

RESEARCH ARTICLE

Investigating the disordered regions (MoRFs, SLiMs and LCRs) and functions of mimicry proteins/peptides *in silico*

Anjali Garg, Govinda Rao Dabburu, Neelja Singhal*, Manish Kumar¹*

Department of Biophysics, University of Delhi South Campus, New Delhi, India

* manish@south.du.ac.in (MK); neelja@south.du.ac.in (NS)**OPEN ACCESS**

Citation: Garg A, Dabburu GR, Singhal N, Kumar M (2022) Investigating the disordered regions (MoRFs, SLiMs and LCRs) and functions of mimicry proteins/peptides *in silico*. PLoS ONE 17(4): e0265657. <https://doi.org/10.1371/journal.pone.0265657>

Editor: Eugene A. Permyakov, Russian Academy of Medical Sciences, RUSSIAN FEDERATION

Received: November 9, 2021

Accepted: March 4, 2022

Published: April 14, 2022

Copyright: © 2022 Garg et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: The work was carried out using the resources funded by the Indian Council of Medical Research [Grant No: VIR(25)/2019/ECD-1] and [Grant No: ISRM/12(33)/2019]. AG is supported by ICMR-JRF scheme [Grant Number: 3/1/3 J.R.F.-2016/LS/HRD-(32262)]. NS is supported by CSIR Senior Research Associateship (Scientists' Pool Scheme) [Grant Number: 13(9089-A)/2019-Pool].

Abstract

Microbial mimicry of the host proteins/peptides can elicit host auto-reactive T- or B-cells resulting in autoimmune disease(s). Since intrinsically disordered protein regions (IDPRs) are involved in several host cell signaling and PPI networks, molecular mimicry of the IDPRs can help the pathogens in substituting their own proteins in the host cell-signaling and PPI networks and, ultimately hijacking the host cellular machinery. Thus, the present study was conducted to discern the structural disorder and intrinsically disordered protein regions (IDPRs) like, molecular recognition features (MoRFs), short linear motifs (SLiMs), and low complexity regions (LCRs) in the experimentally verified mimicry proteins and peptides (mimitopes) of bacteria, viruses and host. Also, functional characteristics of the mimicry proteins were studied *in silico*. Our results indicated that 78% of the bacterial host mimicry proteins and 45% of the bacterial host mimitopes were moderately/highly disordered while, 73% of the viral host mimicry proteins and 31% of the viral host mimitopes were moderately/highly disordered. Among the pathogens, 27% of the bacterial mimicry proteins and 13% of the bacterial mimitopes were moderately/highly disordered while, 53% of the viral mimicry proteins and 21% of the viral mimitopes were moderately/highly disordered. Though IDPR were frequent in host, bacterial and viral mimicry proteins, only a few mimitopes overlapped with the IDPRs like, MoRFs, SLiMs and LCRs. This suggests that most of the microbes cannot use molecular mimicry to modulate the host PPIs and hijack the host cell machinery. Functional analyses indicated that most of the pathogens exhibited mimicry with the host proteins involved in ion binding and signaling pathways. This is the first report on the disordered regions and functional aspects of experimentally proven host and microbial mimicry proteins.

Introduction

Microorganisms might exhibit molecular mimicry with the host by displaying similarities either in the sequences/structures of the proteins or, in the short linear motifs of the proteins or, in the structures of the proteins (without sequence homology). The similarities in the microbial peptides and host epitopes might invoke a cross-activation of the host autoreactive

Competing interests: The authors have declared that no competing interests exist.

T- or B-cells against self-epitopes, resulting in tissue and/or organ destruction and ultimately, autoimmune diseases. The host and the pathogen peptides pairs that show similarity/mimicry with each other are called mimicry peptides or mimitopes.

Several bacteria and viruses like, *Klebsiella pneumoniae*, *Shigella* spp., *Chlamydia trachomatis*, *Helicobacter pylori*, Coxsackievirus, rubella virus etc. have been experimentally implicated in autoimmune diseases like, type I diabetes, multiple sclerosis, autoimmune chronic gastritis, ankylosing spondylitis, reactive arthritis etc [1–5].

Intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs) are those proteins and regions of the proteins respectively, which do not have a well-defined 3D structure. Due to their conformational flexibility, the IDPRs can interact with multiple binding partners, act like hubs in the interaction networks and also perform cellular functions related to signaling and regulation [6–10]. Some conserved functional elements like molecular recognition features (MoRFs) and short linear motifs (SLiMs) have been frequently reported in the IDPRs. MoRFs are amino acid stretches of 10–70 amino acids which are involved in protein-protein interactions (PPIs) and can undergo a disorder-to-order transition upon binding to their partners; this process is termed as coupled folding-binding [11, 12]. IDPs/IDRs can utilize multiple MoRFs simultaneously when interacting with their binding partners [13]. MoRF-containing proteins are present abundantly in the ribosome, nucleus, nucleolus and microtubules and, are involved in translation, protein transport, protein folding, and interactions with DNAs [14]. SLiMs are short stretches of amino acids (3–10 amino acids) which are functionally diverse and mediate signaling interactions [15, 16]. SLiMs regulate the low-affinity interactions and target proteins to a specific subcellular location, recruit enzymes that alter the chemical state of the motif by post-translational modifications, control the stability of a protein, and promote recruitment of binding factors to facilitate complex formation [15, 17]. The short length characteristics of motifs suggest that they might have a high propensity to evolve convergently and emerge in unrelated proteins. Consequently, most of the pathogenic viruses and bacteria evolved to mimic these short linear motifs of the host, allowing them to manipulate cellular processes [18, 19].

Several studies have indicated that viruses manipulate the host cellular machinery by mimicking MoRFs and SLiMs of the host proteins [20, 21]. Similarly, several pathogenic bacteria also mimic the host SLiMs for propagation and sustenance inside the host [22]. Besides MoRFs and SLiMs, disordered regions are also characterized by the presence of low complexity regions (LCRs) which are homo-polymeric repeats of a single amino acid or hetero-polymeric short repeats of a few amino acids residues [23]. Initially, LCRs were considered to be primarily disordered but a few recent studies suggest that LCRs can have a regular secondary structure, also [24]. The LCRs are also associated with important functions like modulation of protein-protein interactions [25], protein-nucleic acid interactions [26], protein subcellular localization [27], antigen processing and diversification [28]. Collectively MoRFs, SLiMs and LCRs facilitate the proteins in adopting several dynamic functional structures, which enables their interaction with multiple binding partners [29]. A few researchers have reported the physico-chemical and structural characteristics of some viral mimicry proteins [20, 21]. However, the presence of disordered regions like MoRFs, SLiMs and LCRs has not been studied in experimentally verified mimicry proteins and mimitopes associated with autoimmune diseases. Since IDPRs are involved in a variety of cell signaling and PPI networks, mimicry of these regions enables the pathogen in substituting its' own proteins in the host PPI networks and eventually hijack the host cellular machinery. Thus, in the present study we have discerned the presence of order/disorderliness, MoRFs, SLiMs and LCRs in mimicry proteins and mimitopes of bacteria, viruses and host using an *in silico* approach. Additionally, the functional annotation of the mimicry proteins was performed using the Gene Ontology (GO) annotations retrieved from the Gene Ontology Consortium [30].

Material and methods

Retrieval of experimentally validated mimicry proteins from miPepBase

The information about bacterial and viral mimicry proteins along with the host mimicry proteins was retrieved from a database of experimentally verified mimicry proteins, miPepBase [31]. The mimicry proteins of the host and pathogen were named as host-protein and pathogen-protein, respectively while, the mimicry peptides (mimitopes) of the host and pathogen were named as host-mimitope and pathogen-mimitope, respectively.

Order/disorder propensity of amino acids in the host and pathogen mimicry proteins

The order/disorder predisposition of amino acids in the host- and pathogen-proteins was predicted using consensus of three different disorder predictors namely, DISOPRED (version 3.16), IUPred (version 1.0) and PONDER VSL2. DISOPRED is a hybrid predictor based on SVM, neural network and nearest neighbor classifiers [32]. For each amino acid, DISOPRED gives a score between 0–1, with 0.50 as the threshold boundary. Amino acids with a DISOPRED score of ≤ 0.50 were considered as ordered while with a score > 0.50 as disordered. IUPred predicts the disordered/unstructured regions in a protein sequence based on total pairwise inter-residue interaction energy and on the assumption that intrinsically unstructured protein sequences do not fold due to their inability to form a sufficient number of stabilizing inter-residue interactions [33]. Similar to DISOPRED, the score of IUPred prediction also ranges from 0 (complete order) to 1 (complete disorder) with a score above 0.5 indicating disorder. PONDR® FIT (Predictor of Natural Disordered Regions) includes six different types of predictors. We used PONDR® VSL2 since it is considered as the most accurate form of PONDR [34]. In PONDR, an amino acid with a score ≥ 0.5 is considered as disordered. In the present study, all three IDPR predictors were used at default parameters. To address the variability in the predictions of the three predictors, the final disorder predisposition of each amino acid was calculated based on the consensus in all the three predictors. An amino acid was annotated ordered/disordered based on consensus in predictions of at least two of the three predictors. On the basis of consensus prediction, we calculated the percentage of disordered residues (PDR) in a protein by dividing the number of residues in a protein that were predicted as disordered by the total number of residues in that protein. On the basis of PDR, all pathogen- and host-proteins were divided into three categories: highly ordered ($\text{PDR} < 10\%$), moderately disordered ($10\% \leq \text{PDR} < 30\%$), and highly disordered ($\text{PDR} \geq 30\%$). Previously, several researchers have also used PDR to classify proteins as ordered/disordered [26–29].

MoRFs, SLiMs and LCRs in the mimicry proteins and mimitopes

The presence of MoRFs in the mimicry proteins and mimitopes was investigated using the MoRFchibi SYSTEM which contains three different modes of MoRF predictions, MoRF-CHiBi, MoRFCHiBi_Light, and MoRFCHiBi_Web [35]. In the present study, MoRFCHiBi_Web mode was used, which though slower than the other two modes, gives highly accurate predictions [35].

To predict SLiMs in the mimicry proteins and mimitopes ANCHOR [36] was used and SEG was used to find LCRs [23]. ANCHOR is one very popular tool to predict the protein binding regions in disordered regions of proteins. It has been used in a large number of work to predict the protein-protein binding motif/SLiM regions in a protein [37–45]. Hence in this work we have used ANCHOR to identify binding motifs present in IDP sequences i.e. SLiM.

For annotating LCRs in a protein sequence, SEG uses three numeric parameters *viz.* window length (L), trigger complexity (K1), and extension complexity (K2). The whole process of LCR identification by SEG undergoes in two-steps. First, it identifies a low complexity segment using a sliding window of length L amino acids with a local sequence complexity K1. Then all overlapping subsequences with sequence complexity K1 are merged in both directions till the complexity of a contig built by overlapping subsequences does not exceed K2. In the present work, MoRFChiBi_Web, ANCHOR and SEG were used at default parameters. MoRFs/SLiMs/LCRs were considered to be present within the host or pathogen mimitopes if at least half of the amino acids of the mimitopes overlapped with these regions.

Functional enrichment analysis

We performed the GO term based functional enrichment analysis of the host and pathogen mimicry proteins using GO mapping tools, OWLTools' Map2Slim (M2S) (<https://github.com/owlcollab/owltools>). The functional enrichment analysis was performed at all three levels of the GO annotations namely, Molecular Function, Cellular Component and Biological Process.

Results

Benchmarking dataset

A total of 147 bacterial mimicry proteins with corresponding 27 host mimicry proteins (named as Bacterial-set) and, 34 viral mimicry proteins with corresponding 22 host mimicry proteins (named as Viral-set) were retrieved from the miPepBase. The bacterial mimicry proteins were involved in 16 while, viral mimicry proteins were involved in 12 different types of autoimmune diseases. The detailed information about the host and the pathogen, UniProtKB ID and name of the mimicry-protein(s), mimitope sequences, associated autoimmune diseases and the scientific literature (PubMed ID) is summarized in [S1 Table](#).

Structural order/disorderliness in the mimicry proteins

To assess the extent of structural disorderliness, the mimicry proteins were divided in three different categories based on the PDR, as was also done in earlier studies [[26](#), [27](#), [41](#)]. Our results revealed that 107 of the 147 bacterial (73%) and 16 of the 34 (47%) viral mimicry proteins were ordered (PDR <10%). This implies that 40 of the 147 (27%) of the bacterial and 18 of the 34 (53%) of the viral mimicry proteins were disordered.

In the Bacterial-set, 21 of the 27 (78%) of the host mimicry proteins and 18 of 40 (45%) of the host mimitopes and, 40 of the 147 (27%) of the bacterial mimicry proteins and 20 of 152 (14%) of the bacterial mimitopes exhibited moderate to high disorderliness.

In the Viral-set 16 of the 22 (73%) of host mimicry proteins and 11 of 35 (31%) of the host mimitopes and, 18 of the 34 (53%) of viral mimicry proteins and 9 of 43 (21%) of the viral mimitopes were moderately/highly disordered ([Table 1](#)). The list of proteins that belonged to each category is shown in [S2 Table](#).

MoRFs, SLiMs and LCRs in mimicry proteins/peptides

In the Bacterial-set, 20 of the 27 (74%) host mimicry proteins and 106 of the 147 (72%) bacterial proteins contain MoRFs. In the Viral-set, 13 of the 22 (59%) host mimicry proteins and 32 of the 34 (94%) viral mimicry proteins had MoRFs. With regards to the mimitopes, 7 of 152 bacterial (5%) and 5 of 40 host (13%) mimitopes of the Bacterial-set and, 3 of 43 viral (7%) and 6 of 35 host (17%) mimitopes of the Viral-set overlapped with the MoRFs ([Table 2](#)).

Table 1. Categorization of mimicry proteins of pathogens and hosts on the basis of PDR (percentage of disordered residues in protein).

Mimicry proteins		Categorization of mimicry proteins on the basis of PDR			No. of mimitopes mapped
		No. of ordered mimicry proteins (PDR <10%)	Number of moderately disordered mimicry proteins (10% ≤ PDR < 30%)	No. of disordered mimicry proteins (PDR ≥ 30%)	
Bacterial-set	Bacterial proteins	107 (73%)	29 (20%)	11 (7%)	20 (14%)
	Host proteins	6 (22%)	8 (30%)	13 (48%)	18 (45%)
Viral-set	Viral proteins	16 (47%)	8 (24%)	10 (29%)	9 (21%)
	Host proteins	6 (27%)	8 (36%)	8 (36%)	11 (31%)

* percent in parentheses depicts the ratio of the number of proteins in each category to the total number of proteins in each dataset.

<https://doi.org/10.1371/journal.pone.0265657.t001>

Analysis of the SLiM regions in the Bacterial-set revealed that 16 of the 27 (59%) host and 41 of the 147 (28%) of the bacterial mimicry proteins had SLiMs. In the Viral-set, 11 of the 22 (50%) of the host and 20 of the 34 (59%) of the viral mimicry proteins had SLiMs. With regards to the mimitopes, 13 of 152 bacterial (9%) and 12 of 40 host (30%) mimitopes of the Bacterial-set and, 4 of 43 (9%) viral and 8 of 35 (23%) host mimitopes of the Viral-set overlapped with the SLiMs.

In the Bacterial-set, 19 of the 27 host (70%) and 87 of the 147 bacteria (59%) mimicry proteins harbored LCRs. In the Viral-set, 16 of the 22 (73%) host and 30 of the 34 viral (88%) mimicry proteins harbored LCRs. With regards to the mimitopes, 8 of the 152 (5%) bacterial and 2 of 40 (5%) host mimitopes of the Bacterial-set while, 6 of the 43 (14%) viral and 3 of 35 (9%) host mimitopes of the viral-set overlapped with the LCRs (Table 2). The list of proteins that contain MoRF, SLiMs, and LCRs is shown in S2 Table and their number with position in each mimicry-protein is shown in S3 Table.

Functional characterization of the host and pathogen mimicry proteins

The GO annotations revealed that the host mimicry proteins of the Bacterial-set were involved in biological processes like, anatomical structure development (10 proteins), ion binding (10 proteins), response to stress (7 proteins), immune system process (6 proteins), signal transduction (6 proteins), cell differentiation (5 proteins), cellular protein modification (5 proteins) and protein transport (5 proteins). The host proteins of the Viral-set were involved in like ion binding (9 proteins), anatomical structure development (8 proteins), biosynthetic process (8

Table 2. Distribution of MoRFs, SLiMs and LCRs in the mimicry proteins and mimitopes of bacteria, viruses and hosts.

Structural feature	Categorization of mimicry proteins		Number of mimicry proteins in which MoRFs, SLiMs and LCRs were present	Number of proteins in which MoRFs, SLiMs and LCRs were present in mimitope region
Molecular recognition features (MoRFs)	Bacterial-set proteins	Bacterial proteins	106	7
		Host proteins	20	5
	Viral-set proteins	Viral proteins	32	3
		Host proteins	13	6
Short linear motifs (SLiMs)	Bacterial-set proteins	Bacterial proteins	41	13
		Host proteins	16	12
	Viral-set proteins	Viral proteins	20	4
		Host proteins	11	8
Low complexity regions (LCRs)	Bacterial-set proteins	Bacterial proteins	87	8
		Host proteins	19	2
	Viral-set proteins	Viral proteins	30	6
		Host proteins	16	3

<https://doi.org/10.1371/journal.pone.0265657.t002>

proteins), immune system process (7 proteins), signal transduction (7 proteins), cell-cell signaling (7 proteins) and catabolic process (7 proteins) (S4 Table). The bacterial proteins were involved in ion binding (45 proteins), biosynthetic process (31 proteins), cellular nitrogen compound metabolic process (22 proteins), carbohydrate metabolism (16 proteins), transmembrane transport (13 proteins), DNA binding (11 proteins) and DNA metabolic process (10 proteins). Functional enrichment of the viral proteins suggested that most of the viral proteins were a part of symbiont process (28 proteins), cellular nitrogen compound metabolic process (11 proteins), biosynthetic process (10 proteins), DNA binding (10 proteins), ion binding (8 proteins), immune system process (7 proteins) and membrane organization (7 proteins) (S4 Table).

In the molecular function category, the functions of the bacterial and viral host proteins were similar, except that the host proteins of the Viral-set did not perform a few functions like transmembrane transporter activity (S4 Table). The bacterial proteins were involved in molecular functions like ion binding (47 proteins), DNA binding (11 proteins), transmembrane transporter activity (10 proteins). The viral proteins were most prominently involved in symbiont process (28 proteins) followed by DNA binding and structural molecule activity (10 proteins in each function).

In the category cellular components, the host mimicry proteins of the Bacterial-set were present in protein-containing complex (10 proteins), plasma membrane (9 proteins), and cell (9 proteins), while the bacterial mimicry-proteins showed a significant presence in cytoplasm (28 proteins) followed by plasma membrane (15 proteins) and protein-containing complex (9 proteins). The functional analysis of Viral-set proteins revealed that host mimicry-proteins were localized at the plasma membrane (13 proteins), cell (10 proteins), and cytoplasm (7 proteins) while no specific subcellular enrichment was observed for viral mimicry-proteins (S4 Table). The biological processes, molecular functions and cellular components enriched in each dataset are shown in Fig 1A–1C.

Discussion

Autoimmune diseases can develop by several mechanisms like, non-specific bystander activation, persistent antigenic stimuli, breach in the central tolerance, host genetics etc [42]. However, microbial molecular mimicry has been recognized as one of the primary mechanisms of pathogen induced autoimmunity. Several studies have indicated that many bacteria and viruses exhibit molecular mimicry with the host proteins and, hijack the host cellular machinery by substituting their own proteins in the host cell signaling and PPI networks [22, 46–49]. Though, the overall fraction of the disordered residues varies in both bacterial and viral proteomes, frequently at least one of the two interacting proteins in viral and host PPIs is structurally disordered [50]. Since, IDPRs are involved in a variety of cell signaling and PPI networks the present study was conducted to discern if the experimentally verified microbial mimicry proteins/mimitopes harbor IDPRs and can thus potentially modulate the host PPIs.

Our results revealed that 78% of the host mimicry proteins and 45% of the host mimitopes involved in bacterial mimicry exhibited moderate to high disorder. Also, 73% of the host proteins and 31% of the host mimitopes involved in viral mimicry exhibited moderate to high disorder. Since disordered regions of the proteins are highly flexible, their presence in the host mimicry proteins might confer a selective advantage to the host in combating the pathogens [51]. With regards to the pathogen, 27% of the bacterial mimicry proteins and 13% of the bacterial mimitopes exhibited moderate to high disorder. In viruses, 53% of the viral mimicry proteins and 21% of the viral mimitopes exhibited moderate to high disorder. Reportedly, the structural disorderliness in viral proteomes might range from as low as 7% in the human coronavirus NL63 to as high as 77% in the avian carcinoma virus, while in bacterial proteomes it

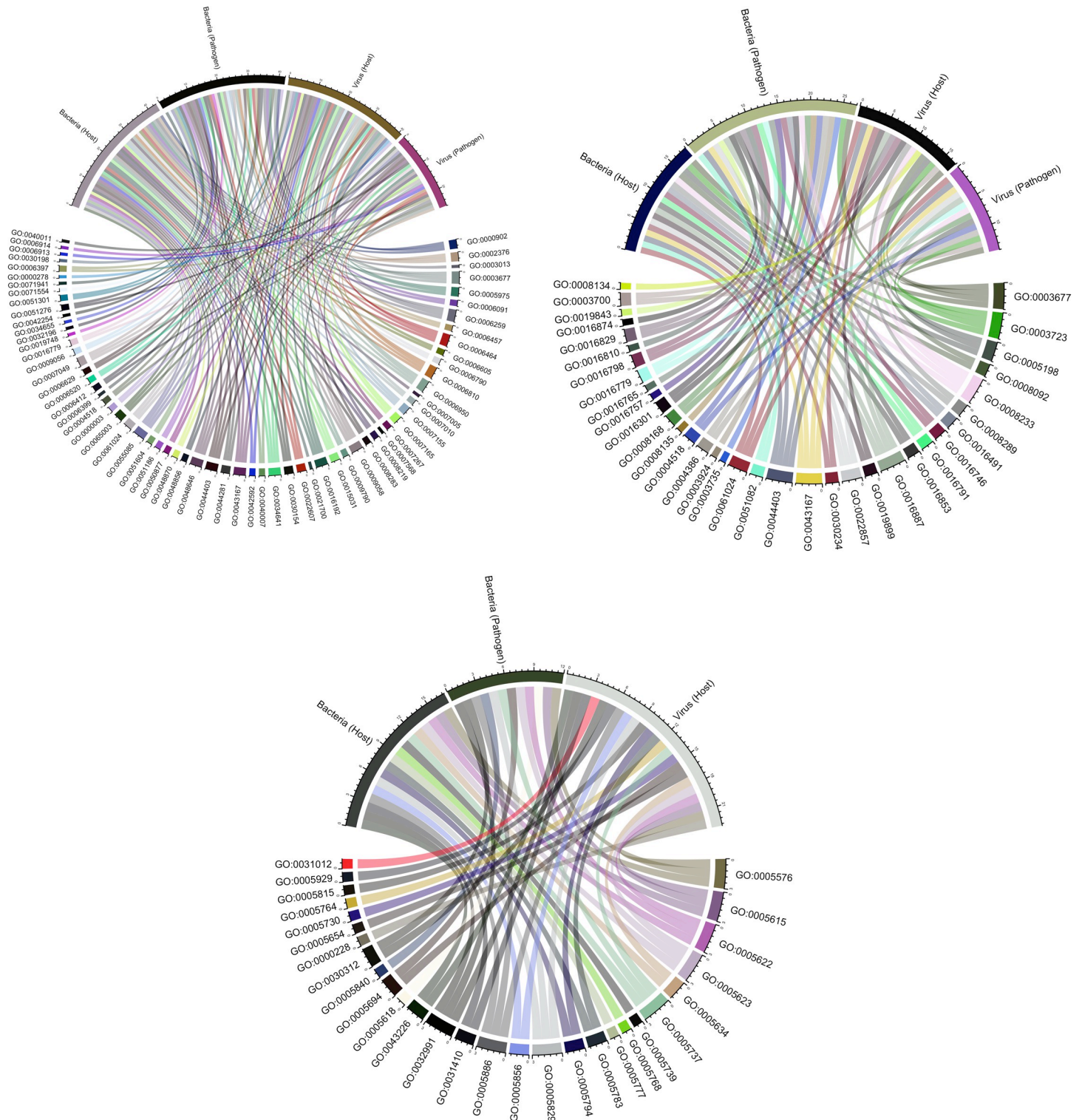


Fig 1. Chord graph representation of gene ontology based functional annotation of bacterial, viral and host mimicry proteins. The GO-term included for: (a) Biological Process, (b) Molecular Function and (c) Cellular Component. Each set is represented by a specific color.

<https://doi.org/10.1371/journal.pone.0265657.g001>

usually varies in a small range of 18–35% [51]. As reported in earlier studies, a similar pattern of structural disorderliness was observed in our study, too. The number of viral mimicry proteins and mimitopes which exhibited disorderliness was much greater than bacterial mimicry proteins and mimitopes.

Analysis of the MoRFs, SLiMs and LCRs in the host and microbial mimicry proteins revealed that these regions were present in both the host and microbial mimicry proteins. However, only a few microbial mimitopes overlapped with these regions. The IDPRs like MoRFs, SLiMs and LCRs are frequently involved in various host processes like, cell signaling and PPI interactions [29]. Substitution of the microbial proteins in the host PPI networks by mimicking these regions can greatly facilitate microbial survival and modulation of the host defense mechanisms. On the contrary, our results indicated that only a few microbial mimitopes overlapped with the IDPRs. This suggests that most of the microbial mimitopes experimentally implicated in autoimmune diseases could not potentially modulate functions of the host proteins.

GO-based functional annotation of the host mimicry proteins of the Bacterial- and Viral-set revealed that most of these proteins were multifunctional. For instance, molecular mimicry between the human Zinc transporter 8 protein (Uniprot Id: Q8IWU4) and MAP_3865c protein of *Mycobacterium avium* subsp. *paratuberculosis* results in Type 1 diabetes mellitus. GO-term analysis of the host mimicry-protein Q8IWU4 revealed its involvement in multiple biological processes like ion binding, stress response, immune system process, protein transport, homeostatic process, transmembrane transport, vesicle-mediated transport and cell-cell signaling. This might also be a probable reason underlying its localization at different organelles like, plasma membrane, golgi apparatus and cytoplasmic vesicles (S4 Table). Additionally, many mimicry proteins of the host were involved in ion-binding, DNA/RNA-binding and signaling pathways. It is well-known that metal-sequestering host-defense proteins and microbial metal acquisition machinery are important players in bacterial pathology and disease outcomes [49]. The host immune system counteracts the bacterial infections by reducing metal availability, but the pathogen outmaneuvers this by hijacking the host metalloproteins. In the present study too, we observed that the bacteria and viruses exhibited molecular mimicry with those host proteins that were mainly involved in ion-binding, DNA/RNA-binding and signaling pathways.

Conclusion

Our results indicated that 78% of the host proteins and 45% of the host mimitopes mimicked by the bacteria were disordered. Also, 73% of the host proteins and 31% of the host mimitopes mimicked by viruses were disordered. In viruses, 53% of the mimicry proteins and 21% of the mimitopes were disordered and in bacteria, only 27% of the mimicry proteins and 13% of the mimitopes were disordered. Moreover, only a few microbial mimitopes overlapped with the IDPRs like MoRFs, SLiMs and LCRs. This suggests that most of the microbial mimitopes experimentally implicated in autoimmune diseases might not potentially modulate the functions of the host proteins. Functional analyses indicated that the host mimicry proteins were mostly involved in ion binding, DNA/RNA-binding and signaling pathways, reflecting that pathogens might preferentially mimic the host proteins which are multi-functional. This is the first *in silico* study investigating the disordered regions and functional aspects of bacterial, viral and host mimicry proteins. We hope our study might serve as a useful platform for further studies on pathogen induced molecular mimicry.

Supporting information

S1 Table. Detailed information about mimicry proteins, peptides and the associated autoimmune diseases of bacteria, viruses and hosts. The data was obtained from miPepBase

database (Garg et al. (2016), *Frontiers in Microbiology*).
(XLSX)

S2 Table. Bacterial and Viral-set host and pathogen mimicry proteins containing IDPRs, SLiMs, MoRFs, and LCR. Proteins in which mimicry epitope overlapped with the IDPRs, SLiMs, MoRFs, and LCR are also shown (PDR: Percentage of disordered residues).

(XLSX)

S3 Table. The sequence and position (shown in parenthesis) of MoRFs/SLiMs/LCRs into the host and pathogen mimicry proteins.

(XLSX)

S4 Table. Functional annotation of Bacterial and Viral-set mimicry proteins on the basis of gene ontology (GO) term. The GO-term is mentioned for all three categories namely (i) Biological Process (ii) Molecular Function (iii) Cellular Component.

(XLSX)

Author Contributions

Conceptualization: Neelja Singhal, Manish Kumar.

Data curation: Anjali Garg.

Formal analysis: Anjali Garg.

Investigation: Neelja Singhal, Manish Kumar.

Resources: Manish Kumar.

Supervision: Neelja Singhal, Manish Kumar.

Validation: Govinda Rao Dabburu.

Visualization: Anjali Garg.

Writing – original draft: Anjali Garg.

Writing – review & editing: Govinda Rao Dabburu, Neelja Singhal, Manish Kumar.

References

1. Fujinami RS. Viruses and autoimmune disease—two sides of the same coin? *Trends Microbiol.* 2001; 9: 377–381. [https://doi.org/10.1016/s0966-842x\(01\)02097-2](https://doi.org/10.1016/s0966-842x(01)02097-2) PMID: 11514220
2. Libbey JE, Fujinami RS. Potential triggers of MS. *Results Probl Cell Differ.* 2010; 51: 21–42. https://doi.org/10.1007/400_2008_12 PMID: 19130026
3. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. *Ann Neurol.* 2007; 61: 504–513. <https://doi.org/10.1002/ana.21141> PMID: 17492755
4. Tarakhovsky A, Prinjha RK. Drawing on disorder: How viruses use histone mimicry to their advantage. *J Exp Med.* 2018; 215: 1777–1787. <https://doi.org/10.1084/jem.20180099> PMID: 29934321
5. Cusick MF, Libbey JE, Fujinami RS. Molecular mimicry as a mechanism of autoimmune disease. *Clin Rev Allergy Immunol.* 2012; 42: 102–111. <https://doi.org/10.1007/s12016-011-8294-7> PMID: 22095454
6. O'Shea C, Staby L, Bendtsen SK, Tidemand FG, Redsted A, Willemoës M, et al. Structures and Short Linear Motif of Disordered Transcription Factor Regions Provide Clues to the Interactome of the Cellular Hub Protein Radical-induced Cell Death1. *J Biol Chem.* 2017; 292: 512–527. <https://doi.org/10.1074/jbc.M116.753426> PMID: 27881680
7. Yang J, Gao M, Xiong J, Su Z, Huang Y. Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding. *Protein Sci.* 2019; 28: 1952–1965. <https://doi.org/10.1002/pro.3718> PMID: 31441158

8. An SY, Ka D, Kim I, Kim E-H, Kim N-K, Bae E, et al. Intrinsic disorder is essential for Cas9 inhibition of anti-CRISPR AcrIIA5. *Nucleic Acids Res.* 2020; 48: 7584–7594. <https://doi.org/10.1093/nar/gkaa512> PMID: [32544231](https://pubmed.ncbi.nlm.nih.gov/32544231/)
9. Du Z, Uversky VN. Functional roles of intrinsic disorder in CRISPR-associated protein Cas9. *Molecular BioSystems.* 2017. pp. 1770–1780. <https://doi.org/10.1039/c7mb00279c> PMID: [28692085](https://pubmed.ncbi.nlm.nih.gov/28692085/)
10. Ruan H, Sun Q, Zhang W, Liu Y, Lai L. Targeting intrinsically disordered proteins at the edge of chaos. *Drug Discov Today.* 2019; 24: 217–227. <https://doi.org/10.1016/j.drudis.2018.09.017> PMID: [30278223](https://pubmed.ncbi.nlm.nih.gov/30278223/)
11. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol.* 2002; 12: 54–60. [https://doi.org/10.1016/s0959-440x\(02\)00289-0](https://doi.org/10.1016/s0959-440x(02)00289-0) PMID: [11839490](https://pubmed.ncbi.nlm.nih.gov/11839490/)
12. Wright PE, Dyson HJ. Linking folding and binding. *Curr Opin Struct Biol.* 2009; 19: 31–38. <https://doi.org/10.1016/j.sbi.2008.12.003> PMID: [19157855](https://pubmed.ncbi.nlm.nih.gov/19157855/)
13. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, et al. Analysis of Molecular Recognition Features (MoRFs). *J Mol Biol.* 2006; 362: 1043–1059. <https://doi.org/10.1016/j.jmb.2006.07.087> PMID: [16935303](https://pubmed.ncbi.nlm.nih.gov/16935303/)
14. Yan J, Dunker AK, Uversky VN, Kurgan L. Molecular recognition features (MoRFs) in three domains of life. *Mol Biosyst.* 2016; 12: 697–710. <https://doi.org/10.1039/c5mb00640f> PMID: [26651072](https://pubmed.ncbi.nlm.nih.gov/26651072/)
15. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, et al. Attributes of short linear motifs. *Mol Biosyst.* 2012; 8: 268–281. <https://doi.org/10.1039/c1mb05231d> PMID: [21909575](https://pubmed.ncbi.nlm.nih.gov/21909575/)
16. Dinkel H, Van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, et al. The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.* 2014; 42: D259–66. <https://doi.org/10.1093/nar/gkt1047> PMID: [24214962](https://pubmed.ncbi.nlm.nih.gov/24214962/)
17. Diella F. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci.* 2008; Volume: 6580. <https://doi.org/10.2741/3175> PMID: [18508681](https://pubmed.ncbi.nlm.nih.gov/18508681/)
18. Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. *Trends Biochem Sci.* 2011; 36: 159–169. <https://doi.org/10.1016/j.tibs.2010.10.002> PMID: [21146412](https://pubmed.ncbi.nlm.nih.gov/21146412/)
19. Jürgens MC, Vörös J, Rautureau GJP, Shepherd DA, Pye VE, Muldoon J, et al. The hepatitis B virus preS1 domain hijacks host trafficking proteins by motif mimicry. *Nat Chem Biol.* 2013; 9: 540–547. <https://doi.org/10.1038/nchembio.1294> PMID: [23851574](https://pubmed.ncbi.nlm.nih.gov/23851574/)
20. Xue B, Blocquel D, Habchi J, Uversky AV, Kurgan L, Uversky VN, et al. Structural disorder in viral proteins. *Chem Rev.* 2014; 114: 6880–6911. <https://doi.org/10.1021/cr4005692> PMID: [24823319](https://pubmed.ncbi.nlm.nih.gov/24823319/)
21. Dolan PT, Roth AP, Xue B, Sun R, Dunker AK, Uversky VN, et al. Intrinsic disorder mediates hepatitis C virus core-host cell protein interactions: HCV Core MoRFs Target Cellular Proteins. *Protein Sci.* 2015; 24: 221–235. <https://doi.org/10.1002/pro.2608> PMID: [25424537](https://pubmed.ncbi.nlm.nih.gov/25424537/)
22. Sámano-Sánchez H, Gibson TJ. Mimicry of Short Linear Motifs by Bacterial Pathogens: A Drugging Opportunity. *Trends Biochem Sci.* 2020; 45: 526–544. <https://doi.org/10.1016/j.tibs.2020.03.003> PMID: [32413327](https://pubmed.ncbi.nlm.nih.gov/32413327/)
23. Wootton JC, Federhen S. Statistics of local complexity in amino acid sequences and sequence databases. *Computers & Chemistry.* 1993. pp. 149–163. [https://doi.org/10.1016/0097-8485\(93\)85006-x](https://doi.org/10.1016/0097-8485(93)85006-x)
24. Kumari B, Kumar R, Kumar M. Low complexity and disordered regions of proteins have different structural and amino acid preferences. *Mol Biosyst.* 2015; 11: 585–594. <https://doi.org/10.1039/c4mb00425f> PMID: [25468592](https://pubmed.ncbi.nlm.nih.gov/25468592/)
25. Xiao H, Jeang KT. Glutamine-rich domains activate transcription in yeast *Saccharomyces cerevisiae*. *J Biol Chem.* 1998; 273: 22873–22876. <https://doi.org/10.1074/jbc.273.36.22873> PMID: [9722505](https://pubmed.ncbi.nlm.nih.gov/9722505/)
26. Shen H, Kan JLC, Green MR. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote pre-spliceosome assembly. *Mol Cell.* 2004; 13: 367–376. [https://doi.org/10.1016/s1097-2765\(04\)00025-5](https://doi.org/10.1016/s1097-2765(04)00025-5) PMID: [14967144](https://pubmed.ncbi.nlm.nih.gov/14967144/)
27. Salichs E, Ledda A, Mularoni L, Albà MM, de la Luna S. Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. *PLoS Genet.* 2009; 5: e1000397. <https://doi.org/10.1371/journal.pgen.1000397> PMID: [19266028](https://pubmed.ncbi.nlm.nih.gov/19266028/)
28. Coletta A, Pinney JW, Solís DYW, Marsh J, Pettifer SR, Attwood TK. Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst Biol.* 2010; 4: 19. <https://doi.org/10.1186/1752-0509-4-19> PMID: [20202198](https://pubmed.ncbi.nlm.nih.gov/20202198/)
29. Cumberworth A, Lamour G, Babu MM, Gsponer J. Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochem J.* 2013; 454: 361–369. <https://doi.org/10.1042/BJ20130545> PMID: [23988124](https://pubmed.ncbi.nlm.nih.gov/23988124/)
30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000; 25: 25–29. <https://doi.org/10.1038/75556> PMID: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/)

31. Garg A, Kumari B, Kumar R, Kumar M. miPepBase: A Database of Experimentally Verified Peptides Involved in Molecular Mimicry. *Front Microbiol.* 2017; 8: 2053. <https://doi.org/10.3389/fmicb.2017.02053> PMID: 29109711
32. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics.* 2015; 31: 857–863. <https://doi.org/10.1093/bioinformatics/btu744> PMID: 25391399
33. Dosztányi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.* 2005; 21: 3433–3434. <https://doi.org/10.1093/bioinformatics/bti541> PMID: 15955779
34. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta.* 2010; 1804: 996–1010. <https://doi.org/10.1016/j.bbapap.2010.01.011> PMID: 20100603
35. Malhis N, Jacobson M, Gsponer J. MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res.* 2016; 44: W488–93. <https://doi.org/10.1093/nar/gkw409> PMID: 27174932
36. Dosztányi Z, Mészáros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics.* 2009; 25: 2745–2746. <https://doi.org/10.1093/bioinformatics/btp518> PMID: 19717576
37. Khan W, Duffy F, Pollastri G, Shields DC, Mooney C. Predicting binding within disordered protein regions to structurally characterised peptide-binding domains. *PLoS One.* 2013; 8: e72838. <https://doi.org/10.1371/journal.pone.0072838> PMID: 24019881
38. Olabode AS, Jiang X, Robertson DL, Lovell SC. Ebola virus is evolving but not changing: No evidence for functional change in EBOV from 1976 to the 2014 outbreak. *Virology.* 2015; 482: 202–207. <https://doi.org/10.1016/j.virol.2015.03.029> PMID: 25880111
39. Yang L-N, Liu H, Duan G-H, Huang Y-M, Liu S, Fang Z-G, et al. The AVR2 Effector Escapes R2 Recognition Through Effector Disordering. *Mol Plant Microbe Interact.* 2020; 33: 921–931. <https://doi.org/10.1094/MPMI-07-19-0179-R> PMID: 32212906
40. Latysheva NS, Oates ME, Maddox L, Flock T, Gough J, Buljan M, et al. Molecular Principles of Gene Fusion Mediated Rewiring of Protein Interaction Networks in Cancer. *Mol Cell.* 2016; 63: 579–592. <https://doi.org/10.1016/j.molcel.2016.07.008> PMID: 27540857
41. Xue R, Zakharov MN, Xia Y, Bhasin S, Costello JC, Jasuja R. Research resource: EPSLiM: ensemble predictor for short linear motifs in nuclear hormone receptors. *Mol Endocrinol.* 2014; 28: 768–777. <https://doi.org/10.1210/me.2014-1006> PMID: 24678734
42. Marquez Y, Höpfler M, Ayatollahi Z, Barta A, Kalyna M. Unmasking alternative splicing inside protein-coding exons defines exons and their role in proteome plasticity. *Genome Res.* 2015; 25: 995–1007. <https://doi.org/10.1101/gr.186585.114> PMID: 25934563
43. Weisz J, Uversky VN. Zooming into the Dark Side of Human Annexin-S100 Complexes: Dynamic Alliance of Flexible Partners. *Int J Mol Sci.* 2020; 21. <https://doi.org/10.3390/ijms21165879> PMID: 32824294
44. Vallet SD, Guérout M, Belloy N, Dauchez M, Ricard-Blum S. A Three-Dimensional Model of Human Lysyl Oxidase, a Cross-Linking Enzyme. *ACS Omega.* 2019; 4: 8495–8505. <https://doi.org/10.1021/acsomega.9b00317> PMID: 31459939
45. Hernández-García J, Briones-Moreno A, Dumas R, Blázquez MA. Origin of Gibberellin-Dependent Transcriptional Regulation by Molecular Exploitation of a Transactivation Domain in DELLA Proteins. *Mol Biol Evol.* 2019; 36: 908–918. <https://doi.org/10.1093/molbev/msz009> PMID: 30668817
46. Anand P, Puranik A, Aravamudan M, Venkatakrishnan AJ, Soundararajan V. SARS-CoV-2 selectively mimics a cleavable peptide of human ENaC in a strategic hijack of host proteolytic machinery. <https://doi.org/10.1101/2020.04.29.069476>
47. Spriggs CC, Harwood MC, Tsai B. How non-enveloped viruses hijack host machineries to cause infection. *Virus Entry.* 2019. pp. 97–122. <https://doi.org/10.1016/bs.aivir.2019.05.002> PMID: 31439154
48. Zhu L, Qin J. A Viral Protein Mimics Histone to Hijack Host MORC3. *Structure.* 2019. pp. 883–885. <https://doi.org/10.1016/j.str.2019.05.007> PMID: 31167123
49. Neumann W, Hadley RC, Nolan EM. Transition metals at the host-pathogen interface: how exploit human metalloproteins for acquiring iron and zinc. *Essays Biochem.* 2017; 61: 211–223. <https://doi.org/10.1042/EBC20160084> PMID: 28487398
50. Halehalli RR, Nagarajaram HA. Molecular principles of human virus protein-protein interactions. *Bioinformatics.* 2015; 31: 1025–1033. <https://doi.org/10.1093/bioinformatics/btu763> PMID: 25417202
51. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn.* 2012; 30: 137–149. <https://doi.org/10.1080/07391102.2012.675145> PMID: 22702725