Intrinsically disordered proteins ($IDPs$) in $SARS\text{-}CoV\text{-}2$ interactome (manuscript)

Lazar Vasović (student)
Gordana Pavlović-Lažetić (professor)

April 30, 2022

Abstract

This manuscript discusses the properties of proteins and their relations in the interactomes of the selected subsets of $SARS\text{-}CoV\text{-}2$ proteome – membrane, non-structural proteins, and, finally, full proteome. Protein disorder and respective node degrees in the protein interaction networks were singled out as the features of interest. Viral interactomes were also combined with that of human lung tissue so as to establish which connections in the resulting viral-host interactome are interesting and how they are linked to protein disorder. Results, which include several scatter plots and correlation analysis, indicate that hubs (highly connected nodes) and their neighbors tend to be on average more ordered than proteins with a small number of connections and their neighbors. This is in contrast to the similar studies conducted on eukaryotic interactomes and possibly opens the path to a new direction of research on viral interactomes.

Contents

1 Introduction

Biological systems are typically viewed as networks of mutually interacting **basic building blocks** [1, 2]. Units that operate independently are rare, so **interactions** are a key part of any biological system. In fact, no system can be viewed as a simple sum of elements. This is true both for small cells and huge ecosystems, and the only significant difference between them is the type of basic blocks. Example microblocks (small basic blocks) are genes, proteins, and metabolites, while individual organisms and generally species are considered macroblocks (big basic blocks).

In order to conduct **computational analysis** of biological systems, it is necessary to choose an appropriate **mathematical model**. One natural model of interacting blocks is a **graph**. The use of graph formalism enables the effective representation of the real-life entities and the complex and mutual interrelationships among those entities. Basic blocks are represented as graph **nodes** and interactions as graph **edges**. Of course, this representation has its flaws, mostly because it oversimplifies some complex non-binary interactions. However, it is still quite a useful abstraction, since it opens the path to effective computational analysis.

Protein interaction networks (_PINs_), also known as protein-protein interaction networks (_PPINs_ or _PPI_ networks), are an important class of biological systems. Each their node represents a **protein**, while **undirected** edges encapsulate binary physical interaction between two proteins, also known as **protein-protein interactions (_PPIs_)**. As a part of larger _omics_ studies, _PINs_ are also known as **interactomes** and graph nodes as **interactors**. The importance of these networks lies in the fact that protein molecules are the workhorses of the cell, performing and controlling almost all activities in an organism. For example, highly connected proteins (hubs) within a network are most probably vital molecules and are in many cases more essential for survival than proteins with lower connectivity. Related _PINs_ can also be aligned with each other as a way of identifying substructures that have been preserved throughout evolution. Densely connected parts of the network may represent potential protein complexes, which is an additional benefit of using graph formalism.

One particularly interesting class of proteins are **intrinsically disordered proteins (_IDPs_)**. These proteins lack a fixed and well-defined three-dimensional structure and are highly abundant in nature [3]. Their conformational flexibility (intrinsic disorder) may be present in a molten globular form, in a form of a random coil, or a pre-molten globule. Only a few proteins are fully disordered in their native state, so _IDPs_ also include semi-disordered proteins, whose flexible parts are called **intrinsically disordered regions (_IDRs_)**. _IDPs_ and _IDRs_ often facilitate protein-protein interactions, being flexible linkers and binding to multiple partners, thus _IDPs_ frequently serve as hubs in complex _PPI_ networks.

In the last years, many authors have researched the topic of interactomics and intrinsically disordered proteins. Much has been revealed about the biological importance of _IDPs/IDRs_, as well as their **markedly big** interactomes [4, 5]. Most studies in the field of bioinformatics and computational analysis focused on exploring _PINs_ of eukaryotic organisms such as human (_Homo sapiens_) and baker's yeast (_Saccharomyces cerevisiae_). There is an agreement between these papers that hubs, which are highly connected nodes, are **on average more disordered** than ends, which are proteins with a small number of connections [6, 7, 8, 9, 10].

Contrary to the related work, this manuscript focuses on analyzing the

interactomes of the selected subsets of **SARS-CoV-2 proteome**. *SARS-CoV-2* is a virus, so its proteins could possibly act differently than those of eukaryotes. Protein **disorder degree** and respective **node degrees** in the *PINs* were singled out as the features of interest. Additionally, since merged **viral-host interactomes** had been found as useful in establishing which interactions are biologically interesting [11], viral interactomes were combined with that of human lung tissue. This made it possible to discuss if links in the viral-host interplay are related to protein disorder. The methodology is described in section 2, while results, which are based on several scatter plots and correlation analysis, are shown in section 3.

2   Materials and methods

The paper aimed to examine several *SARS-CoV-2* interactomes, all constructed from **publicly available data**. Four different graphs were chosen as the starting point, with their core properties shown in table 1.

Table 1: Analyzed interactomes and their properties

	Main protein(s)	Source database	Nodes	Edges
M_IntAct	membrane (M)	IntAct	198	294
M_iRef	membrane (M)	iRefIndex	211	451
Nsp_iRef	non-structural (NSPs)	iRefIndex	311	1066
SARS2_iRef	full proteome	iRefIndex	1419	8722

As an example, the interactome of non-structural proteins (*NSPs*) is shown in figure 1. Nodes are **colored** by origin – human are blue, while viral are orange. Node **size** is proportional to its degree in the network.
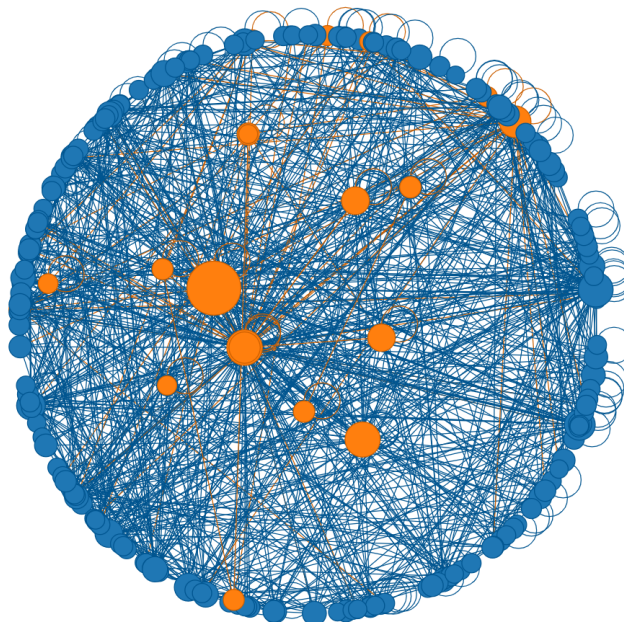

Figure 1: Nsp_iRef, interactome of non-structural proteins

Database ***IntAct***[1] was initially chosen as one of the most popular in the field of protein-protein interaction warehousing. It, however, doesn't include separate data for the *NSPs* of the considered *SARS-CoV-2*[2], which makes it inappropriate for conducting full-scale analysis. Instead, ***iRefIndex***[3] was chosen as the only significant database that meets the coverage criterion. Both databases were queried by protein *IDs* and the resulting lists of *PPIs* were used for building the networks. It can be noted that *iRefIndex* produces **noticeably denser** networks in comparison to *IntAct*, i.e. it contains many additional interactions. Another interesting fact is that the graphs containing non-structural proteins (*Nsp* and *SARS2_iRef*) are much **more connected** than those without them (*M*).

Two additional interactomes were taken into consideration so as to analyze the dynamics of viral-host interplay. They are shown in table 2.

Table 2: Additional interactomes and their properties

	Main proteins	Source database	Nodes	Edges
SARS_iRef	only viral proteins	iRefIndex	30	42
lungs_HuRI	human lung tissue	NDEx	175	162

The first graph, containing only **viral proteins**, is shown in figure 2.
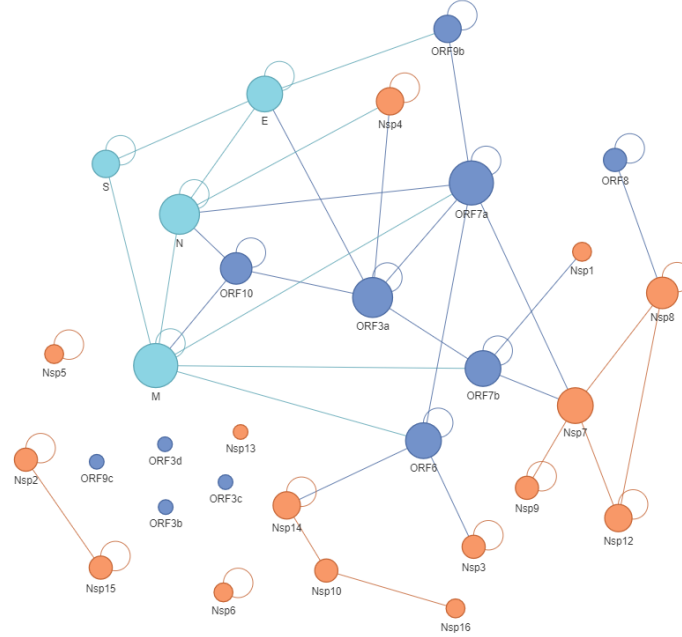

Figure 2: SARS_iRef, interactome with only viral proteins

Nodes are colored by their type. Light blue denotes structural proteins, which are spike (S), envelope (E), membrane (M), and nucleocapsid (N). Orange stands for non-structural proteins, which are *Nsp1* to *Nsp16* (not all are present in *iRefIndex*). Dark blue represents accessory proteins, which are *ORFs*. Once again, size is proportional to node degree.

The second graph, which represents the interactome of the **human lung tissue**, was taken from a reference map of the human binary protein interactome known as ***HuRI*** (*Human Reference Interactome*) [12]. The network itself was downloaded from the ***NDEx***[4] database. It was chosen because *SARS-CoV-2* is a respiratory virus and is shown in figure 3. It is made of two distinct types of nodes – hubs (green) and ends (blue).
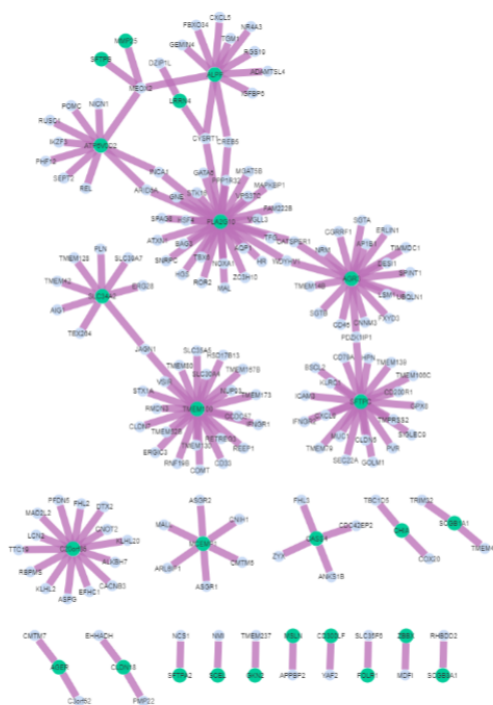

Figure 3: lungs_HuRI, interactome of human lung tissue

The additional interactomes were combined and the nature of the links in the merged network was examined. Discussion is given in section 3.

Since the idea was to analyze the protein disorder in the studied interactomes, the next step was to choose appropriate **disorder predictors**. Amino acid composition[5], *PONDR*[6] family, and charge-hydropathy (CH)[6] were the methods of choice. Number of short (at least five residues in length) and long (more than thirty residues in length) disordered regions, as well as the average disorder of neighbors ("the other side of interactions") were additionally considered. Specialized databases like *DisProt* and D^2P^2 lack many *SARS-CoV-2* proteins, so they were disregarded.

---

[4]NDEx: `https://www.ndexbio.org/#/network/289c6fe3-5fb9-11e9-9f06-0ac135e8bacf`
[5]Composition Profiler: `http://www.cprofiler.org/help.html`
[6]Predictor of Natural Disordered Regions (PONDR): `http://www.pondr.com/`

Correlation analysis was the main method used. The relationship between node degree and disorder degree of the respective protein and its neighbors according to all the listed predictors was represented on the plane, as a series of **scatter plots**. Node degree was put on the horizontal axis, while the vertical axis was reserved for the output of disorder predictors. Concrete examples and discussion are given in section 3, dedicated to the results. The entire processing, which ranges from retrieving the data to the final analyses, was done as a **Python**[7] script.

3 Results

---

[7]Python programming language: `https://www.python.org/`

References

[1] Björn H. Junker and Falk Schreiber. *Analysis of Biological Networks.* John Wiley & Sons, Inc., 2008.

[2] Pietro Hiram Guzzi and Swarup Roy. *Biological Network Analysis.* Elsevier Inc., 2020.

[3] Vladimir N. Uversky. Analyzing IDPs in Interactomes. In Birthe B. Kragelund and Karen Skriver, editors, *Intrinsically Disordered Proteins*, chapter 46, pages 895–945. Springer Science+Business Media, LLC, 2020.

[4] Kaare Teilum, Johan G. Olsen, and Birthe B. Kragelund. On the specificity of protein–protein interactions in the context of disorder. *Biochemical Journal*, 478(11):2035–2050, 06 2021.

[5] Alexander Cumberworth, Guillaume Lamour, M. Madan Babu, and Jörg Gsponer. Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochemical Journal*, 454(3):361–369, 08 2013.

[6] Chad Haynes, Christopher J Oldfield, Fei Ji, Niels Klitgord, Michael E Cusick, Predrag Radivojac, Vladimir N Uversky, Marc Vidal, and Lilia M Iakoucheva. Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes. *PLOS Computational Biology*, 2(8):1–12, 08 2006.

[7] Zsuzsanna Dosztányi, Jake Chen, A. Keith Dunker, István Simon, and Peter Tompa. Disorder and Sequence Repeats in Hub Proteins and Their Implications for Network Evolution. *Journal of Proteome Research*, 5(11):2985–2995, 2006.

[8] Gajinder Pal Singh and Debasis Dash. Intrinsic disorder in yeast transcriptional regulatory network. *Proteins: Structure, Function, and Bioinformatics*, 68(3):602–605, 2007.

[9] Ashwini Patil, Kengo Kinoshita, and Haruki Nakamura. Domain distribution and intrinsic disorder in hubs in the human protein–protein interaction network. *Protein Science*, 19(8):1461–1468, 2010.

[10] Gang Hu, Zhonghua Wu, Vladimir N. Uversky, and Lukasz Kurgan. Functional Analysis of Human Hub Proteins and Their Interactors Involved in the Intrinsic Disorder-Enriched Interactions. *International Journal of Molecular Sciences*, 18(12):2761, 2017.

[11] Ziyang Gao, Senbao Lu, Oleksandr Narykov, Suhas Srinivasan, and Dmitry Korkin. Structural Genomics and Interactomics of SARS-COV2: Decoding Basic Building Blocks of the Coronavirus. In Dmitrij Frishman and Manja Marz, editors, *Virus Bioinformatics*, chapter 7, pages 121–139. CRC Press, 2021.

[12] Katja Luck, Dae-Kyum Kim, Luke Lambourne, Kerstin Spirohn, Bridget Begg, Wenting Bian, Ruth Brignall, Tiziana Cafarelli, Francisco Campos-Laborie, Benoit Charloteaux, Dongsic Choi, Atina Cote, Meaghan Daley, Steven Deimling, Alice Desbuleux, Amelie Dricot, Marinella Gebbia, Madeleine Hardy, Nishka Kishore, and Michael Calderwood. A reference map of the human binary protein interactome. *Nature*, 580(7803):402–408, 2020.