1

DO VIRAL PROTEINS POSSESS UNIQUE FEATURES?

BIN XUE, ROBERT W. WILLIAMS, CHRISTOPHER J. OLDFIELD, GERARD K.-M. GOH, A. KEITH DUNKER, AND VLADIMIR N. UVERSKY

1.1 INTRODUCTION

Many proteins (or protein regions) are intrinsically disordered. They lack unique 3D structures in their native, functional states under physiological conditions *in vitro* (Wright and Dyson, 1999; Uversky et al., 2000; Dunker et al., 2001, 2002a,b; Tompa, 2002, 2003; Uversky, 2002a,b, 2003; Minezaki et al., 2006). The major functions of such proteins and regions are signaling, recognition, and regulation activities (Wright and Dyson, 1999, 2009; Dunker et al., 2002a,b; 2005; 2008a,b; Dyson and Wright, 2005; Uversky et al., 2005; Radivojac et al., 2007; Dunker and Uversky, 2008; Oldfield et al., 2008; Tompa et al., 2009). Owing to these crucial functional roles, intrinsically disordered proteins (IDPs) are highly abundant in all species. According to computational predictions, typically 7–30% prokaryotic proteins contain long disordered regions of more than 30 consecutive residues, whereas in eukaryotes the amount of such proteins reaches 45–50% (Romero et al., 1997, 2001; Dunker et al., 2001; Ward et al., 2004; Oldfield et al., 2005a,b; Feng et al., 2006). Furthermore, almost 70% of proteins in the PDB (which is biased to structured proteins) have intrinsically disordered regions (IDRs), which are indicated by missing electron density (Obradovic et al., 2003). Numerous disordered proteins have been shown to be associated with cancer (Iakoucheva et al., 2002), cardiovascular disease (Cheng et al., 2006), amyloidoses (Uversky, 2008a), neurodegenerative diseases (Uversky, 2008b), diabetes, and other human diseases (Uversky et al., 2008), an

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

observation that was used to introduce the "disorder in disorders" or D^2 concept (Uversky et al., 2008).

Recently, we showed also that IDPs are abundant in the human diseasome (Midic et al., 2009), a framework that systematically linked the human disease phenome (which includes all the human genetic diseases) with the human disease genome (which contains all the disease-related genes) (Goh et al., 2007). This framework was constructed from the analysis of two networks, a network of genetic diseases, the "human disease network," where two diseases are directly linked if there is a gene that is directly related to both of them, and a network of disease genes, the "disease gene network," where two genes are directly linked if there is a disease to which they are both directly related (Goh et al., 2007). Our analysis revealed that there were noticeable differences in the abundance of intrinsic disorder in human disease-related as compared to disease-unrelated proteins (Midic et al., 2009). Furthermore, various disease classes were significantly different with respect to the content of disordered proteins.

Furthermore, we have shown that intrinsic disorder is highly abundant in proteins of the parasitic protozoa (Mohan et al., 2008). Since viruses are common infectious pathogens, here, we summarize some literature data on the abundance of intrinsic disorder in viruses and explore the functional roles of intrinsic disorder in these intriguing "organisms at the edge of life."

Viruses are the most abundant living entities (Breitbart and Rohwer, 2005). For example, 1 mL of natural water contains up to 2.5×10^8 viral particles (Bergh et al., 1989), and the total number of viral particles exceeds the number of cells by at least an order of magnitude (Sano et al., 2004; Edwards and Rohwer, 2005). They are common parasitic organisms that live in the infected cells of Eukarya, Archaea, and Bacteria (or even inside other viruses) and produce virions to disseminate their genes (Breitbart and Rohwer, 2005; Edwards and Rohwer, 2005; Lawrence et al., 2009). Viruses do not have a defined cellular structure and are structurally very simple consisting of two or three parts. This includes two common components found in all viruses, DNA- (double-stranded or single-stranded) or RNA-based genes, and a protein coat protecting the genetic material (this proteinaceous coat is known as the *capsid*), and a lipid-based envelope surrounding some of the viruses when they are outside the host cells. In addition to the capsid proteins, some complex viruses also contain the so-called nonstructural proteins that assist in the construction of their capsid and viral regulatory and accessory proteins. Furthermore, enveloped viruses contain several integral membrane proteins, and matrix proteins forming the so-called matrix, another biologically active proteinaceous coat located right beneath the envelope.

Historically, there is no uniform opinion on whether the viruses are a form of life or just simple nonliving organic structures that interact with living organisms, or are yet the "organisms at the edge of life" (Rybicki, 1990). This difference in opinion originates from the facts that although viruses possess genes, evolve by natural selection, and reproduce by creating multiple copies of themselves through self-assembly, they do not have a defined cellular structure, as well as they lack

their own metabolism, require a host cell to make new products, and therefore cannot reproduce outside the host cell (Holmes, 2007).

In the evolutionary history of life, the origin of viruses is unclear. Currently, there are three major hypotheses for virus origin (Forterre, 2006):

1. Coevolution or the virus first hypothesis (here, viruses appeared simultaneously with the cells early in the history of earth and since that time are dependent on cellular life for many millions of years);
2. Cellular origin or vagrancy hypothesis (here, viruses evolved from pieces of pieces of RNA or DNA (e.g., plasmids, pieces of naked DNA that can move between, or transposons, pieces of DNA that replicate and move around to different positions within the genes) that "escaped" from the genes of a larger organism);
3. Regressive or degeneracy hypothesis (here, viruses originally were small cells that parasitized larger cells and that, with time, lost all the genes unused because of their parasitism).

It is suggested that RNA viruses may have originated in the nucleoprotein world (which followed the RNA world) by escaping or reduction from the primordial RNA-containing cells, whereas DNA viruses (at least some of them) might have evolved directly from RNA viruses (Forterre, 2006). Irrespective of the virus origin hypothesis, the facts that viruses infect cells from the three domains of life, Archaea, Bacteria, and Eukarya, share homologous features, and have probably existed since living cells first evolved (Iyer et al., 2006), clearly suggest that viruses originated very early in the evolution of life (Koonin et al., 2006). This antiquity of viruses can explain why most viral proteins have no homologs in cellular organisms or have only distantly related ones (Forterre, 2006).

Importantly, viruses are suggested to play a number of crucial roles in the general evolution of life. For example, they are responsible for the so-called horizontal gene transfer, a process by which an organism incorporates genetic material from another organism without being the offspring of that organism, which increases genetic diversity (Canchaya et al., 2003). In fact, 3–8% of the human genome is suggested to be composed of fragments of viral DNA. Furthermore, since it is believed that some DNA replication proteins originated in the virosphere and were later transferred to cellular organisms, viruses could play a vital role in the invention of DNA and DNA replication mechanisms and therefore could serve as crucial drives of the origin of the eukaryotic nucleus, and even of the formation of the three domains of life (Forterre, 2006).

Since viruses are believed to play a major role in the evolution of life, and since they are very different from all other life forms on earth, recently, a division was proposed to biological entities into two groups of organisms, namely, ribosome-encoding organisms, which include eukaryotic, archaeal, and bacterial organisms, and capsid-encoding organisms, which include viruses (Raoult and Forterre, 2008). Therefore, viruses are defined now as capsid-encoding organisms, which contain proteins and nucleic acids, self-assemble in the nucleocapsids, and use a ribosome-encoding organism for the completion of their life cycle (Raoult and Forterre, 2008).

This chapter illustrates some structural peculiarities of viral proteins and discusses the role of intrinsic disorder in their functions.

1.2 CLASSIFICATION AND FUNCTIONS OF VIRAL PROTEINS

Viral genomes are typically rather small ranging in size from 6 to 8 proteins (e.g., human papilloma virus (HPV)) to ∼1000 proteins (e.g., *Acanthamoeba polyphaga* mimivirus (APMV)). Functionally, viral proteins are grouped into structural, nonstructural (NS), regulatory, and accessory proteins. For example, there are eight major proteins encoded by HPV. Proteins E1 and E2 are involved in viral replication as well as in the regulation of early transcription. E1 binds to the origin of replication and exhibits ATPase as well as helicase activity (Ustav and Stenlund, 1991; Hughes and Romanos, 1993), whereas E2 forms a complex with E1, facilitating its binding to the origin of viral replication (Mohr et al., 1990; Ustav and Stenlund, 1991; Frattini and Laimins, 1994). Furthermore, E2 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) (Cripe et al., 1987; Gloss et al., 1987). E4 is the most highly expressed protein in the productive life cycle of HPVs, and it plays a number of important roles in promoting the differentiation-dependent productive phase of the viral life cycle (Wilson et al., 2005; Brown et al., 2006; Davy et al., 2006). The E5 protein has weak transforming capabilities *in vitro* (Leechanachai et al., 1992; Straight et al., 1993), supports HPV late functions (Fehrmann et al., 2003; Genther et al., 2003), and disrupts MHC class II maturation (Zhang et al., 2003). Finally, L1 and L2 are the major and the minor capsid proteins, respectively.

Two early proteins (E6 and E7 oncoproteins) are mainly responsible for HPV-mediated malignant cell progression, leading ultimately to an invasive carcinoma. Proteins E6 and E7 function as oncoproteins in high risk HPVs, at least in part, by targeting the cell cycle regulators p53 and Rb, respectively. E7 has been shown to be involved in cellular processes such as cell growth and transformation (McIntyre et al., 1996), gene transcription (Massimi et al., 1997), apoptosis, and DNA synthesis, among other processes (Halpern and Münger, 1995). It interacts with many important proteins including the Rb tumor suppressor and its family members, p107 and p130 (Dyson et al., 1989), glycolytic enzymes (Zwerschke et al., 1999; Mazurek et al., 2001), histone deacetylase (Brehm et al., 1999), kinase p33CDK2, and cyclin A (Tommasino et al., 1993), as well as the cyclin-dependent kinase inhibitor $p21^{cip1}$ protein (Jian et al., 1998). Furthermore, it has been shown that E7 also binds to a protein phosphatase 2A (PP2A) (Pim et al., 2005). Formation of this complex sequesters PP2A, inhibiting its interaction with protein kinase B (PKB) or Akt (which is one of the several second messenger kinases that are activated by cell attachment and growth factor signaling and that transmit signals to the cell nucleus to inhibit apoptosis and thereby increase cell survival during proliferation (Brazil and Hemmings, 2001)), thereby maintaining PKB/Akt signaling by inhibiting its dephosphorylation.

E6 primarily promotes tumorigenesis by stimulating cellular degradation of the tumor suppressor p53 via formation of a trimeric complex comprising E6, p53, and the cellular ubiquitination enzyme E6AP (Scheffner et al., 1990, 1993). Besides this crucial role in the regulation of p53 degradation, E6 displays numerous activities unrelated to p53. These include but are not limited to recognition of a variety of other cellular proteins: transcription coactivators p300/CBP (Patel et al., 1999; Zimmermann et al., 1999) and ADA3 (Kumar et al., 2002), transcription factors c-Myc (Gross-Mesilaty et al., 1998) and IRF3 (Ronco et al., 1998), replication protein hMCM7 (Kukimoto et al., 1998), DNA repair proteins MGMT (Srivenugopal and Ali-Osman, 2002), protein kinases PKN (Gao et al., 2000) and Tyk2 (Li et al., 1999), Rap-GTPase activating protein E6TP1 (Gao et al., 1999), tumor necrosis factor receptor TNF-R1 (Filippova et al., 2002), apoptotic protein Bak (Thomas and Banks, 1999), clathrin-adaptor complex AP-1 (Tong et al., 1998), focal adhesion component paxillin (Tong and Howley, 1997), calcium-binding proteins E6BP (Chen et al., 1995) and fibulin-1 (Du et al., 2002), and several members of the PDZ protein family, including hDLG (Kiyono et al., 1997), hScrib (Nakagawa and Huibregtse, 2000), MAGI-1 (Glaunsinger et al., 2000), and MUPP1 (Lee et al., 2000). Furthermore, E6 activates or represses several cellular or viral transcription promoters (Sedman et al., 1991; Morosov et al., 1994; Dey et al., 1997; Ronco et al., 1998), such as transcriptional activation of the gene encoding the retrotranscriptase of human telomerase (Gewin and Galloway, 2001; Oh et al., 2001). In addition, it has been recently established that E6 recognizes four-way DNA junctions (Ristriani et al., 2000, 2001). The function of the low risk HPV E6 is less well studied. However, the low risk E6 lacks a number of activities that correlate with the oncogenic activity of the high risk HPV E6. For example, low risk E6 neither binds PDZ proteins (Kiyono et al., 1997) or E6TP1 (Gao et al., 1999) nor targets p53 for degradation (Scheffner et al., 1990; Li and Coffino, 1996). Like the high risk E6, low risk E6 does bind MCM7 (Kukimoto et al., 1998) and Bak (Thomas and Banks, 1999) and inhibits p300 acetylation of p53 (Thomas and Chiang, 2005).

1.2.1 Structural Proteins Form the Viral Capsid and Envelope

1.2.1.1 Capsid The capsid is the proteinaceous shell of the virus, which consists of several protomers (also known as *capsomers*), oligomeric protein subunits. Often, capsid proteins are conjugated with DNA or RNA, forming the viral nucleoprotein complex. It is important to remember that such viral nucleoproteins are multifunctional, being able to interact with nucleic acid and other proteins. For example, the transcription and replication of the measles virus, the RNA genome of which is encapsidated by the nucleoprotein (N), are initiated by the RNA-dependent RNA polymerase binding to the nucleocapsid via the phosphoprotein (P) (Longhi, 2009).

The packing of capsomers defines the shape of a viral capsid, which can be helical, icosahedral, or complex. Capsids of the *helical* or *filamentous* viruses are highly ordered helical structures consisting of a single type of capsomer stacked

around a central axis. The genetic material of these viruses, single-stranded RNA or, in some cases, single-stranded DNA is located inside a central cavity of the capsid, where it is bound to the capsid proteins via the electrostatic interactions between negative charges on nucleic acid and positive charges on the protein. The length of a helical capsid is dependent on the length of the viral nucleic acid, whereas its diameter is determined by the size and arrangement of capsomers. These rod-shaped or filamentous viruses can be short and highly rigid, or long and very flexible. Illustrative examples of the filamentous or helical viruses are tobacco mosaic virus (TMV), *Sulfolobus islandicus* filamentous virus (SIFV), *Acidianus* filamentous virus 1 (AFV1), filamentous bacteriophage fd, and others.

Capsids of the majority of viruses are *icosahedral* or near-spherical with icosahedral symmetry. A regular icosahedron is the optimal way to pack identical subunits to form a closed shell (Fig. 1.1a). Since there are 20 identical equilateral triangular faces in an icosahedron, the minimal number of identical subunits to form such a structure is 60 (Fig. 1.1b). Here, each triangular face is made up of three identical subunits. The capsomer of the icosahedral virus includes the five identical subunits that surround each vertex and are arranged in a fivefold symmetry (Fig. 1.1c).

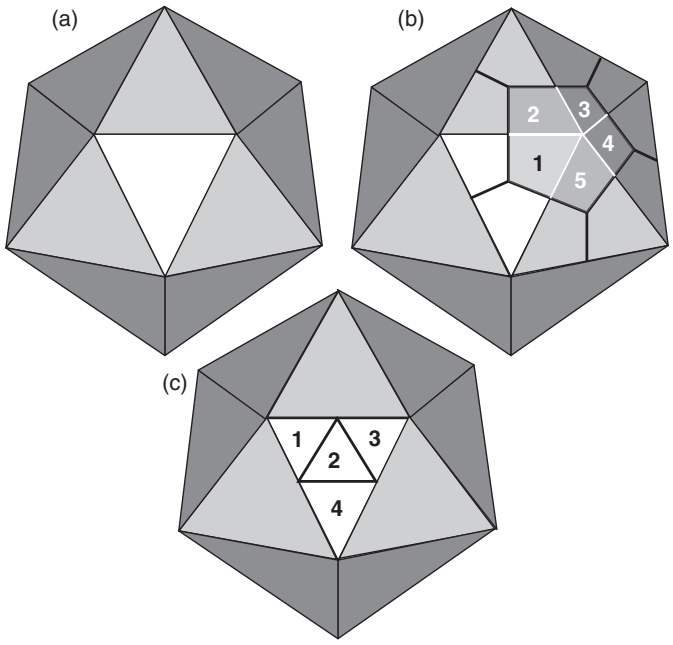

Figure 1.1 Icosahedron and virus capsid. (a) An icosahedron has 20 identical equilateral triangular faces. (b) In most icosahedral capsids of viruses, each triangular face is made up of three identical subunits. As a result, a typical viral capsid contains 60 subunits. The five subunits surrounding each vertex are arranged in a fivefold symmetry. (c) An icosahedral capsid of large viruses can consist of more than 60 subunits. Some of the triangular faces are made up of four subunits.

Therefore, there are typically 12 capsomers in the icosahedral virus. Many viruses have more than 60 subunits. In these viruses with large icosahedral capsids, the triangular faces are made up of four subunits (Fig. 1.1c).

Capsids of several viruses are neither purely helical nor purely icosahedral. These *complex* capsids may include extra structures, such as protein tails or complex outer walls. An illustrative example of such a complex virus is the well-known bacteriophage T4, which has an icosahedral head bound to a helical tail, which may have a hexagonal base plate with protruding protein tail fibers. This peculiar tail structure helps T4 to attach to the bacterial host and acts as a molecular syringe injecting the viral genome into the cell (Rossmann et al., 2004).

1.2.1.2 Viral Envelope In some viruses, the capsid is coated with a lipid membrane, known as the *viral envelope*, which is acquired by the capsid from an intracellular membrane of the virus host. Typically, in addition to the lipid membrane derived from a host, viral envelopes contain viral glycoproteins (e.g., hemagglutinin (HA), neuraminidase, and M2 protein, a proton-selective ion channel in influenza virus, or gp160 protein in human immunodeficiency virus (HIV), which consists of the structural subunit gp120, and the transmembrane subunit gp41). Some of these surface viral glycoproteins (HA, neuraminidase, and gp120) protrude from the viral lipid bilayer and play important roles in its attachment to and penetration into the target cells (Suzuki, 2005). Other viral envelope proteins are involved in various functions related to the virus life cycle. For example, a proton-selective ion channel M2 protein of influenza A virus enables hydrogen ions to enter the viral particle from the endosome, thus lowering pH of the inside of the virus. This decrease in pH triggers the dissociation of the viral matrix protein M1 from the ribonucleoprotein, therefore initiating the uncoating of the virus and exposing its content to the cytoplasm of the host cell (Cady et al., 2009).

1.2.1.3 Matrix In addition to membrane glycoproteins, enveloped viruses have matrix proteins, which link the viral envelope with the virus core. In general, viral matrix proteins are responsible for expelling the genetic material after a virus has entered a cell. However, they have several other biological functions. For example, in the influenza virus, one side of the matrix M1 protein possesses a specific affinity to the glycoproteins of the host cell membrane, whereas another side of this protein has nonspecific affinity for the viral RNA. As a result, a specific proteinaceous layer, or matrix, is formed under the membrane. The assembled complexes of viral ribonucleoprotein and viral RNA bind to the matrix and are enveloped and bud out of the cell as new mature viruses (Nayak et al., 2004, 2009). M1 protein also has multiple regulatory functions performed by interaction with the components of the host cell. These regulatory functions include a role in the export of the viral ribonucleoproteins from the host cell nucleus, inhibition of viral transcription, and a role in virus assembly and budding (Nayak et al., 2004, 2009).

1.2.2 Viral Nonstructural Proteins

Viral NS proteins are virus-encoded proteins that are not a part of the viral particle. Some of these proteins may play roles within the infected cell during virus replication, whereas others act in the regulation of virus replication or virus assembly. Specific functions of six NS HPV proteins were briefly introduced above. Below, three illustrative examples of the functions of viral NS proteins, namely, replicon formation, immunomodulation, and transactivation of genes encoding structural proteins are described.

1.2.2.1 Replicon Formation The hepatitis C virus's (HCV's) RNA replication complex formation requires interactions between the HCV NS proteins and a human cellular vesicle membrane transport protein hVAP-33 (Gao et al., 2004). The formation of this HCV replicon is initiated by the precursor of NS4B, which is able to anchor to the lipid raft membrane. Most of the other HCV NS proteins, including NS5A, NS5B, and NS3, are also localized to these lipid raft membranes, suggesting that protein–protein interactions among the various HCV NS proteins and hVAP-33 are important for the formation of HCV replication complex (Gao et al., 2004).

1.2.2.2 Immunomodulation The immunomodulatory function of West Nile virus NS protein NS1 was demonstrated by showing that the soluble and cell-surface-associated NS1 was able to bind to and recruit the complement regulatory protein factor H. This interaction led to decreased complement activation, minimizing immune system targeting of West Nile virus by decreasing complement recognition of infected cells (Chung et al., 2006). In rinderpest virus, the viral NS C protein was shown to block specifically the actions of type 1 and type 2 interferons, therefore suppressing the induction of the innate immune response (Boxer et al., 2009).

1.2.2.3 Transactivation of Genes Encoding Structural Proteins In the autonomous parvovirus minute virus of mice (MVM), whose genome contains two overlapping transcription units, the genes coding for the two NS proteins (NS1 and NS2) are transcribed from a promoter P04, whereas the promoter P39 controls the transcription of capsid protein genes. Intriguingly, the P39 promoter was shown to be activated by a viral NS protein NS1 (Doerig et al., 1988).

1.2.3 Viral Regulatory and Accessory Proteins

Viral regulatory and accessory proteins play a number of indirect roles in the viral function, for example, some of these proteins regulate the rate of transcription of viral structural genes. These proteins either regulate the expression of viral genes or are involved in modifying host cell functions. Many viral regulatory and accessory proteins serve multiple functions. For example, the active replication of HIV-1 is controlled by the production of several regulatory (Tat and Rev) and accessory

(Vpr, Vif, Vpu, and Nef) proteins (Seelamgari et al., 2004). Accessory proteins are important for the efficient *in vivo* infection. It is believed that Vif has evolved to overcome the antiviral defense mechanisms of the host, whereas accessory proteins such as Nef increase virus pathogenesis by targeting bystander cells. Therefore, these proteins control many aspects of the virus life cycle as well as host cell function, namely, gene regulation and apoptosis, mostly via interactions with other viral and cellular components (Seelamgari et al., 2004).

1.3 INTRINSIC DISORDER IN VIRAL PROTEINS

Most viral proteins (e.g., proteins involved in replication and morphogenesis of viruses and the major capsid proteins of icosahedral virions), being shared by many groups of RNA and DNA viruses, have no homologs in modern cells (Koonin et al., 2006). This clearly suggests that viruses are very antique and that viral genes primarily originated in the virosphere during replication of viral genomes and/or recruited from cellular lineages are now extinct (Forterre and Prangishvili, 2009). Viruses represent an interesting example of adaptation to extreme conditions, which include both environmental peculiarities and biological and genetic features of the hosts. Viruses have to survive outside and within the host cell (some viruses infect Archaea that are isolated from geothermally heated hot environments (Prangishvili et al., 2006)) and need to infect the host organism and replicate their genes while avoiding the host's countermeasures (Reanney, 1982). Genomes of many viruses are characterized by unusually high rates of mutation, which, being estimated as exchanges per nucleotide, per generation can be as high as 10^{-5}–10^{-3} for RNA viruses, 10^{-5} for single-stranded DNA viruses, and 10^{-8}–10^{-7} for double-stranded DNA viruses, compared to 10^{-10}–10^{-9} in bacteria and eukaryotes (Drake et al., 1998). Viral genomes are unusually compact and contain overlapping reading frames. Therefore, a single mutation might affect more than one viral protein (Reanney, 1982).

All these peculiarities raised an intriguing question on whether the viral proteins possess unique structural features. In an attempt to answer this question, a detailed analysis of viral proteins was undertaken (Tokuriki et al., 2009). First, 3D protein crystal structures of 123 representative single domain proteins of 70–250 amino acids that contain no covalent cofactors, and with a high resolution crystal structure, were analyzed. Of these 123 proteins, 26 were RNA viral proteins, 19 were DNA viral proteins (18 double-stranded and one single-stranded DNA virus), 26 were hypothermophilic, 26 were mesophilic eukaryotes, and 26 were mesophilic prokaryotes. The analysis revealed that viral proteins, especially RNA viral proteins, possessed systematically lower van der Waals contact densities than proteins from other groups. Furthermore, viral proteins were shown to have a larger fraction of residues that are not arranged in well-defined secondary structural elements such as helixes and strands. Finally, the effects of mutations on protein conformational stability ($\Delta\Delta G$ values) were compared for all these proteins. This analysis showed that viral proteins show lower average $\Delta\Delta G$ per residue

than proteins from other organisms. RNA viral proteins show particularly low average $\Delta\Delta G$ values, 0.20 kcal/mol lower than the mesophilic proteins of the same size and 0.26 kcal/mol lower than the thermophilic proteins (Tokuriki et al., 2009).

At the next stage, peculiarities of viral proteins were analyzed using approaches that are independent of structures, namely, amino acid composition profiling and disorder propensity calculations. These tools were applied to all available open reading frames (ORFs) in the relevant proteomes of 19 hyperthermophilic archaea, 35 mesophilic bacteria, 20 eukaryotes, and 30 single-stranded RNA, 30 single-stranded DNA, and 29 double-stranded DNA viruses (Tokuriki et al., 2009). In these analyses, viral proteomes were filtered to remove all annotated capsid/coat/envelope/structural proteins. Figure 1.2a represents the relative composition profiles calculated for various species as described by Vacic and colleagues (Vacic et al., 2007). Here, the fractional difference in composition between a given protein set and a set of completely ordered proteins was calculated for each amino acid residue. The fractional difference was evaluated 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 (Xue et al., 2009a,b). In addition to the filtered data set of viral proteins from 89 proteomes, this figure also includes compositional profile calculated for the nonfiltered data set containing all viral proteins from ~2400 viral species. In general, viral proteins show a reduced fraction of hydrophobic and charged residues and a significantly increased proportion of polar resides. Figure 1.2b clearly shows that viral proteomes exhibit a very high propensity for an intrinsic disorder. In general, the amount of disorder in viruses was comparable with that in eukaryotes, which from previous studies were already known to possess the highest levels of disorder (Romero et al., 1998; Dunker et al., 2000, 2001; Ward et al., 2004; Oldfield et al., 2005a). Figure 1.2b illustrates that there was a fundamental difference between viral and eukaryotic proteomes since eukaryotes contained more proteins with long disordered regions, whereas viral proteomes were characterized by the dominance of short disordered segments (Tokuriki et al., 2009).

On the basis of these observations it has been concluded that in comparison with proteins from their hosts, viral proteins are less densely packed, possess a much weaker network of interresidue interactions (manifested by the lower contact density parameters, the increased fraction of residues not involved in secondary structure elements, and the abundance of short disordered regions), have an unusually high occurrence of polar residues, and are characterized by the lower destabilizing effects of mutations (Tokuriki et al., 2009). It has been concluded that the adaptive forces that shape viral proteins were different form those responsible for evolution of proteins of their hosts. In fact, as discussed, the abundance of polar residues, the lower van der Waals contact densities, high resistance to mutations, and the relatively high occurrence of flexible "coils" and numerous short disordered regions suggested that viral proteins are not likely to have evolved for higher thermodynamic stability but rather to be more adaptive for fast change in their biological and physical environments (Tokuriki et al., 2009).

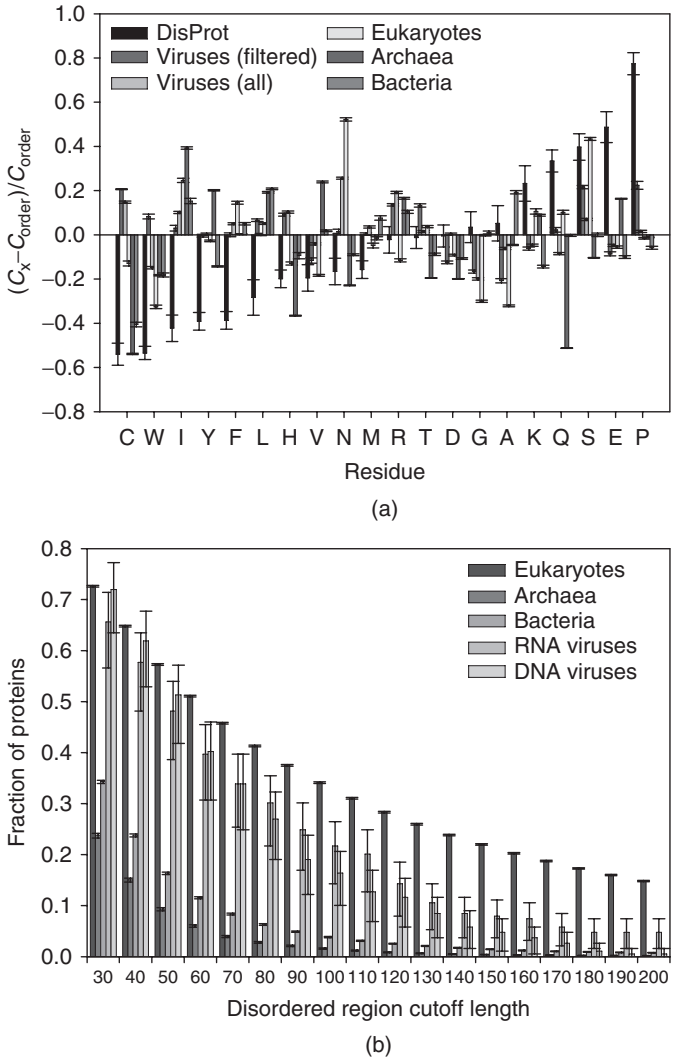

Figure 1.2 Evaluation of the uniqueness and abundance of intrinsic disorder in viral proteins. (a) Composition profile of amino acids for proteins from different organisms. Residues on the x-axis are arranged according to the increasing disorder tendency. The y-axis represents the relative compositional profile compared to a fully disordered data set. (b) The fraction of disordered regions within viral, mesophilic eukaryotic and prokaryotic, and thermophilic proteomes. Shown are the fractions of continuous disordered segments predicted for these proteomes for stretches of varying lengths, from $\geq$30 to $\geq$200 amino acids. (*See insert for color representation of the figure*.)

1.4 FUNCTIONALITY OF INTRINSIC DISORDER IN VIRAL PROTEINS

1.4.1 Intrinsic Disorder and Viral Pfam Domain Seeds

Proteins often contain one or more functional domains, different combinations of which give rise to the diverse range of proteins found in Nature. It has been recognized that the identification of domains that occur within proteins can therefore provide insights into their function. To find a correlation between intrinsic disorder and function in the viral proteins, we analyzed the abundance of intrinsic disorder in the Pfam database, which contains information on protein domains and families and uses hidden Markov models (HMMs) and multiple sequence alignments to identify members of its families emphasizing the evolutionary conservation of protein domains (Bateman et al., 2002, 2004; Finn et al., 2008). Each curated family in Pfam is represented by a seed and full alignment. The seed contains representative members of the family, while the full alignment contains all members of the family as detected with a profile HMM (Bateman et al., 2002). Since Pfam represents an important tool for understanding protein structure and function and since this database contains large amount of information on functional domains, the viral seed domains in the version 23.0 of the Pfam database were analyzed. There are 6360 Pfam domain seeds of viral origin. Figure 1.3 shows that intrinsic disorder is rather abundant among the viral Pfam seed domains. In fact, 535 Pfam domain seeds of viral origin were 50–98% disordered, and the length of disordered regions in the domains varied from 11 to 738 residues (Fig. 1.3a). Figure 1.3b shows that >100 domains ranging in length from 14 to 324 residues were almost completely disordered. Our analysis revealed that many Pfam domain

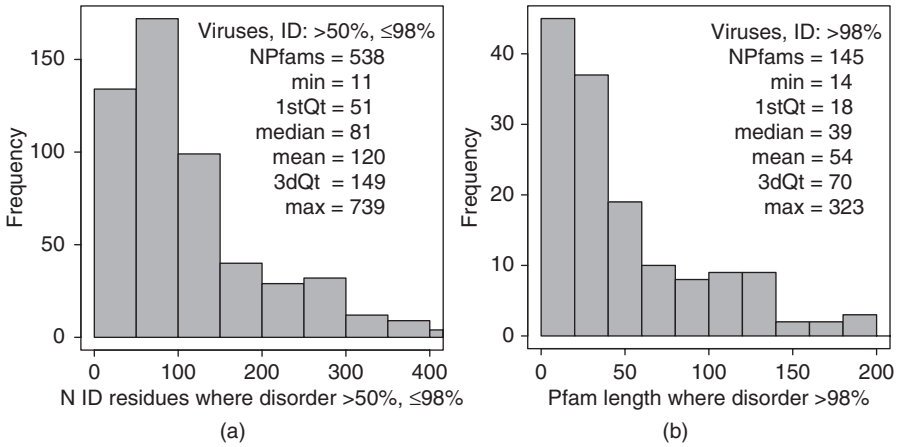

Figure 1.3 Intrinsic disorder distribution in Pfam domain seeds of the viral origin. (a) The length distribution of Pfam domain seeds of viral origin which are 50–98% disordered. (b) The length distribution of Pfam domains where disorder is observed for >98% residues.

seeds of the viral origin were completely disordered (they contain >98% disordered residues) but still possessed a number of crucial biological functions, mostly related to interaction with proteins, as well as recognition, regulation, and signal transduction (Xue et al., 2010). In other words, viral disordered domains possess functions similar to those of prokaryotic, archaeal, and eukaryotic proteins (Wright and Dyson, 1999; Uversky et al., 2000; Dunker et al., 2001, 2002a,b).

1.4.2 Intrinsic Disorder in Viral Structural Proteins

1.4.2.1 Capsids Capsids represent an economical use of multiple copies of a single or a few proteins to build a specific cage for genome transfer. In fact, this approach helps viruses to minimize the coding space for the capsid and also determines an easy and self-controlled mechanism of shell assembly, where only the fitting pieces can work. Since an icosahedral symmetry provides a low energy solution for the shell formation, it is commonly used by many isomeric (or icosahedral) viruses (Caspar and Klug, 1962). Sixty identical units can form an icosahedron. Although the majority of capsid proteins are relatively small, many viruses have very large capsids. These large capsids are built from a high number of building blocks, many times exceeding 60 units. The theory of quasi-equivalence, according to which the capsid is stabilized by the same type of interactions that are perturbed in slightly different ways in the non-symmetry-related environments, explained this apparent contradiction since multiples of 60 proteins can be arranged such that they will all be in nearly identical environments (Caspar and Klug, 1962).

Intriguingly, already in the first virus structures determined by X-ray crystallography (Harrison et al., 1978; Abad-Zapatero et al., 1980), the coat proteins appeared as globular parts (C-terminal domains) formed by two antiparallel four-stranded sheets with a jelly-roll or Swiss-roll topology and extended, partially invisible N-terminal segments (Liljas, 2004). In the polyoma virus and simian virus 40 (SV40), whose capsids are described by an icosahedral surface lattice with the triangulation number $T = 7$d, there are 360 units in the capsid (Baker et al., 1983; Liddington et al., 1991), which is noticeably less than the 420 units expected from the Caspar–Klug rules (Caspar and Klug, 1962). In these viruses, all 72 capsomers are pentamers of the structural protein VP1 (in polyoma virus) or of the coat protein (in SV40) rather than an expected mixture of pentamers and hexamers. Therefore, pentamers are found at the positions predicted to have a hexamer of subunits according to the Caspar–Klug hypothesis (Caspar and Klug, 1962). The apparent contradiction is resolved by intrinsically disordered arms of the capsid proteins: the intercapsomer contacts are established by the folded C-terminal domain, whereas the N-terminal domain of the capsid protein, the so-called arm, is extended and is present in six totally different conformations depending on its position in the lattice (Rayment et al., 1982; Liljas, 2004).

In agreement with the icosahedral symmetry, in the capsid of foot-and-mouth-disease virus there are 60 identical subunits, each of which is made up of four proteins: VP1, VP2, VP3, and VP4 (Fry et al., 2005). VP1, VP2, and VP3 are wedge-shaped, eight-stranded β-sandwiches. The loops connecting strands at the

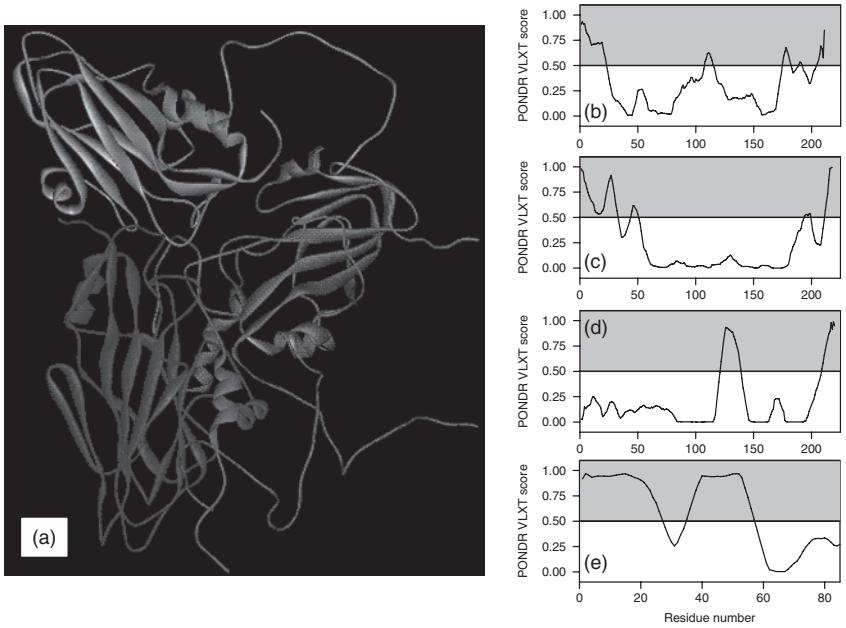

Figure 1.4 Intrinsic disorder in viral structural proteins. (a) Structure of the capsomer in the icosahedral capsid of the foot-and-mouth-disease virus. This capsid contains 60 capsomers, each of which is made up of four capsid proteins: VP1, VP2, VP3, and VP4. Distribution of predicted intrinsic disorder in capsid proteins of the foot-and-mouth-disease virus: VP1 (b), VP2 (c), VP3 (d), and VP4 (e).

narrow end of the wedge are less constrained by structural interactions and tend to mediate host interactions. VP4 and the N-termini of VP1 and VP3 are located at the capsid interior. It has been pointed out that the capsomer structure comprises residues 1–137 and 155–208 of VP1, 12–218 of VP2, 1–221 of VP3, and 15–39 and 62–85 of VP4, whereas residues 138–154 and 209–212 of VP1, 1–11 of VP2, and 1–14 and 40–61 of VP4 were too flexible to be modeled reliably (Fry et al., 2005). Therefore, the three major capsid proteins are mostly ordered and have a conserved iβ-barrel fold, whereas VP4 has a little regular secondary structure. This mostly disordered protein, VP4, is involved in the initial disassembly and final assembly stages of the virus.

Functional roles of disordered regions of capsid proteins extend far away from simple structural roles. In the excellent review by Liljas (2004), the functionality of various disordered arms of viral capsid proteins is systemized to show that these intrinsically disordered fragments can be used to stabilize the structure of a capsid, to control the capsid assembly and disassembly, and for the interaction with nucleic acids. To finish this part, Fig. 1.4 represents the results of disorder prediction for the capsid proteins discussed above together with the crystal structure of the foot-and-mouth-disease virus capsomer.

1.4.2.2 Viral Envelope Specific surface glycoproteins are used by the enveloped viruses, such as influenza, HIV-1, and Ebola, to enter target cells via fusion of the viral membrane with the target cellular membrane (Skehel and Wiley, 1998, 2000; Eckert and Kim, 2001). One of the most well-studied membrane fusion proteins is the influenza virus HA, which is a homotrimeric type I transmembrane surface glycoprotein responsible for virus binding to the host receptor, internalization of the virus, and subsequent membrane-fusion events within the endosomal pathway in the infected cell. HA is also the most abundant antigen on the viral surface and harbors the primary neutralizing epitopes for antibodies. Each 70-kDa HA subunit contains two disulfide-linked polypeptide chains, HA_1 and HA_2, created by proteolytic cleavage of the precursor protein HA_0 (Wiley and Skehel, 1987). Such a cleavage is absolutely crucial for membrane fusion (Wiley and Skehel, 1987). During membrane fusion, HA binds the virus to sialic acid receptors on the host cell surface, and following endocytosis, the acidic pH (pH 5–6) of endosomal compartments induces dramatic and irreversible reorganization of the HA structure (Skehel et al., 1982).

The HA trimer has a tightly intertwined "stem" domain at its membrane-proximal base, which is composed of HA_1 residues 11–51 and 276–329 and HA_2 residues 1–176. The dominant 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, that associate to form a triple-stranded coiled coil. The membrane-distal domain consists of a globular "head," which is formed by HA_1 and which can be further subdivided into the R region (residues 108–261), containing the receptor-binding site and major epitopes for neutralizing antibodies, and the E region (residues 56–108 and 262–274), with close structural homology to the esterase domain of influenza C HA esterase fusion (HEF) protein (Stevens et al., 2004). The HA_2 chain contains two membrane-interacting hydrophobic peptide sequences: an N-terminal "fusion peptide" (residues 1–23), which interacts with the target membrane bilayer (Durrer et al., 1996), and a C-terminal transmembrane segment, which passes through the viral membrane.

Crystallographic studies suggested that the interaction with the host cell involves a dramatic structural reorganization of HA_2, which moves the fusion peptide from the interior approximately 100 AA toward the target membrane (Wilson et al., 1981; Bullough et al., 1994). In this process, the middle of the original long α-helix unfolds to form a reverse turn, jackknifing the C-terminal half of the long α-helix backward toward the N-terminus. These molecular rearrangements place the N-terminal fusion peptide and the C-terminal transmembrane anchor at the same end of the rod-shaped HA_2 molecule (Weber et al., 1994; Wharton et al., 1995), facilitating membrane fusion by bringing the viral and cellular membranes together.

Our recent analysis revealed that although many viral membrane glycoproteins are ordered, intrinsic disorder still is crucial for the biology of these proteins. For example, we have found some distinct differences in the disorder propensity between HA proteins of the virulent and nonvirulent strains of influenza A, especially in the region near residues 68–79 of the HA_2. This region represents the tip of the stalk that is in contact with the receptor chain, HA_1, and is therefore

likely to provide the greatest effect on the motions of the exposed portion of HA. Comparison of this region between virulent strains (1918 H1N1 and H5N1) and less virulent ones (H3N2 and 1930 H1N1) showed that this region is characterized by the increased level of intrinsic disorder in more virulent strains and subtypes of the virus but is predicted to be mostly ordered in less virulent strains (Goh et al., 2009).

1.4.2.3 Matrix We also analyzed the predisposition of several viral matrix proteins to intrinsic disorder (Goh et al., 2008a,b). These studies revealed that the matrix protein p17 from simian immunodeficiency virus (SIV_{mac}) and HIV-1 possesses high levels of predicted intrinsic disorder, whereas matrix proteins of the equine infectious anemia virus (EIAV) were characterized by noticeably lower levels of predicted disorder (Goh et al., 2008a).

1.4.3 Intrinsic Disorder in Viral Nonstructural, Regulatory, and Accessory Proteins

Since these proteins are responsible for the wide range of recognition- and regulation-based functions, including communication with the hosts and regulation of virus replication and assembly, they are frequently disordered. As illustrative examples, we are presenting below a brief overview of intrinsic disorder in several NS proteins from various viruses and the regulatory and accessory proteins from HIV-1.

1.4.3.1 Disorder in Viral Nonstructural Proteins As discussed earlier, an NS oncoprotein E7 of HPV is involved in regulation of cell growth and transformation, gene transcription, apoptosis, and DNA synthesis. It is known to interact with a number of cellular proteins, such as the Rb, p107 and p130, glycolytic enzymes, histone deacetylase, kinase p33CDK2, and cyclin A, and the cyclin-dependent kinase inhibitor $p21^{cip1}$ tumor suppressor. Importantly, E7 is involved in the pathogenesis and maintenance of human cervical cancers. The analysis of the E7 dimer from HPV45 by NMR revealed that each monomer contained an unfolded N-terminus and a well-structured C-terminal domain (Ohlenschlager et al., 2006). Later, a fragment of the oncoprotein E7 comprising the highly acidic N-terminal domain was confirmed to be intrinsically disordered by far-UV CD (circular dichroism) hydrodynamic analyses. Importantly, the N-terminal domain of this protein (residues 1–40) includes the retinoblastoma tumor suppressor binding and casein kinase II phosphorylation sites (Garcia-Alai et al., 2007).

There are more than 100 different types of HPVs, which are the causative agents of benign papillomas/warts and the cofactors in the development of carcinomas of the genital tract, head and neck, and epidermis. In respect to their association with cancer, HPVs are grouped into two classes, low risk (e.g., HPV-6 and HPV-11) and high risk (e.g., HPV-16 and HPV-18) types. Recently, in order to understand whether intrinsic disorder plays a role in the oncogenic potential of different HPV types, the bionformatics analysis of proteomes of high risk and low risk HPVs

with a major focus on E6 and E7 oncoproteins has been performed (Uversky et al., 2006). On the basis of the results of this analysis, it has been concluded that high risk HPVs are characterized by an increased amount of intrinsic disorder in transforming proteins E6 and E7 (Uversky et al., 2006).

Influenza virus NS protein 2 (NS2, or NEP) is known to interact with the nuclear export machinery during viral replication and serves as an adapter molecule between the nuclear export machinery and the viral ribonucleoprotein complex. Structural analysis of the recombinant NS2 by spectroscopy, differential scanning calorimetry, limited proteolysis, and hydrodynamic techniques revealed that this monomeric protein shows characteristics of the native molten globule under near physiological conditions being compact and highly flexible (Lommer and Luo, 2002).

1.4.3.2 Disorderedness of Viral Regulatory and Accessory Proteins Protein Tat is the HIV-1 transactivator of viral transcription and is an important factor in viral pathogenesis. Tat binds to a short nascent stem-bulge loop leader RNA, termed the *transactivation responsive region* (TAR), that is present at the 5′ extremity of all viral transcripts via its basic region and recruits the complex of cyclin T1 and cyclin-dependent kinase 9 (CDK9) forming the positive transcription elongation factor B complex. CDK9 hyperphosphorylates the carboxy terminus domain of RNA polymerase II, leading to the enhanced elongation of transcription from the viral promoter. However, Tat not only acts as the key transactivator of viral transcription but is also secreted by the infected cell and is taken up by neighboring cells where it has an effect both on infected and uninfected cells (Campbell and Loret, 2009).

The Tat amino acid sequence is characterized by a low overall hydrophobicity and a high net positive charge. This protein was predicted to be natively unfolded by several algorithms (Shojania and O'Neil, 2006). These predictions were in agreement with the lack of ordered secondary structure in this protein found by the CD analysis (Vendel and Lumb, 2003), and NMR chemical shifts and coupling constants suggested that Tat existed in a random coil conformation (Shojania and O'Neil, 2006).

Rev is another regulatory protein in HIV-1. This is a 116-residue basic protein that binds to multiple sites in the Rev-response element (RRE) of viral mRNA transcripts in nuclei of host cells, leading to transport of incompletely spliced and unspliced viral mRNA to the cytoplasm of host cells in the later phases of the HIV-1 life cycle. Therefore, Rev is absolutely required for viral replication (Blanco et al., 2001). On the basis of the detailed spectroscopic and hydrodynamic studies, it has been concluded that monomeric Rev is in a molten globule state (Surendran et al., 2004).

Vpr is a 96-residue HIV-1 accessory protein that shows multiple activities, including nuclear transport of the preintegration complex to the nucleus, activation of transcription, cell cycle arrest at the G2/M transition, and triggering of apoptosis. This protein controls many host cell functions through a variety of biological activities and by interaction with cellular biochemical pathways. For example, nuclear

import of Vpr may be due to its interaction with nuclear transport factors and components of the nuclear pore complex. Cell cycle arrest has been correlated with the binding to DCAF1, a cullin-4A-associated factor, and apoptosis may be facilitated by interaction with mitochondrial proteins in a caspase-dependent mechanism. Vpr also plays a critical role in long-term AIDS by inducing viral infection in nondividing cells such as monocytes and macrophages (Morellet et al., 2009). On the basis of the dynamic light scattering (DLS), CD, and ^{1}H NMR spectroscopy analyses, it has been concluded that Vpr was unstructured at neutral pH, whereas under acidic conditions or on addition of trifluorethanol it adopts α-helical structures (Henklein et al., 2000). On the basis of this pH-dependent folding switch, it has been suggested that the Vpr structure is dependent on the presence of specific binding factors (such as nucleic acids, proteins, or membrane components) or the environment of the cytosol, nucleus, mitochondrion, cellular membranes, and the extracellular space (Bruns et al., 2003).

Vif is another HIV-1 accessory protein that neutralizes the cellular defense mechanism against the virus. Many of the interactions of Vif are mediated via its C-terminal domain (residues 141–192). Detailed structural analysis has revealed that this fragment is mostly disordered, a conclusion based on the coil-like far-UV CD spectrum with some residual helical structure, unfolded features of the ^{15}N-HSQC NMR spectrum, and the extended size evaluated by size-exclusion chromatography. These findings have been further supported by the results of the computational analyses of the Vif C-terminal domain sequence. Importantly, CD analysis has revealed that this domain is able to fold upon interaction with membrane micelles, clearly showing that this natively unfolded domain may gain structure on binding its natural ligands (Reingewertz et al., 2009).

Vpu is an oligomeric type I integral membrane phosphoprotein that amplifies the release of virus particles from infected cells by mediation of degradation of the HIV receptor CD4 by the proteasome in the endoplasmic reticulum. Phosphorylation of Vpu at two sites, Ser52 and Ser56, on the motif DSGXXS is required for the interaction of Vpu with the ubiquitin ligase SCF-βTrCP, which triggers CD4 degradation by the proteasome. Vpu consists of a hydrophobic N-terminal membrane-anchoring domain and a polar C-terminal cytoplasmic domain (Gramberg et al., 2009). CD and NMR analyses of nine overlapping 15 amino acid fragments and 3 longer fragments in aqueous solutions have revealed that the C-terminal hydrophilic domain of Vpu is mostly disordered with some limited amounts of stable secondary structure. However, in the presence of trifluoroethanol, this domain protein is shown to fold into a helical conformation composed of two α-helices joined by a flexible region of six or seven residues, which contains the phosphorylation sites of Vpu at positions 52 and 55 (Wray et al., 1995).

Nef is an HIV-1 accessory protein that is known to interact with multiple cellular partners during the course of infection. The interactions of this viral protein with various cellular partners are mediated by the occurrence of ligand-induced conformational changes that direct the binding of Nef to subsequent partners. On the basis of the analysis of the available experimental data, it has been hypothesized that the binding-promoted conformational changes underwent by this protein

define a novel allosteric paradigm, namely, changes that involve conformations with large disordered regions (Leavitt et al., 2004). Importantly, these regions, being devoid of stable secondary or tertiary structure, contain the binding determinants for subsequent partners and only become functionally competent by ligand-induced folding and unfolding. This model of switching binding epitopes between structured and unstructured conformations provides a unique ability to modulate the binding affinity by several orders of magnitude (Leavitt et al., 2004).

1.5 INTRINSIC DISORDER, ALTERNATIVE SPLICING, AND OVERLAPPING READING FRAMES IN VIRAL GENOMES

Viruses have evolved a complex genetic organization for optimal use of their limited genomes and production of all necessary structural and regulatory proteins. The use of alternative splicing is essential for balanced expression of multiple viral regulators from one genomic polycistronic RNA. Furthermore, viruses use both sense and antisense transcriptions. For example, the genome of human T-cell lymphotropic virus type 1 (HTLV-1), which is a causative agent of adult T-cell leukemia (ATL), HTLV-1-associated myelopathy, and *Strongyloides stercoralis* hyperinfection, encodes common structural and enzymatic proteins found in many retroviruses (Gag, Pro, Pol, and Env). *Gag, Pro*, and *Pol* genes are translated as a series of polyproteins, Gag, Gag-Pro, and Gag-Pro-Pol, which are then cleaved posttranslationally to generate seven proteins. *Gag* gene encodes a polyprotein (Gag) whose cleavage products are the major structural proteins (matrix (MA), capsid (CA), and nucleocapsid (NC)) of the virus core. *Pro* encodes a middle part of a polyprotein (Gag-Pro or Gag-Pro-Pol) whose cleavage products include protease (PR). Finally, *Pol* encodes the last part of a polyprotein (Gag-Pro-Pol) whose cleavage products include reverse transcriptase (RT) and integrase (IN). *Env* encodes a polyprotein (Env) whose cleavage products SU (surface) and TM (transmembrane) are the structural proteins of the viral envelope. In addition to these common retroviral proteins, HTLV-1 encodes multiple regulatory and accessory proteins in four overlapping ORFs located in the pX region of the viral genome (Ciminale et al., 1992; Koralnik et al., 1992). The HTLV-1 basic leucine zipper factor (HBZ) is the product of the antisense transcription (Larocca et al., 1989). There are two transcripts that encode the *HBZ* gene, spliced (*sHBZ*) and unspliced (*usHBZ*). *sHBZ* has multiple transcriptional initiation sites in the U5 and R regions of the 3′ long terminal repeat (LTR), whereas the *usHBZ* gene initiates within the *tax* gene (Matsuoka and Green, 2009). Furthermore, expression of the various ORFs is controlled by differential splicing of the single genomic mRNA, producing unspliced, singly spliced, and multiply spliced mRNA (Kashanchi and Brady, 2005). Therefore, the replication of HTLV-1 is controlled by a group of nuclear and cytoplasmic processes, including transcription, splicing, alternative splicing, mRNA nuclear export, RNA stability, and translation (Baydoun et al., 2008).

Let us consider in more detail Tax and Rex, two HTLV-1 regulatory proteins needed for the viral genome expression. Tax is a transcriptional activator of the

viral promoter (Sodroski et al., 1984; Cann et al., 1985; Felber et al., 1985; Seiki et al., 1986; Derse, 1987; Boxus et al., 2008). Rex affects posttranscriptional regulatory steps by promoting transport of the unspliced and singly spliced mRNA from the nucleus to the cytoplasm and by promoting expression of the Gag, Pol, and Env proteins (Kiyokawa et al., 1985; Inoue et al., 1986, 1987; Derse, 1988; Seiki et al., 1988). Tax and Rex were shown to be expressed from two overlapping ORFs located in the distal part of the pX region of the virus by a bicistronic viral mRNA consisting of three exons (Seiki et al., 1983; Nagashima et al., 1986; Ciminale et al., 1992). There are two alternatively spliced isoforms of Rex, p27Rex and p21Rex. In p21Rex, residues 1–78 are missing. Furthermore, alternative splicing of the pX region in the ORF II generates two accessory proteins, p30 and p13.

Figure. 1.5 represents the HTLV-1 proteome map where each major product described above is present as a bar whose location corresponds to the location of the corresponding gene within the HTLV-1 genome. PONDR® VLXT disorder predictions for each (poly)protein are shown as solid lines inside the corresponding bars. A residue is considered to be disordered if its score is above 0.5. The top half of each bar shaded in gray corresponds to disorder scores >0.5. Therefore, inside each bar, pieces of the PONDR plots located in these shaded area correspond to protein fragments predicted to be disordered. Cleavage sites producing Gag, Pro, and Pol polyproteins are indicated by angled arrows and lettered. Cleavage sites, which are responsible for the posttranslational production of MA, CA, NC, RT, IN, SU, and TM proteins, are marked by short straight arrows and numbered. Gag, Pro, Pol, Env, p12, Tax, p27Rex, p21Rex, p30, and p13 are all the products of the genes produced by the sense transcription. Proteins usHBZ and sHBZ are produced from genes generated by antisense transcription. In Fig. 1.5, this fact is indicated by a long bold arrow marked with letters N and C to indicate the location of the beginnings and ends of the corresponding proteins, respectively. Obviously, the numbering of residues for the usHBZ and sHBZ presentation was inverted. p27Rex, p30, and sHBZ proteins are translated from the spliced genes. There are three alternatively spliced pairs of proteins in HTLV-1: p27Rex and p21Rex, p30 and p13, and usHBZ and sHBZ.

Analysis of this figure clearly shows that the economic usage of genetic material by HTLV-1 is translated into very important implementations of intrinsic disorder for the corresponding proteins.

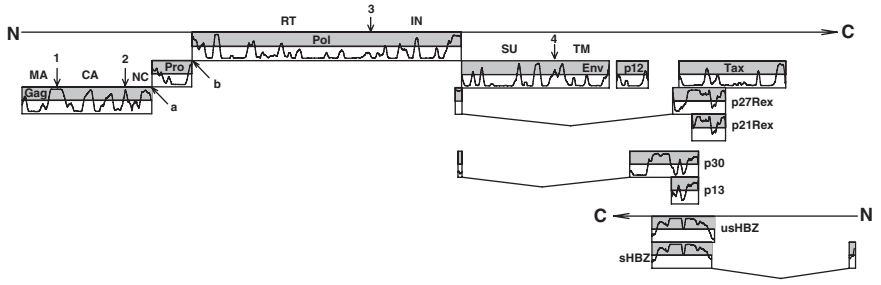

Figure 1.5 The proteome map of HTLV-1. See explanation in the text.

1. Prevailing intrinsic disorder is observed in posttranslational cleavage sites producing polyproteins Gag, Pro, and Pol from the grand-polyproteins Gag–Pro and Gar-Pro-Pol, as well as in the cleavage site producing MA, CA, NC, RT, IN, SU, and TM proteins from the corresponding polyproteins.
2. Proteins affected by alternative splicing are in general highly disordered. Furthermore, protein fragments removed by alternative splicing are mostly disordered.
3. Protein fragments corresponding to the overlapping genes are either disordered or possess complementary disorder distribution if these protein regions are not translated from the genes transcribed from the identical ORFs. For example, the N-terminal fragment of Tax, which overlaps with a significant portion of Rex, is mostly ordered, whereas the corresponding region in the Rex proteins is predominantly disordered. Similarly, the C-terminal region of p30, which overlaps with the ordered N-terminal fragment of Tax, is mostly disordered, as disordered is the p13 protein, which completely overlaps with the ordered N-terminal fragment of Tax.
4. Proteins translated from genes generated by antisense transcription are highly disordered.

The conclusion that the HTLV-1 proteins produced by the overlapping genes are intrinsically disordered in agreement with a recent study where the protein coded by overlapping genes from 43 genera of unspliced RNA viruses infecting eukaryotes has been analyzed (Rancurel et al., 2009). This study has revealed that overlapping proteins have a sequence composition globally biased toward disorder-promoting amino acids and are predicted to contain significantly more structural disorder than nonoverlapping proteins (Rancurel et al., 2009).

Importantly, many of the specific implementations of intrinsic disorder listed above are not unique to the viral proteins. In fact, sites of proteolytic cleavage of proteins from other organisms are frequently located in the disordered regions (Fontana et al., 1986, 1997a,b, 2004; Polverino de Laureto et al., 1995; Iakoucheva et al., 2001; de Laureto et al., 2006). It has been also shown that regions of mRNA that undergo alternative splicing code for disordered proteins much more often than they code for structured proteins (Romero et al., 2006). Finally, the so-called retro-proteins, that is, proteins whose sequence is read backward providing a new polypeptide that does not align with its parent sequence, were shown to lack an ordered 3D structure (Lacroix et al., 1998).

1.6 CONCLUDING REMARKS

The modern literature on protein intrinsic disorder in viral proteomes has been systematically analyzed. Published data clearly show that viral proteins are both different and similar to proteins from their hosts. On one hand, viral proteins are less densely packed, possess a much weaker network of interresidue interactions (manifested by the lower contact density parameters, the increased fraction of

residues not involved in secondary structure elements, and the abundance of short disordered regions), have an unusually high occurrence of polar residues, and are characterized by the lower destabilizing effects of mutations. On the basis of these peculiar features, it has been concluded that viral proteins are not likely to have evolved for higher thermodynamic stability but rather to be more adaptive for fast change in their biological and physical environments. On the other hand, recent studies clearly show that intrinsic disorder is widespread in viral proteomes and has a number of important functional implementations. In fact, almost all viral proteins, irrespective of their functions, have biologically important disordered regions. The list of functions attributed to these disorder regions of viral proteins overlaps with disorder-based activities of proteins from other organisms. In fact, many functional Pfam seed domains of the viral origin were shown to possess various levels of intrinsic disorder, with ∼150 such seeds being completely disordered. Disordered Pfam domains were involved in various crucial functions, such as signaling, regulation, and interaction with nucleic acids and proteins, suggesting that similar to proteins from all domains of life, intrinsic disorder is heavily used by viral proteins in their functions. Therefore, although viral proteins possess a number of unique features, they still rely intensively on intrinsic disorder at almost all stages of their intriguing life cycle.

ACKNOWLEDGMENT

This work was supported in part by the grant EF 0849803 (to A. K. D and V. N. U.) from the National Science Foundation and the Program of the Russian Academy of Sciences for the "Molecular and Cellular Biology" (to V. N. U.). We gratefully acknowledge the support of the IUPUI Signature Centers Initiative. This chapter is based on a previous publication (Xue et al., 2010).

ABBREVIATIONS

AFV1	*Acidianus* filamentous virus 1
APMV	*Acanthamoeba polyphaga* mimivirus
ATL	adult T-cell leukemia
CA	capsid
CD	circular dichroism
CDK9	cyclin-dependent kinase 9
DLS	dynamic light scattering
EIAV	equine infectious anemia virus
HA	hemagglutinin
HBZ	basic leucine zipper factor
HCV	hepatitis C virus
HEF	HA esterase fusion
HIV-1	human immunodeficiency virus-1

HMM	hidden Markov model
HPV	human papilloma virus
HTLV-1	human T-cell lymphotropic virus type 1
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IN	integrase
MA	matrix
MVM	minute virus of mice
NC	nucleocapsid
NS	nonstructural
ORF	open reading frame
PKB	protein kinase P
PONDR	predictor of natural disordered regions
PP2A	protein phosphatase 2A
RRE	Rev-response element
RT	reverse transcriptase
SIFV	*Sulfolobus islandicus* filamentous virus
SIV_{mac}	simian immunodeficiency virus
SU	surface
TM	transmembrane
TMV	tobacco mosaic virus

REFERENCES

Abad-Zapatero C, Abdel-Meguid SS, Johnson JE, Leslie AG, Rayment I, Rossmann MG, Suck D, Tsukihara T. Structure of southern bean mosaic virus at 2.8 Å resolution. Nature 1980;286:33–39.

Baker TS, Caspar DL, Murakami WT. Polyoma virus hexamer tubes consist of paired pentamers. Nature 1983;303:446–448.

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. The Pfam protein families database. Nucleic Acids Res 2002;30:276–280.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Nucleic Acids Res 2004;32:D138–D141.

Baydoun HH, Bellon M, Nicot C. HTLV-1 Yin and Yang: Rex and p30 master regulators of viral mRNA trafficking. AIDS Rev 2008;10:195–204.

Bergh O, Borsheim KY, Bratbak G, Heldal M. High abundance of viruses found in aquatic environments. Nature 1989;340:467–468.

Blanco FJ, Hess S, Pannell LK, Rizzo NW, Tycko R. Solid-state NMR data support a helix-loop-helix structural model for the N-terminal half of HIV-1 Rev in fibrillar form. J Mol Biol 2001;313:845–859.

Boxer EL, Nanda SK, Baron MD. The rinderpest virus non-structural C protein blocks the induction of type 1 interferon. Virology 2009;385:134–142.

Boxus M, Twizere JC, Legros S, Dewulf JF, Kettmann R, Willems L. The HTLV-1 Tax interactome. Retrovirology 2008;5:76.

Brazil DP, Hemmings BA. Ten years of protein kinase B signalling: a hard Akt to follow. Trends Biochem Sci 2001;26:657–664.

Brehm A, Nielsen SJ, Miska EA, McCance DJ, Reid JL, Bannister AJ, Kouzarides T. The E7 oncoprotein associates with Mi2 and histone deacetylase activity to promote cell growth. EMBO J 1999;18:2449–2458.

Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? Trends Microbiol 2005;13:278–284.

Brown DR, Kitchin D, Qadadri B, Neptune N, Batteiger T, Ermel A. The human papillomavirus type 11 E1–E4 protein is a transglutaminase 3 substrate and induces abnormalities of the cornified cell envelope. Virology 2006;345:290–298.

Bruns K, Fossen T, Wray V, Henklein P, Tessmer U, Schubert U. Structural characterization of the HIV-1 Vpr N terminus: evidence of cis/trans-proline isomerism. J Biol Chem 2003;278:43188–43201.

Bullough PA, Hughson FM, Skehel JJ, Wiley DC. Structure of influenza haemagglutinin at the pH of membrane fusion. Nature 1994;371:37–43.

Cady SD, Luo W, Hu F, Hong M. Structure and function of the influenza A M2 proton channel. Biochemistry 2009;48:7356–7364.

Campbell GR, Loret EP. What does the structure-function relationship of the HIV-1 Tat protein teach us about developing an AIDS vaccine? Retrovirology 2009;6:50.

Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brussow H. Phage as agents of lateral gene transfer. Curr Opin Microbiol 2003;6:417–424.

Cann AJ, Rosenblatt JD, Wachsman W, Shah NP, Chen IS. Identification of the gene responsible for human T-cell leukaemia virus transcriptional regulation. Nature 1985;318:571–574.

Caspar DL, Klug A. Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol 1962;27:1–24.

Chen JJ, Reid CE, Band V, Androphy EJ. Interaction of papillomavirus E6 oncoproteins with a putative calcium-binding protein. Science 1995;269:529–531.

Cheng Y, LeGall T, Oldfield CJ, Dunker AK, Uversky VN. Abundance of intrinsic disorder in protein associated with cardiovascular disease. Biochemistry 2006;45:10448–10460.

Chung KM, Liszewski MK, Nybakken G, Davis AE, Townsend RR, Fremont DH, Atkinson JP, Diamond MS. West Nile virus nonstructural protein NS1 inhibits complement activation by binding the regulatory protein factor H. Proc Natl Acad Sci USA 2006;103:19111–19116.

Ciminale V, Pavlakis GN, Derse D, Cunningham CP, Felber BK. Complex splicing in the human T-cell leukemia virus (HTLV) family of retroviruses: novel mRNAs and proteins produced by HTLV type I. J Virol 1992;66:1737–1745.

Cripe TP, Haugen TH, Turk JP, Tabatabai F, Schmid PG, 3rd, Durst M, Gissmann L, Roman A, Turek LP. Transcriptional regulation of the human papillomavirus-16 E6-E7 promoter by a keratinocyte-dependent enhancer, and by viral E2 transactivator and repressor gene products: implications for cervical carcinogenesis. Embo J 1987;6:3745–3753.

Davy CE, Ayub M, Jackson DJ, Das P, McIntosh P, Doorbar J. HPV16 E1;E4 protein is phosphorylated by Cdk2/cyclin A and relocalizes this complex to the cytoplasm. Virology 2006;349:230–244.

de Laureto PP, Tosatto L, Frare E, Marin O, Uversky VN, Fontana A. Conformational properties of the SDS-bound state of alpha-synuclein probed by limited proteolysis: unexpected rigidity of the acidic C-terminal tail. Biochemistry 2006;45:11523–11531.

Derse D. Bovine leukemia virus transcription is controlled by a virus-encoded trans-acting factor and by cis-acting response elements. J Virol 1987;61:2462–2471.

Derse D. trans-acting regulation of bovine leukemia virus mRNA processing. J Virol 1988;62:1115–1119.

Dey A, Atcha IA, Bagchi S. HPV16 E6 oncoprotein stimulates the transforming growth factor-beta 1 promoter in fibroblasts through a specific GC-rich sequence. Virology 1997;228:190–199.

Doerig C, Hirt B, Beard P, Antonietti JP. Minute virus of mice non-structural protein NS-1 is necessary and sufficient for trans-activation of the viral P39 promoter. J Gen Virol 1988;69(Pt 10):2563–2573.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. Genetics 1998;148:1667–1686.

Du M, Fan X, Hong E, Chen JJ. Interaction of oncogenic papillomavirus E6 proteins with fibulin-1. Biochem Biophys Res Comm 2002;296:962–969.

Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002a;41:6573–6582.

Dunker AK, Brown CJ, Obradovic Z. Identification and functions of usefully disordered proteins. Adv Protein Chem 2002b;62:25–49.

Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 2005;272: 5129–5148.

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. J Mol Graph Model 2001;19:26–59.

Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic protein disorder in complete genomes. Genome Inform Ser Workshop Genome Inform 2000;11:161–171.

Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN. The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics 2008a;9(Suppl 2): S1.

Dunker AK, Silman I, Uversky VN, Sussman JL. Function and structure of inherently disordered proteins. Curr Opin Struct Biol 2008b;18:756–764.

Dunker AK, Uversky VN. Signal transduction via unstructured protein conduits. Nat Chem Biol 2008;4:229–230.

Durrer P, Galli C, Hoenke S, Corti C, Gluck R, Vorherr T, Brunner J. H^+-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region. J Biol Chem 1996;271:13417–13421.

Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6:197–208.

Dyson N, Howley PM, Munger K, Harlow E. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science 1989;243:934–937.

Eckert DM, Kim PS. Mechanisms of viral membrane fusion and its inhibition. Annu Rev Biochem 2001;70:777–810.

Edwards RA, Rohwer F. Viral metagenomics. Nat Rev Microbiol 2005;3:504–510.

Fehrmann F, Klumpp DJ, Laimins LA. Human papillomavirus type 31 E5 protein supports cell cycle progression and activates late viral functions upon epithelial differentiation. J Virol 2003;77:2819–2831.

Felber BK, Paskalis H, Kleinman-Ewing C, Wong-Staal F, Pavlakis GN. The pX protein of HTLV-I is a transcriptional activator of its long terminal repeats. Science 1985;229:675–679.

Feng ZP, Zhang X, Han P, Arora N, Anders RF, Norton RS. Abundance of intrinsically unstructured proteins in P. Falciparum and other apicomplexan parasite proteomes. Mol Biochem Parasitol 2006;150:256–267.

Filippova M, Song H, Connolly JL, Dermody TS, Duerksen-Hughes PJ. The human papillomavirus 16 E6 protein binds to tumor necrosis factor (TNF) R1 and protects cells from TNF-induced apoptosis. J Biol Chem 2002;277:21730–21739.

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. The Pfam protein families database. Nucleic Acids Res 2008;36:D281–D288.

Fontana A, de Laureto PP, Spolaore B, Frare E, Picotti P, Zambonin M. Probing protein structure by limited proteolysis. Acta Biochim Pol 2004;51:299–321.

Fontana A, Fassina G, Vita C, Dalzoppo D, Zamai M, Zambonin M. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. Biochemistry 1986;25:1847–1851.

Fontana A, Polverino de Laureto P, De Filippis V, Scaramella E, Zambonin M. Probing the partly folded states of proteins by limited proteolysis. Fold Des 1997a;2:R17–R26.

Fontana A, Zambonin M, Polverino de Laureto P, De Filippis V, Clementi A, Scaramella E. Probing the conformational state of apomyoglobin by limited proteolysis. J Mol Biol 1997b;266:223–230.

Forterre P. The origin of viruses and their possible roles in major evolutionary transitions. Virus Res 2006;117:5–16.

Forterre P, Prangishvili D. The origin of viruses. Res Microbiol 2009;160:466–472.

Frattini MG, Laimins LA. Binding of the human papillomavirus E1 origin-recognition protein is regulated through complex formation with the E2 enhancer-binding protein. Proc Natl Acad Sci USA 1994;91:12398–12402.

Fry EE, Newman JW, Curry S, Najjam S, Jackson T, Blakemore W, Lea SM, Miller L, Burman A, King AM, Stuart DI. Structure of Foot-and-mouth disease virus serotype A10 61 alone and complexed with oligosaccharide receptor: receptor conservation in the face of antigenic variation. J Gen Virol 2005;86:1909–1920.

Gao L, Aizaki H, He JW, Lai MM. Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft. J Virol 2004;78:3480–3488.

Gao Q, Kumar A, Srinivasan S, Singh L, Mukai H, Ono Y, Wazer DE, Band V. PKN binds and phosphorylates human papillomavirus E6 oncoprotein. J Biol Chem 2000;275:14824–14830.

Gao Q, Srinivasan S, Boyer SN, Wazer DE, Band V. The E6 oncoproteins of high-risk papillomaviruses bind to a novel putative GAP protein, E6TP1, and target it for degradation. Mol Cell Biol 1999;19:733–744.

Garcia-Alai MM, Alonso LG, de Prat-Gay G. The N-terminal module of HPV16 E7 is an intrinsically disordered domain that confers conformational and recognition plasticity to the oncoprotein. Biochemistry 2007;46:10405–10412.

Genther SM, Sterling S, Duensing S, Munger K, Sattler C, Lambert PF. Quantitative role of the human papillomavirus type 16 E5 gene during the productive stage of the viral life cycle. J Virol 2003;77:2832–2842.

Gewin L, Galloway DA. E box-dependent activation of telomerase by human papillomavirus type 16 E6 does not require induction of c-myc. J Virol 2001;75:7198–7201.

Glaunsinger BA, Lee SS, Thomas M, Banks L, Javier R. Interactions of the PDZ-protein MAGI-1 with adenovirus E4-ORF1 and high-risk papillomavirus E6 oncoproteins. Oncogene 2000;19:5270–5280.

Gloss B, Bernard HU, Seedorf K, Klock G. The upstream regulatory region of the human papilloma virus-16 contains an E2 protein-independent enhancer which is specific for cervical carcinoma cells and regulated by glucocorticoid hormones. Embo J 1987;6:3735–3743.

Goh GK, Dunker AK, Uversky VN. A comparative analysis of viral matrix proteins using disorder predictors. Virol J 2008a;5:126.

Goh GK, Dunker AK, Uversky VN. Protein intrinsic disorder toolbox for comparative analysis of viral proteins. BMC Genomics 2008b;9(Suppl 2): S4.

Goh GK, Dunker AK, Uversky VN. Protein intrinsic disorder and influenza virulence: the 1918 H1N1 and H5N1 viruses. Virol J 2009;6:69.

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci USA 2007;104:8685–8690.

Gramberg T, Sunseri N, Landau NR. Accessories to the crime: recent advances in HIV accessory protein biology. Curr HIV/AIDS Rep 2009;6:36–42.

Gross-Mesilaty S, Reinstein E, Bercovich B, Tobias KE, Schwartz AL, Kahana C, Ciechanover A. Basal and human papillomavirus E6 oncoprotein-induced degradation of Myc proteins by the ubiquitin pathway. Proc Natl Acad Sci USA 1998;95:8058–8063.

Halpern AL, Münger K. HPV-16 E7: Correspondence between primary structure and biological properties, HPV Sequence Database. Los Alamos: Los Alamos National Laboratory; 1995. pp. III–58–III–73.

Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G. Tomato bushy stunt virus at 2.9 Å resolution. Nature 1978;276:368–373.

Henklein P, Bruns K, Sherman MP, Tessmer U, Licha K, Kopp J, de Noronha CM, Greene WC, Wray V, Schubert U. Functional and structural characterization of synthetic HIV-1 Vpr that transduces cells, localizes to the nucleus, and induces G2 cell cycle arrest. J Biol Chem 2000;275:32016–32026.

Holmes EC. Viral evolution in the genomic age. PLoS Biol 2007;5:e278.

Hughes FJ, Romanos MA. E1 protein of human papillomavirus is a DNA helicase/ATPase. Nucleic Acids Res 1993;21:5817–5823.

Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002;323:573–584.

Iakoucheva LM, Kimzey AL, Masselon CD, Bruce JE, Garner EC, Brown CJ, Dunker AK, Smith RD, Ackerman EJ. Identification of intrinsic order and disorder in the DNA repair protein XPA. Protein Sci 2001;10:560–571.

Inoue J, Seiki M, Yoshida M. The second pX product p27 chi-III of HTLV-1 is required for gag gene expression. FEBS Lett 1986;209:187–190.

Inoue J, Yoshida M, Seiki M. Transcriptional (p40x) and post-transcriptional (p27x-III) regulators are required for the expression and replication of human T-cell leukemia virus type I genes. Proc Natl Acad Sci USA 1987;84:3653–3657.

Iyer LM, Balaji S, Koonin EV, Aravind L. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res 2006;117:156–184.

Jian Y, Schmidt-Grimminger DC, Chien WM, Wu X, Broker TR, Chow LT. Post-transcriptional induction of p21cip1 protein by human papillomavirus E7 inhibits unscheduled DNA synthesis reactivated in differentiated keratinocytes. Oncogene 1998;17:2027–2038.

Kashanchi F, Brady JN. Transcriptional and post-transcriptional gene regulation of HTLV-1. Oncogene 2005;24:5938–5951.

Kiyokawa T, Seiki M, Iwashita S, Imagawa K, Shimizu F, Yoshida M. p27x-III and p21x-III, proteins encoded by the pX sequence of human T-cell leukemia virus type I. Proc Natl Acad Sci USA 1985;82:8359–8363.

Kiyono T, Hiraiwa A, Fujita M, Hayashi Y, Akiyama T, Ishibashi M. Binding of high-risk human papillomavirus E6 oncoproteins to the human homologue of the Drosophila discs large tumor suppressor protein. Proc Natl Acad Sci USA 1997;94:11612–11616.

Koonin EV, Senkevich TG, Dolja VV. The ancient Virus World and evolution of cells. Biol Direct 2006;1:29.

Koralnik IJ, Gessain A, Klotman ME, Lo Monico A, Berneman ZN, Franchini G. Protein isoforms encoded by the pX region of human T-cell leukemia/lymphotropic virus type I. Proc Natl Acad Sci USA 1992;89:8813–8817.

Kukimoto I, Aihara S, Yoshiike K, Kanda T. Human papillomavirus oncoprotein E6 binds to the C-terminal region of human minichromosome maintenance 7 protein. Biochem Biophys Res Comm 1998;249:258–262.

Kumar A, Zhao Y, Meng G, Zeng M, Srinivasan S, Delmolino LM, Gao Q, Dimri G, Weber GF, Wazer DE, Band H, Band V. Human papillomavirus oncoprotein E6 inactivates the transcriptional coactivator human ADA3. Mol Cell Biol 2002;22:5801–5812.

Lacroix E, Viguera AR, Serrano L. Reading protein sequences backwards. Fold Des 1998;3:79–85.

Larocca D, Chao LA, Seto MH, Brunck TK. Human T-cell leukemia virus minus strand transcription in infected T-cells. Biochem Biophys Res Comm 1989;163:1006–1013.

Lawrence CM, Menon S, Eilers BJ, Bothner B, Khayat R, Douglas T, Young MJ. Structural and functional studies of archaeal viruses. J Biol Chem 2009;284:12599–12603.

Leavitt SA, SchOn A, Klein JC, Manjappara U, Chaiken IM, Freire E. Interactions of HIV-1 proteins gp120 and Nef with cellular partners define a novel allosteric paradigm. Curr Protein Pept Sci 2004;5:1–8.

Lee SS, Glaunsinger B, Mantovani F, Banks L, Javier RT. Multi-PDZ domain protein MUPP1 is a cellular target for both adenovirus E4-ORF1 and high-risk papillomavirus type 18 E6 oncoproteins. J Virol 2000;74:9680–9693.

Leechanachai P, Banks L, Moreau F, Matlashewski G. The E5 gene from human papillomavirus type 16 is an oncogene which enhances growth factor-mediated signal transduction to the nucleus. Oncogene 1992;7:19–25.

Li S, Labrecque S, Gauzzi MC, Cuddihy AR, Wong AH, Pellegrini S, Matlashewski GJ, Koromilas AE. The human papilloma virus (HPV)-18 E6 oncoprotein physically associates with Tyk2 and impairs Jak-STAT activation by interferon-alpha. Oncogene 1999;18:5727–5737.

Li X, Coffino P. High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation. J Virol 1996;70:4509–4516.

Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC. Structure of simian virus 40 at 3.8-Å resolution. Nature 1991;354:278–284.

Liljas, L. The role of disordered segments in viral coat proteins. In: Cheng, RH, Hammar, L, editors. Conformational proteomics of macromolecular architecture. New Jersy, London, Singapore, Beijing, Shanghai, Hong Kong, Taipei, Chennai: World Scientific Publishing Co.; 2004. pp.53–77.

Lommer BS, Luo M. Structural plasticity in influenza virus protein NS2 (NEP). J Biol Chem 2002;277:7108–7117.

Longhi S. Nucleocapsid structure and function. Curr Top Microbiol Immunol 2009;329:103–128.

Massimi P, Pim D, Banks L. Human papillomavirus type 16 E7 binds to the conserved carboxy-terminal region of the TATA box binding protein and this contributes to E7 transforming activity. J Gen Virol 1997;78(Pt 10):2607–2613.

Matsuoka M, Green PL. The HBZ gene, a key player in HTLV-1 pathogenesis. Retrovirology 2009;6:71.

Mazurek S, Zwerschke W, Jansen-Durr P, Eigenbrodt E. Effects of the human papilloma virus HPV-16 E7 oncoprotein on glycolysis and glutaminolysis: role of pyruvate kinase type M2 and the glycolytic-enzyme complex. Biochem J 2001;356:247–256.

McIntyre MC, Ruesch MN, Laimins LA. Human papillomavirus E7 oncoproteins bind a single form of cyclin E in a complex with cdk2 and p107. Virology 1996;215:73–82.

Midic U, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. Protein disorder in the human diseasome: Unfoldomics of human genetic diseases. BMC Genomics 2009;10(Suppl 1):S12.

Minezaki Y, Homma K, Kinjo AR, Nishikawa K. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. J Mol Biol 2006;359:1137–1149.

Mohan A, Sullivan WJ, Radivojac P, Dunker AK, Uversky VN, Jr. Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. Mol Biosyst 2008;4:328–340.

Mohr IJ, Clark R, Sun S, Androphy EJ, MacPherson P, Botchan MR. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. Science 1990;250:1694–1699.

Morellet N, Roques BP, Bouaziz S. Structure-function relationship of Vpr: biological implications. Curr HIV Res 2009;7:184–210.

Morosov A, Phelps WC, Raychaudhuri P. Activation of the c-fos gene by the HPV16 oncoproteins depends upon the cAMP-response element at -60. J Biol Chem 1994;269:18434–18440.

Nagashima K, Yoshida M, Seiki M. A single species of pX mRNA of human T-cell leukemia virus type I encodes trans-activator p40x and two other phosphoproteins. J Virol 1986;60:394–399.

Nakagawa S, Huibregtse JM. Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. Mol Cell Biol 2000;20:8244–8253.

Nayak DP, Balogun RA, Yamada H, Zhou ZH, Barman S. Influenza virus morphogenesis and budding. Virus Res 2009;143:147–161.

Nayak DP, Hui EK, Barman S. Assembly and budding of influenza virus. Virus Res 2004;106:147–165.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. Proteins 2003;53(Suppl 6):566–572.

Oh ST, Kyo S, Laimins LA. Telomerase activation by human papillomavirus type 16 E6 protein: induction of human telomerase reverse transcriptase expression through Myc and GC-rich Sp1 binding sites. J Virol 2001;75:5559–5566.

Ohlenschlager O, Seiboth T, Zengerling H, Briese L, Marchanka A, Ramachandran R, Baum M, Korbas M, Meyer-Klaucke W, Durst M, Gorlach M. Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. Oncogene 2006;25:5953–5959.

Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK. Comparing and combining predictors of mostly disordered proteins. Biochemistry 2005a;44:1989–2000.

Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 2005b;44:12454–12470.

Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics 2008;9(Suppl 1): S1.

Patel D, Huang SM, Baglia LA, McCance DJ. The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. Embo J 1999;18:5061–5072.

Pim D, Massimi P, Dilworth SM, Banks L. Activation of the protein kinase B pathway by the HPV-16 E7 oncoprotein occurs through a mechanism involving interaction with PP2A. Oncogene 2005;24:7830–7838.

Polverino de Laureto P, De Filippis V, Di Bello M, Zambonin M, Fontana A. Probing the molten globule state of alpha-lactalbumin by limited proteolysis. Biochemistry 1995;34:12596–12604.

Prangishvili D, Forterre P, Garrett RA. Viruses of the Archaea: a unifying view. Nat Rev Microbiol 2006;4:837–848.

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. Biophys J 2007;92:1439–1456.

Rancurel C, Khosravi M, Dunker AK, Romero PR, Karlin D. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. J Virol 2009;83:10719–10736.

Raoult D, Forterre P. Redefining viruses: lessons from Mimivirus. Nat Rev Microbiol 2008;6:315–319.

Rayment I, Baker TS, Caspar DL, Murakami WT. Polyoma virus capsid structure at 22.5 Å resolution. Nature 1982;295:110–115.

Reanney DC. The evolution of RNA viruses. Annu Rev Microbiol 1982;36:47–73.

Reingewertz TH, Benyamini H, Lebendiker M, Shalev DE, Friedler A. The C-terminal domain of the HIV-1 Vif protein is natively unfolded in its unbound state. Protein Eng Des Sel 2009;22:281–287.

Ristriani T, Masson M, Nomine Y, Laurent C, Lefevre JF, Weiss E, Trave G. HPV oncoprotein E6 is a structure-dependent DNA-binding protein that recognizes four-way junctions. J Mol Biol 2000;296:1189–1203.

Ristriani T, Nomine Y, Masson M, Weiss E, Trave G. Specific recognition of four-way DNA junctions by the C-terminal zinc-binding domain of HPV oncoprotein E6. J Mol Biol 2001;305:729–739.

Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Dunker AK. Identifying Disordered regions in proteins from amino acid sequences. IEEE Int Conf Neural Networks 1997;1:90–95.

Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guilliot S, Dunker AK. Thousands of proteins likely to have long disordered regions. Pac Symp Biocomput 1998;437–448.

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42:38–48.

Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proc Natl Acad Sci USA 2006;103:8390–8395.

Ronco LV, Karpova AY, Vidal M, Howley PM. Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity. Genes Dev 1998;12:2061–2072.

Rossmann MG, Mesyanzhinov VV, Arisaka F, Leiman PG. The bacteriophage T4 DNA injection machine. Curr Opin Struct Biol 2004;14:171–180.

Rybicki EP. The classification of organisms at the edge of life, or problems with virus systematics. S Afr J Sci 1990;86:182–186.

Sano E, Carlson S, Wegley L, Rohwer F. Movement of viruses between biomes. Appl Environ Microbiol 2004;70:5842–5846.

Scheffner M, Huibregtse JM, Vierstra RD, Howley PM. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell 1993;75:495–505.

Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell 1990;63:1129–1136.

Sedman SA, Barbosa MS, Vass WC, Hubbert NL, Haas JA, Lowy DR, Schiller JT. The full-length E6 protein of human papillomavirus type 16 has transforming and trans-activating activities and cooperates with E7 to immortalize keratinocytes in culture. J Virol 1991;65:4860–4866.

Seelamgari A, Maddukuri A, Berro R, de la Fuente C, Kehn K, Deng L, Dadgar S, Bottazzi ME, Ghedin E, Pumfery A, Kashanchi F. Role of viral regulatory and accessory proteins in HIV-1 replication. Front Biosci 2004;9:2388–2413.

Seiki M, Hattori S, Hirayama Y, Yoshida M. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc Natl Acad Sci USA 1983;80:3618–3622.

Seiki M, Inoue J, Hidaka M, Yoshida M. Two cis-acting elements responsible for post-transcriptional trans-regulation of gene expression of human T-cell leukemia virus type I. Proc Natl Acad Sci USA 1988;85:7124–7128.

Seiki M, Inoue J, Takeda T, Yoshida M. Direct evidence that p40x of human T-cell leukemia virus type I is a trans-acting transcriptional activator. EMBO J 1986;5:561–565.

Shojania S, O'Neil JD. HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1–72) by NMR spectroscopy. J Biol Chem 2006;281:8347–8356.

Skehel JJ, Bayley PM, Brown EB, Martin SR, Waterfield MD, White JM, Wilson IA, Wiley DC. Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion. Proc Natl Acad Sci USA 1982;79:968–972.

Skehel JJ, Wiley DC. Coiled coils in both intracellular vesicle and viral membrane fusion. Cell 1998;95:871–874.

Skehel JJ, Wiley DC. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem 2000;69:531–569.

Sodroski J, Trus M, Perkins D, Patarca R, Wong-Staal F, Gelmann E, Gallo R, Haseltine WA. Repetitive structure in the long-terminal-repeat element of a type II human T-cell leukemia virus. Proc Natl Acad Sci USA 1984;81:4617–4621.

Srivenugopal KS, Ali-Osman F. The DNA repair protein, O(6)-methylguanine-DNA methyltransferase is a proteolytic target for the E6 human papillomavirus oncoprotein. Oncogene 2002;21:5940–5945.

Stevens J, Corper AL, Basler CF, Taubenberger JK, Palese P, Wilson IA. Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. Science 2004;303:1866–1870.

Straight SW, Hinkle PM, Jewers RJ, McCance DJ. The E5 oncoprotein of human papillomavirus type 16 transforms fibroblasts and effects the downregulation of the epidermal growth factor receptor in keratinocytes. J Virol 1993;67:4521–4532.

Surendran R, Herman P, Cheng Z, Daly TJ, Ching Lee J. HIV Rev self-assembly is linked to a molten-globule to compact structural transition. Biophys Chem 2004;108:101–119.

Suzuki Y. Sialobiology of influenza: molecular mechanism of host range variation of influenza viruses. Biol Pharm Bull 2005;28:399–408.

Thomas M, Banks L. Human papillomavirus (HPV) E6 interactions with Bak are conserved amongst E6 proteins from high and low risk HPV types. J Gen Virol 1999;80(Pt 6):1513–1517.

Thomas MC, Chiang CM. E6 oncoprotein represses p53-dependent gene activation via inhibition of protein acetylation independently of inducing p53 degradation. Mol Cell 2005;17:251–264.

Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS. Do viral proteins possess unique biophysical features? Trends Biochem Sci 2009;34:53–59.

Tommasino M, Adamczewski JP, Carlotti F, Barth CF, Manetti R, Contorni M, Cavalieri F, Hunt T, Crawford L. HPV16 E7 protein associates with the protein kinase p33CDK2 and cyclin A. Oncogene 1993;8:195–202.

Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–533.

Tompa P. The functional benefits of protein disorder. J Mol Struct - Theochem 2003; 666–667:361–371.

Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN. Close encounters of the third kind: disordered domains and the interactions of proteins. Bioessays 2009;31:328–335.

Tong X, Boll W, Kirchhausen T, Howley PM. Interaction of the bovine papillomavirus E6 protein with the clathrin adaptor complex AP-1. J Virol 1998;72:476–482.

Tong X, Howley PM. The bovine papillomavirus E6 oncoprotein interacts with paxillin and disrupts the actin cytoskeleton. Proc Natl Acad Sci USA 1997;94:4412–4417.

Ustav M, Stenlund A. Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames. EMBO J 1991;10:449–457.

Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002a;11:739–756.

Uversky VN. What does it mean to be natively unfolded? Eur J Biochem 2002b;269:2–12.

Uversky VN. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? Cell Mol Life Sci 2003;60:1852–1871.

Uversky VN. Amyloidogenesis of natively unfolded proteins. Curr Alzheimer Res 2008a;5:260–287.

Uversky, VN. Intrinsic disorder in proteins associated with neurodegenerative diseases. In: Ovádi J, Orosz F, editors. Protein folding and misfolding: neurodegenerative diseases. New York, USA: Springer; 2008b.

Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41:415–427.

Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recogn 2005;18:343–384.

Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. Annu Rev Biophys Biomol Struct 2008;37:215–246.

Uversky VN, Roman A, Oldfield CJ, Dunker AK. Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. J Proteome Res 2006;5:1829–1842.

Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. BMC Bioinformatics 2007;8:211.

Vendel AC, Lumb KJ. Molecular recognition of the human coactivator CBP by the HIV-1 transcriptional activator Tat. Biochemistry 2003;42:910–916.

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 2004;337:635–645.

Weber T, Paesold G, Galli C, Mischler R, Semenza G, Brunner J. Evidence for H(+)-induced insertion of influenza hemagglutinin HA2N-terminal segment into viral membrane. J Biol Chem 1994;269:18353–18358.

Wharton SA, Calder LJ, Ruigrok RW, Skehel JJ, Steinhauer DA, Wiley DC. Electron microscopy of antibody complexes of influenza virus haemagglutinin in the fusion pH conformation. EMBO J 1995;14:240–246.

Wiley DC, Skehel JJ. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Annu Rev Biochem 1987;56:365–394.

Wilson IA, Skehel JJ, Wiley DC. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 1981;289:366–373.

Wilson R, Fehrmann F, Laimins LA. Role of the E1–E4 protein in the differentiation-dependent life cycle of human papillomavirus type 31. J Virol 2005;79:6732–6740.

Wray V, Federau T, Henklein P, Klabunde S, Kunert O, Schomburg D, Schubert U. Solution structure of the hydrophilic region of HIV-1 encoded virus protein U (Vpu) by CD and 1H NMR spectroscopy. Int J Pept Protein Res 1995;45:35–43.

Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999;293:321–331.

Wright PE, Dyson HJ. Linking folding and binding. Curr Opin Struct Biol 2009;19:31–38.

Xue B, Li L, Meroueh SO, Uversky VN, Dunker AK. Analysis of structured and intrinsically disordered regions of transmembrane proteins. Mol Biosyst 2009a;5:1688–1702.

Xue B, Oldfield CJ, Dunker AK, Uversky VN. CDF it all: consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions. FEBS Lett 2009b;583:1469–1474.

Xue B, Williams RW, Oldfield CJ, Goh GK, Dunker AK, Uversky VN. Viral disorder or disordered viruses: do viral proteins possess unique features? Protein Pept Lett 2010;17:932–951.

Zhang B, Li P, Wang E, Brahmi Z, Dunn KW, Blum JS, Roman A. The E5 protein of human papillomavirus type 16 perturbs MHC class II antigen maturation in human foreskin keratinocytes treated with interferon-gamma. Virology 2003;310:100–108.

Zimmermann H, Degenkolbe R, Bernard HU, O'Connor MJ. The human papillomavirus type 16 E6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator CBP/p300. J Virol 1999;73:6209–6219.

Zwerschke W, Mazurek S, Massimi P, Banks L, Eigenbrodt E, Jansen-Durr P. Modulation of type M2 pyruvate kinase activity by the human papillomavirus type 16 E7 oncoprotein. Proc Natl Acad Sci USA 1999;96:1291–1296.