the FEBS Journal

Unstructural biology of the dengue virus proteins

Fanchi Meng[1], Reaid A. Badierah[2], Hussein A. Almehdar[2], Elrashdy M. Redwan[2,3], Lukasz Kurgan[1] and Vladimir N. Uversky[2,4,5]

1 Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada
2 Biological Department, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia
3 Therapeutic and Protective Proteins Laboratory, Protein Research Department, Genetic Engineering and Biotechnology Research Institute, City for Scientific Research and Technology Applications, New Borg El-Arab, Alexandria, Egypt
4 Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA
5 Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, St Petersburg, Russia

Keywords
dengue virus; intrinsically disordered protein; protein function; protein structure; proteome

Correspondence
E. M. Redwan, Biological Sciences Department, Faculty of Science, King Abdulaziz University, PO Box 80203, Jeddah 21589, Saudi Arabia
Fax: +96 602 640 0736
Tel: +96 656 702 8262
E-mail: redwan1961@yahoo.com
L. Kurgan, Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta T6G 2V4, Canada
Fax: +1 780 492 1811
Tel: +1 780 492 5488
E-mail: lkurgan@ece.ualberta.ca
V. N. Uversky, Department of Molecular Medicine, University of South Florida, 12901 Bruce B. Downs Boulevard MDC07, Tampa, FL 33612, USA
Fax: +1 813 974 7357
Tel: +1 813 974 5816
E-mail: vuversky@health.usf.edu

(Received 17 January 2015, revised 1 June 2015, accepted 15 June 2015)

doi:10.1111/febs.13349

In this study, we used a wide spectrum of bioinformatics techniques to evaluate the extent of intrinsic disorder in the complete proteomes of genotypes of four human dengue virus (DENV), to analyze the peculiarities of disorder distribution within individual DENV proteins, and to establish potential roles for the structural disorder with respect to their functions. We show that several proteins (ER, E, 1, 2A and 4A) are predicted to be mostly ordered, whereas four proteins (C, 2k, NS3 and NS5) are expected to have high disorder levels. The profiles of disorder propensities are similar across the four genotypes, except for the NS5 protein. Cleavage sites are depleted in polymorphic sites, and have a high propensity for disorder, especially relative to neighboring residues. Disordered regions are highly polymorphic in type 1 DENV but have a relatively low number of polymorphic sites in the type 4 virus. There is a high density of polymorphisms in proteins 2A and 4A, which are depleted in disorder. Thus, a high density of polymorphism is not unique to disordered regions. Analysis of disorder/function association showed that the predominant function of the disordered regions in the DENV proteins is protein–protein interaction and binding of nucleic acids, metals and other small molecules. These regions are also associated with phosphorylation, which may regulate their function.

Abbreviations
CLV, cleavage site; DENV, dengue virus; ELM, eukaryotic linear motif; IDP, intrinsically disordered protein; IDPR, intrinsically disordered protein region; MoRF, molecular recognition feature; NS, non-structural protein.

Introduction

Dengue fever virus (DENV) is a member of the family *Flaviviridae* in the genus *Flavivirus*, that, among other members, includes hepatitis C virus, West Nile virus and yellow fever virus [1]. The *Flavivirus* genus consists of nearly 80 viruses, many of which are arthropod-borne human pathogens that cause a variety of diseases, including dengue fever, plus the associated dengue hemorrhagic fever and dengue shock syndrome, Japanese encephalitis and yellow fever [2]. There are four antigenically related serotypes of the dengue virus (DENV-1, DENV-2, DENV-3 and DENV-4). All four serotypes are known to cause the full spectrum of disease [3]. Infection with one of these serotypes provides immunity for life, but to only that serotype [4]. Therefore, persons living in a dengue endemic area are at risk of encountering secondary infection with other DENV serotypes.

DENV is an arbovirus (arthropod-borne virus) that is primarily transmitted between humans and *Aedes aegyptim* which breed in domestic and peridomestic water containers. A sylvatic cycle (whereby jungle primates and mosquito vectors perpetuate the virus) has been documented in Southeast Asia and West Africa, but it is presently uncertain to what extent this cycle contributes to human infections [5]. It is believed that this virus displays enzootic maintenance cycles that involve *Aedes* mosquitoes, which breed in tree holes and transmit the virus between monkeys, from monkeys to humans, and between humans. Zoonotic cycles of dengue virus transmission involving monkeys and forest *Aedes* species have been documented in Malaysia [6], Sri Lanka [7] and West Africa [5]. The vector in Malaysia is *Aedes niveus*; the species implicated in West Africa are *Aedes furcifer*, *Aedes taylori*, *Aedes luteocephalus*, *Aedes opok* and *Aedes africanus*. Therefore, although *Aedes aegypti* is considered as the most important vector, other *Aedes* species also may play a role in transmission of the infection [8]. Curiously, another member of the *Flavivirus* genus, *Aedes* flavivirus, has been shown to be transmitted vertically from one mosquito to another [9].

Dengue occurs in epidemic and endemic proportions throughout tropical and sub-tropical regions of the world [10]. For example, dengue epidemics involving thousands of people and multiple strains recur in areas of tropical Asia, Africa, Australia and the Americas, where the *Aedes aegypti* mosquito is present. Since the 1980s, dengue fever and a severe form of the disease described for the first time in 1954, dengue hemorrhagic fever, have emerged as the most important arthropod-borne viral diseases of humans, as far as morbidity, mortality and economic cost are concerned [11,12]. More than 2.5 billion people inhabit tropical areas, where they are at risk of dengue infection, and an estimated 100 million cases of dengue fever [13] and 250 000 cases of life-threatening dengue hemorrhagic fever are reported annually on a worldwide basis [14].

The clinical manifestations of dengue virus infection are varied, ranging from mild sub-clinical symptoms to hemorrhagic fever and/or to dengue shock syndrome. The disease is graded according to severity as follows: non-specific febrile illness, classic dengue fever, dengue hemorrhagic fever (grades I and II) and dengue shock syndrome (grades III and IV) [15]. Classic dengue fever is characterized by sudden onset of fever in combination with various symptoms such as headache, retro-orbital pain, myalgia, arthralgia, rash, hemorrhagic manifestations and leucopenia [16].

In the Arabian peninsula, a pandemic of dengue-like disease appeared in the late 19th century (1870–1873), affecting populations in Yemen and Saudi Arabia [17,18]. It is believed that this outbreak started and ended in East Africa, appearing first in Tanzania in 1870, then spreading to Egypt, Saudi Arabia, Yemen, India, China, Indonesia, Vietnam, Laos and Cambodia, and finally ending in Mauritius in 1873 [19]. In 1994, dengue virus was isolated in Jeddah, Saudi Arabia, for the first time since the outbreaks of 19th century. The isolates were confirmed to be DENV-2 [20]. Between 1994 and 2006, there were four dengue outbreaks in Saudi Arabia. A new outbreak in Jeddah was reported in 2009 [18,21]. As three serotypes of the dengue virus have been found to be circulating in Saudi Arabia, it is likely that dengue has become endemic in this country [18,21], particularly in Jeddah province [22]. In view of the recent continuous epidemic in Jeddah province, and because dengue infection is currently considered as one of the major health problems in Jeddah, systematic and effective sero-epidemiological studies are required to determine the prevalence and incidence of dengue infection among the population within Jeddah province, and to monitor the effectiveness of the ongoing effort to control the epidemic. The requirement for such studies is supported by the current lack of the baseline prevalence data for dengue infection among various populations in Jeddah.

DENV is an enveloped positive-sense single-stranded RNA virus. The virion possesses an icosahedral envelope organization and a spherical nucleocapsid core, and has a diameter of 30 nm [23,24]. The DENV genome is approximately 11 kb long, with a 5′ cap

(m7G5′ppp5′A) at the 5′ end, but lacks a polyadenylate tail [25]. This genomic RNA includes an mRNA containing a single open reading frame for translation into a large polyprotein, that is cleaved into three structural proteins (capsid protein C, membrane protein M and envelope protein E), seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) [3], and two peptides (ER and 2k). Surrounding the open reading frame are 5′ and 3′ non-coding regions that contain conserved sequences. The existing RNA structures of non-coding regions may serve as *cis*-acting elements directing the processes of genome amplification, translation or packaging. Although it is not yet feasible to directly study the structures of flavivirus RNAs *in vivo*, the ability of RNA to adapt alternatively folded structures may regulate these competing processes [26].

Dengue viruses need to be able to survive in two different hosts, the vector and the reservoir, where they must replicate their genes while avoiding the hosts' counter-measures [27]. This dual host requirement is thought to limit the degree of evolution, because the virus must conserve amino acids that are critical for survival in the vector and the reservoir. Furthermore, many viral proteins are multi-functional because of the extremely small viral proteomes, which are typically limited to a bare minimum of structural proteins required for viral particle assembly and a set of non-structural proteins that are used to hijack many functional pathways of the host cell. This multi-functionality has been attributed to the structural uniqueness of the viral proteins, which often contain functional intrinsically disordered regions [28,29].

In fact, many features expected for the viral proteins have their roots in protein intrinsic disorder, in line with the numerous recent studies that clearly indicated the wide spread of intrinsic disorder, whereby many biologically active proteins do not have unique 3D structures as a whole or in part [30–38]. These intrinsically disordered proteins (IDPs) and hybrid proteins containing ordered domains and intrinsically disordered protein regions (IDPRs) exist as dynamic conformational ensembles [30,32,36,39–42], with varying levels of residual structure, ranging from collapsed (molten globule-like), to partially collapsed (pre-molten globule-like), and even highly extended (coil-like) conformations [33,35,42,43]. IDPs/IDPRs are typically involved in regulation, signaling and control pathways [32,35,37,42,44,45], complementing the traditional functions of ordered proteins [40,46–50]. These proteins are often involved in the pathogenesis of various human diseases [51,52]. IDPs/IDPRs are very common in all proteomes analyzed so far, with viruses possessing the largest variation in the content of disordered residues in their proteomes [53,54]. The high levels of intrinsic disorder allow many viral proteins to fulfill their biological roles, as multi-functionality is critically dependent on binding promiscuity and the ability to be involved in numerous interactions with host membranes, host nucleic acids and host proteins. It is also likely that flexible structures may help viral proteins to evade the host immune system [55–58]. Finally, the lack of structural constraints in IDPs/IDPRs may represent a means to support the high mutation rates that are characteristic of viruses [28]. In agreement with these considerations, our recent comprehensive *in silico* analysis revealed that intrinsic disorder is abundant in the completed proteomes of several human hepatitis C virus genotypes [59], as well as in the proteomes of the HIV-1 [60], and various serotypes of human papillomaviruses [61,62], that intrinsic disorder is peculiarly distributed within the individual viral proteins, and that there is an intricate connection between the structural disorder and the functions of the viral proteins [59–62].

In this study, to shed additional light on dengue virus biology, we used a wide spectrum of bioinformatics techniques to evaluate the abundance and functional roles of intrinsic disorder in DENV proteins. This analysis was performed in line with previous studies on the computational characterization of viral intrinsically disordered proteins [59–62].

Results

Translation of the DENV genome produces a single polypeptide, a DENV polyprotein consisting of approximately 3390 residues that requires further proteolytic processing to generate active viral proteins, which are classified as either structural or non-structural. All the structural proteins are located within the N–terminal part of the polyprotein. Among these structural proteins are the highly basic capsid protein C, the membrane precursor protein prM, and the envelope protein E. The remainder of the genome encodes a set of non-structural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5), which play various roles in controlling, coordinating and regulating the various intracellular processes of the virus life cycle.

Overall disorder-based analysis of dengue polyproteins

Structural and functional annotations of the dengue polyproteins, including polymorphisms, trans- and intra-membrane regions, topological domains, func-

tional sites, eukaryotic linear motifs (ELMs), IDPRs (using all binary predictions from MFDp and predictions with a low false-positive rate, MFDp_H) and molecular recognition features (MoRFs), are summarized in Fig. 1. Figure 1A shows annotations over the 20 polyproteins from all dengue virus serotypes.

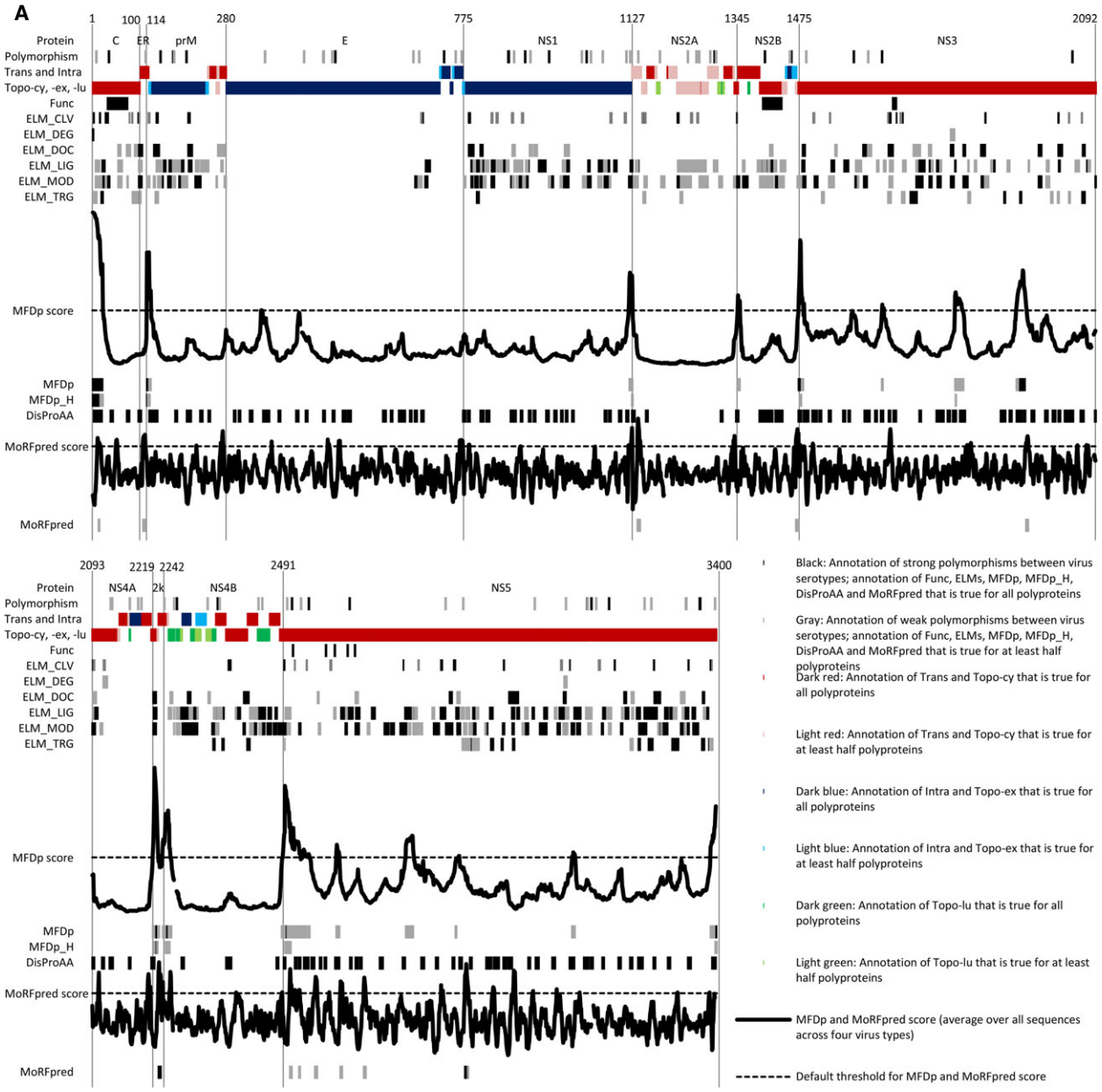

Fig. 1. Structural and functional annotations for the dengue virus polyproteins. Annotations are summarized over all 20 polyproteins in all virus serotypes (A), and over polyproteins for each of the four serotypes of this virus (B). From top to bottom: annotations of polymorphisms, transmembrane and intra-membrane regions ('Trans' and 'Intra'), three types of topological domains (Topo-cy, -ex and -lu), functional sites (Func), six types of ELMs (ELM-CLV, -DEG, -DOC, -LIG, -MOD and -TRG), intrinsic disorder predicted by MFDp, including propensity scores and binary predictions using the default cut-off at 0.37, and by MFDp_H, based on a cut-off of 0.57, which corresponds to a low false-positive rate (0.05), disorder-promoting amino acids (DisProAA), and MoRFs predicted using MoRFpred, including propensity scores and binary predictions (using a default score threshold of 0.5). Dark colors (black, dark red, dark blue and dark green) indicate strong polymorphisms and annotations that are true for all polyproteins over the four virus serotypes or all polyproteins within a given virus serotype; light colors (gray, light red, light blue and light green) indicate weak polymorphisms and annotations that are true for at least half but not all polyproteins across the four virus serotypes or within a given virus serotype. Gray vertical lines indicate cleavage sites.

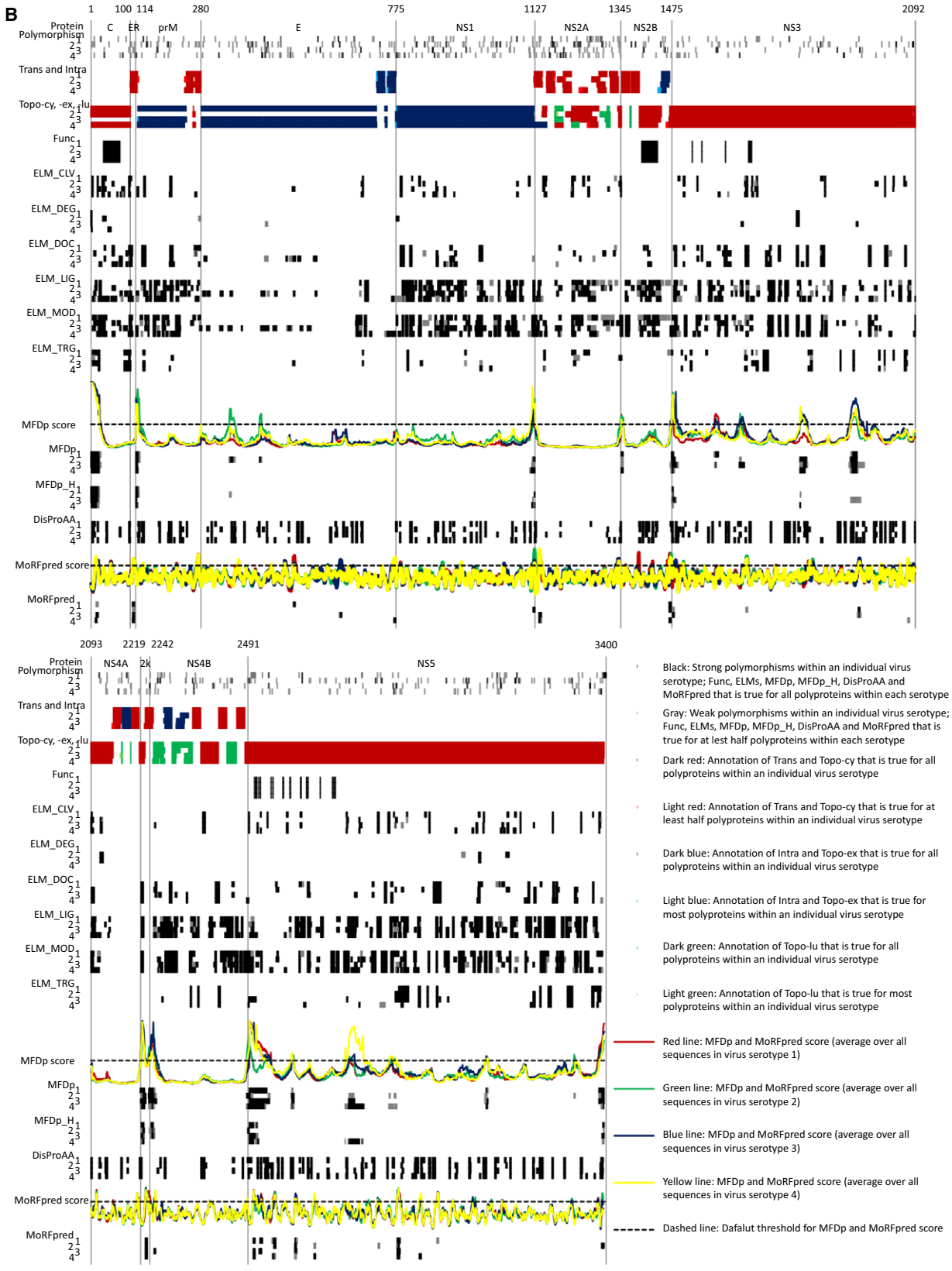

Fig. 1. (Continued)

Figure 1B focuses on individual serotypes. The results are color-coded, whereby dark colors (black, dark red, dark blue and dark green) indicate strong polymorphisms and annotations that are true for all polyproteins in the considered dengue viruses or in a given serotype, whereas light colors (gray, light red, light blue, and light green) indicate weak polymorphisms or annotations that are true for at least half but not all of the polyproteins in all dengue viruses or in a given serotype.

Analysis of the data summarized in Fig. 1 shows that several proteins are predicted to be highly structured (e.g. proteins ER, E, NS1, NS2A and NS4A), whereas four proteins are expected to have relatively high amounts of disorder (proteins C, 2k, NS3 and NS5). Also, proteolytic cleavage sites have a high propensity for disorder, especially relative to neighboring residues. This is a very important observation that clearly indicates the role of intrinsic disorder in processing of the viral polyprotein and the maturation of individual viral proteins. This is in agreement with the important notion that proteolytic digestion is orders of magnitude faster in unstructured protein regions compared with structured protein regions [63–68]. Therefore, it is extremely important for the protein cleavage process that sites of cleavage are located in regions that lack structure or possess high structural flexibility.

Figure 1B also shows that the profiles of disorder propensities are similar across the four serotypes of the virus, except for the parts of the DENV polyprotein corresponding to the viral NS5 protein. Curiously, our analysis revealed that sites that have been annotated functionally are depleted in disorder. On the other hand, MoRF-containing regions are enriched in disorder. Such enrichment was expected, as MoRFs were originally defined as disorder-based potential binding sites. The fact that annotated functional sites (i.e. active sites, binding sites, other functional sites except for cleavage sites, and nucleotide binding regions) are depleted in disorder indicates that functional sites in disordered regions currently lack annotations. Although it is believed that the presence of intrinsic disorder is typically associated with fast evolutionary rates [42,69–73], our analysis revealed that there is a high density of polymorphisms in proteins NS2A and NS4A, which are among the more ordered DENV proteins.

Next, we numerically quantified the patterns of enrichment and depletion of the disordered and MoRF-containing regions, and their relationship with functional and structural annotations.

Content of IDPRs and MoRFs in proteins from the dengue virus

Figure 2 summarizes the content of disordered and MoRF residues from IDPRs and MoRF regions, respectively, for each of the 12 proteins expressed by the dengue virus. The content is defined as the fraction of disordered or MoRF residues in a given protein (the number of disordered or MoRF residues divided by the total number of residues in a given protein). The content values were calculated for the same protein over the 20 genomes and separately for each virus serotype. This analysis provides further support for the observations made based on study of the DENV polyproteins, namely that DENV proteins C, 2k, NS3 and NS5 have a relatively high disorder content, whereas proteins ER, E, NS1, NS2A and NS4A are relatively more structured. As ER and 2k are enriched in disorder-based protein binding sites, it is likely that ER may also have some disorder in the C-terminus. Overall, the proteins in the dengue virus have a relatively low amount of disorder and MoRF regions. These results are consistent across all the DENV serotypes.

Content	Virus types	Proteins and peptides											
		C	ER	prM	E	NS1	NS2A	NS2B	NS3	NS4A	2k	NS4B	NS5
content of IDPRs	All types	0.23	0.00	0.05	0.02	0.02	0.00	0.06	0.09	0.03	0.67	0.06	0.10
	type 1	0.25	0.00	0.07	0.00	0.00	0.00	0.06	0.06	0.04	0.61	0.04	0.08
	type 2	0.22	0.00	0.07	0.04	0.02	0.01	0.06	0.08	0.01	0.75	0.08	0.07
	type 3	0.25	0.00	0.03	0.00	0.02	0.00	0.07	0.14	0.03	0.71	0.06	0.11
	type 4	0.21	0.00	0.04	0.00	0.04	0.00	0.07	0.07	0.03	0.53	0.06	0.14
content of MoRFs	All types	0.04	0.42	0.00	0.01	0.01	0.02	0.04	0.02	0.00	0.27	0.01	0.04
	type 1	0.04	0.43	0.00	0.01	0.00	0.00	0.05	0.00	0.00	0.26	0.00	0.05
	type 2	0.02	0.43	0.00	0.00	0.01	0.00	0.05	0.02	0.00	0.27	0.00	0.04
	type 3	0.06	0.36	0.00	0.01	0.01	0.03	0.00	0.02	0.00	0.26	0.02	0.03
	type 4	0.05	0.46	0.00	0.01	0.00	0.03	0.05	0.01	0.00	0.26	0.00	0.04

Fig. 2. Content of disordered and MoRF residues for each of the 12 proteins and peptides expressed by the dengue virus. The content values (the number of disordered or MoRF residues divided by the number of all residues in a given protein) were computed over all serotypes and for each of the four virus serotypes. The values are color-coded: a dark red background indicates a high content (> 0.1), light red indicates moderate content (between 0.05 and 0.1, inclusive), and blue indicates a low content (< 0.05).

Enrichment in intrinsic disorder, MoRF regions, and polymorphisms of the functional regions in the dengue virus

We analyzed the amount of polymorphic sites, disordered and MoRF residues in the functional sites in the dengue virus. The results of this analysis are shown in Fig. 3. We categorized these quantities as enriched, depleted or neutral based on the ratio of the rates of occurrence of polymorphic, disordered or MoRF residues in a given type of functional region versus the entire polyprotein. First, we calculate the number of polymorphic, disordered or MoRF residues in a given type of functional region, and divide it by the total number of the residues in this region type. Then, we calculate the number of polymorphic, disordered or MoRF residues in the entire polyprotein, and divide it by the total number of residues in these polyproteins. Finally, we divide the first result by the second. We performed these calculations over the 20 genomes (columns labeled 'all' in Fig. 3) and separately over the genomes from each of the four virus serotype (columns numbered 1, 2, 3 and 4 in Fig. 3). Ratios above 1 and below 1 indicate enrichment and depletion, respectively. For example, IDPRs in cleavage sites (CLV regions) show a ratio of 6.2, which means that they occur 6.2 times more often in these functional sites compared to the overall genome. This analysis revealed that cleavage sites are substantially enriched in disorder and depleted in polymorphic sites. Transmembrane regions are enriched in MoRFs and polymorphic sites, whereas intra-membrane regions are enriched in polymorphic sites. Figure 3 also shows that the topological luminal domains are enriched in polymorphisms, whereas cytoplasmic topological domains are enriched in disorder and MoRF sites. For individual serotypes, the functional sites lack enrichment of polymorphic sites, and are characterized by depletion in disorder and MoRF residues. Targeting ELM sites are enriched in MoRFs, and several types of ELMs (cleavage sites, degrons, sites of post-translational modification, and targeting sites) are enriched in disorder. MoRF regions are enriched in polymorphic sites. Curiously, disordered regions are highly polymorphic in DENV-1 but have a low number of polymorphic sites in DENV-2 and DENV-4. MoRF residues are enriched in disorder, and disordered residues are enriched in MoRFs, as expected.

Amino acid composition

We separated the amino acid types into order-promoting residues (W, F, Y, I, M, L, V, N, C and T)

Function	Enrichment in strong polymorphism					Enrichment in strong and weak polymorphism					Enrichment in IDPRs					Enrichment in MoRFs				
	all	1	2	3	4	**all**	1	2	3	4	**all**	1	2	3	4	**all**	1	2	3	4
CLV	**0.0**	0.0	3.0	0.0	0.0	**0.0**	0.0	0.9	0.0	0.0	**6.2**	6.9	7.0	5.1	6.2	**1.7**	0.0	3.1	2.6	0.0
Trans	**0.6**	0.7	1.1	1.0	2.2	**1.1**	0.9	1.1	1.5	1.3	**1.1**	1.2	1.4	0.9	0.8	**2.0**	1.7	2.1	2.5	1.6
Intra	**2.1**	1.1	0.6	0.9	0.6	**2.2**	1.6	1.3	1.3	1.2	**0.0**	0.0	0.0	0.0	0.0	**0.8**	0.9	0.8	0.0	1.7
Topo-cy	**0.9**	1.2	0.9	0.9	0.7	**0.8**	0.9	0.9	0.8	0.8	**1.4**	1.5	1.3	1.5	1.5	**1.2**	1.3	1.3	1.1	1.3
Topo-ex	**1.2**	0.9	0.9	1.2	1.3	**1.1**	1.1	1.0	1.2	1.2	**0.3**	0.1	0.5	0.1	0.2	**0.3**	0.4	0.1	0.4	0.3
Topo-lu	**1.4**	0.0	2.5	0.9	0.0	**2.4**	1.8	1.7	1.0	1.0	**0.7**	0.5	0.8	0.9	0.4	**0.5**	0.0	0.0	1.8	0.0
Func	**2.2**	0.0	0.0	1.0	1.1	**0.7**	0.6	0.5	0.7	0.5	**0.5**	1.0	0.0	0.7	0.7	**0.6**	1.2	0.1	0.8	0.7
ELM_CLV	**1.0**	0.0	1.0	0.0	1.2	**1.0**	0.8	1.1	1.3	1.3	**1.5**	1.3	1.5	1.6	1.5	**1.0**	1.1	1.7	1.1	0.2
ELM_DEG	**1.1**	0.0	0.2	1.6	1.3	**1.2**	0.0	1.8	1.1	1.2	**2.1**	5.0	0.7	4.2	2.3	**0.3**	0.0	0.0	1.2	0.0
ELM_DOC	**0.6**	0.9	0.7	0.3	1.3	**0.8**	0.7	1.2	1.3	0.8	**0.9**	0.7	1.0	0.8	1.1	**0.4**	0.7	0.8	0.4	0.0
ELM_LIG	**1.2**	1.2	1.3	0.9	1.0	**1.3**	1.3	1.1	1.3	0.9	**1.0**	0.6	1.0	0.9	1.2	**1.0**	0.8	1.0	1.0	1.2
ELM_MOD	**1.1**	1.2	1.3	1.1	0.9	**1.0**	1.2	1.0	1.2	1.2	**1.3**	1.5	1.1	1.2	1.5	**0.7**	0.8	0.5	0.7	1.0
ELM_TRG	**0.4**	1.0	1.3	0.9	0.3	**0.8**	1.3	1.5	1.5	0.6	**1.6**	2.4	1.6	2.1	0.8	**2.2**	2.7	2.4	2.9	0.7
IDPRs	**1.1**	6.9	0.6	1.4	1.0	**1.5**	2.0	1.3	0.9	0.9					0.0	**4.7**	5.5	4.5	6.4	2.7
MoRFs	**1.2**	3.3	0.0	1.5	0.0	**1.4**	2.2	1.3	1.8	1.4	**4.7**	5.5	4.5	6.4	2.7					

Fig. 3. Number of polymorphic sites, intrinsically disordered residues and MoRF residues in functionally annotated regions. The functional sites include cleavage sites (CLV), transmembrane regions ('Trans'), the intra-membrane region ('Intra'), topological cytoplasmic, extracellular and luminal domains (Topo-cy, Topo-ex and Topo-lu), functional sites (Func), six types of ELMs (ELM_CLV, DEG, DOC, LIG, MOD and TRG), IDPRs and MoRFs. The amount is quantified as the ratios of the rate of occurrence of polymorphic, disordered and MoRF residues between a given type of functional region and the overall rate in the entire polyprotein. Ratios above 1 and below 1 correspond to the enrichment and depletion, respectively. A dark red background indicates high enrichment (ratio ≥ 1.2), light red background indicates neutral (slight enrichment) (ratio between 1 and 1.2, inclusive), light blue indicates neutral (slight depletion) (ratio between 0.8 and 1, exclusive), and dark blue indicates depletion (ratio ≤ 0.8).

and disorder-promoting residues (A, G, R, D, H, Q, K, S, E and P) [74]. We have annotated the disorder-promoting amino acids (DisProAA) over all virus serotypes and within each serotype in Fig. 1A,B, respectively. Black vertical bars show regions of at least four consecutive disorder-promoting residues (the minimum size of disordered regions) for all polyproteins, and gray bars indicate regions that are found in at least half but not all polyproteins. Proteins ER and NS2A have few disorder-promoting regions, which supports our observation that these two proteins are highly structured. Also, the regions that are predicted as disordered are rich in disorder-promoting residues. We also quantified relative differences between the amino acid compositions of the 12 proteins that constitute the dengue polyprotein and the composition of a large set of structured proteins from the Protein Data Bank [75]. The structured proteins are a subset of all proteins in the Protein Data Bank that share < 25% sequence identity. The relative differences were computed using Composition Profiler [76]. We found that the two proteins that are enriched in disorder, C and 2k, are enriched in some disorder-promoting residues: R and K for protein C, and A and Q for 2k. Structured proteins in the dengue virus are depleted in disorder-promoting residues. The enrichment of the protein C in the positively charged residue R results in the overall high positive charge of this protein, which is needed for RNA binding.

Functional analysis of intrinsically disordered regions

We found 404 IDPRs in the 240 proteins from the 20 dengue polyproteins. We annotated their putative functions based on the protocol described previously [77,78]. Briefly, the function was predicted based on high similarity calculated based on local pairwise alignment against functionally annotated disordered regions obtained from the DisProt database [79]. We note that the same disordered region may be annotated with multiple functions, in agreement with studies that show that the same disordered region may bind multiple partners and perform multiple functions [30,37,80–83]. In total, we annotated 18 functions for 44 IDPRs. After eliminating functions that were predicted for fewer than three IDPRs, which have a higher propensity of being spurious, we obtained 12 functions for 44 IDPRs. Figure 4 summarizes these functions. It shows the fractions of putative IDPRs with a given function and the fraction of virus serotypes with a given function. Our analysis revealed that the predominant function of the DENV IDPRs is protein–protein binding. Disordered regions are also involved in various protein–ligand binding events, such as interactions of viral proteins with nucleic acids, metals and other small molecules. Viral IDPRs are associated with phosphorylation. The fact that disordered regions of the DENV proteins contain numerous phosphorylation sites is in agreement with observations

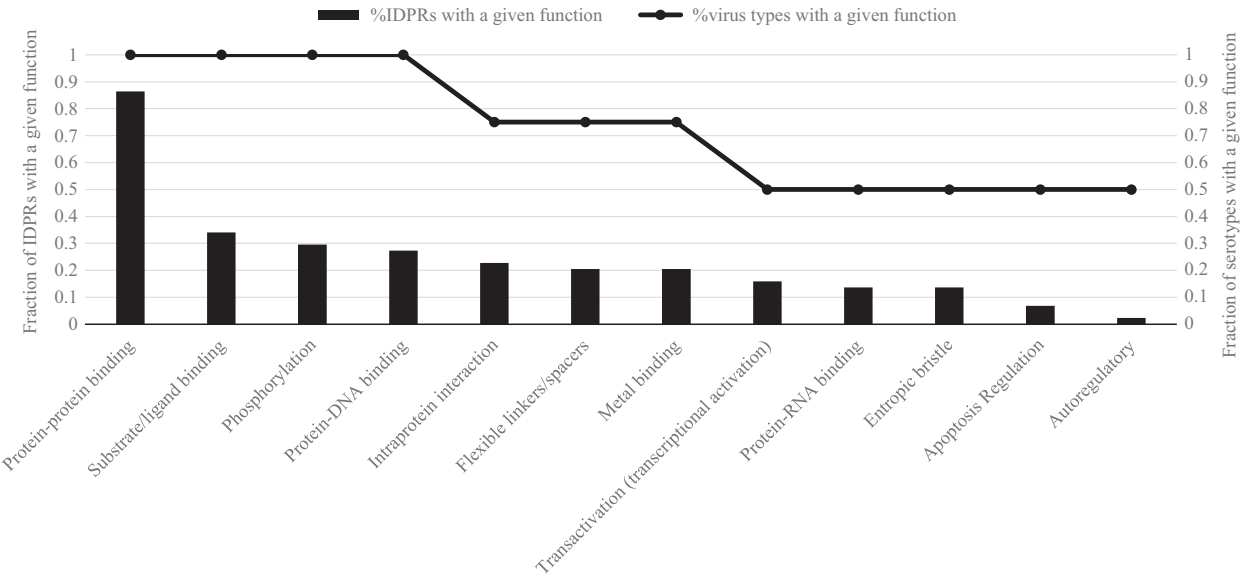

Fig. 4. Functional annotations of putative IDPRs among 20 polyproteins. Values represent indicate proportions of putative IDPRs/virus types having functional rules shown on the *x* axis.

that phosphorylation [84] and many other enzymatically catalyzed post-translational modifications are preferentially located within IDPRs [85].

Discussion

The primary site of replication of the DENV genome is the cytoplasm of infected host cells. Here, the viral RNA is translated into a polyprotein, which is then directed to the endoplasmic reticulum. The DENV proteome includes three structural proteins (C, prM/M and E), which are responsible for formation of the components of the virion, and seven non-structural proteins (NS1, NS2A/B, NS3, NS4A/B and NS5), which play various roles in viral RNA replication. Signal sequences within the polyprotein translocate NS1 and the ectodomains of prM and E into the lumen of the endoplasmic reticulum, whereas the C, NS3 and NS5 proteins are localized to the cytoplasm [86]. Figure 5 shows an over-simplified model of the DENV polyprotein embedded into the membrane of the endoplasmic reticulum, and schematically shows the cleavage sites of various proteases involved in the maturation process [86]. It is interesting to compare the computationally predicted disorder propensities and predicted disorder-based functions of the dengue virus proteins with the available functional and structural information obtained for these proteins by experiment.

Structural proteins

There are three structural proteins in the mature DENV virion (capsid protein C, membrane protein M and envelope protein E), while the immature intracellular virus also contains prM protein, which is a precursor of M [87]. The gene order for the structural proteins from the 5′ terminus of the DENV genome is C–prM/M–E. The viral particles consist of an outer glycoprotein shell and an internal host-derived lipid bilayer that encapsulates the RNA/protein core consisting of genome RNA and capsid protein C.

The capsid protein C

Located at the very end of the 5′ terminus of the DENV genome, protein C is the first viral protein synthesized during translation. In the sequence of the DENV polyprotein, protein C corresponds to the first 113 residues (e.g. residues 2–114 in the genome polyprotein of DENV-2; UniProt ID Q91H74). This highly basic protein is enriched in lysine and arginine residues, which account for 25% of its sequence, and has a major function of interaction with and protection of the virion RNA [87]. It is predicted to be one of the most disordered DENV proteins (see Table 1 and Fig. 6A). Despite little sequence conservation overall, protein C contains a conserved internal hydrophobic segment (around region 45–65) that acts as a membrane anchor domain [88,89]. In infected cells, protein C has been shown to be integrated in the membrane of the endoplasmic reticulum in a 'hairpin' conformation, with both positively charged N- and C-terminal tails protruding into the cytoplasm and the centrally located hydrophobic signal/anchor segment being embedded into the membrane [88,89].

There are several experimental observations supporting the intrinsically disordered nature of protein C. For example, in SDS/PAGE experiments, protein C migrates at an apparent molecular mass of 16–18 kDa, compared with its expected molecular mass of 11.4 kDa [88]. Abnormal electrophoretic mobility, whereby the aberrant molecular mass noticeably exceeds the expected molecular mass of a target protein, is a characteristic feature of IDPs [90,91]. Also, both the N- and C-terminal regions of membrane-

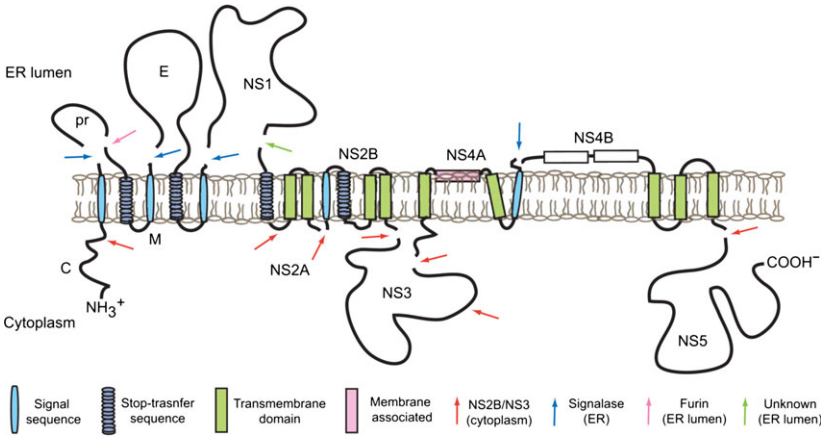

Fig. 5. Schematic representation of the DENV polyprotein embedded in the membrane of the endoplasmic reticulum. Cleavage sites of various proteases are indicated by colored arrows. Adapted with permission from [86].

Table 1. Names and abbreviated names of proteins from the dengue virus. The disorder level was evaluated for the genome polyprotein of DENV-2 (UniProt ID P29990).

Protein name	Abbreviation	Disorder level predicted using PONDR® VSL2	
		Percentage disordered	Averaged score
Capsid protein C	C	36.0	0.45
Endoplasmic reticulum anchor for protein C	ER	21.4	0.34
prM	prM	21.1	0.32
Envelope protein E	E	18.4	0.34
Non-structural protein 1	NS1	17.9	0.34
Non-structural protein 2A	NS2A	9.2	0.23
Serine protease subunit NS2B	NS2B	30.8	0.35
Serine protease NS3	NS3	24.8	0.38
Non-structural protein 4A	NS4A	9.5	0.25
Peptide 2k	2k	30.4	0.30
Non-structural protein 4B	NS4B	14.5	0.27
RNA-directed RNA polymerase NS5	NS5	22.3	0.37

bound protein C are easily accessible to proteolysis [88], which is another characteristic feature of IDPs/IDPRs [67,91]. Recent analysis of a conserved segment of DENV capsid protein C that inhibits the interaction of protein C with host intracellular lipid droplets using a combination of bioinformatics and biophysical techniques revealed that this membrane-binding peptide is disordered in the unbound form and folds to an α-helical conformation upon binding to negatively charged phospholipid membranes [92].

In addition to the membrane-bound form, protein C may exist as a soluble homodimer with pronounced helical structure [93,94]. For example, the analysis of the ^{15}N HSQC spectrum of DENV-2 protein C revealed that although the major part of this protein is characterized by a well-defined fold, approximately 20% of the chemical shifts approximate random-coil values, indicating the presence of an unstructured region within the N-terminal tail of the protein [93,94]. In fact, in the solution NMR structure of the fragment corresponding to residues 1–100 of DENV-2 protein C, the fragment comprising residues 21–100 behaved as an ordered domain and was included in the calculation, whereas the fragment comprising residues 1–20 was intrinsically disordered. The ordered domain was involved in formation of a symmetric dimer characterized by an unusual fold consisting of a large dimerization interface that includes two pairs of helices, one of which has characteristics of a coiled coil (see Fig. 6) [94]. Large dimer interfaces are typical of two-state dimers; i.e. dimers that are formed from the mostly unfolded monomers as a result of concomitant folding and binding events [95]. In agreement with this notion, Fig. 6C,D show the results of computational disassembly of the protein C dimer, indicating that the monomers possess very unusual shapes that are not consistent with simple globular structure, but contain long protrusions or extensions. Therefore, it is likely that individual chains in the DENV-2 protein C dimer are disordered in their unbound forms and fold upon dimer formation. In conclusion, the DENV capsid protein possesses functional and structural plasticity, being able to exist in membrane-bound and unbound forms and forming helical dimers in solution via a coupled folding-binding mechanism.

Membrane glycoprotein prM/M

The glycoprotein shell has a well-defined structure that includes 180 copies each of envelope protein E and membrane protein prM/M [86]. The DENV particles are able to exist in mature and immature forms that are very different morphologically (immature particles are 'spiky', whereas mature particles are 'smooth') [86]. In the immature virion, 90 heterodimers of prM and E extend as 60 trimeric spikes from the particle surface, whereas protein E of the mature virion forms 90 homodimers that lie flat against the viral surface, forming a 'smooth' protein shell [86]. The transition between these two forms of viral particle was shown to be driven by conformational changes in the viral E and prM/M proteins, with the predominant role being played by structural changes in protein E [86]. The immature virions are protected against premature fusion with the host membrane by the 'pr' peptide [96,97]. During transition from the immature to the mature form, the pr peptide is cleaved from the prM protein, and protein M acts in the mature particle as a transmembrane protein located beneath the E protein shell [86]. Figure 7A,B shows that both proteins are predicted to have some degree of disorder.

The dengue virus prM glycoprotein consists of 166 amino acids. During the maturation process, this protein is cleaved by furin, releasing the N-terminal 'pr' polypeptide (residues 1–91) and the mature M protein containing the ectodomain (residues 92–130) and the C-terminal transmembrane region (residues 131–166) [96,97]. Figure 7C shows crystal structures of a hybrid protein consisting of prM protein in which the transmembrane region of prM has been replaced by an

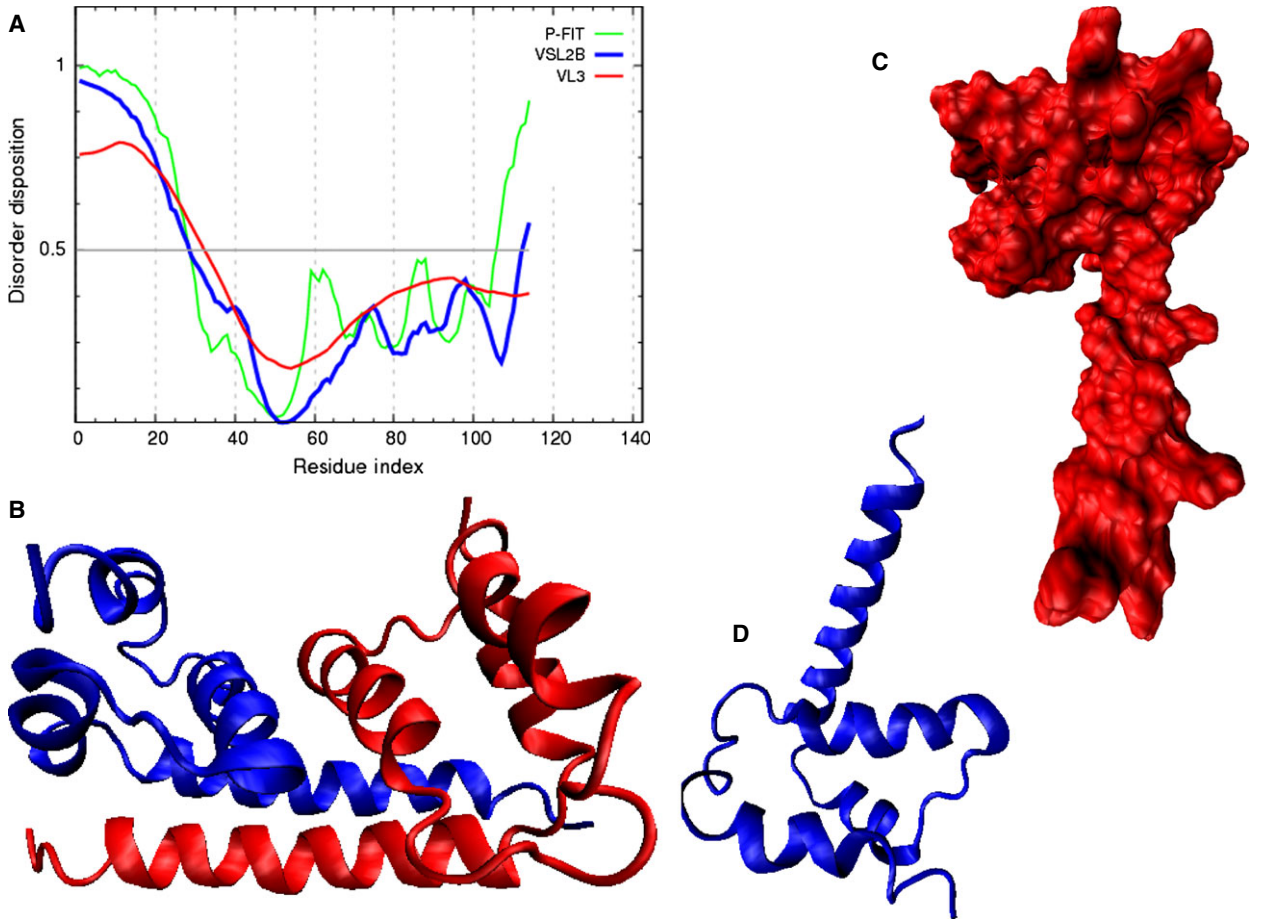

Fig. 6. Predicted intrinsic disorder (A) and solution NMR structure of DENV-2 protein C (B–D) (PDB ID 1R6R). The structure of a dimer (B) and two monomers presented as accessible surface (C) and in ribbon form (D) are shown to emphasize the unusual shape of the protein C monomers. The predisposition for intrinsic disorder in this part of the DENV-2 polyprotein (residues 2–114 of UniProt ID P29990) was evaluated using PONDR® VSL2, PONDR® VL3 and PONDR-FIT. Scores > 0.5 correspond to disordered residues/regions. These predictors were chosen because PONDR® VSL2B is a relatively accurate stand-alone disorder predictor suitable for prediction of short and long IDPRs [159,162,163], PONDR® VL3 possesses high accuracy for long IDPRs [164], and meta-predictor PONDR-FIT is one of the more accurate disorder predictors [165].

eight amino acid linker and DENV-2 protein E, solved at pH 5.5 and pH 7.0 [98]. No significant structural difference was found between these two structures. Although the prM protein used in this crystallographic study contained 130 residues, only the first 81 residues corresponding to the portion of the 'pr' polypeptide were resolved in the crystal structure, and no coordinates were determined for the remaining part of the protein, indicating that prM residues 82–130 are disordered.

Envelope protein E

The structural transition from immature ('spiky') to mature ('smooth') morphology of the DENV particle occurs while particles are in transit through the *trans*-Golgi network and is driven by conformational changes in protein E [86,99–101]. Figure 7D shows the crystal structure of the head-to-tail dimer of the ecto-domain of DENV-2 protein E (sE, corresponding to a large tryptic fragment of protein E), indicating that monomers of this protein are characterized by a highly extended shape. Structurally, the sE glycoprotein consists of three domains: domain I (DI), which is the N-terminal, but structurally central, domain; domain II (DII), which is the fusion (or dimerization) domain, containing the hydrophobic fusion peptide (residues 98–110); domain III (DIII), which is the putative receptor-binding domain III [101]. Of the 395 amino acids in the polypeptide chain, 390 (residues 1–16, 19–224 and

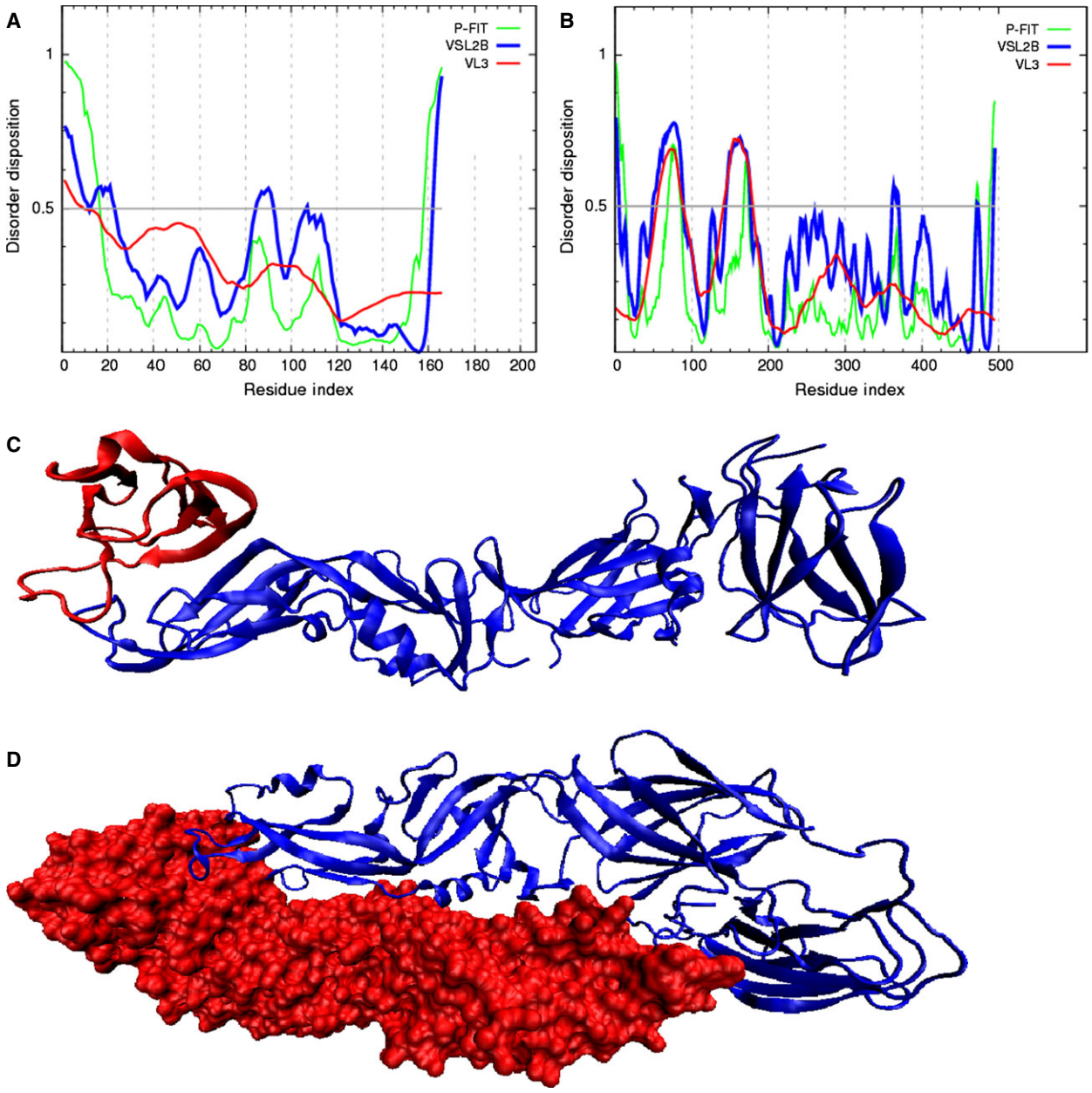

Fig. 7. Predicted disorder (A, B) and crystal structure of DENV-2 proteins prM and E (C, D). (A, B) The predisposition for intrinsic disorder in residues 115–280 of the DENV-2 polyprotein (prM) (A) and residues 281–775 of the DENV-2 polyprotein (protein E) (B) (UniProt ID P29990) was evaluated using PONDR® VSL2, PONDR® VL3 and PONDR-FIT. Scores > 0.5 correspond to disordered residues/regions. (C) Crystal structure for a construct comprising prM in which the transmembrane region has been replaced by an eight amino acid linker (red ribbon) and protein E from DENV-2 (blue ribbon) (PDB ID 3C6E) [98]. (D) Crystal structure of the dimer of the ectodomain of DENV-2 protein E (residues 1–395, PDB ID 1OKE). One monomer is shown as a surface and the monomer is shown in ribbon form.

228–395) were visible in the electron density map [101]. sE consists of two rigid bodies (DI + DIII and DII) connected by a flexible hinge. Protein E is a multi-functional protein that is able to form heterodimers with the prM protein in the immature virion or homodimers in the mature virus. Protein E also represents the first point of contact between the virus and the host cell, and is able to interact with several cellular proteins and carbohydrate molecules [102]. Among host proteins targeted by protein E are binding immunoglobulin protein (BiP) also known as 78 kDa glucose-regulated protein (GRP-78) or heat shock 70 kDa protein 5

(HSPA5) [103], Rab5 [104], heat shock protein 70 family proteins [105], dendric cell-specific ICAM3 [intercellular adhesion molecule 3 also known as CD50 (Cluster of Differentiation 50)]-grabbing non-integrin or DC-SIGN2 [106,107], and macrophage mannose receptor [108]. Curiously, large parts of these binding partners of protein E are intrinsically disordered. Analysis using PONDR® VSL2 (http://www.pondr.com/cgi-bin/PONDR/pondr.cgi) showed that GRP-78/BiP (UniProt ID P110121), Rab5 (UniProt ID Q9UJ41), heat shock protein 70 (UniProt ID P08107), DC-SIGN2 (UniProt ID Q9H2X3) and mannose receptor (UniProt ID Q9UBG0) contain 41.3%, 54.5%, 29.6%, 62.2% and 26.3% disordered residues, respectively.

Non-structural proteins

Seven non-structural (NS) proteins are derived from the C-terminal sequence of DENV polyproteins in the order NS1–NS2A–NS2B–NS3–NS4A–NS4B–NS5 (see Fig. 5).

Non-structural protein NS1

NS1 is a 45 kDa glycoprotein that is translocated into the lumen of the endoplasmic reticulum and secreted from the cell. NS1 glycoproteins exhibit a high degree of sequence homology among various flaviviruses [109]. Although NS1 contains no N- or C-terminal membrane anchoring sequences [110], this protein exists predominantly in the dimeric form, associated with intracellular and cell-surface membranes [111] via a glycosylphosphatidylinositol anchor [112]. The NS1 protein is involved in viral RNA replication [113]. It exists as immature hydrophilic monomers in the lumen of the endoplasmic reticulum, as a stable hydrophobic homodimer that interacts with membranes, as a secreted soluble hexamer, or may be expressed on the surface of infected cells and serve as a target for human antibody responses to DENV infection. Maturation of NS1 involves N-linked glycosylation at two conserved N-linked glycosylation sites, and high-mannose oligosaccharide moieties are present at both these sites in the cell-associated form, whereas in the secreted form, one site binds complex glycans whereas the other binds a high-mannose oligosaccharide moiety [111,114,115]. NS1 dimers bind to other non-structural viral proteins, and, via this association, with viral RNA. The dimeric form of this protein may be also involved in assembly of the viral replicase complex and its localization to cytoplasmic membranes [116,117].

Mature DENV NS1 contains 352 amino acids, including 12 cysteine residues that are absolutely conserved among all flavivirus NS1 proteins [118,119] and form six invariant intramolecular disulfide bonds [120]. Disulfide-bonded NS1 monomers form non-covalent dimers [119] after post-translational modification in the lumen of the endoplasmic reticulum [121]. The dimer is processed in the *trans*-Golgi network, and secreted into the extracellular space as a hexamer [121]. This extracellular hexameric NS1 exists in the form of a lipoprotein particle with an open-barrel protein shell and a prominent central channel that is rich in lipids, such as triglycerides, cholesteryl esters and phospholipids [122]. This suggests that the DENV NS1 protein may hijack lipid metabolic pathways of the host, thereby contributing to a key feature of severe dengue disease: endothelium dysfunction [122]. Furthermore, this secreted form of NS1 plays a significant role in immune evasion and modulation during infection [123]. In the sera of some DENV-infected patients, high levels of secreted NS1 are accumulated, reaching concentrations of up to 50 $\mu g \cdot mL^{-1}$ [124–126].

For a long time, no high-resolution structural information on the full-length NS1 protein was available, and the only source of such information was a low-resolution structure of the hexameric secreted form of NS1 (resolution 23 Å) obtained from electron microscopy images [123]. However, the crystal structures of the C-terminal fragments (residues 172–352) of West Nile virus and Dengue virus NS1 proteins at 1.85 and 2.7 Å resolution, respectively, have recently been reported [120]. $NS1_{172–352}$ is folded into a unique rod-shaped head-to-head dimer with a long axis of approximately 9.9 nm [120] (see Fig. 8A,B). This novel fold is characterized by a highly asymmetric spatial organization, with one face being composed of a 16-stranded β-platform (Fig. 8A) and the other face representing a complex arrangement of protruding loops and 3_{10} helices [120] (Fig. 8B). The β-structural core of the $NS1_{172–352}$ monomer is formed by 11 β-strands (residues 185–189, 192–196, 200–206, 201–217, 273–277, 284–287, 298–299, 310–313, 321–324, 329–331 and 336–337). The loops connecting these β-strands range in length from 2 to 55 residues, with the longest loop between β-strands 4 and 5 (residues 218–272) containing four β-turns and five 3_{10} helices [120]. Residues 172–177 are not visible in the electron density map. Figure 8C shows that, although DENV NS1 is predicted to be mostly structured, this protein is expected to have several disordered or flexible regions (i.e. regions with disorder scores exceeding or approaching 0.5, correspondingly).

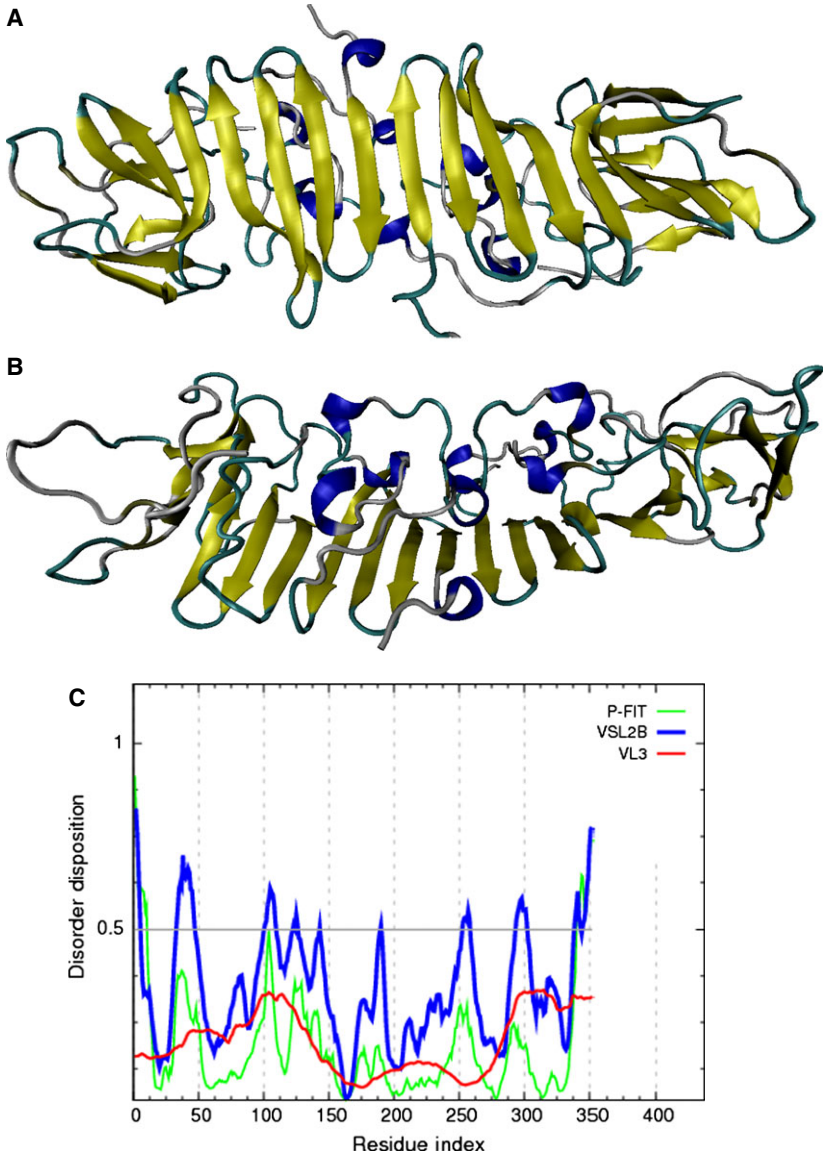

Fig. 8. Structural characterization of the DENV-1 NS1 protein. (A,B) Crystal structure of the head-to-head dimer of a C-terminal fragment (residues 172–352) of DENV-1 NS1 protein (PDB ID 4OIG): (A) top view; (B) side view. (C) Predicted per-residue disorder predisposition of the full-length DENV-1 NS1 protein. The predisposition for intrinsic disorder in this part of the DENV-2 polyprotein (residues 776–1127; UniProt ID P29990) was evaluated using PONDR® VSL2, PONDR® VL3 and PONDR-FIT. Scores > 0.5 correspond to disordered residues/regions.

Non-structural protein NS2A

No structural information is yet available on the hydrophobic NS2A protein (approximately 22 kDa, with 42% of amino acids of the DENV NS2A protein being hydrophobic). The NS2A protein serves as a component of the viral replication complex that functions in virion assembly and antagonizes the host immune response [127,128]. Two distinct sets of NS2A molecules were shown to be responsible for DENV RNA synthesis (NS2A molecules located in the viral replication complex) and virion assembly (NS2A molecules located in the virion assembly/budding site) [128]. The DENV-2 NS2A protein contains five integral transmembrane segments (residues 69–93, 100–118, 143–163, 165–186 and 189–209) that span the lipid bilayer of the membrane of the endoplasmic reticulum [129]. Furthermore, this protein has two membrane-associated segments (residues 32–51 and 120–140) that interact with the membrane of the endoplasmic reticulum without traversing the lipid bilayer [129]. NMR analysis of the first transmembrane segment (residues 69–93) of the DENV-2 NS2A protein revealed that this region contains two helices connected by a Pro85-mediated 'helix breaker' [129]. In agreement with its highly hydrophobic nature and predominant transmembrane localization, the DENV-2 NS2A protein is predicted to be mostly ordered (Fig. 9A).

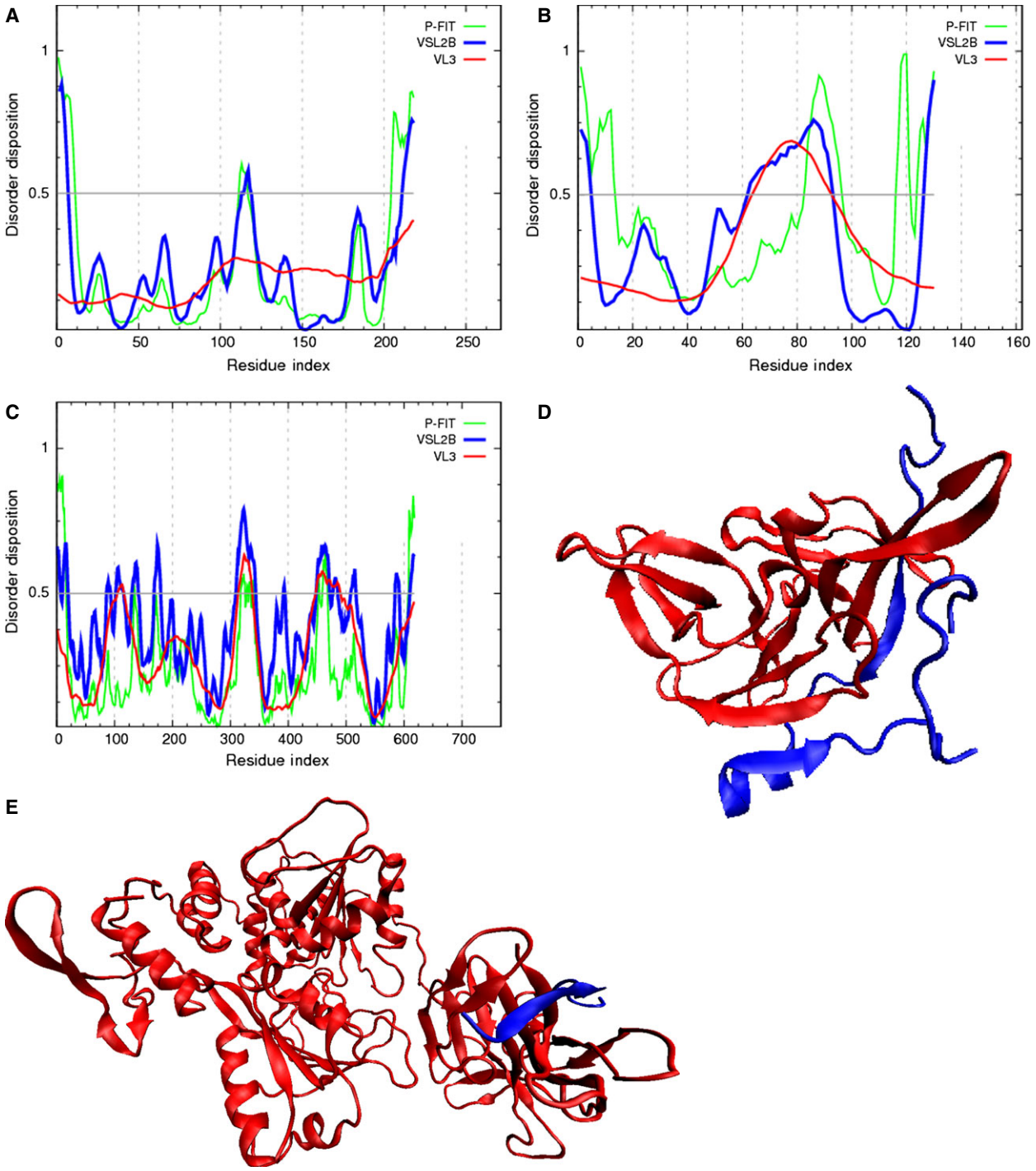

Fig. 9. Predicted disorder (A–C) in DENV-2 proteins NS2A, NS2B and NS3, and crystal structures of the DENV-2 NS2B–NS3 complex (D,E). (A–C) The predisposition for intrinsic disorder in residues 1128–1345 of the DENV-2 polyprotein (NS2A), residues 1346–1475 of the DENV-2 polyprotein (NS2B) and residues 1476–2093 of the DENV-2 polyprotein (NS3) (UniProt ID P29990) was evaluated using PONDR® VSL2, PONDR® VL3 and PONDR-FIT. Scores > 0.5 correspond to disordered residues/regions. (D) Crystal structure of the NS2B–NS3 complex comprising the C-terminal fragment of the NS2B protein (residues 1394–1440, blue ribbon) and the N-terminal protease domain of the NS3 protein (residues 1493–1641, red ribbon) (PDB ID 2FOM). (E) Crystal structure for the complex between full-length DENV-4 NS3 (red ribbon) and a short fragment corresponding to residues 45–75 of the DENV-4 NS2B protein (PDB ID 2VBC).

Non-structural proteins NS2B and xml: the two-component NS2B–NS3 protease and helicase NS3

In flaviviruses, the viral NS3 protease is required for processing of the viral polyprotein and subsequently for virus replication. NS2B is an integral membrane protein that serves as a cofactor that plays an essential role in regulation of the catalytic activity of the NS3 protease.

Figure 9B shows that the DENV-2 NS2B protein is predicted to have substantial levels of disorder. No information is available about the structure of the NS2B protein in the unbound form. According to NMR and circular dichroism analyses, the intact DENV-2 NS2B protein adopts α-helical structures in lipid micelles [130]. Curiously, in the NS2B–NS3 complex, a fragment corresponding to the central part of the NS2B protein (residues 1394–1440) is loosely folded, being wrapped around the protease domain of the NS3 protein (Fig. 9D). The ability of the NS2B to wrap around the protease domain of NS3 in a 'belt-like' structure defines the ability of this protein to serve as an integral part of the protease active site [86]. Furthermore, while the central region of NS2B (residues 67–80) interacts with the protease, flanking hydrophobic regions of NS2B are expected to be responsible for anchoring the NS2B–NS3 complex in the membrane of the endoplasmic reticulum [86].

Figure 9D shows that, in its NS3-bound form, NS2B has 9% helical and 16% β-sheet structure, being folded in a mixed structure containing two short β-strands, a short α-helix, and a long region with irregular structure [131]. The bound form of this protein is characterized by a very large interface. Therefore, there is a high probability that this NS2B fragment is disordered in the unbound state and folds upon binding to NS3. In agreement with this hypothesis, Fig. 9B shows that the central region of the DENV NS2B protein is predicted to be mostly disordered.

As far as DENV NS3 is concerned, this protein is a multi-functional enzyme of 618 amino acids that functions both as a chymotrypsin-like serine protease and as an RNA helicase and RTPase/NTPase. The protease domain is located at the N-terminal domain of NS3 (residues 1–180), and cleaves the DENV polyprotein at several sites, generating ends of several viral proteins (see Fig. 5). Figure 9D shows the PONDR-based disorder profile of DENV-2 NS3, indicating that this protein has several disordered regions. Figure 9E shows the crystal structure of a complex between full-length DENV-4 NS3 and a short fragment of the DENV-4 NS2B protein (residues 49–66), which are linked to the N-terminus of full-length NS3 by a Gly–Ser linker. Curiously, although the NS2B fragment and linker used in this study were 31 residues long, only half of these residues were visible in the complex, with the remaining parts being disordered [132]. This observation provides further support for the highly disordered nature of the NS2B protein.

Figure 9E shows that the DENV-4 NS3 protein is an extended molecule (with approximate overall dimensions of 100 Å × 60 Å × 40 Å), with the protease domain located next to the entrance of the ATPase active site between helicase sub-domains I and II (residues 181–326 and 327–481, respectively) and opposite to sub-domain III (residues 482–618) [86,132]. Importantly, linkers connecting the protease and helicase domains, as well as linking the helicase sub-domains, are predicted to be located within the disordered regions. It was also observed that an amino acid sequence of the interdomain linker connecting the protease and helicase domains has little conservation between the flaviviruses [86]. Analysis of the NS3 crystal structure revealed that the protease domain, the interdomain (linker) region, and sub-domain II from the helicase domain are the most mobile parts of the protein [132].

In addition to serving as a dual protease–helicase enzyme that interacts with specific cleavage sites of the DENV polyprotein and with RNA, NS3 protein binds to DENV polymerase NS5 [133]. Also, in addition to being responsible for maturation of several DENV proteins, the NS3 protein was shown to undergo autocleavage at two sites located at the NS2B–NS3 junction and within the helicase C-terminal region, respectively [134]. As the NS3 polypeptide substrate must be accommodated into its own protease active site during this *cis* cleavage event, intramolecular proteolysis requires significant structural plasticity of the NS3 protein and order/disorder transitions within the NS3 polypeptide chain [86,134]. Also, within the helicase sub-domain I, a segment of 10 residues, Gln243 to Thr252, which was disordered in the structure of the isolated DENV-2 helicase domain, in the full-length NS3 protein structure, forms a β-strand that runs antiparallel to a β-strand from the helicase sub-domain II of a neighboring molecule [132].

Non-structural proteins NS4A and NS4B

NS4A (16 kDa) and NS4B (27 kDa) are highly hydrophobic integral membrane proteins, but structural characterization attempts have been mostly unsuccessful so far. Functionally, NS4A plays a role in induction of the membrane alterations required for virus replication [135,136], whereas NS4B is involved in

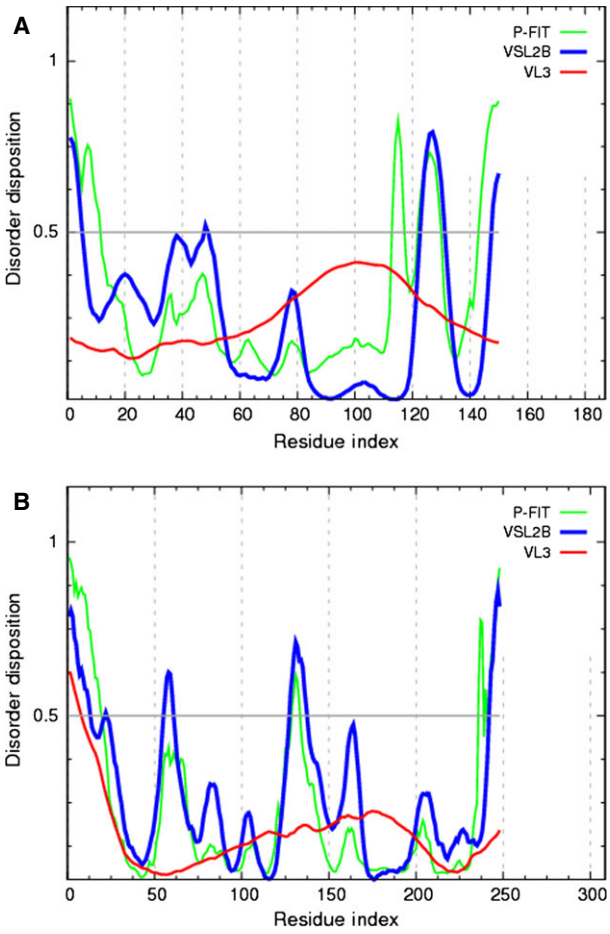

Fig. 10. Predicted disorder in DENV-2 proteins NS4A (A) and NS4B (B). The predisposition for intrinsic disorder in residues 2094–2243 of the DENV-2 polyprotein, which include the NS4A protein (residues 2094–2220) and the 2k peptide (residues 2221–2243), and residues 2244–2491 of the DENV-2 polyprotein (NS4B protein) (UniProt ID P29990) was evaluated using PONDR® VSL2, PONDR® VL3 and PONDR-FIT. Scores > 0.5 correspond to disordered residues/regions.

direct interactions with the NS3 protein, dissociates NS3 from single-stranded RNA, and consequently enhances the helicase activity of NS3, assisting viral RNA replication [137].

Figure 10 shows that, despite their highly hydrophobic nature and transmembrane localization, both NS4A and NS4B are predicted to have several disordered regions, including disordered N- and C-terminal tails of both proteins. In agreement with these predictions, the N-terminal cytoplasmic tail of the NS4A protein (residues 1–48) was shown to be mostly disordered in solution, but was able to form an amphipathic α-helix (i.e. α-helical protein regions in which one face of the helix is hydrophobic while the opposite face is hydrophilic) in the presence of membranes [138,139]. The DENV NS4B protein was recently shown to homodimerize *in vitro*, as evidenced by gel filtration, chemical cross-linking, and multi-angle light scattering experiments [140]. This dimerization is driven by the cytosolic loop (amino acids 129–165) and the C-terminal region (amino acids 166–248) [140]. Figure 10B shows that the cytosolic loop crucial for NS4B dimerization is predicted to be mostly disordered.

Bi-functional non-structural protein xml: viral methyltransferase polymerase

The NS5 protein is the largest DENV protein (900 residues, 104 kDa) and the most conserved DENV protein, wuth 67% sequence identity among the four DENV serotypes [86]. This bi-functional enzyme comprises an N-terminal methyltransferase (residues 1–265) and a C-terminal RNA-dependent RNA polymerase (RdRp) domain (residues 266–900). The crystal structure of the full-length protein became available very recently (PDB ID 4V0Q, Fig. 11A) [141], but structures for the methyltransferase and RdRp domains have been known for longer time [e.g. methyltransferase domain of DENV-2 NS5 (PDB ID 1L9K) [142]; RdRp domain of DENV-3 NS5 (PDB ID 2J7W) [143]]. As structural information is available for the NS5 protein from the DENV2 and DENV3, we compared their intrinsic disorder predispositions. Figure 11B,C shows the disorder profiles of the DENV-2 and DENV-3 NS5 proteins, respectively, indicating that these NS5s have remarkably similar disorder propensities. Furthermore, these plots show that the linker region of the NS5 protein connecting its methyltransferase and RdRp domains is predicted to be highly disordered. It is likely that the highly disordered and flexible nature of this linker precludes crystallization of the full-length protein. Although the linker region located between the two functional domains of NS5 may be flexible, it also plays an important role in regulation of the polymerase initiation activities of the RdRp domain by stabilizing this domain and enhancing turnover of the RNA and NTP substrates [144]. This conclusion was drawn based on comparative analysis of the DENV-3 RdRp fragments spanning residues 265–900 (RdRp with enhanced catalytic properties) and residues 272–900 [144]. The analysis revealed that DENV-3 NS5 residues 269–271 interact with the RdRp domain, suggesting that the flexibility between the methyltransferase and RdRp domains is determined by residues 263–268 of this NS5 protein [144]. Figure 11 shows that both func-

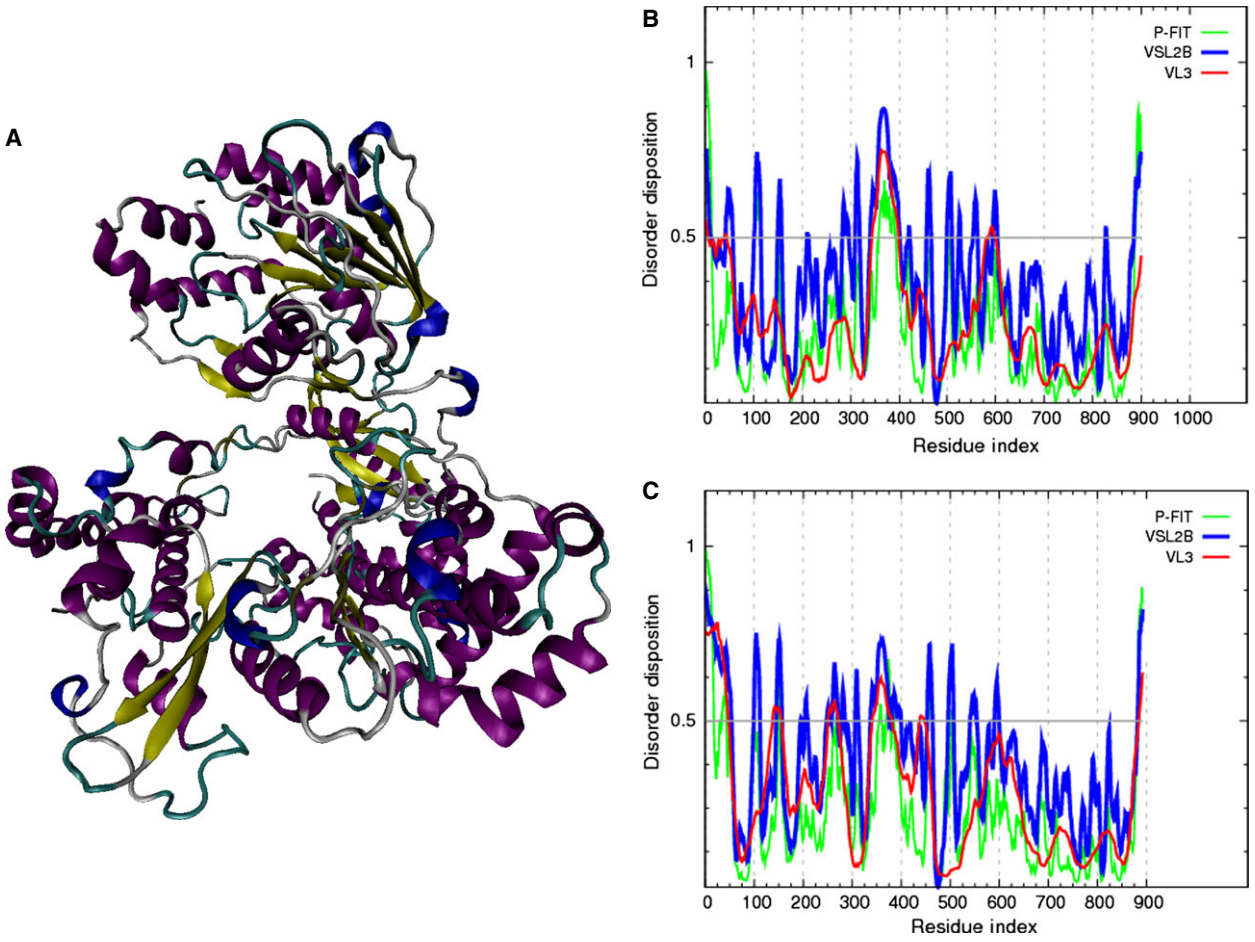

Fig. 11. Crystal structures of the DENV-3 NS5 protein (PDB ID 4V0Q) (A). (B, C) The predisposition for intrinsic disorder of the bi-functional methyltransferase/RNA-dependent RNA polymerase from DENV-2 (residues 2492–3391 of the DENV-2 polyprotein, UniProt ID P29990) (B) or DENV-3 (residues 2494–3395 of the DENV-3 polyprotein, UniProt ID Q6DLV0) (C) was evaluated using PONDR® VSL2, PONDR® VL3 and PONDR-FIT. Scores > 0.5 correspond to disordered residues/regions.

tional domains are rather loopy. For example, the secondary structure content of the methyltransferase is 30% α-helical, 13% β-structural and 57% irregular secondary structure elements [142], whereas the RdRp domain contains 42% α-helical, 10% β-structural and 50% irregular secondary structure elements [143].

Conclusions

Many viral proteins or their biologically important regions do not have unique 3D structures, instead being intrinsically disordered. Numerous functions of viral proteins rely heavily on intrinsic disorder. Although knowledge of the 3D structures of various viral proteins is used in the rational design of antiviral drugs, the biology of viruses is critically dependent on the intrinsically disordered protein regions. In fact, these highly flexible regions are used by viruses as specific tools to invade host organisms and hijack various host systems. Disorder also helps viruses to accommodate to their hostile habitats. We evaluated the applicability of these general conclusions to DENV proteins using a set of bioinformatics tools to assess the abundance and distribution of intrinsic disorder in these proteins, and to understand the potential functional roles and evolvability of such disorder. This analysis revealed that several proteins (ER, E, 1, 2A and 4A) are expected to be mostly ordered, whereas four proteins (C, 2k, NS3 and NS5) are predicted to have high disorder levels. We also showed that, although disordeerd regions are typically very polymorphic, the high density of polymorphism is not unique to the disordered regions, and two of the mostly structured proteins, 2A and 4A, contain a high density of polymorphisms. Established disorder-related functions in the DENV proteins include protein–protein interac-

tions and binding of nucleic acids, metals and other small molecules. The IDPRs in the DENV proteins are frequently post-translational modification sites, and known cleavage sites are depleted in polymorphic sites and have a high propensity for disorder, especially relative to the neighboring residues. Therefore, many DENV proteins, irrespective of their functions, have biologically important disordered regions. The list of functions attributed to IDPRs of DENV proteins overlaps with the disorder-based activities of proteins from other organisms.

Experimental procedures

Dataset

We collected all complete genomes of the dengue virus from UniProt [145] in June 2014. The query consisted of 'dengue virus' as the organism keyword, and was restricted to reviewed entries. The latter criterion ensures that the genomes, particularly the corresponding polyproteins, have been reviewed manually, and that they include functional annotations. The query returned 28 polyprotein sequences, including eight fragments that were excluded. The remaining 20 polyproteins cover all four serotypes of dengue virus [146]. They include three sequences from DENV-1 (UniProt ID P33478, P27909 and P17763), seven from DENV-2 (UniProt ID P29990, P29991, P14337, P07564, P14340, Q9WDA6 and P12823), five from DENV-3 (UniProt ID Q99D35, Q5UB51, Q6YMS3, P27915 and Q6YMS4), and five from DENV-4 (UniProt ID P09866, Q2YHF2, Q58HT7, Q5UCB8 and Q2YHF0). The lengths of the 20 polyproteins are similar, and range between 3387 and 3396 residues. Each polyprotein encodes 12 protein chains (Table 1) for which cleavage sites were annotated in UniProt [145]. In total, there are 240 protein chains from the 20 polyproteins of the four serotypes. The lengths of these individual viral proteins vary between 14 and 904 residues.

Functional and structural annotations

We collected 20 types of structural and functional annotations for each of the considered 20 polyproteins. These were obtained from a variety of resources, including multiple sequence alignment using Clustal Omega [147], Uniprot [145], the ELM resource server [148], and two prediction methods: MFDp, which was used to annotate putative intrinsic disorder [149], and MoRFpred, which predicts MoRFs [150]. These annotations include polymorphisms within and between virus serotypes, cleavage sites (CLV), transmembrane regions ('Trans'), intra-membrane regions ('Intra'), topological cytoplasmic, extracellular and luminal domains (Topo-cy, Topo-ex and Topo-lu), functional sites, ELMs, IDPRs, disorder-promoting amino acids (DisPro-AA) [74] and MoRFs.

Annotations of polymorphisms, i.e. changes in amino acid type between aligned polyproteins, were derived by performing a multiple sequence alignment using Clustal Omega [147]. We assessed polymorphisms in polyproteins within each individual virus serotype and between virus serotypes. For the individual serotypes, we define three types of annotations: conserved positions; strong polymorphisms and weak polymorphisms. The conserved positions have identical amino acid type across all aligned sequences. The strong polymorphisms are defined as substitutions that involve amino acid types that are strongly dissimilar based on the PAM substitution matrix [151]; these are indicated by a space in the Clustal Omega output. The weak polymorphisms involve substitutions of amino acid types that are similar; these are indicated by colon and points in the Clustal Omega output and correspond to conserved and semi-conserved substitutions. For the polymorphisms over multiple serotypes, we combined the polymorphism status of each virus serotype to define conserved positions, strong polymorphisms and weak polymorphisms. For a given position, if two or more virus serotypes were annotated as having polymorphisms, and at least one of them is strong, then we defined this position as a strong polymorphism. If two or more serotypes were annotated as having weak polymorphisms (i.e. neither is strong), we defined this position as a weak polymorphism. We define the remaining positions as conserved. We analyzed strongly polymorphic positions and all polymorphic positions (weak and strong combined) separately.

Annotations of cleavage sites, 'Trans' and 'Intra' regions, the three types of topological domains and functional sites were derived directly from UniProt [145] for each of the 20 polyproteins. The functional sites are a combination of annotations of regions of interest, active sites, binding sites, other functional sites (except cleavage sites), and nucleotide binding regions. In other words, a given position is annotated as functional if any of these annotations is true.

ELMs are short conserved functional sequence motifs, usually between 3 and 11 residues in length [152] that are often found in IDPRs [148]. We include annotations of all six types of ELMs as defined by the ELM server [148]. They include motifs that serve as proteolytic cleavage sites (ELM_CLV), post-translational modification sites (ELM_MOD), motifs for recognition and targeting to subcellular compartments (ELM_TRG), generic ligand binding motifs (ELM_LIG), degron motifs that are involved in polyubiquitylation and targeting the protein to the proteasome for degradation (ELM_DEG), and docking motifs that correspond to site of interactions with modifying enzymes that are distinct from active sites (ELM_DOC). The ELM_DEG and ELM_DOC motifs are specific sub-types of the ligand-binding motifs that were introduced to improve discrimination of functions of ELMs [148]. These annotations were parsed

from the results of the ELM motif search after globular domain filtering, structural filtering and context filtering.

The putative intrinsically disordered residues were determined using the MFDp webserver (http://biomine.ece.ualberta.ca/MFDp/). MFDp is a sophisticated consensus predictor that combines disorder predictions generated by IUPred [153], DISOclust [154] and DISOPRED2 [155], sequence profiles and predictions of secondary structure generated by PSIPRED [156], relative solvent accessibility and backbone dihedral torsion angles generated by Real-SPINE3 [157], B-factors generated by PROFbval [158], and globular domains generated by IUPred. The MFDp predictor is characterized by a highly competitive predictive performance, with an AUC (Areas Under the Receiver Operating characteristic Curve) > 0.81 based on multiple benchmark tests [149,159]. We utilize both propensity scores (a larger score means that a given residues is more likely to be disordered) and binary annotations (residues with propensities above a cut-off value are classified as disordered; otherwise they are assumed to be structured). We consider two types of binary annotations: using the default propensity score cut-off of 0.37, and using a higher cut-off of 0.57 that corresponds to higher-quality disorder predictions with a low false-positive rate of 0.05 (MFDp_H). These cut-off values are based on a comprehensive benchmark test [159].

We also predict MoRFs, which are short protein-binding regions (5–25 consecutive amino acids) located within longer IDPRs that undergo coupled folding and binding, i.e. disorder-to-order transition, upon binding [160,161]. These regions were predicted using the MoRFpred webserver [150]. Similar to MFDp, this method is characterized by a state-of-the-art design that combines information from pairwise alignments, sequence profiles, and predictions of disorder generated by IUPred [153], DISOPRED2 [155], DISOclust [154] and MFDp [149], solvent accessibility generated by Real-SPINE3 [157], and B-factors generated by PROFbval [158]. We utilize both propensity scores and binary predictions, with the latter being computed using the default cut-off of 0.5. We consider a region to be a MoRF only if it is at least five residues long; shorter putative regions were removed. Both MFDp and MoRFpred use sequences of individual proteins for prediction, and thus we first divided polyproteins into proteins, performed predictions for each protein, and combined these predictions to annotate the full polyproteins.

Per-residue disorder analysis in DENV-2 proteins

For illustration purposes, the per-residue predisposition for intrinsic disorder in individual DENV-2 proteins was evaluated by predictors from the PONDR family. Here, scores above 0.5 were considered to correspond to disordered residues/regions. PONDR® VSL2B is one of the more accurate stand-alone disorder predictors [162], and comprehensive assessments of *in silico* predictors of intrinsic disorder showed that this tool performs reasonably well [159,163]. PONDR® VL3 possesses high accuracy for finding long IDPRs [164], whereas PONDR-FIT is a meta-predictor that is moderately more accurate than each of the component predictors [165].

Acknowledgements

This work was supported in part by a grant from King Abdulaziz University (56-130-35-HiCi).

Author contributions

FM and RAB performed experiments and analyzed data. HAA and EMR analyzed data and wrote the paper. LK and VNU planned experiments, performed experiments, analyzed data and wrote the paper.

References

1 Alexander N, Balmaseda A, Coelho IC, Dimaano E, Hien TT, Hung NT, Jänisch T, Kroeger A, Lum L & Martinez E (2011) Multicentre prospective study on dengue classification in four South-east Asian and three Latin American countries. *Tropical Med Int Health* **16**, 936–948.

2 Kalayanarooj S (2011) Dengue classification: current WHO vs. the newly suggested classification for better clinical application? *J Med Assoc Thai* **94**, S74–S84.

3 Rodenhuis-Zybert IA, Wilschut J & Smit JM (2010) Dengue virus life cycle: viral and host factors modulating infectivity. *Cell Mol Life Sci* **67**, 2773–2786.

4 Mier-y-Teran-Romero L, Schwartz IB & Cummings DAT (2013) Breaking the symmetry: immune enhancement increases persistence of dengue viruses in the presence of asymmetric transmission rates. *J Theor Biol* **332**, 203–210.

5 Franco L, Palacios G, Martinez JA, Vázquez A, Savji N, De Ory F, Sanchez-Seco MP, Martín D, Lipkin WI & Tenorio A (2011) First report of sylvatic DENV-2-associated dengue hemorrhagic fever in West Africa. *PLoS Negl Trop Dis* **5**, e1251.

6 Azami NAM, Salleh SA, Neoh H-M, Zakaria SZS & Jamal R (2011) Dengue epidemic in Malaysia: not a predominantly urban disease anymore. *BMC Res Notes* **4**, 216.

7 Weerakoon KG, Kularatne SA, Edussuriya DH, Kodikara SK, Gunatilake LP, Pinto VG, Seneviratne AB & Gunasena S (2011) Histopathological diagnosis of myocarditis in a dengue outbreak in Sri Lanka, 2009. *BMC Res Notes* **4**, 268.

8 Wan S-W, Lin C-F, Yeh T-M, Liu C-C, Liu H-S, Wang S, Ling P, Anderson R, Lei H-Y & Lin Y-S (2013) Autoimmunity in dengue pathogenesis. *J Formos Med Assoc* **112**, 3–11.

9 Haddow AD, Guzman H, Popov VL, Wood TG, Widen SG, Haddow AD, Tesh RB & Weaver SC (2013) First isolation of Aedes flavivirus in the Western hemisphere and evidence of vertical transmission in the mosquito *Aedes* (*Stegomyia*) *albopictus* (Diptera: Culicidae). *Virology* **440**, 134–139.

10 Madani TA, Abuelzein E-TM, Al-Bar H, Azhar EI, Kao M, Alshoeb HO & Bamoosa AR (2013) Outbreak of viral hemorrhagic fever caused by dengue virus type 3 in Al-Mukalla, Yemen. *BMC Infect Dis* **13**, 136.

11 Wright WF & Pritt BS (2012) Update: The diagnosis and management of dengue virus infection in North America. *Diagn Microbiol Infect Dis* **73**, 215–220.

12 Dammert AC, Galdo JC & Galdo V (2014) Preventing dengue through mobile phones: evidence from a field experiment in Peru. *J Health Econ* **35**, 147–161.

13 Jain B, Chaturvedi UC & Jain A (2014) Role of intracellular events in the pathogenesis of dengue; an overview. *Microb Pathog* **69–70**, 45–52.

14 Karunakaran A, Ilyas WM, Sheen S, Jose NK & Nujum ZT (2013) Risk factors of mortality among dengue patients admitted to a tertiary care setting in Kerala, India. *J Infect Public Health* **7**, 114–120.

15 Sreekanth GP, Chuncharunee A, Sirimontaporn A, Panaampon J, Srisawat C, Morchang A, Malakar S, Thuwajit P, Kooptiwut S, Suttitheptumrong A *et al.* (2014) Role of ERK1/2 signaling in dengue virus-induced liver injury. *Virus Res* **188**, 15–26.

16 Tseng C-H, Lin C-K, Chen Y-L, Hsu C-Y, Wu H-N, Tseng C-K & Lee J-C (2014) Synthesis, antiproliferative and anti-dengue virus evaluations of 2-aroyl-3-arylquinoline derivatives. *Eur J Med Chem* **79**, 66–76.

17 Fakeeh M & Zaki AM (2001) Virologic and serologic surveillance for dengue fever in Jeddah, Saudi Arabia, 1994-1999. *Am J Trop Med Hyg* **65**, 764–767.

18 Van Kleef E, Bambrick H & Hales S (2010) The geographic distribution of dengue fever and the potential influence of global climate change. *TropIKA.net* http://journal.tropika.net/scielo.php?script=sci_arttext&pid=S2078-86062010005000001&lng=en&nrm=iso.

19 Centers for Disease Control and Prevention (2010) Locally acquired dengue – Key West, Florida, 2009-2010. *MMWR Morb Mortal Wkly Rep* **59**, 577–581.

20 Azhar E, Kao M, Niedrig M, Masri B, Godus A, Badierah R, Khan N, Almazrooa A, Ashshi A & Jamjoom G (2010) Virological diagnosis of dengue fever in Jeddah, Saudi Arabia: comparison between RT-PCR and virus isolation in cell culture. *J Infect Dis Immun* **2**, 24–29.

21 Zaki A, Perera D, Jahan SS & Cardosa MJ (2008) Phylogeny of dengue viruses circulating in Jeddah, Saudi Arabia: 1994 to 2006. *Tropical Med Int Health* **13**, 584–592.

22 Ayyub M, Khazindar AM, Lubbad EH, Barlas S, Alfi AY & Al-Ukayli S (2006) Characteristics of dengue fever in a large public hospital, Jeddah, Saudi Arabia. *J Ayub Med Coll Abbottabad* **18**, 9–13.

23 Kuhn RJ, Zhang W, Rossmann MG, Pletnev SV, Corver J, Lenches E, Jones CT, Mukhopadhyay S, Chipman PR, Strauss EG *et al.* (2002) Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* **108**, 717–725.

24 Hadinegoro SRS (2012) The revised WHO dengue case classification: does the system need to be modified? *Paediatr Int Child Health* **32**, 33–38.

25 Soares ROS & Caliri A (2013) Stereochemical features of the envelope protein domain III of dengue virus reveals putative antigenic site in the five-fold symmetry axis. *Biochim Biophys Acta* **1834**, 221–230.

26 Mendonça BAA, de Sousa ACB, de Souza AP & Scarpassa VM (2014) Temporal genetic structure of major dengue vector *Aedes aegypti* from Manaus, Amazonas, Brazil. *Acta Trop* **134**, 80–88.

27 Reanney DC (1982) The evolution of RNA viruses. *Annu Rev Microbiol* **36**, 47–73.

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

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

30 Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW *et al.* (2001) Intrinsically disordered protein. *J Mol Graph Model* **19**, 26–59.

31 Dunker AK, Obradovic Z, Romero P, Garner EC & Brown CJ (2000) Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* **11**, 161–171.

32 Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* **27**, 527–533.

33 Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **11**, 739–756.

34 Uversky VN (2010) The mysterious unfoldome: structureless, underappreciated, yet vital part of any given proteome. *J Biomed Biotechnol* **2010**, 568068.

35 Uversky VN & Dunker AK (2010) Understanding protein non-folding. *Biochim Biophys Acta* **1804**, 1231–1264.

36 Uversky VN, Gillespie JR & Fink AL (2000) Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427.
37 Dunker AK, Cortese MS, Romero P, Iakoucheva LM & Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* **272**, 5129–5148.
38 Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF & Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**, 635–645.
39 Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C & Villafranca JE (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput*, 473–484.
40 Wright PE & Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J Mol Biol* **293**, 321–331.
41 Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS & Dunker AK (2005) Natively disordered proteins. In Handbook of Protein Folding (Buchner J & Kiefhaber T, eds), pp. 271–353. Wiley-VCH, Weinheim, Germany.
42 Uversky VN (2013) Unusual biophysics of intrinsically disordered proteins. *Biochim Biophys Acta* **1834**, 932–951.
43 Dunker AK & Obradovic Z (2001) The protein trinity – linking function and disorder. *Nat Biotechnol* **19**, 805–806.
44 Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z & Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* **323**, 573–584.
45 Dyson HJ & Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197–208.
46 Tompa P (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett* **579**, 3346–3354.
47 Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN & Dunker AK (2007) Intrinsic disorder and functional proteomics. *Biophys J* **92**, 1439–1456.
48 Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z & Uversky VN (2007) 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* **6**, 1899–1916.
49 Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z & Uversky VN (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J Proteome Res* **6**, 1917–1932.
50 Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN & Obradovic Z (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res* **6**, 1882–1898.
51 Uversky VN, Oldfield CJ & Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* **37**, 215–246.
52 Vacic V, Markwick PR, Oldfield CJ, Zhao X, Haynes C, Uversky VN & Iakoucheva LM (2012) Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. *PLoS Comput Biol* **8**, e1002709.
53 Xue B, Dunker AK & Uversky VN (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn* **30**, 137–149.
54 Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, Wang K, Hu G, Uversky VN & Kurgan L (2015) Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* **72**, 137–151.
55 Goh GK, Dunker AK & Uversky VN (2008) A comparative analysis of viral matrix proteins using disorder predictors. *Virol J* **5**, 126.
56 Goh GK, Dunker AK & Uversky VN (2009) Protein intrinsic disorder and influenza virulence: the 1918 H1N1 and H5N1 viruses. *Virol J* **6**, 69.
57 Goh GK, Dunker AK & Uversky VN (2012) Understanding viral transmission behavior via protein intrinsic disorder prediction: coronaviruses. *J Pathog* **2012**, 738590.
58 Goh GK, Dunker AK & Uversky V (2013) Prediction of intrinsic disorder in MERS-CoV/HCoV-EMC supports a high oral–fecal transmission. *PLoS Curr* **5**, ecurrents.outbreaks.22254b58675cdebc256dbe3c5aa6498b.
59 Fan X, Xue B, Dolan PT, LaCount DJ, Kurgan L & Uversky VN (2014) The intrinsic disorder status of the human hepatitis C virus proteome. *Mol BioSyst* **10**, 1345–1363.
60 Xue B, Mizianty MJ, Kurgan L & Uversky VN (2012) Protein intrinsic disorder as a flexible armor and a weapon of HIV-1. *Cell Mol Life Sci* **69**, 1211–1259.
61 Uversky VN, Roman A, Oldfield CJ & Dunker AK (2006) Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. *J Proteome Res* **5**, 1829–1842.
62 Xue B, Ganti K, Rabionet A, Banks L & Uversky VN (2014) Disordered interactome of human papillomavirus. *Curr Pharm Des* **20**, 1274–1292.

63 de Laureto PP, Tosatto L, Frare E, Marin O, Uversky VN & Fontana A (2006) Conformational properties of the SDS-bound state of α-synuclein probed by limited proteolysis: unexpected rigidity of the acidic C-terminal tail. *Biochemistry* **45**, 11523–11531.
64 Fontana A, de Laureto PP, Spolaore B, Frare E, Picotti P & Zambonin M (2004) Probing protein structure by limited proteolysis. *Acta Biochim Pol* **51**, 299–321.
65 Fontana A, Fassina G, Vita C, Dalzoppo D, Zamai M & Zambonin M (1986) Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* **25**, 1847–1851.
66 Fontana A, Polverino de Laureto P, De Filippis V, Scaramella E & Zambonin M (1997) Probing the partly folded states of proteins by limited proteolysis. *Fold Des* **2**, R17–R26.
67 Iakoucheva LM, Kimzey AL, Masselon CD, Bruce JE, Garner EC, Brown CJ, Dunker AK, Smith RD & Ackerman EJ (2001) Identification of intrinsic order and disorder in the DNA repair protein XPA. *Protein Sci* **10**, 560–571.
68 Polverino de Laureto P, De Filippis V, Di Bello M, Zambonin M & Fontana A (1995) Probing the molten globule state of α-lactalbumin by limited proteolysis. *Biochemistry* **34**, 12596–12604.
69 Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ & Dunker AK (2002) Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol* **55**, 104–110.
70 Borcherds W, Kashtanov S, Wu H & Daughdrill GW (2013) Structural divergence is more extensive than sequence divergence for a family of intrinsically disordered proteins. *Proteins* **81**, 1686–1698.
71 Xue B, Brown CJ, Dunker AK & Uversky VN (2013) Intrinsically disordered regions of p53 family are highly diversified in evolution. *Biochim Biophys Acta* **1834**, 725–738.
72 Uversky VN (2013) A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci* **22**, 693–724.
73 Gitlin L, Hagai T, LaBarbera A, Solovey M & Andino R (2014) Rapid evolution of virus sequences in intrinsically disordered protein regions. *PLoS Pathog* **10**, e1004529.
74 Campen A, Williams RM, Brown CJ, Meng J, Uversky VN & Dunker AK (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett* **15**, 956–963.
75 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN & Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242.
76 Vacic V, Uversky V, Dunker AK & Lonardi S (2007) Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* **8**, 211.
77 Peng Z, Xue B, Kurgan L & Uversky VN (2013) Resilience of death: intrinsic disorder in proteins involved in the programmed cell death. *Cell Death Differ* **20**, 1257–1267.
78 Peng Z, Oldfield C, Xue B, Mizianty M, Dunker AK, Kurgan L & Uversky V (2014) A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. *Cell Mol Life Sci* **71**, 1477–1504.
79 Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN *et al.* (2007) DisProt: the Database of Disordered Proteins. *Nucleic Acids Res* **35**, D786–D793.
80 Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM & Obradovic Z (2002) Intrinsic disorder and protein function. *Biochemistry* **41**, 6573–6582.
81 Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN & Dunker AK (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genom* **9**, S1.
82 Uversky VN, Oldfield CJ & Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* **18**, 343–384.
83 Hsu WL, Oldfield CJ, Xue B, Meng J, Huang F, Romero P, Uversky VN & Dunker AK (2013) Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci* **22**, 258–273.
84 Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z & Dunker AK (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* **32**, 1037–1049.
85 Pejaver V, Hsu WL, Xin F, Dunker AK, Uversky VN & Radivojac P (2014) The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci* **23**, 1077–1093.
86 Perera R & Kuhn RJ (2008) Structural proteomics of dengue virus. *Curr Opin Microbiol* **11**, 369–377.
87 Henchal EA & Putnak JR (1990) The dengue viruses. *Clin Microbiol Rev* **3**, 376–396.
88 Markoff L, Falgout B & Chang A (1997) A conserved internal hydrophobic domain mediates the stable membrane integration of the dengue virus capsid protein. *Virology* **233**, 105–117.
89 Nemesio H, Palomares-Jerez F & Villalain J (2011) The membrane-active regions of the dengue virus proteins C and E. *Biochim Biophys Acta* **1808**, 2390–2402.
90 Iakoucheva LM, Kimzey AL, Masselon CD, Smith RD, Dunker AK & Ackerman EJ (2001) Aberrant

mobility phenomena of the DNA repair protein XPA. *Protein Sci* **10**, 1353–1362.
91 Uversky VN & Dunker AK (2012) Multiparametric analysis of intrinsically disordered proteins: looking at intrinsic disorder through compound eyes. *Anal Chem* **84**, 2096–2104.
92 Faustino AF, Guerra GM, Huber RG, Hollmann A, Domingues MM, Barbosa GM, Enguita FJ, Bond PJ, Castanho MA, Poian AT *et al.* (2014) Understanding dengue virus capsid protein disordered N-terminus and pep14-23-based inhibition. *ACS Chem Biol* **10**, 517–526.
93 Jones CT, Ma L, Burgner JW, Groesch TD, Post CB & Kuhn RJ (2003) Flavivirus capsid is a dimeric α-helical protein. *J Virol* **77**, 7143–7149.
94 Ma L, Jones CT, Groesch TD, Kuhn RJ & Post CB (2004) Solution structure of dengue virus capsid protein reveals another fold. *Proc Natl Acad Sci USA* **101**, 3414–3419.
95 Gunasekaran K, Tsai CJ & Nussinov R (2004) Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *J Mol Biol* **341**, 1327–1341.
96 Guirakhoo F, Bolin RA & Roehrig JT (1992) The Murray Valley encephalitis virus prM protein confers acid resistance to virus particles and alters the expression of epitopes within the R2 domain of E glycoprotein. *Virology* **191**, 921–931.
97 Zhang Y, Corver J, Chipman PR, Zhang W, Pletnev SV, Sedlak D, Baker TS, Strauss JH, Kuhn RJ & Rossmann MG (2003) Structures of immature flavivirus particles. *EMBO J* **22**, 2604–2613.
98 Li L, Lok SM, Yu IM, Zhang Y, Kuhn RJ, Chen J & Rossmann MG (2008) The flavivirus precursor membrane–envelope protein complex: structure and maturation. *Science* **319**, 1830–1834.
99 Modis Y, Ogata S, Clements D & Harrison SC (2003) A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc Natl Acad Sci USA* **100**, 6986–6991.
100 Modis Y, Ogata S, Clements D & Harrison SC (2004) Structure of the dengue virus envelope protein after membrane fusion. *Nature* **427**, 313–319.
101 Zhang Y, Zhang W, Ogata S, Clements D, Strauss JH, Baker TS, Kuhn RJ & Rossmann MG (2004) Conformational changes of the flavivirus E glycoprotein. *Structure* **12**, 1607–1618.
102 Liu H, Chiou SS & Chen WJ (2004) Differential binding efficiency between the envelope protein of Japanese encephalitis virus variants and heparan sulfate on the cell surface. *J Med Virol* **72**, 618–624.
103 Jindadamrongwech S, Thepparit C & Smith DR (2004) Identification of GRP 78 (BiP) as a liver cell expressed receptor element for dengue virus serotype 2. *Arch Virol* **149**, 915–927.
104 Krishnan MN, Sukumaran B, Pal U, Agaisse H, Murray JL, Hodge TW & Fikrig E (2007) Rab 5 is required for the cellular entry of dengue and West Nile viruses. *J Virol* **81**, 4881–4885.
105 Ren J, Ding T, Zhang W, Song J & Ma W (2007) Does Japanese encephalitis virus share the same cellular receptor with other mosquito-borne flaviviruses on the C6/36 mosquito cells? *Virol J* **4**, 83.
106 Navarro-Sanchez E, Altmeyer R, Amara A, Schwartz O, Fieschi F, Virelizier JL, Arenzana-Seisdedos F & Despres P (2003) Dendritic-cell-specific ICAM3-grabbing non-integrin is essential for the productive infection of human dendritic cells by mosquito-cell-derived dengue viruses. *EMBO Rep* **4**, 723–728.
107 Tassaneetrithep B, Burgess TH, Granelli-Piperno A, Trumpfheller C, Finke J, Sun W, Eller MA, Pattanapanyasat K, Sarasombath S, Birx DL *et al.* (2003) DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. *J Exp Med* **197**, 823–829.
108 Miller JL, de Wet BJ, Martinez-Pomares L, Radcliffe CM, Dwek RA, Rudd PM & Gordon S (2008) The mannose receptor mediates dengue virus infection of macrophages. *PLoS Pathog* **4**, e17.
109 Mackow E, Makino Y, Zhao BT, Zhang YM, Markoff L, Buckler-White A, Guiler M, Chanock R & Lai CJ (1987) The nucleotide sequence of dengue type 4 virus: analysis of genes coding for nonstructural proteins. *Virology* **159**, 217–228.
110 Wright PJ, Cauchi MR & Ng ML (1989) Definition of the carboxy termini of the three glycoproteins specified by dengue virus type 2. *Virology* **171**, 61–67.
111 Winkler G, Maxwell SE, Ruemmler C & Stollar V (1989) Newly synthesized dengue-2 virus nonstructural protein NS1 is a soluble protein but becomes partially hydrophobic and membrane-associated after dimerization. *Virology* **171**, 302–305.
112 Jacobs MG, Robinson PJ, Bletchly C, Mackenzie JM & Young PR (2000) Dengue virus nonstructural protein 1 is expressed in a glycosyl-phosphatidylinositol-linked form that is capable of signal transduction. *FASEB J* **14**, 1603–1610.
113 Mackenzie JM, Jones MK & Young PR (1996) Immunolocalization of the dengue virus nonstructural glycoprotein NS1 suggests a role in viral RNA replication. *Virology* **220**, 232–240.
114 Winkler G, Randolph VB, Cleaves GR, Ryan TE & Stollar V (1988) Evidence that the mature form of the flavivirus nonstructural protein NS1 is a dimer. *Virology* **162**, 187–196.
115 Parrish CR, Woo WS & Wright PJ (1991) Expression of the NS1 gene of dengue virus type 2 using vaccinia virus. Dimerisation of the NS1 glycoprotein. *Arch Virol* **117**, 279–286.

116 Muylaert IR, Chambers TJ, Galler R & Rice CM (1996) Mutagenesis of the N-linked glycosylation sites of the yellow fever virus NS1 protein: effects on virus replication and mouse neurovirulence. *Virology* **222**, 159–168.
117 Lindenbach BD & Rice CM (1999) Genetic interaction of flavivirus nonstructural proteins NS1 and NS4A as a determinant of replicase function. *J Virol* **73**, 4611–4621.
118 Gibson CA, Schlesinger JJ & Barrett AD (1988) Prospects for a virus non-structural protein as a subunit vaccine. *Vaccine* **6**, 7–9.
119 Wallis TP, Huang CY, Nimkar SB, Young PR & Gorman JJ (2004) Determination of the disulfide bond arrangement of dengue virus NS1 protein. *J Biol Chem* **279**, 20729–20741.
120 Edeling MA, Diamond MS & Fremont DH (2014) Structural basis of flavivirus NS1 assembly and antibody recognition. *Proc Natl Acad Sci USA* **111**, 4285–4290.
121 Muller DA & Young PR (2013) The flavivirus NS1 protein: molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker. *Antiviral Res* **98**, 192–208.
122 Gutsche I, Coulibaly F, Voss JE, Salmon J, d'Alayer J, Ermonval M, Larquet E, Charneau P, Krey T, Megret F *et al.* (2011) Secreted dengue virus nonstructural protein NS1 is an atypical barrel-shaped high-density lipoprotein. *Proc Natl Acad Sci USA* **108**, 8003–8008.
123 Muller DA, Landsberg MJ, Bletchly C, Rothnagel R, Waddington L, Hankamer B & Young PR (2012) Structure of the dengue virus glycoprotein non-structural protein 1 by electron microscopy and single-particle analysis. *J Gen Virol* **93**, 771–779.
124 Young PR, Hilditch PA, Bletchly C & Halloran W (2000) An antigen capture enzyme-linked immunosorbent assay reveals high levels of the dengue virus protein NS1 in the sera of infected patients. *J Clin Microbiol* **38**, 1053–1057.
125 Alcon S, Talarmin A, Debruyne M, Falconar A, Deubel V & Flamand M (2002) Enzyme-linked immunosorbent assay specific to Dengue virus type 1 nonstructural protein NS1 reveals circulation of the antigen in the blood during the acute phase of disease in patients experiencing primary or secondary infections. *J Clin Microbiol* **40**, 376–381.
126 Libraty DH, Young PR, Pickering D, Endy TP, Kalayanarooj S, Green S, Vaughn DW, Nisalak A, Ennis FA & Rothman AL (2002) High circulating levels of the dengue virus nonstructural protein NS1 early in dengue illness correlate with the development of dengue hemorrhagic fever. *J Infect Dis* **186**, 1165–1168.
127 Chambers TJ, McCourt DW & Rice CM (1989) Yellow fever virus proteins NS2A, NS2B, and NS4B: identification and partial N-terminal amino acid sequence analysis. *Virology* **169**, 100–109.
128 Xie X, Zou J, Puttikhunt C, Yuan Z & Shi PY (2014) Two distinct sets of NS2A molecules are responsible for dengue virus RNA synthesis and virion assembly. *J Virol*. doi:10.1128/JVI.02882-14.
129 Xie X, Gayen S, Kang C, Yuan Z & Shi PY (2013) Membrane topology and function of dengue virus NS2A protein. *J Virol* **87**, 4609–4622.
130 Huang Q, Chen AS, Li Q & Kang C (2011) Expression, purification, and initial structural characterization of nonstructural protein 2B, an integral membrane protein of Dengue-2 virus, in detergent micelles. *Protein Expr Purif* **80**, 169–175.
131 Erbel P, Schiering N, D'Arcy A, Renatus M, Kroemer M, Lim SP, Yin Z, Keller TH, Vasudevan SG & Hommel U (2006) Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nat Struct Mol Biol* **13**, 372–373.
132 Luo D, Xu T, Hunke C, Gruber G, Vasudevan SG & Lescar J (2008) Crystal structure of the NS3 protease-helicase from dengue virus. *J Virol* **82**, 173–183.
133 Kapoor M, Zhang L, Ramachandra M, Kusukawa J, Ebner KE & Padmanabhan R (1995) Association between NS3 and NS5 proteins of dengue virus type 2 in the putative RNA replicase is linked to differential phosphorylation of NS5. *J Biol Chem* **270**, 19100–19106.
134 Bera AK, Kuhn RJ & Smith JL (2007) Functional characterization of *cis* and *trans* activity of the flavivirus NS2B-NS3 protease. *J Biol Chem* **282**, 12883–12892.
135 Miller S, Kastner S, Krijnse-Locker J, Buhler S & Bartenschlager R (2007) The non-structural protein 4A of dengue virus is an integral membrane protein inducing membrane alterations in a 2K-regulated manner. *J Biol Chem* **282**, 8873–8882.
136 Roosendaal J, Westaway EG, Khromykh A & Mackenzie JM (2006) Regulated cleavages at the West Nile virus NS4A-2K-NS4B junctions play a major role in rearranging cytoplasmic membranes and Golgi trafficking of the NS4A protein. *J Virol* **80**, 4623–4632.
137 Umareddy I, Chao A, Sampath A, Gu F & Vasudevan SG (2006) Dengue virus NS4B interacts with NS3 and dissociates it from single-stranded RNA. *J Gen Virol* **87**, 2605–2614.
138 Stern O, Hung YF, Valdau O, Yaffe Y, Harris E, Hoffmann S, Willbold D & Sklan EH (2013) An N-terminal amphipathic helix in dengue virus nonstructural protein 4A mediates oligomerization and is essential for replication. *J Virol* **87**, 4080–4085.
139 Hung YF, Valdau O, Schunke S, Stern O, Koenig BW, Willbold D & Hoffmann S (2014) Recombinant

production of the amino terminal cytoplasmic region of dengue virus non-structural protein 4A for structural studies. *PLoS One* **9**, e86482.
140 Zou J, Xie X, le Lee T, Chandrasekaran R, Reynaud A, Yap L, Wang QY, Dong H, Kang C, Yuan Z *et al.* (2014) Dimerization of flavivirus NS4B protein. *J Virol* **88**, 3379–3391.
141 Zhao Y, Soh TS, Zheng J, Chan KW, Phoo WW, Lee CC, Tay MY, Swaminathan K, Cornvik TC, Lim SP, *et al.* (2015) A crystal structure of the Dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog* **11**, e1004682.
142 Egloff MP, Benarroch D, Selisko B, Romette JL & Canard B (2002) An RNA cap (nucleoside-2′-O-)-methyltransferase in the flavivirus RNA polymerase xml: crystal structure and functional characterization. *EMBO J* **21**, 2757–2768.
143 Yap TL, Xu T, Chen YL, Malet H, Egloff MP, Canard B, Vasudevan SG & Lescar J (2007) Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-Angstrom resolution. *J Virol* **81**, 4753–4765.
144 Lim SP, Koh JH, Seh CC, Liew CW, Davidson AD, Chua LS, Chandrasekaran R, Cornvik TC, Shi PY & Lescar J (2013) A crystal structure of the dengue virus non-structural protein 5 (NS5) polymerase delineates interdomain amino acid residues that enhance its thermostability and de novo initiation activities. *J Biol Chem* **288**, 31105–31114.
145 UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **42**, D191–D198.
146 Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, Bhatt S, Katzelnick L, Howes RE, Battle KE *et al.* (2014) Global spread of dengue virus types: mapping the 70 year history. *Trends Microbiol* **22**, 138–146.
147 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.
148 Dinkel H, Van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, Speck T, Krüger D, Grebnev G, Kubań M *et al.* (2014) The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res* **42**, D259–D266.
149 Mizianty MJ, Stach W, Chen K, Kedarisetti KD, Disfani FM & Kurgan L (2010) Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. *Bioinformatics* **26**, i489–i496.
150 Disfani FM, Hsu W-L, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN & Kurgan L (2012) MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* **28**, i75–i83.
151 Dayhoff MO, Schwartz RM & Orcutt BC (1978) A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure* (Dayhoff MO, ed.), pp. 345–352. National Biomedical Research Foundation, Washington, DC.
152 Gould CM, Diella F, Via A, Puntervoll P, Gemünd C, Chabanis-Davidson S, Michael S, Sayadi A, Bryne JC, Chica C *et al.* (2010) ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res* **38**, D167–D180.
153 Dosztanyi Z, Csizmok V, Tompa P & Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434.
154 McGuffin LJ (2008) Intrinsic disorder prediction from the analysis of multiple protein fold recognition models. *Bioinformatics* **24**, 1798–1804.
155 Ward JJ, McGuffin LJ, Bryson K, Buxton BF & Jones DT (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138–2139.
156 McGuffin LJ, Bryson K & Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405.
157 Faraggi E, Xue B & Zhou Y (2009) Improving the prediction accuracy of residue solvent accessibility and real-value backbone torsion angles of proteins by guided-learning through a two-layer neural network. *Proteins* **74**, 847–856.
158 Schlessinger A, Yachdav G & Rost B (2006) PROFbval: predict flexible and rigid residues in proteins. *Bioinformatics* **22**, 891–893.
159 Peng ZL & Kurgan L (2012) Comprehensive comparative assessment of in-silico predictors of disordered regions. *Curr Protein Pept Sci* **13**, 6–18.
160 Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK & Uversky VN (2006) Analysis of molecular recognition features (MoRFs). *J Mol Biol* **362**, 1043–1059.
161 Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, Uversky VN & Dunker AK (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res* **6**, 2351–2366.
162 Obradovic Z, Peng K, Vucetic S, Radivojac P & Dunker AK (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* **61** (Suppl 7), 176–182.
163 Fan X & Kurgan L (2014) Accurate prediction of disorder in protein chains with a comprehensive and

empirically designed consensus. *J Biomol Struct Dyn* **32**, 448–464.

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

165 Xue B, Dunbrack RL, Williams RW, Dunker AK & Uversky VN (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* **1804**, 996–1010.