Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

CHAPTER ONE

Intrinsically disordered proteins of viruses: Involvement in the mechanism of cell regulation and pathogenesis

Pushpendra Mani Mishra[a], Navneet Chandra Verma[a], Chethana Rao[a], Vladimir N. Uversky[b,c,*], Chayan Kanti Nandi[a,*]

[a]School of Basic Sciences, Indian Institute of Technology, Mandi, Mandi, Himachal Pradesh, Mandi, India
[b]Department of Molecular Medicine and Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, United States
[c]Laboratory of New Methods in Biology, Institute for Biological Instrumentation of the Russian Academy of Sciences, Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino, Moscow Region, Russia
*Corresponding authors: e-mail address: vuversky@health.usf.edu; chayan@iitmandi.ac.in

Contents

Progress in Molecular Biology and Translational Science, Volume 174
ISSN 1877-1173
https://doi.org/10.1016/bs.pmbts.2020.03.001

Abstract

Intrinsically disordered proteins (IDPs) possess the property of inherent flexibility and can be distinguished from other proteins in terms of lack of any fixed structure. Such dynamic behavior of IDPs earned the name "*Dancing Proteins*." The exploration of these dancing proteins in viruses has just started and crucial details such as correlation of rapid evolution, high rate of mutation and accumulation of disordered contents in viral proteome at least understood partially. In order to gain a complete understanding of this correlation, there is a need to decipher the complexity of viral mediated cell hijacking and pathogenesis in the host organism. Further there is necessity to identify the specific patterns within viral and host IDPs such as aggregation; Molecular recognition features (MoRFs) and their association to virulence, host range and rate of evolution of viruses in order to tackle the viral-mediated diseases. The current book chapter summarizes the aforementioned details and suggests the novel opportunities for further research of IDPs senses in viruses.

Abbreviations

AFV1 acidianus filamentous virus 1
APC/C anaphase-promoting complex/cyclosome
APMV *Acanthamoeba polyphaga* mimivirus
ATL adult T-cell leukemia
CBP CREB-binding protein
CDKs cyclin-dependent kinases

CypA cyclophilin A
DENV dengue virus
DNA deoxyribonucleic acid
DSC differential scanning spectroscopy
E1B early transcription unit 1B
EIAV equine Infectious anemia Virus
EMI-1 early mitotic inhibitor protein-1
EPR electron paramagnetic resonance
EV-71 enterovirus-71
FMDV foot-and-mouth disease virus
FRET fluorescence resonance energy transfer
HA hemagglutinin
HDAg hepatitis delta antigen
HDV hepatitis delta virus
HIV human immunodeficiency virus
HIV-1 human immunodeficiency virus-1
HPV human papillomavirus
HS-AFM high-speed atomic force microscope
HTLV-1 human T-cell lymphotropic virus type 1
IDP intrinsically disordered protein
IDR intrinsically disordered region
IRF3 interferon regulatory factor 3
JEV Japanese encephalitis
KID kinase-inducible domain
KIX kinase-inducible domain (KID) interacting domain
LMV lettuce mosaic virus
MAP 2 microtubule-associated protein 2
Mdm2 mouse double minute 2
MeV measles virus
MG molten globule
MHC major histocompatibility complex
MoRF molecular recognition feature
MPs movement proteins
MVM minute virus of mice
NCBD nuclear-receptor co-activator binding domain
NEP non-structural protein 2
NLP nucleocapsid like particle
NLS nuclear localization signal
NMR nuclear Magnetic Resonance
NTAIL C-terminal domain of nucleoprotein
ORF open reading frame
PACS phosphofurin acidic cluster sorting protein
PI isoelectric point
PKB protein kinase B
PMG pre-molten globule
PNT N-terminal region of P
PP2A protein phosphatase 2A
PPI protein-protein interaction

PPID	predicted percentage of intrinsic disorder
PTM	posttranslational modification
PVA	potato virus A
PVS	papillomavirus
PVY	potato virus Y
RC	random coil
RDRP	RNA dependent RNA polymerase
RNA	ribonucleic acid
RRE	rev. response element
RYMV	rice yellow mottle virus
SARS-CoA	severe acute respiratory syndrome coronavirus
SAXS	small-angle X-ray scattering
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SFV	Semliki forest virus
SIFV	*Sulfolobus islandicus* filamentous virus
SLiM	short linear motif
SM	single-molecule
SM-FRET	single-molecule fluorescence resonance energy transfer
SeMV	sesbania mosaic virus
SeV	Sendai virus
SRC1	steroid receptor coactivator 1
TAD	transactivation domain
TAR	transactivation response region
TGB	triple gene black
TMV	tobacco mosaic virus
URR	upstream regulatory region
VPg	viral genome-linked protein
VSV	vesicular stomatitis virus
WNV	West Nile virus
XD	C-terminal X domain
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
pRb	retinoblastoma protein
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
ZIKV	Zika virus

Rationale and importance of the book chapter

This book chapter entitled "Intrinsically disordered proteins of viruses: involvement in the mechanism of cell regulation and pathogenesis" discusses extensively the intrinsically disordered protein (IDP)-mediated functional mechanisms, pathogenesis, structural regulation and cellular regulation of host cell by complex viral proteome. For a complete

understanding of IDPs and their role in Viruses, this chapter starts with the brief introuction of IDPs and their associated atypical properties and different instrumental and computational techniques to characterize IDPs. Next, chapter describes the IDP-related aspect of viruses. Different possible modes of viral IDP molecular mimicry and host IDP-mediated regulation of host cells have been discussed and a diagrammatic model is proposed. Subsequently, the origin of viruses and their special properties have been described. Further, the importance of viral structural, non-structural and other proteins is emphasized. Furthermore, the IDP prevalence in viruses and their comparison to three distinct domains of life (Archaea, Bacteria, and Eukarya) are discussed in detail. The last portion of this book chapter explains various IDP-associated patterns in viruses and their relation to the host range, pathogenicity, and protein aggregation. Next, the structural and functional importance of IDPs in different viruses (Bacteriophage, Plant and Animal virus) is discussed. The examples of the aforementioned viruses and description of their IDP-associated mechanisms have been taken from the different referenced publications. Lastly, this chapter summarizes the conversed contents and further discusses the future outlook for the purpose of studying IDP prevalence, distribution, and disorder-related mechanisms in the proteomes of viruses.

We hope that this chapter will help in grasping the concept of IDPs and IDPs' perspective of viruses and spawning many novel ideas in relation to deciphering the complexity of viral pathogenesis and drug discovery. For instance, the prevalence of IDPs and patterns of pathogenesis and host range have been explored and proven in a few viruses; however other related patterns have not been explored completely. Additionally, the mechanisms of cell regulation via disordered viral proteome have not been completely understood. The proposed model will form the basis for further research and understanding.

By authors.

1. A general introduction to intrinsically disordered proteins (IDPs) and their major properties

1.1 Intrinsically disordered proteins (IDPs)

The concept of structure-function paradigm that was widely accepted for more than a century tells us that the biological functions of proteins are linked to their rigid three-dimensional (3D) structures.[1]

The normal functioning of most of the globular proteins (e.g. enzymes) requires the orderly arrangement of various functional groups of amino acids in protein's unique 3D structure to facilitate the catalysis of chemical reactions or other related functions. However, recent research demonstrated that the large fraction of genome-encoded proteins of many organisms lack the well-defined 3D structures, but still play various important roles in cellular functionality. The group of such proteins is generally known as intrinsically disordered proteins (IDPs).[2–4] However, they have multiple alternative names, such as natively denatured,[5] natively unfolded,[6] intrinsically unstructured,[7] natively disordered,[8] dancing proteins,[9] protein clouds,[10,11] 4D,[12] malleable,[13–15] chameleon,[16] vulnerable,[17] intrinsically disordered,[18] intrinsically unfolded, intrinsically denatured, flexible,[19] mobile,[20] pliable,[21] rheomorphic,[22] and partially folded proteins.[23] The different name identities of IDPs are based on their properties observed in different experiments conducted at a different time. The computational analysis reveals that greater than one-third of eukaryotic proteins harbor the intrinsically disordered regions (IDRs) of greater than 30 residues in length.[24–31] In solution, when IDPs are kept alone, they lack a unique 3D structure either in parts or completely.[32]

The high abundance of IDPs is associated with their functional importance for many crucial cellular processes, such as signaling, recognition, and regulation by means of high specificity and low-affinity interaction and binding to multiple partners. The disorder-based signaling interaction can be mediated as many to one and one to many interactions. The functional tuning of IDPs induced by various post-translational modifications (PTMs), Alternative splicing and induced folding. The high prevalence of IDPs in various diseases suggests the root cause is not only the protein misfolding but beyond it and also caused by mis-signaling and misidentification. The peculiar behavior of IDPs draws attention to drug targets, which temper the protein-protein interactions.[8]

1.2 Properties of IDPs

1.2.1 Sequence of amino acid define disorderliness

Although the IDPs are biologically active molecules, they tend to adopt an extended mobile dynamic or collapsed conformational ensemble either at the tertiary or secondary structure. A comparative analysis of amino acid sequence of IDPs with respect to those of ordered proteins demonstrate the noticeable enrichment in the content of disorder-promoting amino acids, such as Ala, Arg, Gly, Gln, Ser, Glu, Lys, and Pro, paralleled by

the significant depletion in the content of order-promoting amino acids, Ile, Leu, Val, Trp, Tyr, Phe, Cys, and Asn. In addition to the above-observed criteria, several other disorder-promoting factors are involved that are; 14 Å contact numbers, coordination number, hydropathy, Cys + Phe + Tyr + Trp, volume, Arg + Glu + Ser + Pro, bulkiness, net charge and β-sheet propensity that provide the reliable basis for differentiating disorder and other proteins.[18,33–36]

1.2.2 Intrinsic disorder and binding promiscuity

One of the important properties of IDPs is being promiscuous in nature. This involves interaction with multiple partners and ability to act as highly connected nodes, or hubs, most frequently within the protein-protein interaction (PPI) networks. Hubs are vital for the normal functioning and stability of PPI networks in any organism. It has been shown that the deletion of hub protein could be lethal for that organism.[37–45] The illustrative examples of disordered hub proteins that bind to around 10–100 binding partners are p21, p27, p53, BRCA1, XPA, **α**-synuclein, estrogen receptor,[46] etc. IDRs within disordered hub protein present in at least one of the two functional forms; the one functional form defines the ability of the disordered binding site to interact with specific partner and, upon interaction, to adopt an ordered conformation, another functional form is a flexible linker that connects two ordered domain and allows their unrestricted movement.[47]

1.2.3 Properties of charge

The presence of charge in amino acid residues helps in establishing the structure and function of proteins. The high content of charged residue in the highly disordered proteins (native pre-molten globules (PMG) and native coils) is an important, conspicuous feature.[48] The high net charge is important for extended conformation of IDPs,[49] because it has been observed repeatedly for proteins in aqueous environment, that the sequences lacking in certain hydrophobic residues and rich in polar uncharged amino acids form the heterogeneous ensemble of collapsed structures.[50–56] Analysis of the number of highly charged polypeptides revealed that the intrinsic preference of a polypeptide backbone for the formation of collapsed structure depends on charge content.[49]

1.2.4 Human disease association

The analysis of various protein databases of human diseases and other observations determine the contribution of IDPs to the pathogenesis of many

human ailments and their role as common player in between the diseases.[8,57] Few examples of diseases, where IDPs/IDRs are involved are listed below. Cancer: Intrinsic disorder has been observed in many cancer-associated proteins, such as p53,[58] p57kip2,[59] c-Fos,[60] Bcl-2 and Bcl-XL,[61] thyroid cancer-associated protein TC-1,[62] and protein components of cancer-causing viruses.[1] Down's syndrome: Non-filamentous deposits of intrinsically disordered amyloid-β (Aβ).[63] Alzheimer's disease: IDPs associated with this disease are depositions of Aβ, Tau, and α-synuclein NAC fragment.[64–67] Other diseases, where intrinsic disorder in protein components was reported are family of polyQ diseases[68]; variant of Alzheimer's disease, dementia with Lewy body, diffuse Lewy body disease, Parkinson's disease, Hallervorden–Spatz disease and multiple system atrophy[69]; prion disease[70]; argyrophilic grain disease, myotonic dystrophy, and motor neuron disease with neurofibrillary tangles, subacute sclerosing panencephalitis, Niemann-Pick disease type C.[66] Intrinsic disorder is also reported in protein components of viruses causing various human diseases, such as AIDS and Cutaneous diseases.[71]

1.2.5 Ensemble structures of IDPs

IDPs do not have specific fixed structures, hence they exist as dynamic ensembles, quite similar to the clouds of proteins. In these protein cloud structures, the atomic position and the backbone Ramachandran angle does not have the fixed value and vary significantly over time. Despite being dynamic in nature, these protein clouds could be represented by a fairly limited number of low-energy conformations (but still significantly more than one low-energy state typical for ordered proteins).[10,72,73] To understand the regulatory mechanism and cellular functions involvement, structural details of IDPs are necessary. Various methods have been developed to construct the ensemble modeling of IDPs.[74–76]

1.2.6 Hydration property

Due to the difference in structure and structure-associated properties, ordered and disordered proteins possess different hydration degrees. The degree of hydration is significantly higher for the IDPs in comparison to the similar size globular proteins. Furthermore, the hydration degree also varies for the partially and fully intrinsically disordered proteins.[77–79] In addition to retaining a high amount of water content, IDPs also possess a high propensity of binding to charged solute ions. Both properties play an important protective role in biological systems. For example, under the adverse water-stressed conditions, *D. radioduran* is able to protect its enzyme nudix hydrolase from denaturation

due to the aforementioned properties of the IDRs of this protein.[80] Several plants and free-living insect species also protect themselves by using the ability of IDPs and IDRs for excessive hydration and absorption of solute ion.[46]

1.2.7 Property of induced folding

Many IDPs can undergo (at least partial) disorder-to-order transitions upon binding to the specific partners. The free energy required for the transition comes from the interface contacts, which results in the formation of low net free energy association for the high specific interaction combination.[18,38,39,81,82] In IDPs/IDRs, coupled properties of high specificity and low affinity seems to ensure specific binding and reversibility to complete the signaling cascade.[46] IDPs/IDRs can change their shapes to readily bind multiple different partners. Also, it has been shown that in their unbound conformational ensembles, IDPs/IDRs have a preference for the structure they most likely to adopt after binding.[81,83,84]

1.2.8 Interactability of IDPs

Interactions of IDPs with their partners are characterized by a diverse range of binding modes, due to which the formation of many unusually shaped complexes takes place, with some of these complexes being relatively static hence their structure could be determined by the x-ray crystallography method.[11] The most common binding modes of IDPs that have been studied extensively relative to others are *Molecular Recognition Features* (*MoRFs*). MoRFs are intrinsically disordered protein segments, which are short and interaction-prone. These regions also have intrinsic propensity for order, which is not strong enough to ensure their folding in the unbound state. However, upon binding to specific partners, MoRFs undergo disorder-to-order transition. Such regions are chiefly involved in molecular recognition. The classification of MoRFs is based on their structures in the bound state. As a result, they are classified into α-helix-forming α-MoRFs, β-strand forming β-MoRFs, ordered regions without any regular structure or irregular ι-MoRFs, and complex MoRF that contain two or more types of secondary structure.[85,86] In addition to MoRFs, other known binding modes are Pullers,[87] Penetrators,[88] Flexible Wrapper,[89–92] Connectors and Armature,[93–97] Huggers,[98–100] Stackers or β-Arcs,[101] Intertwined Strings,[102–104] Long Cylindrical Containers,[105] Tweezers and a Forceps,[106] Grabbers,[107] Tentacles,[108] and Chameleons.[16,109–112]

1.3 Roles of IDPs in protein interaction and PPI networks

IDP/IDR can play its roles by contributing to the binding diversity in three different ways, as it may serve as the structural basis for hub protein promiscuity, secondly, it may bind to structured hub proteins, and thirdly, IDR can act as a flexible linker between the functional domains and facilitate the binding diversity through the linker-enabling mechanism.[38] A vast range of functional importance of IDPs/IDRs has been found by the researchers. Few examples are given here to illustrate the type of biological activities carried by the IDPs/IDRs. (1) IDPs contain sites for various posttranslational modifications (PTMs), such as phosphorylation, methylation, glycosylation, ADP-ribosylation or acetylation; (2) Entropic spring (rubber-like) property can be provided by IDRs; (3) IDPs contain autoinhibitory domains; (4) IDPs/IDRs possess binding sites for DNA, rRNA, mRNA, tRNA, metal ions, and other proteins; (5) IDRs include regulatory protease digestion site; (6) Signal for the nuclear localization is located within IDRs; (7) IDRs provide flexible linkers between structured domains[112]; and (8) IDPs, such as p21 and p27, mediate cell regulation.[113] Fig. 1 provides details of the involvement of IDPs in crucial cellular functions and processes.

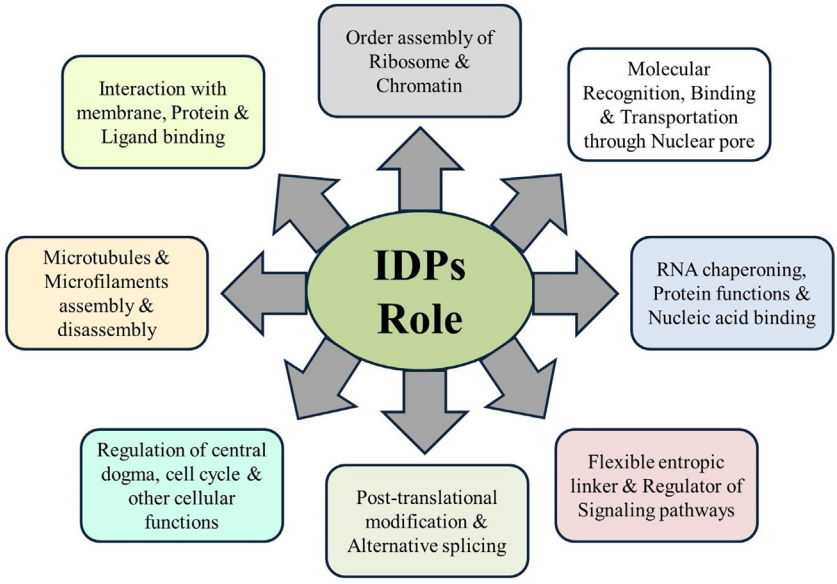

Fig. 1 IDPs involvement in various cellular processes.[114,115]

1.4 Predictors of intrinsic disorder

The compositional differences between ordered proteins and IDPs facilitated the development of various disorder predictors. These predictors were initially elaborated based on amino acid composition. Later, the predictors were developed on the basis of some basic physical principles and machine learning algorithms, which use the characteristic features of IDPs/IDRs, such as net charge, hydrophobicity, and other sequence features. As of 2009, more than 50 predictors for intrinsic disorder prediction have been developed and published,[116] and currently, this list is likely to be doubled. There are the good chances for the development of improved predictors for intrinsic disorder, if the proper sequence information is encoded into the prediction algorithm. The example of few common predictors are as follows: various members of PONDR family,[34] DISOPRED,[117] FoldIndex,[118] IUPRED,[119] DisEMBL,[120] DISOPRED2,[117] and RONN[121,122] to name a few.

1.5 Structural assessment of IDPs through biophysical techniques

There are three functional conformational states, in which IDPs could globally exist, depending upon the environment and content of residual structure. These are, in a range of the increasing depth of disorder, molten globule (MG), pre-molten globule (PMG), and random-coil-like (RC-like) states. Therefore, IDPs could adopt either extended conformations (RC and PMG) or remain globally collapsed (MG).[123]

So far, the conformational and spectroscopic study of IDPs confirmed the important notion that the IDPs could not be represented by a homogeneous structural class, but it would be in the range of fully extended (RC-like) to compact (MG-like) conformations. Protein trinity hypothesis given by the Keith Dunker to accommodate three most known conformations of a protein molecule in a functional framework, which postulated that there a biologically active protein molecule can exist in three conformationally different native states, an ordered form, a state with collapsed-disorder (molten globule, MG) and a state with extended disorder (RC). Functional form is represented by any of the three conformations or transitions between them. Subsequently, this model was extended to accommodate an extra conformation that is the PMG, which is an intermediate conformation between MG and RC.[18]

Many biophysical techniques can be applied for the conformational analysis and structure determination of IDPs. Some of these techniques provide

outputs in an indirect way, while others are useful in providing more quantitative structural data. Nuclear Magnetic Resonance (NMR) is one of the most powerful techniques for deriving quantitative structural information.[124] A wide line NMR relaxation experiment characterizes the IDPs and provides details about the presence of the hydrated layer in the vicinity of disordered regions in the extended and open state. Additionally, the diffusion coefficient of protein can be measured by the pulse field gradient NMR, from which the hydrodynamic parameters could be derived.[4] Structural transition in IDPs can be mapped and documented by the electron paramagnetic resonance (EPR) spectroscopy. The introduction of new generation spin-labels EPR that target the residues other than the cysteine expanded the approach of this technique.[125–129] Small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS), which are the experimental techniques for the extraction of quantitative information, lead to an investigation of transient intermediates and provide detailed information about the nature of IDPs. The techniques of a single-molecule approach such as fluorescence resonance energy transfer (FRET),[130–132] High-Speed Atomic Force Microscope (HS-AFM),[133,134] and AFM-based force spectroscopy (FS)[135] are the tools to explore the dynamics and structure of IDPs. The change in distance between two residues and study of conformational equilibria in time length of less than a second based on the intramolecular distance distribution is done by the Single-molecule fluorescence resonance energy transfer (SM-FRET). Formation of secondary structures and probing of time scales from milliseconds to seconds is particularly sensed by AFM-based SM-FS. HS-AFM is used for the direct observation of dynamic processes and structural dynamics of biological molecules, with the temporal resolution of subsecond to sub-100 ms.[134,136] To date, the various dynamic processes have been visualized successfully by using this approach. HS-AFM is applicable to both IDPs and well-structured protein.

Various other complementary methods that can be used to study protein disorder are sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), gel filtration or size exclusion chromatography-based analysis, and specific behavior analysis in acidic and high-temperature environments. In the SDS-PAGE analysis, the observed mobility of IDPs appears to be anomalous. This phenomenon is explained by the less efficient binding of SDS molecules to highly charged IDPs in comparison to the globular proteins of similar molecular masses. The apparent molecular mass determine by this method is up to 1.2- to 1.8-fold higher than the molecular mass determined from the protein sequences or by mass spectrometry.[137]

The unusually high apparent molecular mass of IDPs is also observed by gel filtration or size exclusion chromatography techniques.[123] The specific behavior of IDPs in different sets of environmental conditions, such as their stability in an acidic environment and insensitivity to high temperature, has been described for several IDPs, such as caldesmon,[138] microtubule-associated protein-2 MAP2,[139] involucrin,[140] and α-synuclein[6] to name a few. These environmental conditions usually cause the denaturation or/and precipitation of globular proteins out of solution. This difference in the behavior of IDPs and globular protein in the various sets of environmental condition form the basis of purification of IDPs.[141–144] This uniqueness of IDPs provides the first clue of their unusual structural conformation.[145]

2. The dark proteomes of viruses

IDPs offer high flexibility to viral proteins[146] either in the wholly or partially disordered form. This provides viral proteins with the capability for quick adaption in the changing environment, survival in host body environments, and invasion of the defense mechanism of the host. To accomplish aforementioned tasks, a high mutation rate is exhibited by the viral genomes. For example, the rate of nucleotide exchange per position per generation exhibited by ribonucleic acid (RNA) viruses fall in the range of 10^{-5} to 10^{-3}, for deoxyribonucleic acid (DNA) viruses it is 10^{-8} to 10^{-5}, while eukaryotes and bacteria demonstrate mutation rate of 10^{-9}.[147] Even a single mutation has high potency to affect more than one viral protein, due to high compactness of viral genomes and the existence of the overlapping reading frames, which often is observed in the viral system. Throughout the life cycle of the virus, many interactions are made to various components of the host cell. It begins with the attachment, entry, and proceeding for the hijacking of the cellular machinery and further viral components synthesis, viral particle assembly, and end by exiting the host cell in the form of new infectious particles.[148] And all these stages are heavily relying on the intrinsic disorder of viral proteins.[148]

3. Involvement of IDPs in the pathogen-host mediated regulation of cell cycle

There are numerous pathways in host cells, where IDPs are involved in the controlling of the cell cycle. Several illustrative examples are discussed here and shown in Fig. 2. (1) Preformed helical portion of the

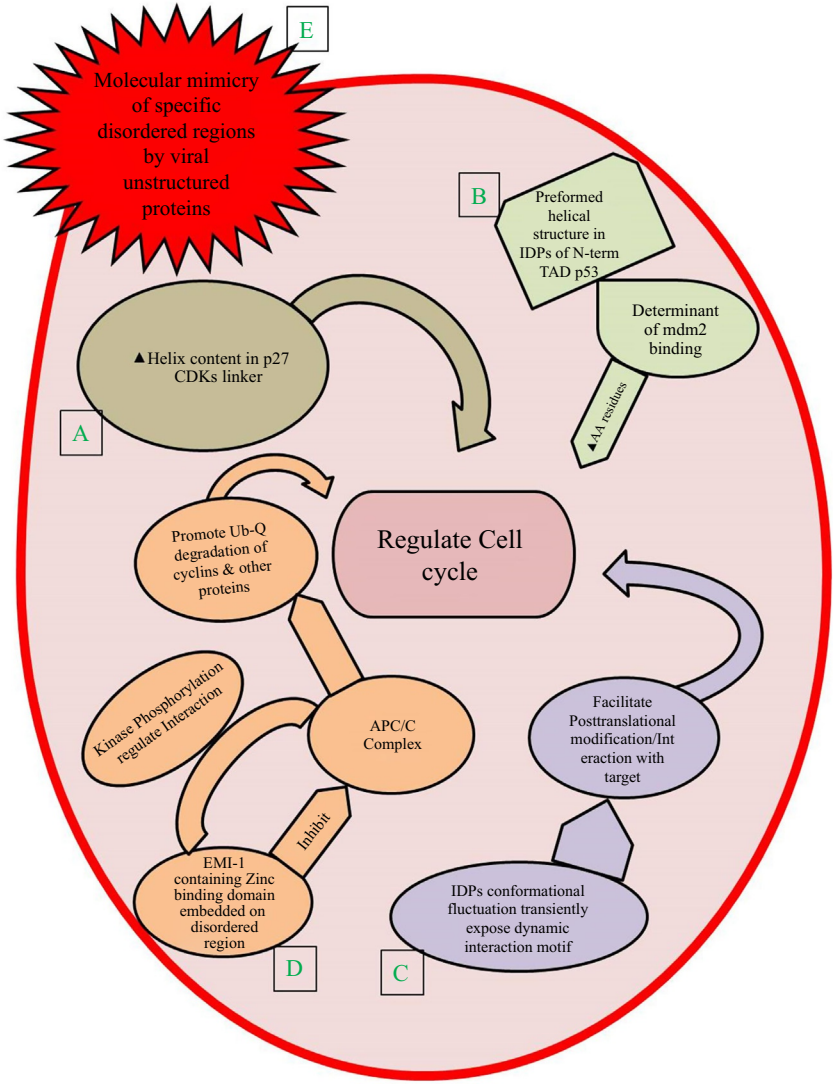

Fig. 2 Intrinsic disorder controlled (A) a structural change in the cyclin-dependent kinase inhibitor 1B ($p27^{Kip1}$); (B) TADp53-mediated Mdm2 binding have an association with the cell cycle regulation and apoptosis. (C) Conformational fluctuation and transient exposure of interaction motif lead to PTM that subsequently controls the target binding and ultimately regulation of cell cycle. (D) IDRs control zinc-binding domain of EMI-1 that upon interaction with APC/C complex controls UBQ-mediated degradation of cyclins and other proteins that are related to the cell cycle regulation. (E) Viruses hijack the cellular machinery using one or more cell regulation pathways by using their proteins to mimic the host IDPs/IDRs in cell cycle pathways.

cyclin-dependent kinase inhibitor 1B ($p27^{Kip1}$) protein is associated with positive and negative regulation of cell cycle.[149] (2) Preformed helical structure in disordered N-terminal transactivation domain (TAD) of p53 determines interaction of this protein with Mouse double minute 2 homolog (Mdm2), any change in amino acid residues in the molecular determinant region affects the binding of Mdm2 and subsequently cell regulation and apoptosis. (3) Conformational fluctuation in the intrinsically disordered cell proteins transiently exposes dynamic interaction motif that leads to post-translational modifications (PTMs) and interaction with various target protein that affects cell cycle control. (4) Early mitotic inhibitor protein-1 (EMI-1) containing zinc-binding domain embedded on IDPs inhibits anaphase-promoting complex/cyclosome (APC/C) that controls cell division by promoting ubiquitin-mediated degradation of cyclins and other proteins involved in the regulation of the cell cycle. Kinase phosphorylations regulate the interaction between EMI-1 and APC/C.[114] Intrinsically Unstructured viral protein components, through molecular mimicry, could invade the host IDPs position involved in various cell regulatory processes (few of them as discussed above) and hijack host cell machinery.[114,149–152]

In addition to the aforementioned pathways, viruses through histone mimicry can control the expression of the gene and ultimately cell cycle regulations network of host cells.[153] Many unconventional RNA binding proteins containing IDRs can also play an important role in the control of cellular machinery by viruses in the battlefront of host and pathogens.[154] Besides mimicking the function of host cells, viral protein complex directly attacks the cellular components and disrupt their normal functions, for instance, the disordered viral oncoproteins of many cancer-causing viruses attack the Retinoblastoma protein (pRb) and E2F complex and affect the normal cell regulation mechanism as shown in Fig. 3.

Phosphofurin acidic cluster sorting protein (PACS) acting as a traffic modulator first appeared in lower metazoans. Later evolution of this protein in vertebrate makes integration of cytoplasmic trafficking and inter-organellar communication with nuclear gene expression. In due course of evolution, PACS functional diversity increases in the vertebrate by acquiring the phosphorylation sites and nuclear trafficking signals within its disordered regions. PACS proteins variants PACS-1 and PACS-2 mediated protein trafficking pathways hijacked by viruses for immune evasion, multiplication, and pathogenesis. However, the complete mechanism is yet to decipher.[156]

To accomplish all the above functions, viral proteins made interactions with many cellular components, with few of them being Nucleic acids,

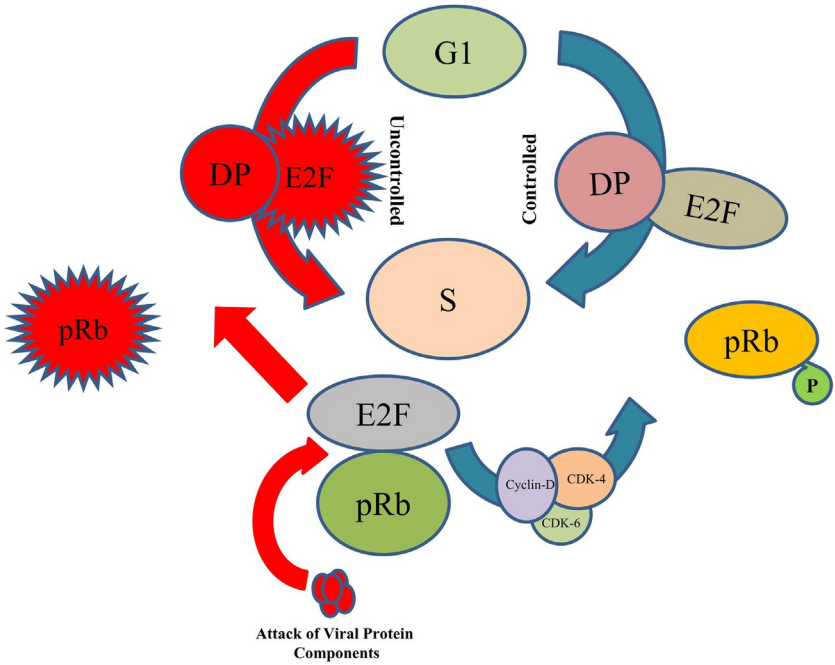

Fig. 3 Host cell cycle regulation influenced by the attack of viral protein components on pRb and E2F complex. Viral protein complex forcibly releases the E2F from pRb and E2F complex and abruptly increases the cell cycle progression in an uncontrolled way. The blue color shows the normal pathway of G1 to S progression, while red color shows virus-induced uncontrolled cell progression from G1 to S.[155]

Proteins, and Membranes. The presence of intrinsic disorder in viral proteins is advantageous for their interactions with the host cell components. The easiness of said interactions could be explained on the basis of the lack of rigid 3D structure and the presence of high structural flexibility in these viral IDPs or proteins containing IDRs to allow their interaction with many binding partners at a time. Linking of functional domains and their promiscuity is achieved by interaction with partner IDRs, where the flexibility plays a major role in bringing two or more domains in proximity in order to perform a particular function. Flexible linker functions of IDRs in viral protein confer the advantage of escaping the recognition by the host immune system; the viral protein interacts with host protein in such a way that the recognition of viral epitope becomes difficult to be recognized by the components of the host immune system. A mutation rate that is typically high in the viral system could be tolerated by the presence of these flexible regions in viral proteins that forbid the structural constraints, hence

avoid the susceptibility to mutation. The expected explanation behind all these incidents points toward the involvement of IDPs. The first and pivotal observation of an abundance of intrinsic disorder in the replicative complex of paramyxoviruses had been confirmed.[157–159] Availability and use of bioinformatics tools in the last decades and their continuous growth and the development of sensitive biophysical experimental techniques lead to the identification of an abundance of IDPs in Viruses.[31,160–163]

4. Origin of viruses and their exclusive properties

Among all replicating organisms, the highest number is demonstrated by the viruses, which, therefore, are considered as the most abundant biological entities on the Earth,[164] For instance, if we compare the count of cells of all living creature present on the earth to number of the viral particles, it will be less than at least an order of magnitude.[165,166] The number of viruses can be estimated by counting the number of virus-like particles in the environment. For example, 1 mL of natural water contains as many as 2.5×10^8 viral particles.[167] Viruses are parasitic in nature, and in high abundance could be found in infected cells of Bacteria, Archaea, and Eukarya or even in other viruses.[164,166,168] The discovery of a small icosahedral virophage named Sputnik established the concept of infection of the virus by another virus[169] Sputnik virus infects the *Acanthamoeba polyphaga* Mimivirus (APMV) that in turn infects amoeba. APMV is a member of the Megaviridae family.[170–172] Infection of APMV by Sputnik virus is damaging and produces many deleterious effects in APMV, e.g., the assembly of capsid becomes abnormal and abortive viral forms appear.[169] This breach in the normal morphogenesis of APMV is explained on the basis of cytoplasm-independent replication center of APMV, where final morphogenesis normally takes place. However infection with Sputnik and multiplication of this virus at this center hinder its normal function.[173]

From the structural perspective, viruses demonstrate very simple structural organization. However, they display various shapes and strictly do not possess a unique common morphology. The genome of all viruses either made up of double or single-stranded DNA or RNA. It is encapsulated within a protective protein coat known as the capsid. An additional lipid envelope contains a number of membrane proteins found in Enveloped viruses. The position of the envelope is above the matrix protein, which is an additional proteinaceous coat. Some complex viruses in addition to the non-structural proteins contain numerous accessory and regulatory

proteins all that help in the assembly of the viral capsid. Viruses in reference to the structure of their genome, mechanism of replication, and transcription display a wide array of diversity. The viral genome could be of single or double-stranded DNA or single or double-stranded RNA and transcribed via a negative sense, positive sense, or ambisense mechanism. The diversity of the viruses either in genomic structure or mechanism of function leads to their classification in seven major classes.[174] Following this classification, all DNA based viruses kept in class I, II, and VII that contain dsDNA viruses, ssDNA viruses and dsDNA viruses that replicate via an intermediate single-stranded RNA (ssRNA) respectively. The remaining four classes, that is III, IV, V and VI, contain various RNA viruses, such as double-stranded RNA viruses (dsRNA), ssRNA viruses of positive (+) sense, ssRNA virus of negative (−) sense, and ssRNA virus of positive (+) sense that replicate via DNA intermediate, respectively. The certain features of viruses that typically oppose them to the living organisms are the absence of cell-like defined structure and inability of maintaining homeostasis and reproduce outside of the cellular environment due to the absence of their own metabolism and essential dependence on the host cell to make new products. The other features, such as the presence of a genome, replication ability and self-assembling creation of their own copies, and continuous evolution by natural selection make viruses similar to other living organisms.[175]

The presence of unusual properties makes it difficult to agree on the common view on the viruses. It is difficult to elaborate on whether viruses are some organisms at the edge of life, different and special with respect to other living cellular organisms, or nonliving organic structures that have a self-driven property to interact with living organisms.[176] The recent discovery of the presence of the metabolic protein-encoding genes in giant viruses challenged the previous view of the lack of these genes in viruses.[177] Certain bacterial species, such as Mycoplasma, Rickettsia, and Chlamydia are obligate intracellular parasites exactly as viruses. All this approves the reconsideration of criteria describing the living organisms.

There is an incomplete understanding of virus origin, three chief hypotheses have put forth to explain the understanding of their beginning.[178] The first hypothesis is the coevolution theory, according to which viruses and cells appeared simultaneously in the early history of the Earth. Since their emergence, viruses have a dependency on cellular life. The second hypothesis is known as the cellular origin hypothesis or the vagrancy. According to this hypothesis it is assumed that the evolution of viruses occurs from the DNA or RNA pieces that escaped from the genes of the larger

organisms. Examples of potential candidates for this escaped genetic material are (1) physically separated chromosomal DNA that is naked and can replicate independently called plasmid, (2) DNA pieces that have the ability to move from one place to other within the gene and replicate, termed transposon. Last, the third hypothesis of virus origin is a regressive or degeneracy hypothesis that proposes the origin of viruses take place from a parasitic cell that sheds all genes that were not required for the support of parasitism. The root of viral origin also traced from the nucleoprotein world that transiently existed during the transition of the RNA world to the modern DNA-RNA-Protein world according to different hypotheses. The appearance of RNA viruses took place either due to reduction or escape from the RNA containing primitive cells. These RNA viruses are also considered as the evolutionary starting point for some of the DNA viruses.[178] The origin of viruses considered to be in the early phase of the evolution of life,[179] when the living cells first evolved. Since then the existence of viruses has been proposed. This could be a reason why viruses have the ability to affect the cells from all three kingdoms of life that are Eukarya, Archaea, and Bacteria. The primitive viruses and their quick evolution propose a possible explanation for the lack of homology among the major viral proteins and proteins of cellular organisms.[178]

5. Classification of viral protein and their functions

Viruses contribute to the evolution of life through their ability to promote horizontal gene transfer and discovering DNA and its mechanism of replication among different life forms. The amalgamation of foreign genes often from unrelated organisms and modification in replication machinery leads to continuous evolution and genetic diversity.[180] The contribution of virally originated DNA fragments in the genetic material of humans is between 3% and 8%. Origin of few DNA replicating proteins through viral sources and their successive transfer in the cellular organisms advocate the key role of viruses in the formation of DNA and subsequent development of replication mechanism. These viral-mediated developmental processes were essential for the evolution of the eukaryotic nucleus and potentially the development of three domains of life.[178] A new classification for the life forms present on the Earth has been proposed. According to this classification all ribosome encoding organisms that include Archaea, Bacteria and Eukaryotes are kept in one class and all viruses are included in separate class of capsid encoding organisms that dependent on ribosome-encoding host

for completion of their life cycle and contain nucleic acids and proteins and also possess the ability of self-assembly into nucleocapsids.[181]

5.1 Viral structural proteins

The viral capsid is the protective coat surrounding the viral genome. Protein monomeric subunits termed as capsomers or protomers combine together to build the shell structure of the capsid. A tight association of RNA or DNA based genome to capsid protein results in the formation of a nucleoprotein complex. Nucleoprotein complex of viruses has the capability to interact with both nucleic acids and proteins, thereby possessing multifunctionality. Capsid structure is determined by the arrangement of capsomers. On this basis capsid could be of helical, icosahedral, or complex in shape. Highly ordered helical structure is a property of the capsids of helical, rod-shaped and filamentous viruses that are generally formed around a central axis with a single type of capsomer packaging. The genetic material of viruses made of RNA or DNA occupies the central cavity of the capsid, the positive charge of capsid protein and negative charge of viral genome maintain an electrostatic interaction between them. There is a variation in the size of helical viruses, which could be very long and flexible or very short and rigid. Capsid length of the helical viruses is defined by their genome size, whereas their diameter is defined by the size and arrangement of capsomers. Well-known illustrative examples among filamentous viruses are *Sulfolobus islandicus* filamentous virus (SIFV), Tobacco mosaic virus (TMV), Acidianus filamentous virus 1 (AFV1), and bacteriophage fd. In icosahedral viruses, the capsids are either icosahedral or nearly spherical with icosahedral symmetry. Although the number of capsomers required in the formation of such an icosahedral structure theoretically is calculated to be 60, in reality, in the majority of icosahedral viruses it is above the 60.[148]

Viral capsids are often made up of more than one capsid protein. For instance, capsid of Human papillomavirus (HPV) is made of major (L1) and minor (L2) capsid proteins. In the case of icosahedral viruses, the capsid is made up of more than 60 identical subunits. To develop the icosahedral shape, the same protein in different sites shows different symmetries. This intriguing puzzle has been the topic of long-lasting debates on how the identical subunits with identical unique 3D structures fit into different symmetries in different environments.[182]

Few viruses have complex capsid structures that are neither completely helical nor icosahedral and contain some extra structures, such as protein tails

or complex outer walls. The example of one of the best-studied complex viruses is T4 bacteriophage. The characteristic feature of this virus is an icosahedral head on the top of the helical tail. A structure of a hexagonal base plate with extended and protruding proteinaceous fibers occurs at the end of its tail. T4 virus attains the ability to bind host bacterium and successfully transfer its genome into it due to this tail structure that acts as a molecular syringe.[183]

Lipid membrane of viral capsid is acquired from the host by certain viruses. The membrane-coated capsid of these viruses is known as the viral envelope that might also contain the viral glycoprotein, for example, gp160 in Human Immunodeficiency Virus (HIV) that contains transmembrane subunit gp41 and structural subunit gp120, proton-selective ion channel and M2 protein of influenza virus and Hemagglutinin (HA) and neuraminidase in other enveloped viruses. The functional role of these surface-incorporated viral glycoproteins is rather diverse. Few among these glycoprotein that protrudes from the lipid bilayer of the virus, for example, neuraminidase (NA), HA and gp120, play a number of important roles in early-stage viral infection typically associated with attachment and penetration of the viruses into the host cells.[184]

As stated earlier, viral glycoprotein performs diverse functions related to the life cycle of enveloped viruses. For instance, the M2 proton channel of influenza A virus has a crucial role in the early and late replication cycle of influenza. The exposure of viral content to host cytoplasm requires hydrogen ion to lower the pH. At lower pH, M1 dissociate from the ribonucleoprotein and initiate viral uncoating. The supply of hydrogen ion into the viral particle from endosomes is mediated through the integral homotetrameric membrane protein (M2 proton channel), situated in the viral envelope. This ion channel is proton-selective and is gated by low pH conditions.[185] In enveloped viruses, the viral envelope is attached to their core via matrix proteins. Matrix protein plays their role once the virus enters into the host cell. In addition to expelling the genetic material from the viral core, matrix proteins have various regulatory roles via interacting with host components. For instance, influenza virus matrix protein M1 controls inhibition of viral transcription, its ribonucleoprotein export from nucleus and budding.[186,187]

5.2 Viral non-structural proteins

Non-structural proteins do not form the capsid structure. Instead, they participate in viral multiplication and have multiple regulatory functions.

Below are some illustrative examples of the non-structural proteins of a few viruses and their involvement in crucial viral functions. HPV open reading frames (ORFs) are classified in early (E) and late (L) types on the basis of location within the viral genome. HPV early ORFs code for non-structural proteins. Both E1 and E2 proteins participate in viral replication and regulation of transcription at an early stage. E1 binds to the origin of replication and unveil helicase and ATPase activity,[188,189] while E2 facilitates E1 binding to the origin of replication by forming the complex with it.[189–191] E2 also plays a role of a transcription factor by regulating (both positively and negatively) early gene expression by attaching to the specific recognition sites within the upstream regulatory region (URR).[192,193] A differentiation-dependent productive phase of the viral life cycle is promoted by the highly expressed protein E4 that is involved in a number of important functions.[194–196] In vitro studies found that E5 has weak transforming capabilities.[197,198] It disrupts the MHC class II maturation[199] and is involved in the HPV late functions.[200,201] E6 and E7 proteins are primarily involved in the progression of HPV-mediated malignant cells that ultimately cause invasive carcinoma. Their role in high-risk HPVs is to act as partial oncoproteins at least by targeting the cell cycle regulator/tumor suppressor p53 and Rb. Another example that demonstrates the diversity of functional roles attributed to the non-structural proteins is given by Hepatitis C Virus (HCV), where the interaction of non-structural protein with the hVAP-33 (VAMP-Associated Protein A), which is a human cellular vesicle membrane transport protein, lipid raft membranes, and with each other leads to the formation of the HCV RNA replication complex also called HCV replicon.[202]

In the diversity of their functional roles, immunomodulation is also demonstrated by non-structural proteins. The non-structural protein NS1 of West Nile virus (WNV) has displayed is presence in the immunomodulation, as concluded by experimental finding that both cell-surface associated, as well as soluble NS1 was able to bind and recruit the complement regulatory protein factor H. Due to this activity, there is a decrease in the complement activation that minimizes the targeting of WNV by immune system via decrease in the infected cells complement recognition.[203] The immune modulation role is also exhibited by rinderpest virus non-structural C protein, but via a different mechanism. In rinderpest virus action of type 1 and type 2, interferons, which are responsible for the induction of innate immune response, are specifically blocked by non-structural C protein.[204] It has been determined that many non-structural V proteins of paramyxovirus have shown their roles in countering the response of

antivirals.[205] At last, gene transactivation may require viral non-structural proteins. For instance, the autonomous parvovirus minute virus of mice (MVM) non-structural protein NS1 is required the activation of p39 promoter that controls the transcription of a gene that encodes capsid protein. Gene that code for NS-1 also codes NS-2 due to overlapping transcription unit in MVM virus. This gene is transcribed by a P04 promoter.[206]

5.3 Viral accessory and regulatory proteins

Many of the crucial functions of viruses are performed by various accessory and regulatory proteins through their involvement in an indirect functional role that ranges from transcription rate regulation of viral gene encoding structural proteins to modification of host cell functions. For instance, the replication of HIV-1 is actively controlled by the production of several accessory (Nef, Vpu, Vif, and Vpr) and two regulatory proteins (Rev and Tat). These regulatory and accessory proteins control the various aspects of the viral life cycle, in addition to regulating the host cell functions, such as gene regulation and apoptosis.[207] A number of accessory proteins are, in fact, responsible for in vivo infection. For instance, Vif protein overcomes the host defense mechanism, while Nef increases the viral pathogenesis by targeting the bystander cells.[207]

6. Role of bioinformatics in divulging the dark proteome of viruses

Viral proteins contain many unusual features that are lacking in the cellular proteins of other organisms.[160,179] The presence of a specific feature in a viral proteome helps them to adopt to a hostile environment quickly, while providing means for controlling the cellular machinery easily.[208] The absence of corresponding features in the proteome of other organisms might reflect the ancient origin of viruses and their genome from the cellular lineage that is extinct now.[209] In addition to demonstrating various peculiar features, as enlisted in,[160] viral proteome contains frequent short disordered regions that generally lack the hydrophobic residues and lysine, while containing the polar residues and residues that are not involved in the regular secondary structure formation.[148,160] The polar residues are required for the specific recognition and stabilizing the interaction with partner molecule through hydrogen bonding in a bound state and maintain randomness in an isolated state.[210] The loosely packed and disorder-enriched viral proteome resists the negative effect of mutations that is a quite common event in

viruses.[148] In order to evaluate the correlation of structure, function, and extent of disorder in the proteome of viruses, Pfam database analysis was carried out.[162] The disordered regions of viruses are mainly attributed to the protein-protein interaction, recognition, signal transduction, and regulation.[148] Viruses hijack the host cell machinery and use it for their specific functions on the basis of their ability to mimic the host protein short linear motif (SLIMs).[152] SLIMs are embedded in disordered regions and play a great number of diverse roles, such as directing proteins to the correct subcellular localization, targeting host proteins for proteasomal degradation, cell signaling, deregulating cell cycle checkpoints, and altering transcription of host proteins.[211] Based on the requirements, the proportion of SLIMs varies; hence the number of disordered regions could vary from one viral family to other. Recent studies also determined that there is no specific correlation in the genome size and disordered content in viruses.[212]

Bioinformatics plays an important role in divulging the Intrinsic disorderness of small biological machinery owning the replication ability in the host, and establishing the structural, functional and regulation networking as discussed and referenced in the aforementioned paragraph.

7. Prevalence of IDPs in viruses in context to three distinct domains of life

Many different studies and evaluations of IDPs fraction in evolutionarily distant species were conducted in the last decade.[24–29,213] Based on the major outcomes, in general, it was concluded that in comparison to prokaryotic proteomes, proteomes of Eukaryotic species have a higher portion of IDPs and IDRs. The basis for the justification of these observations was the repertoire of the specific function of IDPs/IDRs which are mainly involved in the events of recognition, regulation, and signaling. The regulatory network of eukaryotic organisms, especially those who are multicellular, is explicitly depend on the ability of IDPs/IDRs to perform multiple vital functions.[2,38,39] Although, as much as the functional basis considered as an important component that acts as a driving force for evolutionary changes, the change of proteome by itself cannot be ignored. The assumption to establish the relationship between morphological complexity and proteome size of the organism is alluring. Although this trend is valid in the case of establishing the difference between eukaryotes and prokaryotes, but cannot be implemented among species of eukaryotes, where the wide variations in nuclear genome size have been reported and termed as the C-value paradox. C-value, which is simply described as the amount of haploid DNA present in the cells of an organism, was described as a significant

quantity that could be used to estimate and look into the nature of the gene.[214–216] In comparison to the human genome. The genome size of a plant *Paris japonica* is nearly 50 times greater; genome sizes of some unicellular Protista are much larger than the human genome. For instance, *Polychaos dubium* genome is 210 times of human genome and is the largest known genome.[217] Cells of some salamanders contain 40 times more DNA than cells of humans.[218] The mystery of complexity of the relation of eukaryotic genome size and gene number is solved with the discovery of non-coding DNA revealing that the most of the DNA of eukaryotes is non-coding in nature hence cannot be incorporated in genes. This discovery also proposed that the description of organisms should not be solely based on a total number of protein-encoding genes, but the number of encoded proteins should be taken into account. However, the recent finding evidenced the poor correlation between the complexity of a given organism and its proteome size for instance number of proteins in the whole proteome of Nematode, *Caenorhabditis elegans* is ~20,000[219] and is similar to the number of proteins encoded by the human genome. A study focused on the analysis of predicted intrinsic disorder in the proteome of 3484 organisms including viruses conducted in 2012 revealed the number of significant details of the proteomes of various organisms.[31] Table 1 lists the details of the prevalence of intrinsic Disorder in protein contents of different viruses deposited into the DisProt database.[220]

7.1 Continuous spectrum of the proteome size space

Analysis of IDPs of the 3484 proteomes of different species resulted in the observation of the continuous spectrum of the proteome size space among the proteomes of eukaryotes, bacteria, archaea, and viruses, as wonderfully depicted in Fig. 1A.[31] Eukaryotes demonstrate wide-scale variations in the size of their proteome that form proteins whose number ranges from 4000 for unicellular species to ~20,000 for multicellular species. Bacterial proteomes have a number of proteins in the range of 500–8000, with only a small portion of bacterial species having proteome size less than 1500 proteins. The archaeal proteomes are condensed to the much narrow range of 1500–3000 proteins. Proteomes of viruses are very compact, being limited to less than 1000 proteins. Log-based plot analysis (Fig. 1B of [31]) determines that the only one polyprotein is possessed by the greater than 200 viruses and the number of viruses whose genome encode proteins between 15 and 30 is limited in comparison to the viruses with other sizes of the proteome.[31] So far, nine large mimiviruses are known, each containing more than 500 proteins. The size of the proteome of these mimiviruses is so large, that

Table 1 Details the DisProt ID, Uniprot ID, protein name, source organism and identified disordered content.

No.	DisProt ID	UniProt accession	Protein name	Organism	Disorder content (%)
1	DP00003	P03265	DNA-binding protein	Human adenovirus C serotype 5	9.83
2	DP00005	P03045	Antitermination protein N	Escherichia phage lambda	100.00
3	DP00024	P03129	Protein E7	Human papillomavirus type 16	100.00
4	DP00034	P03661	Attachment protein G3P	Enterobacteria phage fd	5.66
5	DP00048	P03406	Protein Nef	Human immunodeficiency virus type 1 group M subtype B (isolate BRU/LAI)	59.22
6	DP00064	P03607	Capsid protein	Southern cowpea mosaic virus	22.94
7	DP00066	P27285	Structural polyprotein	Sindbis virus subtype Ockelbo (strain Edsbyn 82-5)	9.08
8	DP00087	P68336	Tegument protein VP16	Human herpesvirus 2 (strain HG52)	27.14
9	DP00101	P12493	Gag polyprotein	Human immunodeficiency virus type 1 group M subtype B (isolate NY5)	5.60
10	DP00133	P03422	Phosphoprotein	Measles virus (strain Edmonston)	62.72
11	DP00148	P03347	Gag polyprotein	Human immunodeficiency virus type 1 group M subtype B (isolate BH10)	10.74
12	DP00160	P04851	Nucleoprotein	Measles virus (strain Edmonston)	23.71
13	DP00182	P03087	Major capsid protein VP1	Simian virus 40	14.09

14	DP00189	P04324	Protein Nef	Human immunodeficiency virus type 1 group M subtype B (isolate PCV12)	21.84
15	DP00284	P16009	Baseplate central spike complex protein gp5	Enterobacteria phage T4	17.74
16	DP00288	Q06253	Antitoxin phd	Escherichia phage P1	100.00
17	DP00410	P12497	Gag-Pol polyprotein	Human immunodeficiency virus type 1 group M subtype B (isolate NY5)	6.41
18	DP00419	P03176	Thymidine kinase	Human herpesvirus 1 (strain 17)	15.69
19	DP00424	P04325	Protein Rev	Human immunodeficiency virus type 1 group M subtype B (isolate PCV12)	62.07
20	DP00447	P12579	Phosphoprotein	Human respiratory syncytial virus A (strain Long)	100.00
21	DP00566	P13102	Hemagglutinin	Influenza A virus (strain A/Whale/Maine/328/1984 H13N2)	6.18
22	DP00573	P03305	Genome polyprotein	Foot-and-mouth disease virus (isolate Bovine/Germany/O1Kaufbeuren/1966 serotype O)	1.67
23	DP00583	P16006	Deoxycytidylate deaminase	Enterobacteria phage T4	10.88
24	DP00588	P27958	Genome polyprotein	Hepatitis C virus genotype 1a (isolate H)	2.72
25	DP00615	Q9WMX2	Genome polyprotein	Hepatitis C virus genotype 1b (isolate Con1)	3.39
26	DP00627	Q05323	Hexameric zinc-finger protein VP30	Zaire ebolavirus (strain Mayinga-76)	4.86

Continued

Table 1 Details the DisProt ID, Uniprot ID, protein name, source organism and identified disordered content.—cont'd

No.	DisProt ID	UniProt accession	Protein name	Organism	Disorder content (%)
27	DP00629	Q07097	Nucleoprotein	Sendai virus (strain Fushimi)	23.66
28	DP00640	Q89933	Nucleoprotein	Measles virus (strain Edmonston B)	24.00
29	DP00673	P06935	Genome polyprotein	West Nile virus	3.06
30	DP00674	Q69422	Genome polyprotein	Hepatitis GB virus B	5.38
31	DP00675	P19711	Genome polyprotein	Bovine viral diarrhea virus (isolate NADL)	2.56
32	DP00685	Q98157	Viral macrophage inflammatory protein 2	Human herpesvirus 8 type P (isolate GK18)	25.53
33	DP00686	Q9IH62	Glycoprotein G	Nipah virus	3.65
34	DP00697	Q9IK92	Nucleoprotein	Nipah virus	25.00
35	DP00698	O89339	Nucleoprotein	Hendra virus (isolate Horse/Autralia/Hendra/1994)	25.00
36	DP00699	Q9IK91	Phosphoprotein	Nipah virus	57.26
37	DP00700	O55778	Phosphoprotein	Hendra virus (isolate Horse/Autralia/Hendra/1994)	57.14
38	DP00726	Q5UPJ7	Tyrosine—tRNA ligase	*Acanthamoeba polyphaga* mimivirus	6.07
39	DP00741	P03040	Regulatory protein cro	Escherichia phage lambda	16.67
40	DP00750	Q38151	39 protein	Bacillus phage SPP1	46.83

41	DP00764	O89467	Protein Tat	Equine infectious anemia virus	69.23
42	DP00808	P24937	Pre-protein VI	Human adenovirus C serotype 5	57.60
43	DP00820	O73557	RING finger protein Z	Lassa virus (strain Mouse/Sierra Leone/Josiah/1976)	58.59
44	DP00842	P12506	Protein Tat	Human immunodeficiency virus type 1 group M subtype D (isolate Z2/CDC-Z34)	100.00
45	DP00847	P20220	Protein F-112	Sulfolobus spindle-shape virus 1	34.82
46	DP00849	Q9Q8E9	M156R	Myxoma virus (strain Lausanne)	43.14
47	DP00850	P36932	Integrase	Escherichia phage P2	34.42
48	DP00871	A4ZNR2	Nuclear export protein	Influenza A virus	100.00
49	DP00875	P69723	Virion infectivity factor	Human immunodeficiency virus type 1 group M subtype B (isolate HXB2)	27.08
50	DP00876	P14340	Genome polyprotein	Dengue virus type 2 (strain Thailand/NGS-C/1944)	2.95
51	DP00895	P03421	Phosphoprotein	Human respiratory syncytial virus A (strain A2)	42.32
52	DP00898	P13338	RNA polymerase-associated protein Gp33	Enterobacteria phage T4	36.61
53	DP00919	C6KEI3	Protein Nef	Human immunodeficiency virus 1	50.00

Continued

Table 1 Details the DisProt ID, Uniprot ID, protein name, source organism and identified disordered content.—cont'd

No.	DisProt ID	UniProt accession	Protein name	Organism	Disorder content (%)
54	DP00929	P04608	Protein Tat	Human immunodeficiency virus type 1 group M subtype B (isolate HXB2)	100.00
55	DP00932	P35926	Recombination enhancement function protein	Escherichia phage P1	40.86
56	DP00939	P04859	Phosphoprotein	Sendai virus (strain Harris)	7.57
57	DP00947	O10609	Protein E7	Human papillomavirus type 45	71.70
58	DP00948	P59595	Nucleoprotein	Human SARS coronavirus	42.42
59	DP00965	P0C6L3	Small delta antigen	Hepatitis delta virus genotype I (isolate D380)	75.90
60	DP00976	P04578	Envelope glycoprotein gp160	Human immunodeficiency virus type 1 group M subtype B (isolate HXB2)	12.27
61	DP00978	P35961	Envelope glycoprotein gp160	Human immunodeficiency virus type 1 group M subtype B (isolate YU-2)	55.28
62	DP00986	Q8QWD4	VP4	Enterovirus D68	40.58
63	DP00998	Q05127	Polymerase cofactor VP35	Zaire ebolavirus (strain Mayinga-76)	8.53
64	DP00999	P03315	Structural polyprotein	Semliki forest virus	11.33
65	DP01012	P27392	Protein P16	Enterobacteria phage PRD1	31.62
66	DP01013	P68927	Excisionase	Escherichia phage HK022	30.56

67	DP01016	Q20MD5	Matrix protein 2	Influenza A virus (strain A/Udorn/1972 H3N2)	21.88
68	DP01031	Q99IB8	Genome polyprotein	Hepatitis C virus genotype 2a (isolate JFH-1)	3.10
69	DP01039	Q85258	Polyprotein	Potato virus Y	31.65
70	DP01043	Q80FJ1	Membrane fusion protein p14	Reptilian orthoreovirus	24.00
71	DP01059	Q71FK2	Coat protein	Pepino mosaic virus	8.44
72	DP01060	A8CDV5	Latent membrane protein 2A	Epstein-Barr virus (strain GD1)	100.00
73	DP01087	Q1PAB4	Protein Tat	Human immunodeficiency virus 1	100.00
74	DP01129	P12296	Genome polyprotein	Mengo encephalomyocarditis virus	0.61
75	DP01142	O92972	Genome polyprotein	Hepatitis C virus genotype 1b (strain HC-J4)	8.54
76	DP01150	P03255	Early E1A protein	Human adenovirus C serotype 5	48.10
77	DP01151	P03259	Early E1A protein	Human adenovirus A serotype 12	100.00
78	DP01186	P03086	Agnoprotein	JC polyomavirus	43.66
79	DP01188	Q5XXP4	Polyprotein P1234	Chikungunya virus (strain 37997)	0.69
80	DP01245	P12823	Genome polyprotein	Dengue virus type 2 (strain Puerto Rico/ PR159-S1/1969)	0.59
81	DP01256	Q32ZE1	Genome polyprotein	Zika virus	2.22
82	DP01295	Q98XH7	Protein Tat	Human immunodeficiency virus 1	100.00
83	DP01305	P08392	Major viral transcription factor ICP4	Human herpesvirus 1 (strain 17)	2.23

Continued

Table 1 Details the DisProt ID, Uniprot ID, protein name, source organism and identified disordered content.—cont'd

No.	DisProt ID	UniProt accession	Protein name	Organism	Disorder content (%)
84	DP01336	P03709	DNA-packaging protein FI	Escherichia phage lambda	29.55
85	DP01391	P03520	Phosphoprotein	Vesicular stomatitis Indiana virus (strain San Juan)	22.64
86	DP01393	P04880	Phosphoprotein	Vesicular stomatitis Indiana virus (strain Mudd-Summers)	22.64
87	DP01394	P04879	Phosphoprotein	Vesicular stomatitis Indiana virus (strain Glasgow)	22.64
88	DP01395	Q8B0H3	Phosphoprotein	Vesicular stomatitis Indiana virus (strain94GUB Central America)	22.64
89	DP01405	Q5V913	Nucleoprotein	Influenza B virus	12.50
90	DP01428	P03120	Regulatory protein E2	Human papillomavirus type 16	21.92
91	DP01466	A4L7I2	Non-structural polyprotein	Chikungunya virus	8.04
92	DP01468	A3RMR8	Non-structural polyprotein	Chikungunya virus	8.04
93	DP01469	A4L7I4	Non-structural polyprotein	Chikungunya virus	8.04
94	DP01481	Q5UPT2	Probable uracil-DNA glycosylase	Acanthamoeba polyphaga mimivirus	25.41
95	DP01512	P03050	Transcriptional repressor arc	Salmonella phage P22	100.00
96	DP01539	O57173	Protein F1	Vaccinia virus (strain Ankara)	22.52

97	DP01615	P03126	Protein E6	Human papillomavirus type 16	13.29
98	DP01616	P10104	Fibritin	Enterobacteria phage T4	23.20
99	DP01618	P03070	Large T antigen	Simian virus 40	11.72
100	DP01621	E5LC01	LANA	Human herpesvirus 8	4.74
101	DP01625	P07567	Gag polyprotein	Mason-Pfizer monkey virus	3.04
102	DP01642	P06492	Tegument protein VP16	Human herpesvirus 1 (strain 17)	27.14
103	DP01759	Q0GBY3	Phosphoprotein	Rabies virus (strain China/MRV)	22.90
104	DP01762	P03714	Head-tail connector protein FII	Escherichia phage lambda	35.04
105	DP01780	P21736	Protein E7	Human papillomavirus type 45	50.94
106	DP01806	Q67953	Large envelope protein	Hepatitis B virus	24.27
107	DP01843	P03404	Protein Nef	Human immunodeficiency virus type 1 group M subtype B (isolate BH10)	12.14
108	DP01928	P03254	Early E1A protein	Human adenovirus C serotype 2	81.66
109	DP01929	P17763	Genome polyprotein	Dengue virus type 1 (strain Nauru/West Pac/ 1974)	0.29
110	DP01930	P29990	Genome polyprotein	Dengue virus type 2 (strain Thailand/ 16681/1984)	0.62
111	DP01931	Q2YHF0	Genome polyprotein	Dengue virus type 4 (strain Thailand/ 0348/1991)	0.86
112	DP01983	Q9Q8N4	Probable host range protein 2-3	Myxoma virus (strain Lausanne)	26.11

Continued

Table 1 Details the DisProt ID, Uniprot ID, protein name, source organism and identified disordered content.—cont'd

No.	DisProt ID	UniProt accession	Protein name	Organism	Disorder content (%)
113	DP01984	B4Y891	Capsid protein VP1	Adeno-associated virus	29.06
114	DP02042	Q98325	Viral CASP8 and FADD-like apoptosis regulator	Molluscum contagiosum virus subtype 1	22.41
115	DP02051	P14335	Genome polyprotein	Kunjin virus (strain MRM61C)	0.52
116	DP02071	P04383	Capsid protein	Carnation mottle virus	23.28
117	DP02128	P06437	Envelope glycoprotein B	Human herpesvirus 1 (strain KOS)	12.39
118	DP02194	P68466	Protein K7	Vaccinia virus (strain Western Reserve)	16.78
119	DP02203	Q9Q6P4	Genome polyprotein	West Nile virus (strain NY-99)	0.61
120	DP02204	Q5UB51	Genome polyprotein	Dengue virus type 3 (strain Singapore/8120/1995)	0.77
121	DP02208	A0A140GKJ0	TAP transporter inhibitor ICP47	Human herpesvirus 1	37.50
122	DP02212	P05769	Genome polyprotein	Murray valley encephalitis virus (strain MVE-1-51)	0.70
123	DP02256	P26554	Protein E6	Human papillomavirus type 51	7.28
124	DP02261	P13848	Capsid assembly scaffolding protein	Bacillus phage phi29	20.41
125	DP02291	P04486	Tegument protein VP16	Human herpesvirus 1 (strain F)	16.12
126	DP02334	Q98148	Kaposi's sarcoma-associated herpes-like virus ORF73 homolog	Human herpesvirus 8	4.56

The table details have been obtained from the DisProt database at https://www.disprot.org/browse.

we can say that it is of nearly equal size to the proteome size of some small bacteria. Therefore, the continuous spectrum of a size of proteome arrange in the order of viruses to archaea, to unicellular eukaryotes and lastly to multicellular eukaryotes. The proteome of bacterial species overlapped with the proteome of viruses, archaea, and unicellular eukaryotes.[31]

7.2 Disordered residues fraction in various proteomes

7.2.1 Bacteria

Disorder protein content in the majority of bacterial species is estimated to be between 18% to 28%, which is quite low. Although the small number of bacteria shows disorder content as high as 35%, this value represents the lower boundary of the fraction of disordered residues predicted for both unicellular and multicellular eukaryotic organisms (Figs. 1 and 2 of [31]).

7.2.2 Archaea

Based on the estimated disordered content in Archaea, this kingdom can be split into three classes. Class one consists of the organism whose proteomic disordered content range from 12% to 21%, and 61 organisms such have been analyzed. Class two consists of 4 organisms whose disordered content varies from 21% to 32%. The last class has the 8 organisms with the estimated variation in their disordered content being reported to range from 32% to 38%. The comparatively higher percentage of disorder in the class three species is attributed to the peculiarities of their habitats. As confirmed by the studies, the high disordered bearing archaeal species are halophiles and methanophiles.[29] Generally, the global disorder predictors are developed on the basis of the training set of non-halophilic proteins under the normal physiological conditions of 100–150 mM NaCl. The accuracy of determined IDRs for the proteins of the extremophilic microorganisms surviving under the hypersaline conditions with the help of such predictors might vary. Actually, since halophilic microorganisms are the salt-loving extremophilic organism, their optimum growth occurs in the salt-rich environment. A strategy used by these microorganisms to maintain an appropriate osmotic environment in their cytoplasm is "salting- in". Through this, they accumulate molar concentration of chloride and potassium.[221] Extensive adaptation in the intracellular proteins is required for this strategy to tackle the presence of excessive salt concentration, as at near saturating salt concentration they should maintain proper conformation and activity. The proteomes of these "salting-in" organisms are highly acidic in nature and corresponding proteins possess remarkable structural instability in low salt

conditions, while possessing soluble and active conformations in a hypersaline (Salt rich) condition that are usually detrimental to proteins of non-halophilic organisms. Furthermore, a salt-rich environment determines the structure to function capability. In similarity to their physiological environment, excessive salts and water bind to proteins of these organisms in solvent conditions that depend upon the acidic amino acid residues present on the protein surface.[222–230] Considering the aforementioned reasons, it could be suggested that prediction of high disorder in these organisms may simply represents prediction error.[31]

7.2.3 Eukaryotes

The analyzed disorder levels among non-viral proteomes revealed that unicellular and multicellular eukaryotes generally have the highest amount of IDPs/IDRs in their proteomes. Comparative fractional analysis of disorder for them range between 35% and 45%. However, a group of unicellular eukaryotes has levels of disordered residues in the range of 45–50%. The organisms included in this group are *Cryptococcus neoformans* (CRYNE, DISORDER%, 47.1), *Neurospora crassa* (NEUCR, DISORDER%, 48.2), *Plasmodium falciparum* (PLAF7, DISORDER%, 49.5), *Plasmodium yoelii* (PLAYO, 46.0%), and *Ustilago maydis* (USTMA, 49.9%). The observed high variability and high levels of predicted disorder are in line with the earlier study that revealed enrichment of predicted disorder in early-branching protein, while comparing it to typical eukaryotic proteins structure submitted in Swiss-Prot database and ordered proteins from PDB.[231] As much as twice the fraction of IDRs with $\geq$30 disordered residues is found in some protozoa, in comparison to Swiss-Prot database-based representative set of proteins. If it will be compared with similar regions from a PDB select 25 set of proteins, it would be sevenfold increase.[231] It is noteworthy that more disordered proteins were found in parasitic protozoa than in non-parasitic protists.[231] For instance, 35% proteins encoded by genes present on the chromosomes 2 and 3 of *P. falciparum* were predicted to contain long IDRs (i.e., longer than 40 residues).[24] Although more recent study revealed that the data on the amount of disorder in *P. falciparum* was underestimated, proposing that 52–67% proteins of this organism contain long disorder regions.[232] The latest study examines the prevalence of disorder in the proteome of many apicomplexan parasites, the obtained result demonstrated that the primate malaria parasite (*P. knowlesi*) and human malaria parasites (*P. falciparum and P. vivax*) contain more disordered regions in comparison to rodent malaria parasite.[25] Additionally, more disorder was reported in the

proteins expressed at a sporozite stage of *P. falciparum* in comparison to those expressed in the other stages of their life cycle.[25] It has been proposed that a high abundance of disorder in the proteome of this unicellular organism is related to its adaptation to changing environment during its whole life-cycle, as it is able to affect many different hosts.[231] In simple words, we may say that the abundance of intrinsic disorder in the apicomplexan parasite evolves as a way to adopt a parasitic life style.[231] Overall observance of various proteomes of different life forms and their disorder contents revealed that with the increase in the proteome size, the lower bound fractions of disordered content appear to increase continuously, whereas the upper bound fractions of disordered residues decrease in viruses and increase among the bacteria, archaea, and eukaryote. Therefore, the species whose proteome size falls between 1000 and 2000 proteins have the least variance of the fraction of disordered residues. Nevertheless, if the variance of a fraction of disordered residues is measured by different domains of life, the largest variance comes to 70% and it would be for viruses, whereas for multicellular eukaryotes variance comes to 12% which is smallest.[31]

7.2.4 Viruses

There is a variation in the fraction of disordered protein residues among viral proteomes as shown in Fig. 1 of reference .[31] For example, avian carcinoma virus proteome has the highest fraction of disordered residues (77.3%), while human coronavirus NL63 has very low fraction of disordered residues (7.3%). Few species of viruses are highly rich in disordered residues. There are 20 small viruses that encode $\leq$5 proteins in their proteomes and that have disorder content 50% or greater. In viruses, it appears that with increasing proteome size, the disorder content converges in the range of 20–40%. The prediction of the high content of intrinsically disordered residues in viruses found to be in great agreement with a study showing that many proteins of bacteriophage, viruses, bacteria, and archaea are significantly depleted in the hydrophobic residues and enriched in polar (hydrophilic) residues in their sequences.[210] A portion of IDRs in viruses is likely to evolve to support their ability to deal with their hostile habitat, in addition to be profoundly involved in functioning of their proteins. Still, other IDRs have evolved to deal with the alternative splicing, antisense transcription, and gene overlapping in a way that makes more efficient use of genetic material.[162] Polar residues have the ability of specific recognition and could establish a strong hydrogen bond with partner molecules contrary to the non-specific hydrophobic interactions. An increased amount of polar

residues in viral proteins could be linked to increasing demand for disorder in the unbound state and specific recognition and stabilization inbound states.[31,210]

8. Predicted IDPs pattern relation to viral transmission and host tropism

A model has been proposed to categorize the different coronaviruses on the basis of the distribution patterns of IDPs within their Nucleocapsid (N) and Membrane (M) proteins. This categorization allows the quick determination of transmission behaviors (Route, mode, and mechanisms) of various coronaviruses regardless of their genetic proximity. For instance, the shell rigidity has been reported in the viruses transmitted by the oral-fecal route because rigidity in shell protein protects the virions from damage, rigidity in shell protein is directly linked to intrinsic disorder of N and M protein.[233]

Envelope protein gp120 of HIV-1 contains both ordered and disordered regions. V3 loop represents a disordered region that is important for controlling the immune cell receptor chemokines co-receptor mediated entry. Chemokines co-receptors CCR5 (R5), CXCR4 (X4) or Both (R5X4) used by the viruses are known as R5, X4, and dual tropic respectively. HIV-1 variant, while infecting the host, uses the different chemokine receptors. Switch from R5 to X4 is related to disease progression and pathogenesis, however, the reason for switching is majorly unknown. Xiaowei Jiang et al. hypothesized that this change is associated with sequence variation and intrinsic disorder. Detailed analysis by the same group using the nonparametric statistical approach determined that there is an increased disordered propensity in the V3 domain, while switching from the dual/R5 tropic to the X4 tropic virus. This increased structural disorder of the V3 domain is associated with HIV-1 cell tropism.[234] The aforementioned study forms the basis for the identification of different hidden patterns with respect to IDPs and their association with viral distinguished characteristics.

9. Aggregation in viral protein and its relation to intrinsic disorderness

Host cellular machinery hijacking and modulation of regulation network/components often results in the formation of insoluble inclusions/aggregates that usually contains the viral structural components. These

viral-mediated aggregates utilize the viruses to build the large complex containing both viral and host protein assembly for promoting viral replication, transcription, and translation and Intra/Intercellular transport. The aggregated structure housing the viral-host assembled complex protects it from cellular degradation mainly. Although the complete role and mechanism of function of these aggregates with respect to specific viruses are not completely understood,[235] however, in most cases, the pattern of aggregates and their associated characteristics helps in unraveling the behavior, quantification, and identification of viruses.[236] However, deep understanding and establishing an association between aggregation behavior and intrinsic disorder might provide the surplus information pertaining to the viral infection. Fig. 4 demonstrates the analysis of intrinsic disorder predisposition and intrinsic propensity for aggregation (and intrinsic solubility) in Japanese encephalitis (JEV), Enterovirus-71 (EV-71) and ZIKV genome polyproteins.

10. Functional prominences of disordered viral proteins: Examples from bacteriophages, plant, and animal viruses

Viral proteins are atypical in nature due to their poor homology to the proteins of modern cells, which proposed viruses are very primitive.[179] While evading the defense mechanisms of the host, it is compulsory for the viruses that they are able to survive outside and inside the host and also be able to quickly adapt to fast-changing surroundings. In order to keep the pact of quick adaptation with the fast-changing environment, viruses undergo a very high mutation (for RNA viruses it is 10^{-5} to 10^{-3} nucleotide exchange per generation and for DNA viruses it is in the range of 10^{-8} to 10^{-5}).[147] This much higher rate of mutation in viruses is due to the lack of RNA repair mechanisms. On average, mutation rate in Bacteria and eukaryotes is 10^{-9} nucleotide exchange per generation, which is comparatively low.[147] The viral genome is quite compact, and there is an overlap of many reading frames, a single mutation might affect more than one viral protein.[240] During various stages of their life cycle, viral proteins usually interact with multiple components of the host cells, starting from the early entry to formation and exit of new infectious viral particles. In order to perform crucial functions associated with their life cycle events, viruses interact with host nucleic acid and proteins, even though the large gaps exist in between viral and host protein.[178,240] The aforementioned features incite curiosity to look

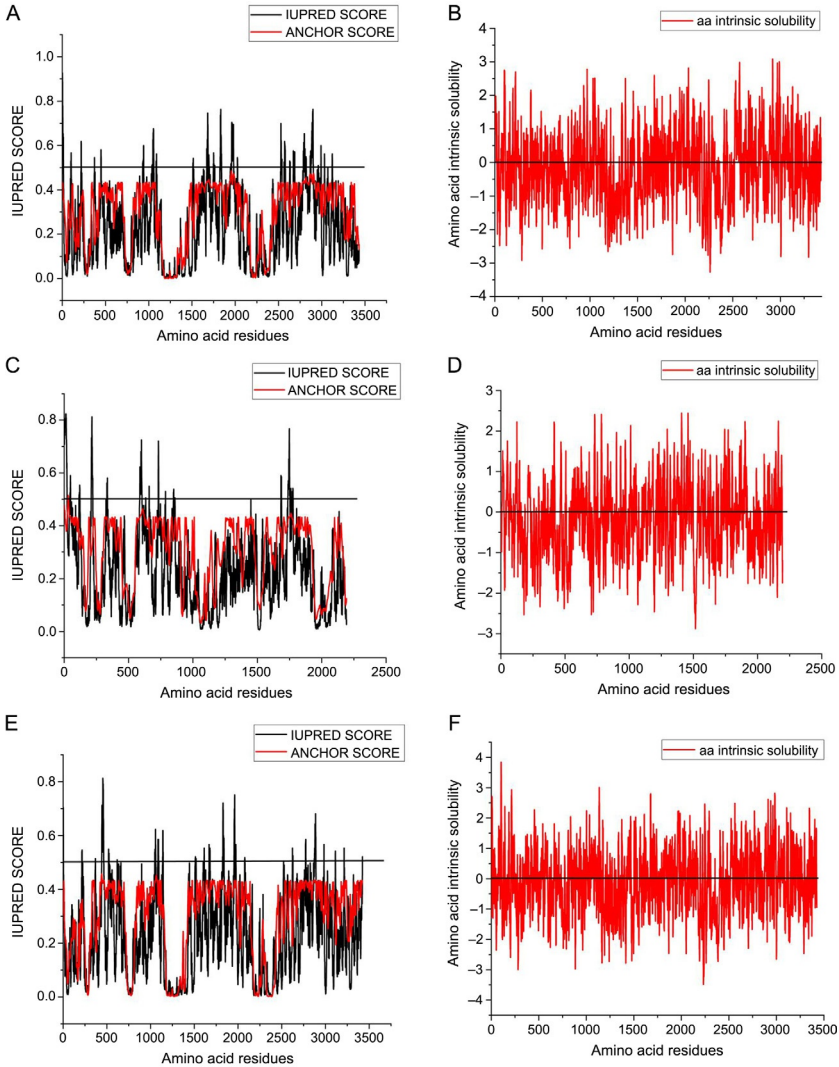

Fig. 4 Comparative analysis of IDPs contents to intrinsic aggregation (and Intrinsic solubility) propensity in three viruses genome polyprotein; JEV (UniProt id: P27395), EV-71 (UniProt id: Q66478) and ZIKV (UniProt id: A0A024B7W1) that impacted India). A, C, E represents IDPs content (and MoRF) propensity in JEV, EV-71, and ZIKV respectively determined by IUPred2A[237] while B, D, F show Aggregation propensity (and Intrinsic solubility) in JEV, EV-71 and ZIKV respectively determined by CamSol method Vendruscolo lab software.[238,239] A relation between IDPs content and aggregation propensity and viral infection pattern could be established.

into more details of their unique characteristics from the biophysical perspectives. The extent of the presence of intrinsic disorder in the viral proteome provides the corresponding plasticity that confers numerous functional advantages. The flexibility of an IDP/IDR and the lack of compact rigid structure enable it for multiple interactions. IDR binding promiscuity is facilitated by various mechanisms, with the operability of theses mechanisms depending upon its extent of the flexible linking property. This property of flexible linking provides an additional advantage to the viral proteins for eluding the host immune system and making it difficult for the host immune system to properly recognize the epitope. High disorder in viral proteome can be a way to deal with high frequency of mutations. Deleterious effects of mutation buffered by the high adaptability and low interaction between amino acids (flexibility) of IDPs. This is because the unstructured IDPs has less to lose when substitution takes place than a highly ordered structure that might have more impact on substitution. It is clearly evident that viral proteins can be benefited from flexibility garnered by disordered residues but not all the viral proteins have IDRs nor they are IDPs.

There is a relation between disorder content and location of a protein within the virion, and a comparative analysis of disorder predictors used in the analysis of viral proteins confirms it.[241] Such a study has begun with the construction of a database including viral proteins from HIV and Influenza-related viruses that followed by the protein sequence comparison, structure prediction, as well as function and location within the virion. The outcomes (particularly for influenza virus) demonstrated a correlation between the proximity to the RNA core of the virion and the levels of disorder in protein, where the closer protein is located to core the higher disordered percentage it would have. This finding of a relation between disorder and proximity to the core could be explained on the basis of more interactions with viral RNA. It has been found that nucleic acid-binding proteins are commonly disordered or at least have disordered regions at the site of nucleic acid binding.[242] In the case of the HIV, the correlation between proximity to the core and high disorder content has not to be observed possibly due to the presence of enzymes around the core region that are predominantly structured proteins.[243–245] The matrix protein of both HIV and influenza A viruses have rather different disorder contents. The HIV matrix protein is predicted to be highly disordered, while Influenza A virus protein is less disordered or somewhat ordered.[241] Concerning the Surface protein disorder, it was found that the surface protein gp120 of HIV has less disorder content across all analyzed strains, while

gp41 found to be highly disordered.[241] Surface proteins of Influenza A virus NA and HA are predicted to be mostly disordered.[241] However, the subsequent studies revealed that predicted disordered content vary among subtypes and suggested that this variability could have a link to the virulence level.[241,246]

10.1 Flexible promiscuity of viral proteins

IDPs can made interaction with several distinct partners due to their conformational flexibility and property of interaction adaptability. When a single IDR binds to many partners, then it converts themselves in many different structural forms.[110] IDPs demonstrate different interaction modes, either being able to form a very stable complex structure or transiting between the interacting partner as dynamic bound and unbound state acting as an on-off switch in signaling pathways.[11] Depending upon the surrounding environment, IDPs adopt different conformations and functions accordingly. Binding promiscuity is an important characteristic and required feature of the for viral proteomes, since despite encoding many proteins, viruses explicitly require host cell machinery to complete their life cycle, and in doing so, binding promiscuity is helping them to fulfill this role. The binding promiscuity and interaction types have well explained in the earlier paragraphs of this chapter. The compact genomes of viruses restrict them to encoding fewer proteins, but the presence of IDRs or global disorder allowed proteins to be involved in different tasks by interacting with various partners. With a few given examples, it would be easy to understand how the binding promiscuity of viruses is related to their intrinsic disorder. The replication of the RNA genome of hepatitis delta virus (HDV) requires the translation of a single basic protein known as the delta antigen (δAg). δAg is a small protein containing 195 amino acid residues and has no known enzymatic activity, although being essential for the replication of viral genome.[247] Experimental CD measurement and computational research via disordered protein meta-predictors have proven this protein to be an IDP.[248] Completion of the HDV replication cycle of the depends on this protein and various components of the host cell. Therefore, it is easy to understand the importance of binding promiscuity of δAg, that interact with multiple components in the host cell for various reasons and through a different approach, although the exact purpose of these interactions is still unclear and studied widely.[249,250] In an in vitro analysis, it was found that δAg binds to RNAs and even dsDNA in addition to binding to HDV RNAs that shows a lack of specificity in δAg protein.[248]

HCV NS5A protein that is involved in viral replication and viral particle assembly makes another example.[110,251] NS5A is a membrane-associated protein that has both disordered and ordered regions, an anchor attaches its N-terminal region to the membrane, but its cytoplasmic regions are mostly disordered and contain three domains. Among these three domains, domain I (D1) is highly conserved and has ordered sequence,[252] while domain II (D2) and III (D3) are highly disordered and less conserved.[253,254] Promiscuity of NS5A is well studied, and some of the interactions that involve its disordered domain have been identified.[255] D2-associated binding motifs that appear to affect the host regulation pathways, such as apoptosis and signaling demonstrate distinct interaction patterns described in detail.[256]

A third example of binding promiscuity was described for the Measles virus (MeV) Nucleoprotein (N) that forms the nucleocapsid of the virus. Intrinsically disordered regions are located at the C-terminal of N-protein,[159] that make interaction with phosphoprotein of the viral polymerase complex and perform functions required in replication and transcription.[257] Besides interacting with phosphoprotein for crucial processes, N-protein interacts with several host components, including cellular receptor and cellular cytoskeleton through its C-terminal tail.[258] Phosphoprotein of MeV is an important cofactor of polymerase complex and requires for recruitment of transcriptional machinery through its long disordered regions that it contained.[258] It has been observed that when IDRs of both phosphoprotein and N-protein binds, the major extent of flexibility disappeared, although some flexibility still presents that represent remaining disorder within the complex.[259] This finding in N-protein suggests that its IDRs act as a platform for the interaction with various protein partners for the completion of cellular processes.[259] The common feature of the structural disorder has successively shown in the nucleoprotein of Paramyxoviruses.[259] Disordered (Intrinsically unstructured) components were found together with structural components in proteins like nucleoprotein and phosphoprotein of Hendra and Nipah viruses.[259]

10.2 Intrinsic disorder in viral proteome regions affected by alternative splicing and overlapping reading frames

In due course of evolution to maximize the use of the limited genome in regulatory and structural protein, viruses adapted sophisticated genetic organization and mechanisms such as alternative splicing of polycistronic RNA which are necessary for the expression of the regulatory viral proteins in controlled manners. Viruses also evolve their genetic constitution, genomic

structure and mechanism of transcription and replication to efficiently use both positive and negative and even ambisense transcription. Among examples of such viruses are human T-cell lymphotropic virus type 1 (HTLV-1), a delta-retrovirus that causes HTLV-1-associated myelopathy, adult T-cell leukemia (ATL), and *Strongyloides stercoralis* hyperinfection. Economic usage of the genetic material of HTLV-1 is due to the wide accumulation of intrinsically disordered proteins in its proteome. This is paralleled to the occurrences of intrinsic disorder in HIV-1 protein, where intrinsic disorder was observed in post-translational cleavage sites leading to the production of Gag, Pro and Pol from Gag-pro and Gag-pro-pol grand polyproteins and cleavage sites of polyproteins that yield MA, NC, CA, RT, TM, IN, and SU proteins.[148]

10.3 Intrinsic disorders in viral genome-linked proteins

In few viruses, a protein named viral genome-linked protein (VPg) is bound to 5′ end of their RNA genome through a phosphodiester bond formed between the hydroxyl group of Thr/Ser/Tyr residues and 5′ phosphate group of RNA.[260–262] VPg's are highly diverse in terms of their size and sequence. For example, in Comoviridae and Picornaviridae members it is 2–4 kDa, Caliciviridae, Sobemoviruses, and Potyviridae members it is 10–26 kDa, while it is up to 90 kDa in Birnaviridae members.[263] VPg plays a key role in major steps of the viral life cycle, such as cell-cell movement, replication, and translation. Since VPg performs these crucial functions either in its mature or precursor form, VPg precursor processing represents one of the regulatory mechanisms of its multi-functionality.[262] The multitude of interactions with different viral and host proteins define VPg multifunctional role. The different interactions made by VPgs are: VPg to itself, cylindrical inclusion helicase, cylindrical inclusion protein, nuclear inclusion protein b, helper component protease, coat protein or eukaryotic translation initiation factors eIF4A, eIF4E, eIF3, and eIF4G, and the poly(A)-binding protein.[262,264–272] Poly-functionality and binding promiscuity of VPs' at least to some extent is due to its intrinsically disordered nature. Intrinsically disordered nature of VPg was reported for many viruses through their individual protein characterization. These viruses are: rice yellow mottle virus (RYMV), Sesbania mosaic virus (SeMV), potato virus Y (PVY), potato virus A (PVA), and lettuce mosaic virus (LMV).[262,273–276] The computational analysis showed that functionally important disordered VPg representative of viral diversity includes four members of the

Caliciviridae family, six potyviruses and six sobemoviruses.[276] The disordered VPg components associated with the regulation of enzymatic activity in different viruses[273,277] in addition to performing specific regulation and transportation of viral RNA from one cell to another.[278]

10.4 Intrinsic disorder in matrix proteins and nucleocapsid of HIV-related viruses

In order to determine the intrinsic disorder content in viral proteins, bioinformatic studies were carried on a few viruses matrix proteins.[241,279] This study revealed that matrix proteins p17 of SIVmac and HIV-I possess high disorder content, while low disorder was observed in the matrix protein of equine infectious anemia virus (EIAV).[279] Matrix protein p17 of HIV-I, also known as MA protein, is 132 amino acid long polypeptide that lines the inner surface of the virion membrane and holds the RNA containing viral core at its place. The N-terminal part of the p17 matrix protein is myristylated.[280,281] p17 associated with the inner leaflet of the viral membrane and form the protective shell and participate in virion assembly.[282] A targeting signal for the Gag polyprotein transport to plasma membrane is provided by co-translational myristylation of p17 N terminus.[280,281] A specific feature; i.e., the presence of a set of basic residues within the first 50 amino acid residues of p17, enable its involvement in membrane targeting.[283] In addition to performing the number of functions in the viral replication cycle, it could be involved in nuclear import possibly through its specific nuclear localization sequence.[284]

HIV-I nucleocapsid protein is 55 residues long protein that contains two zinc finger domains flanked by linker comprised of basic amino acids, which is required for nucleic acid interaction.[285,286] This nucleocapsid covers the genomic RNA inside the virion core. The important function of nucleocapsid is in viral genomic RNA assembly; it binds to the signal sequence of full-length RNAs and transports them into the assembling virion.[283] Within the virion, nucleocapsid binds to ssRNA non-specifically due to its highly charged basic regions and protects it from nuclease besides compacting it. Nucleocapsid also acts as a chaperone for viral RNA and facilitates the several steps of the viral life cycle associated with a nucleic acid, such as the melting of secondary structure within RNA, annealing of t-RNA primer, stimulating integration[287] and promoting the DNA exchange reactions during reverse transcription.[157–159] Computational prediction reveals that p7 is a highly disordered protein except for a few regions that are corresponding to the zinc finger domain and possess ordered structure identified as

α-MoRFs.[23] Flexible nature of p7 (NC) explains its multiple functional roles, such as participation in RNA chaperoning and viral replication.[288]

10.5 Replicative complex of Paramyxoviridae and Rhabdoviridae members: Intrinsic disorder and disorder-to-order transitions

Paramyxoviridae and Rhabdoviridae are the members of the mononegavirales order consisting of viruses with non-segmented ssRNA genome of negative polarity.[289] In mononegavirales, genome is tightly encapsidated by the nucleoprotein within a helical nucleocapsid. The viral nucleocapsid serves as a substrate for both replication and transcription. Both replication and transcription are performed by the viral RNA-dependent RNA polymerase (RDRP) that consists of complex formed between the viral large protein (L) and phosphoprotein (P). P protein acts as an essential polymerase cofactor and recruits the L-protein onto the nucleocapsid template. Beyond its role as a polymerase cofactor, it also acts as chaperone for the N-protein in a way that it prevents their illegitimate self-assembly when genomic RNA synthesis does not occur and maintain them in a soluble form (N°) within a complex (N°−P) and used for the encapsidation of Nascent RNA chain during replication.[290] The significant functional importance of N and P protein appears due to their involvement in numerous protein-protein interactions within the internal (viral) and external (Host) PPI networks. Multiple biological functions occur due to this interactability. Including modulation of both acquired and innate immunity. Experiments have proven the abundance of disorder in the N and P protein of these viruses. The persistence of disorder in the C-terminal domain of nucleoprotein (N_{TAIL}), even after complex formation, indicates potential role of this region in binding,[259,291,292] as described in case of MeV NTAIL, whose first 20 amino acids interacts with cellular nucleoprotein receptor[293,294] and C-terminal region interact with the major inducible heat shock protein Hsp 70 that leads to both viral replication and transcription.[295] The disordered nature of N_{TAIL} in measles and Hendra viruses has also confirmed in the context of full-length N protein that formed Nucleocapsid like particle (NLP) when expressed in the heterologous system.[296–298] Initially, it was thought that the C-terminal X domain (XD) of the phosphoprotein triggers major conformational rearrangement within nucleocapsid, and this leads to the access of the viral polymerase to RNA genome.[259,292,299] However recent NMR studies rule out these possibilities and provide the first direct observation of the interaction between XD and intact

nucleocapsid in the Paramyxoviridae. The disordered N_{TAIL} region is partially exposed at the surface of the nucleocapsid and provides a way for interaction with numerous protein partners. Indeed, MeV N_{TAIL} interacts with various viral protein partners, such as P, P-L complex, and matrix protein.[300] Besides interaction with viral components, it also interacts with host cellular components, such as Interferon regulatory factor 3 (IRF3),[301] hsp70,[295] peroxiredoxin 1,[302] casein kinase II,[303] the cell protein responsible for the nuclear export of N,[304] and possibly the components of the cell cytoskeleton.[305,306] Additionally, the N_{TAIL} of MeV nucleocapsid released from infected cells binds to the cell receptors involved in MeV-induced immunosuppression.[293,294] The P protein disorder was reported in both Paramyxoviridae[157,158,307–311] and Rhabdoviridae.[312–314] P protein in the members of these families possesses a very high modular organization that consists of alternate ordered and disordered regions. In Paramyxoviridae, P protein possesses a large disordered region (4000 residues) at its N terminal (PNT) domain. Several interactions made by the PNT domain of MeV and Sendai virus (SeV) have been reported, such as the PNT domain of MeV interactions with N and cellular protein,[315,316] SeV interacts with an unassembled form of N (N°) and L protein.[317,318] While the C-terminus nucleocapsid binding region of P adopts compact folded stable conformations in members of Rhabdoviridae and majority of Paramyxovirinae, it remains disordered in the respiratory syncytial virus which is a member of Pneumovirinae subfamily.[311,319] The N-terminal region of P protein from Rhabdoviridae and Paramyxoviridae that is involved in binding to N° has been reported to contain the α-MoRF.[158,312,313,320] This induced folding upon the binding effect in a form of the α-MoRF is limited to vesicular stomatitis virus (VSV), a rhabdovirus. The structure of VSV N°−P complex was solved and verified that although the binding region adopts an α-helical configuration, the flanking regions remain flexible. P protein α-MoRF binding occurs at the same site that is responsible for RNA and different N protein binding, thereby preventing the polymerization of N protein. These results provide a link between different processes and possibly explain the mechanism of initiation for viral RNA synthesis.[321] In MeV, limited proteolysis study carried out in secondary structure stabilizer (TFE) provided evidence for the disorder to order transition of disordered N-terminal region of P (PNT).[157] The presence of disordered domains in both P and N proteins leads to the controlled dynamic interactions in a coordinated manner between template nucleocapsid surface and polymerase complex that could extend further over the successive turns of the helix. The long

disordered regions in viral proteins enable them to act as a potential linker between the binding partner and participate in large macromolecular assembly acting as a scaffolding engine.[322,323]

10.6 Intrinsic disorder in capsid proteins

IDRs provide more flexibility, hence help in the quick conformational changes of proteins required for the capsid assembly of viruses. For instance, the VP-4 protein of the Foot-and-mouth disease virus (FMDV) contains low structure content however plays a crucial role in capsid assembly.[148] As most viral proteins have synthesized in the form of the polyprotein, the presence of IDRs at the proteolytic sites make digestion easy and faster and generate independent functional chains.[324,325] The presence of IDRs in viral proteins provides a self-driven mechanism of self-assembly due to the aforementioned property.

10.6.1 Intrinsic disorder in Flaviviridae core proteins

Flaviviridae family members are non-segmented single-stranded positive-sense viruses, whose genome size varies between 9.6 and 12.3 kb. Viral genera Flavivirus, Hepacivirus, and Pestivirus come under the family of Flaviviridae.[326] N-terminal region of viral core protein is highly basic and makes interaction in a sequence-specific manner with RNA to accomplish the various functions. The core protein is released from the rest of the polyprotein to initiate the functions required for further maturation and multiplication of viruses. RNA chaperoning activity of core protein is confirmed in in vitro assays, additionally it is responsible for packaging and condensation of viral genomic RNA during viral morphogenesis. Core protein mediates several interactions with host proteins for viral persistence and pathogenicity and simultaneously involves itself in functions related to viral replication.[327] Biophysical and biochemical studied done so far on the Flaviviridae family confirmed the widespread use of core protein IDRs in its member viruses despite having the low sequence similarity and other pronounced differences in their modular organization.[148]

10.6.2 Disordered capsid protein of ZIKV and DENV

Capsid protein of DENV and ZIKV are found to be highly disordered with respect to other proteins encoded by their genome. The disorder content is found to be 33.3% and 36% in ZIKV[146] and DENV, respectively.[328] This high amount of disorder suggests the exclusive involvement of these regions in the mechanism of viral-mediated functions at the battlefront

of host and pathogens. The ZIKV capsid major functions are nucleocapsid assembly and involvement in the viral infection processes by interacting with cellular proteins, modulating cellular metabolism, apoptosis, and immune response.[329] Major functions of the Capsid protein of DENV are RNA binding and RNA chaperone activity, nucleocapsid assembly, lipid droplet accumulation and interaction with host components.[330] Despite major knowledge on the functions and disorder status of capsid proteins of DENV and ZIKV, the exact mechanism of IDR-mediated control of various functions of this protein is yet to be discovered. Fig. 5 demonstrates the MoRF position of (A) ZIKV and (B) DENV capsid proteins predicted by the MoRFchibi SYSTEM HTML server.[331] A pattern of position and number of MoRFs could be analyzed in detail in the capsid proteins of these viruses to identify the factors associated with their specific functions.

10.6.3 The fd phage coat protein pVIII undergoes transitions from order to disorder form

Fd bacteriophage, filamentous in shape, belongs to the Invorus genus and infects enterobacteria, such as *E. coli*.[332,333] The coat protein of Fd phage undergoes the transition from the state of disordered to ordered and ordered to disordered to regulate the molecular mechanism of its penetration and assembly.[148] The structural transition in FdpVIII coat protein indicates that there is involvement of MG (partial disorder) intermediate in the process of macromolecular assembly and disassembly.[334]

10.6.4 Capsid protease: An illustrative example of an intrinsically disordered enzyme in Semliki forest virus

The IDRs play their role in the activation and deactivation of the enzymatic property of viral proteins, as in the case of the Semliki forest virus (SFV). SFV belongs to the Alphavirus genus that has enveloped positive-strand RNA with an icosahedral nucleocapsid and spherical morphology.[335,336] The N-terminal region of SFV polyprotein (residues 1–267) is an intramolecular serine protease that cleaves itself off after the Trp267 from the rest of the polyprotein segment and provides a mature capsid protein. After this auto cleavage process, the free carboxyl group of Trp267 interacts with catalytic triad consisting of amino acid His145, Asp167, and Ser219 and leads to inactivation of the enzyme.[337]

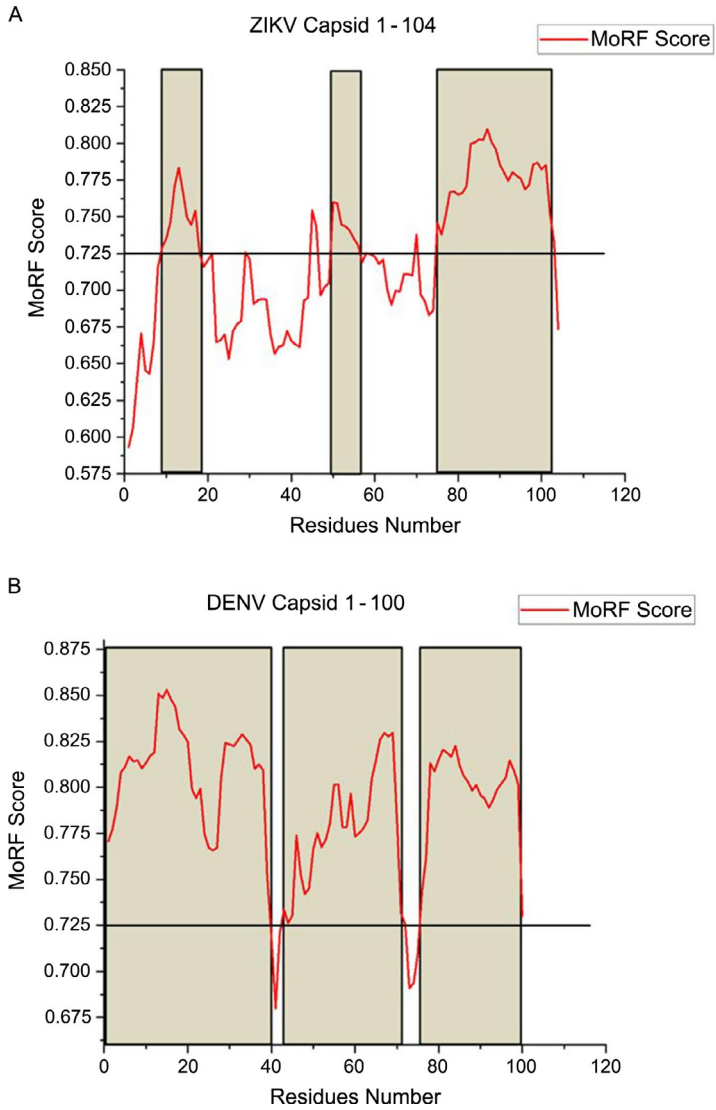

Fig. 5 The capsid protein MoRF position (toggled gray bars) predicted by the MoRFchibi system for (A) ZIKV (UniProt id: Q32ZE1|1-104); (B) DENV (UniProt id: P33478|1-100). Three MoRF regions of different lengths have observed in the capsid protein of both viruses, located within disordered areas, these MoRF regions play a crucial role by recognizing, interacting and inducing a conformational change to viral as well as host proteins.

10.7 Intrinsic disorder in the nucleocapsid protein of SARS-CoA

Nucleocapsid protein (N) of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoA) plays a crucial role in its viability and packaging of its genomic RNA. However, the exact mechanism of binding of N protein to genomic RNA is not completely understood. Two domains present in N protein NTD and CTD are flanked by long stretches of disordered regions that counts for almost half of the entire length. Both domains through their flanking disordered regions bind to RNA. Although low sequence homology reported in different coronavirus N protein through bioinformatics study, flexible linker region of N protein of all coronaviruses started with SR-rich region and end with region enriched with basic residues. These features are the hallmarks of the protein disorder. The overall isoelectric point (pI) of these flexible linkers is high, which is self-explanatory for their RNA binding abilities. The aforementioned findings suggest that the physiochemical features are likely to be conserved across different groups of Coronaviridae. This observation highlights the role of intrinsic disorder in N protein whether it be multisite nucleic acid binding or RNP packaging.[338]

10.8 Intrinsic disorder in influenza virus surface glycoproteins

Surface glycoprotein is required for the fusion of viral membrane with host membrane, hence mediating the way of entrance to the target cell.[339–341] One of the best examples of the most studied membrane fusion proteins is the influenza virus HA. HA is homotrimeric type I transmembrane surface glycoprotein responsible for the binding of viruses to the host receptor, their internalization and subsequent membrane fusion events within the endosome of the infected cell. Presence of HA at the viral surface in high numbers make it the most abundant antigen that contains primary neutralizing epitopes for antibodies.[342] Recent bioinformatics study revealed that although many viral membrane proteins are universally ordered, intrinsic disorder is still present in these proteins pointing out that IDRs might have crucial functions. For instance, influenza A virus virulent strain 1918 H1N1 and H5N1 differ from less virulent or nonvirulent strain H3N2 and 1930 H1N1 in their disordered content of the HA protein.[148]

10.9 Intrinsic disorder influenza virus non-structural protein 2

It has been observed that during viral replication, non-structural protein 2 of the influenza virus interacts with nuclear export machinery. It behaves as an

adaptor molecule between viral ribonucleoprotein complex and the viral nuclear export machinery. Various techniques such as differential scanning calorimetry (DSC), hydrodynamic techniques, and limited proteolysis demonstrated the presence of high levels of disorder in this protein.[343]

10.10 Intrinsic disorder in human adenovirus type 5 early transcription unit 1B

A set of proteins comprises early transcription unit 1B (E1B) encoded by human adenovirus type 5. These proteins participate in several important viral functions, such as viral replication and adenoviral-mediated cell transformation.[344,345] An interesting feature demonstrated by this set of proteins is that they are expressed from the overlapping reading frames of the 2.28 kb E1B-mRNA through alternative splicing that takes place between common splice donor and one splice acceptor site among three possible sites. This results in the encoding of proteins from mRNAs having common N-terminus and different C-terminus.[346,347] This feature determines one of the names of these proteins, E1BN proteins. Computational analysis along with NMR and CD determines that E1B-93R is a typical IDP, and the N-terminal region within E1B and other E1BN proteins is likely to be intrinsically disordered.[345]

10.11 Intrinsic disorder in non-structural HCV proteins

HCV NS5A, a key protein involved viral replication that plays a role in viral particle assembly.[251] Numerous interactions made by NS5A with viral and host proteins have been reported.[348] NS5A is a membrane associated protein that possesses an anchor at its N-terminal region with C-terminal region being divided into three different domains, D1, D2, and D3. D1 is highly conserved and is less disordered, while D2 and D3 are less conserved and are highly disordered.[252–254,349] High disorder content defines the dynamic behavior of D2 and D3 that makes them a hub-like a center for multiple interactions. NS5A-D2 is important for NS5A function and is involved in molecular interaction with RDRP (NS5B) and PKR. The interaction established by NS5A-D2 interferes with host signaling pathways and apoptosis.[256] Although NS5A-D3 is mostly disordered, it contains short ordered elements at its N-terminus. In a recent study, NS5A-D3 proteins from two HCV strains were found to exhibit a propensity to partial folding into an α-helix.[350] NMR analysis revealed two putative α-helices for that a

molecular model could be proposed. The first α-helix conservation in all genotypes and its amphipathic character suggest that it could be corresponding to MoRE and hence promote the interaction with a suitable biological partner(s). One such partner is Cyclophilin A (CypA). Cyclophilins are cell factors crucial in HCV replication. Interestingly, Cyclosporin completely abrogates the interaction between HCV NS5A-D3 and CypA. CypA together with NS5A and NS5B forms the crucial component of multi-protein complex and supports RNA transcription and replication.[350]

10.12 Intrinsic disorder in the HDV basic protein δAg

Among many animal viruses know so far, HDV has the smallest RNA genome that code for single protein known as δ-antigen (δAg).[351] From a structural perspective, this protein comprises of the coiled-coil domain, a nuclear localization signal (NLS) and RNA binding domain.[352] δAg is self-oligomerize to yield dodecamers structure associated with HDV genomic RNA.[248,353] Computational and experimental analysis of eight clades of HDV shows the high disorder of this protein.[248]

10.13 Intrinsic disorder in HIV-1 accessory and regulatory proteins

Tat protein of HIV-1 is an important factor in viral pathogenesis that serves as a transactivator of viral transcription. The activity of Tat is dependent on its interaction with the Transactivation response region (TAR), whose example is a short nascent stem bulge loop leader RNA. TAR present at 5′ extreme of all viral transcripts. Tat protein display typical characteristics of IDPs that include the high net charge to low global hydrophobicity.[148] Intrinsic disorder of Tat is also proven by CD and NMR studies.[354] Rev. protein also plays a regulatory role in HIV-1. This is a basic protein of 116 residues in length that belongs to the ARM family of RNA-binding proteins. Rev. binds to the Rev. Response element (RRE) of viral mRNA in the cytoplasm of the host cell, and, therefore, Rev. is essential for viral replication[355] Monomeric Rev. adopts MG state as confirm by Hydrodynamic and Spectroscopic studies.[356] Recent biophysical studies of Rev. ARM associated with RNA binding suggest it is intrinsically disordered not only in the isolated state but also when embedded into oligomerization deficient Rev. Mutant.[357]

10.14 Intrinsic disorder in non-structural HPV E6 and E7 proteins

The large family of papillomavirus (PV) includes small DNA viruses infecting mammals, reptile, and birds. At least 100 different types of HPV are reported to date that act as a cofactor in the development of carcinoma of head, neck, genital tract and epidermis and also cause the papillomas and benign wart. HPV classified into two classes on the basis of its association with cancer. The first category includes low-risk viruses (HPV-6, HPV-11), and the second category contains high-risk viruses (HPV-16, HPV-18, and HPV-45) types. Similar to all DNA tumor viruses, HPV hijacks the replication machinery and forces the infected cell to enter into the S phase of the cell cycle. The transforming activity of high-risk HPVs is mainly exerted through their E7, which is one of their two oncoproteins. E7 is responsible for pathogenesis and maintenance of human cervical cancer and has been determined to participate in numerous cellular processes including DNA synthesis, transcription, transformation, cell growth, and apoptosis.[358] E7 interacts with Rb, which is a tumor suppressor protein, and interferes with its tumor suppression activity. Rb acts as guardian of the cell cycle due to its involvement into the control of G1/S transition.[359] Therefore, Rb is critical for determining the progression of the cell into the normal phase or transformation. Besides interacting with proteins of Rb family, E7 also interacts with histone deacetylase,[360] kinase p33CDK2 and cyclin A,[361] protein phosphatase 2A (PP2A),[362] and the cyclin-dependent kinase inhibitor p21cip1 protein.[363] PP2A is sequestered and excluded from its interaction with protein kinase B (PKB) or Akt due to its involvement in the formation of a complex with E7.[364] PKB is one of several second messenger kinases that is activated via cell attachment and growth factor signaling and that sends a signal to the cell nucleus to prevent apoptosis, thus leading the way toward cell survival during proliferation. The interaction between PP2A and E7 leads to the inhibition of PKB/Akt dephosphorylation that keeps the PKB/Akt signaling activated. E7 protein broad range molecular interactions depend on the flexible disordered region present within the E7. Previous studies performed on recombinant E7 reveal that its structure can be described as the elongated dimer that changes conformation upon a small change in pH, while gaining α-helicity by exposure to solvents.[365] Biophysical characterization of E7 from HPV-45 with far-UV CD and NMR revealed that its N-terminal region (E7N, amino acids 1–40) is disordered, while its C-terminal domain is well structured (41–98)

with a unique zinc-binding fold. The Intrinsically unstructured N-terminal region of E7 contains binding and Casein kinase II phosphorylation sites.[292,366,367] The CD spectra recorded for the different conformations as a function of temperature and pH indicated a polyproline II-like structure.[366] The structural stability is maintained by phosphorylation that results in increased transformation activity in the cell. Transforming protein E6 and E7 of high-risk HPVs incorporate high amounts of intrinsic disorder.[368]

10.15 Intrinsically unstructured N protein of λ bacteriophage

In λ bacteriophage, its N protein (λN) plays an important role in the transcription of the gene. The absence of this protein leads to the reduction in the phage genome transcription to 2% with the only transcription of the early gene.[369] λN protein positively regulates the transcription of λ bacteriophage and promotes the expression of a gene located downstream to the termination signal. λN acts as an anti-terminator transcription factor and in doing so, it binds to an RNA sequence (the box B segment) and multiple proteins in the transcription complex, where it serves as an important regulator of anti-iterminator complex that allows transcription through termination sites during phage gene expression. The interaction between host bacteria RNA polymerase and factor NusA to λN has been also observed.[161] λN demonstrate all features of unstructured flexible protein that are typical to IDPs. These features include high net charge and low hydrophobicity,[161] as well as structural asymmetry determined through various experiments.[370–374]

10.16 Intrinsic disorder in the Hordeivirus movement TGBp1 protein

Plant viral infection spreads from one infected position to another through special proteins known as movement proteins (MPs) that facilitate the movement of viruses within the plant body. These MPs possess a wide range of functions. They interact with the viral proteins and RNA to form ribonucleoprotein complex that facilitates cell to cell and long-distance movement of the viral genome in the plant and helps in the interaction with cytoskeleton components and endoplasmic reticulum.[161] Three types of movement protein that are TGBp1 (528 residues), TGBp2 (204 residues), and TGBp3 (155 residues), encoded by "triple gene black" (TGB) are reported in hordeiviruses.[375] The N-terminal region of TGB1 of *Barley stripe mosaic virus* (residues 1–180) are predicted to be highly disordered, whereas C-terminal is not as shown in Fig. 6.

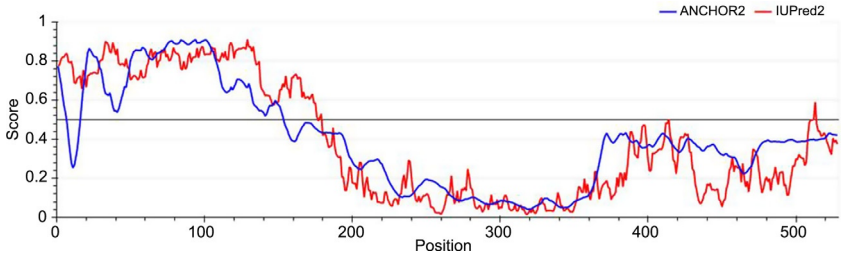

Fig. 6 Intrinsic disorder prediction in TGB1 protein (Uniprot Id: P04867) of *Barley stripe mosaic virus* by IUpred2A server.[237]

11. Summary and outlook

This chapter summarizes the current knowledge on the protein intrinsic disorder phenomenon, discusses various peculiar features of IDPs, including their involvement in PPI networks, other biological roles and introduces different disorder predictors. It also discusses some details of the intrinsic disorder perspective of viruses, the role of IDPs and IDRs in the virus-facilitated host mechanisms, prevalence of the intrinsic disorder in viral proteomes, and functional prominence of disordered viral proteins. The role of IDRs in various structural and non-structural proteins of viruses, such as capsid, nucleocapsid, genome-linked surface glycoproteins, matrix and accessory, and regulatory proteins have been summarized. IDPs/IDRs role in specific function-oriented proteins in different viruses have been elaborated, such as membrane-binding protein λN of bacteriophage, hordeivirus movement protein TGBp1, influenza virus non-structural protein 2, bBasic protein δAg of HDV, and Human adenovirus type 5 early transcription unit 1B. Also, the importance of intrinsic disorder for the alternative splicing and overlapping reading frames of viral proteome is discussed. Viruses mainly cause pathogenesis by hijacking the cell machinery and modulating its functions, e.g., by altering IDP components involved in the host cell cycle control mechanism. Viral IDPs mediate successful infection and regulate pathogenesis at multiple levels. Therefore, the knowledge of intrinsic disorder and structural flexibility in processes of virus-host interaction and associated functions is crucial for better understanding of viral pathogenesis. The involvement of IDPs/IDRs in the mechanism of viral infection is not completely understood. Therefore, this chapter would allow readers to get better understanding of the importance of IDPs/IDRs in various functional mechanisms/viral components, which are essential for the

completion of crucial phases of the viral life cycle. Finally, the IDPs/IDPRs of viruses are considered as potential drug targets, due to their high prevalence in viral proteomes and ubiquitous involvement in host-pathogen mediated regulations. In conclusion, the involvement of IDPs in viral pathogenesis should be solemnly considered for unlocking the complex riddles of viral infection and associated patterns, their cellular control, and exploitation strategies, and drug development approach in near future by targeting their disordered regions.

Acknowledgments

The authors acknowledge the funding support from the Council of Scientific and Industrial Research (CSIR), India. The file no. for the same is 09/1058(0013)/2019-EMR-I. The authors also acknowledge the host support and facilities provided by the Indian Institute of Technology Mandi, Himachal Pradesh, India. PMM acknowledge the cooperation of Mr. Kundlik while preparing the list of abbreviation and citing the part of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Tamarozzi ER, Giuliatti S. Understanding the role of intrinsic disorder of viral proteins in the oncogenicity of different types of HPV. *Int J Mol Sci*. 2018;19:E198. https://doi.org/10.3390/ijms19010198.
2. Uversky VN, Dunker AK. Understanding protein non-folding. *Biochim Biophys Acta Protein Proteomics*. 2010;1804:1231–1264. https://doi.org/10.1016/j.bbapap.2010.01.017.
3. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*. 2005;6:197–208. https://doi.org/10.1038/nrm1589.
4. Tompa P. *Structure and Function of Intrinsically Disordered Proteins*. Chapman & Hall/CRC Press; 2010.
5. Schweers O, Schönbrunn-Hanebeck E, Marx A, Mandelkow E. Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for β-structure. *J Biol Chem*. 1994;269:24290–24297.
6. Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry*. 1996;35:13709–13715. https://doi.org/10.1021/bi961799n.
7. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*. 1999;293:321–331. https://doi.org/10.1006/jmbi.1999.3110.
8. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D 2 concept. *Annu Rev Biophys*. 2008;37:215–246. https://doi.org/10.1146/annurev.biophys.37.032807.125924.
9. Livesay DR. Protein dynamics: dancing on an ever-changing free energy stage. *Curr Opin Pharmacol*. 2010;10:706–708. https://doi.org/10.1016/j.coph.2010.09.015.

10. Dunker AK, Uversky VN. Drugs for "protein clouds": targeting intrinsically disordered transcription factors. *Curr Opin Pharmacol*. 2010;10:782–788. https://doi.org/10.1016/j.coph.2010.09.005.
11. Uversky VN. Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chem Soc Rev*. 2011;40:1623–1634. https://doi.org/10.1039/c0cs00057d.
12. Tsvetkov P, Asher G, Paz A, et al. Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome. *Proteins Struct Funct Genet*. 2008;70:1357–1366. https://doi.org/10.1002/prot.21614.
13. Uversky VN. Flexible nets of malleable guardians: Intrinsically disordered chaperones in neurodegenerative diseases. *Chem Rev*. 2011;111:1134–1166. https://doi.org/10.1021/cr100186d.
14. Tóth-Petróczy Á, Oldfield CJ, Simon I, et al. Malleable machines in transcription regulation: the mediator complex. *PLoS Comput Biol*. 2008;4:e1000243. https://doi.org/10.1371/journal.pcbi.1000243.
15. Fuxreiter M, Tompa P, Simon I, Uversky VN, Hansen JC, Asturias FJ. Malleable machines take shape in eukaryotic transcriptional regulation. *Nat Chem Biol*. 2008;4:728–737. https://doi.org/10.1038/nchembio.127.
16. Uversky VN. A protein-chameleon: conformational plasticity of α-synuclein, a disordered protein involved in neurodegenerative disorders. *J Biomol Struct Dyn*. 2003;21:211–234. https://doi.org/10.1080/07391102.2003.10506918.
17. Chen J, Liang H, Fernández A. Protein structure protection commits gene expression patterns. *Genome Biol*. 2008;9:R107. https://doi.org/10.1186/gb-2008-9-7-r107.
18. Dunker AK, Obradovic Z. The protein trinity—linking function and disorder. *Nat Biotechnol*. 2001;19:805–806. https://doi.org/10.1038/nbt0901-805.
19. Pullen RA, Jenkins JA, Tickle IJ, Wood SP, Blundell TL. The relation of polypeptide hormone structure and flexibility to receptor binding: the relevance of X-ray studies on insulins, glucagon and human placental lactogen. *Mol Cell Biochem*. 1975;8:5–20. https://doi.org/10.1007/BF01731645.
20. Cary PD, Moss T, Bradbury EM. High-resolution proton-magnetic-resonance studies of chromatin core particles. *Eur J Biochem*. 1978;89:475–482. https://doi.org/10.1111/j.1432-1033.1978.tb12551.x.
21. Dziedzic-Letka A, Rymarczyk G, Kapłon TM, et al. Intrinsic disorder of Drosophila melanogaster hormone receptor 38 N-terminal domain. *Proteins Struct Funct Bioinf*. 2011;79:376–392. https://doi.org/10.1002/prot.22887.
22. Holt C, Sawyer L. Caseins as rheomorphic proteins: interpretation of primary and secondary structures of the αs1-, β- and κ-caseins. *J Chem Soc Faraday Trans*. 1993;89:2683–2692. https://doi.org/10.1039/FT9938902683.
23. Xue B, Mizianty MJ, Kurgan L, Uversky VN. Protein intrinsic disorder as a flexible armor and a weapon of HIV-1. *Cell Mol Life Sci*. 2012;69:1211–1259. https://doi.org/10.1007/s00018-011-0859-3.
24. 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. https://doi.org/10.11234/gi1990.11.161.
25. 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. https://doi.org/10.1016/j.molbiopara.2006.08.011.
26. Galea CA, High AA, Obenauer JC, et al. Large-scale analysis of thermostable, mammalian proteins provides insights into the intrinsically disordered proteome. *J Proteome Res*. 2009;8:211–226. https://doi.org/10.1021/pr800308v.

27. Tompa P, Dosztányi Z, Simon I. Prevalent structural disorder in E. coli and S. cerevisiae proteomes. *J Proteome Res*. 2006;5:1996–2000. https://doi.org/10.1021/pr0600881.
28. 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. https://doi.org/10.1016/j.jmb.2004.02.002.
29. Xue B, Williams RW, Oldfield CJ, Dunker AK, Uversky VN. Archaic chaos: intrinsically disordered proteins in archaea. *BMC Syst Biol*. 2010;4:S1. https://doi.org/10.1186/1752-0509-4-S1-S1.
30. Peng Z, Yan J, Fan X, et al. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci*. 2014;72:137–151. https://doi.org/10.1007/s00018-014-1661-9.
31. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn*. 2012;30:137–149. https://doi.org/10.1080/07391102.2012.675145.
32. Madan Babu M, Van Der Lee R, Sanchez De Groot N, Rg Gsponer J, Gough J, Dunker K. Intrinsically disordered proteins: regulation and disease this review comes from a themed issue on sequences and topology edited. *Curr Opin Struct Biol*. 2011;21:1–9. https://doi.org/10.1016/j.sbi.2011.03.011.
33. Williams RM, Obradovi Z, Mathura V, et al. The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac Symp Biocomput*. 2001;6:89–100. https://doi.org/10.1142/9789814447362_0010.
34. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins Struct Funct Genet*. 2001;42:38–48. https://doi.org/10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3.
35. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. *Biophys J*. 2007;92:1439–1456. https://doi.org/10.1529/biophysj.106.094045.
36. Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinfor*. 2007;8:211. https://doi.org/10.1186/1471-2105-8-211.
37. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001;411:41–42. https://doi.org/10.1038/35075138.
38. 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. https://doi.org/10.1111/j.1742-4658.2005.04948.x.
39. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit*. 2005;18:343–384. https://doi.org/10.1002/jmr.747.
40. Ekman D, Light S, Björklund ÅK, Elofsson A. What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? *Genome Biol*. 2006;7:R45. https://doi.org/10.1186/gb-2006-7-6-r45.
41. Haynes C, Oldfield CJ, Ji F, et al. Intrinsic disorder is a common feature of hub proteins from four eukaryotic Interactomes. *PLoS Comput Biol*. 2006;2:e100. https://doi.org/10.1371/journal.pcbi.0020100.
42. Patil A, Nakamura H. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett*. 2006;580:2041–2045. https://doi.org/10.1016/j.febslet.2006.03.003.
43. Dosztányi Z, Chen J, Dunker AK, Simon I, Tompa P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res*. 2006;5:2985–2995. https://doi.org/10.1021/pr060171o.

44. Singh GP, Ganapathi M, Dash D. Role of intrinsic disorder in transient interactions of hub proteins. *Proteins Struct Funct Bioinf*. 2006;66:761–765. https://doi.org/10.1002/prot.21281.
45. Singh GP, Dash D. Intrinsic disorder in yeast transcriptional regulatory network. *Proteins Struct Funct Bioinf*. 2007;68:602–605. https://doi.org/10.1002/prot.21497.
46. Uversky VN. Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol*. 2011;43:1090–1103. https://doi.org/10.1016/j.biocel.2011.04.001.
47. Patil A, Kinoshita K, Nakamura H. Hub promiscuity in protein-protein interaction networks. *Int J Mol Sci*. 2010;11:1930–1943. https://doi.org/10.3390/ijms11041930.
48. Uversky VN, Gillespie JR, Fink AL. Why are natively unfolded proteins unstructured under physiologic conditions? *Proteins Struct Funct Genet*. 2000;41:415–427. https://doi.org/10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7.
49. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci U S A*. 2010;107:8183–8188. https://doi.org/10.1073/pnas.0911107107.
50. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc Natl Acad Sci U S A*. 2006;103:16764–16769. https://doi.org/10.1073/pnas.0608175103.
51. Wang X, Vitalis A, Wyczalkowski MA, Pappu RV. Characterizing the conformational ensemble of monomeric polyglutamine. *Proteins Struct Funct Genet*. 2006;63:297–311. https://doi.org/10.1002/prot.20761.
52. Möglich A, Joder K, Kiefhaber T. End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc Natl Acad Sci U S A*. 2006;103:12394–12399. https://doi.org/10.1073/pnas.0604748103.
53. Vitalis A, Wang X, Pappu RV. Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories. *Biophys J*. 2007;93:1923–1937. https://doi.org/10.1529/biophysj.107.110080.
54. Mukhopadhyay S, Krishnan R, Lemke EA, Lindquist S, Deniz AA. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc Natl Acad Sci U S A*. 2007;104:2649–2654. https://doi.org/10.1073/pnas.0611503104.
55. Dougan L, Li J, Badilla CL, Berne BJ, Fernandez JM. Single homopolypeptide chains collapse into mechanically rigid conformations. *Proc Natl Acad Sci U S A*. 2009;106:12605–12610. https://doi.org/10.1073/pnas.0900678106.
56. Walters RH, Murphy RM. Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. *J Mol Biol*. 2009;393:978–992. https://doi.org/10.1016/j.jmb.2009.08.034.
57. Midic U, Oldfield CJ, Keith AK, Obradovic Z, Uversky VN. Protein disorder in the human diseasome: unfoldomics of human genetic diseases. *BMC Genomics*. 2009;10:S12. https://doi.org/10.1186/1471-2164-10-S1-S12.
58. Lee H, Mok KH, Muhandiram R, et al. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J Biol Chem*. 2000;275:29426–29432. https://doi.org/10.1074/jbc.M003107200.
59. Adkins JN, Lumb KJ. Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57Kip2. *Proteins Struct Funct Genet*. 2002;46:1–7. https://doi.org/10.1002/prot.10018.
60. Campbell KM, Terrell AR, Laybourn PJ, Lumb KJ. Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos. *Biochemistry*. 2000;39:2708–2713. https://doi.org/10.1021/bi9923555.

61. Chang BS, Minn AJ, Muchmore SW, Fesik SW, Thompson CB. Identification of a novel regulatory domain in Bcl-x(L) and Bcl-2. *EMBO J*. 1997;16:968–977. https://doi.org/10.1093/emboj/16.5.968.
62. Sunde M, McGrath KCY, Young L, et al. TC-1 is a novel tumorigenic and natively disordered protein associated with thyroid cancer. *Cancer Res*. 2004;64:2766–2773. https://doi.org/10.1158/0008-5472.can-03-2093.
63. Wisniewski KE, Dalton AJ, McLachlan DRC, Wen GY, Wisniewski HM. Alzheimer's disease in down's syndrome: clinicopathologic studies. *Neurology*. 1985;35:957–961. https://doi.org/10.1212/wnl.35.7.957.
64. Glenner GG, Wong CW. Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. *Biochem Biophys Res Commun*. 1984;122:1131–1135. https://doi.org/10.1016/0006-291X(84)91209-9.
65. Masters CL, Multhaup G, Simms G, Pottgiesser J, Martins RN, Beyreuther K. Neuronal origin of a cerebral amyloid: neurofibrillary tangles of Alzheimer's disease contain the same protein as the amyloid of plaque cores and blood vessels. *EMBO J*. 1985;4:2757–2763. https://doi.org/10.1002/j.1460-2075.1985.tb04000.x.
66. Lee VMY, Balin BJ, Otvos L, Trojanowski JQ. A68: a major subunit of paired helical filaments and derivatized forms of normal tau. *Science*. 1991;251:675–678. https://doi.org/10.1126/science.1899488.
67. Ueda K, Fukushima H, Masliah E, et al. Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc Natl Acad Sci U S A*. 1993;90:11282–11286. https://doi.org/10.1073/pnas.90.23.11282.
68. Zoghbi HY, Orr HT. Polyglutamine diseases: protein cleavage and aggregation. *Curr Opin Neurobiol*. 1999;9:566–570. https://doi.org/10.1016/S0959-4388(99)00013-6.
69. Dev KK, Hofele K, Barbieri S, Buchman VL, Van Der Putten H. Part II: α-synuclein and its molecular pathophysiological role in neurodegenerative disease. *Neuropharmacology*. 2003;45:14–44. https://doi.org/10.1016/S0028-3908(03)00140-0.
70. Prusiner SB. Shattuck lecture—neurodegenerative diseases and prions. *N Engl J Med*. 2001;344:1516–1526. https://doi.org/10.1056/NEJM200105173442006.
71. Goh GKM, Dunker AK, Foster JA, Uversky VN. HIV vaccine mystery and viral shell disorder. *Biomolecules*. 2019;9:E178. https://doi.org/10.3390/biom9050178.
72. Choy WY, Forman-Kay JD. Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J Mol Biol*. 2001;308:1011–1032. https://doi.org/10.1006/jmbi.2001.4750.
73. Huang A, Stultz CM. The effect of a ΔK280 mutation on the unfolded state of a microtubule-binding repeat in tau. *PLoS Comput Biol*. 2008;4:e1000155. https://doi.org/10.1371/journal.pcbi.1000155.
74. Gong H, Zhang S, Wang J, Gong H, Zeng J. Constructing structure ensembles of intrinsically disordered proteins from chemical shift data. *J Comput Biol*. 2016;23:300–310. https://doi.org/10.1089/cmb.2015.0184.
75. Fisher CK, Stultz CM. Constructing ensembles for intrinsically disordered proteins. *Curr Opin Struct Biol*. 2011;21:426–431. https://doi.org/10.1016/j.sbi.2011.04.001.
76. Ziegler Z, Schmidt M, Gurry T, Burger V, Stultz CM. Mollack: a web server for the automated creation of conformational ensembles for intrinsically disordered proteins. *Bioinformatics*. 2016;32:2545–2547. https://doi.org/10.1093/bioinformatics/btw200.
77. Bokor M, Csizmók V, Kovács D, et al. NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. *Biophys J*. 2005;88:2030–2037. https://doi.org/10.1529/biophysj.104.051912.
78. Csizmók V, Bokor M, Bánki P, et al. Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. *Biochemistry*. 2005;44:3955–3964. https://doi.org/10.1021/bi047817f.

79. Tompa P, Bánki P, Bokor M, et al. Protein-water and protein-buffer interactions in the aqueous solution of an intrinsically unstructured plant dehydrin: NMR intensity and DSC aspects. *Biophys J*. 2006;91:2243–2249. https://doi.org/10.1529/biophysj.106.084723.
80. Awile O, Krisko A, Sbalzarini IF, Zagrovic B. Intrinsically disordered regions may lower the hydration free energy in proteins: a case study of nudix hydrolase in the bacterium Deinococcus radiodurans. *PLoS Comput Biol*. 2010;6:12. https://doi.org/10.1371/journal.pcbi.1000854.
81. Dunker AK, Lawson JD, Brown CJ, et al. Intrinsically disordered protein. *J Mol Graph Model*. 2001;3263:26–59. https://doi.org/10.1016/S1093-3263(00)00138-8.
82. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol*. 2002;323:573–584. https://doi.org/10.1016/S0022-2836(02)00969-5.
83. Karush F. Heterogeneity of the binding sites of bovine serum albumin [1]. *J Am Chem Soc*. 1950;72:2705–2713. https://doi.org/10.1021/ja01162a099.
84. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A*. 1996;93:11504–11509. https://doi.org/10.1073/pnas.93.21.11504.
85. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with α-helix-forming molecular recognition elements. *Biochemistry*. 2005;44:12454–12470. https://doi.org/10.1021/bi050736e.
86. Mishra PM, Uversky VN, Giri R. Molecular recognition features in Zika virus proteome. *J Mol Biol*. 2018;430:2372–2388. https://doi.org/10.1016/j.jmb.2017.10.018.
87. Lee S, Sowa ME, Watanabe YH, et al. The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell*. 2003;115:229–240. https://doi.org/10.1016/S0092-8674(03)00807-9.
88. Brodersen DE, Clemons WM, Carter AP, Wimberly BT, Ramakrishnan V. Crystal structure of the 30 S ribosomal subunit from Thermus thermophilus: structure of the proteins and their interactions with 16 S RNA. *J Mol Biol*. 2002;316:725–768. https://doi.org/10.1006/jmbi.2001.5359.
89. Luscombe NM, Austin SE, Berman HM, Thornton JM. An overview of the structures of protein-DNA complexes. *Genome Biol*. 2000;1:reviews001.1–001.37. https://doi.org/10.1186/gb-2000-1-1-reviews001.
90. Graham TA, Weaver C, Mao F, Kimelman D, Xu W. Crystal structure of a β-catenin/Tcf complex. *Cell*. 2000;103:885–896. https://doi.org/10.1016/S0092-8674(00)00192-6.
91. Galea CA, Nourse A, Wang Y, Sivakolundu SG, Heller WT, Kriwacki RW. Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27Kip1. *J Mol Biol*. 2008;376:827–838. https://doi.org/10.1016/j.jmb.2007.12.016.
92. Galea CA, Wang Y, Sivakolundu SG, Kriwacki RW. Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. *Biochemistry*. 2008;47:7598–7609. https://doi.org/10.1021/bi8006803.
93. Low HH, Moncrieffe MC, Löwe J. The crystal structure of ZapA and its modulation of FtsZ polymerisation. *J Mol Biol*. 2004;341:839–852. https://doi.org/10.1016/j.jmb.2004.05.031.
94. Xun Z, Saghi G, Harvey L, Malashkevich VN, Kim PS. Structure of the bcr-abl oncoprotein oligomerization domain. *Nat Struct Biol*. 2002;9:117–120. https://doi.org/10.1038/nsb747.
95. Liu X, Xu L, Liu Y, et al. Crystal structure of the hexamer of human heat shock factor binding protein 1. *Proteins Struct Funct Bioinf*. 2009;75:1–11. https://doi.org/10.1002/prot.22216.

96. Malashkevich VN, Schneider BJ, Mcnally ML, Milhollen MA, Pang JX, Kim PS. Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-Å resolution. *Proc Natl Acad Sci U S A*. 1999;96:2662–2667. https://doi.org/10.1073/pnas.96.6.2662.
97. Malashkevich VN, Singh M, Kim PS. The trimer-of-hairpins motif in membrane fusion: visna virus. *Proc Natl Acad Sci U S A*. 2001;98:8502–8506. https://doi.org/10.1073/pnas.151254798.
98. Teschke CM, King J. Folding and assembly of oligomeric proteins in Escherichia coli. *Curr Opin Biotechnol*. 1992;3:468–473. https://doi.org/10.1016/0958-1669(92)90073-r.
99. Xu D, Tsai C-J, Nussinov R. Mechanism and evolution of protein dimerization. *Protein Sci*. 1998;7:533–544. https://doi.org/10.1002/pro.5560070301.
100. Gunasekaran K, Tsai CJ, Nussinov R. Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *J Mol Biol*. 2004;341:1327–1341. https://doi.org/10.1016/j.jmb.2004.07.002.
101. Kajava AV, Baxa U, Steven AC. β arcades: recurring motifs in naturally occurring and disease-related amyloid fibrils. *FASEB J*. 2010;24:1311–1319. https://doi.org/10.1096/fj.09-145979.
102. Wolf E, Kim PS, Berger B. MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci*. 1997;6:1179–1189. https://doi.org/10.1002/pro.5560060606.
103. Mason JM, Arndt KM. Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem*. 2004;5:170–176. https://doi.org/10.1002/cbic.200300781.
104. Liu J, Zheng Q, Deng Y, Cheng CS, Kallenbach NR, Lu M. A seven-helix coiled coil. *Proc Natl Acad Sci U S A*. 2006;103:15457–15462. https://doi.org/10.1073/pnas.0604871103.
105. Özbek S, Engel J, Stetefeld J. Storage function of cartilage oligomeric matrix protein: the crystal structure of the coiled-coil domain in complex with vitamin D3. *EMBO J*. 2002;21:5960–5968. https://doi.org/10.1093/emboj/cdf628.
106. Glover JNM, Harrison SC. Crystal structure of the heterodimeric bZIP transcription factor c-Fos–c-Jun bound to DNA. *Nature*. 1995;373:257–261. https://doi.org/10.1038/373257a0.
107. Im YJ, Kang GB, Lee JH, et al. Structural basis for asymmetric association of the βPIX coiled coil and shank PDZ. *J Mol Biol*. 2010;397:457–466. https://doi.org/10.1016/j.jmb.2010.01.048.
108. Siegert R, Leroux MR, Scheufler C, Hartl FU, Moarefi I. Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. *Cell*. 2000;103:621–632. https://doi.org/10.1016/S0092-8674(00)00165-3.
109. Uversky VN. Neuropathology, biochemistry, and biophysics of α-synuclein aggregation. *J Neurochem*. 2007;103:17–37. https://doi.org/10.1111/j.1471-4159.2007.04764.x.
110. Oldfield CJ, Meng J, Yang JY, Qu 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:S1. https://doi.org/10.1186/1471-2164-9-S1-S1.
111. V.N. Uversky, Alpha-synuclein misfolding and neurodegenerative diseases., Curr Protein Pept Sci 9. (2008) 507–40. https://www.ncbi.nlm.nih.gov/pubmed/18855701 (accessed November 10, 2019).
112. Dunker AK, Bondos SE, Huang F, Oldfield CJ. Intrinsically disordered proteins and multicellular organisms. *Semin Cell Dev Biol*. 2015;37:44–55. https://doi.org/10.1016/j.semcdb.2014.09.025.
113. Yoon MK, Mitrea DM, Ou L, Kriwacki RW. Cell cycle regulation by the intrinsically disordered proteins p21 and p27. *Biochem Soc Trans*. 2012;40:981–988. https://doi.org/10.1042/BST20120092.

114. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol*. 2015;16:18–29. https://doi.org/10.1038/nrm3920.
115. Uversky VN. Functional roles of transiently and intrinsically disordered regions within proteins. *FEBS J*. 2015;282:1182–1189. https://doi.org/10.1111/febs.13202.
116. He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK. Predicting intrinsic disorder in proteins: an overview. *Cell Res*. 2009;19:929–949. https://doi.org/10.1038/cr.2009.87.
117. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. *Bioinformatics*. 2004;20:2138–2139. https://doi.org/10.1093/bioinformatics/bth195.
118. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, et al. FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*. 2005;21:3435–3438. https://doi.org/10.1093/bioinformatics/bti537.
119. Dosztányi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*. 2005;21:3433–3434. https://doi.org/10.1093/bioinformatics/bti541.
120. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. *Structure*. 2003;11:1453–1459. https://doi.org/10.1016/j.str.2003.10.002.
121. Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*. 2005;21:3369–3376. https://doi.org/10.1093/bioinformatics/bti534.
122. Pazos F, Pietrosemoli N, García-Martín JA, Solano R. Protein intrinsic disorder in plants. *Front Plant Sci*. 2013;4:363. https://doi.org/10.3389/fpls.2013.00363.
123. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*. 2002;11:739–756. https://doi.org/10.1110/ps.4210102.
124. Kosol S, Contreras-Martos S, Cedeño C, Tompa P. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules*. 2013;18:10802–10828. https://doi.org/10.3390/molecules180910802.
125. Belle V, Rouger S, Costanzo S, Longhi S, Fournel A. Site-directed spin labeling EPR spectroscopy. In: *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation*. John Wiley and Sons; 2010:131–169. https://doi.org/10.1002/9780470602614.ch6.
126. Habchi J, Martinho M, Gruet A, Guigliarelli B, Longhi S, Belle V. Monitoring structural transitions in IDPs by site-directed spin labeling EPR spectroscopy. *Methods Mol Biol*. 2012;895:361–386. https://doi.org/10.1007/978-1-61779-927-3_21.
127. Drescher M. EPR in protein science: intrinsically disordered proteins. *Top Curr Chem*. 2012;321:91–120. https://doi.org/10.1007/128_2011_235.
128. Lorenzi M, Puppo C, Lebrun R, et al. Tyrosine-targeted spin labeling and EPR spectroscopy: an alternative strategy for studying structural transitions in proteins. *Angew Chemie—Int Ed*. 2011;50:9108–9111. https://doi.org/10.1002/anie.201102539.
129. Mileo E, Etienne E, Martinho M, et al. Enlarging the panoply of site-directed spin labeling electron paramagnetic resonance (SDSL-EPR): sensitive and selective spin-labeling of tyrosine using an Isoindoline-based Nitroxide. *Bioconjug Chem*. 2013;24:1110–1117. https://doi.org/10.1021/bc4000542.
130. Ferreon ACM, Moran CR, Gambin Y, Deniz AA. Single-molecule fluorescence studies of intrinsically disordered proteins. *Methods Enzymol*. 2010;472:179–204. https://doi.org/10.1021/cr400297g.
131. Milles S, Koehler C, Gambin Y, Deniz AA, Lemke EA. Intramolecular three-colour single pair FRET of intrinsically disordered proteins with increased dynamic range. *Mol Biosyst*. 2012;8:2531–2534. https://doi.org/10.1039/c2mb25135c.

132. Ferreon ACM, Ferreon JC, Wright PE, Deniz AA. Modulation of allostery by protein intrinsic disorder. *Nature*. 2013;498:390–394. https://doi.org/10.1038/nature12294.
133. Miyagi A, Tsunaka Y, Uchihashi T, et al. Visualization of intrinsically disordered regions of proteins by high-speed atomic force microscopy. *ChemPhysChem*. 2008;9:1859–1866. https://doi.org/10.1002/cphc.200800210.
134. Ando T, Kodera N. Visualization of mobility by atomic force microscopy. *Methods Mol Biol*. 2012;896:57–69. https://doi.org/10.1007/978-1-4614-3704-8_4.
135. Sandal M, Valle F, Tessari I, et al. Conformational equilibria in monomeric α-synuclein at the single-molecule level. *PLoS Biol*. 2008;6:e6. https://doi.org/10.1371/journal.pbio.0060006.
136. Ando T, Uchihashi T, Kodera N. High-speed AFM and applications to biomolecular systems. *Annu Rev Biophys*. 2013;42:393–414. https://doi.org/10.1146/annurev-biophys-083012-130324.
137. Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci*. 2002;27:527–533. https://doi.org/10.1016/S0968-0004(02)02169-2.
138. Lynch WP, Riseman VM, Bretscher A. Smooth muscle caldesmon is an extended flexible monomeric protein in solution that can readily undergo reversible intra- and intermolecular sulfhydryl cross-linking. A mechanism for caldesmon's F-actin bundling activity. *J Biol Chem*. 1987;262:7429–7437. http://www.ncbi.nlm.nih.gov/pubmed/3584120. Accessed 10 November 2019.
139. Hernandez MA, Avila J, Andreu JM. Physicochemical characterization of the heat-stable microtubule-associated protein MAP2. *Eur J Biochem*. 1986;154:41–48. https://doi.org/10.1111/j.1432-1033.1986.tb09356.x.
140. Etoh Y, Simon M, Green H. Involucrin acts as a transglutaminase substrate at multiple sites. *Biochem Biophys Res Commun*. 1986;136:51–56. https://doi.org/10.1016/0006-291x(86)90875-2.
141. Kalthoff C. A novel strategy for the purification of recombinantly expressed unstructured protein domains. *J Chromatogr B Anal Technol Biomed Life Sci*. 2003;786:247–254. https://doi.org/10.1016/S1570-0232(02)00908-X, Elsevier.
142. Cortese MS, Baird JP, Uversky VN, Dunker AK. Uncovering the unfoldome: enriching cell extracts for unstructured proteins by acid treatment. *J Proteome Res*. 2005;4:1610–1618. https://doi.org/10.1021/pr050119c.
143. Receveur-Bréhot V, Bourhis JM, Uversky VN, Canard B, Longhi S. Assessing protein disorder and induced folding. *Proteins Struct Funct Genet*. 2006;62:24–45. https://doi.org/10.1002/prot.20750.
144. Livernois AM, Hnatchuk DJ, Findlater EE, Graether SP. Obtaining highly purified intrinsically disordered protein by boiling lysis and single step ion exchange. *Anal Biochem*. 2009;392:70–76. https://doi.org/10.1016/j.ab.2009.05.023.
145. Habchi J, Tompa P, Longhi S, Uversky VN. Introducing protein intrinsic disorder. *Chem Rev*. 2014;114:6561–6588. https://doi.org/10.1021/cr400514h.
146. Giri R, Kumar D, Sharma N, Uversky VN. Intrinsically disordered side of the Zika virus proteome. *Front Cell Infect Microbiol*. 2016;6:144. https://doi.org/10.3389/fcimb.2016.00144.
147. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics*. 1998;148:1667–1686.
148. Xue B, Blocquel D, Habchi J, et al. Structural disorder in viral proteins. *Chem Rev*. 2014;114:6880–6911. https://doi.org/10.1021/cr4005692.
149. Otieno S, Kriwacki R. Probing the role of nascent helicity in p27 function as a cell cycle regulator. *PLoS One*. 2012;7:e47177. https://doi.org/10.1371/journal.pone.0047177.
150. Hagai T, Azia A, Babu MM, Andino R. Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. *Cell Rep*. 2014;7:1729–1739. https://doi.org/10.1016/j.celrep.2014.04.052.

151. Duro N, Miskei M, Fuxreiter M. Fuzziness endows viral motif-mimicry. *Mol Biosyst*. 2015;11:2821–2829. https://doi.org/10.1039/c5mb00301f.
152. Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. *Trends Biochem Sci*. 2011;36:159–169. https://doi.org/10.1016/j.tibs.2010.10.002.
153. Tarakhovsky A, Prinjha RK. Drawing on disorder: how viruses use histone mimicry to their advantage. *J Exp Med*. 2018;215:1777–1787. https://doi.org/10.1084/jem.20180099.
154. Garcia-Moreno M, Järvelin AI, Castello A. Unconventional RNA-binding proteins step into the virus–host battlefront. *Wiley Interdiscip Rev RNA*. 2018;9:e1498. https://doi.org/10.1002/wrna.1498.
155. Dyson HJ, Wright PE. How do intrinsically disordered viral proteins hijack the cell? *Biochemistry*. 2018;57:4045–4046. https://doi.org/10.1021/acs.biochem.8b00622.
156. Thomas G, Aslan JE, Thomas L, Shinde P, Shinde U, Simmen T. Caught in the act - protein adaptation and the expanding roles of the PACS proteins in tissue homeostasis and disease. *J Cell Sci*. 2017;130:1865–1876. https://doi.org/10.1242/jcs.199463.
157. Karlin D, Longhi S, Receveur V, Canard B. The n-terminal domain of the phosphoprotein of morbilliviruses belongs to the natively unfolded class of proteins. *Virology*. 2002;296:251–262. https://doi.org/10.1006/viro.2001.1296.
158. Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in Paramyxovirinae N and P. *J Gen Virol*. 2003;84:3239–3252. https://doi.org/10.1099/vir.0.19451-0.
159. Longhi S, Receveur-Bréchot V, Karlin D, et al. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J Biol Chem*. 2003;278:18638–18648. https://doi.org/10.1074/jbc.M300518200.
160. Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS. Do viral proteins possess unique biophysical features? *Trends Biochem Sci*. 2009;34:53–59. https://doi.org/10.1016/j.tibs.2008.10.009.
161. Uversky VN, Longhi S. *Flexible Viruses : Structural Disorder in Viral Proteins*. Wiley; 2012.
162. Xue B, Williams RW, Oldfield CJ, Kian-Meng Goh G, Keith Dunker A, Uversky VN. Viral disorder or disordered viruses: do viral proteins possess unique features? *Protein Pept Lett*. 2010;17:932–951. https://doi.org/10.2174/092986610791498984.
163. Alves C, Cunha C. Order and disorder in viral proteins: new insights into an old paradigm. *Future Virol*. 2012;7:1183–1191. https://doi.org/10.2217/fvl.12.114.
164. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol*. 2005;13:278–284. https://doi.org/10.1016/j.tim.2005.04.003.
165. Sano E, Carlson S, Wegley L, Rohwer F. Movement of viruses between biomes. *Appl Environ Microbiol*. 2004;70:5842–5846. https://doi.org/10.1128/AEM.70.10.5842-5846.2004.
166. Edwards RA, Rohwer F. Opinion: viral metagenomics. *Nat Rev Microbiol*. 2005;3:504–510. https://doi.org/10.1038/nrmicro1163.
167. Bergh Ø, Børsheim KY, Bratbak G, Heldal M. High abundance of viruses found in aquatic environments. *Nature*. 1989;340:467–468. https://doi.org/10.1038/340467a0.
168. Lawrence CM, Menon S, Eilers BJ, et al. Structural and functional studies of archaeal viruses. *J Biol Chem*. 2009;284:12599–12603. https://doi.org/10.1074/jbc.R800078200.
169. La Scola B, Desnues C, Pagnier I, et al. The virophage as a unique parasite of the giant mimivirus. *Nature*. 2008;455:100–104. https://doi.org/10.1038/nature07218.
170. La Scola B, Audic S, Robert C, et al. A giant virus in amoebae. *Science*. 2003;299:2033. https://doi.org/10.1126/science.1081867.
171. Raoult D, Audic S, Robert C, et al. The 1.2-megabase genome sequence of Mimivirus. *Science*. 2004;306:1344–1350. https://doi.org/10.1126/science.1101485.

172. Koonin EV. Virology: gulliver among the lilliputians. *Curr Biol*. 2005;15https://doi.org/10.1016/j.cub.2005.02.042.
173. Suzan-Monti M, La Scola B, Barrassi L, Espinosa L, Raoult D. Ultrastructural characterization of the giant volcano-like virus factory of Acanthamoeba polyphaga Mimivirus. *PLoS One*. 2007;2:e328. https://doi.org/10.1371/journal.pone.0000328.
174. Baltimore D. Expression of animal virus genomes. *Bacteriol Rev*. 1971;35:235–241.
175. Holmes EC. Viral evolution in the genomic age. *PLoS Biol*. 2007;5:e278. https://doi.org/10.1371/journal.pbio.0050278.
176. Rybicki E. The classification of organisms at the edge of life or problems with virus systematics. *S Afr J Sci*. 1990;86:182–186.
177. Piacente F, Marin M, Molinaro A, et al. Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (viosamine). *J Biol Chem*. 2012;287:3009–3018. https://doi.org/10.1074/jbc.M111.314559.
178. Forterre P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res*. 2006;117:5–16. https://doi.org/10.1016/j.virusres.2006.01.010.
179. Koonin EV, Senkevich TG, Dolja VV. The ancient virus world and evolution of cells. *Biol Direct*. 2006;1:29. https://doi.org/10.1186/1745-6150-1-29.
180. Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brüssow H. Phage as agents of lateral gene transfer. *Curr Opin Microbiol*. 2003;6:417–424. https://doi.org/10.1016/S1369-5274(03)00086-9.
181. Raoult D, Forterre P. Redefining viruses: lessons from Mimivirus. *Nat Rev Microbiol*. 2008;6:315–319. https://doi.org/10.1038/nrmicro1858.
182. Caspar DL, Klug A. Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol*. 1962;27:1–24. https://doi.org/10.1101/SQB.1962.027.001.005.
183. Rossmann MG, Mesyanzhinov VV, Arisaka F, Leiman PG. The bacteriophage T4 DNA injection machine. *Curr Opin Struct Biol*. 2004;14:171–180. https://doi.org/10.1016/j.sbi.2004.02.001.
184. Suzuki Y. Sialobiology of influenza molecular mechanism of host range variation of influenza viruses. *Biol Pharm Bull*. 2005;28:399–408. https://doi.org/10.1248/bpb.28.399.
185. Cady SD, Luo W, Hu F, Hong M. Structure and function of the influenza A M2 proton channel. *Biochemistry*. 2009;48:7356–7364. https://doi.org/10.1021/bi9008837.
186. Nayak DP, Hui EKW, Barman S. Assembly and budding of influenza virus. *Virus Res*. 2004;106:147–165. https://doi.org/10.1016/j.virusres.2004.08.012.
187. Nayak DP, Balogun RA, Yamada H, Zhou ZH, Barman S. Influenza virus morphogenesis and budding. *Virus Res*. 2009;143:147–161. https://doi.org/10.1016/j.virusres.2009.05.010.
188. Hughes FJ, Romanos MA. E1 protein of human papillomavirus is a DNA helicase/ATPase. *Nucleic Acids Res*. 1993;21:5817–5823. https://doi.org/10.1093/nar/21.25.5817.
189. 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. https://doi.org/10.1002/j.1460-2075.1991.tb07967.x.
190. 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 U S A*. 1994;91:12398–12402. https://doi.org/10.1073/pnas.91.26.12398.
191. 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. https://doi.org/10.1126/science.2176744.

192. Cripe TP, Haugen TH, Turk JP, et al. Transcriptional regulation of the human papillomavirus-16 E6-E7 promoter by a keratinocyte-dependent enhancer, and by viral E2 trans-activator and repressor gene products: implications for cervical carcinogenesis. *EMBO J*. 1987;6:3745–3753. https://doi.org/10.1002/j.1460-2075.1987.tb02709.x.
193. 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. http://www.ncbi.nlm.nih.gov/pubmed/2828035. Accessed 11 November 2019.
194. 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. https://doi.org/10.1128/jvi.79.11.6732-6740.2005.
195. 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. https://doi.org/10.1016/j.virol.2005.09.048.
196. 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. https://doi.org/10.1016/j.virol.2006.02.024.
197. 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.
198. 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. http://www.ncbi.nlm.nih.gov/pubmed/8392596. Accessed 11 November 2019.
199. Zhang B, Li P, Wang E, et al. The E5 protein of human papillomavirus type 16 perturbs MHC class II antigen maturation in human foreskin keratinocytes treated with interferon-γ. *Virology*. 2003;310:100–108. https://doi.org/10.1016/S0042-6822(03)00103-X.
200. 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. https://doi.org/10.1128/jvi.77.5.2819-2831.2003.
201. 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. https://doi.org/10.1128/jvi.77.5.2832-2842.2003.
202. Gao L, Aizaki H, He J-W, Lai MMC. 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. https://doi.org/10.1128/jvi.78.7.3480-3488.2004.
203. Kyung MC, Liszewski MK, Nybakken G, et al. West Nile virus nonstructural protein NS1 inhibits complement activation by binding the regulatory protein factor H. *Proc Natl Acad Sci U S A*. 2006;103:19111–19116. https://doi.org/10.1073/pnas.0605668103.
204. 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. https://doi.org/10.1016/j.virol.2008.11.022.
205. Fontana JM, Bankamp B, Rota PA. Inhibition of interferon induction and signaling by paramyxoviruses. *Immunol Rev*. 2008;225:46–67. https://doi.org/10.1111/j.1600-065X.2008.00669.x.

206. 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;69:2563–2573. https://doi.org/10.1099/0022-1317-69-10-2563.
207. Seelamgari A, Maddukuri A, Berro R, et al. Role of viral regulatory and accessory proteins in HIV-1 replication. *Front Biosci*. 2004;9:2388–2413. https://doi.org/10.2741/1403.
208. Le Romancer M, Gaillard M, Geslin C, Prieur D. Viruses in extreme environments. *Life Extrem Environ*. 2006;99–113. https://doi.org/10.1007/978-1-4020-6285-8_6, Springer Netherlands, Dordrecht.
209. Forterre P, Prangishvili D. The origin of viruses. *Res Microbiol*. 2009;160:466–472. https://doi.org/10.1016/j.resmic.2009.07.008.
210. Berezovsky IN. The diversity of physical forces and mechanisms in intermolecular interactions. *Phys Biol*. 2011;8:035002. https://doi.org/10.1088/1478-3975/8/3/035002.
211. Neduva V, Russell RB. Linear motifs: evolutionary interaction switches. *FEBS Lett*. 2005;579:3342–3345. https://doi.org/10.1016/j.febslet.2005.04.005.
212. Pushker R, Mooney C, Davey NE, Jacqué J-M, Shields DC. Marked variability in the extent of protein disorder within and between viral families. *PLoS One*. 2013;8:e60724. https://doi.org/10.1371/journal.pone.0060724.
213. Burra PV, Kalmar L, Tompa P. Reduction in structural disorder and functional complexity in the thermal adaptation of prokaryotes. *PLoS One*. 2010;5:e12069. https://doi.org/10.1371/journal.pone.0012069.
214. Mirsky AE, Ris H. The desoxyribonucleic acid content of animal cells and its evolutionary significance. *J Gen Physiol*. 1951;34:451–462. https://doi.org/10.1085/jgp.34.4.451.
215. Swift H. The constancy of desoxyribose nucleic acid in plant nuclei. *Proc Natl Acad Sci U S A*. 1950;36:643–654. https://doi.org/10.1073/pnas.36.11.643.
216. Swift HH. The desoxyribose nucleic acid content of animal nuclei. *Physiol Zool*. 1950;23:169–198. https://doi.org/10.1086/physzool.23.3.30152074.
217. Parfrey LW, Lahr DJG, Katz LA. The dynamic nature of eukaryotic genomes. *Mol Biol Evol*. 2008;25:787–794. https://doi.org/10.1093/molbev/msn032.
218. Gregory TR. Macroevolution, hierarchy theory, and the C-value enigma. *Paleobiology*. 2004;30:179–202. https://doi.org/10.1666/0094-8373(2004)030<0179:mhtatc>2.0.co;2.
219. Clamp M, Fry B, Kamal M, et al. Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A*. 2007;104:19428–19433. https://doi.org/10.1073/pnas.0709013104.
220. Sickmeier M, Hamilton JA, LeGall T, et al. DisProt: the database of disordered proteins. *Nucleic Acids Res*. 2007;35:D786–D793. https://doi.org/10.1093/nar/gkl893.
221. Siglioccolo A, Paiardini A, Piscitelli M, Pascarella S. Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC Struct Biol*. 2011;11https://doi.org/10.1186/1472-6807-11-50.
222. Bergqvist S, Williams MA, O'Brien R, Ladbury JE. Halophilic adaptation of protein-DNA interactions. *Biochem Soc Trans*. 2003;31:677–680. https://doi.org/10.1042/bst0310677.
223. Elcock AH, McCammon JA. Electrostatic contributions to the stability of halophilic proteins. *J Mol Biol*. 1998;280:731–748. https://doi.org/10.1006/jmbi.1998.1904.
224. Fukuchi S, Yoshimune K, Wakayama M, Moriguchi M, Nishikawa K. Unique amino acid composition of proteins in halophilic bacteria. *J Mol Biol*. 2003;327:347–357. https://doi.org/10.1016/s0022-2836(03)00150-5.
225. Madern D, Ebel C, Zaccai G. Halophilic adaptation of enzymes. *Extremophiles*. 2000;4:91–98. https://doi.org/10.1007/s007920050142.

226. Mevarech M, Frolow F, Gloss LM. Halophilic enzymes: proteins with a grain of salt. *Biophys Chem*. 2000;155–164. https://doi.org/10.1016/S0301-4622(00)00126-5.
227. Paul S, Bag SK, Das S, Harvill ET, Dutta C. Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol*. 2008;9:R70. https://doi.org/10.1186/gb-2008-9-4-r70.
228. Tadeo X, López-Méndez B, Trigueros T, Laín A, Castaño D, Millet O. Structural basis for the aminoacid composition of proteins from halophilic archea. *PLoS Biol*. 2009;7: e1000257. https://doi.org/10.1371/journal.pbio.1000257.
229. Wright DB, Banks DD, Lohman JR, Hilsenbeck JL, Gloss LM. The effect of salts on the activity and stability of Escherichia coli and Haloferax volcanii dihydrofolate reductases. *J Mol Biol*. 2002;323:327–344. https://doi.org/10.1016/s0022-2836(02)00916-6.
230. Zaccai G, Eisenberg H. Halophilic proteins and the influence of solvent on protein stabilization. *Trends Biochem Sci*. 1990;15:333–337. https://doi.org/10.1016/0968-0004(90)90068-m.
231. Mohan A, Sullivan WJ, Radivojac P, Dunker AK, Uversky VN. Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. *Mol Biosyst*. 2008;4:328–340. https://doi.org/10.1039/b719168e.
232. Vucetic S, Brown CJ, Dunker AK, Obradovic Z. Flavors of protein disorder. *Proteins Struct Funct Genet*. 2003;52:573–584. https://doi.org/10.1002/prot.10437.
233. Goh GK-M, Dunker AK, Uversky VN. Understanding viral transmission behavior via protein intrinsic disorder prediction: coronaviruses. *J Pathog*. 2012;2012:1–13. https://doi.org/10.1155/2012/738590.
234. Jiang X, Feyertag F, Robertson DL. Protein structural disorder of the envelope V3 loop contributes to the switch in human immunodeficiency virus type 1 cell tropism. *PLoS One*. 2017;12:e0185790. https://doi.org/10.1371/journal.pone.0185790.
235. Moshe A, Gorovits R. Virus-induced aggregates in infected cells. *Viruses*. 2012;4:2218–2232. https://doi.org/10.3390/v4102218.
236. Gerba CP, Betancourt WQ. Viral aggregation: impact on virus behavior in the environment. *Environ Sci Technol*. 2017;51:7318–7325. https://doi.org/10.1021/acs.est.6b05835.
237. Mészáros B, Erdös G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res*. 2018;46: W329–W337. https://doi.org/10.1093/nar/gky384.
238. Sormanni P, Aprile FA, Vendruscolo M. The CamSol method of rational design of protein mutants with enhanced solubility. *J Mol Biol*. 2015;427:478–490. https://doi.org/10.1016/j.jmb.2014.09.026.
239. Sormanni P, Amery L, Ekizoglou S, Vendruscolo M, Popovic B. Rapid and accurate in silico solubility screening of a monoclonal antibody library. *Sci Rep*. 2017;7:8200. https://doi.org/10.1038/s41598-017-07800-w.
240. Reanney DC. The evolution of RNA viruses. *Annu Rev Microbiol*. 1982;36:47–73. https://doi.org/10.1146/annurev.mi.36.100182.000403.
241. Goh G, Dunker AK, Uversky VN. Protein intrinsic disorder toolbox for comparative analysis of viral proteins. *BMC Genomics*. 2008;9:S4. https://doi.org/10.1186/1471-2164-9-S2-S4.
242. Dyson HJ. Roles of intrinsic disorder in protein-nucleic acid interactions. *Mol Biosyst*. 2012;8:97–104. https://doi.org/10.1039/c1mb05258f.
243. Xie H, Vucetic S, Iakoucheva LM, et al. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res*. 2007;6:1882–1898. https://doi.org/10.1021/pr060392u.
244. Vucetic S, Xie H, Iakoucheva LM, et al. Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding

sequence diversities correlated with long disordered regions. *J Proteome Res*. 2007;6:1899–1916. https://doi.org/10.1021/pr060393m.
245. Xie H, Vucetic S, Iakoucheva LM, et al. Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J Proteome Res*. 2007;6:1917–1932. https://doi.org/10.1021/pr060394e.
246. Goh G, Dunker AK, Uversky VN. Protein intrinsic disorder and influenza virulence: the 1918 H1N1 and H5N1 viruses. *Virol J*. 2009;6:69. https://doi.org/10.1186/1743-422X-6-69.
247. Rizzetto M. Hepatitis D: thirty years after. *J Hepatol*. 2009;50:1043–1050. https://doi.org/10.1016/j.jhep.2009.01.004.
248. Alves C, Cheng H, Roder H, Taylor J. Intrinsic disorder and oligomerization of the hepatitis delta virus antigen. *Virology*. 2010;407:333–340. https://doi.org/10.1016/j.virol.2010.08.019.
249. Greco-Stewart V, Pelchat M. Interaction of host cellular proteins with components of the hepatitis delta virus. *Viruses*. 2010;2:189–212. https://doi.org/10.3390/v2010189.
250. Casaca A, Fardilha M, Silva EDCE, Cunha C. The heterogeneous ribonuclear protein C interacts with the hepatitis delta virus small antigen. *Virol J*. 2011;8:358. https://doi.org/10.1186/1743-422X-8-358.
251. Foster TL, Belyaeva T, Stonehouse NJ, Pearson AR, Harris M. All three domains of the hepatitis C virus nonstructural NS5A protein contribute to RNA binding. *J Virol*. 2010;84:9267–9277. https://doi.org/10.1128/jvi.00616-10.
252. Love RA, Brodsky O, Hickey MJ, Wells PA, Cronin CN. Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. *J Virol*. 2009;83:4395–4403. https://doi.org/10.1128/JVI.02352-08.
253. Hanoulle X, Verdegem D, Badillo A, Wieruszeski JM, Penin F, Lippens G. Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded. *Biochem Biophys Res Commun*. 2009;381:634–638. https://doi.org/10.1016/j.bbrc.2009.02.108.
254. Hanoulle X, Badillo A, Verdegem D, Penin F, Lippens G. The domain 2 of the HCV NS5A protein is intrinsically unstructured. *Protein Pept Lett*. 2010;17:1012–1018. https://doi.org/10.2174/092986610791498920.
255. Macdonald A, Harris M. Hepatitis C virus NS5A: tales of a promiscuous protein. *J Gen Virol*. 2004;85:2485–2502. https://doi.org/10.1099/vir.0.80204-0.
256. Feuerstein S, Solyom Z, Aladag A, et al. Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. *J Mol Biol*. 2012;420:310–323. https://doi.org/10.1016/j.jmb.2012.04.023.
257. Curran J, Kolakofsky D. Replication of paramyxoviruses. *Adv Virus Res*. 1999;54:403–422. https://doi.org/10.1016/s0065-3527(08)60373-5.
258. Bourhis JM, Canard B, Longhi S. Structural disorder within the replicative complex of measles virus: functional implications. *Virology*. 2006;344:94–110. https://doi.org/10.1016/j.virol.2005.09.025.
259. Habchi J, Longhi S. Structural disorder within paramyxovirus nucleoproteins and phosphoproteins. *Mol Biosyst*. 2012;8:69–81. https://doi.org/10.1039/c1mb05204g.
260. Ambros V, Baltimore D. Protein is linked to the 5′ end of poliovirus RNA by a phosphodiester linkage to tyrosine. *J Biol Chem*. 1978;253:5263–5266.
261. Olspert A, Arike L, Peil L, Truve E. Sobemovirus RNA linked to VPg over a threonine residue. *FEBS Lett*. 2011;585:2979–2985. https://doi.org/10.1016/j.febslet.2011.08.009.
262. Olspert A, Peil L, Hébrard E, Fargette D, Truve E. Protein-RNA linkage and post-translational modifications of two sobemovirus VPgs. *J Gen Virol*. 2011;92:445–452. https://doi.org/10.1099/vir.0.026476-0.

263. Sadowy E, Miłner M, Haenni AL. Proteins attached to viral genomes are multifunctional. *Adv Virus Res*. 2001;57:185–262. https://doi.org/10.1016/s0065-3527(01)57004-9.
264. Daughenbaugh KF, Fraser CS, Hershey JWB, Hardy ME. The genome-linked protein VPg of the Norwalk virus binds eIF3, suggesting its role in translation initiation complex recruitment. *EMBO J*. 2003;22:2852–2859. https://doi.org/10.1093/emboj/cdg251.
265. Goodfellow I, Chaudhry Y, Gioldasi I, et al. Calicivirus translation initiation requires an interaction between VPg and eIF4E. *EMBO Rep*. 2005;6:968–972. https://doi.org/10.1038/sj.embor.7400510.
266. Miyoshi H, Suehiro N, Tomoo K, et al. Binding analyses for the interaction between plant virus genome-linked protein (VPg) and plant translational initiation factors. *Biochimie*. 2006;88:329–340. https://doi.org/10.1016/j.biochi.2005.09.002.
267. Michon T, Estevez Y, Walter J, German-Retana S, Gall O. The potyviral virus genome-linked protein VPg forms a ternary complex with the eukaryotic initiation factors eIF4E and eIF4G and reduces eIF4E affinity for a mRNA cap analogue. *FEBS J*. 2006;273:1312–1322. https://doi.org/10.1111/j.1742-4658.2006.05156.x.
268. Tavert-Roudet G, Abdul-Razzak A, Doublet B, et al. The C terminus of lettuce mosaic potyvirus cylindrical inclusion helicase interacts with the viral VPg and with lettuce translation eukaryotic initiation factor 4E. *J Gen Virol*. 2012;93:184–193. https://doi.org/10.1099/vir.0.035881-0.
269. Daughenbaugh KF, Wobus CE, Hardy ME. VPg of murine norovirus binds translation initiation factors in infected cells. *Virol J*. 2006;3. https://doi.org/10.1186/1743-422X-3-33.
270. Khan MA, Miyoshi H, Gallie DR, Goss DJ. Potyvirus genome-linked protein, VPg, directly affects wheat germ in vitro translation: interactions with translation initiation factors eIF4F and eIFiso4F. *J Biol Chem*. 2008;283:1340–1349. https://doi.org/10.1074/jbc.M703356200.
271. Lin L, Shi Y, Luo Z, et al. Protein-protein interactions in two potyviruses using the yeast two-hybrid system. *Virus Res*. 2009;142:36–40. https://doi.org/10.1016/j.virusres.2009.01.006.
272. Hébrard E, Poulicard N, Gérard C, et al. Direct interaction between the rice yellow mottle virus (RYMV) VPg and the central domain of the rice eIF(iso)4G1 factor correlates with rice susceptibility and RYMV virulence. *Mol Plant Microbe Interact*. 2010;23:1506–1513. https://doi.org/10.1094/MPMI-03-10-0073.
273. Satheshkumar PS, Gayathri P, Prasad K, Savithri HS. "Natively unfolded" VPg is essential for Sesbania mosaic virus serine protease activity. *J Biol Chem*. 2005;280:30291–30300. https://doi.org/10.1074/jbc.M504122200.
274. Grzela R, Szolajska E, Ebel C, et al. Virulence factor of potato virus Y, genome-attached terminal protein VPg, is a highly disordered protein. *J Biol Chem*. 2008;283:213–221. https://doi.org/10.1074/jbc.M705666200.
275. Rantalainen KI, Uversky VN, Permi P, Kalkkinen N, Dunker AK, Mäkinen K. Potato virus A genome-linked protein VPg is an intrinsically disordered molten globule-like protein with a hydrophobic core. *Virology*. 2008;377:280–288. https://doi.org/10.1016/j.virol.2008.04.025.
276. Hébrard E, Bessin Y, Michon T, et al. Intrinsic disorder in viral proteins genome-linked: experimental and predictive analyses. *Virol J*. 2009;6:23. https://doi.org/10.1186/1743-422X-6-23.
277. Nair S, Gayathri P, Murthy MRN, Savithri HS. Stacking interactions of W271 and H275 of SeMV serine protease with W43 of natively unfolded VPg confer catalytic activity to protease. *Virology*. 2008;382:83–90. https://doi.org/10.1016/j.virol.2008.08.034.

278. Chowdhury SR, Savithri HS. Interaction of sesbania mosaic virus movement protein with VPg and P10: implication to specificity of genome recognition. *PLoS One*. 2011;6:e15609. https://doi.org/10.1371/journal.pone.0015609.
279. Goh GKM, Dunker AK, Uversky VN. A comparative analysis of viral matrix proteins using disorder predictors. *Virol J*. 2008;5https://doi.org/10.1186/1743-422X-5-126.
280. Bryant M, Ratner L. Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proc Natl Acad Sci U S A*. 1990;87:523–527. https://doi.org/10.1073/pnas.87.2.523.
281. Gottlinger HG, Sodroski JG, Haseltine WA. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A*. 1989;86:5781–5785. https://doi.org/10.1073/pnas.86.15.5781.
282. Gelderblom HR. Assembly and morphology of HIV: potential effect of structure on viral function. *AIDS*. 1991;5:617–637. https://doi.org/10.1097/00002030-199106000-00001.
283. Frankel AD, Young JAT. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem*. 1998;67:1–25. https://doi.org/10.1146/annurev.biochem.67.1.1.
284. Riviere L, Darlix J-L, Cimarelli A. Analysis of the viral elements required in the nuclear import of HIV-1 DNA. *J Virol*. 2010;84:729–739. https://doi.org/10.1128/jvi.01952-09.
285. Schmalzbauer E, Strack B, Dannull J, Guehmann S, Moelling K. Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding in vitro. *J Virol*. 1996;70:771–777. http://www.ncbi.nlm.nih.gov/pubmed/8551614. Accessed 12 November 2019.
286. Poon DTK, Wu J, Aldovini A. Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *J Virol*. 1996;70:6607–6616. https://doi.org/10.1128/JVI.77.11.6284.
287. Carteau S, Batson SC, Poljak L, et al. Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg2+−dependent DNA integration in vitro. *J Virol*. 1997;71:6225–6229. http://www.ncbi.nlm.nih.gov/pubmed/9223522.
288. Darlix JL, Godet J, Ivanyi-Nagy R, Fossé P, Mauffret O, Mély Y. Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *J Mol Biol*. 2011;410:565–581. https://doi.org/10.1016/j.jmb.2011.03.037.
289. Mononegavirales. *Virus Taxon*. 2012;653–657. https://doi.org/10.1016/B978-0-12-384684-6.00053-7.
290. Blocquel D, Bourhis JM, Éléouët JF, et al. Transcription et réplication des Mononegavirales: Une machine moléculaire originale. *Virologie*. 2012;16:225–257. https://doi.org/10.1684/vir.2012.0458.
291. Longhi S. The measles virus N TAIL-XD complex: an illustrative example of fuzziness. *Adv Exp Med Biol*. 2012;725:126–141. https://doi.org/10.1007/978-1-4614-0659-4_8.
292. *Flexible Viruses: Structural Disorder in Viral Proteins-Vladimir Uversky*. Sonia Longhi: John Wiley Sons, Inc., HOBOKEN, NJ, USA; 2011.
293. Laine D, Bourhis JM, Longhi S, et al. Measles virus nucleoprotein induces cell-proliferation arrest and apoptosis through N TAIL-NR and N CORE-FccRIIB1 interactions, respectively. 86:1771–1784. https://doi.org/10.1099/vir.0.80791-0.
294. Laine D, Trescol-Biemont M-C, Longhi S, et al. Measles virus (MV) nucleoprotein binds to a novel cell surface receptor distinct from Fc RII via its C-terminal domain: role in MV-induced immunosuppression. *J Virol*. 2003;77:11332–11346. https://doi.org/10.1128/jvi.77.21.11332-11346.2003.
295. Zhang X, Bourhis J-M, Longhi S, et al. Hsp72 recognizes a P binding motif in the measles virus N protein C-terminus. *Virology*. 2005;337:162–174. https://doi.org/10.1016/j.virol.2005.03.035.

296. Habchi J, Blangy S, Mamelli L, et al. Characterization of the interactions between the nucleoprotein and the phosphoprotein of henipavirus. *J Biol Chem*. 2011;286:13583–13602. https://doi.org/10.1074/jbc.M111.219857.
297. Jensen MR, Communie G, Ribeiro EA, et al. Intrinsic disorder in measles virus nucleocapsids. *Proc Natl Acad Sci U S A*. 2011;108:9839–9844. https://doi.org/10.1073/pnas.1103270108.
298. Communie G, Habchi J, Yabukarski F, et al. Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus. *PLoS Pathog*. 2013;9 e1003631https://doi.org/10.1371/journal.ppat.1003631.
299. Longhi S, Oglesbee M. Structural disorder within the measles virus nucleoprotein and phosphoprotein. *Protein Pept Lett*. 2010;17:961–978. https://doi.org/10.2174/092986610791498894.
300. Iwasaki M, Takeda M, Shirogane Y, Nakatsu Y, Nakamura T, Yanagi Y. The matrix protein of measles virus regulates viral RNA synthesis and assembly by interacting with the nucleocapsid protein. *J Virol*. 2009;83:10374–10383. https://doi.org/10.1128/JVI.01056-09.
301. Colombo M, Bourhis JM, Chamontin C, et al. The interaction between the measles virus nucleoprotein and the interferon regulator factor 3 relies on a specific cellular environment. *Virol J*. 2009;6:59. https://doi.org/10.1186/1743-422X-6-59.
302. Watanabe A, Yoneda M, Ikeda F, Sugai A, Sato H, Kai C. Peroxiredoxin 1 is required for efficient transcription and replication of measles virus. *J Virol*. 2011;85:2247–2253. https://doi.org/10.1128/JVI.01796-10.
303. Hagiwara K, Sato H, Inoue Y, et al. Phosphorylation of measles virus nucleoprotein upregulates the transcriptional activity of minigenomic RNA. *Proteomics*. 2008;8:1871–1879. https://doi.org/10.1002/pmic.200701051.
304. Sato H, Masuda M, Miura R, Yoneda M, Kai C. Morbillivirus nucleoprotein possesses a novel nuclear localization signal and a CRM1-independent nuclear export signal. *Virology*. 2006;352:121–130. https://doi.org/10.1016/j.virol.2006.04.013.
305. De BP, Banerjee AK. Involvement of actin microfilaments in the transcription/replication of human parainfluenza virus type 3: possible role of actin in other viruses. *Microsc Res Tech*. 1999;47:114–123. https://doi.org/10.1002/(SICI)1097-0029(19991015)47:2<114::AID-JEMT4>3.0.CO;2-E.
306. Moyer SA, Baker SC, Horikami SM. Host cell proteins required for measles virus reproduction. *J Gen Virol*. 1990;71:775–783. https://doi.org/10.1099/0022-1317-71-4-775.
307. Houben K, Marion D, Tarbouriech N, Ruigrok RWH, Blanchard L. Interaction of the C-terminal domains of Sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J Virol*. 2007;81:6807–6816. https://doi.org/10.1128/jvi.00338-07.
308. Habchi J, Mamelli L, Darbon H, Longhi S. Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. *PLoS One*. 2010;5:e11684. https://doi.org/10.1371/journal.pone.0011684.
309. Blanchard L, Tarbouriech N, Blackledge M, et al. Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology*. 2004;319:201–211. https://doi.org/10.1016/j.virol.2003.10.029.
310. Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci U S A*. 2005;102:17002–17007. https://doi.org/10.1073/pnas.0506202102.
311. Llorente MT, García-Barreno B, Calero M, et al. Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an α-helical domain involved

in oligomerization. *J Gen Virol*. 2006;87:159–169. https://doi.org/10.1099/vir.0.81430-0.
312. Gerard FCA, Ribeiro EdA, Leyrat C, et al. Modular organization of rabies virus phosphoprotein. *J Mol Biol*. 2009;388:978–996. https://doi.org/10.1016/j.jmb.2009.03.061.
313. Leyrat C, Jensen MR, Ribeiro EA, et al. The N0-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α-helices. *Protein Sci*. 2011;20:542–556. https://doi.org/10.1002/pro.587.
314. Leyrat C, Schneider R, Ribeiro EA, et al. Ensemble structure of the modular and flexible full-length vesicular stomatitis virus phosphoprotein. *J Mol Biol*. 2012;423:182–197. https://doi.org/10.1016/j.jmb.2012.07.003.
315. Chen M, Cortay JC, Gerlier D. Measles virus protein interactions in yeast: new findings and caveats. *Virus Res*. 2003;98:123–129. https://doi.org/10.1016/j.virusres.2003.09.003.
316. Liston P, DiFlumeri C, Briedis DJ. Protein interactions entered into by the measles virus P, V, and C proteins. *Virus Res*. 1995;38:241–259. https://doi.org/10.1016/0168-1702(95)00067-z.
317. Curran J, Marq JB, Kolakofsky D. An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J Virol*. 1995;69:849–855. http://www.ncbi.nlm.nih.gov/pubmed/7815552. Accessed 12 November 2019.
318. Curran J, Pelet T, Kolakofsky D. An acidic activation-like domain of the Sendai virus P protein is required for RNA synthesis and encapsidation. *Virology*. 1994;202:875–884. https://doi.org/10.1006/viro.1994.1409.
319. Tran TL, Castagné N, Bhella D, et al. The nine C-terminal amino acids of the respiratory syncytial virus protein P are necessary and sufficient for binding to ribonucleoprotein complexes in which six ribonucleotides are contacted per N protein promoter. *J Gen Virol*. 2007;88:196–206. https://doi.org/10.1099/vir.0.82282-0.
320. Karlin D, Belshaw R. Detecting remote sequence homology in disordered proteins: discovery of conserved motifs in the N-termini of Mononegavirales phosphoproteins. *PLoS One*. 2012;7:e31719. https://doi.org/10.1371/journal.pone.0031719.
321. Leyrat C, Yabukarski F, Tarbouriech N, et al. Structure of the vesicular stomatitis virus N0-P complex. *PLoS Pathog*. 2011;7:e1002248. https://doi.org/10.1371/journal.ppat.1002248.
322. Cortese MS, Uversky VN, Keith Dunker A. Intrinsic disorder in scaffold proteins: getting more from less. *Prog Biophys Mol Biol*. 2008;98:85–106. https://doi.org/10.1016/j.pbiomolbio.2008.05.007.
323. Balázs A, Csizmok V, Buday L, et al. High levels of structural disorder in scaffold proteins as exemplified by a novel neuronal protein, CASK-interactive protein1. *FEBS J*. 2009;276:4168–4180. https://doi.org/10.1111/j.1742-4658.2009.07090.x.
324. Fontana A, De Laureto PP, De Filippis V, Scaramella E, Zambonin M. Probing the partly folded states of proteins by limited proteolysis. *Fold Des*. 1997;2:R17–R26. https://doi.org/10.1016/S1359-0278(97)00010-2.
325. Fontana A, De Laureto PP, Spolaore B, Frare E, Picotti P, Zambonin M. Probing protein structure by limited proteolysis. *Acta Biochim Pol*. 2004;299–321. https://doi.org/10.1038/s41598-019-42524-z.
326. Flaviviridae—ScienceDirect, (2012). https://www.sciencedirect.com/science/article/pii/B9780123846846000860 (accessed November 12, 2019).
327. Ivanyi-Nagy R, Darlix J-L. Intrinsic disorder in the core proteins of Flaviviruses. *Protein Pept Lett*. 2010;17:1019–1025. https://doi.org/10.2174/092986610791498911.
328. Meng F, Badierah RA, Almehdar HA, Redwan EM, Kurgan L, Uversky VN. Unstructural biology of the dengue virus proteins. *FEBS J*. 2015;282:3368–3394. https://doi.org/10.1111/febs.13349.

329. Shang Z, Song H, Shi Y, Qi J, Gao GF. Crystal structure of the capsid protein from Zika virus. *J Mol Biol*. 2018;430:948–962. https://doi.org/10.1016/j.jmb.2018.02.006.
330. Byk LA, Gamarnik AV. Properties and functions of the dengue virus capsid protein. *Annu Rev Virol*. 2016;3:263–281. https://doi.org/10.1146/annurev-virology-110615-042334.
331. Malhis N, Jacobson M, Gsponer J. MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res*. 2016;44:W488–W493. https://doi.org/10.1093/nar/gkw409.
332. Marvin DA. Filamentous Bacterial viruses. *J Biosci*. 1985;8:799–813. https://doi.org/10.1007/BF02702778.
333. Enterobacteria Phage Fd, (n.d.). https://www.genome.jp/virushostdb/10864 (accessed November 14, 2019).
334. Dunker AK, Ensign LD, Arnold GE, Roberts LM. Proposed molten globule intermediates in fd phage penetration and assembly. *FEBS Lett*. 1991;292:275–278. https://doi.org/10.1016/0014-5793(91)80883-5.
335. Atkins GJ, Sheahan BJ, Liljeström P. The molecular pathogenesis of Semliki forest virus: a model virus made useful? *J Gen Virol*. 1999;80:2287–2297. https://doi.org/10.1099/0022-1317-80-9-2287.
336. Singh I, Helenius A. Role of ribosomes in Semliki forest virus nucleocapsid uncoating. *J Virol*. 1992;66:7049–7058. http://www.ncbi.nlm.nih.gov/pubmed/1433506. Accessed 12 November 2019.
337. Morillas M, Eberl H, Allain FHT, Glockshuber R, Kuennemann E. Novel enzymatic activity derived from the Semliki forest virus capsid protein. *J Mol Biol*. 2008;376:721–735. https://doi.org/10.1016/j.jmb.2007.11.055.
338. Chang C-K, Hsu Y-L, Chang Y-H, et al. Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. *J Virol*. 2009;83:2255–2264. https://doi.org/10.1128/JVI.02001-08.
339. Skehel JJ, Wiley DC. Coiled coils in both intracellular vesicle and viral membrane fusion. *Cell*. 1998;95:871–874. https://doi.org/10.1016/s0092-8674(00)81710-9.
340. Skehel JJ, Wiley DC. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem*. 2000;69:531–569. https://doi.org/10.1146/annurev.biochem.69.1.531.
341. Eckert DM, Kim PS. Mechanisms of viral membrane fusion and its inhibition. *Annu Rev Biochem*. 2001;70:777–810. https://doi.org/10.1146/annurev.biochem.70.1.777.
342. Wiley DC, Skehel JJ. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu Rev Biochem*. 1987;56:365–394. https://doi.org/10.1146/annurev.bi.56.070187.002053.
343. Lommer BS, Luo M. Structural plasticity in influenza virus protein NS2 (NEP). *J Biol Chem*. 2002;277:7108–7117. https://doi.org/10.1074/jbc.M109045200.
344. Sieber T, Dobner T. Adenovirus type 5 early region 1B 156R protein promotes cell transformation independently of repression of p53-stimulated transcription. *J Virol*. 2007;81:95–105. https://doi.org/10.1128/jvi.01608-06.
345. Sieber T, Scholz R, Spoerner M, Schumann F, Kalbitzer HR, Dobner T. Intrinsic disorder in the common N-terminus of human adenovirus 5 E1B-55K and its related E1BN proteins indicated by studies on E1B-93R. *Virology*. 2011;418:133–143. https://doi.org/10.1016/j.virol.2011.07.012.
346. Anderson CW, Schmitt RC, Smart JE, Lewis JB. Early region 1B of adenovirus 2 encodes two coterminal proteins of 495 and 155 amino acid residues. *J Virol*. 1984;50:387–396. http://www.ncbi.nlm.nih.gov/pubmed/6323739. Accessed 12 November 2019.

347. Lewis JB, Anderson CW. Identification of adenovirus type 2 early region 1B proteins that share the same amino terminus as do the 495R and 155R proteins. *J Virol*. 1987;61:3879–3888. http://www.ncbi.nlm.nih.gov/pubmed/2960832. Accessed 12 November 2019.
348. De Chassey B, Navratil V, Tafforeau L, et al. Hepatitis C virus infection protein network. *Mol Syst Biol*. 2008;4:230. https://doi.org/10.1038/msb.2008.66.
349. Tellinghuisen TL, Marcotrigiano J, Rice CM. Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature*. 2005;435:374–379. https://doi.org/10.1038/nature03580.
350. Verdegem D, Badillo A, Wieruszeski JM, et al. Domain 3 of NS5A protein from the hepatitis C virus has intrinsic α-helical propensity and is a substrate of cyclophilin A. *J Biol Chem*. 2011;286:20441–20454. https://doi.org/10.1074/jbc.M110.182436.
351. Taylor JM. Chapter 3 replication of the hepatitis delta virus RNA genome. *Adv Virus Res*. 2009;74:103–121. https://doi.org/10.1016/S0065-3527(09)74003-5.
352. Han Z, Alves C, Gudima S, Taylor J. Intracellular localization of hepatitis delta virus proteins in the presence and absence of viral RNA accumulation. *J Virol*. 2009;83:6457–6463. https://doi.org/10.1128/jvi.00008-09.
353. Ryu WS, Netter HJ, Bayer M, Taylor J. Ribonucleoprotein complexes of hepatitis delta virus. *J Virol*. 1993;67:3281–3287. http://www.ncbi.nlm.nih.gov/pubmed/8497052. Accessed 12 November 2019.
354. Vendel AC, Lumb KJ. Molecular recognition of the human coactivator CBP by the HIV-1 transcriptional activator tat. *Biochemistry*. 2003;42:910–916. https://doi.org/10.1021/bi0270034.
355. 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. https://doi.org/10.1006/jmbi.2001.5067.
356. Surendran R, Herman P, Cheng Z, Daly TJ, Lee JC. HIV rev self-assembly is linked to a molten-globule to compact structural transition. *Biophys Chem*. 2004;108:101–119. https://doi.org/10.1016/j.bpc.2003.10.013.
357. Casu F, Duggan BM, Hennig M. The arginine-rich RNA-binding motif of HIV-1 rev is intrinsically disordered and folds upon RRE binding. *Biophys J*. 2013;105:1004–1017. https://doi.org/10.1016/j.bpj.2013.07.022.
358. 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. https://doi.org/10.1006/viro.1996.0008.
359. Dyson N, Howley P, 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. https://doi.org/10.1126/science.2537532.
360. Brehm A, Nielsen SJ, Miska EA, et al. The E7 oncoprotein associates with Mi2 and histone deacetylase activity to promote cell growth. *EMBO J*. 1999;18:2449–2458. https://doi.org/10.1093/emboj/18.9.2449.
361. Tommasino M, Adamczewski JP, Carlotti F, et al. HPV16 E7 protein associates with the protein kinase p33(CDK2) and cyclin A. *Oncogene*. 1993;8:195–202.
362. 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. https://doi.org/10.1038/sj.onc.1208935.
363. 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. https://doi.org/10.1038/sj.onc.1202142.

364. Brazil DP, Hemmings BA. Ten years of protein kinase B signalling: a hard Akt to follow. *Trends Biochem Sci*. 2001;26:657–664. https://doi.org/10.1016/s0968-0004(01)01958-2.
365. Alonso LG, García-Alai MM, Nadra AD, et al. High-risk (HPV16) human papillomavirus E7 oncoprotein is highly stable and extended, with conformational transitions that could explain its multiple cellular binding partners. *Biochemistry*. 2002;41:10510–10518. https://doi.org/10.1021/bi025579n.
366. García-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. https://doi.org/10.1021/bi7007917.
367. Chemes LB, Sánchez IE, Smal C, De Prat-Gay G. Targeting mechanism of the retinoblastoma tumor suppressor by a prototypical viral oncoprotein: structural modularity, intrinsic disorder and phosphorylation of human papillomavirus E7. *FEBS J*. 2010;277:973–988. https://doi.org/10.1111/j.1742-4658.2009.07540.x.
368. 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. https://doi.org/10.1021/pr0602388.
369. Roberts JW. Termination factor for RNA synthesis. *Nature*. 1969;224:1168–1174. https://doi.org/10.1038/2241168a0.
370. Greenblatt J, Li J. Properties of the N gene transcription antitermination protein of bacteriophage ?? *J Biol Chem*. 1982;257:362–365.
371. Van Gilst MR, Von Hippel PH. Assembly of the N-dependent antitermination complex of phage λ: NusA and RNA bind independently to different unfolded domains of the N protein. *J Mol Biol*. 1997;274:160–173. https://doi.org/10.1006/jmbi.1997.1389.
372. Van Gilst MR, Rees WA, Das A, von Hippel PH. Complexes of N Antitermination protein of phage λ with specific and nonspecific RNA target sites on the nascent transcript †. *Biochemistry*. 1997;36:1514–1524. https://doi.org/10.1021/bi961920q.
373. Johansen D, Trewhella J, Goldenberg DP. Fractal dimension of an intrinsically disordered protein: small-angle X-ray scattering and computational study of the bacteriophage λ N protein. *Protein Sci*. 2011;20:1955–1970. https://doi.org/10.1002/pro.739.
374. Mogridge J, Legault P, Li J, Van Oene MD, Kay LE, Greenblatt J. Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage λ N protein. *Mol Cell*. 1998;1:265–275. https://doi.org/10.1016/S1097-2765(00)80027-1.
375. Morozov SY, Solovyev AG. Triple gene block: modular design of a multifunctional machine for plant virus movement. *J Gen Virol*. 2003;84:1351–1366. https://doi.org/10.1099/vir.0.18922-0.