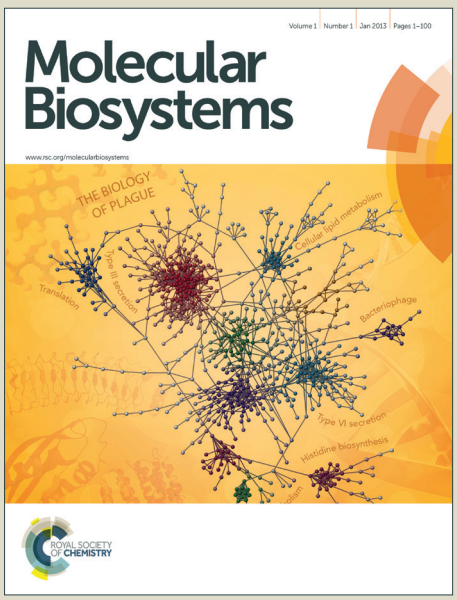
# Molecular BioSystems

Accepted Manuscript

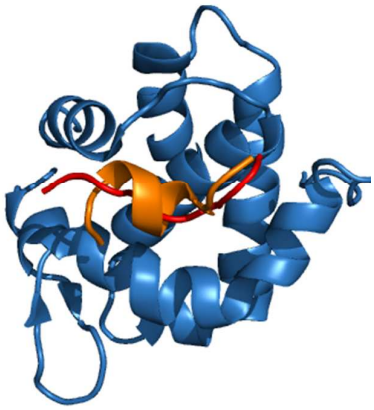







ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems



Local dynamics of host and viral peptide motifs suggests a different scenario for partner recognition. Host peptide motifs serve as molecular recognition elements, while viral motifs preserve structural heterogeneity and remain fuzzy when bound to the host.

# Fuzziness endows viral motif-mimicry

Norbert Duro, Marton Miskei and Monika Fuxreiter
*MTA-DE Momentum, Laboratory of Protein Dynamics, Department of Biochemistry and Molecular Biology, University of Debrecen, Hungary*



## Abstract

Motif-mimicry is exploited by viruses to interfere with host regulatory networks and has also been suggested as a prevalent strategy for eukaryotic and prokaryotic pathogens. Using the same peptide motif however does not guaranty more effective interactions with the host. Motif-mediated interactions require flexible or disordered environment, which structural and dynamic features could differ between the competing host and viral proteins. Using the eukaryotic linear motif (ELM) database we analyzed protein regions, which embed the eukaryotic and viral motifs, including human and human virus ELMs with common target sites. We found that although eukaryotic motifs are associated with a lack of structure, they are more stable than their flanking regions and can serve as molecular recognition elements. In contrast, eukaryotic viral motifs are often located in more ordered regions, but have increased local flexibility or disorder compared to their embedding environment. Most viral ELMs are devoid of stable binding elements and remain fuzzy after binding. Fuzziness reduces the entropic cost of binding and imparts versatile interaction modes to increase binding promiscuity and to compete with multiple host peptides. Fuzzy interactions confer further functional benefits such as combinatorial usage of motifs, and fine-tuning affinity via post-translational modifications.

## Introduction

Functional diversity and adaptability of viruses is intriguing considering their small genome size. Viruses were proposed to have unique biophysical properties to cope with these evolutionary constraints: they possess loosely packed cores and a high

propensity of non-regular secondary structure elements [1]. In accord, viral protein segments often a lack stable tertiary structure, i.e. they are intrinsic disordered (ID) in the absence of a partner [2]. This architecture can tolerate high mutation rates without the loss of the structural framework, which is required for function [3]. Viral proteins were also experimentally demonstrated to maintain their intrinsic disordered state even upon interacting with other components of the replication machinery [4]. This phenomenon is termed as fuzziness [5,6]. Nucleoprotein-phosphoprotein complexes in Measles [7], Hendra [8] and Nipah [9] viruses for example are characterized by a significant degree of conformational heterogeneity, which imparts dynamism on the recognition process. Fuzzy regions for example contribute to organization of the nucleocapsid [10] and also facilitate access to viral RNA [4].

Viruses invade their hosts via exploiting their regulatory networks. To this end, pathogens employ molecular mimicry and interact via short motifs, which resemble to those of the host system [11]. Viral motifs are versatile: they can interfere with signaling pathways; control target protein levels or perturb posttranslational modifications of host proteins. Viral motifs can also tune cooperativity of host proteins and allosterically modulate signaling outputs, as it has been recently characterized in detail in case of E1A oncoprotein in complex with CREB binding protein and retinoblastoma protein [12,13]. Motif-mimicry of host-peptide interactions is a powerful strategy, which is likely exploited by a large set of viral genomes (> 2000) [14] and is also employed by eukaryotic and prokaryotic pathogens [15]. Robustness of short linear motifs (SLiMs) were also proposed to contribute to adaptation and rapid evolution of viruses [11,14].

Motif-mimicry does not necessarily mean that all viral motifs compete with host motifs for the same site. For example, ubiquitin ligase recruitment [16] or masking destruction motifs [17] influences the host protein concentration or turnover, which might be regulated via different pathways of the host. In the following we focus on cases, when viral and host motifs target the same host site. Series of experimental evidence demonstrate that affinities of viral motifs are higher than any of the host motifs, e.g. in case of the PxxP (x could be any residue) SH3 binding motif of HIV Nef [18] or PTAG TSG101 binding motif of GAG-p6 [19]. As both the host and the pathogen utilizes the same motif pattern (e.g. specificity determining residues) higher affinities of viral motifs likely rely on factors, which are located outside the motif.

These could provide additional contacts with the target protein or modulate the structural or dynamic properties of the motifs.

Eukaryotic linear motif (ELM) database is an excellent resource to analyze these features and compare properties virus and host proteins [20]. Previously it has been shown that ELMs tend to be located in protein regions, which lack well-defined tertiary structures [21]. Plasticity of ELM environments can impart versatility on binding modes and also enhance adaptability. It is reasonable to assume that viral motifs employ a similar strategy. Local structural analysis on protein segments, which embed viral motifs has not been carried out yet.

Here we aimed to reveal molecular mechanisms of how viral motifs recognize their targets and those molecular factors, which enable them to outperform host motifs. The following scenarios were considered (Figure 1): *i)* protein regions embedding viral motifs have increased plasticity as compared to eukaryotic motifs *ii)* viral peptide motifs are flanked by longer disordered regions, which can hamper the access of the competing host proteins to the target site by steric exclusion *iii)* environments of viral motifs provide further binding sites, which anchor viral protein to the target and increase affinity. To investigate these mechanisms, the structural and dynamic features of the host and viral ELMs and their respective flanking regions were analyzed using two datasets (Figure S1): *i)* experimentally verified viral and eukaryotic motifs in the ELM database *ii)* human and human virus ELMs with common host target sites based on VirusMentha[22] and ELM interaction databases [23] (Table S1). Similarly to previous studies [24] we found that structural properties of viral motifs and their flanking regions vary in a wide range, but none of the proposed strategies were applicable. Environments of eukaryotic motifs were found to be more flexible and include longer ID regions than those of eukaryotic viral motifs. Flanking regions of eukaryotic ELMs contain more ID binding sites, which could establish buttressing interactions with the target proteins. Albeit these factors were comparable for human and human virus ELMs with common target sites, none of the scenarios could not explain the how viral motifs outcompete the host proteins. Local flexibility/disorder of the host and viral motifs relative to their embedding regions however showed a significant difference. While eukaryotic motifs are more stable and likely fold upon binding, viral motifs have increased flexibility or disorder as compared to their flanking segments. Static or dynamic disorder, i.e. fuzziness of viral ELMs and their neighboring residues is also observed upon interacting with the host

target. Fuzziness decreases the entropic cost of binding and improves affinity. Fuzzy virus-host interactions are amenable to fine-tuning via post-translational modifications or combinatorial usage of motifs, which increases binding versatility to interfere with host regulatory networks.

## Results and discussion

### *ELM containing viral proteins are more ordered than eukaryotic proteins.*

Increasing awareness of structural flexibility of viral proteins gives the misleading impression that global disorder is a key feature for virus adaptability, survival and function. Recently it was shown however, that disorder of viral proteomes extensively vary and it is not correlated to genome size [24]. Comparing viral proteins to mesophilic eukaryotic proteins, no significant differences in contact densities, i.e. tightness of packing were observed [1]. Furthermore, viral proteins contain less regular secondary structure elements and more coils, but are equipped with fewer disordered regions. Recent comparative analysis between animal virus and human proteins corroborated these results [14].

We focused on eukaryotic and virus proteins with experimentally verified ELMs (Table S2) and compared their preference for a well-defined tertiary structure versus disordered state. A significant difference between the average degree of disorder was observed (Figure 2A): ELM-containing eukaryotic proteins are on the borderline between globular and disordered proteins (median value of 0.42; the average degree of disorder of ID segments in the Disprot database v6.02[25] is 0.44 by IUpred program [26]), while ELM-containing viral proteins are mostly structured (median value of 0.24). The fraction of intrinsically disordered residues in eukaryotic ELM proteins is also significantly higher than in virus ELM proteins (Figure 2B). This trend is in accord with the higher flexibility of eukaryotic proteomes as compared to eukaryotic virus proteins including all proteins, not only those with ELMs (Figure S2).

Then we focused on human and human virus ELMs, which belonged to the same ELM classes (Table S3). Human ELM proteins are also more pliable than human virus ELM proteins and have higher fraction of ID residues (not shown).

Finally we compared to the subset of human and human virus ELMs, where common motif-binding domains have been experimentally demonstrated (Table S1, Table S4). Here no significant difference in the average degree of disorder or in the propensity of ID residues was seen between human and human virus ELM proteins. Comparing disorder of different viral families using all proteins, excludes that this is owing to a biased selection of ELMs from given families (Figure S3). These results were consistent with disorder predictions using the PONDR VSL1 algorithm [27] (Figure S4).

***Flanking regions of competing human and human virus ELMs have similar levels of disorder.***

ELMs in general tend to be located in disordered regions, which impart plasticity on linear motifs [21,28]. This feature is also exploited for ELM discovery [29]. In accord to previous observations, 20AA regions flanking eukaryotic ELMs have higher disorder scores than the average of the corresponding proteins (0.58 vs 0.42, Figure 3A). Similarly, regions embedding eukaryotic virus ELMs are also more flexible than other protein regions (0.37 versus 0.24). These results were also corroborated by comparing the flanking environments of eukaryotic and eukaryotic virus ELMs to randomly chosen segments of the corresponding proteins with the same length (Figure S5). In line with their higher level of disorder, ID regions neighboring eukaryotic ELMs are significantly longer than those flanking eukaryotic virus ELMs, likely owing to the smaller genome size (Figure 3B).

In contrast to the marked difference between the disorder properties of eukaryotic and eukaryotic virus ELM environments, regions embedding human and human virus ELMs are rather similar. Neither the degree of disorder nor the length of the motif-flanking ID segments exhibit a significant difference between human and human virus ELMs (Figure 3). These results indicate some sort of constraints on the disorder properties of the competing motifs. This conclusion was supported by comparing the flanking regions to randomly chosen segments of the corresponding proteins (Figure S6) as well as by using a different, PONDR VSL1 disorder prediction method (Figure S7).

***Additional disordered binding sites facilitate binding of both host and virus ELMs.***

ID regions in general are reminiscent of multi-partite interactions, which are mediated either by linear motifs or ID binding sites, which gain structure upon binding. The latter can be transient secondary structure conformations, which are biased for their bound state [30,31], or hydrophobic patches, which are stabilized by intermolecular interactions [32]. ID binding sites have a variety of names, their definitions and relationships are detailed elsewhere [33]. They in general have lower disorder scores than the embedding disordered protein regions [34] and can also mediate non-specific interactions by anchoring short linear motifs [32]. This suggests a plausible scenario to increase the affinity of motif binding via engaging additional binding regions for partner interactions. We compared the number of ID binding sites, which tend to fold upon binding in the 100 AA flanking regions of eukaryotic and eukaryotic virus ELMs using the Anchor program [32].

More ID binding sites were observed in protein segments neighboring eukaryotic ELMs than in regions flanking eukaryotic viral ELMs (Figure 4). This is however not a mere consequence of the higher disorder of the eukaryotic proteins or the longer ID regions flanking the motif (Figure S8). The difference between the number of ID binding sites in the respective 100 AA motif-flanking segments of eukaryotic and eukaryotic virus proteins is higher than the difference comparing other 100AA regions of the corresponding eukaryotic and viral proteins. We should also note that ID binding sites are enriched in the environments of both eukaryotic and eukaryotic virus ELMs as compared to the corresponding proteins in average.

Flanking regions of human and human virus ELMs comprise comparable number of ID binding sites, while this markedly deviates in the corresponding proteins. For this dataset we also repeated the ID binding site calculations using the Disopred3 algorithm [35]. Disopred3 employs a support vector machine approach, which predicts more irregular structural elements to bind than Anchor and excludes potential transient contacts. Owing to these reasons Disopred3 predicts more ID binding sites in 100 AA regions flanking human virus ELMs than those neighboring human ELMs (Figure S9). This indicates that buttressing contacts by ID binding sites may contribute to higher efficiency of human viral motifs. The DGR motif of the capsid protein VP1 of adeno-associated virus 2 for example competes with four NGR integrin binding sites of fibronectin (LIG_Integrin_isoDGR_1 motif in ELM; see

Table S1) for host entry. The 100 AA flanking region of the viral DGR motif contains more ID binding sites than the neighboring segments of the fibronectin NGR motifs of the same size, in accord with the critical role of the flanking residues in facilitating the adhesion of the cell surface and integrin receptor switching [36].

***Eukaryotic ELMs tend to fold, while viral motifs remain fuzzy***

Probing the three proposed scenarios (Figure 1) did not provide conclusive answer how viral motifs compete with their host counterparts. Hence we assumed that the structural or dynamic properties of the viral ELMs themselves are responsible for more efficient partner recognition. By comparing the disorder properties of the human and human virus ELMs with common host motif-binding domains we found that human virus ELMs are significantly more flexible or disordered than their human counterparts (Figure 5A) despite of the comparable disorder level of the embedding regions (Figure 3A). This suggests that human and human virus motifs have different characteristics as compared to their flanking protein regions. Thus we computed the difference between the degree of disorder of the human and human virus ELMs and their flanking regions ($\Delta ID = ID_{ELM} - ID_{Flanking20AA}$). We found that human ELMs have lower degree of disorder than their environment, while human virus ELMs have elevated level of disorder as compared to their embedding environments (Figure 5B). Pair-wise differences of disorder scores between human and human virus ELMs and their 20AA flanking regions underscore this observation ($p = 4.7 \times 10^{-5}$ in Wilcoxon test, $p = 1.5 \times 10^{-4}$ in Kolmogorov-Smirnov test). Taken together human ELMs appear to be more stable, while human virus ELMs are more flexible or dynamic than their 20 AA flanking regions.

Decreased disorder of human ELMs as compared to their flanking regions indicates that they can serve as preformed [30] or molecular recognition elements [31], which exhibit transient secondary structures biased for their bound conformation. These binding sites could fold upon interacting with their partner. In accord, 68% of human ELMs are associated with non-regular secondary structure elements (NORS) [37], while only 37% of human virus ELMs are located in NORS. Along these lines, 54% of the disordered human and 28% of the disordered viral ELMs are predicted to fold upon binding [32] irrespective of their secondary structure preferences. Taken

together, human virus ELMs follow a different strategy for partner recognition than the corresponding host motifs, which is driven by increased local flexibility or disorder. Corroborating this observation 87 % of disordered viral motifs are flanked by short (at least 5AA) fuzzy regions, which remain dynamic even when bound to their partner.

***Experimental evidences for fuzziness in virus-host interactions***

Despite the experimental difficulties in characterizing fuzzy protein regions [38], growing evidence supports the existence of the disordered state, i.e. fuzziness in viral complexes [2]. Two examples how fuzziness contributes to motif mimicry are detailed below.

Nonstructural protein 5A (NS5A) of the hepatitis C virus (HCV) has two PxxP motifs (PP2.1 and PP2.2), out of which PP2.2 motif can interact with a variety of SH3 domains of the Src kinase family. NMR results reveal two additional PxxP motifs, serving as low-affinity sites for noncanonical SH3 binding [39]. All NS5A binding motifs compete for the same pocket on the SH3 domain via mutually exclusive binding modes. Although the noncanonical sites are embedded in transiently structured α-helical regions, the population of helical conformations decreases upon binding. The heteronuclear Overhauser effect (hetNOE) values of NS5A residues at the binding interface also decrease upon interacting with SH3 domains, indicating increased conformational flexibility and more heterogeneous structural ensemble. The fuzzy nature of the complex provides a favorable entropic contribution to the binding free energy and results in 2-3 fold increase in $K_d$ values.

Hepatitis B virus (HBV) preS1 domain contains multiple motifs, which resemble the recognition sites of the cell-surface receptor γ2-adaptin. preS1 does not exhibit a preformed secondary structure, and interactions with γ2- adaptin EAR domain does not induce structure-formation [40]. NOE enhancements show that the binding motifs, which are flanked by proline residues, have distinct dynamic character and remain fuzzy in the context of the binding partner. Deletion experiments demonstrate that flanking residues also contribute to the binding affinity of preS1 to γ2-adaptin EAR domain. In preS1 fuzziness enables combinatorial usage of the

motifs, which increases functional versatility of the viral protein.

A series of further experimental evidence supports that presence of fuzziness in paramyxovirus complexes [3], upon interactions of $N_{TAIL}$ of nucleoprotein with phosphoprotein in Measles, Nipah and Hendra viruses.

***Static fuzziness and binding promiscuity of human virus ELMs***

Owing to genetic compaction of viruses, the encoded proteins should be involved in multiple functions, e.g. capable to establish interactions with different partners. This could be facilitated by the underlying conformational heterogeneity of the viral proteins, which allows the ensemble to shift between different conformational states upon responding to different signals. Our results indicate that viral motifs have increased local flexibility or disorder relative to their 20 AA flanking regions. We discussed examples for dynamic fuzziness, when the protein interconverts between multiple conformations in the bound state [5,6]. In case of static fuzziness however, the ID protein folds upon interacting with the partner, but adopts alternative conformations [5,6]. 17% of the human virus ELMs are located in ordered regions, which were analyzed for static fuzziness. To this end, we collected structures of host-virus complexes of these ELMs and different secondary structure conformations upon targeting the same host protein were identified (Table S5). The adenovirus E1A protein interacts with retinoblastoma protein in multiple locations via short peptides, which despite of the same sequence adopt different secondary structures in the complex (Figure 6)[41]. The E2F transcription factor use the same contacts to inactivate the viral oncoprotein. Another example is the YNSTFF motif (MOD_N-GLC_1) of the SARS coronavirus spike receptor-binding domain (SRBD), which interacts with its receptor using α-helix, turn or β-bridge (PDB codes: 2ajf, 3scj) [42]. In addition, the FNATKF motif of SARS SRBD adopts 3 different conformations upon binding to the receptor, increasing structural and functional versatility of the virus-host interaction. This illustrates that static fuzziness can also increase binding promiscuity and enable viral ELMs to compete with various human motifs (Table S1) for their respective targets.

***Implications of viral motif fuzziness for antiviral strategies***

Disrupting virus-host interactions is an exciting challenge to develop antiviral therapeutics. As motif mimicry is exploited by numerous viral genomes, a possible strategy is to design a motif-mimetic, which targets the same binding site of the host and outcompetes the viral motif [43].

Evolutionary plasticity of viral ELMs however, could be a bottleneck of this approach. Another disadvantage is the low selectivity/binding promiscuity of the motif, which can lead to side-effects via binding to undesired host proteins [44]. One can attempt to block the viral motif directly by specific antibodies and overcome the problem of resistance [45]. Our results indicate that fuzziness or increased local flexibility of viral motifs is critical for their interactions. Consequently, rigidifying viral motifs should hamper their binding to the host proteins, and host peptides become more effective upon targeting the same site. This offers an alternative strategy to developing motif-mimetic drugs.

Experimental and bioinformatics results indicate that disorder properties of a given site could be modulated by a longer protein sequence (approx. 100 AA) [26,32]. Binding a small-molecule compound into this region thus could decrease the local dynamics of the viral motif, which stabilize structural elements and reduce conformational heterogeneity. Targeting the more ordered flanking regions of viral motifs could be more feasible than the dynamic viral ELM itself, yet could impact its function. Rigidifying viral motifs and their proximal residues might also impair capsid formation and disfavor replication. As embedding regions of viral ELMs are more structured, they are likely less mutation-prone than the motif itself, and thereby could be more suitable drug targets.

**Conclusion**

Viral motif-mimicry is a successful strategy to invade the host and reprogram its regulatory networks. While eukaryotic ELMs serve as molecular recognition elements for host-peptide interactions, viral ELMs have increased local flexibility as compared to their environments. Experimental evidence supports that fuzziness of viral ELMs improves binding affinity and promiscuity. Perturbing dynamic properties

of viral ELMs via small-molecule binding at the flanking regions could offer an alternative approach to direct motif-based drug development.

We must note that viral proteins use versatile mechanisms to reprogram the host system by having alternative set of motifs than host proteins or alter turnover owing to presence or absence of degradation signals. This analysis focused only a subset of viral motifs, which use the same functional classes as host peptide motifs.

## Experimental procedures

*Eukaryotic and virus ELM datasets*:

2585 ELMs were downloaded from the ELM database (http://elm.eu.org) [20]. We ignored all cases were the evidence class was 'predicted' or ELMs with 'no further instance evidence', discarding in total 556 ELMs. All the remaining 2029 ELMs were verified with experimental evidence. This dataset contained 1801 eukaryotic, 220 virus and 8 prokaryotic ELMs. Owing to their small number and the absence of the corresponding prokaryotic viral motifs the prokaryotic ELMs were not analyzed. Out of the 1801 eukaryotic ELMs, 1148 were human ELMs and out of the 220 eukaryotic viral ELMs 160 were human virus ELMs. As one protein can contain multiple ELMs, the analyzed ELMs belonged to 1182 eukaryotic, 698 human, 123 eukaryotic virus, 82 human virus proteins.

Distribution of eukaryotic and eukaryotic virus ELMs amongst different ELM classes is displayed in Table S2. Then we paired the human and human virus ELMs based on their ELM classes. First all possible pairs from human and human virus ELMs with identical ELM classes (e.g. integrin_isoDGR_1) were created, resulting in 1159 pairs (from 31 ELM classes, Table S3). Obviously, the same ELM type does not guarantee that human and human virus ELMs have identical host domain targets. Hence using two resources: the iELM database [23] and the VirusMentha database [22] we collected experimental evidences to filter out those pairs, where the human and human virus motifs target the same host protein domain. The selection process is displayed in Figure S1. iELM provided 32 examples, with identical motif-binding domains for human and human virus ELMs. The VirusMentha database provided 202 cases, where the human virus ELM and the human ELM targeted the same protein. The VirusMentha database however does not provide information on the target binding domains. Therefore, the human-human virus ELM pair examples, which were

filtered based on the VirusMentha database, were cross-validated using the iELM database interaction domains. Here domains, which bind a given motif are collected. Cross-validation could assign a motif-binding domain to 90 out of those 202 human-human virus ELM pairs, which were resulted by the VirusMentha database. 14 pairs out of 90 hits overlap with those examples, which were selected from the iELM interactions database. Thus 76 human-human virus ELM pair examples were resulted by the VirusMentha database with interacting motif binding domains, which were validated against the iELM interacting domains. In total, the two databases resulted in 108 (32+76) human and human virus ELM pairs, with identical host binding domains. As in the VirusMentha database we applied the binary interaction filter, and in the 32 iELM cases the experimental binding affinities are provided, we excluded that these 108 motif-domain associations are resulted by indirect interactions. Distribution of the 108 human and human virus pairs amongst different ELM classes are displayed in Table S4.

*Analysis of disorder*:

Preference for a well-defined structure or the lack of a stable structure was estimated based on low-resolution pair-wise potentials by the IUPred program [26]. The results were corroborated using the PONDR VSL1 and VSL2 neural network algorithms [46], which provide very similar results. Hence we only display data computed by VSL1 in the supplementary material. Average degree of disorder was computed by averaging the disorder score of all residues. In the IUPred 'long' algorithm we applied a 0.44 threshold to discriminate between ordered and disordered residues. Using this binary classification, we determined the propensity of disordered residues ($N_{ID}/N_{AA}$, where $N_{ID}$ is the number of disordered residues, and $N_{AA}$ is the length of the sequence). The 0.44 limit is based on the average disorder score of the disordered residues in the 6.02 version of the Disprot database [25]. To analyze local disorder properties, 20AA ELM-flanking regions were considered to make the results comparable to the previous analysis on eukaryotic ELMs [21].

*Analysis of disordered binding sites and fuzzy regions*

Intrinsically disordered binding regions, which fold upon binding were computed using the Anchor program [32] and were defined as continuous stretches of at least 5 residues with Anchor scores > 0.5. Dynamic fuzzy regions do not adopt a well-

defined structure upon interacting with their partners. Fuzzy regions were defined as disordered regions, which do not overlap with intrinsically disordered binding sites. Fuzzy regions were identified applying two conditions: IUPred score > 0.44 (to define an ID region) and Anchor score < 0.5 (to exclude formation of a stable structure).

Non-regular secondary structure (NORS) elements were identified using the PredictProtein server[47].

*Statistical analysis*:

All statistics were performed by the R program (http://www.r-project.org/). Both Wilcoxon rank sum test (Mann-Whitney) and Wilcoxon signed rank test were computed.

**Acknowledgement**

The support of the Hungarian Science Fund program (OTKA NN 106562) and the Momentum program (LP2012-41) of the Hungarian Academy of Sciences is gratefully acknowledged (M.F.). We thank to Sonia Longhi for fruitful discussions.

## References


1. Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS: Do viral proteins possess unique biophysical features? *Trends Biochem Sci* 2009,34(2):53-59.
2. Xue B, Blocquel D, Habchi J, Uversky AV, Kurgan L, Uversky VN, Longhi S: Structural disorder in viral proteins. *Chem Rev* 2014,114(13):6880-6911.
3. Habchi J, Longhi S: Structural disorder within paramyxovirus nucleoproteins and phosphoproteins. *Mol Biosyst* 2012,8(1):69-81.
4. Jensen MR, Communie G, Ribeiro EAJ, Martinez N, Desfosses A, Salmon L, Mollica L, Gabel F, Jamin M, Longhi S, Ruigrok RW, Blackledge M: Intrinsic disorder in measles virus nucleocapsids. *Proc Natl Acad Sci U S A* 2011.
5. Tompa P, Fuxreiter M: Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* 2008,33(1):2-8.
6. Fuxreiter M: Fuzziness: linking regulation to protein dynamics. *Mol Biosyst* 2012,8(1):168-177.
7. D'Urzo A, Konijnenberg A, Rossetti G, Habchi J, Li J, Carloni P, Sobott F, Longhi S, Grandori R: Molecular Basis for Structural Heterogeneity of an Intrinsically Disordered Protein Bound to a Partner by Combined ESI-IM-MS and Modeling. *Journal of the American Society for Mass Spectrometry* 2015,26(3):472-481.
8. Communie G, Habchi J, Yabukarski F, Blocquel D, Schneider R, Tarbouriech N, Papageorgiou N, Ruigrok RW, Jamin M, Jensen MR, Longhi S, Blackledge M: Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus. *PLoS pathogens* 2013,9(9):e1003631.
9. Habchi J, Blangy S, Mamelli L, Jensen MR, Blackledge M, Darbon H, Oglesbee M, Shu Y, Longhi S: Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipavirus. *J Biol Chem* 2011,286(15):13583-13602.
10. Ivanyi-Nagy R, Darlix JL: Fuzziness in the core of the human pathogenic viruses HCV and HIV. *Advances in experimental medicine and biology* 2012,725:142-158.
11. Davey NE, Trave G, Gibson TJ: How viruses hijack cell regulation. *Trends Biochem Sci* 2011,36(3):159-169.
12. Ferreon JC, Martinez-Yamout MA, Dyson HJ, Wright PE: Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc Natl Acad Sci U S A* 2009,106(32):13260-13265.
13. Ferreon AC, Ferreon JC, Wright PE, Deniz AA: Modulation of allostery by protein intrinsic disorder. *Nature* 2013,498(7454):390-394.
14. Hagai T, Azia A, Babu MM, Andino R: Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. *Cell reports* 2014,7(5):1729-1739.
15. Via A, Uyar B, Brun C, Zanzoni A: How pathogens use linear motifs to perturb host cell networks. *Trends Biochem Sci* 2015,40(1):36-48.

16. Stanley BJ, Ehrlich ES, Short L, Yu Y, Xiao Z, Yu XF, Xiong Y: Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly. *Journal of virology* 2008,82(17):8656-8663.

17. Welcker M, Clurman BE: The SV40 large T antigen contains a decoy phosphodegron that mediates its interactions with Fbw7/hCdc4. *The Journal of biological chemistry* 2005,280(9):7654-7658.

18. Stangler T, Tran T, Hoffmann S, Schmidt H, Jonas E, Willbold D: Competitive displacement of full-length HIV-1 Nef from the Hck SH3 domain by a high-affinity artificial peptide. *Biological chemistry* 2007,388(6):611-615.

19. Schlundt A, Sticht J, Piotukh K, Kosslick D, Jahnke N, Keller S, Schuemann M, Krause E, Freund C: Proline-rich sequence recognition: II. Proteomics analysis of Tsg101 ubiquitin-E2-like variant (UEV) interactions. *Molecular & cellular proteomics : MCP* 2009,8(11):2474-2486.

20. Dinkel H, Van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, Speck T, Kruger D, Grebnev G, Kuban M, Strumillo M, Uyar B, Budd A, Altenberg B, Seiler M, Chemes LB, Glavina J, Sanchez IE, Diella F, Gibson TJ: The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res* 2014,42(Database issue):D259-266.

21. Fuxreiter M, Tompa P, Simon I: Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* 2007,23(8):950-956.

22. Calderone A, Licata L, Cesareni G: VirusMentha: a new resource for virus-host protein interactions. *Nucleic acids research* 2015,43(Database issue):D588-592.

23. Weatheritt RJ, Jehl P, Dinkel H, Gibson TJ: iELM--a web server to explore short linear motif-mediated interactions. *Nucleic Acids Res* 2012,40(Web Server issue):W364-369.

24. Pushker R, Mooney C, Davey NE, Jacque JM, Shields DC: Marked variability in the extent of protein disorder within and between viral families. *PloS one* 2013,8(4):e60724.

25. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK: DisProt: the Database of Disordered Proteins. *Nucleic Acids Res* 2007,35(Database issue):D786-793.

26. Dosztanyi Z, Csizmok V, Tompa P, Simon I: The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol* 2005,347(4):827-839.

27. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z: Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* 2006,7:208.

28. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H, Gibson TJ: Attributes of short linear motifs. *Mol Biosyst* 2012,8(1):268-281.

29. Davey NE, Shields DC, Edwards RJ: SLiMDisc: short, linear motif discovery, correcting for common evolutionary descent. *Nucleic Acids Res* 2006,34(12):3546-3554.

30. Fuxreiter M, Simon I, Friedrich P, Tompa P: Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol* 2004,338(5):1015-1026.

31. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK: Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* 2005,44(37):12454-12470.

32. Meszaros B, Simon I, Dosztanyi Z: Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol* 2009,5(5):e1000376.

33. Pancsa R, Fuxreiter M: Interactions via intrinsically disordered regions: what kind of motifs? *IUBMB Life* 2012,64(6):513-520.

34. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN: Analysis of molecular recognition features (MoRFs). *J Mol Biol* 2006,362(5):1043-1059.

35. Jones DT, Cozzetto D: DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* 2015,31(6):857-863.

36. Curnis F, Cattaneo A, Longhi R, Sacchi A, Gasparri AM, Pastorino F, Di Matteo P, Traversari C, Bachi A, Ponzoni M, Rizzardi GP, Corti A: Critical role of flanking residues in NGR-to-isoDGR transition and CD13/integrin receptor switching. *The Journal of biological chemistry* 2010,285(12):9114-9123.

37. Liu J, Tan H, Rost B: Loopy proteins appear conserved in evolution. *Journal of Molecular Biology* 2002,322:53-64.

38. Rezaei-Ghaleh N, Blackledge M, Zweckstetter M: Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery. *Chembiochem* 2012,13(7):930-950.

39. Schwarten M, Solyom Z, Feuerstein S, Aladag A, Hoffmann S, Willbold D, Brutscher B: Interaction of nonstructural protein 5A of the hepatitis C virus with Src homology 3 domains using noncanonical binding sites. *Biochemistry* 2013,52(36):6160-6168.

40. Jurgens MC, Voros J, Rautureau GJ, Shepherd DA, Pye VE, Muldoon J, Johnson CM, Ashcroft AE, Freund SM, Ferguson N: The hepatitis B virus preS1 domain hijacks host trafficking proteins by motif mimicry. *Nature chemical biology* 2013,9(9):540-547.

41. Liu X, Marmorstein R: Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. *Genes & development* 2007,21(21):2711-2716.

42. Kabsch W, Sander C: Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983,22(12):2577-2637.

43. Wells JA, McClendon CL: Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 2007,450(7172):1001-1009.

44. Vavouri T, Semple J, Garcia-Verdugo R, Lehner B: Intrinsic Protein Disorder and Interaction promiscuity are widely associated with Dosage Sensitivity. *Cell* 2009,138:198-208.

45. Chames P, Van Regenmortel M, Weiss E, Baty D: Therapeutic antibodies: successes, limitations and hopes for the future. *British journal of pharmacology* 2009,157(2):220-233.

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

47. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Honigschmid P, Schafferhans A, Roos M, Bernhofer M, Richter L, Ashkenazy H, Punta M, Schlessinger A, Bromberg Y, Schneider R, Vriend G, Sander C, Ben-Tal N, Rost B: PredictProtein--an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res* 2014,42(Web Server issue):W337-343.

**Figure legends**

**Figure 1** Schematic representation of plausible scenarios how viral ELMs (magenta) can outcompete host ELMs (blue). (**A**) Protein regions embedding viral motifs have increased plasticity, higher level of disorder as compared to eukaryotic motifs. (**B**) Viral motifs are flanked by longer disordered regions, which can hamper the access of the competing host proteins to the target site. (**C**) Environments of viral motifs provide further binding sites, which anchor viral proteins to the target and increase binding affinity. ELMs are represented by solid labeled boxes. Disordered regions are displayed by dashed lines or dotted lines, latter designates higher level of disorder. Intrinsically disordered binding sites are represented by solid boxes.

**Figure 2** Average degree of disorder (**A**) and fraction of disordered residues (**B**) of ELM-containing eukaryotic (dark gray), eukaryotic virus (light gray), human (dark red) and human virus (light red) proteins. Human and human virus ELMs target common host motif-binding domains. p values were computed by Wilcoxon test.

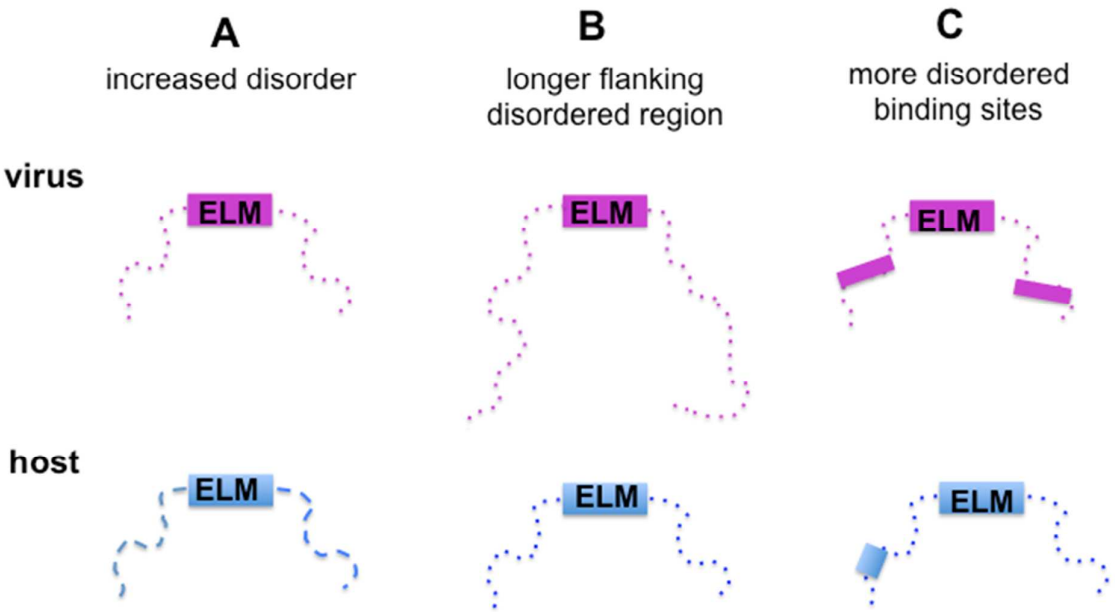
**Figure 3** Disorder properties of protein regions flanking the eukaryotic (dark gray), eukaryotic virus (light gray), human (dark red) and human virus (light red) ELMs. Human and human virus ELMs target common host motif-binding domains. (**A**) Average degree of disorder of the +/- 20 AA flanking region (**B**) Length of the disordered segment flanking the motif . p values were computed by Wilcoxon test.

**Figure 4** Number of intrinsically disordered binding sites within 100 AA flanking regions of the eukaryotic (dark gray), eukaryotic virus (light gray), human (dark red) and human virus (light red) ELMs. Human and human virus ELMs target common host motif-binding domains. p value was computed by Wilcoxon test.

**Figure 5** (A) Average degree of disorder of ELMs of human and human virus ELMs, which target common host motif-binding domains. (B) Difference in the degree of disorder between the human and human virus ELMs and their respective +/- 20 AA flanking regions. p values were computed by Wilcoxon test.

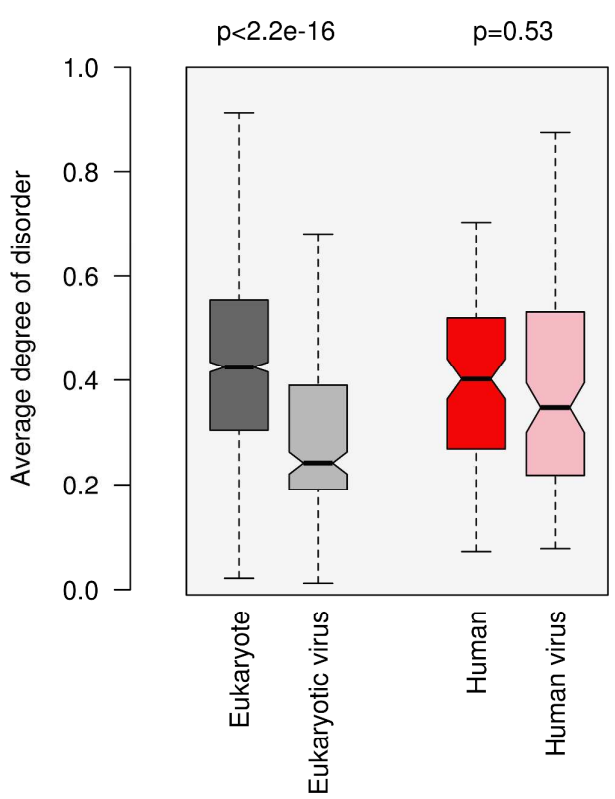
**Figure 6** Interactions of the E1A viral oncoprotein with retinoblastoma protein are realized via an α-helical and a turn secondary structure element (PDB code: 2r7g).

**Figure 1**

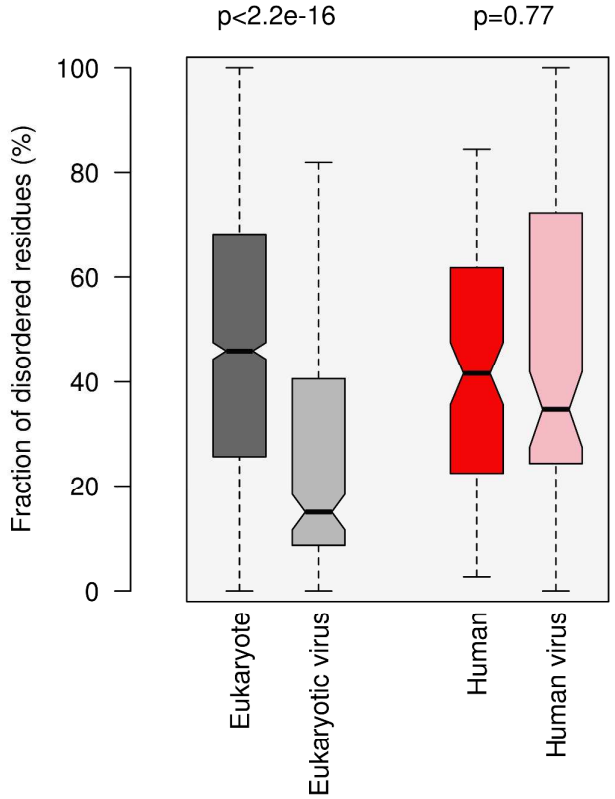


## Figure 2

**A**




**B**



## Figure 3

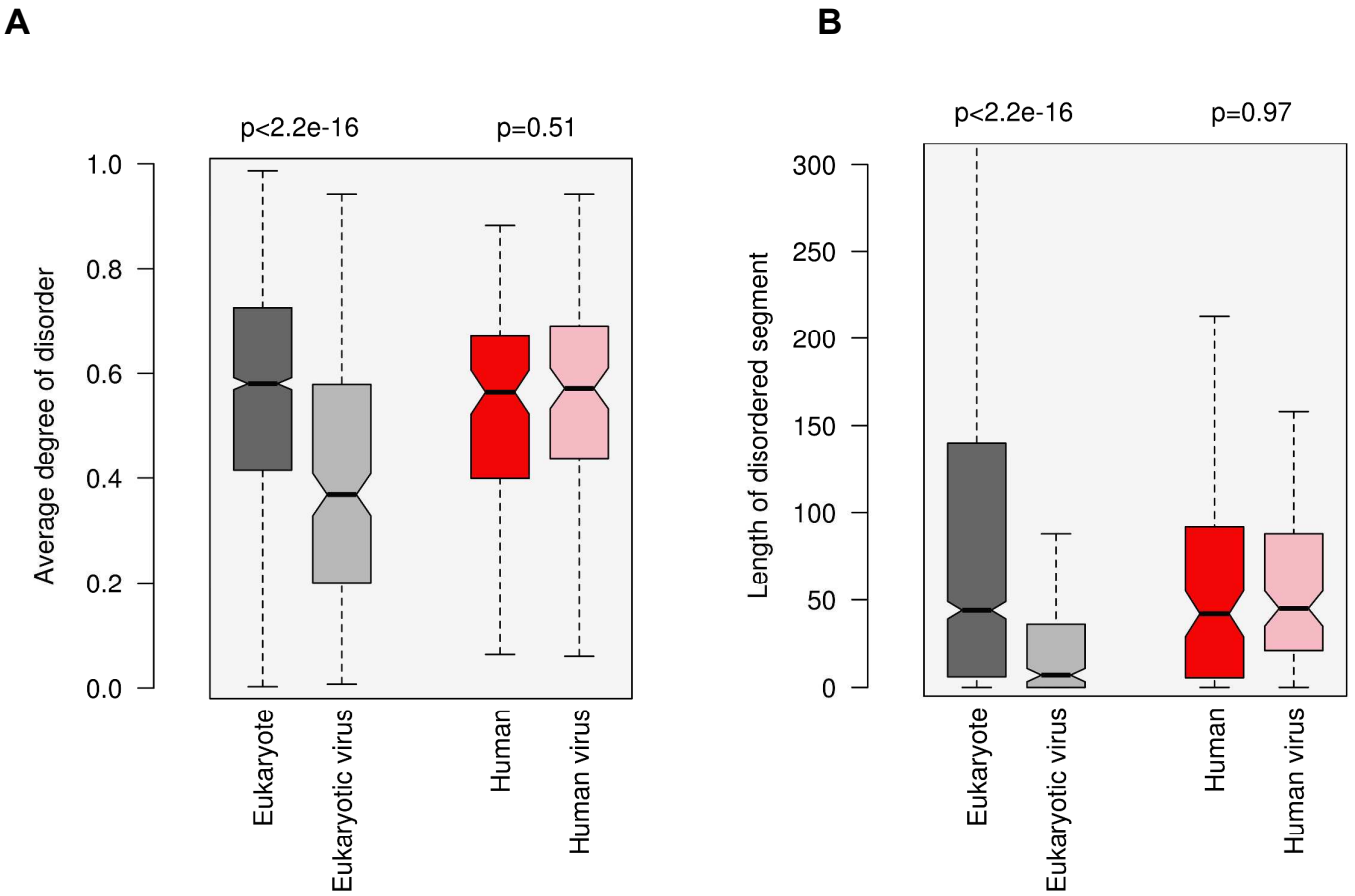
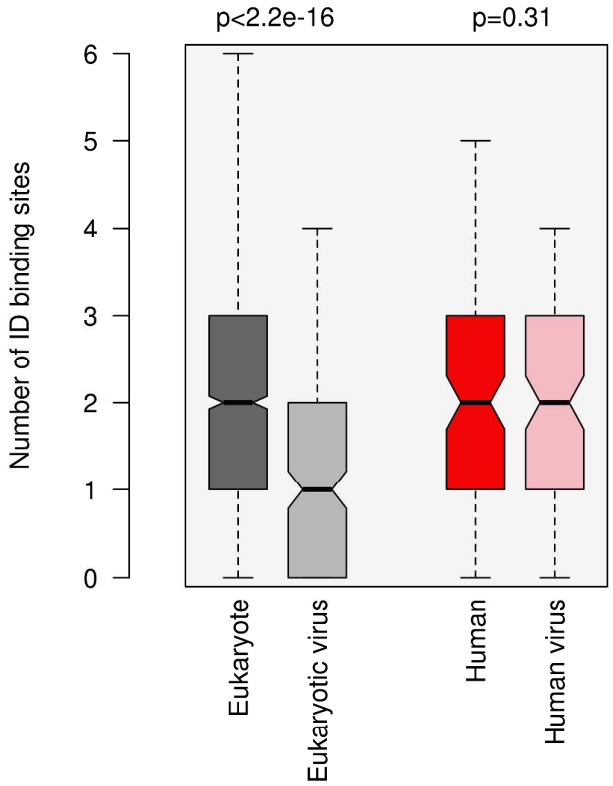
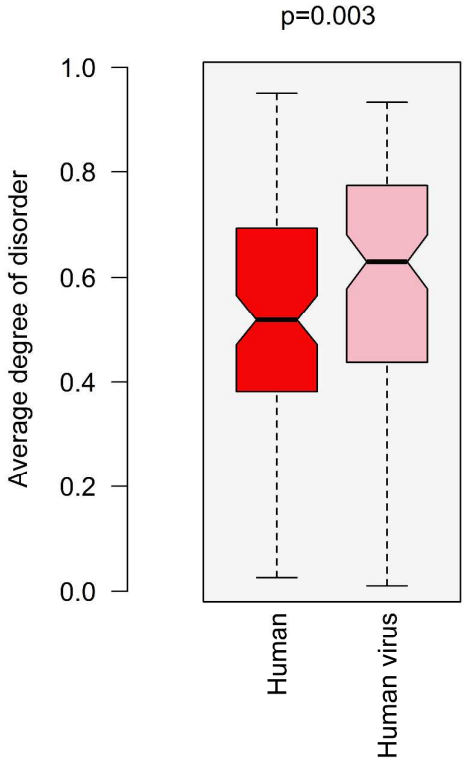


**Figure 4**



## Figure 5

**A**




**B**

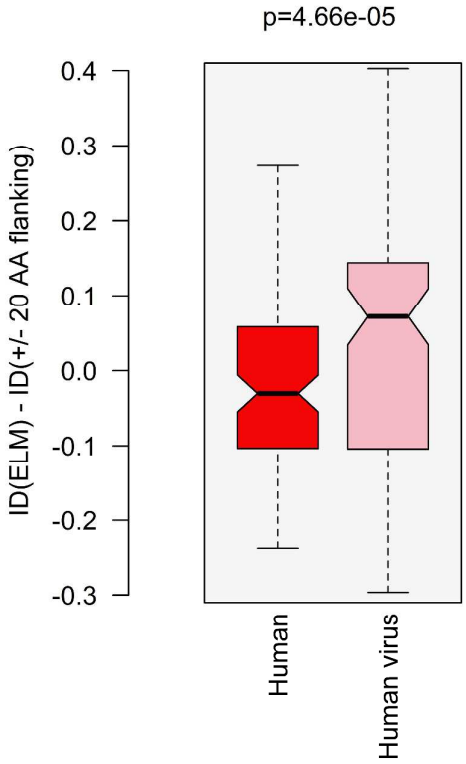


**Figure 6**