For example, as compositions differ across homologs, the residue interactions that contribute free energies should also vary in their corresponding physical systems, especially among those of remote homologs. Consequently, the similarity in free energy is not a necessary condition in protein homology, and the importance of free energy is ultimately decreased. Therefore, it is rational that the fundamental physical principle focusing on the homology of protein evolution is based on, but not limited to, those at single-system level. At present, due to such a shift in the object of research, there is still a gap between the physical principles of the two levels. In this paper, we present a novel level of description for a multiple system, and at-

* Corresponding author. Address: Beisihuanxi Road #15, Beijing, China.
E-mail addresses: liuxin@lnm.imech.ac.cn (X. Liu), yzhao@imech.ac.cn (Y.-P. Zhao).

0014-5793/$36.00 © 2010 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/j.febslet.2010.02.010

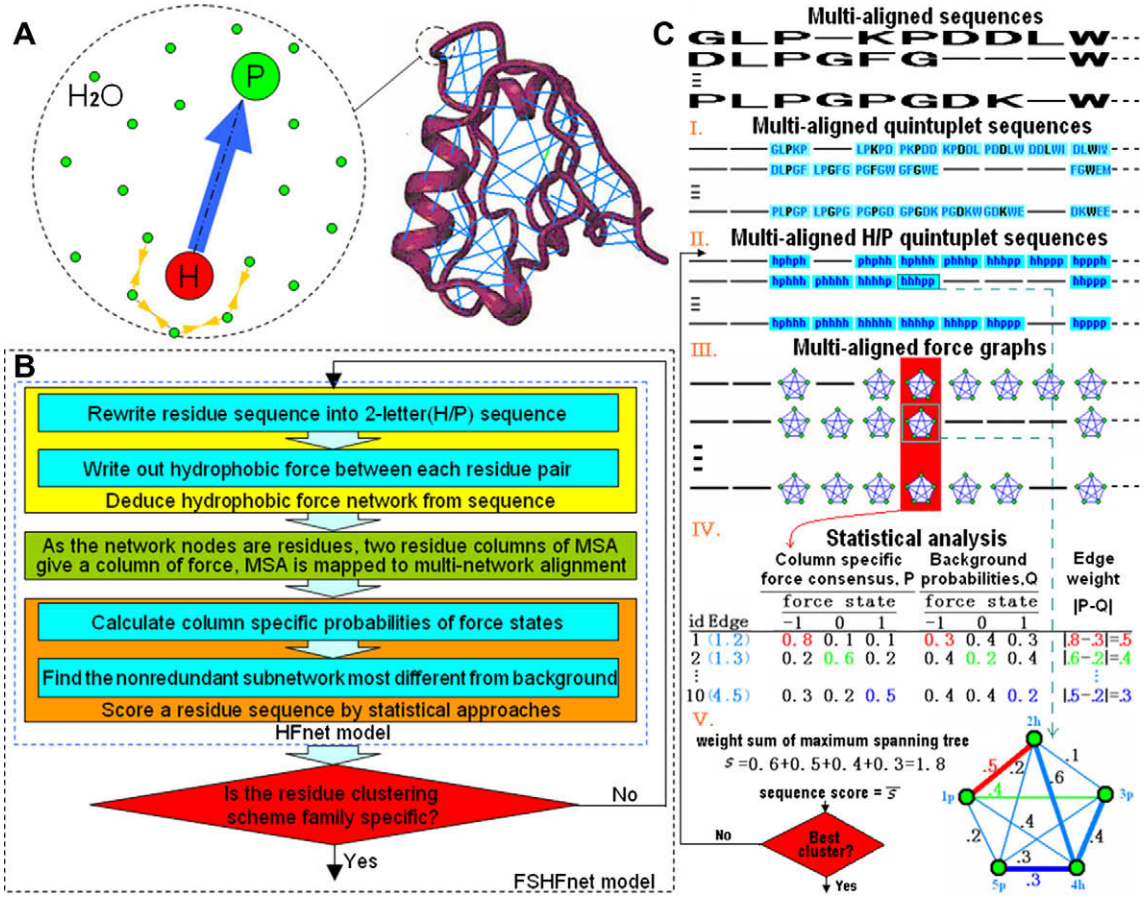

Fig. 1. Illustrations of hydrophobic force and the FSHFnet algorithm. The hydrophobic force along the virtual line of an H–P residue pair is shown in (A), with the sketch map of the hydrophobic force's origin wherein the attractions between water molecules are denoted in gold. As indicated in the flowchart (B), we tried each kind of clustering scheme and evaluated the performance of HFnet. The clustering scheme with the maximum counts of correctly identified samples in the learning set was selected as the family specific amino acid classification scheme in FSHFnet. Some details of FSHFnet are shown by examples in (C), including: (I) rewriting protein sequence into successive overlapping 5-residue units; (II) rewriting quintuplet sequence into H/P quintuplet sequence; (III) drawing a force graph of residue-to-residue interaction in each H/P quintuplet; (IV) calculating edge-specific probabilities of the occurrences of force states for each column of aligned graphs, and those of the background that are evaluated by all graphs in a background sequence set; and (V) finding the maximum spanning tree for each graph (the tree is shown in bold line, with the edge weight reflecting the difference between the occurrence of a force state and that of the background), scoring a sequence, and evaluating the residue clustering scheme.

tempt to make a step in closing up the two with a simple, empirical, but physics-based, mathematical representation.

Evolution has been the focus of protein science for a long time. Many efforts have been devoted to the study of sequence [4,5], structure [6,7] and function [8], which are biological properties that are suitable for a direct comprehension and are relatively easy to observe. But some evolutionary events can not be fully investigated without the analysis of physical mechanism. For instance, the native-structure absent homologs of a disorder protein family can carry out their biological functions by dynamic conformational changes. Since structure or distance is no longer important in these cases, the biological properties of these homologs should be determined by an physical quantity that is responsible for the change of movement state of polypeptide or the change of movement tendency, that is, the force. The molecular mechanics property may be more conserved than the structure [9]. Therefore, there is a requirement to investigate protein evolution in an aspect of physics. In particular, the molecular mechanics property is the basis of side-chain fluctuations, movement of active site loops, structural exchanges and rearrangements, and other processes that are vital to protein biological properties. The investigation of such property has been regarded as a new hotspot of protein evolution [10].

There are vast complexities of interactions in the protein that can be coped with quantum mechanical, molecular mechanical, or other treatments. As multiple systems are jointly investigated, the complexity increases drastically in the study of protein homology. To reduce the difficulty involved, a feasible option is to adopt a coarse-grained scheme that focuses on significant items but still monitors the secondary factors.

Hydrophobic interactions have been suggested as the driving force of protein folding [3], and play an important role in protein function [11]. In an aqueous solution, a hydration shell is formed on a protein surface by at least two layers of water molecules [12]. The water molecules that surround a hydrophobic (H) residue attract one another, resulting in a radial compressive stress on the amino acid. No such force is loaded on a polar (P) residue. As shown in Fig. 1A, this results in a force between each residue pair, and subsequently, a complicated force network in each protein molecule. This network is a representation of the consequence of hydration in a corresponding physical system.

In agreement with Frauenfelder's observation that internal protein motions or dynamic properties are controlled by the hydration shell, we suggest that there are some common and representative family characters in the inbuilt force networks of homologous proteins, which eventually govern the conservation of biological properties during protein evolution [13]. The maintenance of these characters would serve as the fundamental physical principle that potentially governs protein homology. We believe that if this