Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Virology

Intrinsically disordered proteins implicated in paramyxoviral replication machinery

Guillaume Communie[1,2,3], Rob WH Ruigrok[4,5,6], Malene Ringkjøbing Jensen[1,2,3] and Martin Blackledge[1,2,3]

The development of mechanistic insight into the molecular basis of how intrinsically disordered proteins function is a key challenge for contemporary molecular biology. Intrinsic protein disorder is abundant in the replication machinery of paramyxoviruses. In order to study this kind of protein, new methods are required that specifically take account of the highly dynamic nature of the chain, and describe this disorder in quantitative terms. Here we review recent studies of conformational disorder in paramyxoviral phosphoproteins and nucleoproteins using solution-based approaches such as nuclear magnetic resonance.

Addresses
[1] Université Grenoble Alpes, Institut de Biologie Structurale (IBS), Grenoble, France
[2] CEA, DSV, IBS, Grenoble, France
[3] CNRS, IBS, Grenoble, France
[4] Université Grenoble Alpes, UVHCI, Grenoble, France
[5] CNRS, UVHCI, Grenoble, France
[6] Unit for Virus Host-Cell Interactions, Université Grenoble Alpes-EMBL-CNRS, Grenoble, France

Corresponding author: Blackledge, Martin (martin.blackledge@ibs.fr)

Current Opinion in Virology 2014, **5**:72–81

This review comes from a themed issue on **Virus structure and function**

Edited by **Wah Chiu, Thibaut Crépin** and **Rob WH Ruigrok**

For a complete overview see the Issue and the Editorial

Available online 13th March 2014

http://dx.doi.org/10.1016/j.coviro.2014.02.001

Introduction

The realization that a significant fraction of proteins in eukaryotic organisms comprise unstructured or disordered regions of more than 50 amino acids in length has revolutionized our understanding of the molecular basis of biological function [1–3]. Intrinsically disordered proteins (IDPs) are functional in the absence of a stable three-dimensional structural fold, and as such they impose a new perspective on the relationship between primary protein sequence and function. Recent estimates indicate that over 1300 proteins have been classified as intrinsically disordered, on the basis of diverse experimental observations [4], and that these proteins are particularly prevalent in signalling and regulation and play organizational roles in scaffold and hub proteins. IDPs are involved in numerous human pathologies, and rational design of pharmacological solutions to these questions awaits a molecular understanding of the role of protein flexibility in disease [5–7]. The primary sequence of IDPs is characterized by a high prevalence of so-called disorder-promoting amino acids exhibiting low hydrophobicity and high net charge compared to globular, folded proteins [8–10]. This tendency allows for genome-wide prediction of disorder, revealing that eukaryotic proteins have a significantly higher tendency to disorder than prokaryotic proteins [11–16]. Intrinsic disorder has been shown to be abundant in proteins implicated in human diseases such as cancer, neurodegenerative diseases and diabetes [6,17], as well as being present in viruses in highly variable, but on average similar levels as are found in eukaryotic organisms [18–20,21•]. RNA viruses in particular have been predicted to present a high occurrence of weakly packed structural elements [22•] and the level of disorder within RNA viruses has been found to increase with genome size [21•]. In addition to the proposed advantages of protein disorder relating to highly dynamic, promiscuous interaction hubs, disorder in viral proteins has been suggested to provide specific advantages, related to efficient use of limited genetic material, such as alternative splicing and overlapping genes [19]. An excellent review recently highlighted diverse mechanisms by which short linear motifs in disordered viral proteins can mimic, and thereby deregulate host machinery in a highly efficient manner, despite the necessarily compact viral genome [23•].

Intrinsically disordered proteins

The study of the conformational relationship between primary sequence and biological function in IDPs is complicated by the highly dynamic nature of the protein [24]. Classical structural biology techniques, such as crystallization, are therefore in general inappropriate due to the inherent conformational heterogeneity of the chain. Instead, it is necessary to develop descriptions of IDPs that explicitly account for this structural disorder, for example by using an ensemble of rapidly interconverting conformers that can describe the available conformational space sampled by the unfolded chain [25–27], optimally in both free, and bound forms of the protein. Solution state methods, are the most appropriate for studying IDPs, and nuclear magnetic resonance (NMR) spectroscopy is one of the most powerful techniques

[28,29]. IDPs can be expected to adopt many different conformations in solution, but as long as these conformers interchange on timescales faster than tens of microseconds, each NMR peak provides atomic resolution characterization of local and long-range conformational properties. Similarly small angle X-ray and neutron scattering [30–32], or single molecule techniques, such as Förster resonance energy transfer (FRET) [33,34] provide information about the volume space and long-range intramolecular contacts respectively. The remarkable increase in interest in IDPs over the last decade has fostered vigorous development of associated techniques to interpret experimental data in terms of representative conformational ensembles [35–38]. Mechanistic insight into protein interactions involving intrinsically disordered regions is of fundamental importance for understanding many biological processes, and represents one of the key challenges for contemporary molecular biology. The structural plasticity of IDPs is thought to provide functional modes that are inaccessible to folded proteins, including folding-upon-binding [39], the formation of transient complexes that co-localize hosts via non-specific interactions [40], interactions with rapid dissociation rates where the IDP remains dynamic in a so-called 'fuzzy' complex, or entropic bristles that frustrate access to specific sites. However little is known about the molecular basis underpinning interactions between IDPs and their physiological partners, and many interaction pathways have been evoked, both theoretically [41–43] and experimentally [44–46], for example folding-upon-binding via conformational selection or induced fit, or so-called fly-casting [47], where non-specific interactions between distant sites outside the interaction site facilitate the primary interaction.

Intrinsically disordered proteins involved in the replication of paramyxoviruses

In this review we describe intrinsic disorder in paramyxoviruses, non-segmented negative strand RNA viruses that include Measles virus (MeV), and the emerging human pathogens Hendra (HeV) and Nipah (NiV) viruses, as well as Sendai virus (SeV), a paramyxovirus that causes bronchiolitis in mice and primates. Paramyxoviral genomes are packaged into a large helical nucleoprotein assembly, the nucleocapsid [48–50], formed by multiple copies of the nucleoprotein (N) and that have been described in detail using electron microscopy [51–53]. The N-RNA complex is the template for transcription and replication by the viral polymerase complex consisting of the large protein (L) that carries the RNA-dependent RNA polymerase activity and the phosphoprotein (P), the polymerase co-factor [54]. Paramyxoviruses express their own machinery for transcribing and replicating their genome, and although the proteins participating in this machinery have been identified, little is known at a molecular level about transcription and replication. Both processes are thought to be initiated by the interaction between N and the polymerase associated P, and both of these proteins are predicted to comprise significant conformational disorder [55–57].

Disorder in paramyxoviral nucleoproteins and interactions with phosphoproteins

N comprises two distinct regions (Figure 1): a core domain (N_{CORE}) of approximately 400 amino acids, that is essential for the interaction with RNA, and a C-terminal domain, often known as N_{TAIL}, of approximately 125 amino acids, that is predicted to be intrinsically disordered along its entire length with enhanced helical propensity in the C-terminal region that mediates the interaction with P. The atomic structure of N_{CORE} is unknown [50], although crystal structures of related members of the Mononegavirales family, respiratory syncytial virus (RSV) [58], vesicular stomatitis virus (VSV) [59] and rabies virus have been determined [60]. The phosphoprotein is also predicted to comprise both folded and unfolded domains (Figure 2) [55,56]. The N-terminal domain, whose length varies between 300 and 470 amino acids, is predicted to be highly disordered, and comprises annotated phosphorylation sites [61–63] as well as sites that have been identified to bind proteins involved in the innate immune system [64]. This domain also comprises an N_{CORE}-binding site that is thought to chaperone monomeric RNA-free N before nucleocapsid assembly on the RNA genome [54]. This is followed by an oligomerization domain that has been shown in SeV [65], MeV [66] and NiV [67] to constitute a tetrameric coiled-coil domain. In the case of MeV the additional capping domain found in SeV is absent, and the tetramerization domain is constituted uniquely by a single helix. This domain is followed by a shorter unfolded domain (around 100 amino acids), which precedes the small C-terminal domain, known as PX or XD, but that will be hitherto referred to as XD, that represents the N_{TAIL} binding domain. XD has been shown, using NMR for SeV [68] and X-ray crystallography for MeV [69,70] and HeV [71•] to adopt a structurally conserved three-helix bundle.

Early SAXS and unassigned homonuclear NMR measurements on MeV N_{TAIL} indicated its intrinsically disordered nature, while circular dichroism (CD) measurements were used to infer an increase in structure upon interaction with the XD domain [72,73]. Subsequent NMR studies of the XD domain of MeV demonstrated helical folding of the N_{TAIL} peptide upon binding and the helical peptide was co-crystallized with XD using a chimeric construct [69]. More detailed studies of a 120 amino acid domain from SeV, that included the folded XD domain, using NMR residual dipolar couplings (RDCs) and SAXS, demonstrated that the predicted disorder in the 60 amino acids preceding XD were indeed devoid of structure and highly dynamic in solution [74]. The XD domain itself, although folded in solution, was also subsequently shown to be highly dynamic, exhibiting significant conformational