Biochemical and Biophysical Research Communications xxx (2016) 1–5

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

Interface property responsible for effective interactions of protean segments: Intrinsically disordered regions that undergo disorder-to-order transitions upon binding

Divya Shaji, Takayuki Amemiya, Ryotaro Koike, Motonori Ota*

Graduate School of Information Science, Nagoya University, Nagoya 464-8601, Japan

ARTICLE INFO

Article history:
Received 13 July 2016
Accepted 19 July 2016
Available online xxx

Keywords:
Intrinsically disordered proteins
Molecular recognition
Protein interface
Relative accessible surface area (rASA)
Protein-protein interactions
Protean segments (ProSs)

ABSTRACT

Proteins that lack a well-defined conformation under native conditions are referred to as intrinsically disordered proteins. When interacting with partner proteins, short regions in disordered proteins can undergo disorder-to-order transitions upon binding; these regions are called protean segments (ProSs). It has been indicated that interactions of ProSs are effective: the number of contacts per residue of ProS interface is large. To reveal the properties of ProS interface that are responsible for the interaction efficiency, we classified the interface into core, rim and support, and analyzed them based on the relative accessible surface area (rASA). Despite the effective interactions, the ProS interface is mainly composed of rim residues, rather than core. The ProS rim is more effective than the rim of heterodimers, because the average rASAs of ProS rim, which is significantly large in the monomeric state, provides a large area to be used for the interactions. The amino acid composition of ProSs correlated well with those of heterodimers in both the core and rim. Therefore, the composition cannot explain why the rASAs of the ProS rim are large in the monomeric state. The balance between a small core and a large rim, and the large solvent exposure of the rim in the monomeric state, are the key to the disorder-to-order transition and the effective interactions of ProSs.

1. Introduction

Intrinsically disordered proteins (IDPs) or intrinsically unstructured proteins (IUPs) are proteins that lack stable 3-dimensional structures under physiological conditions [1,2]. These IDPs contain long or short intrinsically disordered regions (IDRs) [3]. IDPs are more abundant in eukaryotic proteomes than in archaea and prokaryotes, and are preferentially localized in the nucleus [4]. They are involved in numerous biological activities such as signal transduction and transcriptional regulation [2,5]. The dysfunctions or unnatural interactions of IDPs are associated with human diseases, including cancer, cardiovascular disease, neurodegenerative diseases and amyloidoses [6,7]. IDPs usually use short segments of IDRs that undergo disorder-to-order transitions upon binding to their partners (i.e., coupled folding and binding) [8,9]. We call these short segments protean segments (ProSs) and deposit them in the intrinsically disordered proteins with extensive annotations and literature (IDEAL) database [10,11]. The concepts of molecular recognition features (MoRFs) [12–15] and eukaryotic linear motifs (ELMs) or short linear motifs (SLiMs) [16,17] are similar to ProSs, but the definitions are partially different from each other [10,11]. As such binding regions (e.g., ProSs) are essential for the molecular function of IDPs, more attention has been paid to their interactions, and several characteristics have been revealed [12–15,18–22]. In particular, it has been indicated that the interactions of ProSs are effective [18]: on average, the number of contacts of ProS interface with its interaction partners is larger than that of globular proteins (e.g., heterodimers). This has been explained by their unique interaction mode employing coupled folding and binding [18], but the details are still unclear. In this study, we focused on the interface of ProSs and compared it with that of heterodimers. The interface residues were further classified into core, rim and support [19,23], and their relative solvent accessible surface areas (rASA) were analyzed in detail. The residues in the interface core are the most buried residues upon protein binding, generally at the central region of the interface, and play an important role in the interaction, like hot spots [24,25]. The residues in the interface rim are located on the outer edges of the interface that remain partially

* Corresponding author.
E-mail address: mota@is.nagoya-u.ac.jp (M. Ota).

http://dx.doi.org/10.1016/j.bbrc.2016.07.082
0006-291X/

exposed to the solvent [23]. The support residues play an insignificant role in the interaction, which represents the intersection between the interior and the interface. The major finding of our work is that the ProS interface is mainly composed of rim residues even though it has effective interactions. The key to effective interactions of ProSs is the solvent exposure of rim residues in the monomeric state.

2. Materials and methods

2.1. ProSs and heterodimers

All ProSs (210) in 70 protein sequences were collected from the IDEAL database (as of August 2013) [10,11]. If more than one ProS were found in a protein and their positions overlapped, we chose the longest ProS. The sequence redundancy was removed with 80% sequence similarity (based on the CLUSTALW alignment [26]). Hierarchical clustering was done with R [27] using complete-linkage clustering and the longest ProS in a cluster was selected as the representatives. A non-redundant set contained 99 ProSs (Table S1). DNA-binding ProSs and one-to-many binding ProSs (a single ProS binds to two or more different partners, [22]) were discarded. Both the X-ray and NMR structures were used in this study.

A non-redundant dataset of 276 heterodimers was selected from the Protein Data Bank (PDB) [28], using the PDB's advanced search interface (as of July 2014). The search criteria satisfied the following conditions: (1) less than 30% sequence identity; (2) the macromolecule type contained only proteins; (3) the oligomeric state was heterodimer; (3) each chain was greater than 100 residues; and (4) structures determined by X-ray crystallography had higher than 3 Å resolution. Only smaller protomers were analyzed as the reference for ProSs.

2.2. Amino acid propensity

The propensities of amino acids are represented as the Chou–Fasman parameters [29] $CF(a,P) = \frac{N^a(P)/N(P)}{N^a_{all}/N_{all}}$, where N^a (P) is the number of amino acid residue a in place P, N (P) is the total number of residues in P, N^a_{all} is the total number of amino acid residue a in the protein sequence, and N_{all} is the total number of residues in the protein sequence. In P, we considered the interface, core and rim residues in ProSs and heterodimers. To calculate the reference states (the denominator), we used SCOP25 proteins (version 1.75) [30].

2.3. Relative ASA and residue contact

We classified the residues into surface, interior and interface. Based on the definitions by Levy [23], the interfaces were further classified into core, rim and support. The relative solvent accessible surface area (rASA) is defined as the total accessible surface area (ASA) of the residues in a protein structure normalized by the ASA of the residues in the most exposed state to a solvent molecule, generally water [31]. The rASAs of each residue, in the monomeric and complex states (rASAm and rASAc, respectively) were computed for ProSs and heterodimers using the program Naccess [32], which is an implementation of Lee and Richard's algorithm [33]. ΔrASA = rASAm − rASAc. The rASAs were averaged for interface, core and rim residues, to derive the average rASAs of proteins.

Two residues, i and j, were considered to be in contact if any atom of residue i was within a distance of <4.5 Å with any atom of residue j [34,35]. We calculated the number of external contacts for ProSs and heterodimers at the interface, core and rim. External contacts are defined as the contacts between the proteins and their interaction partners. The average number of contacts at the interface, core and rim was calculated for each ProS and heterodimer.

2.4. Statistical analysis

Wilcoxon rank-sum test was performed by RStudio [36] to calculate the P-values (Table S2). $P < 0.01$ was considered statistically significant.

3. Results and discussion

3.1. Composition of interfaces and effective interactions of ProSs

Based on the protein dimeric structures, amino acid residues in ProSs and heterodimers were classified into surface, interior and interface residues (Fig. S1). As in nature ProSs have a small number of intra-chain contacts, and only adopt structures when interacting with partner proteins, ProSs have a larger number of interface residues and a smaller number of interior residues than heterodimers. Fig. 1A and B further breaks down the composition of the interface residues into core, rim and support residues in ProS and heterodimer interfaces, respectively. In the ProS interface, core residues are less abundant (33.7%) compared with the heterodimer interface (36.8%). Moreover, in ProSs, the interface is mainly composed of rim (64.7%), which is nearly double that observed in heterodimers (35.3%). The distribution of the rates of core and rim (Fig. S2) is significantly different in ProSs and heterodimers as assessed by the Wilcoxon rank-sum test (P-values: core = 4.5e-05, rim = 1.3e-40). In summary, the ProS interface is composed of a small core and a large rim. This statement based on the rates gives a slightly different representation from the results of a previous report on absolute values [19], denoting that the number of residues in the core of the 1D segment (corresponding to ProS) is smaller than that of the 3D complex proteins (heterodimers), but in the rim, the numbers of residues are almost equal.

To examine the efficiency of interactions, we calculated the average number of (inter-chain) contacts of interface residues for each ProS and heterodimer, and compared their distributions. As was shown in Fig. 1C as well as in a previous report [18], the ProS interface can be in contact with a larger number of residues of the interaction partners compared with the heterodimer interface, confirming that the ProS interaction is effective. However, this result seems to be inconsistent with our results of the ProS interface composition (Fig. 1A and B), because for effective interactions, a large core and a small rim are expected. To analyze the contribution of residues, the average number of contacts was derived individually for the core and the rim of the interfaces (Fig. 1D and E). Apparently, on average core residues have a larger number of contacts compared with the rim, confirming that having a core should be reasonable for effective interactions. Moreover, it is noticeable that in both the core and rim cases, the average number of contacts by ProS residues is larger than that by heterodimers (see the P-values in Table S2). This indicates that ProS residues contribute to effective interactions not only through the core, but also through the rim. In particular, because of their abundance, the efficiency of the ProS rim is remarkable. To prove this hypothesis, we ignored the interactions of the core, rim or support individually, and calculated the average number of contacts again (Fig. S3). When we took into account the rim contacts and ignored those of the core and support, the average number of contacts of the ProS was different from that of the heterodimer (Fig. S3A and C). Only when we ignored the rim contacts, the average number of ProS contacts was almost equal to that of the heterodimer (Fig. S3B),

Please cite this article in press as: D. Shaji, et al., Interface property responsible for effective interactions of protean segments: Intrinsically disordered regions that undergo disorder-to-order transitions upon binding, Biochemical and Biophysical Research Communications (2016), http://dx.doi.org/10.1016/j.bbrc.2016.07.082