Biochimica et Biophysica Acta 1834 (2013) 932–951

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Review

Unusual biophysics of intrinsically disordered proteins ☆

Vladimir N. Uversky *

Department of Molecular Medicine, USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, Florida 33612, USA
Institute for Biological Instrumentation, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

ARTICLE INFO

Article history:
Received 23 October 2012
Received in revised form 21 November 2012
Accepted 12 December 2012
Available online 23 December 2012

Keywords:
Intrinsically disordered protein
Structural heterogeneity
Complex systems
Protein function

ABSTRACT

Research of a past decade and a half leaves no doubt that complete understanding of protein functionality requires close consideration of the fact that many functional proteins do not have well-folded structures. These intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered protein regions (IDPRs) are highly abundant in nature and play a number of crucial roles in a living cell. Their functions, which are typically associated with a wide range of intermolecular interactions where IDPs possess remarkable binding promiscuity, complement functional repertoire of ordered proteins. All this requires a close attention to the peculiarities of biophysics of these proteins. In this review, some key biophysical features of IDPs are covered. In addition to the peculiar sequence characteristics of IDPs these biophysical features include sequential, structural, and spatiotemporal heterogeneity of IDPs; their rough and relatively flat energy landscapes; their ability to undergo both induced folding and induced unfolding; the ability to interact specifically with structurally unrelated partners; the ability to gain different structures at binding to different partners; and the ability to keep essential amount of disorder even in the bound form. IDPs are also characterized by the "turned-out" response to the changes in their environment, where they gain some structure under conditions resulting in denaturation or even unfolding of ordered proteins. It is proposed that the heterogeneous spatiotemporal structure of IDPs/IDPRs can be described as a set of foldons, inducible foldons, semi-foldons, non-foldons, and unfoldons. They may lose their function when folded, and activation of some IDPs is associated with the awaking of the dormant disorder. It is possible that IDPs represent the "edge of chaos" systems which operate in a region between order and complete randomness or chaos, where the complexity is maximal. This article is part of a Special Issue entitled: The emerging dynamic view of proteins: Protein plasticity in allostery, evolution and self-assembly.

1. Introduction

Recently, intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs) gained significant attention of the researchers primarily due to the fact that the existence of such biologically active molecules without unique 3D-structure clearly contradicts to the traditional "one protein–one structure–one function" paradigm [1–7]. Before they were finally recognized as a unique and important extension of the protein kingdom, these highly dynamic proteins with important biological functions were discovered and rediscovered multiple times. The complex and lengthy pathway to recognition left a wide trail of terms used for the description of these proteins, which were depicted as floppy, pliable, rheomorphic [8], flexible [9], mobile [10], partially folded [11], natively denatured [12], natively unfolded [3,13], natively disordered [6], intrinsically unstructured [2,5], intrinsically denatured, [12] intrinsically unfolded [13], intrinsically disordered [4], vulnerable [14], chameleon [15], malleable [16], 4D [17], protein clouds [18], and dancing proteins [19], among several other terms.

This trail of terms can be considered as "prehistory" of intrinsic disorder. For early researchers, it was clear that biologically active but non-folded proteins are different from "normal" globular, transmembrane, and fibrous proteins. For a long time, each such a protein was considered as an exception from a general rule, where unique sequence defined unique 3D-structure that was crucial for unique function. The multitude of terms used to describe IDPs in past not only reflects the creativity of researchers but also indicates difficulties they faced while trying to find an appropriate way of portraying these proteins. Although none of the terms proposed for defining biologically active proteins without unique structure is perfect, the term "intrinsically disordered protein" is currently used more often than any other terms. The qualifier "disordered" is always used in the context of a comparison between a single, ideal, well-defined situation, and the actual situation which we consider to be only one of many different possibilities, none of which deserves to be singled out [20].

☆ This article is part of a Special Issue entitled: The emerging dynamic view of proteins: Protein plasticity in allostery, evolution and self-assembly.

* Department of Molecular Medicine, University of South Florida, Morsani College of Medicine, 12901 Bruce B. Downs Blvd, MDC07, Tampa, Florida 33612, USA. Tel.: +1 813 974 5816; fax: +1 813 974 7357.
E-mail address: vuversky@health.usf.edu.

1570-9639/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2012.12.008

Systematic bioinformatics analyses clearly indicated that IDPs are highly abundant in any given proteome [4,7,21–23]. Therefore, these proteins have moved from a category of obscure and rare exceptions to the novel class of proteins, whose functionality is determined by the lack of stable structure, and which are very common in nature. Functions of IDPs/IDPRs are complementary to functions of ordered proteins and domains [4,24,25], with disordered proteins being typically involved in regulation, signaling and control pathways [26,27]. Because of their unique functionality, dysfunctions of IDPs are known to be associated with various human diseases, such as cancer, cardiovascular disease, amyloidosis and neurodegenerative diseases [28].

One of the goals of this review is to put together a set of old and new concepts (such as the ideas that IDPs are characterized by high spatiotemporal heterogeneity; that they have rough and relatively flat energy landscapes; that IDPs might contain foldons, inducible foldons, semi-foldons and non-foldons; that some ordered proteins might have unfoldons, i.e., regions that have to undergo order-to-disorder transition in order to make protein active; that globally, there is a phenomenon of dormant disorder, where some proteins are inactive when they are ordered, and become activated when they become more dynamic; and that IDPs can be considered as the "edge of chaos" systems) that would inevitably provoke disputes and therefore would initiate new studies. I do realize that some of the concepts are not well-developed and some might be naïve. However, they are present here since they can be clearly taken as "food for thoughts".

2. Unusual biophysics of IDPs

2.1. Behold the root: Peculiarities of the amino acid sequence provide an answer to the question "To fold or not to fold?"

IDPs/IDPRs are different from ordered proteins and domains already at the level of their amino acid sequences. In fact, the sequence peculiarities define both the ability of ordered proteins to fold and the ability of IDPs to stay non-folded. Therefore, the well-known Anfinsen's dogma for foldable proteins stating that information dictating the native fold of protein domains is encoded in their amino acid sequence [29] and therefore at optimal conditions (temperature, solvent concentration and composition, etc.), the native structure represents a unique, stable and kinetically accessible minimum of the free energy, can be converted into similar statement for IDPs/IDPRs, namely, information dictating the lack of folded structure in disordered proteins is encoded in their amino acid sequence. In other words, the absence of rigid structure in IDPs may be somehow encoded in the specific features of their amino acid sequences [1,3,4,7,25,30]. In agreement with this hypothesis, the unusual amino acid sequence compositions were observed for some IDPs, which in extreme cases were unfolded at the physiological conditions due to the presence of numerous uncompensated charged groups (often negative) that defined a high net charge of these proteins at neutral pH [13,31,32], and a low content of hydrophobic amino acid residues [31,32]. In fact, based on the comparative analysis of 275 natively folded and 91 natively unfolded proteins (i.e., proteins which at physiologic conditions have been reported to have the NMR chemical shifts of a random-coil), and/or lack significant ordered secondary structure (as determined by CD or FTIR) it was revealed that the combination of low mean hydropathy and relatively high net charge represents an important prerequisite for the absence of compact structure in proteins under physiological conditions [3]. The resulting charge-hydropathy (CH) plot method can distinguish ordered and disordered proteins based only on their net charges and hydropathies [3]. From the physical viewpoint, such a combination of low hydropathy with high net charge as a prerequisite for intrinsic disorder makes perfect sense: high net charge leads to charge-charge repulsion, and low hydropathy means less driving force for protein compaction. In other words, these features are characteristic for highly disordered IDPs with the coil-like (or close to coil-like) structures, which obviously represent only a small subset of the entire IDP realm.

At the more detailed level, there are numerous differences in the amino acid compositions of ordered and disordered proteins and many IDPs clearly share at least some common sequence features [1,33]. Here, IDPs/IDPRs are significantly depleted in so-called order-promoting residues that include bulky hydrophobic (Ile, Leu, and Val) and aromatic amino acid residues (Trp, Tyr, and Phe), which would normally form the hydrophobic core of a folded globular protein, and also possess low content of Cys (which is often contribute to the protein conformational stability via the disulfide bond formation or coordination of different prosthetic groups) and Asn residues. On the other hand, IDPs/IDPRs were shown to be substantially enriched in disorder-promoting, amino acids, that were polar Arg, Gly, Gln, Ser, Glu, and Lys, and hydrophobic, but structure breaking Pro and hydrophobic Ala [4,7,24,34–36]. Based on the ability of amino acids to promote order and disorder, a special amino acid scale was introduced that was able to discriminates between ordered and intrinsically disordered proteins reasonable well [37]. Here, amino acids were ranked according to their capabilities to promote order or disorder resulting in the following scale (where amino acids are arranged from the most order-promoting to the left to the most disorder-promoting to the right): W, F, Y, I, M, L, V, N, C, T, A, G, R, D, H, Q, K, S, E, P [37].

It is clear that the amino acid sequence peculiarities of IDPs can be blamed for the unusual and unexpected behavior of IDPs. Many early IDP researchers were stunned by the peculiar features of these mysterious then members of the protein kingdom. On a personal note, my journey to the IDP field started when one sunny day, an excited colleague of mine appeared in the lab shaking a tube with a sample in his hand and shouting: "I have a funny protein here. I cannot measure its concentration. And it is extremely stable – I can boil it for a few days, but as soon as I am bringing temperature down it shows 100% activity." That funny protein was prothymosin α. Fig. 1 shows that the unusual behavior of this protein is definitely determined by its amino acid sequence. It does not have any aromatic residues and cysteins. Therefore its concentration cannot be measured spectroscopically. 64 of 111 residues in this protein have charged side groups (there are 19 D, 35 E, 2 R, and 8 K residues), whereas overall content of hydrophobic residues (L, I and V) is very low [38]. Based on this amino acid composition, it was not a big surprise to find that prothymosin α behaved as a highly disordered coil-like chain – you cannot expect that highly charged polypeptide (60% polyE/D) will have a strong tendency to fold under the physiological conditions. This luck of stable structure also explained extreme thermal and acid stability of prothymosin α — you cannot break what is already broken [38].

Differences between ordered proteins and IDPs can be further elaborated by going to the very subtle levels. However, this exercise is outside the scopes of this review. The important message is already obvious from the observations listed above, namely, sequences encoding IDPs/IDPRs are very different from sequences encoding ordered proteins and domains. In fact, these two types of sequences are so different that they can be discriminated reasonably well by numerous computational tools, where comparing and combining several predictors can provide additional insight regarding the predicted disorder [39–46]. This clearly indicates that IDPs are new and specific entities in the protein kingdom.

2.2. Sequential, structural, and spatiotemporal heterogeneity of IDPs

2.2.1. Sequence space and sequence heterogeneity of IDPs

A typical estimate of the size of the protein sequence space is 20^{100} ($\sim 10^{130}$) for a protein of 100 amino acids in which any of the normally occurring 20 amino acids can be found [47]. For a long time, discussion