Structural Disorder in Viral Proteins

ARTICLE *in* CHEMICAL REVIEWS · MAY 2014

Impact Factor: 46.57 · DOI: 10.1021/cr4005692 · Source: PubMed

CHEMICAL REVIEWS

Review

pubs.acs.org/CR

Structural Disorder in Viral Proteins

Bin Xue,[†] David Blocquel,[§,∥] Johnny Habchi,[§,∥] Alexey V. Uversky,[⊥] Lukasz Kurgan,[#] Vladimir N. Uversky,*[,‡,▽] and Sonia Longhi*[,§,∥]

[†]Department of Cell Biology, Microbiology and Molecular Biology, College of Fine Arts and Sciences, and [‡]Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, Florida 33620, United States

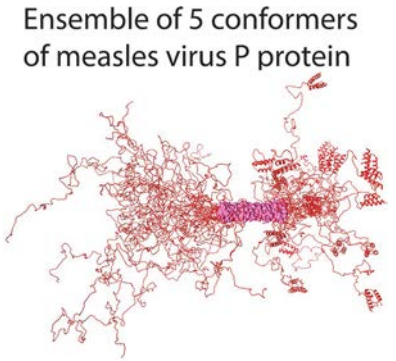
[§]Architecture et Fonction des Macromolécules Biologiques (AFMB) UMR 7257, Aix-Marseille Université, 13288 Marseille, France

[∥]AFMB UMR 7257, CNRS, 13288 Marseille, France

[⊥]Center for Data Analytics and Biomedical Informatics, Department of Computer and Information Sciences, College of Science and Technology, Temple University, Philadelphia, Pennsylvania 19122, United States

[#]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2R3

[▽]Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Russia



CONTENTS

1. INTRODUCTION

One of the core standards of biochemistry is that the function of a protein relies on its 3D structure. This idea posits that all the information required for a protein to perform its biological function(s) is contained in its amino acid sequence and that the protein can only carry out these functions once it has been

Special Issue: 2014 Intrinsically Disordered Proteins (IDPs)

Received: October 10, 2013

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX–XXX

folded into a particular structure.[1] This notion, however, has been challenged over the past 20−25 years due to the finding of a myriad of counterexamples that suggest that many proteins and protein regions in fact possess a partly or entirely disordered structure. These intrinsically disordered proteins (IDPs) or regions (IDRs) do not have a stable or unique 3D structure in solution, in physiological conditions of both pH and salinity, and when lacking a partner. Instead, these proteins exist as dynamic ensembles of conformations that do not have a stable folded structure and still carry out their respective biological activities.[2−13] Proteins and protein regions involved in the establishment of numerous interactions, such as proteins/regions implicated in signaling, recognition, and regulation activities, are enriched in intrinsic disorder (ID).[2,5,6,14−23] These and other crucial functional roles explain the high abundance of IDPs in all species.

IDPs are defined by a distinct set of specific features, in terms of both their composition and the nature of their amino acid sequences. They possess a high content in charged residues and a low content in hydrophobic residues, which consequently distinguishes them from regular, globular proteins. The unique nature of IDP sequences has led to the development of various algorithms designed for disorder prediction. These predictors enabled calculation of an estimate of the amount of disorder in a number of biological systems. The studies that utilized these algorithms showed that ID is abundant in nature, with many proteins being disordered along their entire length. Indeed, predictions on representative genomes from the three kingdoms of life (i.e., bacteria, archaea, and eukaryotes) confirmed the prevalence of disorder, although the disorder amount differs significantly between them.[4,24] More recently, ID was shown to be abundant in proteins of parasitic protozoa.[25] Those studies unveiled that, as the complexity of the studied organism increases, there is also a notable increase in the length and frequency of the disordered regions it possesses.[4,24,26] In agreement, earlier predictions revealed that disordered regions with more than 30 consecutive residues can be found in as many as 7−30% of prokaryotic proteins, with this number further increasing to 45−50% in eukaryotes.[4,24,27−31]

In another study carried out by another group who made use of a different disorder predictor, long IDRs that are wholly disordered were predicted to occur in different amounts, representing as much as 33% of eukaryotic proteins and more than 10% of all eukaryotic proteins.[32] It is to be noted, though, that the level of disorder seems to represent a strong adaptive trait that shows correlation with the environment or pathogenic lifestyle.[33,34] In all, the signaling functions linked with the complexity of the organism seem to represent a major factor determining the general prevalence of disorder in an organism, on which various adaptive changes reflecting the environment and lifestyle of the organism operate.

In particular, viral proteins stand to gain the most from the flexibility that partially or wholly unfolded proteins offer, as viruses have to quickly adapt to changes in their environment, survive in both their hosts and their hosts' environments, and evade the hosts' defense mechanisms. To accomplish this, viral genomes exhibit very high mutation rates, in the range of 10^{-5}−10^{-3} nucleotide exchanges per position per generation for RNA viruses and in the range of 10^{-8}−10^{-5} for DNA viruses; bacteria and eukaryotes, on the other hand, on average have a mutation rate of 10^{-9}.[35] Furthermore, as viruses possess highly compact genomes and, often, overlapping reading frames, one mutation has the potential to affect more than one viral protein.[36] Viral proteins tend to be involved in various interactions with host cell components, with these interactions being established over the life cycle of the virus, beginning with entry, proceeding to virus assembly, and then culminating in the exit of new infectious particles. As such, viral proteins have to interact with various components of the host, including membranes, nucleic acids, and proteins. ID, and the functional advantages that it confers, could explain how a viral protein is able to carry out these functions. Indeed, the lack of a rigid 3D structure allows IDPs/IDRs to perform various interactions with several partners at once. IDRs in particular can serve as a flexible linker between functional domains, which encourages binding and promiscuity. These flexible linkers are also able to aid viral proteins in eluding the host cell's immune system, thanks to interactions with host proteins which make viral epitopes harder to recognize by the host's immune system. Finally, the lack of structural constraints of IDRs may help in tolerating the high mutation rates that are typically encountered in viruses. In agreement with these expectations, after the first and seminal observations that disorder is abundant in proteins of the replicative complex of paramyxoviruses,[37−39] an increasing amount of computational and experimental evidence has been obtained in the past 10 years, suggesting an abundance of disorder in viral proteins (see refs 40−44 and references therein). In this survey we outline some of the peculiar structural characteristics of viral proteins and overview bioinformatics studies pointing out the abundance of ID in viruses along with the relatively few examples where ID was experimentally shown in viral proteins. We also discuss the role of ID in the functions of the various viral proteins.

2. UNIQUE PROPERTIES AND ORIGIN OF VIRUSES

Among replicating organisms, viruses are the most abundant entities:[45] indeed, the total number of cells is less than that of virus particles by at least an order of magnitude.[47,48]For example, there can be as many as 2.5×10^8 virus particles in 1 mL of natural water.[46] Viruses are parasitic organisms that can be found in high abundance in various infected cells, including Eukarya, Archaea, and Bacteria (or even inside other viruses).[45,48,49] Strikingly, the ability of a virus to infect another virus was recently established by the discovery of Sputnik, a small icosahedral virophage.[50] Sputnik infects a member of the Megaviridae family, *Acanthamoeba polyphaga* mimivirus (APMV), which, in turn, infects amoeba.[51−53] Infection by Sputnik is deleterious to the host virus, leading to abnormal APMV capsid assembly and to the appearance of abortive viral forms.[50] This is because of the fact that the multiplication of Sputnik takes place within the cytoplasm-independent APMV replication center, where the final morphogenesis of APMV normally takes place.[54]

Viruses are structurally very simple, display various shapes, and do not possess a unique common morphology. In all viruses, a single- or double-stranded DNA or RNA genome is encapsulated within a capsid, i.e., a protective protein coat. Enveloped viruses contain an additional lipid envelope that includes a number of membrane proteins; this envelope is located above another proteinaceous coat, which is known as the matrix. Finally, some complex viruses possess various regulatory and accessory proteins, together with nonstructural proteins assisting the assembly of the viral capsid.

Viruses display a broad variety in the structure of their genomes and in the mechanisms of their transcription and replication. Viral genomes can be of single- or double-stranded

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX−XXX

DNA nature or single- or double-stranded RNA nature and can be transcribed via positive sense, negative sense, or ambisense transcription mechanisms. This diversity in the nature of genomes and in the replication mechanisms has been used as a basis of the classification of viruses into seven major classes.[55] According to this classification, viruses with DNA-based genomes are grouped into classes I, II, and VII containing double-stranded DNA (dsDNA) viruses, single-stranded DNA (ssDNA) viruses, and dsDNA viruses that replicate though an ssRNA intermediate, respectively. RNA viruses are grouped into the four remaining classes, with double-stranded RNA (dsRNA) viruses, single-stranded RNA viruses of positive (+) sense (ssRNA+), ssRNA viruses of negative (−) sense (ssRNA−), and positive (+) sense ssRNA viruses that replicate through a DNA intermediate being included in classes III, IV, V, and VI, respectively. The lack of a defined cellular structure combined with the inability of viruses to maintain their homeostasis and reproduce outside the host cell due to the lack of their own metabolism and the crucial dependence of viruses on a host cell to make new products is opposed by the presence in viruses of some features typical of living organisms, such as the presence of genes, the ability to replicate by creating copies of themselves via self-assembly, and the ability to evolve by natural selection.[57] This unusual combination of properties defines the reason for the lack of a uniform agreement on whether viruses are "organisms at the edge of life", different and special organisms, or organic structures that are nonliving but can still interact with living organisms.[56]

It should be mentioned however that the statement that viruses are devoid of genes encoding proteins involved in metabolism has been recently challenged by the discovery of such genes in giant viruses.[58] In addition, the fact that some bacteria (e.g., *Rickettsia*, *Chlamidia*, and *Mycoplasma*) are obligate intracellular parasites exactly like viruses prompts a reappraisal of the criteria defining living organisms.

Since the origin of viruses is not completely clear, three major hypotheses are currently put forward to explain the origin of these interesting creatures.[59] In the virus first or coevolution hypothesis, it is suggested that early in the history of Earth, viruses and cells appeared simultaneously, with viruses having been dependent on cellular life since their emergence. In the vagrancy or cellular origin hypothesis, it is assumed that viral evolution is based on RNA or DNA pieces that "escaped" from the genes of a larger organism. Some of the potential candidates for this "escaped" genetic material are pieces of naked DNA that are physically separated from chromosomal DNA and can independently replicate (plasmids) and DNA pieces that are able to move to various positions within genes (transposons) and even replicate. Finally, the degeneracy or regressive hypothesis suggests that viruses originated from small parasitic cells that genetically regressed or degenerated by dropping all the genes that are not needed for successful support of parasitism.

It has also been hypothesized that viruses might have their roots in the nucleoprotein world that transiently existed at the transition from the primordial RNA world to the modern DNA−RNA−protein world. According to this hypothesis, RNA viruses appeared first as a result of reduction or escape from the primitive RNA-containing cells, and these RNA viruses were the starting point of the evolution of at least some of the DNA viruses.[59] Very probably, the origin of viruses occurred very early in the evolution of life,[61] and they have existed since living cells first evolved.[60] This probably explains the ability of viruses to affect cells from all three kingdoms of life, Bacteria, Archaea, and Eukarya. The antiquity of viruses and/or their fast evolution also constitutes a potential explanation for the lack of homologues of the majority of viral proteins in cellular organisms.[59]

It is difficult to overemphasize the importance of viruses, with their ability to promote horizontal gene transfer and with their potential role in inventing DNA and its replication mechanisms and, consequently, in the evolution of all life. In fact, the mentioned virus-driven horizontal gene transfer increases genetic diversity by allowing organisms to incorporate "foreign" genetic material from another, often unrelated, organism.[62] As a result, fragments of viral origin constitute 3−8% of the human genome. Also, the possible viral origin of some DNA replication proteins followed by the transfer of these proteins to cellular organisms suggests a crucial role of viruses in the creation of DNA and DNA replication mechanisms, which are responsible for the evolution of the eukaryotic nucleus and potentially for the origin of the three aforementioned domains of life.[59]

Recently, a new classification of the Earth life forms was proposed, where eukaryotic, archaeal, and bacterial organisms are included in the class of ribosome-encoding organisms, whereas viruses constitute the class of capsid-encoding organisms, which contain nucleic acids and proteins, utilize a ribosome-encoding host organism for the completion of their life cycles, and are able to self-assemble into nucleocapsids.[63]

3. VIRAL PROTEIN CLASSIFICATION AND FUNCTIONS

Viral genomes differ greatly in size, with some viruses encoding up to approximately 1000 proteins (e.g., APMV) or as few as 6−8 proteins as in the case of human papilloma virus (HPV). There are four major functional classes of viral proteins, structural, nonstructural, regulatory, and accessory proteins.

3.1. Structural Proteins

The viral capsid is a shell consisting of several protein subunits known as protomers or capsomers. The capsid represents a protective coat around the viral genome. Often, the DNA- or RNA-based genome is tightly associated with the viral capsid of coat proteins to form a nucleoprotein complex. All viral nucleoproteins can interact with both nucleic acids and other proteins, thereby possessing substantial multifunctionality.

The way that capsomers are packed defines the shape of the capsid, which can be helical, icosahedral, or complex. Capsids of filamentous, rod-shaped, or helical viruses are highly ordered helical structures that in general consist of a single type of capsomer packed around a central axis. The central cavity in the capsid of these viruses contains their genetic material, ssRNA or ssDNA, which is electrostatically bound to the positively charged capsid proteins. These viruses can be long and very flexible or short and highly rigid. The length of the viral genome defines the length of the helical capsid of these helical viruses, whereas the size and arrangement of capsomers define the diameter of the capsid. Among well-known illustrative examples of filamentous viruses are filamentous bacteriophage *fd*, *Acidianus filamentous* virus 1 (AFV1), Tobacco mosaic virus (TMV), and *Sulfolobus islandicus* filamentous virus (SIFV).

In icosahedral viruses, the capsids are icosahedral or nearly spherical with icosahedral symmetry. Although the theoretical minimal number of identical subunits required to form such a structure is 60, the number of capsomers in the majority of icosahedral viruses is over 60. Capsids are often made of more

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX−XXX

than one capsid protein. For example, the HPV capsid is made of L1 and L2 (major and minor capsid proteins, respectively). Since in capsids of icosahedral viruses containing more than 60 identical subunits the same protein can be found in sites with different symmetries, the problem of how to fit identical subunits into different environments represents an intriguing puzzle that has been the topic of considerable debate.[64]

Some viruses have complex capsids that are neither entirely icosahedral nor entirely helical and contain some extra structures, such as complex outer walls or protein tails. One of the best-studied complex viruses is T4 bacteriophage. The characteristic feature of this virus is an icosahedral head topped on a helical tail. This tail structure ends with a hexagonal base plate with extended and protruding proteinaceous tail fibers. Due to the structure of this tail, T4 is able to attach itself to the bacterial host and to inject the viral genome into the cell using the tail as a molecular syringe.[65]

Furthermore, the capsids of some viruses can acquire lipid membrane from the host. These membrane-coated capsids, known as *viral envelopes*, might also contain viral glycoproteins, such as gp160 of human immunodeficiency virus (HIV) (consisting of the transmembrane subunit gp41 and the structural subunit gp120), the proton-selective ion channel, and the M2 protein of the influenza virus, or neuraminidase and hemagglutinin in other enveloped viruses. The functional roles of these surface viral glycoproteins are rather diverse. Some of these proteins, which typically protrude from the virus lipid bilayer (e.g., gp120, neuraminidase, and hemagglutinin) play important roles in early stages of viral infection, being associated with attachment and penetration of the virus into the target cells.[66] There are many other functions of viral envelope proteins related to the life cycle of enveloped viruses. One illustrative example of such functions is given by the M2 proton channel of influenza A virus, which has important roles in the early and late replication cycle of the influenza A virus. This proton-selective, low-pH gated ion channel is an integral homotetrameric membrane protein in the viral envelope that promotes entry of endosomal hydrogen ions into the viral particle. This leads to a decrease in the pH at the interior of the particle, with this lowering being needed for dissociating M1 from the ribonucleoprotein, initiating virus uncoating, and exposing the viral content to the cytoplasm of the host cell.[67]

The viral envelope of enveloped viruses is linked to the virus core via *matrix* proteins, which play a role after the virus has entered a cell, being responsible for expelling the genetic material. Among other biological functions of matrix proteins are their various regulatory roles performed via interactions with the components of the host cell. For example, the matrix M1 protein of the influenza virus is involved in the export of the viral ribonucleoproteins from the host cell nucleus, but also plays a role in the assembly and budding of this virus and controls inhibition of viral transcription.[68,69]

3.2. Nonstructural Proteins

Besides structural proteins that form the capsid, viruses encode viral nonstructural (NS) proteins, e.g., those proteins that are not included in the viral particle. Such nonstructural proteins act inside the infected cell and play various roles during virus replication and virus assembly. We provide below a few illustrative examples of the broad range of functions exerted by nonstructural proteins.

HPV ORFs are classified as early (E) and late on the basis of their location in the genome. The early ORFs code for nonstructural proteins. Both E1 and E2 take part in viral replication, as well as in the regulation of early transcription. While E1 binds to the origin of replication and exhibits ATPase as well as helicase activity,[70,71] E2 forms a complex with E1, thereby facilitating its binding to the origin of viral replication.[71−73] E2 also acts as a transcription factor that positively and negatively regulates early gene expression by binding to specific E2 recognition sites within the upstream regulatory region (URR).[74,75] E4 is the most highly expressed protein and plays a number of important roles in promoting the differentiation-dependent productive phase of the viral life cycle.[76−78] The E5 protein has weak transforming capabilities in vitro,[79,80] supports HPV late functions,[81,82] and disrupts MHC class II maturation.[83] The E6 and E7 proteins are primarily responsible for progression of HPV-mediated malignant cells, which ultimately leads to an invasive carcinoma. They play at least a partial role of oncoproteins in the high-risk HPVs by respectively targeting the cell cycle regulators p53 and Rb.

Another interesting example illustrating the diversity of roles covered by nonstructural proteins is provided by the hepatitis C virus (HCV), where interactions of nonstructural proteins with lipid raft membranes, with each other, and with hVAP-33, a human cellular vesicle membrane transport protein, lead to the formation of the HCV RNA replication complex or HCV replicon.[84]

As already mentioned above, nonstructural proteins can also play a role in immunomodulation. The West Nile virus (WNV) nonstructural protein NS1 was shown to participate in immunomodulation, as judged from experiments that showed that soluble and cell-surface-associated NS1 was able to both bind to and recruit the complement regulatory protein factor H. As a result of this interaction, there is a decreased complement activation, which minimizes the immune system's targeting of WNV by decreasing the complement recognition of infected cells.[85] Likewise, although through a different mechanism, the action of type 1 and type 2 interferons (and hence the induction of the innate immune response) is specifically blocked by the rinderpest virus nonstructural C protein.[86] A role in counteracting the antiviral response has also been shown for many nonstructural paramyxoviral V proteins.[87]

Finally, nonstructural proteins can also be involved in gene transactivation. For example, the genes that code for the nonstructural proteins NS-1 ad NS-2 of the autonomous parvovirus minute virus of mice (MVM), whose genome includes two overlapping transcription units, are transcribed from the P04 promoter, whereas the transcription of capsid protein genes is controlled by the P39 promoter, which is activated by the nonstructural protein NS-1.[88]

3.3. Accessory and Regulatory Proteins

Various accessory and regulatory proteins serve multiple functions and have indirect roles in several viral functions, ranging from transcription rate regulation of viral genes encoding structural proteins to modification of host cell functions. For example, the production of two regulatory (Tat and Rev) and several accessory (Vpr, Vif, Vpu, and Nef) proteins actively controls the replication of human immunodeficiency virus-1 (HIV-1), regulates various aspects of the life cycle of a virus, and controls various host cell functions, including apoptosis and gene regulation.[89] In fact, efficient in vivo infection depends on a number of accessory proteins, with Vif being needed for overcoming the host's defense

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX−XXX

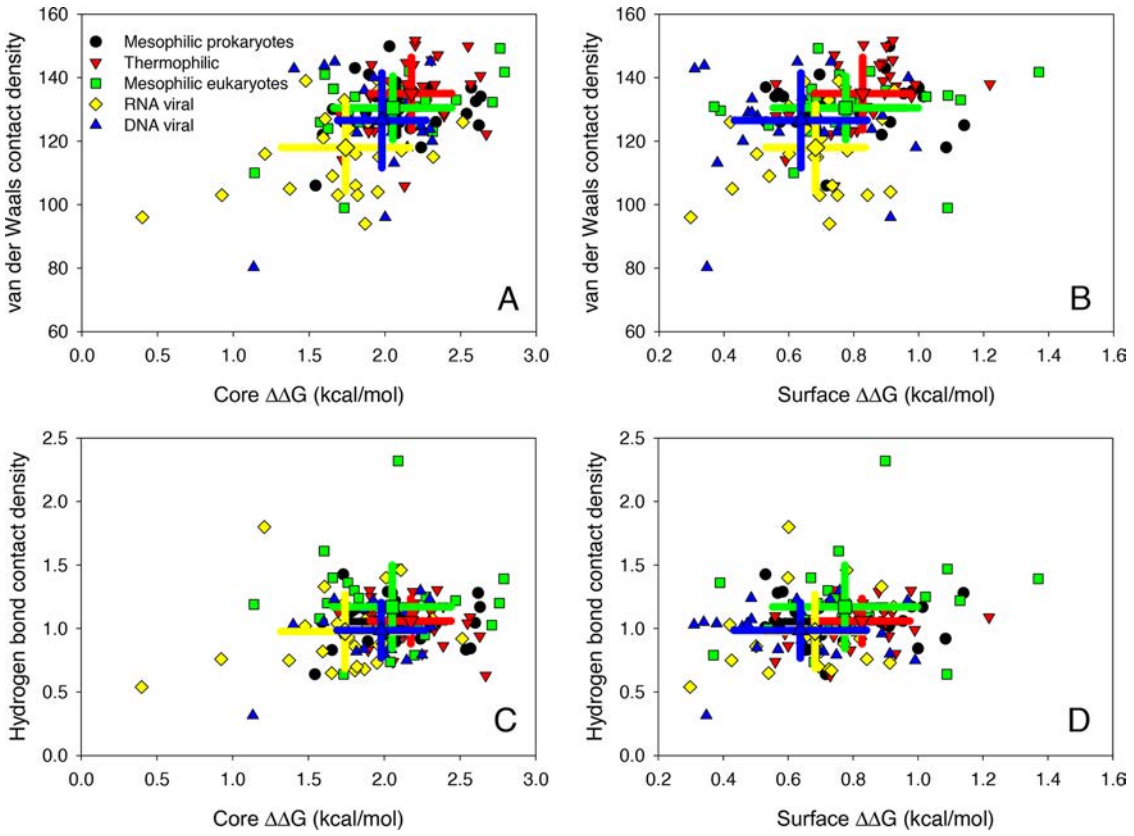


Figure 1. Comparison of the contact densities and conformational stabilities for sets of mesophilic prokaryotic proteins (black circles), mesophilic eukaryotic proteins (green squares), thermophilic proteins (red inverted triangles), RNA viral proteins (yellow tilted squares), and DNA viral proteins (blue triangles). Dependence of the van der Waals contact density on the $\Delta\Delta G$ values (conformational stability) calculated for core (A) and surface (B) residues. Dependence of the hydrogen bond contact density on the $\Delta\Delta G$ values (conformational stability) calculated for core (C) and surface (D) residues. In each plot, the corresponding mean values with standard deviations are shown as large symbols with bidirectional error bars colored according to the corresponding data sets. Data were taken from ref 40.

mechanisms and with Nef increasing the pathogenesis of the virus by targeting bystander cells.[89]

4. PREVALENCE OF INTRINSIC DISORDER IN VIRAL PROTEINS AS UNVEILED FROM BIOINFORMATICS ANALYSES

Although many RNA and DNA viruses contain similar viral proteins, such as the major capsid proteins of icosahedral viruses or special proteins participating in both the replication and morphogenesis of the virus, the majority of viral proteins do not have homologues in present-day cells.[61] Such a general lack of homology between viral and cellular proteins emphasizes the very ancient origin of viruses and suggests that viral genes came from cellular lineages that are now extinct or mainly originated in the virosphere during replication of viral genomes.[90] As discussed before, one of the many noteworthy features of viruses is their ability to adapt to very harsh and hostile environments and to adjust themselves according to the biological and genetic features of the hosts, which in turn are often adapted to exist at extreme conditions. For example, all viruses have to avoid the host's countermeasures while replicating their genes inside the host organisms.[36] Also, many extremophilic archaea isolated from geothermally heated environments are infected with viruses.[91]In addition, the genomes of numerous viruses possess several peculiar characteristics, including high rates of mutation, the structural consequences of which are very difficult to predict since one single mutation might have an effect on more than one viral protein (due to the existence of overlapping reading frames in many viral genomes).[35,36]

On the basis of the findings presented here (as well as many others), viral proteins have been suggested to have distinct and unusual structural features.[40] To check this hypothesis, a detailed comparative examination of viral and nonviral proteins was carried out. To this end, 123 representative single-domain, small (70−250 amino acids) proteins whose crystal structures have been determined at high resolution and that were shown to contain no cofactors were analyzed.[40] Among these proteins, there were 26 hypothermophilic bacterial/archaeal proteins, 26 mesophilic eukaryotic proteins, 26 mesophilic prokaryotic proteins, 26 proteins from RNA viruses, and 19 proteins from DNA viruses (18 from double-stranded DNA viruses and one from a single-stranded DNA virus). For these comparative studies, proteins with similar sizes and folds were selected whenever possible. On the basis of this analysis, it has been concluded that (a) van der Waals contact densities of viral proteins, and especially proteins from RNA viruses, are significantly lower than those of proteins from other groups, (b) a larger fraction of residues of viral proteins are not organized into regular secondary structural elements, and (c) conformational stabilities of viral proteins, evaluated as corresponding $\Delta\Delta G$ per residue values, are less affected by mutations than the stabilities of other proteins.[40] More specifically, the average $\Delta\Delta G$ per residues values measured

dx.doi.org/10.1021/cr4005692 | Chem. Rev. XXXX, XXX, XXX−XXX

for viral proteins were 0.20 and 0.26 kcal/mol lower than the average $\Delta\Delta G$ values calculated for the mesophilic and thermophilic proteins of the same size.[40] These observations are illustrated in Figure 1, which represents the dependencies of the van der Waals contact densities (Figures 1A,B) or hydrogen bond contact densities (Figures 1C,D) on the $\Delta\Delta G$ values calculated for core and surface residues.

Successively, the authors analyzed the open reading frames in the proteomes of 19 hyperthermophilic archaea, 35 mesophilic bacteria, 20 eukaryotes, 30 single-stranded RNA viruses, 30 single-stranded DNA viruses, and 29 double-stranded DNA viruses, and the amino acid compositions and disorder propensities were compared.[40] Since this work was dedicated to the comparative analysis of proteins of different origins (prokaryotic, eukaryotic, and viral) and since structural capsid or coat proteins of viruses constitute a new class that lacks parallels in Eukaryota or Prokaryotae, the viral proteomes were first filtered to get rid of all annotated capsid/coat/envelope/structural proteins.

Figure 2A gives the relative composition profiles that were calculated for different species using the methods described by Vacic and colleagues.[92] The fractional difference in composition of a given protein set and the set of completely ordered proteins was computed for each amino acid residue. This difference was calculated as $(C_X - C_{order})/C_{order}$, where C_X is the content of a given amino acid in a given protein set and C_{order} is the corresponding content in the fully ordered data set.[93,94] This figure shows the compositional profiles computed for the filtered data set of viral proteins from 89 proteomes and for the nonfiltered data set that contains all viral proteins taken from approximately 2400 viral species.

On the basis of this study, viral proteins were found in general to display a reduced fraction of hydrophobic, acidic, and lysine residues, while also having a significantly increased proportion of polar resides. Figure 2B shows that viral proteomes possess a very high propensity for ID. The amount of disorder in viruses was, in general, comparable to that in eukaryotes, which are already known to possess very high amounts of disorder from previous studies.[4,24,30,95,96] Figure 2B shows that there is a clear and fundamental difference between eukaryotic and viral proteomes, as eukaryotes contain more proteins that have long disordered regions, while viral proteomes are characterized primarily by the dominance of disordered segments that are much shorter.[40] Of note, this difference does not reflect a systematic difference in size between eukaryotic and viral proteins.

These observations suggested that viral proteins clearly differ from the proteins of their hosts, being less densely packed and less affected by mutations, being enriched in short disordered regions, possessing a higher content of polar residues and residues that are not involved in regular secondary structure elements, and being characterized by a significantly weaker network of interactions among residues.[40] As such, viral proteins (and in particular those from RNA viruses) were found to be enriched in disordered regions. It was therefore concluded that viral proteins and proteins of their hosts are shaped by very different adaptive forces. The above-mentioned specific structural features of viral proteins indicate that they were shaped by evolution to be endowed with better adaptation to their hostile habitats and to rapid changes in their biological and physical environment, rather than to ensure a higher thermodynamic stability.[40]

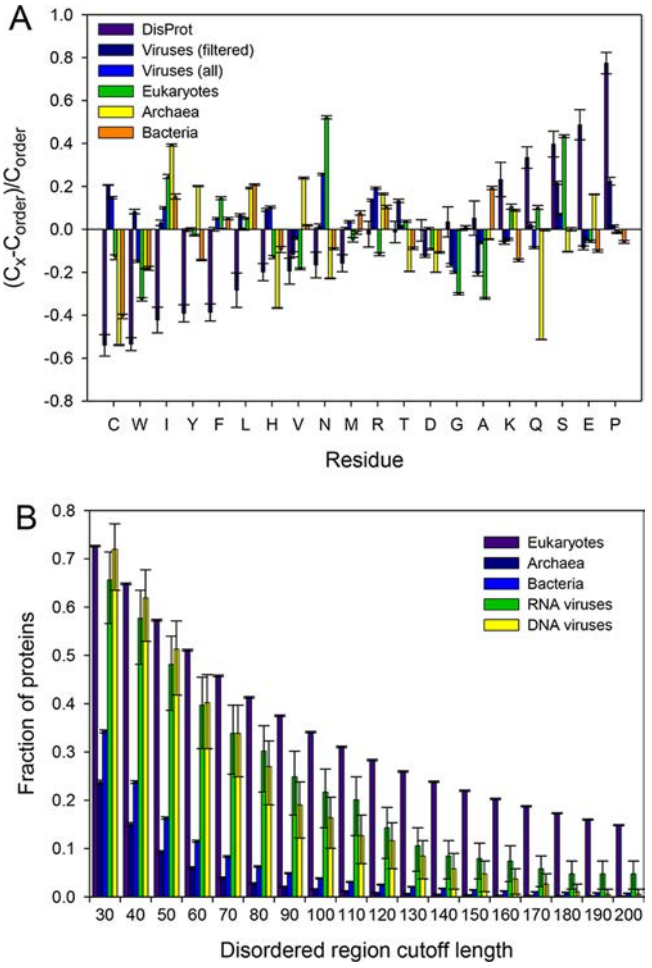


Figure 2. Evaluation of the abundance of ID in viral proteins. (A) Composition profile of amino acids for proteins from different organisms. The *x*-axis shows residues arranged by increasing disorder tendencies. The *y*-axis displays the relative compositional profile compared to an entirely disordered data set. (B) Fractions of disordered regions in viral, eukayotic, and prokaryotic proteomes. The proportions of predicted continuous disordered segments are shown as a function of the length of the disordered region. Data were taken from ref 399.

The authors therefore proposed that, in addition to allowing for a broad partnership, the frequent occurrence of disordered regions in viral proteins might also be connected to the typically high mutation rates of RNA viruses, which represents a strategy for buffering the damaging effects of mutations (i.e., a protein that already has no structure has less to lose from a mutation than one with a high degree of structure, as it is already unfolded).

In a subsequent study, an attempt was made to find possible correlations between ID and function in viral proteins on the basis of the examination of the distribution of ID in the Pfam database.[42] This database holds a large amount of data on protein functional domains and families and therefore provides a key tool to understand protein function and structure. Members of the Pfam database families are identified via evolutionary conservation of protein domains via hidden Markov models (HMMs) as well as multiple sequence alignments.[97−99] Typically, each of the curated families in Pfam is represented by a seed that includes representative members of the family and a full alignment that includes all

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX−XXX

members of the family which were detected via HMMs.[97] The Pfam database (version 23.0) contained 6360 viral Pfam domain seeds[42] when this analysis was carried out. Figure 3

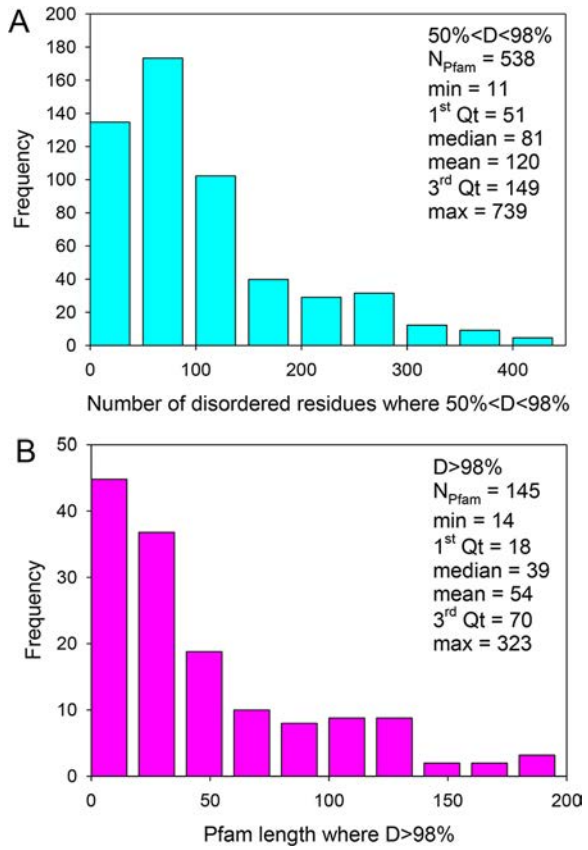


Figure 3. ID frequency in viral Pfam domain seeds. (A) Length distribution of viral Pfam domain seeds with the percentage of disordered residues between 50% and 98%. (B) Length distribution of Pfam domains in which disorder is observed for >98% residues. Data were taken from ref 399.

shows that the Pfam seed domains of viral origin are considerably disordered (e.g., 538 viral Pfam domain seeds are 50−98% disordered) and the length of their disordered regions ranges from 11 to 739 residues (Figure 3A). Furthermore, many domains (>100) were almost completely disordered (Figure 3B). Several completely disordered (i.e., containing >98% disordered residues) viral Pfam domain seed are listed in Table 1, along with their corresponding functions.[42] The biological functions that are ascribed to these disordered viral Pfam domains are mostly related to protein−protein interactions, recognition, regulation, and signal transduction,[42] suggesting that the major biological functions of viral disordered domains are similar to functions typically ascribed to disordered proteins of prokaryotic, archaeal, and eukaryotic origin.[2−6]

In an even more recent study, the authors analyzed the abundance of ID in various organisms and discovered that viruses have the largest variation range in the content of disordered residues in their proteomes.[43] This is illustrated in Figure 4, which shows the correlation between ID content and proteome size for 3484 species from viruses, archaea, bacteria, and eukaryotes. As shown in Figure 4, in human coronavirus NL63 only as few as 7.3% of residues are predicted to be disordered, whereas this percentage reaches a value as high as 77.3% in the case of the avian carcinoma virus proteome. Some viral species are highly enriched in ID. More than 20 small viruses with 5 or less proteins have 50% or higher disordered residues in their proteomes.[43] These small viruses have the highest fraction of ID among all species. When the proteome size increases, the fractions of disordered residues in the proteomes of various viruses seem to converge to a range between 20% and 40%. This high content of predicted ID in viruses is consistent with a different study which showed that, when compared to those of archaea and bacteria, bacteriophagic and viral proteins are enriched in polar residues and depleted in hydrophobic residues.[100] In that study it was also pointed out that eukaryotes are similar to viruses with respect to the amino acid compositions of their proteins; a clear decrease in order-promoting residues was indeed found in bacteriophages, eukaryotes, and viruses as compared to prokaryotes.[100] Since polar residues are able to participate in specific recognition by providing strong stabilizing hydrogen bonds with partner molecules, a higher amount of polar residues in viral proteins could reflect the need for disorder in the unbound state and specific stabilization and recognition in the bound state.[100]

Since viruses are obligate intracellular parasites, they have an exquisitely close relationship with their host, and their genomes are consequently shaped directly by interactions with the host proteome. In fact, these interactions shape every step of the viral life cycle, from entry to budding. Consequently, in the course of evolution, viruses have "learned" to hijack and manipulate host proteins for their benefit. A recent study by Davey and co-workers showed that viruses have achieved this ability through broad mimicry of host protein short linear motifs (SLiMs),[101] where the latter are embedded in disordered regions and play a variety of roles, including targeting host proteins for proteosomal degradation, cell signaling, directing proteins to the correct subcellular localization, deregulating cell cycle checkpoints, and altering transcription of host proteins.[102] In that study, examples of convergent evolution for more than 50 such eukaryotic linear motifs (ELMs) annotated in the ELM database have been experimentally validated in viral proteins.[103] Very recently, 2278 available viral genomes in 41 families were analyzed with respect to their predicted disorder, and the amount of disorder was correlated with the size of the genome, as well as other factors.[104] The amount of protein disorder found between viral families varies strikingly (from 2.9% to 23.1% of residues), as well as within families. This noticeable variation did not follow however the trend established among their hosts, with increasing disorder across eubacteria, archaebacteria, protists, and multicellular eukaryotes. For example, within large mammalian viruses, poxviruses and herpesviruses considerably differed in disorder (5.6% and 17.9%, respectively). Viral families with smaller genome sizes have more disorder within each of the five main viral types (ssDNA, dsDNA, ssRNA+, dsRNA, retroviruses), except for negative single-stranded RNA viruses, in which case disorder increases with the size of the genome. When considering viruses as a whole, however, no notable association was found between protein disorder and genome size. From that study it was concluded that there is a broad disparity in the disorder content of viral proteomes, suggesting the presence of virus-specific and family-specific effects. Varying disorder contents likely have an impact on the modulation of host factors by viruses, as well as the rapidity with which viruses are able to

dx.doi.org/10.1021/cr4005692 | Chem. Rev. XXXX, XXX, XXX−XXX

Table 1. Highly Disordered (>98%) Viral Pfam Domains

ID	type	function
PF08793.2	domain	This virus-specific 2-cysteine adaptor domain was found fused to OTU/A20-like peptidases and S/T protein kinases. The domain associations between these proteins suggest that they may function as viral adaptors which connect the kinases and OTU/A20 peptidases to specific targets.
PF08792.2	domain	This zinc ribbon domain was found to be associated with some viral A2L transcription factors.
PF04623.4	family	This family constitutes the N-termini of E1B 55 kDa, a protein that binds the tumor suppressor p53 protein, converting it from a transcriptional activator that responds to damaged DNA to an unregulated repressor of genes with a p53-binding site. As a result, the virus is protected against p53-induced host antiviral responses; this process also prevents apoptosis induced by the adenovirus E1A protein.
PF00608.9	repeat	This adenovirus fiber protein interacts with host cell receptors, resulting in specific attachment of adenovirus. This interaction is mediated by the globular carboxy-terminal domain of the adenovirus fiber protein.
PF06515.3	family	This family consists of Borna disease virus P10 (or X) proteins. Transcription and replication of Borna disease virus (BDV) takes place in the nucleus of the infected cell. The p10 protein has been suggested to play a role in viral RNA synthesis or ribonucleoprotein transport.
PF02337.9	family	This family consists of various retroviral Gag (core) polyproteins and encompasses the p10 region that produces the p10 protein after proteolytic cleavage of Gag v by the retroviral protease. The p10 or matrix protein (MA) is associated with the virus envelope glycoproteins in most mammalian retroviruses and can be involved in virus budding, particle assembly, and transport.
PF08705.3, pdb_2c55	domain	The p6 protein from HIV contains two late-budding domains (L domains). The latter are short sequence motifs that are crucial for viral particle release. P6 interacts with the endosomal sorting complex and constitutes a docking site for several binding factors. The PTAP motif establishes interactions with the cellular budding factor TSG101. This domain is also found in some proteins from the chimpanzee immunodeficiency virus (SIV-cpz).
PF08720.2, pdb_1flc	domain	This is the stalk segment-forming domain of influenza C virus hemagglutinin.
PF04534.4	family	In herpes simplex virus type 2, UL56 is probably a tail-anchored type II membrane protein involved in vesicular trafficking. The C-terminal region is hydrophobic and necessary for association with the cytoplasmic membrane, while the N-terminal proline-rich region is required for the translocation of UL56 to the Golgi apparatus and to cytoplasmic vesicles.
PF04726.5	family	This small protein is implicated in DNA packaging; e.g., it interacts with DNA by means of its hydrophobic carboxyl terminus. In bacteriophage ϕ-X174, 60 copies of J are found. J forms an S-shaped polypeptide chain devoid of any regular secondary structure.
PF02344.7	family	This family consists of the leucine zipper dimerization domain found in both cellular c-Myc proto-oncogenes and viral v-Myc oncogenes. Dimerization via the leucine zipper motif with other basic helix−loop−helix−leucine zipper (b/HLH/lz) proteins such as Max Swiss:P25912 is required for efficient DNA binding. The Myc−Max dimer is a transactivating complex that activates the expression of growth-related genes, which promotes cell proliferation.
PF07020.3	family	This family consists of several orthopoxviral C10L proteins. These proteins play an important role in vaccinia virus evasion of the immune system of the host. This can take place through the blockade of IL-1 receptors by C10L, a homologue of IL-1 Ra.
PF00519.9	family	This protein forms a complex with the E2 protein Pfam:PF00508. It is a DNA helicase required for the initiation of viral DNA replication.

dx.doi.org/10.1021/cr4005692 | Chem. Rev. XXXX, XXX, XXX−XXX

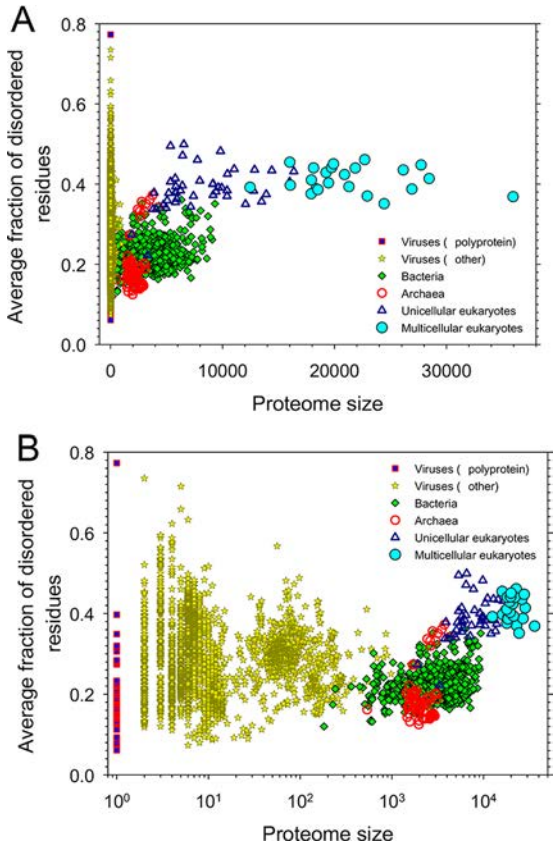


Figure 4. Correlation between ID content and proteome size for 3484 species from viruses, archaea, bacteria, and eukaryotes. Each symbol indicates a species out of six groups: bacteria (small green circles), viruses expressing one polyprotein precursor (small red circles filled with blue), archaea (blue circles), multicellular eukaryotes (pink triangles), unicellular eukaryotes (brown squares), and other viruses (small red circles). Each viral polyprotein was analyzed as a single polypeptide chain, with no parsing done to split it into individual proteins before prediction. The proteome size corresponds to the number of proteins in the proteome of the species, which is displayed in linear scale (A) or in log base (B). The average proportion of disordered residues is calculated as the average fraction of disordered residues of each sequence over all the sequences of that species. Disorder predictions were done using PONDR-VSL2B. Data were taken from ref 43.

create novel instances of SLiMs subverting host functions, such as acquired and innate immunity.[104]

5. INTRINSIC DISORDER IN STRUCTURAL PROTEINS

5.1. Intrinsic Disorder in Capsid Proteins: Evidence from X-ray Crystallography

Capsids are specific cages for genome transfer assembled from multiple copies of a single or a few proteins. In any given virus, the coding space for the capsid is minimized via utilization of a limited number of different capsid proteins. Also, in this way, a simple and self-controlled mechanism of shell assembly is provided in which the only pieces used are those that can fit together. The icosahedral symmetry is frequently used by many isomeric viruses since this symmetry defines a low-energy solution for shell formation.[64] Although the capsids of small icosahedral viruses are formed by 60 identical units, viruses with large, nearly spherical capsids with icosahedral symmetry are constructed from many building blocks. These capsids are made of flat hexamers and pentamers with a convex shape and are characterized by a stoichiometry that typically involves multiples of 60, with the multiplier value being given by the triangulation number T.[64] For example, if $T = 3$, then there are $3 \times 60 = 180$ subunits in the corresponding icosahedral virus. These considerations constitute a basis for the quasi-equivalence theory where bonding relations between the mentioned pentamers and hexamers and their environments in the icosahedrons are not identical and are non-symmetry-related.[64]

Early X-ray crystallography analyses of viral structures have already revealed that the structure of the coat proteins is characterized by the presence of two structurally different regions: a globular C-terminal domain involved in the formation of two antiparallel, four-stranded β-sheets with a jellyroll (e.g., Swiss roll) topology[105,106] and a highly extended N-terminal domain partly not defined in the electron density.[107] Also, although the capsid of the polyoma virus and simian virus 40 (SV40) consists of an icosahedral surface lattice with triangulation number $T = 7d$, it was shown to contain 72 capsomers assembled from 360 units,[108,109] which is considerably smaller than the $7 \times 60 = 420$ units that are expected from the Caspar–Klug rules.[64] Although a mixture of hexamers and pentamers was expected, all 72 capsomers in the capsids of these viruses are pentamers. Therefore, the capsid positions that were predicted to have a hexamer of subunits on the basis of the Caspar–Klug hypothesis[64] are occupied by pentamers. These pentamers comprise the structural protein VP1 or the coat protein in polyoma virus and SV40, respectively. The resolution of the apparent contradiction between the observed capsid geometry and the geometry predicted on the basis of the Caspar–Klug hypothesis is based on the utilization of the intrinsically disordered arms formed by the N-terminal domains of the capsid proteins, with these arms being extended and found to exist in at least six entirely different conformations, depending on their positions in the lattice.[107,110]

The capsid of foot-and-mouth-disease virus (FMDV) is built in agreement with the icosahedral symmetry and has 60 identical subunits. Each of these subunits consists of four proteins, namely, VP1, VP2, VP3, and VP4.[111] Three of these proteins, VP1, VP2, and VP3, are eight-stranded β-sandwiches with wedgelike shapes. Interaction of the FMDV capsid with the host is mediated by the loops that connect the strands at the narrow end of the wedge. These loops are relatively mobile, being less constrained by structural interactions. The capsid interior contains VP4 and the N-termini of VP1 and VP3. Although they are involved in the capsomer formation, three of the four viral proteins are noticeably disordered. In fact, residues 138–154 and 209–212 of VP1, 1–11 of VP2, and 1–14 and 40–61 of VP4 are not visible in the structure of the capsomer.[111] Furthermore, although VP4 has a low content in the regular secondary structure (Figure 5), it plays crucial roles in virus disassembly and assembly.

To further underscore the use of ID in capsid proteins, the Arg-rich motif (ARM) located within the R domain of the coat protein of sobemoviruses was shown to be disordered in all the sobemovirus structures that have been determined thus far.[112] Subsequent mutational and structural studies showed that the flexible nature of this motif is essential to promote correct particle assembly and RNA encapsidation (see ref 113 and references therein).

dx.doi.org/10.1021/cr4005692 | Chem. Rev. XXXX, XXX, XXX–XXX

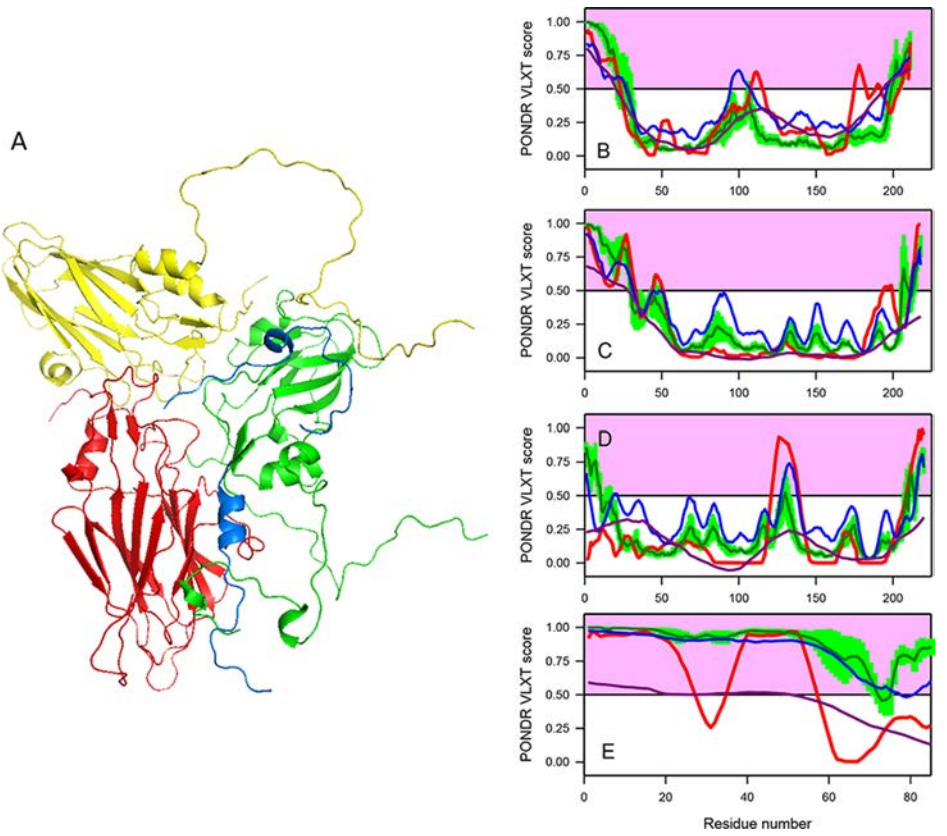


Figure 5. ID in viral structural proteins. (A) Structure of the capsomer in the icosahedral capsid of the foot-and-mouth-disease virus (PDB code 1FMD). This capsid contains 60 capsomers, each of which is composed by 4 capsid proteins: VP1, VP2, VP3, and VP4. Predicted ID in capsid proteins of the foot-and-mouth-disease virus: VP1 (B), VP2 (C), VP3 (D), and VP4 (E). For all proteins, the propensity for disorder was evaluated by a family of PONDR predictors, PONDR VLXT (red lines), PONDR VSL2 (blue lines), PONDR VL3 (dark pink lines), and PONDR-FIT (dark green lines). The light green shadow around the PONDR-FIT curves represents the distribution of errors in evaluating disorder by this meta-predictor. Disorder scores above 0.5 correspond to residues/regions predicted to be disordered. The light pink shaded areas in each plot correspond to the scores associated with intrinsic disorder. Data were taken from ref 399.

Therefore, systematic structural analysis of capsid proteins suggests that many of these proteins comprise a globular part and one (or more) extended arms. The arms may be ordered in the capsid via interactions with other viral components, but at the same time it is evident that they are flexible in the isolated protein. Disordered regions of capsid proteins, commonly known as arms, do not simply play crucial structural roles, but are highly involved in a wide range of biological functions. This idea is reinforced by the excellent reviews by Liljas,[107,114] who analyzed and systemized the functional repertoire of various viral capsid proteins and revealed that viral disordered arms can interact with nucleic acids, are involved in the control of the capsid assembly and disassembly, and, obviously, play a crucial role in the stabilization of the assembled capsid structure.

An illustration of the important regulatory mechanisms in terms of virus assembly provided by the disordered tails of the viral capsid proteins are given by the picornavirus capsid proteins. Here, the assembly process is likely to be initiated by a multidomain monomer, a newly synthesized polyprotein, that contains the four proteins of the capsid of picornaviruses (including FMDV). At a subsequent stage, this polyprotein is cleaved proteolytically to yield VP1, VP2, VP3, and VP4, the generation of which is required for interaction with RNA and completion of assembly. The role of intrinsic disorder in these processes stems from the well-established fact that disordered proteins (or protein regions) are proteolytically digested much faster than ordered proteins or protein domains.[115,116] Therefore, the coupling between protein cleavage and assembly clearly relies on the lack of structure in regions containing cleavage sites. One should keep in mind that many noncapsid viral proteins are also synthesized in the form of polyproteins that are proteolytically cleaved at specific disordered regions to yield independent functional chains (see below for additional details).

5.2. Order-to-Disorder Transitions in *fd* Phage Coat Protein pVIII

The *fd* bacteriophage belongs to the *Invorus* genus. It is a filamentous phage infecting enterobacteria (e.g., *Escherichia coli*). The circular ssDNA genome of this bacteriophage is encapsulated in a filamentous capsid consisting of α-helical subunits arranged to form a helicoidal bundle.[117−119] The major component of this filamentous capsid (both by mass and by protein content, accounting for more than 85% and 96%, respectively) is the major, α-helical coat protein (pVIII).[119] The *fd* pVIII is a small (50 amino acids long),[120] mostly helical protein[121−128] with a hydrophobic core and oppositely charged ends.[129] In particular, the C-terminus of the protein is rich in basic residues which interact with DNA.[130] These pVIII subunits are assembled into a helical sheath, where each subunit is oriented at a small angle to the virion axis and interlocked with the neighboring subunits.[130] During phage penetration into the host cells, the pVIII protein is inserted into the inner membrane of the host. Much of the pVIII from the

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX−XXX

infecting phage is not discarded and instead is reused in progeny.[131] On the basis of the analysis of the relative positioning of the virion with respect to the membrane during infection or viral assembly, it has been concluded that the orientation of the membrane-spanning coat protein, and therefore the positioning of the virion with respect to the membrane, is the same during both infection and assembly.[130] These observations suggest that the same steps, although in reverse order, are involved in penetration and subsequent assembly of the *fd* phage.[119,132]

Large morphological changes are associated with the penetration of *fd* into the host cell. In particular, in the first stage, the initial form, which is highly extended, is converted to the shortened (e.g., contracted) I-form, which is about 1/3 as long as the original phage. This rodlike I-form undergoes a subsequent transition to a membranelike spherical S-form, which, at the final stage, is converted to a form where the pVIII protein is embedded into the host membrane.[133] Subsequent studies revealed that the exposure of the *fd* phage to chloroform at different temperatures in vitro induces global morphological transitions from the original filamentous form to the contracted I-form (at lower temperatures) or to the spheroidal S-form (at higher temperatures), indicating that the conformational changes accompanying *fd* phage penetration into the host cell can be structurally characterized in this model system.[133−136]

On the basis of a comprehensive biophysical analysis, it has been concluded that the filamentous form and I-form are characterized by the same secondary structure of the pVIII protein. Similarly, the pVIII protein in the S-form and membrane-bound form was shown to possess similar secondary structure.[133] Since the *fd* phage pVIII protein in both I- and S-forms was found to be compact and to possess nativelike secondary structure, but not to have a rigid side-chain packing, it has been concluded that this protein in the I- and S-forms is endowed with many of the properties that typify the molten globule state.[137] Furthermore, since these two forms of a viral capsid were characterized not only by the non-native conformations of the pVIII protein, but also by the non-native morphologies of the capsids, it has been proposed that the transition from the ordered to molten-globule-like conformation takes place in the ensemble of the pVIII molecules within the entire capsid.[137] Therefore, both order-to-disorder and disorder-to-order transitions in the pVIII coat protein are functionally important and regulate the molecular mechanisms of penetration and assembly of the *fd* phage. This suggests that molten-globule-like intermediates are involved in macromolecular assembly and disassembly.[137]

Among the multiple cellular factors that can affect the structure of an ordered protein, leading to its (at least partial) denaturation, is the membrane surface. In fact, functional membrane-induced transitions from ordered to molten-globule-like conformations have been reported for various toxins and for some transport proteins.[138] In vitro, some ordered proteins were shown to undergo protein denaturation in the vicinity of a membrane. This process was originally attributed to the denaturing effects of the existing negative electrostatic potential of the membrane surface,[139] which attracts protons from the solution, leading to a decrease in the pH of the membrane surface and to the existence of a noticeable pH gradient in the membrane's nearest surroundings. However, since in salt-free solutions this pH decrease does not typically exceed 2 pH units,[138,139] most of the globular proteins cannot be pH-denatured by such local "acidification" of the media. This indicates that the local pH decrease induced by the electrostatic potential of the membrane surface is not the only denaturing factor of the membrane surface. Simple considerations based on classical electrodynamics suggested that a local decrease in the dielectric constant near the membrane surface was proposed as a new membrane-based denaturing factor.[138] This idea is based on the known fact that the effective dielectric constant of water at a water−hydrophobic medium interface is significantly lower than that of bulk water. Therefore, to mimic the joint action of locally decreased pH and lowered dielectric constant near the membrane surface, novel model systems consisting of water−organic solvent or water−alcohol mixtures at moderately low pH values were proposed.[138,140,141] In agreement with this hypothesis, cytochrome *c* and α-fetoprotein were shown to undergo a methanol-induced transition to a molten globule state under conditions of moderately low pH values.[140,142] Furthermore, a comprehensive biophysical analysis of the effect of a wide spectrum of organic solvents on the structure of β-lactoglobulin revealed that this well-characterized ordered protein clearly undergoes a dielectric-constant-induced transition to a molten-globule-like intermediate.[141] In agreement with these observations, the aforementioned order-to-disorder transitions in pVIII were found to be induced by organic solvents, i.e., conditions mimicking the membrane field effects. All this suggests that membrane-induced order−disorder transitions are important for conferring functions to various proteins, playing a role in the translocation of some proteins through the membrane, in hydrophobic ligand release from some transport proteins, and in the assembly and disassembly of phage and virus capsids.

5.3. Semliki Forest Virus Capsid Protease: An Illustrative Example of an Intrinsically Disordered Enzyme

The Togaviridae family of viruses comprises two genera: *Alphavirus* and *Rubivirus*, which are enveloped positive strand RNA viruses with an icosahedral nucleocapsid and a spherical morphology. The *Alphavirus* genus contains more than 40 recognized members, including Semliki forest virus (SFV), which infects birds, rodents, and humans via mosquito bites, causing rash and arthritis.[143] In the inner capsid of SFV, 240 copies of the capsid protein are arranged in a $T = 4$ icosahedral lattice encapsulating the single molecule of genome RNA.[144] The N-terminal segment (residues 1−267) of the SFV polyprotein is an intramolecular serine protease that cleaves itself off after the invariant Trp267 from a viral polyprotein and generates the mature capsid protein (SFVP). After this autoproteolytic cleavage, the free carboxylic group of Trp267 interacts with the catalytic triad (His145, Asp167, and Ser219) and inactivates the enzyme.[145] Therefore, SFVP performs a single enzymatic reaction before it assembles in the viral capsid shell, where the SFVP and its N-terminal protease domain in their mature form are inactive. SFVP consists of an unstructured basic segment at the N-terminal part (residues 1−118) and a serine protease segment that forms a two-β-barrel domain at the C-terminal part (residues 119−267).[146] Autoinhibition can be efficiently abrogated by the deletion of the autoinhibitory, C-terminal Trp267 residue.[145,147] Thus, although the autoinhibited C-terminal part of the protease domain of SFVP is well folded, it is enzymatically inactive. Strikingly, C-terminally truncated variants of the serine protease segment all possess noticeable enzymatic (esterase) activity, even though they are intrinsically disordered.[145,147] The conclusion on the intrinsically disordered nature of the

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX−XXX

truncated forms stems from the results of the spectroscopic characterization of the C-terminal deletion variants of SFVP by nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopies. These variants are in fact characterized by broad 1D ^{1}H NMR spectra with low dispersion, which is indicative of a natively unfolded protein. Furthermore, the near-UV CD spectra of the C-terminally truncated SFVP variants under native conditions are similar to the spectrum of the unfolded protein under denaturing conditions, and the far-UV CD spectra showed a minimum around 200 nm observed in unfolded polypeptides.[145,147]

5.4. Intrinsic Disorder in Flaviviridae Core Proteins

Flaviviridae members are RNA viruses with a nonsegmented single-stranded RNA genome of positive polarity whose length is between 9.6 and 12.3 kb. All three genera encompassed by the Flaviviridae family, namely, *Hepacivirus*, *Flavivirus*, and *Pestivirus*, comprise severe human pathogens. In Flaviviridae, the core protein (or capsid protein) is released from the N-terminal region of the viral polyprotein. All core proteins are highly basic and bind RNA with broad sequence specificity. They possess RNA chaperone activities in vitro and are in charge of the packaging and condensation of the viral genomic RNA during virion morphogenesis. As such, they play pivotal roles in the viral replication cycle, and at the same time, they orchestrate a complex, dynamic interaction network with host cell proteins, contributing to viral persistence and pathogenicity. As explained below, these promiscuous interactions are made possible by the intrinsic flexibility of viral nucleocapsid proteins, facilitating either simultaneous or sequential binding to a plethora of structurally unrelated substrates, resulting in flexible, ever-changing multiprotein, RNA−protein, and lipid−protein complexes during the viral replicative cycle (see refs 148−151 and references therein).

The hepatitis C virus (HCV) and the closely related GB virus B (GBV-B) are the only members representing the *Hepacivirus* genus of Flaviviridae. The HCV core protein (HCV-C) is positioned at the N-terminus of the polyprotein. The latter is cleaved by host-encoded proteinases, thereby resulting in the generation of an immature core protein and a mature core protein of 191 and 179 amino acids, respectively. HCV-C is larger than the core protein of other Flaviviridae members that are approximately 100 amino acids long. Contrary to the core proteins of other Flaviviridae, HCV-C is predicted to possess only a few structural elements.[152] Mature HCV-C consists of two domains referred to as domain 1 and domain 2. The C-terminal domain is enriched in hydrophobic residues and serves as a membrane-binding module. Domain 1 encompasses residues 1−117 and contains three highly basic amino acid clusters that mediate RNA binding and promote RNA structural rearrangements. Domain 1 of HCV-C was revealed to be sufficient for assembly in nucleocapsid-like particles (NLPs) when in the presence of structured RNA. Beyond RNA binding, most of the mapped protein interaction sites also fall within domain 1, indicating that the latter is the primary organizer of the HCV infection network.

The biophysical features of the N-terminal domain 1 of HCV-C (and fragments thereof) were characterized via several methods. Those studies unveiled that this domain lacks any stable secondary or tightly folded tertiary structure.[149] Indeed, domain 1 was found to be hypersensitive to proteolytic digestion by either chymotrypsin or trypsin[153] and to show abnormal electrophoretic mobility on SDS gels,[153] features characteristic of IDPs.[154] In agreement with this and with bioinformatics studies that predict that domain 1 is mostly unstructured, far-UV CD spectra of the N-terminal 124 (C124), 117 (C117), or 82 (C82) amino acids of HCV-C suggest a random-coil-like conformation, as judged from the marked ellipticity minimum in the spectrum observed at approximately 200 nm.[148,153,155] NMR spectroscopy, as well as chemical shift indexing, provides further support for the proposed highly disordered nature of C82.[153] These findings provide a rationale for the ability of HCV-C to interact with several host proteins, such as the C-terminus of p53, the DEAD-box protein (DDX3, CAP- Rf), the intracellular domain of lymphotoxin β receptor, p21Waf1/Cip1/Sid1, and the 14−3−3 protein (see ref 153 and references therein).

The RNA chaperoning activity of HCV-C was found not to require a folded state of the protein, as judged from the observation that heat denaturation does not abrogate this activity. Strikingly, ID is a well-known hallmark of RNA chaperones.[156] According to the "entropy exchange model" proposed by Tompa and Csermely, the high flexibility of chaperone proteins would facilitate broad-specificity RNA binding and disruption of non-native bonds.[156] In particular, highly flexible protein regions can undergo a disorder-to-order transition after binding to RNA while simultaneously melting the RNA structure via an entropy exchange process. The destabilized RNA can then perform another search in the conformational space, until it reaches its most stable configuration upon cyclic protein binding and release.[156] However, experimental results gathered on the structural changes that the HCV-C core undergoes in the presence of its target RNA support a "mutual induced folding" scenario. Furthermore, Fourier transform infrared (FTIR) spectroscopy suggests that binding of the intrinsically unstructured D1 domain of the core protein to its specific target (the SLIIId subdomain of the HCV internal ribosome entry site, IRES) results in β-sheet formation (∼22%) in HCV-C, even though the majority of the protein stays unstructured.[157]

The structures of core proteins from dengue and West Nile viruses, two *Flavivirus* members, have been characterized by cryo electron microscopy, X-ray crystallography, and NMR.[158−160] These studies showed that the structure of these proteins is primarily α-helical, though the N-terminal 20 amino acids of dengue and West Nile viruses are unstructured as shown by protease digestion, X-ray crystallography, and NMR.[158,159]

Murray and co-workers purified the core protein from bovine viral diarrhea virus (BVDV), a *Pestivirus* member, and characterized it biochemically. Using fluorescence spectroscopy and far-UV CD, the protein was shown to be intrinsically disordered. This protein was shown to bind RNA (albeit with low affinity and barely discernible specificity), which is consistent with the very basic sequence it possesses and with the behavior of all Flaviviridae core proteins; it was also shown to functionally replace the RNA-binding and -condensing region of an unrelated viral capsid protein, a finding that designates BVDV core protein as a central player in virion morphogenesis and RNA packaging.[161]

Similar studies carried out on GBV-B and West Nile virus, unveiled the capacity of their core proteins to bind to RNA and to induce large structural rearrangements, with no requirement for a defined 3D structure.[148] Collectively, biochemical and biophysical studies carried out so far showed that, in spite of low sequence similarity and pronounced differences in their

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX−XXX

modular organization, core proteins from Flaviviridae members make broad use of IDRs (Figure 6) for RNA binding,

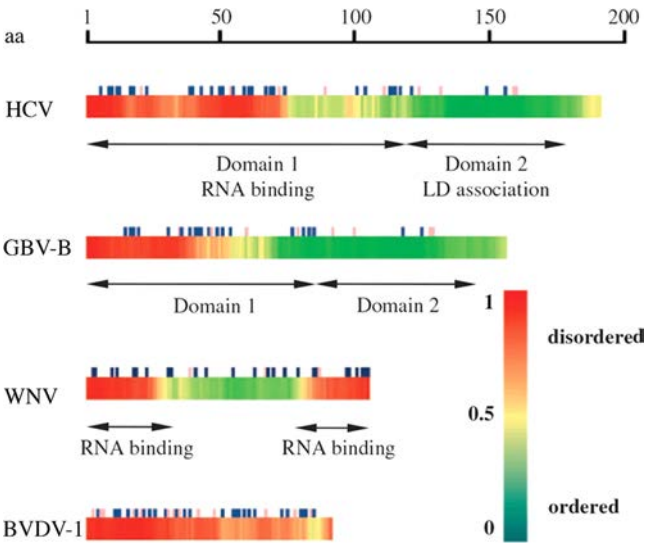


Figure 6. Disorder prediction in Flaviviridae core proteins. Disordered regions in HCV, GBV-B, WNV, and BVDV core proteins (GenBank accession numbers D89872, AF179612, AF481864, and AF220247, respectively) were predicted using DisProt VL3-H (http://www.ist.temple.edu/disprot/predictor.php). A disorder score above 0.5 indicates a disordered amino acid, while a score below 0.5 indicates order. The disorder is displayed using a color scale, with well-folded domains in green and highly flexible segments in red. Basic amino acid residues are displayed in dark blue, and acidic residues are shown in mauve. Adapted with permission from ref 148. Copyright 2008 Oxford University Press.

chaperoning, and particle assembly, with very basic yet flexible protein segments being a trademark of active RNA chaperone domains in Flaviviridae core proteins.

5.5. Intrinsic Disorder and Disorder-to-Order Transitions in the Replicative Complex of Paramyxoviridae and Rhabdoviridae Members

Paramyxoviridae and Rhabdoviridae are members of the Mononegavirales order, which comprises viruses with a nonsegmented single-stranded RNA genome of negative polarity. In Mononegavirales, the genome is tightly encapsidated by the nucleoprotein (N) within a helical nucleocapsid. The latter, rather than naked RNA, serves as a substrate for both transcription and replication. Both transcription and replication are performed by the viral RNA-dependent RNA polymerase (RdRp), which consists of a complex between the large protein (L) and the phosphoprotein (P). The P protein is an essential polymerase cofactor as it recruits the L protein onto the nucleocapsid template. Beyond its role as a polymerase cofactor, the P protein also acts as a chaperone for the N protein in that it prevents illegitimate self-assembly of the latter when ongoing genomic RNA synthesis does not occur and maintains it in a soluble form (N^0) within a complex (N^0–P) that is used for encapsidation of the nascent RNA chain during replication (for a review see ref 162).

The N and P proteins are involved in numerous protein–protein interactions, not only with viral proteins but also with host cell factors, which leads to multiple biological effects, including modulation of both innate and acquired immunity. In line with this promiscuity, a considerable amount of experimental evidence has been assembled in the past 10 years, which suggests an abundance of disorder in the N and P proteins of these viruses. Below, we provide an overview of the structural information available so far, with emphasis on the functional relevance of IDRs within these proteins.

In members of the Paramyxovirinae subfamily, the N protein consists of a structured N-terminal domain (N_{CORE}) responsible for RNA binding and self-assembly[163] and a C-terminal domain (N_{TAIL}) predicted to be intrinsically disordered.[38] The disordered nature of N_{TAIL} was proven in the measles (MeV), Sendai (SeV), Nipah (NiV), and Hendra (HeV) viruses.[39,164−167] In these viruses, N_{TAIL} was shown to undergo α-helical folding upon binding to the C-terminal X domain (XD) of the phosphoprotein,[164,165,168,169] with this transition having been mapped to a molecular recognition element (MoRE) of helical nature encompassing approximately 20 residues in length.[170−181] NMR studies that made use of residual dipolar couplings (RDCs) allowed atomistic descriptions of MeV and SeV N_{TAIL} as conformational ensembles to be obtained and showed that, rather than fraying randomly, their α-MoREs preferentially populate four (MeV) or three (SeV) specific overlapping helical conformers, each of which is stabilized by N-capping interactions. As a result, the unfolded strands that are adjacent to the helix are projected in the direction of the partner protein, providing a mechanism by which they could achieve nonspecific encounter interactions before binding.[166,178,182] Although the α-MoRE of both MeV and SeV N_{TAIL} is partly preconfigured prior to binding, experimental and computational data available in the case of MeV, suggest a mixed "folding before binding" and "folding after binding" mechanism.[176,183]

In spite of the XD-induced α-helical transition of N_{TAIL}, N_{TAIL}–XD complexes display a significant amount of residual disorder and hence provide illustrative examples of "fuzziness",[184] where this term has been recently coined by Tompa and Fuxreiter to designate the persistence of conspicuous regions of disorder, often playing a functional role in binding, within protein complexes implicating IDPs/IDRs.[185] The fuzzy nature of these complexes is illustrated both by the considerable amount of N_{TAIL} residues that are not affected by XD and by the dynamic behavior of the α-MoRE at the surface of XD.[169,171−177,181] The conservation of a significant amount of residual disorder in N_{TAIL}–XD complexes suggests a functional role: it has been proposed that the prevalently disordered nature of N_{TAIL} even after complex formation may serve as a platform for the capture of other binding partners.[184,186,187] In agreement, in the case of MeV N_{TAIL}, the first 20 residues have been shown to be accountable for the interaction with the cellular nucleoprotein receptor (NR),[188,189] and the C-terminal region has been found to interact with the major inducible heat shock protein hsp70, which modulates both viral transcription and replication.[190] Notably, the disordered nature of N_{TAIL} from the measles and Hendra viruses has also been confirmed in the context of the full-length N protein, i.e., in the context of the NLPs that the latter forms when expressed in heterologous systems.[169,178,181] Although XD has long been thought to trigger a major conformational rearrangement within the nucleocapsid, thereby providing access of the viral polymerase to the RNA genome (for reviews see refs 186, 187, and 191–194), recent NMR studies carried out on HeV nucleocapsids ruled out this possibility and provided the first direct observation of the interaction between the X domain of P and intact nucleocapsids in Paramyxoviridae.[181]

dx.doi.org/10.1021/cr4005692 | Chem. Rev. XXXX, XXX, XXX–XXX

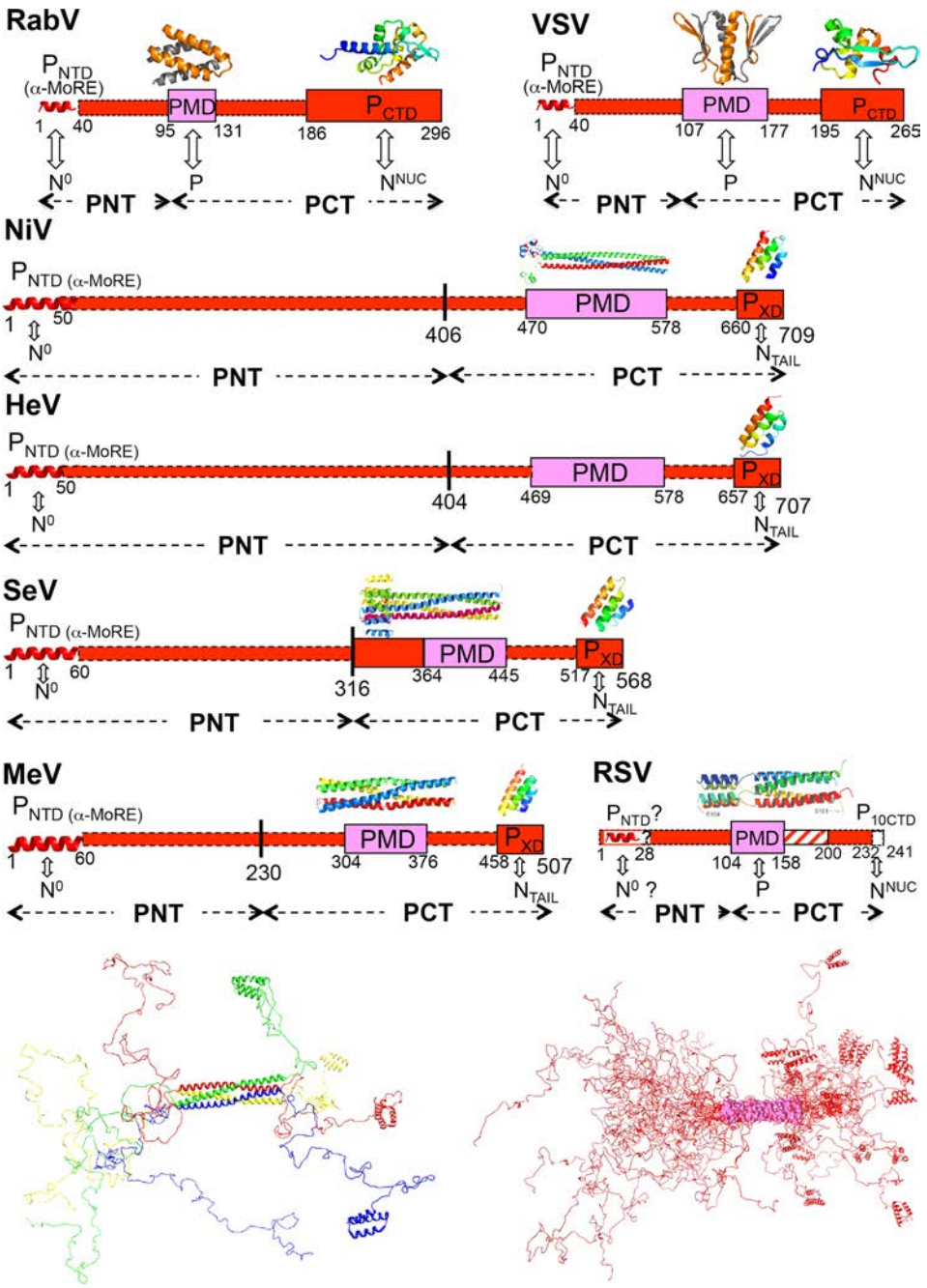


Figure 7. Upper panel: Modular organization of P proteins from rabies virus (RabV), vesicular stomatitis virus (VSV), Nipah virus (NiV), Hendra virus (HeV), Sendai virus (SeV), measles virus (MeV), and respiratory syncytial virus (RSV). Structured and disordered regions are represented as large or narrow boxes, respectively. PNT = N-terminal region of P, PCT = C-terminal region of P, P_{NTD} = P N-terminal domain containing the N°-binding site, P_{CTD} = P C-terminal domain containing the nucleocapsid (N^{NUC})-binding site, PMD = P multimerization domain, and PXD = P X domain adopting a triple α-helical bundle. The α-MoRE partly preconfigured in solution and predicted/shown to adopt a stable α-helical conformation upon binding to N° is shown. Whenever available, the relevant crystal structures are shown above each domain. PDB codes: RabV PMD, 3L32;[400] VSV PMD, 2FQM;[401] SeV PMD, 1EZJ;[402] MeV PMD, 4BHV;[403] SeV XD, 1R4G;[204] HeV XD, 4HEO;[181] MeV XD, 1OKS.[168] The structural models of NiV and RSV PMD are shown,[404,405] as is that of NiV XD.[169] Lower panel: Cartoon representation of the MeV P tetramer either as a single conformer (left) or as an ensemble of five conformers (right), illustrating the very long reach of the molecule. In the left panel each monomer is drawn with a different color. Disordered regions were generated using Flexible-Mecano,[406] and structures were drawn using Pymol.[407] Data on the modular organization of the proteins were taken from ref 162.

The presence of the disordered N_{TAIL} region partly exposed at the surface of the nucleocapsid and projecting far away from the latter is instrumental for the recruitment of the numerous binding partners that have been reported for this protein. Indeed, MeV N_{TAIL} establishes numerous interactions with various viral partners, such as P, the P−L complex, and the matrix protein.[195] Beyond the viral partners, MeV N_{TAIL} also interacts with various cellular proteins, including interferon regulatory factor 3 (IRF3),[196] hsp70,[190] peroxiredoxin 1,[198] the cell protein responsible for the nuclear export of N,[197] casein kinase II,[199] and possibly components of the cell cytoskeleton.[200,201] Additionally, within MeV nucleocapsids released

dx.doi.org/10.1021/cr4005692 | Chem. Rev. XXXX, XXX, XXX−XXX

from infected cells, N_{TAIL} can also bind to cell receptors involved in MeV-induced immunosuppression[188,189] (for reviews see refs 184, 186, 187, 191−194, 202, and 203).

In the case of the P protein, structural disorder could be documented not only in Paramyxoviridae members,[37,38,165,167,204−206] but also in Rhabdoviridae[207−209] (for reviews see refs 186, 187, 193, 194, 210, and 211). The P protein from these viruses was found to possess a highly modular organization that consists of alternating disordered and ordered regions (Figure 7). In Paramyxovirinae, the P protein was shown to possess a large (up to 400 residues) N-terminal disordered domain (PNT). Both MeV and SeV PNT domains were shown to interact with several partners, with MeV interacting with N and cellular proteins[212,213] and SeV interacting with the unassembled form of N (N°) and the L protein.[214,215]

Interestingly, while the C-terminal nucleocapsid-binding region of P in Rhabdoviridae and in the majority of Paramyxovirinae members adopts a stably folded, compact conformation, it is disordered in respiratory syncytial virus (RSV), a Pneumovirinae member.[206,216] Of note, recent experimental data unveiled that in the case of the P proteins from *Rubulavirus* members ID further extends to their X domains, which were found to span a structural continuum, ranging from stable to largely disordered in solution. They share however the ability to adopt, at least transiently, a common fold consisting of a triple α-helical bundle like in other Paramyxovirinae members.[217,218]

The N-terminal region of the P proteins of Paramyxoviridae and Rhabdoviridae members contains an α-MoRE involved in binding to N°.[38,207,208,219] Experimental evidence showing that the N°-binding region of the N-terminal region of P undergoes induced folding upon binding to a partner is limited to vesicular stomatitis virus (VSV), a rhabdovirus: the structure of the VSV N°−P complex was solved and unveiled that this region does adopt an α-helical folding upon binding to N, while the flanking regions remain flexible in the complex. Binding of the α-MoRE of P occurs at the same site responsible for binding to RNA and N, thus preventing N polymerization. Incidentally, these results also suggested a possible mechanism for the initiation of viral RNA synthesis.[220] In the case of MeV, so far, the only hints suggesting that the N-terminal disordered region of P (PNT) could undergo a disorder-to-order transition were provided by limited proteolysis studies carried out in the presence of the secondary structure stabilizer trifluoroethanol (TFE). The use of this organic solvent to locate disordered regions with likelihood to fold is determined by the ability of TFE to strengthen the peptide hydrogen bonds and to provide a favorable environment for hydrophobic side chains of amino acids, thereby increasing the propensity of the latter to gain α-helical structure. Those studies allowed mapping of a fragment resistant to proteolysis in the presence of TFE to the N-terminal region of PNT, with this fragment having been proposed to correspond to a putative α-MoRE involved in binding to N°.[37]

The presence of unstructured domains on both N and P would allow for coordinated and dynamic interactions between the polymerase complex and a large surface area of the nucleocapsid template, which could extend over successive turns of the helix. It was shown indeed that the maximal extension of MeV PNT, as measured by small-angle X-ray scattering (SAXS) (Longhi and Receveur-Bréchot, unpublished data), is 40 nm. On the other hand, one turn of the MeV nucleocapsid is 18 nm in diameter and 6 nm high.[221] Hence, PNT can easily stretch over several turns of the nucleocapsid, and since P is multimeric, N°−P could possess a considerable extension. Similarly, the maximum extension of MeV N_{TAIL} in solution is 13 nm.[39] Consequently, the potentially long reach of disordered regions may enable them to serve as linkers and to bind partners on large macromolecular assemblies, serving as scaffolding engines as already discussed for intrinsically disordered scaffold proteins.[222,223]

5.6. Intrinsic Disorder in the Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein

The nucleocapsid protein (N) of the severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV) packs up the viral genomic RNA and is crucial for the viability of the virus. However, the mechanism by which this protein binds RNA is not well understood. The N protein possesses two domains, an N-terminal domain (NTD; residues 45−181) and a C-terminal dimerization domain (CTD; residues 248−365), flanked by long stretches of disordered regions accounting for almost half of the entire sequence. SAXS data showed that the protein is in an extended conformation and that the two structural domains of SARS-CoV nucleocapsid protein are far away from each other. Both NTD and CTD have been shown to bind RNA, as were the flanking disordered regions. Constructs containing multiple RNA-binding regions showed Hill coefficients greater than 1, which was taken to reflect cooperative binding of the N protein to RNA. This effect can be explained by the "coupled-allostery" model that was devised to explain the allosteric effect in a multidomain regulatory system. While the N proteins of different coronaviruses share very low sequence homology, bioinformatics studies showed that the flexible linker regions of these proteins all start with an SR-rich region and end with a region enriched in basic residues, features that are both hallmarks of protein disorder. The overall isoelectric points (pI's) of these flexible linkers are high, which could explain their RNA-binding abilities. These findings suggest that the physicochemical features described above are likely conserved across different Coronaviridae groups. These observations underscore once the important roles of multisite nucleic acid binding and ID in N protein function and RNP packaging.[224]

5.7. Intrinsic Disorder in Viral Genome-Linked Proteins

Some viruses possess a viral genome-linked protein (VPg) attached to the 5′ end of their RNA genome over a phosphodiester bond formed between the hydroxyl group of a Tyr/Ser/Thr residue and the 5′-phosphate group of RNA.[225−227] VPg's are highly diverse in sequence and in size (2−4 kDa for Picornaviridae and Comoviridae members, 10−26 kDa for Potyviridae, Sobemoviruses, and Caliciviridae members, and up to 90 kDa for Birnaviridae members).[228] Similar to the vast majority of viral proteins, VPg's have multiple functions that were shown to play key roles in the major steps of the viral cycle, such as replication, translation, and cell-to-cell movement.[228] Since these numerous functions are able to be performed by mature VPg's, as well as by their precursors, it is believed that the processing of VPg precursors represents one of the regulatory mechanisms of the VPg multifunctionality.[229] The multifunctionality of VPg's depends on the ability of these proteins to be involved in a multitude of interactions with different viral or host partners such as VPg itself, nuclear inclusion protein b, helper component protease, cylindrical inclusion protein, cylindrical inclusion helicase, coat protein or eukaryotic translation initiation factors eIF4A, eIF4E,

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX−XXX

eIF3, and eIF4G, and the poly(A)-binding protein.[229−238] Binding promiscuity and related polyfunctionality of VPg's at least in part are defined by the intrinsically disordered nature of these proteins. In fact, on the basis of the structural characterization of individual proteins, the intrinsically disordered status has been reported for VPg's from potato virus A (PVA), potato virus Y (PVY), lettuce mosaic virus (LMV), Sesbania mosaic virus (SeMV), and rice yellow mottle virus (RYMV).[229,239−242] In accordance with these experimental studies, it was shown via computational analyses that all the VPg's representative of the viral diversity (including six sobemoviruses, six potyviruses, and four members of the Caliciviridae family) possess functionally important intrinsically disordered regions.[242] Strikingly, in SeMV another IDP was identified, namely, P8, an 8 kDa highly basic protein that is released as a result of a proteolytic cleavage in cis from the C-terminus of polyprotein 2a.[113] In SeMV, both VPg and P8 were shown to regulate the enzymatic activity of two virus-encoded proteins. In fact, when fused to the viral protease, but not when added in trans, VPg was found to render active both in cis and in trans the otherwise inactive viral protease responsible for cleavage of the viral polyproteins.[239,243] This activation relies on a critical interaction between Trp43 of VPg and residues Trp271 and His275 in the protease domain.[243] Likewise, P8 was found to stimulate the ATPase activity of P10, a finding that suggests that both VPg and P8 positively regulate the activities of domains present at their N-terminus.[113] Interestingly, VPg was shown to exert in vitro an inhibitory effect on the polymerase activity of the RdRp that is present at its C-terminus.[113] In SeMV, VPg was also shown to interact with the movement protein (MP), a protein responsible for effective cell-to-cell spread of the virus, a discovery that suggests that VPg might play a role in specific recognition and transport of viral RNA from cell to cell.[244]

5.8. Intrinsic Disorder in Nucleocapsid and Matrix Proteins of HIV-Related Viruses

Bioinformatics studies were carried out on a few viral matrix proteins.[245,246] Those studies showed that the matrix protein p17 from simian immunodeficiency virus (SIV_{mac}) and HIV-1 possesses high levels of predicted ID, while matrix proteins of the equine infectious anemia virus (EIAV) contain noticeably lower levels of predicted disorder.[246]

The HIV-1 matrix protein p17 (also known as MA protein) is a 132 amino acid long polypeptide that lines the inner surface of the virion membrane, holds the RNA-containing viral core (defined as the structure that remains after the lipid bilayer is stripped away) in place, and is myristoylated at its N-terminus.[247,248] p17 participates in the virion assembly, is directly associated with the inner leaflet of the viral membrane, and forms a protective shell.[249] The cotranslational myristylation of the N-terminus of the p17 protein provides a targeting signal for Gag polyprotein transport to the plasma membrane.[247,248] Furthermore, p17 possesses another feature that is involved in membrane targeting, namely, its set of basic residues located within the first 50 amino acids.[250] These features define the importance of p17 for targeting Gag and Gag-Pol precursor polyproteins to the plasma membrane prior to viral assembly.[250] In addition to targeting Gag polyproteins to the plasma membrane by means of its multipartite membrane-binding signal, p17 performs a number of important functions in the viral replication cycle and may be involved (possibly via specific nuclear localization sequences) in nuclear import.[251] Computational analysis of p17 from ∼50 HIV-1 isolates revealed that these proteins are expected to behave as native coils or native pre molten globules.[252] p17 forms trimers, and trimerization is driven by residues 42−77.[253] In the NMR structure, the center of the p17 molecule is an antiparallel coiled coil formed by helices B and C, whereas helices A and D lie parallel to each other on either side of a coiled coil. All the helices are accessible to solvent and are highly amphipathic, except for helix C, which is located at the center of the hydrophobic core. Two regions (fragments 19−23 and 26−29) together with the region between helices C and D form three strands of the β-sheet. Finally, the last C-terminal 20 residues do not adopt any rigid conformation in solution, and there is an ill-defined potential turn in the middle of the N-terminal 14 residues.[254]

HIV-1 nucleocapsid protein (NC or p7) coats the genomic RNA inside the virion core. This 55 residue long protein contains two zinc finger domains (of the CCHC type) flanked by basic amino acids that are required for interaction with nucleic acids.[255,256] The major function of NC is to bind specifically to the packaging signal of the full-length viral RNAs and to transport them into the assembling virion.[250] As a highly charged basic protein, NC binds single-stranded nucleic acids nonspecifically. As a result, it coats the genomic RNA and hence protects it from nucleases and compacting viral RNA within the core. It is possible that, in addition to its other roles, NC also serves as an RNA chaperone that improves several nucleic acid-dependent steps of viral life, such as melting of RNA secondary structures, annealing of the tRNA primer, stimulating integration,[260] and promoting DNA strand exchange reactions during reverse transcription.[257−259]

In the NMR solution structure of NC, regions corresponding to the two zinc fingers (residues 15−28 and 36−49) possessed well-resolved structures, whereas residues 1−13, 32−34, and 52−55 were highly dynamic and did not converge to the unique conformations.[261,262] In agreement with these structural studies, computational ID propensity analysis revealed that p7 is a highly disordered protein, with regions corresponding to the zinc fingers predicted to be more ordered than the remainder of the protein and identified as potential α-MoRFs.[252]

Recently, it has been emphasized that the flexible nature of p7 (NC) is crucial for the interactions of this protein with nucleic acids via the invariant zinc fingers and flanking basic residues. This flexibility also provides the basis for a possible mechanism that defines the multiple functions of p7. In fact, this protein has a pivotal role in the early steps of virus replication, where p7 is known to be involved in the obligatory strand transfer reactions during viral DNA synthesis by the reverse transcriptase enzyme and in other stages of reverse transcription from the initiation to the completion of viral DNA synthesis, acting as a chaperoning partner of the genomic RNA template and of reverse transcriptase.[263]

5.9. Intrinsic Disorder in Surface Glycoproteins from Influenza Virus

Surface glycoproteins are used by enveloped viruses, such as influenza, measles, HIV-1, and Ebola viruses, to enter target cells through fusion of the viral membrane with the target membrane.[264−266] One of the best-studied membrane fusion proteins is the influenza virus hemagglutinin (HA). HA is a homotrimeric type I transmembrane surface glycoprotein that is responsible for binding of the virus to the host receptor, virus

dx.doi.org/10.1021/cr4005692 | Chem. Rev. XXXX, XXX, XXX−XXX

internalization, and subsequent membrane-fusion events within the endosomal compartment of the infected cell. HA is also the most abundant antigen on the viral surface and contains the primary neutralizing epitopes for antibodies. Each 70 kDa HA subunit includes two disulfide-linked polypeptide chains, HA_1 and HA_2, that are created via proteolytic cleavage of the precursor protein HA_0.[267] This cleavage is essential for membrane fusion since it exposes the fusion peptide.[267] During membrane fusion, HA binds to sialic acid receptors on the host cell surface, and following endocytosis, the acidic pH of the endosomal compartments (pH 5−6) induces irreversible and dramatic reorganization of the HA structure.[268]

The HA trimer has a tightly entangled "stem" domain at its membrane-proximal base, which consists of HA_1 residues 11−51 and 276−329 and HA_2 residues 1−176. The main feature of this stalk region in the HA trimer is the three long, parallel α-helices (~50 amino acids in length each), one from each monomer, which interact to form a triple-stranded coiled coil. The membrane-distal domain consists of a globular "head", formed by HA_1 and which can be further subdivided into the R region (residues 108−261), which contains the receptor-binding site and major epitopes for neutralizing antibodies, and the E region (residues 56−108 and 262−274), with high structural homology with the esterase domain of influenza C esterase fusion protein (HEF).[269] The HA_2 chain contains two membrane-interacting hydrophobic peptide sequences: a C-terminal transmembrane segment that passes through the viral membrane and an N-terminal "fusion peptide" (residues 1−23) that interacts with the target membrane bilayer.[270]

Several crystallographic studies have suggested that the interaction with the host cell involves a drastic structural reorganization of HA_2, which makes the fusion peptide move approximately 100 Å from the interior to the target membrane.[271,272] In this process, the middle of the original long α-helix unfolds to form a reverse turn, jack-knifing the C-terminal half of the long α-helix backward toward the N-terminal extremity. As a result of these molecular rearrangements, the N-terminal fusion peptide and the C-terminal transmembrane anchor are placed at the same end of the rod-shaped HA_2 molecule,[273,274] thereby facilitating membrane fusion by bringing the viral and cellular membranes together.

A recent bioinformatics analysis unveiled that if many viral membrane glycoproteins are globally ordered, ID is still present in these proteins, suggesting that it plays a crucial role in their biological functions. For example, this study pointed out that the highly virulent strains of influenza A virus (1918 H1N1 and H5N1) differ from the nonvirulent or less virulent strains (H3N2 and 1930 H1N1) in the disorder propensities of their HA proteins. The most pronounced differences pertain to disorder propensities of the region near residues 68−79 of HA_2, a region that is located at the tip of the stalk participating in the interaction with the receptor chain, HA_1.[275]

6. INTRINSIC DISORDER IN NONSTRUCTURAL, REGULATORY, AND ACCESSORY PROTEINS

Since nonstructural proteins are involved in a wide range of functions related to regulation and recognition, such as regulation of virus replication and assembly and communication with the host, they are often disordered. To illustrate this, we present a brief overview of ID in various nonstructural proteins from different viruses, as well as in the accessory and regulatory proteins from HIV-1.

6.1. Disorder in Nonstructural HCV Proteins

HCV NS5A is a 49 kDa well-studied protein that plays a key role in the replication of the virus, while also being involved in particle assembly.[276] Numerous protein−protein interactions have been reported for NS5A, including viral or host cell proteins (reviewed in ref 277). NS5A is a membrane-associated protein that possesses an anchor on its N-terminal. Its cytoplasmic portion, which is divided into three domains, encompasses disordered regions. Domain 1 (D1) of NS5A is highly conserved, and its structure has been solved, which revealed a structural scaffold with a novel zinc-binding motif and a disulfide bond.[278,279] Conversely, as detailed below, domains 2 and 3 (D2 and D3) are less conserved and possess highly disordered regions.[280,281] Both NS5A-D2 and NS5A-D3 are known to establish a complex molecular partnership (see refs 282 and 283). The absence of an ordered conformation and the dynamic behavior of both NS5A-D2 and NS5-D3 serve as an underlying molecular basis that enables interactions with multiple partners and confers to NS5A a hublike character.

Domain 2 of HCV NS5A (NS5A-D2) is important for NS5A function and is implicated in molecular interactions with the RdRp (NS5B) and PKR, a cellular interferon-inducible serine/threonine-specific protein kinase. Thus, the interactions established by NS5A-D2 interfere with host regulation processes such as signaling pathways and apoptosis.[284] Liang and co-workers carried out a structural analysis of NS5-D2 using NMR spectroscopy. The analysis of the backbone ^{1}H, ^{13}C, and ^{15}N resonances, $^{3}J_{HN\alpha}$ coupling constants, and 3D NOE data indicates that NS5A-D2 does not have stable secondary structural elements and reveals several characteristics typical of unfolded proteins. The lack of a rigid structure in the domain was confirmed by NMR relaxation parameters.[285]

Likewise, sequence analysis indicates that NS5A-D3 is mostly unstructured, although short structural elements may exist at its N-terminus. In agreement, gel filtration chromatography and CD and NMR spectroscopy all pointed out the disordered nature of purified recombinant NS5A-D3.[280] However, in a more recent study by the same group, two NS5A-D3s from two HCV strains were found to exhibit propensity to partially fold into an α-helix.[286] NMR analysis identifies two putative α-helices for which a molecular model could be obtained. The amphipathic character of the first helix and its conservation in all genotypes suggest that it might correspond to a MoRE and as such promote the interaction with relevant biological partner(s). One such partner is cyclophilin A (CypA).[286] Cyclophilins are host cell factors that are vital for HCV replication. NMR heteronuclear exchange experiments demonstrate that CypA has in vitro peptidyl−prolyl cis/trans isomerase (PPIase) activity toward some, but not all, of the peptidyl−prolyl bonds in NS5A-D3.[286] Interestingly, the interaction between HCV NS5A-D3 and CypA is completely abrogated by cyclosporin A (CsA), a discovery that designates inhibitors of CypA, such as CsA or nonimmunosuppressive analogues, as candidates for development of antiviral strategies.[286]

As already mentioned, CypA is critical for HCV replication. In agreement, together with NS5A and NS5B, it is part of the membrane-associated multiprotein complex that supports RNA transcription and replication. In a recent study, the same authors used NMR spectroscopy to characterize at a residue level the molecular interactions between NS5A-D2 and NS5A-D3, CypA, and a truncated form of NS5B (NS5BΔ21). Those

studies unveiled that while NS5A-D2 interacts with NS5BΔ21, NS5A-D3 does not. In addition, both NS5BΔ21 and CypA were found to share a common binding site on NS5A. No direct molecular interaction was detected between HCV NS5BΔ21 and host CypA. Addition of CsA to a sample containing NS5BΔ21, NS5A-D2, and CypA specifically inhibits the interaction between CypA and NS5A-D2 without altering the one between NS5A-D2 and NS5BΔ21. Finally, a high-quality heteronuclear NMR spectrum of HCV NS5BΔ21 was obtained, which allowed characterization of the NS5-D2-binding site on the polymerase.[287]

The HCV Core+1/S polypeptide, also called alternative reading frame protein (ARFP)/S, provides another example of a regulatory HCV protein that is disordered. Core+1/S is an ARFP that is expressed from the Core coding region of the HCV genome. This ORF drives the expression of various ARFPs, also referred to as Core+1 proteins, resulting from mechanisms such as ribosomal frame shifting and internal initiation at alternative AUG or non-AUG codons. Although it was shown that Core+1 proteins are not required for HCV replication, they are expressed during HCV infection and were shown to interfere with both apoptosis and cell cycle regulation, which suggests that these proteins may play a role in HCV pathogenesis. Core+1/S, a highly basic polypeptide, corresponds to the C-terminal fragment of the Core+1 ORF and is the shortest ARFP form described so far. Its translation results from internal initiation at alternative AUG codons (85−87) located downstream from the polyprotein codon initiator. Various studies, including disorder predictions, size exclusion chromatography, fluorescence, dynamic light scattering (DLS), CD, and NMR, showed that Core+1/S lacks any significant secondary structure in vitro, which could be relevant for recognizing diverse molecular partners, as well as for the assembly of Core+1/S. Indeed, that study also showed that Core+1/S has a certain propensity for self-association.[288]

6.2. Disorder in Nonstructural HPV E6 and E7 Proteins

Papillomaviruses (PVs) are a large family of small DNA viruses infecting mammals, along with reptiles and birds. Over 100 different types of human PVs (HPVs) exist, and they are the causative agents of benign warts and papillomas, as well as cofactors in the development of carcinomas of the head, neck, epidermis, and genital tract. On the basis of their association with cancer, HPVs are divided into two classes, namely, low-risk (e.g., HPV-6 and HPV-11) and high-risk (e.g., HPV-16, HPV-18, and HPV-45) types. As all DNA tumor viruses, HPV hijacks the replication machinery of the cell and forces infected cells to enter the S phase of the cell cycle. High-risk papillomaviruses exert their transforming activity mainly through E7, which is one of their two oncoproteins. E7 plays a role in the pathogenesis and maintenance of human cervical cancers and has been shown to participate in numerous cellular processes, including cell apoptosis, DNA synthesis,[291] gene transcription,[290] and cell growth and transformation.[289] In line with its involvement in tumor induction, E7 interacts with the retinoblastoma tumor suppressor protein (Rb), a protein that controls the G1/S transition and that is targeted by several viral oncoproteins.[292] Rb is thus a critical guardian of the cell cycle, and the balance in the interactions involving the Rb protein determines whether the cell will progress into the normal cell cycle or transformation. Beyond proteins of the Rb family, E7 also interacts with histone deacetylase,[293] kinase p33CDK2 and cyclin A,[294] the cyclin-dependent kinase inhibitor $p21^{cip1}$ protein,[295] and protein phosphatase 2A (PP2A).[296] Being involved in the formation of a complex with E7, PP2A is sequestered and excluded from its interaction with protein kinase P (PKB) or Akt.[297] The latter is one of several second messenger kinases which are activated via cell attachment and growth factor signaling and which send signals to the cell nucleus to hinder apoptosis, thus leading to increased cell survival during proliferation. Therefore, interaction between E7 and PP2A maintains PKB/Akt signaling activated through inhibition of PKB/Akt dephosphorylation. Biochemical and computational studies, summarized below, showed that this broad molecular partnership relies on the presence of extended/disordered regions within E7.

Early studies carried out on recombinant E7 from HPV-16 showed that this protein is an elongated dimer capable of undergoing a substantial conformational transition upon a small pH decrease, where the protein gains α-helicity and increased exposure of hydrophobic surfaces to the solvent.[298] The protein was found to have a high resistance to thermal denaturation even in the presence of SDS, with the persistence of residual structure in the monomer being responsible for its anomalous electrophoretic behavior. The dimer was found to become more globular in the presence of 0.3 M guanidinium chloride, with hydrophobic surfaces becoming accessible to the solvent, as judged from the large increase in ANS binding.[298] Although E7 bears properties reminiscent of those of intrinsically disordered proteins, its far-UV CD spectrum, exposure of ANS-binding sites, and cooperative unfolding were interpreted as reflecting an extended and folded, rather than disordered, conformation. The authors proposed that the large increase in the area of solvent-accessible hydrophobic surface following a small pH decrease within a physiological range could indicate conformational properties that could have evolved to enable protein−protein recognition of the numerous cellular partners with which E7 is known to interact.[298]

In subsequent far-UV CD and NMR studies, E7 from HPV-45 was shown to contain an unfolded N-terminal region (E7N, amino acids 1−40) and a well-structured C-terminal domain (amino acids 41−98) with a unique zinc-binding fold.[299,300] The N-terminal domain of this protein includes the Rb-binding and casein kinase II phosphorylation sites.[300−302] The shape, pH, and temperature dependence of the CD spectrum of E7N at pH 7.5 were taken as indicative of a polyproline type II structure.[300] This structure is stabilized by phosphorylation, which results in an increase of transforming activity in the cell. In a more recent study, the authors dissected the structural elements within the intrinsically disordered E7N domain and used a series of proteins and peptide fragments spanning different regions of the HPV-16 E7N domain and the E7 protein.[303] Using far-UV CD and NMR spectroscopy, the authors unveiled that two E7N segments located within the conserved CR1 and CR2 regions present transient α-helical structure. The helix in the CR1 region spans residues 8−13 and overlaps with the E2F mimic linear motif. The helix in the highly acidic CR2 region presents a pH-dependent structural transition: Around neutral pH this helix spans residues 17−29, which include the Rb LxCxE binding motif (residues 21−29), while the acidic CKII-PEST region spanning residues 33−38 populates the polyproline type II structure. At pH 5.0, the second helix extends up to residue 38 at the expense of loss of polyproline type II structure as a result of charge neutralization of acidic residues. It is worth noting that while at both pH values the region encompassing the LxCxE motif adopts an α-

dx.doi.org/10.1021/cr4005692 | Chem. Rev. XXXX, XXX, XXX−XXX

helical structure, the isolated fragment encompassing residues 21−29 and the LxCxE motif cannot populate an α-helix even at high TFE concentrations. Collectively, the results indicate that the E7N domain populates dynamic but discrete structural ensembles by sampling α-helix−coil−polyproline type II−β-sheet structures.[303]

Although the E7 domain is not a cooperatively folded and compact unit, it is a genuine functional domain, which evolved to preserve an extended and dynamic structure in the cell. The intrinsically disordered nature of the N-terminal module of E7 is thus responsible for the structural plasticity of this oncoprotein: the extended structure of this domain allows for adaptation to a variety of protein targets and exposure of the PEST degradation sequence that regulates its turnover in the cell, a modification of which leads to accumulation of E7 with impact on the transformation process. Curiously, E7 from another high-risk HPV, namely, HPV-16, can also form spherical oligomers with amyloid-like properties that accumulate in the cytosol of cancerous cells.[304] These oligomers can serve as platforms for the establishment of interactions with cellular partners.

Recently, NMR characterization of the entire E7 from this high-risk HPV confirmed the disordered nature of E7N and unveiled that residues 26−36 have a certain propensity to adopt an elongated structured conformation.[305] They also showed that even if the C-terminal region is more structured, it is nevertheless characterized by a high degree of local motion that reflects a molten globule state. In addition, the concentration dependence of the resonance behavior indicates that the C-terminal region aggregates while E7N retains its high flexibility.[305] The authors speculated that this would provide a way to increase the effective concentration of E7N, where the C-terminal region would be used to gather different molecules together, leaving the N-terminus free to move and therefore to interact with partners.[305]

A quantitative analysis of the interaction between HPV-16 E7 and the AB domain of Rb (RbAB) in solution showed that the interaction involves multiple motifs.[301] In particular, 90% of the binding energy is afforded by the LxCxE motif, a motif located within the conserved region 2 in the N-terminal disordered half of E7 (E7N), a region that was shown to be the major determinant of E7 binding in three prototypical HPV types.[301] The C-terminal domain of E7 contains an additional binding determinant (1.0 kcal/mol), leading to a dual-contact mode.[301] The stoichiometry and subnanomolar affinity of E7 indicate that it binds RbAB as a monomer. The low-risk HPV-11 E7 protein binds more weakly than the high-risk HPV-16- and HPV-18-type counterparts, though the modularity and binding mode are preserved. Phosphorylation at a conserved casein kinase II site in the intrinsically disordered E7N domain (i.e., Ser31 and Ser32) affects the local conformation by increasing the polyproline II content and hence stabilizing an extended conformation, which allows for a tighter interaction with Rb. These findings offer the first molecular glimpses into the functional role of phosphorylation of E7 by providing direct evidence that phosphorylation increases the affinity of E7 for RbAB. In addition, since phosphorylation-induced polyproline II−coil transitions in disordered PEST regions can modulate the sensitivity of a protein to intracellular degradation,[306] the functional relevance of E7 phosphorylation can embrace intracellular stability.[301]

The E6 oncoproteins of high-risk HPV types 16 and 18 are involved in the development of cervical cancer. E6 mainly promotes tumorigenesis by hastening cellular degradation of the tumor suppressor p53 through formation of a trimeric complex consisting of p53, E6, and the cellular ubiquitination enzyme E6AP.[307,308]

Beyond its key role in the regulation of p53 degradation, E6 also displays numerous activities such as activation or repression of various transcription promoters of cellular or viral origin,[309−312] e.g., transcriptional activation of human telomerase retrotranscriptase.[313,314] The function of low-risk HPV E6 proteins is less studied. The latter lack a number of activities that correlate with the oncogenic activity of the high-risk HPV E6 proteins, being for instance unable to target p53 for degradation.[307,315] Recently, Neveu and co-workers further extended the interactome of E6 and E7 proteins from 11 distinct HPV genotypes, which were selected for their different tropisms and pathologies. Hierarchical clustering of interaction profiles was used to measure the correlation between specific virus−host interaction profiles and various pathological traits, reflecting the different carcinogenic potential of different HPVs.[316]

HPV E6 proteins are quite small (roughly 150 amino acids) and possess a common architecture consisting of two zinc-binding domains (E6N and E6C).[317] Recombinant E6 from high-risk HPV-16 and HPV-18 strains folds into soluble and thermostable oligomers of approximately 1.2 MDa.[318] The strong propensity of E6 to self-associate long precluded its structural analysis. Recently, Zanier and co-workers were able to obtain the high-resolution NMR structure of E6 from HPV-16 by introducing mutations into the N-terminal domain that abrogated the ability of the protein to dimerize.[319]

A bioinformatics analysis of proteomes of high-risk and low-risk HPVs, focusing on E6 and E7 oncoproteins, was performed with the specific purpose of understanding whether ID plays a role in the oncogenic potential of different HPV types.[320] The study revealed that high-risk HPV-16 E7 and low-risk HPV-6 E7 share only 50% of their amino acid sequence and also display a divergence in their functions, with HPV-16 E7 being highly promiscuous.[298] Although E6 has less disorder than E7, an alternative transcript encoding only the first 50 amino acids of E6, which is highly disordered and incredibly promiscuous, is found in high-risk HPVs.[321] This fragment, unique to high-risk HPVs, forms low-molecular-mass species with minimal structure that can oligomerize into different conformations and thereby interact with a wide variety of partners.[321] It can play a direct or indirect role in several cellular processes, interfering with the cellular metabolism and resulting in tumorigenesis.[321]

On the basis of the results of this analysis, the authors concluded that high-risk HPVs are endowed with an increased amount of ID in their transforming proteins E6 and E7.[320] These results underscore the importance of disorder predictions to predict virulence.

6.3. Disorder in HIV-1 Regulatory and Accessory Proteins

The HIV-1 protein Tat is one of the important factors in viral pathogenesis serving as a transactivator of viral transcription. The role of Tat is based on its ability to interact with the transactivation responsive region (TAR), e.g., a short nascent stem−bulge loop leader RNA. TAR can be found at the 5′ extremity of all viral transcripts. Interaction between Tat and TAR is driven by the protein basic region. Once bound to TAR, Tat recruits the complex made of cyclin T1 and cyclin-dependent kinase 9 (CDK9), thereby forming the positive

transcription elongation factor B complex. The C-terminal domain of RNA polymerase II is hyperphosphorylated by CDK9, and this modified form of RNA polymerase II is responsible for the enhanced elongation of viral transcripts. Besides acting as the key transactivator of viral transcription in infected cells, Tat can be secreted by infected cells and successively taken up by uninfected neighboring cells.[322] Since the amino acid sequence of Tat possesses a high net positive charge combined with a low global hydrophobicity, this protein is expected to be a typical IDP. In agreement with predictions, CD and NMR studies suggest that this protein lacks any ordered secondary structure.[323]

Rev, a basic 116-residue protein, also plays a regulatory role in HIV-1. It can bind to multiple sites in the Rev response element (RRE) of viral mRNA transcripts in the nuclei of host cells, which leads to the transport of unspliced and incompletely spliced viral mRNAs to the cytoplasm of host cells in the later phases of the HIV-1 replication cycle. Therefore, Rev is essential for the replication of the virus,[324] serving as the regulatory HIV-1 protein that belongs to the ARM family of RNA-binding proteins.[325] Spectroscopic and hydrodynamic studies have shown that monomeric Rev adopts a molten globule state.[326] Furthermore, the highly basic primary sequences of ARMs were predicted to adopt coil-like structures.[325] Recently, the conformational changes of the Rev ARM associated with RNA binding have been investigated by CD spectroscopy, molecular dynamics simulations, and multidimensional NMR. The combined spectroscopic and simulated results suggested that the Rev ARM is intrinsically disordered not only as an isolated peptide but also when it is embedded into an oligomerization-deficient Rev mutant.[325] However, a crucial coil-to-helix transition in this important protein is promoted by its interaction with the Rev response element of the viral mRNA.[325]

Among HIV-1 accessory proteins is Vpr, a 96-residue multifunctional protein that controls many host cell functions, interferes with numerous cellular biochemical pathways, and plays crucial roles in arrest of the cell cycle at the G2/M transition, nuclear transport of the preintegration complex to the nucleus, apoptosis induction, and transcription activation. The binding promiscuity of Vpr and its ability to interact with various cellular proteins are crucial for its multifunctionality. For example, interactions of Vpr with components of the nuclear pore complex and nuclear transport factors dictate the nuclear import of this protein. Binding of Vpr to cullin 4A-associated factor DCAF1 is correlated with cell cycle arrest, whereas interaction of this viral protein with mitochondrial proteins may facilitate apoptosis in a caspase-dependent manner. The Vpr-mediated induction of viral infection in nondividing cells, such as monocytes and macrophages, may play a key role in long-term AIDS disease.[327] Comprehensive biophysical analyses using DLS, CD, and ^{1}H NMR spectroscopy revealed that although Vpr is mostly unstructured at neutral pH, it can sample an α-helical structure in the presence of TFE or under acidic conditions,[328] suggesting that the Vpr structure might be dependent on some specific binding partners (e.g., proteins, membrane components, or nucleic acids) and further modulated by the peculiarities of its cellular environments in the cytosol, mitochondrion, nucleus, cell membranes, and extracellular space.[329]

Vif is another HIV-1 accessory protein and is responsible for the neutralization of the cell's defense antiviral mechanisms. The intrinsically disordered C-terminal domain (residues 141−192) of Vif mediates the numerous interactions established by this protein. The mostly disordered nature of this domain is supported by its extended size obtained by size exclusion chromatography and the predominant coil-like nature of its far-UV CD spectrum in spite of the presence of some residual helical structure. Further support for its mostly disordered conformation comes from the ^{15}N−H^{1} HSQC NMR spectrum, where the peaks (each corresponding to a distinct residue) were found to be poorly spread (e.g., to have a narrow ^{1}H dispersion). The outcome of computational analyses of the C-terminal domain of Vif is consistent with these experimental findings. Importantly, the ability of this disordered domain to gain structure upon binding to its natural ligands was revealed by CD analysis, where this domain was shown to fold following interaction with membrane micelles.[330]

The major role of Vpu, another HIV-1 accessory protein, resides in the amplification of the release of virus particles from infected cells. This is achieved through Vpu-mediated proteasomal degradation of the HIV-1 receptor CD4 that occurs in the endoplasmic reticulum of infected cells. Interaction of Vpu with the ubiquitin ligase SCF-βTrCP, which is needed for triggering degradation of CD4 by the proteasome, is dependent on Vpu phosphorylation at Ser52 and Ser56, within the DSGXXS motif. Vpu consists of a polar, mostly disordered C-terminal cytoplasmic domain and a hydrophobic N-terminal membrane-anchoring domain.[331] The conclusion on the mostly disordered nature of the cytoplasmic domain is based on NMR and CD analysis of 9 overlapping fragments 15 residues in length and of 3 longer fragments. Importantly, in aqueous solutions, this domain possesses some residual secondary structure, whereas in the presence of TFE, this domain folds into an α-helical conformation with two α-helices joined by a flexible region that contains the two Vpu phosphorylation sites.[332]

Finally, during infection, Nef, another HIV-1 accessory protein, interacts with multiple cellular partners. Intriguingly, the ability of Nef to bind to various cellular partners is controlled via ligand-induced conformational changes that redirect Nef binding among partners, suggesting the existence of a novel disorder-based allosteric paradigm.[333] Here, the identified allosteric changes are located in the intrinsically disordered regions of Nef, whereas in the standard allosteric mechanism, an effector-induced conformational change occurs in well-defined structural elements of a protein. Importantly, the binding determinants for subsequent partners found in these disordered regions of Nef were shown to be devoid of stable secondary or tertiary structure and only become functionally competent by folding and unfolding events induced by the ligand. It was proposed that the unique ability of Nef to regulate its binding affinity by several orders of magnitude is determined by the switching of binding epitopes between structured and unstructured conformations.[333]

6.4. Bacteriophage λ N Protein

In the bacteriophage λ, the N protein (λN) plays a key role in transcribing phage genes, since in the absence of this protein transcription of the phage λ genome drops to about 2%, with only the early genes being transcribed.[334] λN is a positive regulator of transcription that promotes the expression of genes located downstream from the termination signals. The major in vivo λN function is to serve as a transcriptional antitermination factor that binds to a specific RNA sequence (the box B segment) and multiple proteins in the transcription complex,

dx.doi.org/10.1021/cr4005692 | Chem. Rev. XXXX, XXX, XXX−XXX

thereby serving as the key regulator of the assembly of the antitermination complex, which allows transcription through termination sites during gene expression of the phage. λN also interacts with a bacterial host factor (NusA) and the RNA polymerase of the bacterial host.[335]

Structurally, λN is characterized by all the features typical of highly disordered proteins, starting from high net charge and low hydrophobicity[335] and ending with structural asymmetry as determined by sedimentation and gel filtration experiments,[336] the coil-like far-UV CD spectrum,[337,338] and a narrow chemical shift dispersion in the NMR spectra.[338] Later, detailed SAXS analysis revealed that the radius of gyration of native λN is 38 ± 3.5 Å and its fractal dimension 1.76 ± 0.05, exceeding the value predicted for a well-solvated polymer with excluded volume (1.7).[339] These results suggested that due to its large net charge and high content in charged residues, λN is among the more expanded members of the structurally heterogeneous class of IDPs.[339] Also, the heteronuclear magnetic resonance experiments demonstrated that λN is a highly disordered protein.[340] However, interaction with box B RNA induced a folded conformation only within the RNA-binding domain of this protein (the arginine-rich motif (ARM), residues 1−22), whereas the activating regions of the protein, i.e., the NusA-binding region (amino acids 34−47), which suppresses enhancement of termination by NusA, and a C-terminal region (amino acids 73−107) that interacts directly with RNA polymerase, preserve their disordered state in the RNA-bound form in the absence of their target proteins.[340] Therefore, the RNA-binding region represents the minimal structural binding unit of λN proteins. The analysis of the complex between the ARM-containing peptide and the cognate RNA demonstrated that the tertiary structure of RNA forms a scaffold that allows wrapping around the peptide.[341]

6.5. Disorder in the *Hordeivirus* Movement TGBp1 Protein

The spread of plant viruses, which begins in initially infected cells and continues to the rest of the plant, relies on the systemic movement of the virus between organs (long-distance movement) that occurs in vascular tissue, typically phloem sieve elements.[342] This virus transport in plants is facilitated by special multifunctional proteins referred to as movement proteins (MPs) possessing a wide range of activities, such as interaction with viral RNA and some other viral proteins to form a ribonucleoprotein complex for the cell-to-cell and long-distance movement of the virus genome in plants and interaction with the endoplasmic reticulum and cellular cytoskeleton components.[342] In hordeiviruses, viral MPs are represented by three proteins, TGBp1, TGBp2, and TGBp3, encoded by a "triple gene block" (TGB).[343]

Functionally, *Hordeivirus* TGBp1 contains a C-terminally located, conserved NTPase/helicase domain (HELD) with seven conserved motifs of superfamily 1 NTPases/helicases[344] that possesses NTPase and RNA helicase activities and cooperatively binds RNA in vitro.[345−349] The N-terminal region of TGBp1 from the Poe semilatent virus (PSLV), which is a typical representative of the genus *Hordeivirus*, has long intrinsically disordered regions and includes two domains: a slightly structured internal domain with a molten-globule-like structure and a completely disordered extreme N-terminal domain (NTD).[342,350] Importantly, the completely disordered N-terminal domain of the *Hordeivirus* TGBp1 movement protein was recently shown to mediate localization of TGBp1 to the nucleolus and interaction with fibrillarin.[351]

6.6. Influenza Virus Nonstructural Protein 2

Influenza virus nonstructural protein 2 (NS2 or NEP) has been shown to interact with the nuclear export machinery during viral replication. It also serves as an adapter molecule between the nuclear export machinery and the viral ribonucleoprotein complex. Various approaches, such as differential scanning calorimetry, spectroscopy, hydrodynamic techniques, and limited proteolysis have shown that this monomeric protein is highly flexible and compact, thus possessing characteristics typical of the molten globule state under nearly physiological conditions.[352]

6.7. Intrinsic Disorder in the Hepatitis δ Virus Basic Protein δAg

The hepatitis δ virus (HDV) has the smallest RNA genome of any animal virus known so far and encodes only one protein, known as δ antigen (δAg).[353] δAg is a small, basic protein of 195 residues that has no known enzymatic activity and yet is vital for viral replication.[354] This protein contains a coiled-coil domain, a nuclear localization signal (NLS), and an RNA-binding domain.[355] It was shown to self-oligomerize to yield structures as large as dodecamers and to be associated with HDV genomic RNA.[356,357] Both computational and experimental (i.e., CD spectroscopy) approaches have shown that δAg's from eight clades of HDV are highly disordered, with no more than 30% of their residues adopting an α-helical structure.[357] The absence of viral proteins with enzymatic activity suggests that replication of HDV relies entirely on its unique viral protein and on the cell components of the host. In agreement, δAg was shown to be highly promiscuous, being able to bind not only multiple host partners,[358,359] but also nucleic acids.[357]

6.8. Intrinsic Disorder in Early Transcription Unit 1B of Human Adenovirus Type 5

Human adenovirus type 5 (HAdV5) encodes a set of proteins, comprising early transcription unit 1B (E1B), that carry out functions important for viral replication and adenovirus-mediated cell transformation.[360,361] An intriguing feature of this set of proteins (which includes E1B-84R, E1B-19K, E1B-55K, E1B-93R, and E1B-156R) is that they are expressed from overlapping reading frames of the 2.28 kb E1B-mRNA through alternative splicing between a common splice donor (SD1) and one of three possible splice acceptor sites (SA1−3). The resulting mRNAs encode proteins having a common N-terminus (79 residues) and different C-termini,[362,363] a feature determining one of the names of these proteins, E1BN proteins.[361] Comprehensive CD and NMR spectroscopic studies and in silico analyses revealed that E1B-93R is a typical intrinsically disordered protein and that the common N-terminal region within E1B-55K and the other E1BN proteins is likely to be intrinsically disordered.[361]

7. ALTERNATIVE SPLICING, ID, AND OVERLAPPING READING FRAMES

In the course of evolution, viruses have adapted an intricate genetic organization that allows for the optimized use of their limited genomes and enables them to produce all necessary regulatory and structural proteins. The use of alternative splicing is important for regulated expression of multiple viral regulators from one genomic polycistronic RNA. Additionally, as already mentioned, viruses display a broad variety in their genome structure and in their transcription and replication

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX−XXX

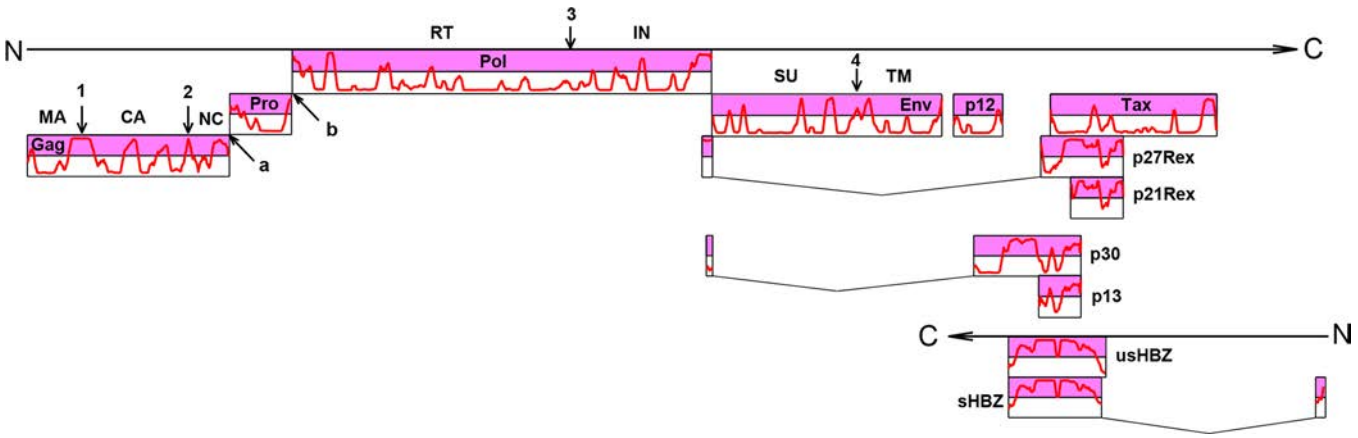


Figure 8. Proteome map of HTLV-1. Each of the products are displayed with a bar, the location of which depends on the location of the corresponding gene within the HTLV-1 genome. Disorder predictions for each (poly)protein, as obtained using PONDR VLXT, are displayed as solid lines inside the corresponding bars. A disorder score of above 0.5, indicated in pink, designates a residue is predicted to be disordered. Hence, in each bar, pieces of the PONDR plots located in these shaded areas correspond to protein fragments predicted to be disordered. Cleavage sites yielding Gag, Pro, and Pol polyproteins are indicated by angled arrows and letters. Cleavage sites, which are responsible for the post-translational production of MA, CA, NC, RT, IN, SU, and TM, are marked by short straight arrows and numbered. Gag, Pro, Pol, Env, p12, Tax, p27Rex, p21Rex, p30, and p13 all result from sense transcription of the viral genes. usHBZ and sHBZ are produced from mRNAs generated by antisense transcription. This is illustrated by a long bold arrow marked with letters N and C to designate the location of the beginnings and ends of the corresponding proteins. Obviously, the numbering of residues for the usHBZ and sHBZ presentation was inverted. The p27Rex, p30, and sHBZ proteins are translated from spliced genes. There are three alternatively spliced pairs of proteins in HTLV-1, namely, p21Rex/p27Rex, p13/p30, and sHBZ/usHBZ. Data were taken from ref 399.

mechanisms, being able to use both positive and negative sense or even ambisense transcription.

One illustrative example of such an ergonomic usage of genetic material is provided by a deltaretrovirus, human T-cell lymphotropic virus type 1 (HTLV-1), which is the causative agent of HTLV-1-associated myelopathy, adult T-cell leukemia (ATL), and *Strongyloides stercoralis* hyperinfection. Among the major products of the HTLV-1 genome are four common structural and enzymatic proteins that can be found in several retroviruses (Gag, Env, Pol, and Pro). Similarly to other retroviruses, various polyproteins, Gag, Gag-Pro, and Gag-Pro-Pol, are produced from the *Gag*, *Pro*, and *Pol* genes. These polyproteins undergo post-translational cleavage to generate seven proteins. The major structural proteins (nucleocapsid (NC), matrix (MA), and capsid (CA)) of the virus core are derived by the proteolytic cleavage of the *Gag*-encoded Gag polyprotein. The last part of the Gag-Pro polyprotein and the middle part of the Gag-Pro-Pol polyprotein are encoded by the *Pro* gene. The corresponding cleavage products include the viral protease (PR). Finally, the last part of the Gag-Pro-Pol polyprotein is encoded by the *Pol* gene. Reverse transcriptase (RT) and integrase (IN) are the result of Pol cleavage. Finally, the structural proteins of the viral envelope, SU (surface) and TM (transmembrane), are the cleavage products of the Env polyprotein encoded by the *Env* gene.

Besides these common retroviral proteins, various accessory and regulatory proteins are encoded by the pX region of the HTLV-1 genome in four overlapping open reading frames.[364,365] Antisense transcription of the *HBZ* gene yields the two forms of the HTLV-1 basic leucine zipper factor (HBZ),[366] spliced (sHBZ) and unspliced (usHBZ). There are multiple transcriptional initiation sites in the U5 and R regions of the 3′ LTR of the *sHBZ* transcript, and there is only one initiation site for the *usHBZ* gene located within the *tax* gene.[367] Additionally, differential splicing of the single genomic mRNA controls expression of the various ORFs and produces unspliced, singly spliced, and multiply spliced mRNAs.[368] Therefore, multiple nuclear and cytoplasmic processes, such as transcription, mRNA nuclear export, splicing, alternative splicing, RNA stability, and translation, control the replication of HTLV-1.[369]

Below we provide a detailed description of two HTLV-1 regulatory proteins needed for viral genome expression, namely Tax, which acts as a transcriptional activator of the viral promoter,[370−375] and Rex. Rex promotes expression of the Env, Gag, and Pol proteins and regulates transport of both the unspliced and singly spliced mRNA from the nucleus to the cytoplasm, thereby affecting post-transcriptional regulatory steps.[376−380] Tax and Rex are expressed from two overlapping ORFs of a bicistronic viral mRNA consisting of three exons and transcribed from the distal part of the pX region of the viral genome.[365,381,382] The Rex isoforms, namely, p27Rex and p21Rex (in which residues 1−78 are missing), are produced by alternative splicing of the corresponding mRNA. Furthermore, two accessory proteins, p30 and p13, are generated by alternative splicing of the pX region in open reading frame II.

Figure 8 shows the proteome map of HTLV-1 along with the corresponding disorder predictions. From this figure it is evident that the economic usage of genetic material by HTLV-1 is paralleled by a wide occurrence of ID in HIV-1 proteins. In particular, prevailing ID is observed in post-translational cleavage sites, leading to the production of the Gag, Pro, and Pol polyproteins from the Gag-Pro and Gar-Pro-Pol grand polyproteins, as well as in the cleavage site, yielding the MA, NC, CA, RT, TM, IN, and SU proteins from the corresponding polyproteins. This conclusion stems from the fact that regions surrounding the cleavage sites are predicted to have an increased flexibility (meaning that their disorder score is higher than 0.5). This observation holds also for the region around cleavage site 3: in fact, although the overall disorder score of the corresponding region is low, the actual cleavage site coincides with a peak in local disorder. In addition, proteins that are

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX−XXX

affected by alternative splicing are typically highly disordered, and protein fragments that have been removed by alternative splicing are mostly disordered. Also, protein fragments corresponding to overlapping genes either are disordered or have a complementary disorder distribution in cases where these protein regions are not translated from the genes transcribed from the identical ORFs. For instance, the N-terminal fragment of Tax, which overlaps with a significant portion of Rex, is mostly ordered, while the corresponding region in Rex is mostly disordered. Likewise, the C-terminal region of p30, which overlaps with the ordered N-terminal fragment of Tax, is predominantly disordered, much like the p13 protein that fully overlaps with the ordered N-terminal fragment of Tax. Finally, proteins that are translated from genes generated by antisense transcription tend to be highly disordered.

The hypothesis that the HTLV-1 proteins that are produced by overlapping genes are intrinsically disordered is supported by similar observations made on paramyxoviral P genes.[38,167] It is also supported by computational analyses of proteins encoded by overlapping genes from 43 genera of unspliced RNA viruses infecting eukaryotes.[383] In fact, sequence compositions of proteins produced from overlapping genes were globally biased toward disorder-promoting amino acids. Furthermore, these proteins were predicted to contain significantly more ID than viral proteins encoded by non-overlapping genes.[383] Further support for a tight relationship between ID and overlapping proteins is provided by a recent study aimed at identifying viral genes created de novo by overprinting of ancient open reading frames.[384] In addition, a study carried out by Kovacs and co-workers unveiled a high propensity for ID in dual-coding regions of human genes, which underscores the generality of the relationship between intrinsic disorder and overlapping reading frames.[385]

Many conserved genomic RNA structures, such as packaging signals, internal ribosome entry sites, ribosomal frame-shift motifs, cis-regulatory elements, tRNA mimics, and pseudo-knots,[386,387] are important for the multilevel regulation of viral replication. Furthermore, on the basis of the analysis of the architecture and secondary structure of the HIV-1 RNA genome, a correlation was recently found between the high levels of RNA structure and sequences that encode protein disorder in HIV proteins.[388] Specifically, at least 10 highly structured RNA regions were identified in the HIV-1 genome using high-throughput RNA analysis, nucleotide pairing probability, and selective 2′-hydroxyl acylation analyzed by primer extension. Intriguingly, flexible peptide linkers at junctions between the individual proteins in the corresponding polyproteins (e.g., matrix−capsid, capsid−nucleocapsid, and nucleocapsid−p6 junctions in Gag, protease−reverse transcriptase and reverse transcriptase−RNase H junctions in Pol, and signal peptide−gp120 and gp120−gp41 junctions in Env) were shown to be encoded by some of these highly structured RNA regions.[388] Furthermore, unstructured linker regions connecting ordered domains in individual HIV-1 proteins, such as integrase, capsid, and reverse transcriptase, were also shown to be encoded by ordered RNA regions, which typically were more highly structured than 95.2% of randomly selected genome regions.[388] These observations suggest that, in the HIV-1 genome, unstructured protein elements are encoded by higher order RNA structure. Such attenuation and pausing in elongation caused by highly structured RNA encoding protein−domain junctions provides a unique mechanism for the precise regulation of cotranslational protein folding by allowing domains to fold independently during translation.[388]

It is worth noting that many of the specific functional implementations of ID listed above are not unique to viral proteins. Indeed, also in other organisms, proteolytic cleavage of proteins often occurs in disordered regions.[115,116,389−393] Also, regions of mRNA that undergo alternative splicing tend to code for disordered proteins much more frequently than for structured proteins.[394] Finally, retro-proteins whose sequences are derived from the mRNA sequences that are read backward, yielding a new polypeptide that does not align with its parent sequence, were also shown to lack ordered 3D structure (see ref 395).

8. CONCLUSIONS

Promiscuity in binding is a key characteristic of viral proteins: while some viruses have a genome that encodes several proteins, that alone is not sufficient to sustain viral replication, which renders viruses dependent on the machinery of the host cell to complete their life cycles. Typically, viruses have highly compact genomes, and if a viral protein contains IDRs or is completely disordered, a single such protein is able to be involved in various tasks since it can interact with different partners. This is particularly well illustrated by the above-described case of HDV δAg.

The computational and experimental evidence collated so far and herein reviewed show that viruses broadly rely on ID. On one hand, viral proteins have a high occurrence of polar residues, possess a much weaker network of inter-residue interactions (illustrated by the increased fraction of residues not involved in regular secondary structure elements, the lower contact density parameters, and the abundance of short disordered regions), are less densely packed, and are characterized by an increased mutational robustness and lowered destabilizing effects of mutations. When these features are taken together, it can be concluded that viral proteins have been shaped in the course of evolution probably not to achieve a higher thermodynamic stability but rather to be able to better adapt to fast changes in their biological and physical environment and to allow an ergonomic usage of genetic material via alternative splicing, overlapping genes, and antisense transcription.

Taking into account the correlation between ID and promiscuity on one hand and the correlation between overlapping genes and disorder on the other hand,[383,384,396,397] it has been proposed that the primary advantage of the wealth of disorder within viruses is related to pleiotropy and genetic compaction.[187,193,202,398] In fact, disorder offers a solution to reduce not only the genome size but also molecular crowding, since a single gene would (i) encode a single (regulatory) protein product which can perform multiple interactions through use of its disordered regions and therefore exert multiple concomitant biological effects and/or (ii) encode several products via overlapping reading frames. Since disordered regions are not as sensitive to structural constraints as ordered regions, disorder within one or both protein products encoded by an overlapping reading frame can in fact represent a strategy to alleviate the evolutionary restrictions imposed by the overlap. As a result, disorder can confer to viruses the ability to "handle" overlaps and, as a result, further expands the coding potential of viral genomes.

dx.doi.org/10.1021/cr4005692 | Chem. Rev. XXXX, XXX, XXX−XXX

In conclusion, ID is of benefit to viruses as it allows an economic usage of genetic material, enables them to tolerate the high mutation rates that viral genomes are subjected to, and renders them more adapted to hostile and changing environmental conditions. Concomitantly, the inherent plasticity of IDRs confers to viral proteins the ability to interact with needed host machinery components while evading the defense mechanisms of the host. As such, targeting IDRs within viral proteins to impair critical protein−protein interactions could constitute a general appealing antiviral strategy.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: vuversky@health.usf.edu.
*E-mail: Sonia.Longhi@afmb.univ-mrs.fr.

Notes

The authors declare no competing financial interest.

Biographies

Bin Xue is an assistant professor in the Department of Cell Biology, Microbiology, and Molecular Biology at the University of South Florida. After obtaining his Ph.D. in soft condensed matter physics from Nanjing University, China (2005), Dr. Xue continued his research in computational structural proteomics at the State University of New York in Buffalo and then at Indiana University-Purdue University in Indianapolis. In 2008, he joined the Center for Computational Biology and Bioinformatics and the Institute for Intrinsically Disordered Protein Research at Indiana University-Purdue University to work on intrinsically disordered proteins and bioinformatics. He was a member of the research faculty in the Department of Molecular Biology at the University of South Florida before taking his current position. Dr. Xue has published more than 60 peer-reviewed papers and several book chapters. He is an associate editor, a regional editor, an academic editor, or an editorial member of many international journals.

David Blocquel began his university studies at the University of Nancy I, where he obtained a master's degree in molecular biology and biochemistry. He then continued his studies with a master's degree in biotechnology in Montpellier. He prepared his Ph.D. thesis within the Structural Disorder and Molecular Recognition team of the AFMB laboratory (Marseille, France) under the supervision of Dr. Sonia Longhi. He defended his thesis in December 2013 and received his Ph.D. degree from the Aix-Marseille University. To date, he has authored six scientific papers focused on the biochemical and biophysical characterization of the nucleoproteins and the phosphoproteins from paramyxoviruses and a review on the molecular mechanisms of transcription and replication in Mononegavirales members.

Johnny Habchi obtained a bachelor's degree from the Lebanese University of Beirut in 2008 and then a master's degree in structural biology in 2009 at the Aix-Marseille University. He prepared his Ph.D. thesis within the Structural Disorder and Molecular Recognition team at the AFMB laboratory (Marseille, France) under the supervision of Dr. Sonia Longhi. He received his Ph.D. in biophysics from the Aix-Marseille University in 2012. Since October 2012 he has been a postdoctoral fellow in Michele Vendruscolo's group within the Department of Chemistry, University of Cambridge (United Kingdom). So far, he has authored 13 papers, with most of them focusing on the characterization of the intrinsically disordered regions within the replicative complex of henipaviruses and the induced folding that these regions undergo in the presence of their partners.

dx.doi.org/10.1021/cr4005692 | Chem. Rev. XXXX, XXX, XXX−XXX

Alexey V. Uversky received bachelor's degrees in mathematics from Butler University and computer engineering from Indiana University-Purdue University in Indianapolis in 2010. He is currently a Ph.D. student being supervised by Dr. Zoran Obradovic at the Temple University Center for Science and Technology (CST), with an expected graduation date of summer 2015. His research is focused on social network analysis in the context of data mining and machine learning tasks, and he has worked on several Defense Advanced Research Projects Agency (DARPA)-sponsored projects related to those fields, which have led to two publications. He has also worked as a research assistant on a drug repositioning project for the pharmaceutical company GlaxoSmithKline.

Lukasz Kurgan received Ph.D. in computer science from the University of Colorado at Boulder in 2003. Currently he is a professor in the Department of Electrical and Computer Engineering at the University of Alberta. His research interests are in the development and application of computational methods for high-throughput analysis of the sequences, structures, and functions of proteins and short RNAs. He has published close to 100 peer-reviewed journal papers. More details are available on the Web site of his laboratory at http://biomine.ece.ualberta.ca/.

Vladimir N. Uversky obtained his Ph.D. in biophysics from the Moscow Institute of Physics and Technology (1991) and D.Sc. in biophysics from the Institute of Experimental and Theoretical Biophysics, Russian Academy of Sciences (1998). He spent his early career working on protein folding at the Institute of Protein Research and the Institute for Biological Instrumentation (Russian Academy of Sciences). In 1998, he moved to the University of California, Santa Cruz, to work on protein folding, misfolding, and protein intrinsic disorder. In 2004, he moved to the Center for Computational Biology and Bioinformatics at Indiana University-Purdue University, Indianapolis, to work on intrinsically disordered proteins. Since 2010, he has been with the Department of Molecular Biology at the University of South Florida.

Sonia Longhi is Director of Research at the Center for National Scientific Research (CNRS), heading the Structural Disorder and Molecular Recognition group within the AFMB laboratory. She obtained a Ph.D. in molecular biology from the University of Milan in 1993. She received an HDR in structural virology from the Aix-Marseille University in 2003. Her scientific focus is on intrinsically disordered proteins and the mechanistic and functional aspects of the interactions they establish with partners using paramyxoviral proteins as model systems. She has authored 90 scientific publications in international peer-reviewed journals, edited a book on measles virus nucleoprotein, and coedited with Prof. Vladimir Uversky a book entitled *Instrumental Analysis of Intrinsically Disordered Proteins* and one entitled *Flexible Viruses—Structural Disorder in Viral Proteins*.

ACKNOWLEDGMENTS

This work was carried out with the financial support of the Agence Nationale de la Recherche, specific programs Physico-Chimie du Vivant, Grant ANR-08-PCVI-0020-01, and ASTRID, Grant ANR-11-ASTR-003-01 (S.L.). This work was also supported in part by the Russian Academy of Sciences

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX–XXX

Molecular and Cellular Biology Program (V.N.U.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ABBREVIATIONS

AFV1	*Acidianus filamentous* virus 1
AIDS	acquired immunodeficiency syndrome
ANS	8-anilino-1-naphthalenesulfonic acid
APMV	*Acanthamoeba polyphaga* mimivirus
ARFP	alternative reading frame protein
ARM	arginine-rich motif
ATL	adult T-cell leukemia
BVDV	bovine viral diarrhea virus
CA	capsid
CCMV	cowpea chlorotic mottle virus
CD	circular dichroism
CDK9	cyclin-dependent kinase 9
CsA	cyclosporin A
CTD	C-terminal domain
CypA	cyclophilin A
δA	δ antigen
D1, D2, and D3	domains 1, 2, and 3
DLS	dynamic light scattering
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
E	early ORFs
E7N	N-terminal domian of E7
EIAV	equine infectious anemia virus
ELMs	eukaryotic linear motifs
Env	envelope protein
FMDV	foot-and-mouth-disease virus
FTIR	Fourier transform infrared
GBV-G	GB virus B
HA	hemagglutinin
HAdV5	human adenovirus type 5
HBZ	basic leucine zipper factor
HCV	hepatitis C virus
HCV-C	HCV core protein
HDV	hepatitis δ virus
HEF	HA esterase fusion
HELD	helicase domain
HeV	Hendra virus
HIV-1	human immunodeficiency virus-1
HMM	hidden Markov model
HPV	human papilloma virus
HTLV-1	human T-cell lymphotropic virus type 1
ID	intrinsic disorder
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IN	integrase
IRES	internal ribosome entry site
IRF3	interferon regulator factor 3
L	large protein
LMV	lettuce mosaic virus
MA	matrix
MeV	measles virus
MoRE	molecular recognition element
MP	movement protein
MVM	minute virus of mice
N	nucleoprotein
N°	monomeric form of N
λN	N protein of bacteriophage λ
NC	nucleocapsid
N_{CORE}	N-terminal domain of N
N_{TAIL}	C-terminal domain of N
NiV	Nipah virus
NLP	nucleocapsid-like particle
NLS	nuclear localization signal
NMR	nuclear magnetic resonance
NR	nucleoprotein receptor
NS	nonstructural
NS5BΔ21	C-terminal truncated form of NS5B
NTD	N-terminal domain
ORF	open reading frame
P	phosphoprotein
PCT	C-terminal domain of P
PNT	N-terminal domain of P
PMD	P multimerization domain
Pol	polymerase
pI	isoelectric point
PKB	protein kinase P
PONDR	predictor of natural disordered regions
PPIase	peptidyl−prolyl cis/trans isomerase
PP2A	protein phosphatase 2A
PSLV	Poe semilatent virus
PR	protease
PV	papilloma virus
PVA	potato virus A
PVY	potato virus Y
RabV	rabies virus
Rb	retinoblastoma tumor suppressor protein
RbAB	AB domain of Rb
RDC	residual dipolar coupling
RdRp	RNA-dependent RNA polymerase
RRE	Rev response element
RSV	respiratory syncytial virus
RT	reverse transcriptase
RYMV	rice yellow mottle virus
SARS	severe acute respiratory syndrome
SARS-CoV	SARS coronavirus
SAXS	small-angle X-ray scattering
SA1−3	splice acceptor sites 1−3
SD1	splice donor 1
SDS	sodium dodecyl sulfate
SeMV	Sesbania mosaic virus
SeV	Sendai virus
SFV	Semliki forest virus
SFVP	SFV capsid protein
SIFV	*Sulfolobus islandicus* filamentous virus
SIV_{mac}	simian immunodeficiency virus
SLiMs	short linear motifs
ssDNA	single-stranded DNA
ssRNA+	single-stranded RNA of positive (+) sense
ssRNA−	single-stranded RNA of negative (−) sense
SU	surface
SV40	simian virus 40
TFE	trifluoroethanol
TGB	triple gene block
TM	transmembrane
TMV	tobacco mosaic virus
XD	X domain of P
VPg	viral genome-linked protein
VSV	vesicular stomatitis virus
WNV	West Nile virus

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX−XXX

REFERENCES

(1) Anfinsen, C. B. *Science* **1973**, *181*, 223.
(2) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1999**, *293*, 321.
(3) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins* **2000**, *41*, 415.
(4) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. *J. Mol. Graphics Modell.* **2001**, *19*, 26.
(5) Dunker, A. K.; Brown, C. J.; Obradovic, Z. *Adv. Protein Chem.* **2002**, *62*, 25.
(6) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. *Biochemistry* **2002**, *41*, 6573.
(7) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739.
(8) Uversky, V. N. *Eur. J. Biochem.* **2002**, *269*, 2.
(9) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527.
(10) Tompa, P. *J. Mol. Struct.: THEOCHEM* **2003**, *666*−*667*, 361.
(11) Uversky, V. N. *Cell. Mol. Life Sci.* **2003**, *60*, 1852.
(12) Minezaki, Y.; Homma, K.; Kinjo, A. R.; Nishikawa, K. *J. Mol. Biol.* **2006**, *359*, 1137.
(13) Dunker, A. K.; Babu, M. M.; Barbar, E.; Blackledge, M.; Bondos, S. E.; Dosztányi, Z.; Dyson, H. J.; Forman-Kay, J.; Fuxreiter, M.; Gsponer, J.; Han, K.-H.; Jones, D. T.; Longhi, S.; Metallo, S. J.; Nishikawa, K.; Nussinov, R.; Obradovic, Z.; Pappu, R. V.; Rost, B.; Selenko, P.; Subramaniam, V.; Sussman, J. L.; Tompa, P.; Uversky, V. N. *Intrinsically Disord. Proteins* **2013**, *1*, e24157.
(14) Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. *FEBS J.* **2005**, *272*, 5129.
(15) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *J. Mol. Recognit.* **2005**, *18*, 343.
(16) Radivojac, P.; Iakoucheva, L. M.; Oldfield, C. J.; Obradovic, Z.; Uversky, V. N.; Dunker, A. K. *Biophys. J.* **2007**, *92*, 1439.
(17) Dunker, A. K.; Oldfield, C. J.; Meng, J.; Romero, P.; Yang, J. Y.; Chen, J. W.; Vacic, V.; Obradovic, Z.; Uversky, V. N. *BMC Genomics* **2008**, *9* (Suppl.2), S1.
(18) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756.
(19) Dunker, A. K.; Uversky, V. N. *Nat. Chem. Biol.* **2008**, *4*, 229.
(20) Oldfield, C. J.; Meng, J.; Yang, J. Y.; Yang, M. Q.; Uversky, V. N.; Dunker, A. K. *BMC Genomics* **2008**, *9* (Suppl.1), S1.
(21) Tompa, P.; Fuxreiter, M.; Oldfield, C. J.; Simon, I.; Dunker, A. K.; Uversky, V. N. *Bioessays* **2009**, *31*, 328.
(22) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197.
(23) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31.
(24) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* **2004**, *337*, 635.
(25) Mohan, A.; Sullivan, W. J., Jr.; Radivojac, P.; Dunker, A. K.; Uversky, V. N. *Mol. Biosyst.* **2008**, *4*, 328.
(26) Schad, E.; Tompa, P.; Hegyi, H. *Genome Biol.* **2011**, *12*, R120.
(27) Romero, P.; Obradovic, Z.; Kissinger, C. R.; Villafranca, J. E.; Dunker, A. K. *IEEE Int. Conf. Neural Networks* **1997**, *1*, 90.
(28) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins* **2001**, *42*, 38.
(29) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Romero, P.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2005**, *44*, 12454.
(30) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Brown, C. J.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2005**, *44*, 1989.
(31) Feng, Z. P.; Zhang, X.; Han, P.; Arora, N.; Anders, R. F.; Norton, R. S. *Mol. Biochem. Parasitol.* **2006**, *150*, 256.
(32) Bogatyreva, N. S.; Finkelstein, A. V.; Galzitskaya, O. V. *J. Bioinf. Comput. Biol.* **2006**, *4*, 597.
(33) Burra, P. V.; Kalmar, L.; Tompa, P. *PLoS One* **2010**, *5*, e12069.
(34) Pancsa, R.; Tompa, P. *PLoS One* **2012**, *7*, e34687.
(35) Drake, J. W.; Charlesworth, B.; Charlesworth, D.; Crow, J. F. *Genetics* **1998**, *148*, 1667.
(36) Reanney, D. C. *Annu. Rev. Microbiol.* **1982**, *36*, 47.
(37) Karlin, D.; Longhi, S.; Receveur, V.; Canard, B. *Virology* **2002**, *296*, 251.
(38) Karlin, D.; Ferron, F.; Canard, B.; Longhi, S. *J. Gen. Virol.* **2003**, *84*, 3239.
(39) Longhi, S.; Receveur-Brechot, V.; Karlin, D.; Johansson, K.; Darbon, H.; Bhella, D.; Yeo, R.; Finet, S.; Canard, B. *J. Biol. Chem.* **2003**, *278*, 18638.
(40) Tokuriki, N.; Oldfield, C. J.; Uversky, V. N.; Berezovsky, I. N.; Tawfik, D. S. *Trends Biochem. Sci.* **2009**, *34*, 53.
(41) Uversky, V. N.; Longhi, S. *Flexible Viruses: Structural Disorder in Viral Proteins*; John Wiley and Sons: Hoboken, NJ, 2012.
(42) Xue, B.; Williams, R. W.; Oldfield, C. J.; Goh, G. K.; Dunker, A. K.; Uversky, V. N. *Protein Pept. Lett.* **2010**, *17*, 932.
(43) Xue, B.; Dunker, A. K.; Uversky, V. N. *J. Biomol. Struct. Dyn.* **2012**, *30*, 137.
(44) Alves, C.; Cunha, C. *Future Virol.* **2012**, *7*, 1183.
(45) Breitbart, M.; Rohwer, F. *Trends Microbiol.* **2005**, *13*, 278.
(46) Bergh, O.; Borsheim, K. Y.; Bratbak, G.; Heldal, M. *Nature* **1989**, *340*, 467.
(47) Sano, E.; Carlson, S.; Wegley, L.; Rohwer, F. *Appl. Environ. Microbiol.* **2004**, *70*, 5842.
(48) Edwards, R. A.; Rohwer, F. *Nat. Rev. Microbiol.* **2005**, *3*, 504.
(49) Lawrence, C. M.; Menon, S.; Eilers, B. J.; Bothner, B.; Khayat, R.; Douglas, T.; Young, M. J. *J. Biol. Chem.* **2009**, *284*, 12599.
(50) La Scola, B.; Desnues, C.; Pagnier, I.; Robert, C.; Barrassi, L.; Fournous, G.; Merchat, M.; Suzan-Monti, M.; Forterre, P.; Koonin, E.; Raoult, D. *Nature* **2008**, *455*, 100.
(51) La Scola, B.; Audic, S.; Robert, C.; Jungang, L.; de Lamballerie, X.; Drancourt, M.; Birtles, R.; Claverie, J. M.; Raoult, D. *Science* **2003**, *299*, 2033.
(52) Raoult, D.; Audic, S.; Robert, C.; Abergel, C.; Renesto, P.; Ogata, H.; La Scola, B.; Suzan, M.; Claverie, J. M. *Science* **2004**, *306*, 1344.
(53) Koonin, E. V. *Curr. Biol.* **2005**, *15*, R167.
(54) Suzan-Monti, M.; La Scola, B.; Barrassi, L.; Espinosa, L.; Raoult, D. *PLoS One* **2007**, *2*, e328.
(55) Baltimore, D. *Bacteriol. Rev.* **1971**, *35*, 235.
(56) Rybicki, E. P. *S. Afr. J. Sci.* **1990**, *86*, 182.
(57) Holmes, E. C. *PLoS Biol.* **2007**, *5*, e278.
(58) Piacente, F.; Marin, M.; Molinaro, A.; De Castro, C.; Seltzer, V.; Salis, A.; Damonte, G.; Bernardi, C.; Claverie, J. M.; Abergel, C.; Tonetti, M. *J. Biol. Chem.* **2012**, *287*, 3009.
(59) Forterre, P. *Virus Res.* **2006**, *117*, 5.
(60) Iyer, L. M.; Balaji, S.; Koonin, E. V.; Aravind, L. *Virus Res.* **2006**, *117*, 156.
(61) Koonin, E. V.; Senkevich, T. G.; Dolja, V. V. *Biol. Direct* **2006**, *1*, 29.
(62) Canchaya, C.; Fournous, G.; Chibani-Chennoufi, S.; Dillmann, M. L.; Brussow, H. *Curr. Opin. Microbiol.* **2003**, *6*, 417.
(63) Raoult, D.; Forterre, P. *Nat. Rev. Microbiol.* **2008**, *6*, 315.
(64) Caspar, D. L.; Klug, A. *Cold Spring Harbor Symp. Quant. Biol.* **1962**, *27*, 1.
(65) Rossmann, M. G.; Mesyanzhinov, V. V.; Arisaka, F.; Leiman, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 171.
(66) Suzuki, Y. *Biol. Pharm. Bull.* **2005**, *28*, 399.
(67) Cady, S. D.; Luo, W.; Hu, F.; Hong, M. *Biochemistry* **2009**, *48*, 7356.
(68) Nayak, D. P.; Hui, E. K.; Barman, S. *Virus Res.* **2004**, *106*, 147.
(69) Nayak, D. P.; Balogun, R. A.; Yamada, H.; Zhou, Z. H.; Barman, S. *Virus Res.* **2009**, *143*, 147.
(70) Hughes, F. J.; Romanos, M. A. *Nucleic Acids Res.* **1993**, *21*, 5817.
(71) Ustav, M.; Stenlund, A. *EMBO J.* **1991**, *10*, 449.
(72) Frattini, M. G.; Laimins, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12398.
(73) Mohr, I. J.; Clark, R.; Sun, S.; Androphy, E. J.; MacPherson, P.; Botchan, M. R. *Science* **1990**, *250*, 1694.
(74) Cripe, T. P.; Haugen, T. H.; Turk, J. P.; Tabatabai, F.; Schmid, P. G., 3rd; Durst, M.; Gissmann, L.; Roman, A.; Turek, L. P. *EMBO J.* **1987**, *6*, 3745.
(75) Gloss, B.; Bernard, H. U.; Seedorf, K.; Klock, G. *EMBO J.* **1987**, *6*, 3735.

(76) Wilson, R.; Fehrmann, F.; Laimins, L. A. *J. Virol.* **2005**, *79*, 6732.
(77) Brown, D. R.; Kitchin, D.; Qadadri, B.; Neptune, N.; Batteiger, T.; Ermel, A. *Virology* **2006**, *345*, 290.
(78) Davy, C. E.; Ayub, M.; Jackson, D. J.; Das, P.; McIntosh, P.; Doorbar, J. *Virology* **2006**, *349*, 230.
(79) Leechanachai, P.; Banks, L.; Moreau, F.; Matlashewski, G. *Oncogene* **1992**, *7*, 19.
(80) Straight, S. W.; Hinkle, P. M.; Jewers, R. J.; McCance, D. J. *J. Virol.* **1993**, *67*, 4521.
(81) Fehrmann, F.; Klumpp, D. J.; Laimins, L. A. *J. Virol.* **2003**, *77*, 2819.
(82) Genther, S. M.; Sterling, S.; Duensing, S.; Munger, K.; Sattler, C.; Lambert, P. F. *J. Virol.* **2003**, *77*, 2832.
(83) Zhang, B.; Li, P.; Wang, E.; Brahmi, Z.; Dunn, K. W.; Blum, J. S.; Roman, A. *Virology* **2003**, *310*, 100.
(84) Gao, L.; Aizaki, H.; He, J. W.; Lai, M. M. *J. Virol.* **2004**, *78*, 3480.
(85) Chung, K. M.; Liszewski, M. K.; Nybakken, G.; Davis, A. E.; Townsend, R. R.; Fremont, D. H.; Atkinson, J. P.; Diamond, M. S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19111.
(86) Boxer, E. L.; Nanda, S. K.; Baron, M. D. *Virology* **2009**, *385*, 134.
(87) Fontana, J. M.; Bankamp, B.; Rota, P. A. *Immunol. Rev.* **2008**, *225*, 46.
(88) Doerig, C.; Hirt, B.; Beard, P.; Antonietti, J. P. *J. Gen. Virol.* **1988**, *69* (Part 10), 2563.
(89) Seelamgari, A.; Maddukuri, A.; Berro, R.; de la Fuente, C.; Kehn, K.; Deng, L.; Dadgar, S.; Bottazzi, M. E.; Ghedin, E.; Pumfery, A.; Kashanchi, F. *Front. Biosci.* **2004**, *9*, 2388.
(90) Forterre, P.; Prangishvili, D. *Res. Microbiol.* **2009**, *160*, 466.
(91) Prangishvili, D.; Forterre, P.; Garrett, R. A. *Nat. Rev. Microbiol.* **2006**, *4*, 837.
(92) Vacic, V.; Uversky, V. N.; Dunker, A. K.; Lonardi, S. *BMC Bioinf.* **2007**, *8*, 211.
(93) Xue, B.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N. *FEBS Lett.* **2009**, *583*, 1469.
(94) Xue, B.; Li, L.; Meroueh, S. O.; Uversky, V. N.; Dunker, A. K. *Mol. Biosyst.* **2009**, *12*, 1688.
(95) Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. *Genome Inf. Ser. Workshop Genome Inf.* **2000**, *11*, 161.
(96) Romero, P.; Obradovic, Z.; Kissinger, C. R.; Villafranca, J. E.; Garner, E.; Guilliot, S.; Dunker, A. K. *Pac. Symp. Biocomput.* **1998**, 437.
(97) Bateman, A.; Birney, E.; Cerruti, L.; Durbin, R.; Etwiller, L.; Eddy, S. R.; Griffiths-Jones, S.; Howe, K. L.; Marshall, M.; Sonnhammer, E. L. *Nucleic Acids Res.* **2002**, *30*, 276.
(98) Bateman, A.; Coin, L.; Durbin, R.; Finn, R. D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E. L.; Studholme, D. J.; Yeats, C.; Eddy, S. R. *Nucleic Acids Res.* **2004**, *32*, D138.
(99) Finn, R. D.; Tate, J.; Mistry, J.; Coggill, P. C.; Sammut, S. J.; Hotz, H. R.; Ceric, G.; Forslund, K.; Eddy, S. R.; Sonnhammer, E. L.; Bateman, A. *Nucleic Acids Res.* **2008**, *36*, D281.
(100) Berezovsky, I. N. *Phys. Biol.* **2011**, *8*, 035002.
(101) Davey, N. E.; Trave, G.; Gibson, T. J. *Trends Biochem. Sci.* **2011**, *36*, 159.
(102) Neduva, V.; Russell, R. B. *FEBS Lett.* **2005**, *579*, 3342.
(103) Dinkel, H.; Michael, S.; Weatheritt, R. J.; Davey, N. E.; Van Roey, K.; Altenberg, B.; Toedt, G.; Uyar, B.; Seiler, M.; Budd, A.; Jodicke, L.; Dammert, M. A.; Schroeter, C.; Hammer, M.; Schmidt, T.; Jehl, P.; McGuigan, C.; Dymecka, M.; Chica, C.; Luck, K.; Via, A.; Chatr-Aryamontri, A.; Haslam, N.; Grebnev, G.; Edwards, R. J.; Steinmetz, M. O.; Meiselbach, H.; Diella, F.; Gibson, T. J. *Nucleic Acids Res.* **2012**, *40*, D242.
(104) Pushker, R.; Mooney, C.; Davey, N. E.; Jacque, J. M.; Shields, D. C. *PLoS One* **2013**, *8*, e60724.
(105) Harrison, S. C.; Olson, A. J.; Schutt, C. E.; Winkler, F. K.; Bricogne, G. *Nature* **1978**, *276*, 368.
(106) Abad-Zapatero, C.; Abdel-Meguid, S. S.; Johnson, J. E.; Leslie, A. G.; Rayment, I.; Rossmann, M. G.; Suck, D.; Tsukihara, T. *Nature* **1980**, *286*, 33.
(107) Liljas, L. In *Conformational Proteomics of Macromolecular Architecture*; Cheng, R. H., Hammar, L., Eds.; World Scientific Publishing Co.: Hackensack, NJ, London, Singapore, Beijing, Shanghai, Hong Kong, Taipei, Taiwan, Chennai, India, 2004.
(108) Baker, T. S.; Caspar, D. L.; Murakami, W. T. *Nature* **1983**, *303*, 446.
(109) Liddington, R. C.; Yan, Y.; Moulai, J.; Sahli, R.; Benjamin, T. L.; Harrison, S. C. *Nature* **1991**, *354*, 278.
(110) Rayment, I.; Baker, T. S.; Caspar, D. L.; Murakami, W. T. *Nature* **1982**, *295*, 110.
(111) Fry, E. E.; Newman, J. W.; Curry, S.; Najjam, S.; Jackson, T.; Blakemore, W.; Lea, S. M.; Miller, L.; Burman, A.; King, A. M.; Stuart, D. I. *J. Gen. Virol.* **2005**, *86*, 1909.
(112) Choi, Y. G.; Grantham, G. L.; Rao, A. L. *Virology* **2000**, *270*, 377.
(113) Nair, S.; Murthy, M. R. N.; Savithri, H. S. In *Flexible Viruses: structural Disorder in Viral Proteins*; Uversky, V. N., Longhi, S., Eds.; John Wiley and Sons: Hoboken, NJ, 2012.
(114) Liljas, L. In *Flexible Viruses: Structural Disorder in Viral Proteins*; Uversky, V. N., Longhi, S., Eds.; John Wiley and Sons: Hoboken, NJ, 2012.
(115) Fontana, A.; Polverino de Laureto, P.; De Filippis, V.; Scaramella, E.; Zambonin, M. *Folding Des.* **1997**, *2*, R17.
(116) Fontana, A.; de Laureto, P. P.; Spolaore, B.; Frare, E.; Picotti, P.; Zambonin, M. *Acta Biochim. Polym.* **2004**, *51*, 299.
(117) Marvin, D. A.; Hohn, B. *Bacteriol. Rev.* **1969**, *33*, 172.
(118) Model, P.; Russel, M. In *The Bacteriophages*; Calendar, R., Ed.; Plenum Press: New York, 1988; Vol. *2*.
(119) Marvin, D. A. In *The Single Stranded DNA Phages*; Denhardt, D. T., Dressier, D., Ray, D., Eds.; Cold Spring Harbor Laboratories: Cold Spring Harbor, NY, 1978.
(120) Nakashima, Y.; Konigsberg, W. *J. Mol. Biol.* **1974**, *88*, 598.
(121) Day, L. A. *J. Mol. Biol.* **1969**, *39*, 265.
(122) Williams, R. W.; Dunker, A. K. *J. Biol. Chem.* **1977**, *252*, 6253.
(123) Williams, R. W.; Dunker, A. K. *J. Mol. Biol.* **1981**, *152*, 783.
(124) Williams, R. W.; Dunker, A. K.; Peticolas, W. L. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1984**, *791*, 131.
(125) Williams, R. W.; Dunker, A. K.; Peticolas, W. L. *Biophys. J.* **1980**, *32*, 232.
(126) Thomas, G. J., Jr.; Prescott, B.; Day, L. A. *J. Mol. Biol.* **1983**, *165*, 321.
(127) Grygon, C. A.; Perno, J. R.; Fodor, S. P.; Spiro, T. G. *BioTechniques* **1988**, *6*, 50.
(128) Clack, B. A.; Gray, D. M. *Biopolymers* **1989**, *28*, 1861.
(129) Asbeck, V. F.; Beyreuther, K.; Kohler, H.; von Wettstein, O.; Braunitzer, G. *Hoppe-Seyler's Z. Physiol. Chem.* **1969**, *350*, 1047.
(130) Marvin, D. A.; Symmons, M. F.; Straus, S. K. *Prog. Biophys. Mol. Biol.* **2014**, *114*, 80.
(131) Trenkner, E.; Bonhoeffer, F.; Gierer, A. *Biochem. Biophys. Res. Commun.* **1967**, *28*, 932.
(132) Marvin, D. A. *Int. J. Biol. Macromol.* **1989**, *11*, 159.
(133) Dunker, A. K.; Ensign, L. D.; Arnold, G. E.; Roberts, L. M. *FEBS Lett.* **1991**, *292*, 271.
(134) Griffith, J.; Manning, M.; Dunn, K. *Cell* **1981**, *23*, 747.
(135) Manning, M.; Chrysogelos, S.; Griffith, J. *J. Virol.* **1981**, *40*, 912.
(136) Lopez, J.; Webster, R. E. *J. Virol.* **1982**, *42*, 1099.
(137) Dunker, A. K.; Ensign, L. D.; Arnold, G. E.; Roberts, L. M. *FEBS Lett.* **1991**, *292*, 275.
(138) Ptitsyn, O. B.; Bychkova, V. E.; Uversky, V. N. *Philos. Trans. R. Soc. London, B* **1995**, *348*, 35.
(139) Endo, T.; Schatz, G. *EMBO J.* **1988**, *7*, 1153.
(140) Bychkova, V. E.; Dujsekina, A. E.; Klenin, S. I.; Tiktopulo, E. I.; Uversky, V. N.; Ptitsyn, O. B. *Biochemistry* **1996**, *35*, 6058.
(141) Uversky, V. N.; Narizhneva, N. V.; Kirschstein, S. O.; Winter, S.; Lober, G. *Folding Des.* **1997**, *2*, 163.
(142) Narizhneva, N. V.; Uversky, V. N. *Protein Pept. Lett.* **1997**, *4*, 243.
(143) Strauss, J. H.; Strauss, E. G. *Microbiol. Rev.* **1994**, *58*, 491.

(144) Harrison, S. C.; Strong, R. K.; Schlesinger, S.; Schlesinger, M. J. *J. Mol. Biol.* **1992**, *226*, 277.
(145) Morillas, M.; Eberl, H.; Allain, F. H.; Glockshuber, R.; Kuennemann, E. *J. Mol. Biol.* **2008**, *376*, 721.
(146) Choi, H. K.; Lu, G.; Lee, S.; Wengler, G.; Rossmann, M. G. *Proteins* **1997**, *27*, 345.
(147) Morillas, M.; Eberl, H.; Allain, F. H.; Glockshuber, R.; Kuennemann, E. In *Flexible Viruses: Structural Disorder in Viral Proteins*; Uversky, V. N., Longhi, S., Eds.; John Wiley and Sons: Hoboken, NJ, 2012.
(148) Ivanyi-Nagy, R.; Lavergne, J. P.; Gabus, C.; Ficheux, D.; Darlix, J. L. *Nucleic Acids Res.* **2008**, *36*, 712.
(149) Ivanyi-Nagy, R.; Darlix, J. L. *Protein Pept. Lett.* **2010**, *17*, 1019.
(150) Ivanyi-Nagy, R.; Darlix, J. L. *Adv. Exp. Med. Biol.* **2012**, *725*, 142.
(151) Ivanyi-Nagy, R.; Pécheur, E.; Darlix, J. L. In *Flexible Viruses: Structural Disorder in Viral Proteins*; Uversky, V. N., Longhi, S., Eds.; John Wiley and Sons: Hoboken, NJ, 2012.
(152) Kunkel, M.; Watowich, S. J. *FEBS Lett.* **2004**, *557*, 174.
(153) Duvignaud, J. B.; Savard, C.; Fromentin, R.; Majeau, N.; Leclerc, D.; Gagne, S. M. *Biochem. Biophys. Res. Commun.* **2009**, *378*, 27.
(154) Receveur-Bréchot, V.; Bourhis, J. M.; Uversky, V. N.; Canard, B.; Longhi, S. *Proteins* **2006**, *62*, 24.
(155) Boulant, S.; Vanbelle, C.; Ebel, C.; Penin, F.; Lavergne, J. P. *J. Virol.* **2005**, *79*, 11353.
(156) Tompa, P.; Csermely, P. *FASEB J.* **2004**, *18*, 1169.
(157) Carmona, P.; Molina, M. *Biochemistry* **2010**, *49*, 4724.
(158) Dokland, T.; Walsh, M.; Mackenzie, J. M.; Khromykh, A. A.; Ee, K. H.; Wang, S. *Structure* **2004**, *12*, 1157.
(159) Jones, C. T.; Ma, L.; Burgner, J. W.; Groesch, T. D.; Post, C. B.; Kuhn, R. J. *J. Virol.* **2003**, *77*, 7143.
(160) Ma, L.; Jones, C. T.; Groesch, T. D.; Kuhn, R. J.; Post, C. B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3414.
(161) Murray, C. L.; Marcotrigiano, J.; Rice, C. M. *J. Virol.* **2008**, *82*, 1294.
(162) Blocquel, D.; Bourhis, J. M.; Eléouët, J. F.; Gerlier, D.; Habchi, J.; Jamin, M.; Longhi, S.; Yabukarski, F. *Virologie* **2012**, *16*, 225.
(163) Karlin, D.; Longhi, S.; Canard, B. *Virology* **2002**, *302*, 420.
(164) Bourhis, J. M.; Johansson, K.; Receveur-Brechot, V.; Oldfield, C. J.; Dunker, K. A.; Canard, B.; Longhi, S. *Virus Res.* **2004**, *99*, 157.
(165) Houben, K.; Marion, D.; Tarbouriech, N.; Ruigrok, R. W.; Blanchard, L. *J. Virol.* **2007**, *81*, 6807.
(166) Jensen, M. R.; Houben, K.; Lescop, E.; Blanchard, L.; Ruigrok, R. W.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 8055.
(167) Habchi, J.; Mamelli, L.; Darbon, H.; Longhi, S. *PLoS One* **2010**, *5*, e11684.
(168) Johansson, K.; Bourhis, J. M.; Campanacci, V.; Cambillau, C.; Canard, B.; Longhi, S. *J. Biol. Chem.* **2003**, *278*, 44567.
(169) Habchi, J.; Blangy, S.; Mamelli, L.; Jensen, M. R.; Blackledge, M.; Darbon, H.; Oglesbee, M.; Shu, Y.; Longhi, S. *J. Biol. Chem.* **2011**, *286*, 13583.
(170) Kingston, R. L.; Hamel, D. J.; Gay, L. S.; Dahlquist, F. W.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8301.
(171) Bourhis, J. M.; Receveur-Brechot, V.; Oglesbee, M.; Zhang, X.; Buccellato, M.; Darbon, H.; Canard, B.; Finet, S.; Longhi, S. *Protein Sci.* **2005**, *14*, 1975.
(172) Morin, B.; Bourhis, J. M.; Belle, V.; Woudstra, M.; Carriere, F.; Guigliarelli, B.; Fournel, A.; Longhi, S. *J. Phys. Chem. B* **2006**, *110*, 20596.
(173) Belle, V.; Rouger, S.; Costanzo, S.; Liquiere, E.; Strancar, J.; Guigliarelli, B.; Fournel, A.; Longhi, S. *Proteins* **2008**, *73*, 973.
(174) Bernard, C.; Gely, S.; Bourhis, J. M.; Morelli, X.; Longhi, S.; Darbon, H. *FEBS Lett.* **2009**, *583*, 1084.
(175) Bischak, C. G.; Longhi, S.; Snead, D. M.; Costanzo, S.; Terrer, E.; Londergan, C. H. *Biophys. J.* **2010**, *99*, 1676.
(176) Gely, S.; Lowry, D. F.; Bernard, C.; Jensen, M. R.; Blackledge, M.; Costanzo, S.; Bourhis, J. M.; Darbon, H.; Daughdrill, G.; Longhi, S. *J. Mol. Recognit.* **2010**, *23*, 435.
(177) Kavalenka, A.; Urbancic, I.; Belle, V.; Rouger, S.; Costanzo, S.; Kure, S.; Fournel, A.; Longhi, S.; Guigliarelli, B.; Strancar, J. *Biophys. J.* **2010**, *98*, 1055.
(178) Jensen, M. R.; Communie, G.; Ribeiro, E. A., Jr.; Martinez, N.; Desfosses, A.; Salmon, L.; Mollica, L.; Gabel, F.; Jamin, M.; Longhi, S.; Ruigrok, R. W.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 9839.
(179) Blocquel, D.; Habchi, J.; Gruet, A.; Blangy, S.; Longhi, S. *Mol. Biosyst.* **2012**, *8*, 392.
(180) Martinho, M.; Habchi, J.; El Habre, Z.; Nesme, L.; Guigliarelli, B.; Belle, V.; Longhi, S. *J. Biomol. Struct. Dyn.* **2013**, *31*, 453.
(181) Communie, G.; Habchi, J.; Yabukarski, F.; Blocquel, D.; Schneider, R.; Tarbouriech, N.; Papageorgiou, N.; Ruigrok, R. W.; Jamin, M.; Ringkjøbing-Jensen, M.; Longhi, S.; Blackledge, M. *PLoS Pathog.* **2013**, *9*, e1003631.
(182) Jensen, M. R.; Bernado, P.; Houben, K.; Blanchard, L.; Marion, D.; Ruigrok, R. W.; Blackledge, M. *Protein Pept. Lett.* **2010**, *17*, 952.
(183) Wang, Y.; Chu, X.; Longhi, S.; Roche, P.; Han, W.; Wang, E.; Wang, J. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E3743.
(184) Longhi, S. *Adv. Exp. Med. Biol.* **2012**, *725*, 126.
(185) Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* **2008**, *33*, 2.
(186) Habchi, J.; Longhi, S. *Mol. Biosyst.* **2012**, *8*, 69.
(187) Habchi, J.; Mamelli, L.; Longhi, S. In *Flexible Viruses: Structural Disorder in Viral Proteins*; Uversky, V. N., Longhi, S., Eds.; John Wiley and Sons: Hoboken, NJ, 2012.
(188) Laine, D.; Bourhis, J. M.; Longhi, S.; Flacher, M.; Cassard, L.; Canard, B.; Sautes-Fridman, C.; Rabourdin-Combe, C.; Valentin, H. *J. Gen. Virol.* **2005**, *86*, 1771.
(189) Laine, D.; Trescol-Biemont, M. C.; Longhi, S.; Libeau, G.; Marie, J. C.; Vidalain, P. O.; Azocar, O.; Diallo, A.; Canard, B.; Rabourdin-Combe, C.; Valentin, H. *J. Virol.* **2003**, *77*, 11332.
(190) Zhang, X.; Bourhis, J. M.; Longhi, S.; Carsillo, T.; Buccellato, M.; Morin, B.; Canard, B.; Oglesbee, M. *Virology* **2005**, *337*, 162.
(191) Longhi, S. *Measles Virus Nucleoprotein*; Nova Publishers Inc.: Hauppauge, NY, 2007.
(192) Longhi, S. *Curr. Top. Microbiol. Immunol.* **2009**, *329*, 103.
(193) Longhi, S.; Oglesbee, M. *Protein Pept. Lett.* **2010**, *17*, 961.
(194) Longhi, S. In *Negative Strand RNA Virus*; Luo, M., Ed.; World Scientific Publishing: Singapore, 2011.
(195) Iwasaki, M.; Takeda, M.; Shirogane, Y.; Nakatsu, Y.; Nakamura, T.; Yanagi, Y. *J. Virol.* **2009**, *83*, 10374.
(196) Colombo, M.; Bourhis, J. M.; Chamontin, C.; Soriano, C.; Villet, S.; Costanzo, S.; Couturier, M.; Belle, V.; Fournel, A.; Darbon, H.; Gerlier, D.; Longhi, S. *Virol. J.* **2009**, *6*, 59.
(197) Sato, H.; Masuda, M.; Miura, R.; Yoneda, M.; Kai, C. *Virology* **2006**, *352*, 121.
(198) Watanabe, A.; Yoneda, M.; Ikeda, F.; Sugai, A.; Sato, H.; Kai, C. *J. Virol.* **2011**, *85*, 2247.
(199) Hagiwara, K.; Sato, H.; Inoue, Y.; Watanabe, A.; Yoneda, M.; Ikeda, F.; Fujita, K.; Fukuda, H.; Takamura, C.; Kozuka-Hata, H.; Oyama, M.; Sugano, S.; Ohmi, S.; Kai, C. *Proteomics* **2008**, *8*, 1871.
(200) De, B. P.; Banerjee, A. K. *Microsc. Res. Technol.* **1999**, *47*, 114.
(201) Moyer, S. A.; Baker, S. C.; Horikami, S. M. *J. Gen. Virol.* **1990**, *71*, 775.
(202) Bourhis, J. M.; Canard, B.; Longhi, S. *Virology* **2006**, *344*, 94.
(203) Bourhis, J. M.; Longhi, S. In *Measles Virus Nucleoprotein*; Longhi, S., Ed.; Nova Publishers Inc.: Hauppauge, NY, 2007.
(204) Blanchard, L.; Tarbouriech, N.; Blackledge, M.; Timmins, P.; Burmeister, W. P.; Ruigrok, R. W.; Marion, D. *Virology* **2004**, *319*, 201.
(205) Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002.
(206) Llorente, M. T.; Garcia-Barreno, B.; Calero, M.; Camafeita, E.; Lopez, J. A.; Longhi, S.; Ferron, F.; Varela, P. F.; Melero, J. A. *J. Gen. Virol.* **2006**, *87*, 159.
(207) Gerard, F. C.; Ribeiro Ede, A., Jr.; Leyrat, C.; Ivanov, I.; Blondel, D.; Longhi, S.; Ruigrok, R. W.; Jamin, M. *J. Mol. Biol.* **2009**, *388*, 978.
(208) Leyrat, C.; Jensen, M. R.; Ribeiro, E. A., Jr.; Gerard, F. C.; Ruigrok, R. W.; Blackledge, M.; Jamin, M. *Protein Sci.* **2011**, *20*, 542.

dx.doi.org/10.1021/cr4005692 | *Chem. Rev.* XXXX, XXX, XXX–XXX

(209) Leyrat, C.; Schneider, R.; Ribeiro, E. A., Jr.; Yabukarski, F.; Yao, M.; Gerard, F. C.; Jensen, M. R.; Ruigrok, R. W.; Blackledge, M.; Jamin, M. *J. Mol. Biol.* **2012**, *423*, 182.
(210) Leyrat, C.; Gerard, F. C.; de Almeida Ribeiro, E., Jr.; Ivanov, I.; Ruigrok, R. W.; Jamin, M. *Protein Pept. Lett.* **2010**, *17*, 979.
(211) Ivanov, I.; Yabukarski, F.; Ruigrok, R. W.; Jamin, M. *Virus Res.* **2011**, *162*, 126.
(212) Chen, M.; Cortay, J. C.; Gerlier, D. *Virus Res.* **2003**, *98*, 123.
(213) Liston, P.; DiFlumeri, C.; Briedis, D. J. *Virus Res.* **1995**, *38*, 241.
(214) Curran, J.; Marq, J. B.; Kolakofsky, D. *J. Virol.* **1995**, *69*, 849.
(215) Curran, J.; Pelet, T.; Kolakofsky, D. *Virology* **1994**, *202*, 875.
(216) Tran, T. L.; Castagne, N.; Bhella, D.; Varela, P. F.; Bernard, J.; Chilmonczyk, S.; Berkenkamp, S.; Benhamo, V.; Grznarova, K.; Grosclaude, J.; Nespoulos, C.; Rey, F. A.; Eleouet, J. F. *J. Gen. Virol.* **2007**, *88*, 196.
(217) Kingston, R. L.; Gay, L. S.; Baase, W. S.; Matthews, B. W. *J. Mol. Biol.* **2008**, *379*, 719.
(218) Yegambaram, K.; Bulloch, E. M.; Kingston, R. L. *Protein Sci.* **2013**, *22*, 1502.
(219) Karlin, D.; Belshaw, R. *PLoS One* **2012**, *7*, e31719.
(220) Leyrat, C.; Yabukarski, F.; Tarbouriech, N.; Ribeiro, E. A., Jr.; Jensen, M. R.; Blackledge, M.; Ruigrok, R. W.; Jamin, M. *PLoS Pathog.* **2011**, *7*, e1002248.
(221) Bhella, D.; Ralph, A.; Murphy, L. B.; Yeo, R. P. *J. Gen. Virol.* **2002**, *83*, 1831.
(222) Cortese, M. S.; Uversky, V. N.; Dunker, A. K. *Prog. Biophys. Mol. Biol.* **2008**, *98*, 85.
(223) Balazs, A.; Csizmok, V.; Buday, L.; Rakacs, M.; Kiss, R.; Bokor, M.; Udupa, R.; Tompa, K.; Tompa, P. *FEBS J.* **2009**, *276*, 3744.
(224) Chang, C. K.; Hsu, Y. L.; Chang, Y. H.; Chao, F. A.; Wu, M. C.; Huang, Y. S.; Hu, C. K.; Huang, T. H. *J. Virol.* **2009**, *83*, 2255.
(225) Ambros, V.; Baltimore, D. *J. Biol. Chem.* **1978**, *253*, 5263.
(226) Olspert, A.; Arike, L.; Peil, L.; Truve, E. *FEBS Lett.* **2011**, *585*, 2979.
(227) Olspert, A.; Peil, L.; Hebrard, E.; Fargette, D.; Truve, E. *J. Gen. Virol.* **2011**, *92*, 445.
(228) Sadowy, E.; Milner, M.; Haenni, A. L. *Adv. Virus Res.* **2001**, *57*, 185.
(229) Olspert, A.; Peil, L.; Hebrard, E.; Fargette, D.; Truve, E. *J. Gen. Virol.* **2011**, *92*, 445.
(230) Daughenbaugh, K. F.; Fraser, C. S.; Hershey, J. W.; Hardy, M. E. *EMBO J.* **2003**, *22*, 2852.
(231) Goodfellow, I.; Chaudhry, Y.; Gioldasi, I.; Gerondopoulos, A.; Natoni, A.; Labrie, L.; Laliberte, J. F.; Roberts, L. *EMBO Rep.* **2005**, *6*, 968.
(232) Miyoshi, H.; Suehiro, N.; Tomoo, K.; Muto, S.; Takahashi, T.; Tsukamoto, T.; Ohmori, T.; Natsuaki, T. *Biochimie* **2006**, *88*, 329.
(233) Michon, T.; Estevez, Y.; Walter, J.; German-Retana, S.; Le Gall, O. *FEBS J.* **2006**, *273*, 1312.
(234) Tavert-Roudet, G.; Abdul-Razzak, A.; Doublet, B.; Walter, J.; Delaunay, T.; German-Retana, S.; Michon, T.; Le Gall, O.; Candresse, T. *J. Gen. Virol.* **2012**, *93*, 184.
(235) Daughenbaugh, K. F.; Wobus, C. E.; Hardy, M. E. *Virol. J.* **2006**, *3*, 33.
(236) Khan, M. A.; Miyoshi, H.; Gallie, D. R.; Goss, D. J. *J. Biol. Chem.* **2008**, *283*, 1340.
(237) Lin, L.; Shi, Y.; Luo, Z.; Lu, Y.; Zheng, H.; Yan, F.; Chen, J.; Adams, M. J.; Wu, Y. *Virus Res.* **2009**, *142*, 36.
(238) Hebrard, E.; Poulicard, N.; Gerard, C.; Traore, O.; Wu, H. C.; Albar, L.; Fargette, D.; Bessin, Y.; Vignols, F. *Mol. Plant-Microbe Interact.* **2010**, *23*, 1506.
(239) Satheshkumar, P. S.; Gayathri, P.; Prasad, K.; Savithri, H. S. *J. Biol. Chem.* **2005**, *280*, 30291.
(240) Grzela, R.; Szolajska, E.; Ebel, C.; Madern, D.; Favier, A.; Wojtal, I.; Zagorski, W.; Chroboczek, J. *J. Biol. Chem.* **2008**, *283*, 213.
(241) Rantalainen, K. I.; Uversky, V. N.; Permi, P.; Kalkkinen, N.; Dunker, A. K.; Makinen, K. *Virology* **2008**, *377*, 280.
(242) Hebrard, E.; Bessin, Y.; Michon, T.; Longhi, S.; Uversky, V. N.; Delalande, F.; Van Dorsselaer, A.; Romero, P.; Walter, J.; Declerck, N.; Fargette, D. *Virol. J.* **2009**, *6*, 23.
(243) Nair, S.; Gayathri, P.; Murthy, M. R.; Savithri, H. S. *Virology* **2008**, *382*, 83.
(244) Roy Chowdhury, S.; Savithri, H. S. *PLoS One* **2011**, *6*, e15609.
(245) Goh, G. K.; Dunker, A. K.; Uversky, V. N. *BMC Genomics* **2008**, *9* (Suppl. 2), S4.
(246) Goh, G. K.; Dunker, A. K.; Uversky, V. N. *Virol. J.* **2008**, *5*, 126.
(247) Bryant, M.; Ratner, L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 523.
(248) Gottlinger, H. G.; Sodroski, J. G.; Haseltine, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5781.
(249) Gelderblom, H. R. *AIDS* **1991**, *5*, 617.
(250) Frankel, A. D.; Young, J. A. *Annu. Rev. Biochem.* **1998**, *67*, 1.
(251) Riviere, L.; Darlix, J. L.; Cimarelli, A. *J. Virol.* **2010**, *84*, 729.
(252) Xue, B.; Mizianty, M. J.; Kurgan, L.; Uversky, V. N. *Cell. Mol. Life Sci.* **2012**, *69*, 1211.
(253) Hill, C. P.; Worthylake, D.; Bancroft, D. P.; Christensen, A. M.; Sundquist, W. I. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3099.
(254) Matthews, S.; Barlow, P.; Boyd, J.; Barton, G.; Russell, R.; Mills, H.; Cunningham, M.; Meyers, N.; Burns, N.; Clark, N.; Kingsman, S.; Kingsman, A.; Campbell, I. *Nature* **1994**, *370*, 666.
(255) Schmalzbauer, E.; Strack, B.; Dannull, J.; Guehmann, S.; Moelling, K. *J. Virol.* **1996**, *70*, 771.
(256) Poon, D. T.; Wu, J.; Aldovini, A. *J. Virol.* **1996**, *70*, 6607.
(257) Huang, Y.; Khorchid, A.; Wang, J.; Parniak, M. A.; Darlix, J. L.; Wainberg, M. A.; Kleiman, L. *J. Virol.* **1997**, *71*, 4378.
(258) Guo, J.; Henderson, L. E.; Bess, J.; Kane, B.; Levin, J. G. *J. Virol.* **1997**, *71*, 5178.
(259) Cameron, C. E.; Ghosh, M.; Le Grice, S. F.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6700.
(260) Carteau, S.; Batson, S. C.; Poljak, L.; Mouscadet, J. F.; de Rocquigny, H.; Darlix, J. L.; Roques, B. P.; Kas, E.; Auclair, C. *J. Virol.* **1997**, *71*, 6225.
(261) Summers, M. F.; Henderson, L. E.; Chance, M. R.; Bess, J. W., Jr.; South, T. L.; Blake, P. R.; Sagi, I.; Perez-Alvarado, G.; Sowder, R. C., 3rd; Hare, D. R.; L, A. *Protein Sci.* **1992**, *1*, 563.
(262) Morellet, N.; Jullian, N.; De Rocquigny, H.; Maigret, B.; Darlix, J. L.; Roques, B. P. *EMBO J.* **1992**, *11*, 3059.
(263) Darlix, J. L.; Godet, J.; Ivanyi-Nagy, R.; Fosse, P.; Mauffret, O.; Mely, Y. *J. Mol. Biol.* **2011**, *410*, 565.
(264) Skehel, J. J.; Wiley, D. C. *Cell* **1998**, *95*, 871.
(265) Skehel, J. J.; Wiley, D. C. *Annu. Rev. Biochem.* **2000**, *69*, 531.
(266) Eckert, D. M.; Kim, P. S. *Annu. Rev. Biochem.* **2001**, *70*, 777.
(267) Wiley, D. C.; Skehel, J. J. *Annu. Rev. Biochem.* **1987**, *56*, 365.
(268) Skehel, J. J.; Bayley, P. M.; Brown, E. B.; Martin, S. R.; Waterfield, M. D.; White, J. M.; Wilson, I. A.; Wiley, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 968.
(269) Stevens, J.; Corper, A. L.; Basler, C. F.; Taubenberger, J. K.; Palese, P.; Wilson, I. A. *Science* **2004**, *303*, 1866.
(270) Durrer, P.; Galli, C.; Hoenke, S.; Corti, C.; Gluck, R.; Vorherr, T.; Brunner, J. *J. Biol. Chem.* **1996**, *271*, 13417.
(271) Wilson, I. A.; Skehel, J. J.; Wiley, D. C. *Nature* **1981**, *289*, 366.
(272) Bullough, P. A.; Hughson, F. M.; Skehel, J. J.; Wiley, D. C. *Nature* **1994**, *371*, 37.
(273) Weber, T.; Paesold, G.; Galli, C.; Mischler, R.; Semenza, G.; Brunner, J. *J. Biol. Chem.* **1994**, *269*, 18353.
(274) Wharton, S. A.; Calder, L. J.; Ruigrok, R. W.; Skehel, J. J.; Steinhauer, D. A.; Wiley, D. C. *EMBO J.* **1995**, *14*, 240.
(275) Goh, G. K.; Dunker, A. K.; Uversky, V. N. *Virol. J.* **2009**, *6*, 69.
(276) Foster, T. L.; Belyaeva, T.; Stonehouse, N. J.; Pearson, A. R.; Harris, M. *J. Virol.* **2010**, *84*, 9267.
(277) de Chassey, B.; Navratil, V.; Tafforeau, L.; Hiet, M. S.; Aublin-Gex, A.; Agaugue, S.; Meiffren, G.; Pradezynski, F.; Faria, B. F.; Chantier, T.; Le Breton, M.; Pellet, J.; Davoust, N.; Mangeot, P. E.; Chaboud, A.; Penin, F.; Jacob, Y.; Vidalain, P. O.; Vidal, M.; Andre, P.; Rabourdin-Combe, C.; Lotteau, V. *Mol. Syst. Biol.* **2008**, *4*, 230.

(278) Tellinghuisen, T. L.; Marcotrigiano, J.; Rice, C. M. *Nature* **2005**, *435*, 374.
(279) Love, R. A.; Brodsky, O.; Hickey, M. J.; Wells, P. A.; Cronin, C. N. *J. Virol.* **2009**, *83*, 4395.
(280) Hanoulle, X.; Verdegem, D.; Badillo, A.; Wieruszeski, J. M.; Penin, F.; Lippens, G. *Biochem. Biophys. Res. Commun.* **2009**, *381*, 634.
(281) Hanoulle, X.; Badillo, A.; Verdegem, D.; Penin, F.; Lippens, G. *Protein Pept. Lett.* **2010**, *17*, 1012.
(282) Macdonald, A.; Crowder, K.; Street, A.; McCormick, C.; Harris, M. *J. Gen. Virol.* **2004**, *85*, 721.
(283) Macdonald, A.; Harris, M. *J. Gen. Virol.* **2004**, *85*, 2485.
(284) Feuerstein, S.; Solyom, Z.; Aladag, A.; Favier, A.; Schwarten, M.; Hoffmann, S.; Willbold, D.; Brutscher, B. *J. Mol. Biol.* **2012**, *420*, 310.
(285) Liang, Y.; Ye, H.; Kang, C. B.; Yoon, H. S. *Biochemistry* **2007**, *46*, 11550.
(286) Verdegem, D.; Badillo, A.; Wieruszeski, J. M.; Landrieu, I.; Leroy, A.; Bartenschlager, R.; Penin, F.; Lippens, G.; Hanoulle, X. *J. Biol. Chem.* **2011**, *286*, 20441.
(287) Rosnoblet, C.; Fritzinger, B.; Legrand, D.; Launay, H.; Wieruszeski, J. M.; Lippens, G.; Hanoulle, X. *J. Biol. Chem.* **2012**, *287*, 44249.
(288) Boumlic, A.; Nomine, Y.; Charbonnier, S.; Dalagiorgou, G.; Vassilaki, N.; Kieffer, B.; Trave, G.; Mavromara, P.; Orfanoudakis, G. *FEBS J.* **2010**, *277*, 774.
(289) McIntyre, M. C.; Ruesch, M. N.; Laimins, L. A. *Virology* **1996**, *215*, 73.
(290) Massimi, P.; Pim, D.; Banks, L. *J. Gen. Virol* **1997**, *78* (Part 10), 2607.
(291) Halpern, A. L.; Münger, K. *HPV Sequence Database*; Los Alamos National Laboratory: Los Alamos, NM, 1995.
(292) Dyson, N.; Howley, P. M.; Munger, K.; Harlow, E. *Science* **1989**, *243*, 934.
(293) Brehm, A.; Nielsen, S. J.; Miska, E. A.; McCance, D. J.; Reid, J. L.; Bannister, A. J.; Kouzarides, T. *EMBO J.* **1999**, *18*, 2449.
(294) Tommasino, M.; Adamczewski, J. P.; Carlotti, F.; Barth, C. F.; Manetti, R.; Contorni, M.; Cavalieri, F.; Hunt, T.; Crawford, L. *Oncogene* **1993**, *8*, 195.
(295) Jian, Y.; Schmidt-Grimminger, D. C.; Chien, W. M.; Wu, X.; Broker, T. R.; Chow, L. T. *Oncogene* **1998**, *17*, 2027.
(296) Pim, D.; Massimi, P.; Dilworth, S. M.; Banks, L. *Oncogene* **2005**, *24*, 7830.
(297) Brazil, D. P.; Hemmings, B. A. *Trends Biochem. Sci.* **2001**, *26*, 657.
(298) Alonso, L. G.; Garcia-Alai, M. M.; Nadra, A. D.; Lapena, A. N.; Almeida, F. L.; Gualfetti, P.; Prat-Gay, G. D. *Biochemistry* **2002**, *41*, 10510.
(299) Ohlenschlager, O.; Seiboth, T.; Zengerling, H.; Briese, L.; Marchanka, A.; Ramachandran, R.; Baum, M.; Korbas, M.; Meyer-Klaucke, W.; Durst, M.; Gorlach, M. *Oncogene* **2006**, *25*, 5953.
(300) Garcia-Alai, M. M.; Alonso, L. G.; de Prat-Gay, G. *Biochemistry* **2007**, *46*, 10405.
(301) Chemes, L. B.; Sanchez, I. E.; Smal, C.; de Prat-Gay, G. *FEBS J.* **2010**, *277*, 973.
(302) Chemes, L. B.; Sanchez, I. E.; Alonso, L. G.; de Prat-Gay, G. In *Flexible Viruses: Structural Disorder in Viral Proteins*; Uversky, V. N., Longhi, S., Eds.; John Wiley and Sons: Hoboken, NJ, 2012.
(303) Noval, M. G.; Gallo, M.; Perrone, S.; Salvay, A. G.; Chemes, L. B.; de Prat-Gay, G. *PLoS One* **2013**, *8*, e72760.
(304) Dantur, K.; Alonso, L.; Castano, E.; Morelli, L.; Centeno-Crowley, J. M.; Vighi, S.; de Prat-Gay, G. *Int. J. Cancer* **2009**, *125*, 1902.
(305) Calcada, E. O.; Felli, I. C.; Hosek, T.; Pierattelli, R. *ChemBioChem* **2013**, *14*, 1876.
(306) Garcia-Alai, M. M.; Gallo, M.; Salame, M.; Wetzler, D. E.; McBride, A. A.; Paci, M.; Cicero, D. O.; de Prat-Gay, G. *Structure* **2006**, *14*, 309.
(307) Scheffner, M.; Werness, B. A.; Huibregtse, J. M.; Levine, A. J.; Howley, P. M. *Cell* **1990**, *63*, 1129.
(308) Scheffner, M.; Huibregtse, J. M.; Vierstra, R. D.; Howley, P. M. *Cell* **1993**, *75*, 495.
(309) Ronco, L. V.; Karpova, A. Y.; Vidal, M.; Howley, P. M. *Genes Dev.* **1998**, *12*, 2061.
(310) Sedman, S. A.; Barbosa, M. S.; Vass, W. C.; Hubbert, N. L.; Haas, J. A.; Lowy, D. R.; Schiller, J. T. *J. Virol.* **1991**, *65*, 4860.
(311) Morosov, A.; Phelps, W. C.; Raychaudhuri, P. *J. Biol. Chem.* **1994**, *269*, 18434.
(312) Dey, A.; Atcha, I. A.; Bagchi, S. *Virology* **1997**, *228*, 190.
(313) Gewin, L.; Galloway, D. A. *J. Virol.* **2001**, *75*, 7198.
(314) Oh, S. T.; Kyo, S.; Laimins, L. A. *J. Virol.* **2001**, *75*, 5559.
(315) Li, X.; Coffino, P. *J. Virol.* **1996**, *70*, 4509.
(316) Neveu, G.; Cassonnet, P.; Vidalain, P. O.; Rolloy, C.; Mendoza, J.; Jones, L.; Tangy, F.; Muller, M.; Demeret, C.; Tafforeau, L.; Lotteau, V.; Rabourdin-Combe, C.; Trave, G.; Dricot, A.; Hill, D. E.; Vidal, M.; Favre, M.; Jacob, Y. *Methods* **2012**, *58*, 349.
(317) Nomine, Y.; Charbonnier, S.; Ristriani, T.; Stier, G.; Masson, M.; Cavusoglu, N.; Van Dorsselaer, A.; Weiss, E.; Kieffer, B.; Trave, G. *Biochemistry* **2003**, *42*, 4909.
(318) Garcia-Alai, M. M.; Dantur, K. I.; Smal, C.; Pietrasanta, L.; de Prat-Gay, G. *Biochemistry* **2007**, *46*, 341.
(319) Zanier, K.; ould M'hamed ould Sidi, A.; Boulade-Ladame, C.; Rybin, V.; Chappelle, A.; Atkinson, A.; Kieffer, B.; Trave, G. *Structure* **2012**, *20*, 604.
(320) Uversky, V. N.; Roman, A.; Oldfield, C. J.; Dunker, A. K. *J. Proteome Res.* **2006**, *5*, 1829.
(321) Heer, A.; Alonso, L. G.; de Prat-Gay, G. *Biochemistry* **2011**, *50*, 1376.
(322) Campbell, G. R.; Loret, E. P. *Retrovirology* **2009**, *6*, 50.
(323) Vendel, A. C.; Lumb, K. J. *Biochemistry* **2003**, *42*, 910.
(324) Blanco, F. J.; Hess, S.; Pannell, L. K.; Rizzo, N. W.; Tycko, R. *J. Mol. Biol.* **2001**, *313*, 845.
(325) Casu, F.; Duggan, B. M.; Hennig, M. *Biophys. J.* **2013**, *105*, 1004.
(326) Surendran, R.; Herman, P.; Cheng, Z.; Daly, T. J.; Ching Lee, J. *Biophys. Chem.* **2004**, *108*, 101.
(327) Morellet, N.; Roques, B. P.; Bouaziz, S. *Curr. HIV Res.* **2009**, *7*, 184.
(328) Henklein, P.; Bruns, K.; Sherman, M. P.; Tessmer, U.; Licha, K.; Kopp, J.; de Noronha, C. M.; Greene, W. C.; Wray, V.; Schubert, U. *J. Biol. Chem.* **2000**, *275*, 32016.
(329) Bruns, K.; Fossen, T.; Wray, V.; Henklein, P.; Tessmer, U.; Schubert, U. *J. Biol. Chem.* **2003**, *278*, 43188.
(330) Reingewertz, T. H.; Benyamini, H.; Lebendiker, M.; Shalev, D. E.; Friedler, A. *Protein Eng., Des. Sel.* **2009**, *22*, 281.
(331) Gramberg, T.; Sunseri, N.; Landau, N. R. *Curr. HIV/AIDS Rep.* **2009**, *6*, 36.
(332) Wray, V.; Federau, T.; Henklein, P.; Klabunde, S.; Kunert, O.; Schomburg, D.; Schubert, U. *Int. J. Pept. Protein Res.* **1995**, *45*, 35.
(333) Leavitt, S. A.; Schon, A.; Klein, J. C.; Manjappara, U.; Chaiken, I. M.; Freire, E. *Curr. Protein Pept. Sci.* **2004**, *5*, 1.
(334) Roberts, J. W. *Nature* **1969**, *224*, 1168.
(335) Schweimer, K.; Rosch, P. In *Flexible Viruses: Structural Disorder in Viral Proteins*; Uversky, V. N., Longhi, S., Eds.; John Wiley and Sons: Hoboken, NJ, 2012.
(336) Greenblatt, J.; Li, J. *J. Biol. Chem.* **1982**, *257*, 362.
(337) Van Gilst, M. R.; von Hippel, P. H. *J. Mol. Biol.* **1997**, *274*, 160.
(338) Van Gilst, M. R.; Rees, W. A.; Das, A.; von Hippel, P. H. *Biochemistry* **1997**, *36*, 1514.
(339) Johansen, D.; Trewhella, J.; Goldenberg, D. P. *Protein Sci.* **2011**, *20*, 1955.
(340) Mogridge, J.; Legault, P.; Li, J.; Van Oene, M. D.; Kay, L. E.; Greenblatt, J. *Mol. Cell* **1998**, *1*, 265.
(341) Patel, D. J. *Curr. Opin. Struct. Biol.* **1999**, *9*, 74.
(342) Makarov, V.; Taliansky, M.; Dobrov, E.; Kalinina, N. In *Flexible Viruses: Structural Disorder in Viral Proteins*; Uversky, V. N., Longhi, S., Eds.; John Wiley and Sons: Hoboken, NJ, 2012.
(343) Morozov, S. Y.; Solovyev, A. G. *J. Gen. Virol.* **2003**, *84*, 1351.

(344) Caruthers, J. M.; McKay, D. B. *Curr. Opin. Struct. Biol.* **2002**, *12*, 123.
(345) Bleykasten, C.; Gilmer, D.; Guilley, H.; Richards, K. E.; Jonard, G. *J. Gen. Virol* **1996**, *77* (Part 5), 889.
(346) Donald, R. G.; Lawrence, D. M.; Jackson, A. O. *J. Virol.* **1997**, *71*, 1538.
(347) Kalinina, N. O.; Rakitina, D. A.; Yelina, N. E.; Zamyatnin, A. A., Jr.; Stroganova, T. A.; Klinov, D. V.; Prokhorov, V. V.; Ustinova, S. V.; Chernov, B. K.; Schiemann, J.; Solovyev, A. G.; Morozov, S. Y. *J. Gen. Virol.* **2001**, *82*, 2569.
(348) Kalinina, N. O.; Rakitina, D. V.; Solovyev, A. G.; Schiemann, J.; Morozov, S. Y. *Virology* **2002**, *296*, 321.
(349) Leshchiner, A. D.; Solovyev, A. G.; Morozov, S. Y.; Kalinina, N. O. *J. Gen. Virol.* **2006**, *87*, 3087.
(350) Makarov, V. V.; Rybakova, E. N.; Efimov, A. V.; Dobrov, E. N.; Serebryakova, M. V.; Solovyev, A. G.; Yaminsky, I. V.; Taliansky, M. E.; Morozov, S. Y.; Kalinina, N. O. *J. Gen. Virol.* **2009**, *90*, 3022.
(351) Semashko, M. A.; Gonzalez, I.; Shaw, J.; Leonova, O. G.; Popenko, V. I.; Taliansky, M. E.; Canto, T.; Kalinina, N. O. *Biochimie* **2012**, *94*, 1180.
(352) Lommer, B. S.; Luo, M. *J. Biol. Chem.* **2002**, *277*, 7108.
(353) Taylor, J. M. *Adv. Virus Res.* **2009**, *74*, 103.
(354) Rizzetto, M. *J. Hepatol.* **2009**, *50*, 1043.
(355) Han, Z.; Alves, C.; Gudima, S.; Taylor, J. *J. Virol.* **2009**, *83*, 6457.
(356) Ryu, W. S.; Netter, H. J.; Bayer, M.; Taylor, J. *J. Virol.* **1993**, *67*, 3281.
(357) Alves, C.; Cheng, H.; Roder, H.; Taylor, J. *Virology* **2010**, *407*, 333.
(358) Greco-Stewart, V.; Pelchat, M. *Viruses* **2010**, *2*, 189.
(359) Casaca, A.; Fardilha, M.; da Cruz e Silva, E.; Cunha, C. *Virol. J.* **2011**, *8*, 358.
(360) Sieber, T.; Dobner, T. *J. Virol.* **2007**, *81*, 95.
(361) Sieber, T.; Scholz, R.; Spoerner, M.; Schumann, F.; Kalbitzer, H. R.; Dobner, T. *Virology* **2011**, *418*, 133.
(362) Anderson, C. W.; Schmitt, R. C.; Smart, J. E.; Lewis, J. B. *J. Virol.* **1984**, *50*, 387.
(363) Lewis, J. B.; Anderson, C. W. *J. Virol.* **1987**, *61*, 3879.
(364) Koralnik, I. J.; Gessain, A.; Klotman, M. E.; Lo Monico, A.; Berneman, Z. N.; Franchini, G. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8813.
(365) Ciminale, V.; Pavlakis, G. N.; Derse, D.; Cunningham, C. P.; Felber, B. K. *J. Virol.* **1992**, *66*, 1737.
(366) Larocca, D.; Chao, L. A.; Seto, M. H.; Brunck, T. K. *Biochem. Biophys. Res. Commun.* **1989**, *163*, 1006.
(367) Matsuoka, M.; Green, P. L. *Retrovirology* **2009**, *6*, 71.
(368) Kashanchi, F.; Brady, J. N. *Oncogene* **2005**, *24*, 5938.
(369) Baydoun, H. H.; Bellon, M.; Nicot, C. *AIDS Rev.* **2008**, *10*, 195.
(370) Cann, A. J.; Rosenblatt, J. D.; Wachsman, W.; Shah, N. P.; Chen, I. S. *Nature* **1985**, *318*, 571.
(371) Derse, D. *J. Virol.* **1987**, *61*, 2462.
(372) Felber, B. K.; Paskalis, H.; Kleinman-Ewing, C.; Wong-Staal, F.; Pavlakis, G. N. *Science* **1985**, *229*, 675.
(373) Seiki, M.; Inoue, J.; Takeda, T.; Yoshida, M. *EMBO J.* **1986**, *5*, 561.
(374) Sodroski, J.; Trus, M.; Perkins, D.; Patarca, R.; Wong-Staal, F.; Gelmann, E.; Gallo, R.; Haseltine, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 4617.
(375) Boxus, M.; Twizere, J. C.; Legros, S.; Dewulf, J. F.; Kettmann, R.; Willems, L. *Retrovirology* **2008**, *5*, 76.
(376) Derse, D. *J. Virol.* **1988**, *62*, 1115.
(377) Inoue, J.; Seiki, M.; Yoshida, M. *FEBS Lett.* **1986**, *209*, 187.
(378) Inoue, J.; Yoshida, M.; Seiki, M. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3653.
(379) Kiyokawa, T.; Seiki, M.; Iwashita, S.; Imagawa, K.; Shimizu, F.; Yoshida, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 8359.
(380) Seiki, M.; Inoue, J.; Hidaka, M.; Yoshida, M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7124.
(381) Seiki, M.; Hattori, S.; Hirayama, Y.; Yoshida, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3618.
(382) Nagashima, K.; Yoshida, M.; Seiki, M. *J. Virol.* **1986**, *60*, 394.
(383) Rancurel, C.; Khosravi, M.; Dunker, A. K.; Romero, P. R.; Karlin, D. *J. Virol.* **2009**, *83*, 10719.
(384) Pavesi, A.; Magiorkinis, G.; Karlin, D. G. *PLoS Comput. Biol.* **2013**, *9*, e1003162.
(385) Kovacs, E.; Tompa, P.; Liliom, K.; Kalmar, L. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5429.
(386) Cann, A. J. *Principles of Molecular Virology*; Elsevier: Amsterdam, 2005; Vol. 4.
(387) *Retroviruses*; Coffin, J. M., Hughes, S. H., Varmus, H. E., Eds.; Cold Spring Harbor Laboratory Press: New York, 1997.
(388) Watts, J. M.; Dang, K. K.; Gorelick, R. J.; Leonard, C. W.; Bess, J. W., Jr.; Swanstrom, R.; Burch, C. L.; Weeks, K. M. *Nature* **2009**, *460*, 711.
(389) de Laureto, P. P.; Tosatto, L.; Frare, E.; Marin, O.; Uversky, V. N.; Fontana, A. *Biochemistry* **2006**, *45*, 11523.
(390) Fontana, A.; Fassina, G.; Vita, C.; Dalzoppo, D.; Zamai, M.; Zambonin, M. *Biochemistry* **1986**, *25*, 1847.
(391) Fontana, A.; Zambonin, M.; Polverino de Laureto, P.; De Filippis, V.; Clementi, A.; Scaramella, E. *J. Mol. Biol.* **1997**, *266*, 223.
(392) Polverino de Laureto, P.; De Filippis, V.; Di Bello, M.; Zambonin, M.; Fontana, A. *Biochemistry* **1995**, *34*, 12596.
(393) Iakoucheva, L. M.; Kimzey, A. L.; Masselon, C. D.; Bruce, J. E.; Garner, E. C.; Brown, C. J.; Dunker, A. K.; Smith, R. D.; Ackerman, E. J. *Protein Sci.* **2001**, *10*, 560.
(394) Romero, P. R.; Zaidi, S.; Fang, Y. Y.; Uversky, V. N.; Radivojac, P.; Oldfield, C. J.; Cortese, M. S.; Sickmeier, M.; LeGall, T.; Obradovic, Z.; Dunker, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8390.
(395) Lacroix, E.; Viguera, A. R.; Serrano, L. *Folding Des.* **1998**, *3*, 79.
(396) Jordan, I. K.; Sutter, B. A.; McClure, M. A. *Mol. Biol. Evol.* **2000**, *17*, 75.
(397) Narechania, A.; Terai, M.; Burk, R. D. *J. Gen. Virol.* **2005**, *86*, 1307.
(398) Longhi, S. *Protein Pept. Lett.* **2010**, *17*, 930.
(399) Xue, B.; Williams, R. W.; Oldfield, C. J.; Goh, G. K.; Dunker, A. K.; Uversky, V. N. In *Flexible Viruses: Structural Disorder in Viral Proteins*; Uversky, V. N., Longhi, S., Eds.; John Wiley and Sons: Hoboken, NJ, 2012.
(400) Ivanov, I.; Crepin, T.; Jamin, M.; Ruigrok, R. W. *J. Virol.* **2010**, *84*, 3707.
(401) Ding, H.; Green, T. J.; Lu, S.; Luo, M. *J. Virol.* **2006**, *80*, 2808.
(402) Tarbouriech, N.; Curran, J.; Ruigrok, R. W.; Burmeister, W. P. *Nat. Struct. Biol.* **2000**, *7*, 777.
(403) Blocquel, D. H. J.; Durand, E.; Sevajol, M.; Ferron, F.; Erales, J.; Papageorgiou, N.; Longhi, S. *Acta Crystallogr., D* **2014**, *70*, 10.1107/S139900471400234X.
(404) Blocquel, D.; Beltrandi, M.; Erales, J.; Barbier, P.; Longhi, S. *Virology* **2013**, *446*, 162.
(405) Llorente, M. T.; Taylor, I. A.; Lopez-Vinas, E.; Gomez-Puertas, P.; Calder, L. J.; Garcia-Barreno, B.; Melero, J. A. *Proteins* **2008**, *72*, 946.
(406) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J. R.; Jensen, M. R.; Segard, S.; Bernado, P.; Charavay, C.; Blackledge, M. *Bioinformatics* **2012**, *28*, 1463.
(407) DeLano, W. L. *Proteins* **2002**, *30*, 442.