Interfaces of Intrinsically Disordered Proteins

by

Buşra Topal

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science
in
Chemical and Biological Engineering

Koç University

March 2013

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Buşra Topal

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Prof. Dr. Özlem Keskin (Advisor)

_____

Prof. Dr. Attila Gürsoy

_____

Assist. Prof. Dr. Mehmet Sayar

Date: _____

To my mother...

Abstract

Until quite recently, the common belief was that every protein has stable tertiary structure. However, now it is widely known that some proteins or regions of the proteins can be unfolded under physiological conditions. The proteins that carry these properties are called as intrinsically disordered proteins or natively unfolded proteins. Disordered proteins play a key role in many critical biological functions. Especially for signalling and regulation, intrinsically disordered proteins are of crucial importance.

This project mainly focuses on the differences between ordered and disordered proteins and their interfaces. For this purpose, we separate the proteins in our database into two groups: (1) Disordered, (2) Ordered proteins. We do the analysis according to their structure and chemical nature. Concerning the structural comparison, the accessible surface areas and per residue areas are typical features that we use. Also, we investigate the change of hot spot residues with respect to disorder changes. For comparing the chemical nature, the ratio of the polar amino acids to non-polar amino acids in proteins and interfaces are calculated. Since low hydrophobic content is an important indicator for disordered proteins, polar/apolar ratios are used for comparing hydrophobicity. In addition, binding preferences of intrinsically disordered proteins are analyzed.

Also, another aim of this work is to create new interface type for disordered proteins. Interfaces separated into three groups according to their structures and functions. These groups are called as Type 1, Type 2 and Type 3. Thus, our aim is forming Type 4 for disordered proteins.

Özet

Yakın zamana kadar, literatürde bütün proteinlerin sabit üçüncül (tersiyer) yapıya sahip olduğu görüşü ağırlık kazanmış durumdaydı. Ancak yapılan çalışmalar bazı proteinlerin ya da protein bölümlerinin fizyolojik şartlar altında katlanmamış olarak bulunduğunu göstermektedir. Bu özelliği gösteren proteinlere doğal düzensiz proteinler ya da aslen katlanmamış proteinler denilmektedir. Düzensiz proteinler birçok önemli biyolojik fonksiyonda anahtar roller almaktadırlar. Özellikle sinyal aktarımı ve regülasyon için, düzensiz proteinler oldukça önemlidir.

Bu proje temel olarak düzenli ve düzensiz proteinler ve onların arayüzlerinin farkları üzerinde durmaktadır. Bu amaçla veri setimizdeki proteinleri (1) düzenli ve (2) düzensiz proteinler olmak üzere ikiye ayırdık. Daha sonar bu proteinleri yapısal özellikler ve kimyasal doğalarına göre inceledik. Erişilebilir yüzey alanlarını ve amino asit başına düşen alanları yapısal özelliklerini karşılaştırmak için kullandık. Ayrıca proteinlerin düzenli ya da düzensiz oluşuna göre sıcak nokta sayısının nasıl değiştiğini de inceledik. Kimyasal doğalarını karşılaştırmak için, protein ve arayüzlerdeki polar amino asitlerin apolar amino asitlere oranını hesapladık. Düşük hidrofobik içerik düzensiz proteinler için önemli bir gösterge olduğu için, hidrofobi karşılaştırmalarında polar/apolar oranı kullanıldı. Ek olarak, düzensiz proteinlerin bağlanma tercihleri de analiz edildi.

Bu projede ikinci amacımız düzensiz proteinler için yeni bir arayüz tipi oluşturmaktı. Arayüzler yapılarına ve fonksiyonlarına göre 3 gruba ayrılmaktadırlar. Bu gruplar Tip 1, Tip 2 ve Tip 3 olarak adlandırılmıştır. Bizim amacımız da düzensiz proteinlerin arayüzlerinden oluşan Tip 4'ü oluşturmaktır.

Acknowledgements

This thesis project would not have been possible without the support of many people.

At first, I would like to extend my sincerest thanks my advisors **Prof. Özlem Keskin** and **Prof. Attila Gürsoy** for their support to this work. Also, I am thankful to my thesis committee member **Assist. Prof. Mehmet Sayar** for his comments and suggestions.

I would like to express my sincere gratitude to **Prof. Ruth Nussinov** for her invaluable guidance in specifying my thesis topic and her support.

I am also deeply indebted to **Engin Çukuroğlu** for his friendship and sharing his knowledge whenever I need help.

I would also like to thank my all the friends, especially to my close friends **Pelin Atıcı**, **Tuğçe Yıldızoğlu** and **Serap Beldar** for the precious time we shared and will be shared. They made these two years more enjoyable.

I am thankful to all members of the COSBI Group members **Emine Güven**, **Halil Peynirci**, **Güray Kuzu**, **Ömer An**, **Selin Karagülle**, **Ece Acuner**, **Billur Engin**, **Gözde Kar** for their support and friendship. I am also thankful to my friend **Berat Denizdurduran** for his patient proofreading.

My infinite thanks go to **Ömer Deniz Akyıldız**. Without him finishing this thesis would not be possible. Whenever I feel depressed, he is always there for giving me courage to continue. With his help I am able to find purpose of my life.

Finally, special recognition goes out to my family, for their support, encouragement and patience. Words cannot express how grateful I am to my mother and father for all of the sacrifices that you have made on my behalf. I would also like to thank to my beloved sister for being shoulder to cry on and always cheering me up.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Disordered proteins can be defined as proteins that have not stable tertiary structure. Disordered proteins or disordered regions in proteins have not known until end of the 80's [1]. Although ordered proteins are studied intensively and comprehensively in the last decades, importance of intrinsically disordered proteins is understood recently. As a result, there are lots of studies about ordered proteins and their interactions; however intrinsically disordered proteins have received much less attention. In recent years, few works concerning disordered proteins are published [2–4]. In this work, we compare the ordered proteins with disordered proteins by promoting differences on their interfaces and surfaces.

In this thesis, we have two aims. Our first aim is to show how disordered proteins and their interfaces structurally and chemically different than ordered proteins and their interfaces. We try to show this by comparing the ordered proteins and their interfaces with intrinsically disordered proteins and their interfaces. For examining structural differences, accessible surface areas (ASA) of proteins and their interfaces is used. In addition, for showing chemical differences, hydrophobicity of proteins' surfaces and interfaces is compared. Interfaces of ordered proteins can be defined in 3 groups according to their structures and functions and global structures of proteins that interact each other. These groups are called as Type 1, Type 2 and Type 3. Details of these groups are given in Chapter 2. Our second aim is creating Type 4 that is comprised from disordered proteins. We aim to establish a characterization of a new type of interface by using our PRISM template dataset.

We process the topics as follows.

In Chapter 2, we give a literature review about types of interactions and

background information about disordered proteins. To be more precise, we explain detailed information about signatures, benefits and functions of the intrinsically disordered proteins.

In the next chapter, we present an explanation of dataset and methodology that is used to construct this thesis. First of all, we give the details of generating the dataset. By this, we aim to give an intuition about the dataset. Then DISOPRED2, which is a tool used for predicting unstructured residues of proteins, is explained in methods. At last interface and disorder features and parameters are given with explanations.

Results obtained after different analyses provided in Chapter 4. First, analysis of data is given with histograms. Then, binding preferences of disordered proteins are analyzed. Differences between ordered and disordered proteins and their interfaces are presented according to their structure properties and their chemical nature. In structure properties, accessible surface areas of these two groups are calculated and evaluated. In the final part, hydrophobicity of the interfaces and the surfaces of disordered and ordered proteins are observed.

At final chapter, discussion of the results is briefly summarized.

Chapter 2

Literature Review

2.1 Protein Interfaces and Interface Types

Proteins carry out their biological functions with forming complexes. To form complexes, proteins bind peptides, proteins, DNA and RNA. All residues that interact with target called as interface.

Interface structures are very important in protein-protein interactions since proteins interact through their interfaces. Hence, the better we understand interfaces, the better we understand mechanisms of protein interactions.

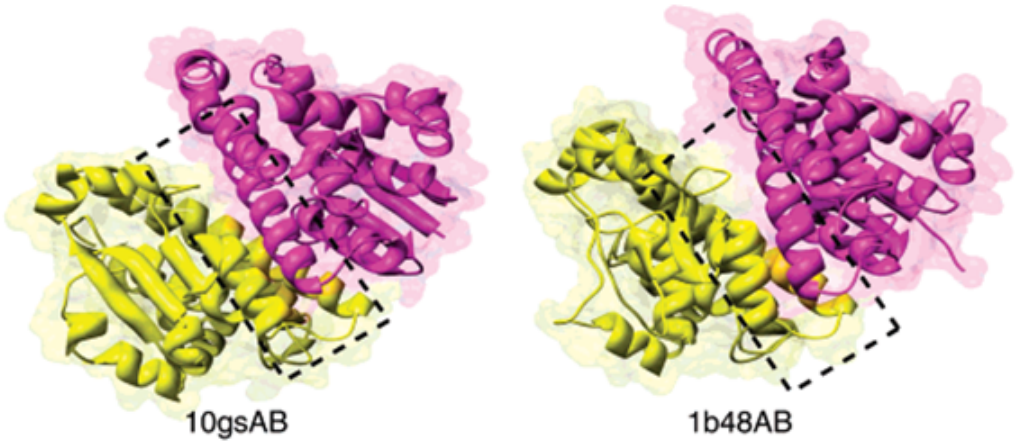

Figure 2.1: Exapmle for Type 1 [5]. Interface architecture of 10gsAB and 1b48AB similar. Also they have similar folds and functions.

Up to now, three types of interfaces are defined [5]. These characterizations are

done according to structure of the interfaces and the global structures of the interacting proteins. As seen in the Figure 2.1 members of Type 1 have similar interfaces, global folds and functions.

In Type 2 members often do not share similar functions and do not have globally similar structures, however they have similar interfaces (Figure 2.2).

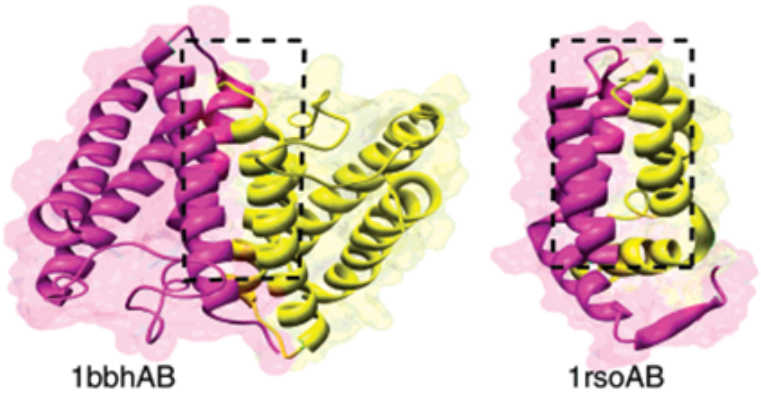

Figure 2.2: Example of Type 2 [5]. Interface architecture of 1bbhAB and 1rsoAB similar. However, their functions and folds are different.

Members of a Type 3 cluster have similar binding sites on one side of the interface, but the partner proteins are different (Figure 2.3).

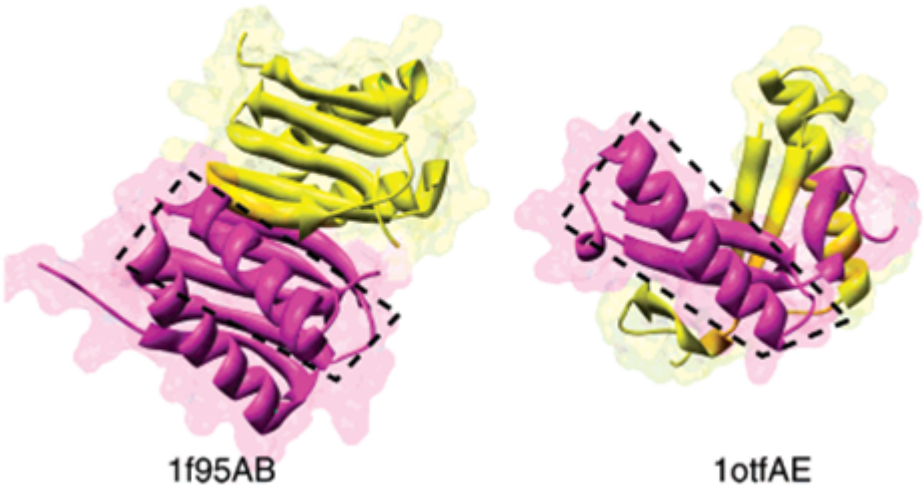

Figure 2.3: Example of Type 3 [5]. 1f95AB and 1otfAB only have similar binding site on the one side of interface. Their functions and structures of partner proteins are also different.

In this work, we will try to define an additional type, Type 4 for interfaces of disordered proteins.

2.2 Disordered Proteins

Disordered regions are defined as entire proteins or regions of proteins that lack a defined tertiary structure at neutral pH. They are also known as intrinsically disordered or natively unfolded proteins (Figure 2.4).

Disordered regions in transcriptional regulatory proteins were known since 1989 [1]. Since then an extensive amount of research about 'natively denatured/unfolded' or 'intrinsically unstructured/disordered' proteins are appeared in the literature [2–4]. Also it was shown that commonly functional proteins have disordered regions that constitute more than 50 residues [2, 4]. These functional proteins with disordered regions shows that classical protein structure- function paradigm needs to be changed [6]. Since intrinsically disordered proteins (IDPs) cannot form hydrophobic core, they cannot fold into stable 3D structures [7].

Figure 2.4: Example of fully unstructured protein [8].

There are two groups intrinsically disordered proteins can be separated. One of these groups is proteins that have regions with more than 30-40 disordered residues (Figure 2.5). The other one is natively unfolded proteins which are fully disordered (Figure 2.4) [4].

Due to the genomic analysis it is shown that percentage of the disordered proteins is higher in the complex organisms [6, 9–11]. In a work, organisms that have different complexity levels are compared [12]. Disorder in six archaeal, thirteen bacterial and five eukaryotic genomes was predicted. According to the prediction results 33.0% of eukaryotic proteins have long disordered regions. On the other hand, only 2.0% of archaeal and 4.2% of bacterial proteins have disordered regions [12]. These results are consistent with previous works that show relation between disorder and complexity of organism. Also it is shown that 35-51% of eucaryotic proteins have one or more disordered regions that have more than 50 residues [4, 9] and at 2001 11% of the proteins in Swiss-Prot were disordered [3, 9].

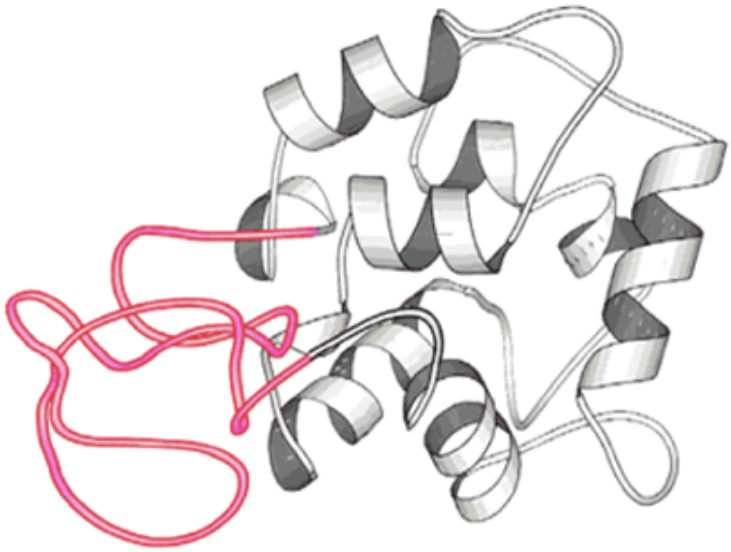

Figure 2.5: Example of partially disordered protein [8]. Some proteins are mostly folded and they have only local disorder.

In addition, as can be seen in Figure 2.6, with binding most of intrinsically disordered proteins undergo structural transition to stable 3D states and form secondary and tertiary structures [6, 13, 14]. This disorder-to-order transition makes disordered proteins suitable to bind them multiple partners. IDPs are capable of to form different 3D structures with different targets. This is called with different terms in different works as a binding promiscuity [15] or one-to-many signaling [3].

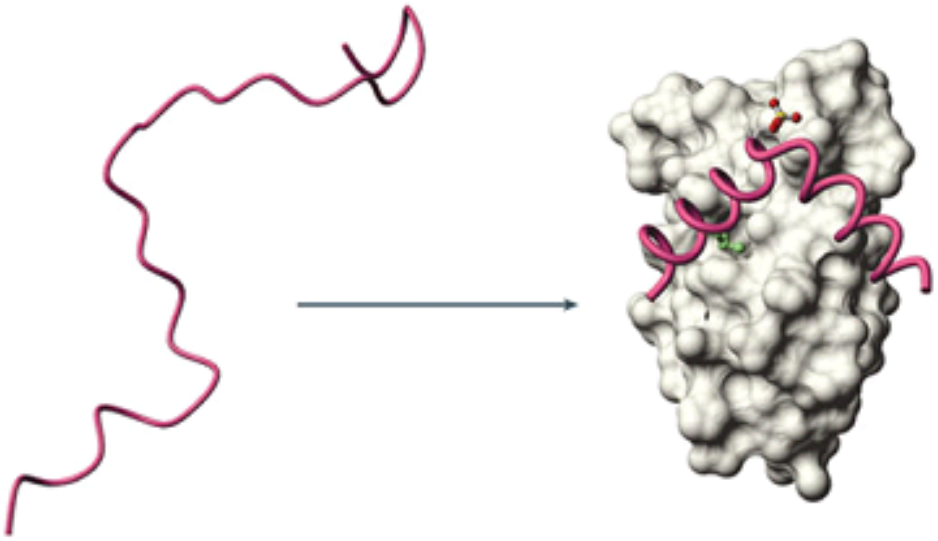

Figure 2.6: Disorder to order transition [16]. Conformation of the disordered protein changes after forming complex with another protein. After binding IDPs become more compact.

2.3 Functional Benefits of IDPs

As it was mentioned before, IDPs cannot form tertiary structures. However they can undergo disorder to order transition upon binding. With this transition IDPs can bind multiple targets and with every target their conformations can be changed. These make IDPs functionally advantageous [2, 3, 6, 14, 17].

An additional advantage of IDPs is that, their large surface area [18, 19]. Although IDPs undergo disorder to order transition, after binding, conformation of IDPs is preserved mostly. In spite of the fact that IDPs become ordered, they remain their larger surface area. Therefore, when IDPs bind their partner, their interface areas become larger compared to the ordered proteins. Also, because of disproportionately larger interface areas, IDPs can have multiple contact points [18–20]. As seen in the cyclin-dependent-kinase inhibitor p27Kip1 and cyclin A-cyclindependent kinase 2 (Cdk2) complex, having larger surface gives IDP ability of interact distant residues on partner [20]. Protein stability is determined by the size of interface. In order to make the same interface area with IDP, an ordered protein would have 2-3 times

larger protein size. Not only sequence length but also cell size would be increase % 15-30 and this would also increase cellular crowding [21]. However IDPs have larger interface area with smaller protein size which makes cells more compact.

2.4 Signatures of probable intrinsic disorder

Generally amino acid composition and size of the IDPs are different from ordered proteins that are present in the PDB [22]. It has been suggested that low mean hydrophobicity and high net charge are signatures of probable intrinsic disorder [4,17]. It is known that hydrophobic amino acid content of the protein forms the core of protein structure. Also low compositional complexity is another signature of the IDPs [23]. Thus it is expected to see low content of hydrophobic amino acids (V, L, I, M, F, W and Y) and high content of the charged and polar amino acids (Q, S, P, E, K, G and A) [24, 25].

Based on the observations on amino acid compositions of proteins, IDPs are considerably rich with P, E, K, S and Q but they have lower amount of W, Y, F, C, I, L and N from average ordered protein in PDB [3, 26]. P, E, K, S and Q are called as disorder-promoting amino acids and W, Y, F, C, I, L and N are called as order-promoting amino acids by Dunker's group [3, 26].

2.5 Functions of Disordered Proteins

It is shown that with sequence analysis, occurrence of IDPs in an organism is increasing with the complexity and nearly half of the eukaryotic proteins have extensive disordered regions [2,12]. Also, IDPs take key roles in cell. This situation makes IDPs functionally important in the cellular progress. Due to IDPs capability to make complex with binding single or multiple partners [27], IDPs make efficient interactions and make available regulation of most of the cellular process more easily [10, 28, 29]. Recent investigations indicate that IDPs involve more than 30 different types of functions which are generally essential like transcriptional and translational regulation, and cell cycle control [6, 10, 25, 28, 30].

Molecular recognition is the most important function of IDPs. This function occurs with disorder-order transition. This situation is observed firstly in site-specific

DNA binding [31] and then investigated in detail [13, 14]. To be more precise, transcription, transposition, packaging, repair and replication processes that proteins with long disordered regions take role and make them easier [32]. Also it is known that with disorder-order transition conformational entropy is decreasing largely.

Beside signaling, cell cycle control, development, multiprotein complex assembly and endocytosis are other processes which are either done with the help of the functional disordered proteins or proteins with disordered regions. Moreover some proteins regulated by phosphorylation have disordered regions [12]. Also it is shown that cancer-associated proteins contain disordered structure [10].

Furthermore, RNA and protein chaperones have unstructured regions commonly. As a matter of fact RNA chaperones have higher disordered residue percentage than other protein classes. It is assumed that disordered segments in chaperones may be crucial for the function of the chaperones [33].

2.6 Prediction and Identification of Protein Disorder

Generally to identify disordered residues X-ray crystallography is used. If there is a missing electron density of backbone residues in three dimensional structures, these residues are accepted disordered [3]. There are also several tools for identifying the lack of structures. Heteronuclear multidimensional nuclear magnetic resonance [7], far-UV circular dichroism (CD) spectroscopy which is used for detecting amount of lack of structure, near-UV CD, UV spectroscopy, Fourier transform infrared spectroscopy are some examples of detection methods of protein disorder.

There are also lots of studies for developing algorithms to predict disordered regions accurately [8, 28, 34, 34–42]. In addition there is a database of protein disorder (DISPROT) which is updated frequently [8]. With using algorithms that was developed for predicting disordered regions from amino acid sequences, many software tools are developed. Numbers of the predictors are increasing rapidly year by year. DISOPRED2 [12], FoldIndex [43], DisEMBL [39], GLOBPLOT 2 [44, 45], PONDR [24, 46, 47], RONN [48], NORSp [49] and IUPred [50] are some examples of the these software tools.

Predictors use different ways to predict disorder. Every predictor has advantages and disadvantages. However, generally they give similar results but they differ in

details according to parameters they chose to use [51–53]. And yet none of them gives %100 reliable results. Also it is observed that combining results of these methods gives accurate results [51].

However, when we look over predictors, they can be separated into two groups. First type of predictors use disorder datasets for training, and second type of predictors use charge/hydropathy method [53]. PONDR, Globplot, DisEMBL, Disopred2, RONN can be given as an examples for first group. FoldIndex, NORSp and IUPred are examples of second group.

As it is mentioned before, lack of structure is predicted from amino acid sequence. Low hydrophobicity [54] and compositional complex, high net charge and flexibility [32] are characteristics of IDPs. Furthermore density of some amino acids in disordered regions is lower or higher than density of amino acids in ordered regions. To be more precise, Q, S, P, E, K, G and A are the amino acids that are found in disordered regions with high frequency [24, 25]. Success of the algorithms and programs that predict disorder from amino acid sequence is evidence of this situation [8,12,34,35,37–40,42,55].

2.7 Hot Spots

In protein complexes, some amino acids make more contribution to the free binding energy. If this contribution is more than 2kcal/mol for an amino acid, this residue called as hot spot [56–58]. Hot spots are more important than other residues in interfaces. It is shown that tryptophan, arginine, tyrosine, aspartic acid, proline and histidine are most frequent amino acids that found as a hot spots [58]. Alanine mutations used as an experimental methods for identification of hot spots.

Chapter 3

Methods

3.1 Introduction

Disordered proteins cannot form stable 3D structures. However generally when they bind to target proteins, disorder to order transitions occur and proteins become more stable (Figure 3.1) [6,13,14,59]. Their specifications like having larger surface area, low mean hydrophobicity and high net charge, low compositional complexity give a different characterization to disordered proteins when compared to ordered proteins [4, 17, 22]. Naturally, the expectation is that, these differences lead to important differences in the sense of structural properties and chemical nature.

In this work, we compare the disordered proteins and their interfaces with ordered proteins and their interfaces. In order to compare proteins, we form two groups of proteins. First group consists of disordered chains. This group comprises chains that have more than 80% disordered residues. Second group is ordered chains. Members of this group consists smaller chains of complexes that have 0% disordered residue.

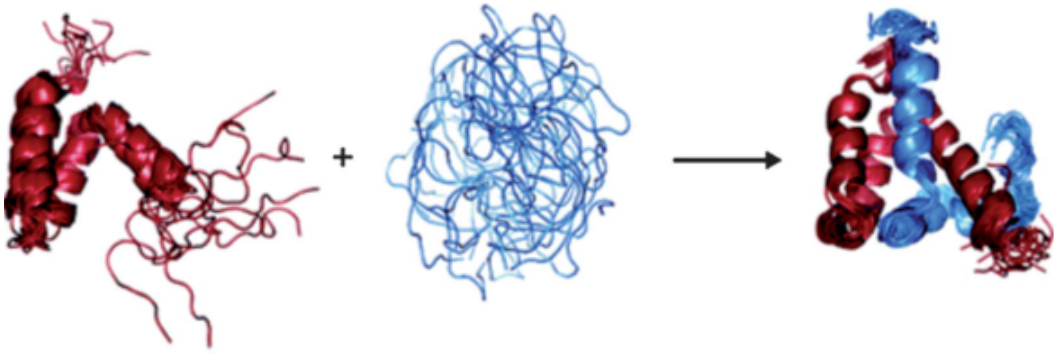

Figure 3.1: The disorder to order transition of ACTR and CBP with binding [59].

By this comparison, we aim to characterize a different type of protein complexes and thus interfaces. Note that, there are three types of protein interfaces, which are already defined. By using proteins whose chains comprises high ratio of disordered residues, we aim to characterize a different type of protein interfaces that we will call as Type 4.

3.2 Interface Dataset

In order to form Type 4, we need a dataset to work on and the interface dataset used in this study is template set of the PRISM. In Keskin's work, for generating this dataset, 49512 two chain interfaces extracted from PDB from available protein complexes. For grouping proteins into clusters, structures of the proteins were compared. 8205 interface clusters that contain interfaces with similar architectures were formed. For each cluster one representative interface were selected. This representative interface share similar interface structure with members of the cluster [60]. In this work these 8205 protein interfaces are selected as the non-redundant interface dataset.

3.3 Prediction of Disorder in Proteins

Another important problem, which is central for defining the disordered proteins, is how can we tackle the prediction of disorder in proteins, since we first need to determine that a protein have high disorder. Then, we move to examine the structural properties of these proteins.

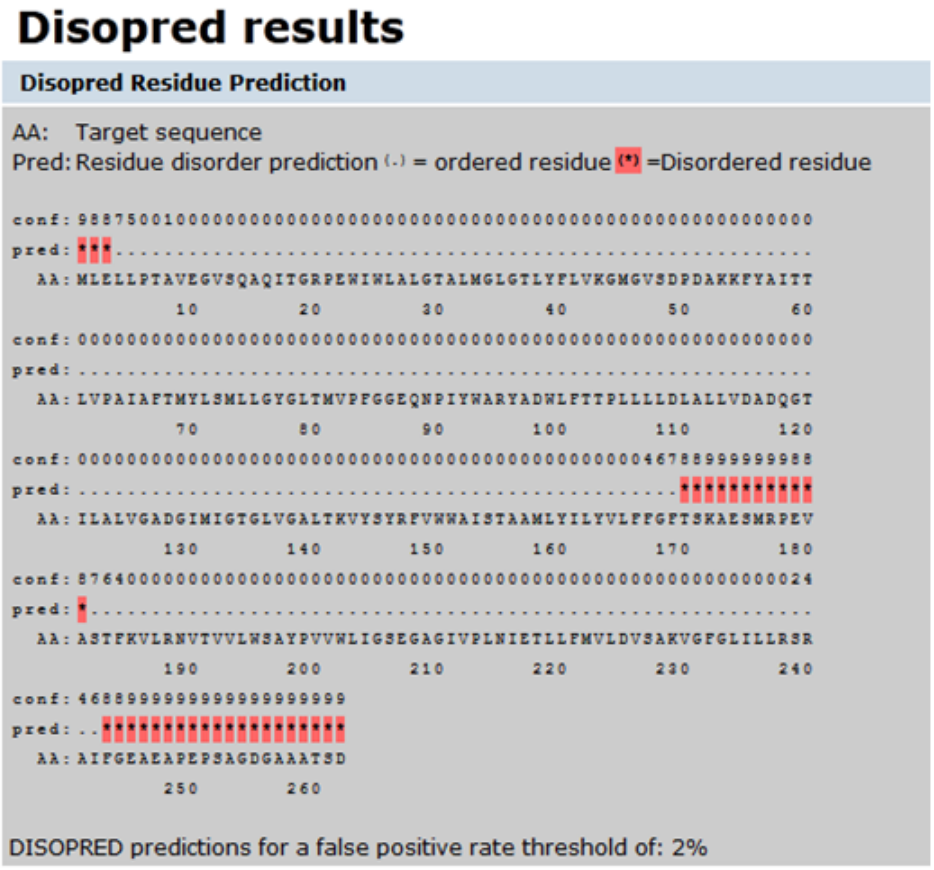

Figure 3.2: Example of DISOPRED2 results ('*' shows disordered residues, '.' Shows ordered residues).

In order to overcome this problem, DISOPRED2 is used for the prediction of disordered residues in proteins [34]. DISOPRED2 uses sequences for prediction and gives residue based results (Figure 3.2). For predicting disordered residues, 750 non-redundant sequences were used for training DISOPRED2. Proteins in this set have high resolution X-ray structures. In addition, disordered residues in this set of proteins were identified from missing coordinates from electron density map [12]. Although this method leads to false positives, it is the best and simplest way to predict disorder.

In order to predict disorder residues with DISOPRED2, PDBs of proteins are mapped to the UNIPROT IDs. Sequences of proteins are taken from the UNIPROT. Sequences of the chains are given to the DISOPRED2. Residue based results are

obtained. These results are mapped to the PDBs again for making structure based comparisons.

According to prediction results, 31 chains from the dataset have more than 80% disordered residues. On the other hand, 934 of them have not got any disordered residues. In results chapter, analysis of data was done in detail.

3.4 Prediction of Hot Spots

In this part, HotRegion (A database of predicted hot spot clusters) is used for detecting hot spots in interfaces [61]. After detecting, percentages of the hot spots in the interfaces are calculated. For comparison, same number of ordered protein complexes with the disordered protein complexes selected randomly from all ordered protein complexes.

3.5 Interface and Disorder Features

In this section, we introduce the parameters, which are used in comparison of disordered and ordered proteins.

3.5.1 Structure Properties

Surface area per residue: The ratio of the surface area to number of the surface residues in the chain.

Interface area per residue: The ratio of the interface area to number of the interface residues in the chain.

$DI\delta ASA$: δASA of the interface in the disordered complexes

$OI\delta ASA$: δASA of the interface in the ordered complexes.

$DS\delta ASA$: δASA of the surface in the disordered complexes.

$OS\delta ASA$: δASA of the surface in the ordered complexes.

3.5.2 Chemical Nature

Polar amino acid percentage of interfaces: The ratio of polar amino acids to the all amino acids in interface. It is used as measure for hydrophobicity.

Polar amino acid percentage of chains: The ratio of polar amino acids to the all amino acids in chain.

Total hydropathy index of interfaces: Sum of the hydropathy indexes of all amino acids in the interface.

Total hydropathy index of proteins: Sum of the hydropathy indexes of all amino acids in the chain.

3.6 Conclusion

In this chapter, we briefly stated the problem of defining a new type of protein. Then, we explained the datasets we used in order to characterize Type 4. In the next section, we present the results that are obtained from the methods explained in this section.

Chapter 4

Results

4.1 Analysis of Data

As mentioned before PRISM template dataset is used for this analysis. After predictions are done with DISOPRED2 for the PRISM template dataset, disordered residues of proteins were obtained. DISOPRED2 is a downloadable server that predicts protein disorder and gives residue-based results.

When fractions of disordered residues were calculated, we see that dataset mainly consist of ordered proteins. In Figure 4.1 and Figure 4.2, we can easily see that majority of protein complexes and chains have only 0-10% disordered residue fraction. Because of high ordered protein proportion in dataset, it is hard to get good results. Different thresholds for disorder residue fraction are used for deciding whether a protein is disordered or not. Although the number of proteins, which is decided as disordered, are increased with lower thresholds, specifications get closer to ordered proteins. For getting sufficient amount of IDP data and getting better results, threshold of disorder residue fraction was specified as 80% after several trials.

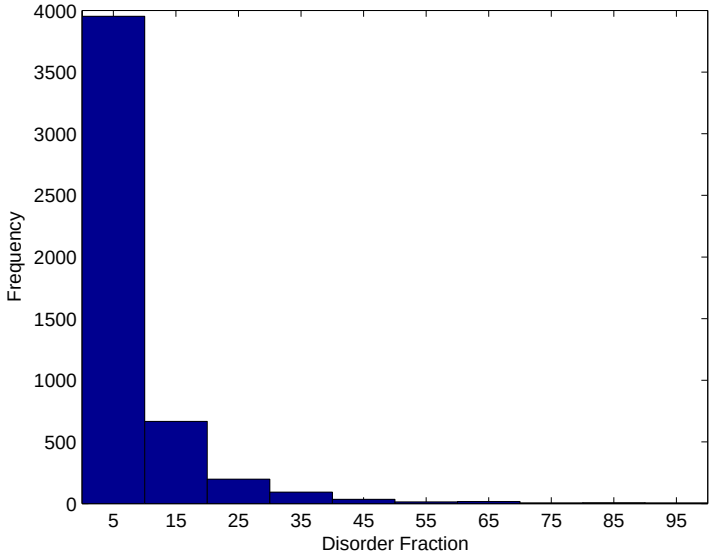

Figure 4.1: Disordered residue fraction in complexes in our dataset.

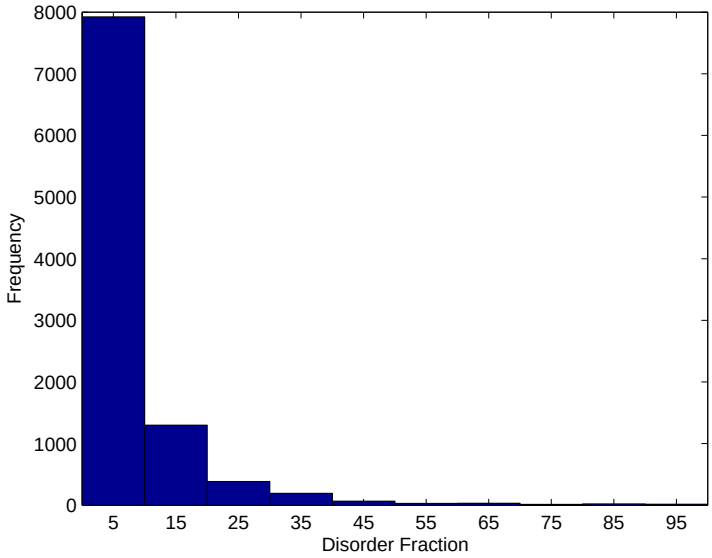

Figure 4.2: Disordered residue fraction in chains in our dataset.

With using this data two groups were created. These groups are ordered chains group and disordered chains group. In disordered chains group, number of disordered

chains that have more than 80% disordered residue fraction is 31. To create ordered chains group, 150 ordered chains selected from the dataset for making more reliable comparison with disordered chains.

4.2 Binding Preferences of Disordered Proteins

When we examine quaternary structure of the proteins in dataset, we can easily see in the Figure 4.3 that majority of proteins in dataset are homodimer. Homodimer complexes are formed by two identical chains. As expected, disordered residue fractions are similar in homodimer proteins. Thus, when we compare chains of the homodimer proteins, we can see the diagonal, which is formed by chains have same disorder fraction, in the Figure 4.3.

Surprisingly, in heterodimer proteins, we compared the disordered fractions of two chains of a protein and we have found that proteins that have higher disordered residue fraction tend to bind proteins that have lower disordered residue fraction (Figure 4.4).

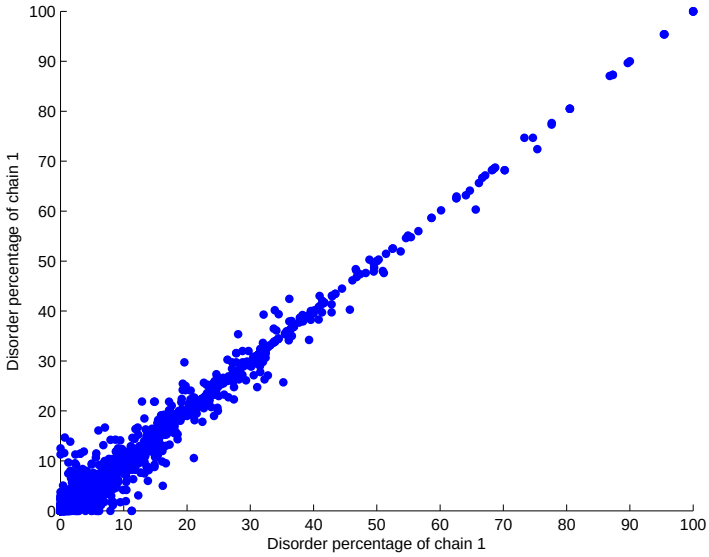

Figure 4.3: Disordered residue fraction comparison of chains of homodimer proteins in dataset.

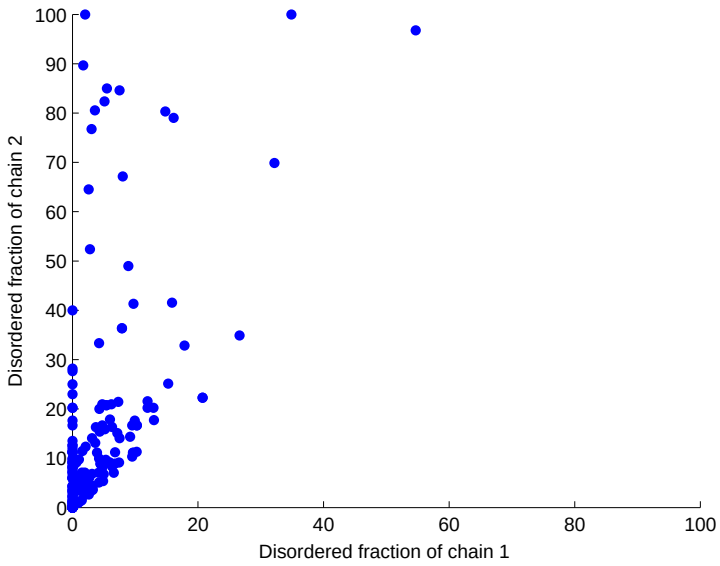

Figure 4.4: Disordered residue fractions of chains of heterodimer proteins.

In Table 4.1 it can be seen easily that, proteins that have more than 50% disorder residue fraction tends to bind proteins that have very low disorder residue fraction. In this Table 4.1 chain 1 and chain 2 are determined according to alphabetical order of the chain names.

4.3 Structure Properties

To understand differences between ordered proteins and IDPs, firstly geometrical analyses were done. For deciding threshold of disorder residue fraction, same analyses were done with different thresholds from 40% to 80%. After these calculations, 80% disordered residue fraction is decided as a threshold.

It is known that when an IDP binds to another protein, it becomes more stable and forms secondary structures after disorder to order transitions. Since, IDPs have 3D structures in a complex, their architectures have similarities with ordered proteins. However, their structure properties have significant differences. In this section we aim to show these differences between IPDs and ordered proteins.

Figure 4.5 shows the difference between interface areas of IDPs and ordered chains

Table 4.1: Heterodimer proteins that have more than 50% disordered residue fraction

			Disorder Fraction(%)	
PDB ID	Chain 1	Chain 2	c1	c2
1o3p	A	B	100.000	2.033
2b2w	A	C	100.000	34.884
1pnb	A	B	96.774	54.667
1qd6	A	C	84.615	7.500
1c5n	L	H	80.556	3.571
1htr	P	B	76.744	3.040
1f3m	A	C	67.143	8.014
1avf	P	A	52.381	2.795
1ef1	B	D	1.730	89.655
1ncq	B	D	5.490	85.000
1hxs	3	4	5.106	82.353
1d4m	1	4	14.789	80.328
1pvc	1	4	16.129	79.032
1rke	A	B	32.171	69.886
1mec	3	4	2.597	64.516
1pnb	A	B	96.774	54.667

as a box plot. For getting more reliable results, data analyzed with Matlab's statistic tool Anova (Analysis of variance). According to P values, interface and surface areas and per residue areas of IDPs are significantly different than ordered proteins. As it can be followed from the Figure 4.5 and Figure 4.6, interface areas and surface areas of IUPs are larger than ordered proteins. To find out reason of this situation, per residue areas of interfaces calculated. It is clearly shown that, per residue areas of IDPs are larger than ordered proteins Figure 4.7 and Figure 4.8. This can be explained as IDPs form larger areas with same chain length, instead of using larger part of surfaces to interact with other proteins.

To claim this result, same calculations were done for surface areas of IDPs and ordered proteins. As expected, IDPs have larger surface and per residue surface areas when compared to the ordered proteins.

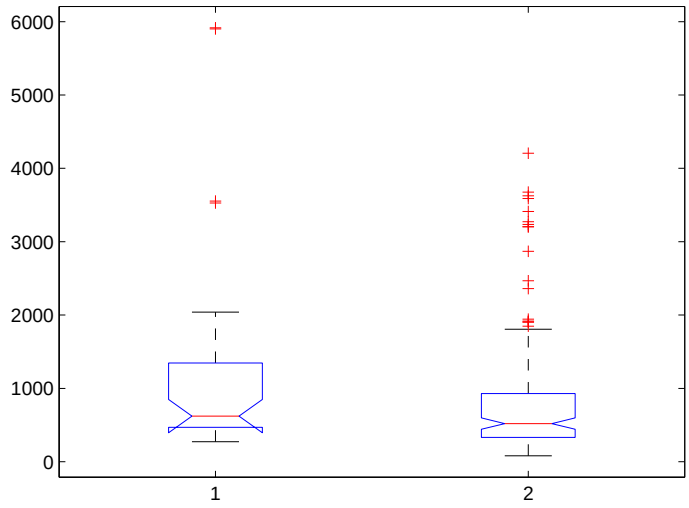

Figure 4.5: Box plot of interface areas. 1= IDPs 2=ordered proteins.

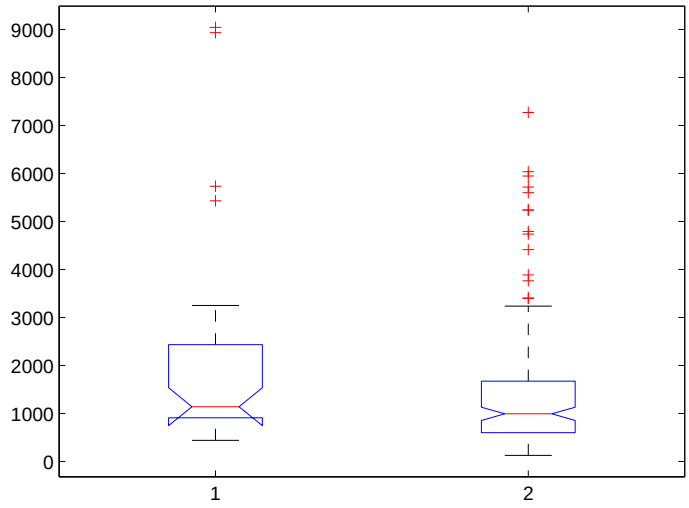

Figure 4.6: Box plot of surface areas. 1= IDPs 2=ordered proteins.

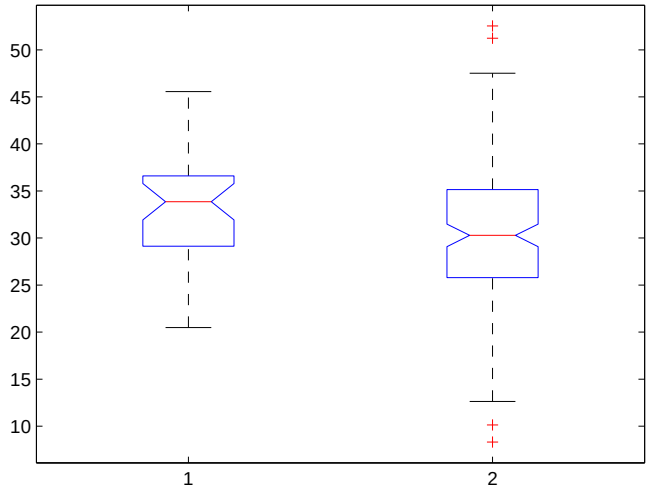

Figure 4.7: Box plot of per residue interface areas. 1= IDPs 2=ordered proteins.

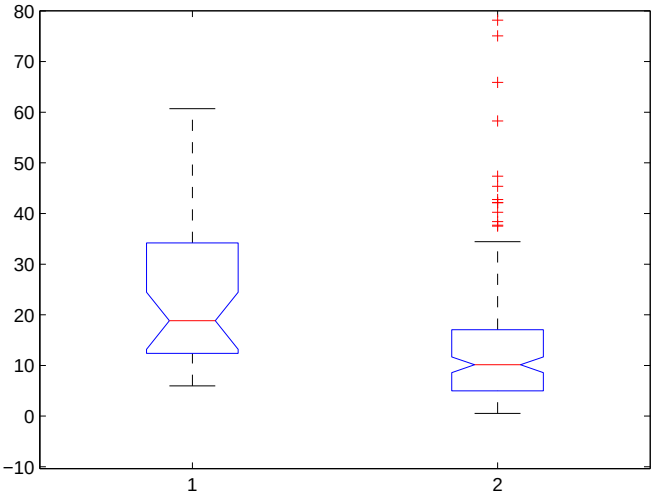

Figure 4.8: Box plot of per residue surface areas. 1= IDPs 2=ordered proteins.

When we plotted per residue surface area with respect to chain length of the protein (Figure 4.9), we see that IDPs that have shorter chains have larger per

residue areas. Also, ordered proteins become more compact when they have larger chains. Therefore, to form same surface area with IDPs ordered proteins need more residues than disordered proteins.

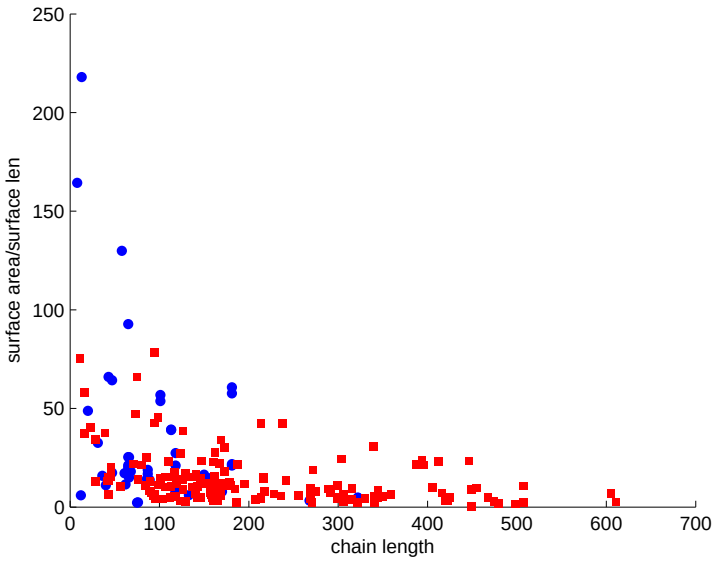

Figure 4.9: Per residue surface area against chain length. Red squares represent the ordered proteins and blue circles represent IDPs.

4.3.1 Hot spots in Interfaces

In interfaces some residues make more contribution to binding free energy. These residues are more important than others and they called as hot spots. In this part, we want to observe that how amount of the hot spots change in ordered proteins and IDPs. For this purpose, hot spots are detected with using Hotregion. For dataset, our disordered complexes and randomly selected same number ordered complexes are used.

Table 4.2: Hot spot fractions of IDPs and ordered proteins

	IDPs		Ordered Proteins	
	Complex	Disordered Chain	Complex	Ordered Chain
Average	0.206	0.158	0.249	0.269
S. Deviation	0.162	0.151	0.174	0.192

According to Table 4.2 interfaces of ordered complexes have more hot spots than

IDPs. However, number of samples is not high enough and standard deviation is slightly high. Thus we need more reliable results to draw an inference.

4.4 Chemical Nature

Because of structure properties of IDPs and ordered proteins different from each other, it can be expected that their chemical nature differ too. As mentioned before, it is known that IDPs lack of enough hydrophobic residues to form compact 3D structures. Therefore, hydrophobicity of ordered proteins and IDPs is examined. For calculating hydrophobicity, we count polar (R, N, D, E, Q, H, K, S, T, Y) and apolar (A, C, G, I, L, M, F, P, W, V) residues and we divide polar residue number to apolar residue number. Apolar residues are showing hydrophobic property and polar residues are showing hydrophilic property. In accordance, while p/a ratio of proteins decreases, hydrophobicity increases.

4.4.1 Hydrophobicity of Interfaces

As can be seen in the Figure 4.10, interfaces of the ordered proteins have higher polar/apolar ratio. According to this result we can say that interfaces of IDPs are more hydrophobic than ordered proteins. From Figure 4.10, we can claim that IDPs are using their hydrophobic residues for interact target proteins instead of forming hydrophobic core.

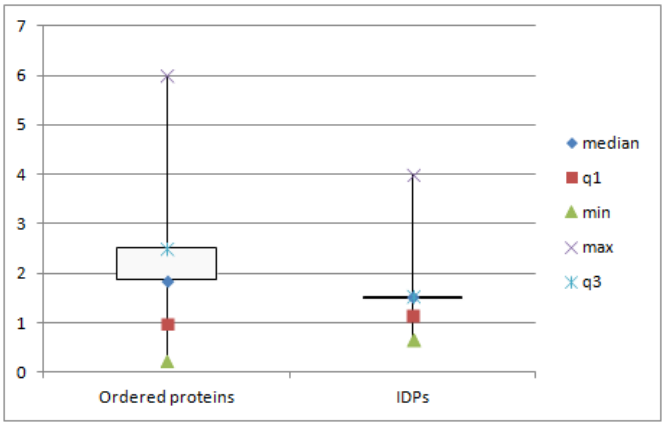

Figure 4.10: Hydrophobicty of interfaces.

However, when we look at the surface of the proteins, we observe that disordered proteins have higher polar/apolar ratio(Figure 4.11). As expected, ordered proteins are more hydrophobic than IDPs. Thus, we can mention that IDPs prefer using their hydrophobic residues for interactions not for forming compact 3D structure.

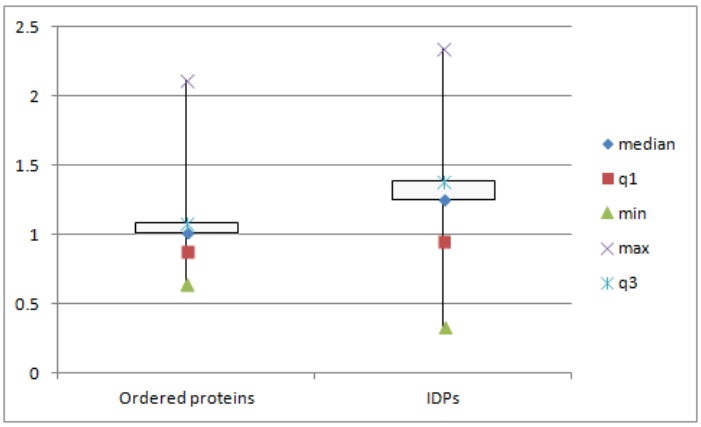

Figure 4.11: Hydrophobicty of surfaces.

4.5 Type 4 Proteins

In conclusion, in this work we showed that IDPs and their interfaces are different from ordered proteins. These differences can be seen in their structure properties and chemical nature. We also mention that before, interfaces can be separated into three groups according to their structures. Since we observed that interfaces of unstructured proteins have significant properties, we can describe Type4 for these proteins. List of proteins in Type 4 can be seen in Table 4.3.

Table 4.3: Type 4: Interfaces of Disordered proteins

Disordered Chain Names			
complex	chain	complex	chain
1D7MAB	1D7MB	1D7MAB	1D7MA
1DYNAB	1DYNB	1DYNAB	1DYNA
1HBXBE	1HBXE	1O3PAB	1O3PA
1ZY7AB	1ZY7B	1HBXBE	1HBXB
1GCPAD	1GCPD	2B2WAC	2B2WA
1GCPBD	1GCPD	1PNBAB	1PNBA
1A1RCD	1A1RD	1GCPBD	1GCPB
1EF1CD	1EF1D	1GCPAD	1GCPA
1EF1BD	1EF1D	1EF1CD	1EF1C
1P93AC	1P93C	1P93AC	1P93A
1P93CD	1P93D	1P93CD	1P93C
1SJ7AB	1SJ7B	1SJ7AB	1SJ7A
1NCQBD	1NCQD	1QD6AC	1QD6A
1HXS34	1HXS4	1XNIAB	1XNIA
1XNIAB	1XNIB	1XNIAE	1XNIA
1XNIAE	1XNIE		

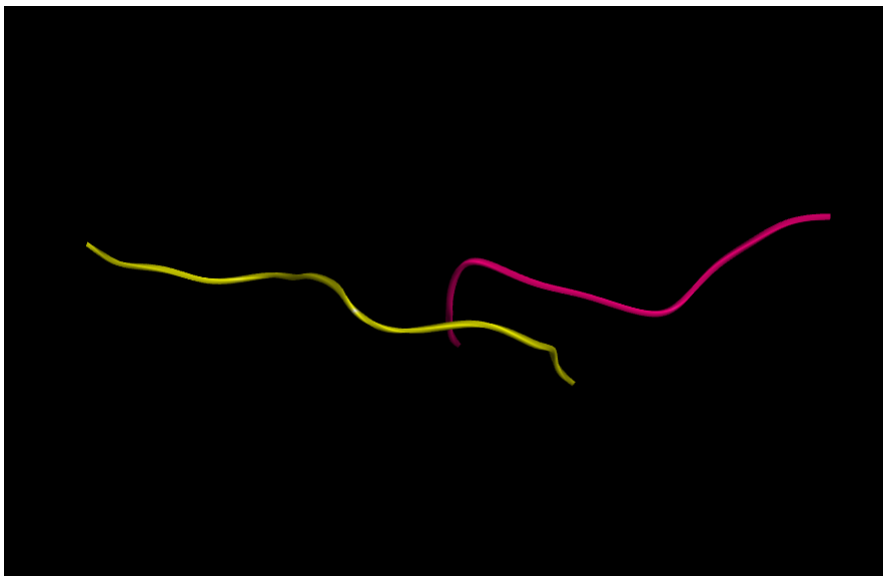

Figure 4.12: NS3 protease domain complexed with a synthetic NS4A cofactor peptide (PDB ID:1A1R). According to prediction results 1A1R_C (yellow) and 1A1R_D(magenta) are both disordered.

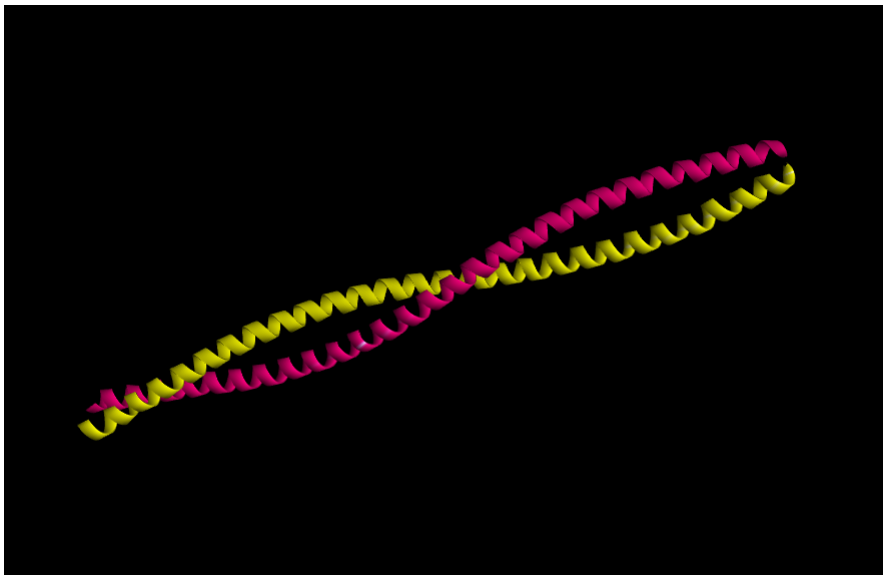

Figure 4.13: Coiled-coil dimerization domain from cortexillin (PDB ID: I1D7M). According to prediction results 1D7M_A (yellow) and 1D7M_B (magenta) are both disordered.

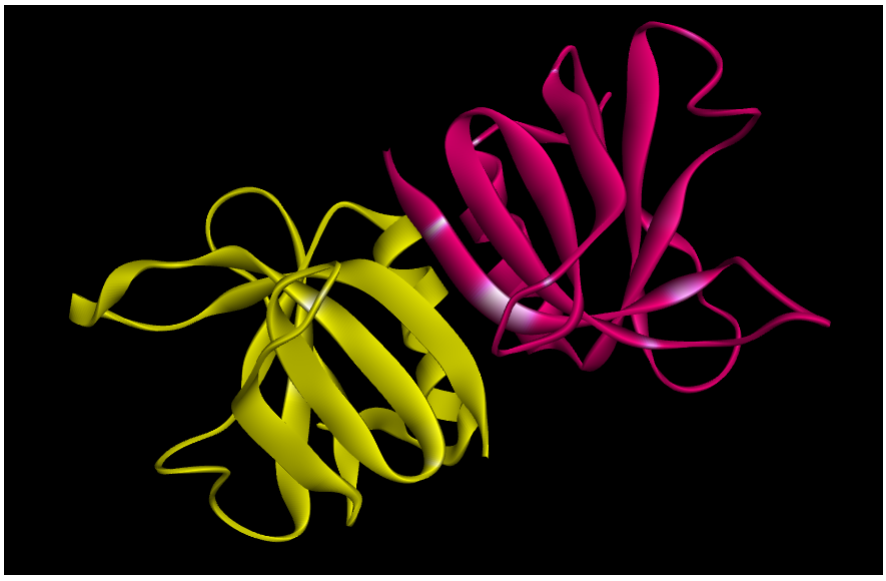

Figure 4.14: Pleckstrin homology domain from human dynamin (PDB ID: 1DYNAB). According to prediction results 1DYN_A (yellow) and 1DYN_B (magenta) are both disordered.

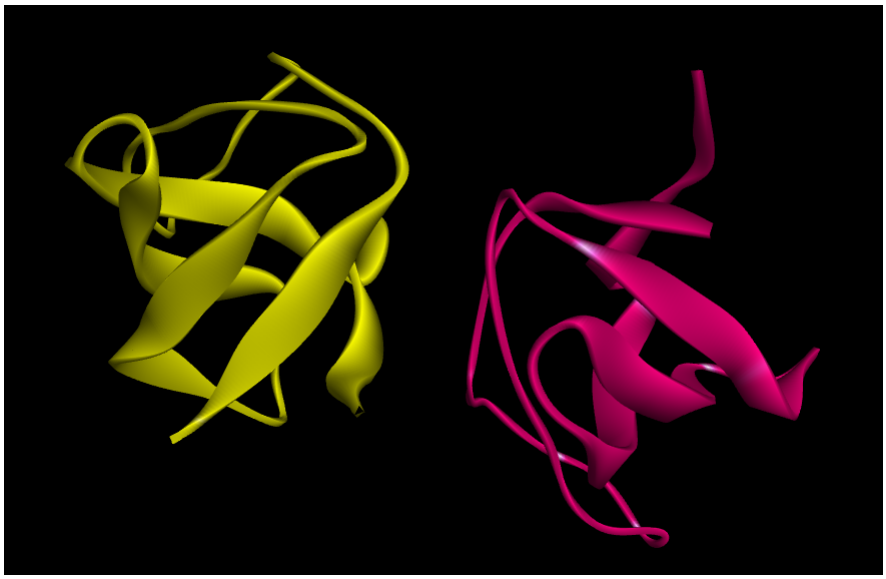

Figure 4.15: Vav SH3 domain (PDB ID: 1GCP). According to prediction results 1GCP_B (yellow) is disordered and 1GCP_D (magenta) is ordered.

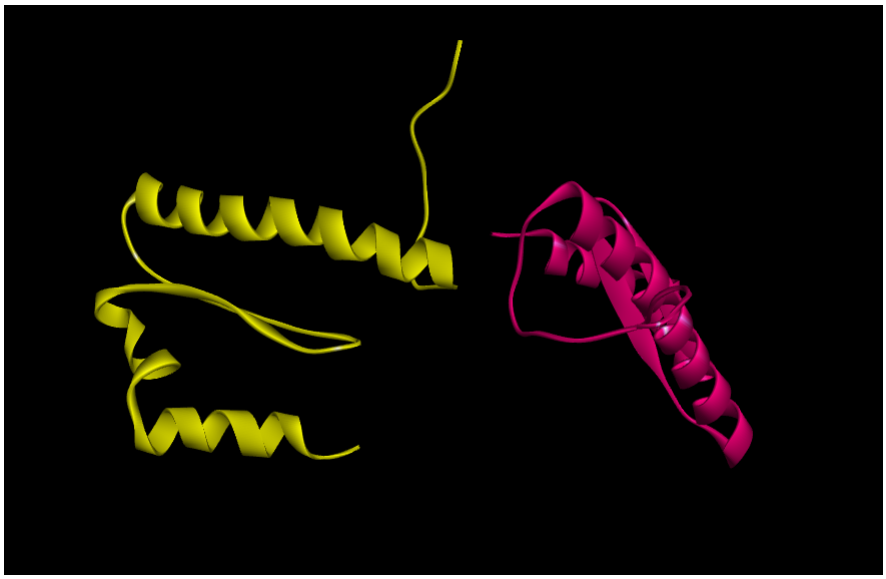

Figure 4.16: Ternary complex of Sap-1 and Srf (PDB ID: 1HBX). According to prediction results 1HBX_B (yellow) and 1HBX_E (magenta) are both disordered.

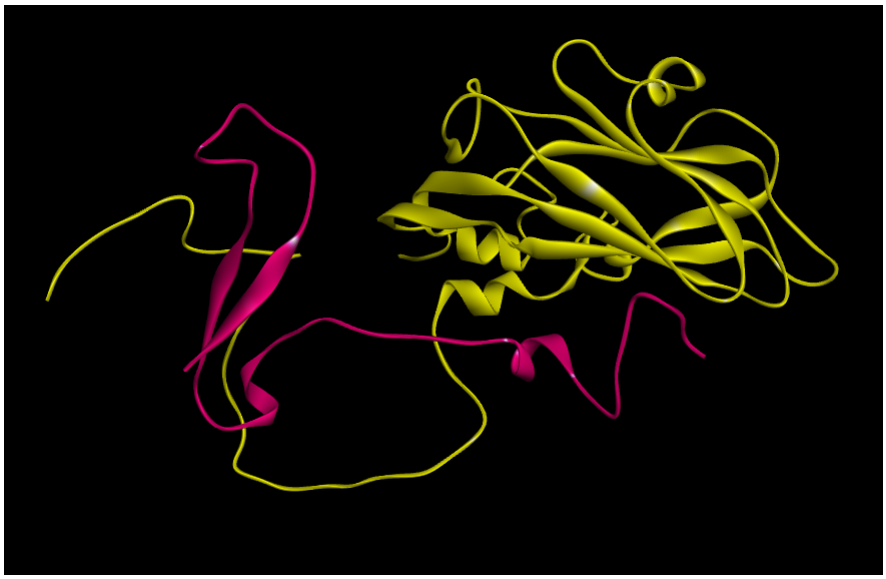

Figure 4.17: Crystal structure of mahoney strain of poliovirus (PDB ID: 1HXS). According to prediction results 1HXS_3 (yellow) is ordered and 1HXS_4 (magenta) is disordered.

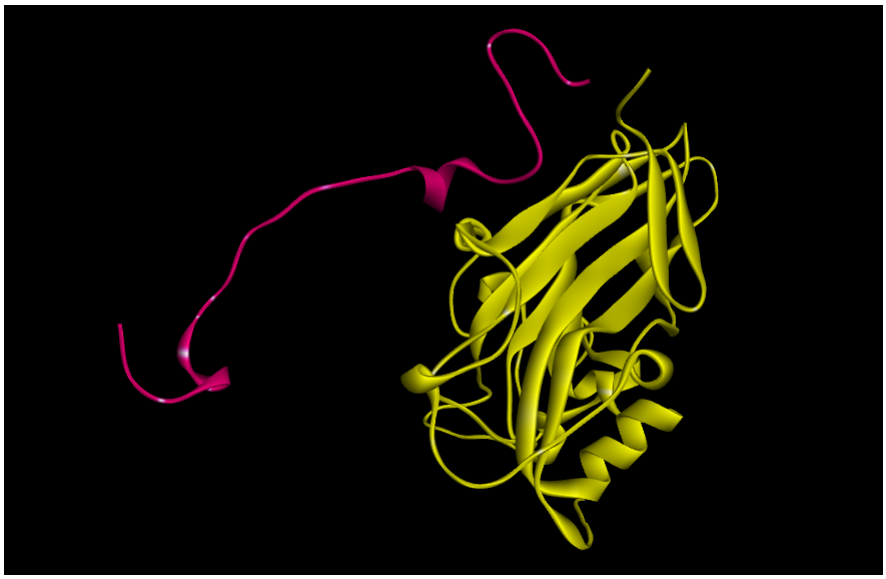

Figure 4.18: The structure of HRV14 complexed with pleconaril (PDB ID: 1NCQ). According to prediction results 1NCQ_B (yellow) is ordered and 1NCQ_D (magenta) is disordered.

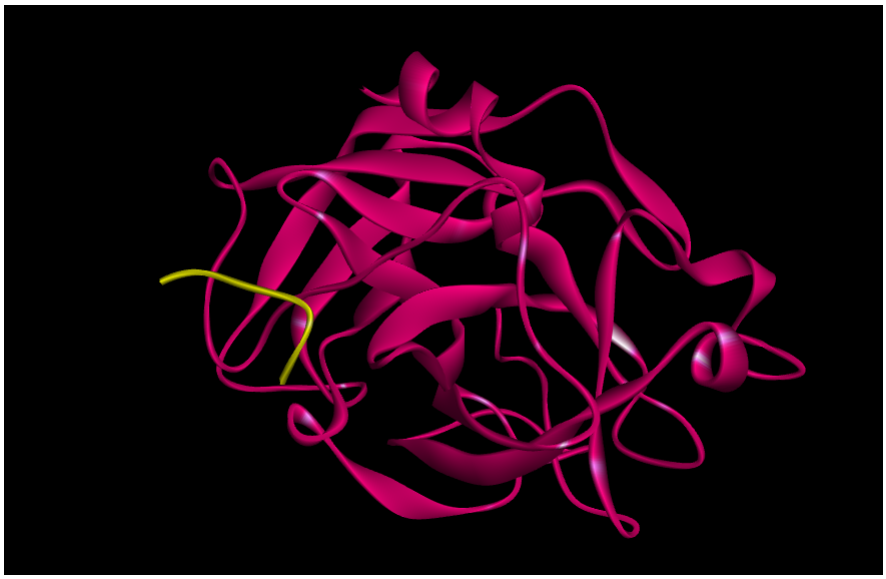

Figure 4.19: Elaborate Manifold of Short Hydrogen Bond Arrays Mediating Binding of Active Site-Directed Serine Protease Inhibitors (PDB ID: 1O3P). According to prediction results 1O3P_A (yellow) is disordered and 1O3P_B (magenta) is ordered

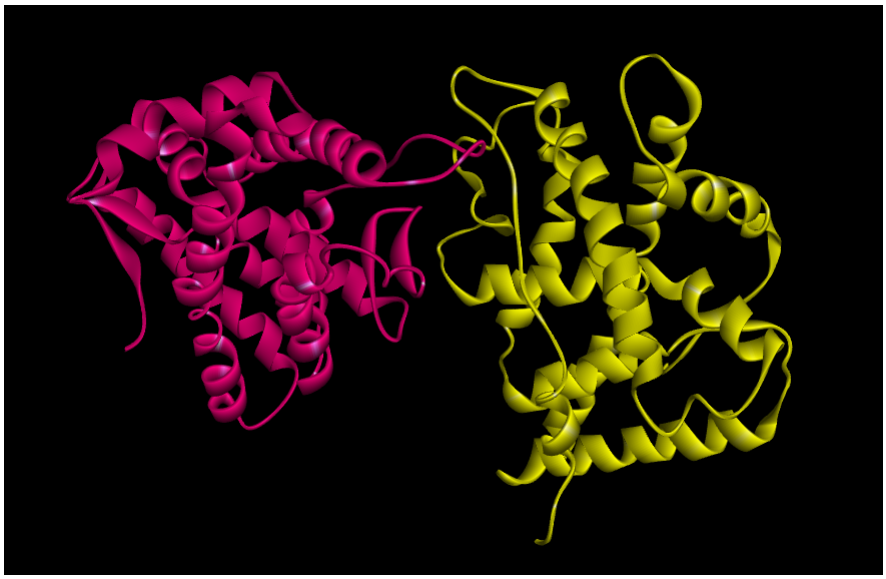

Figure 4.20: Agonistic form of the glucocorticoid receptor ligand-binding domain (PDB ID: 1P93). According to prediction results 1P93_A (yellow) and 1P93_C (magenta) are both disordered.

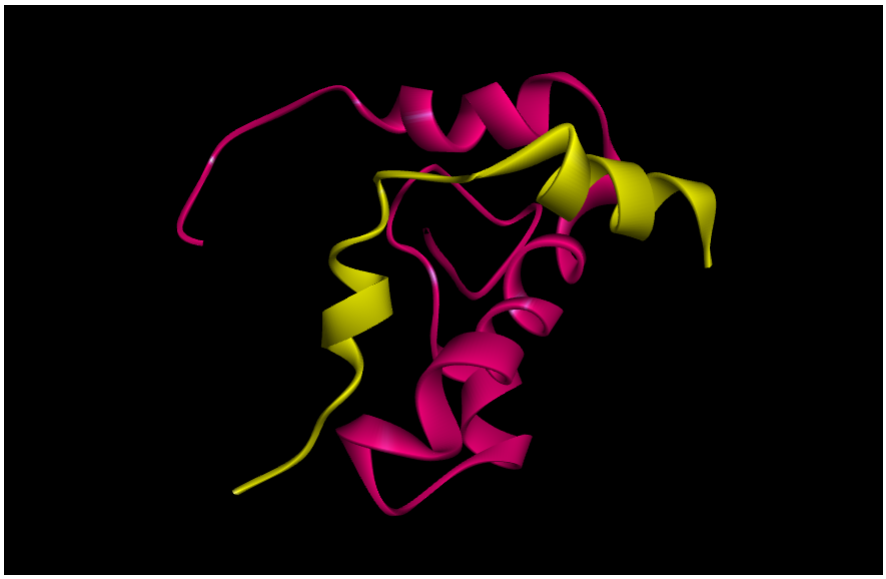

Figure 4.21: Structure of napin BnIb (PDB ID: 1PNB). According to prediction results 1PNB_A (yellow) is disordered and 1PNB_B (magenta) is ordered.

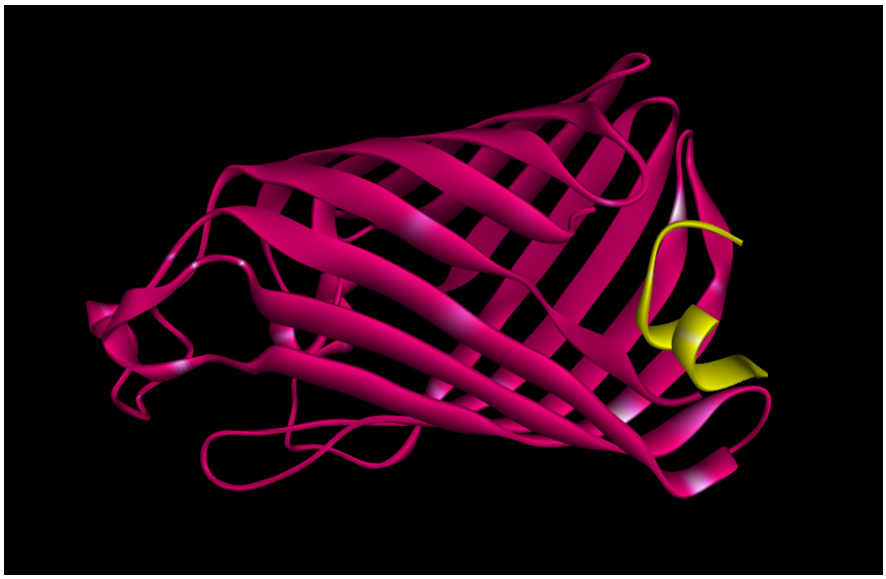

Figure 4.22: Outer membrane phospholipase A from e. coli (PDB ID: 1QD6). According to prediction results 1QD6_A (yellow) is disordered and 1QD6_C (magenta) is ordered.

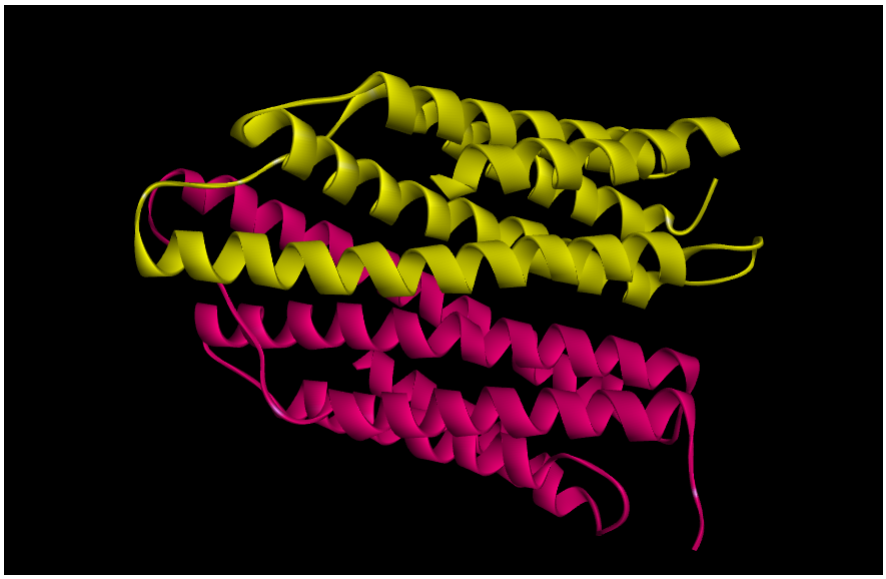

Figure 4.23: Crystal Structure of Talin Rod 482-655 (PDB ID: 1SJ7). According to prediction results 1SJ7_A (yellow) and 1SJ7_B (magenta) are both disordered.

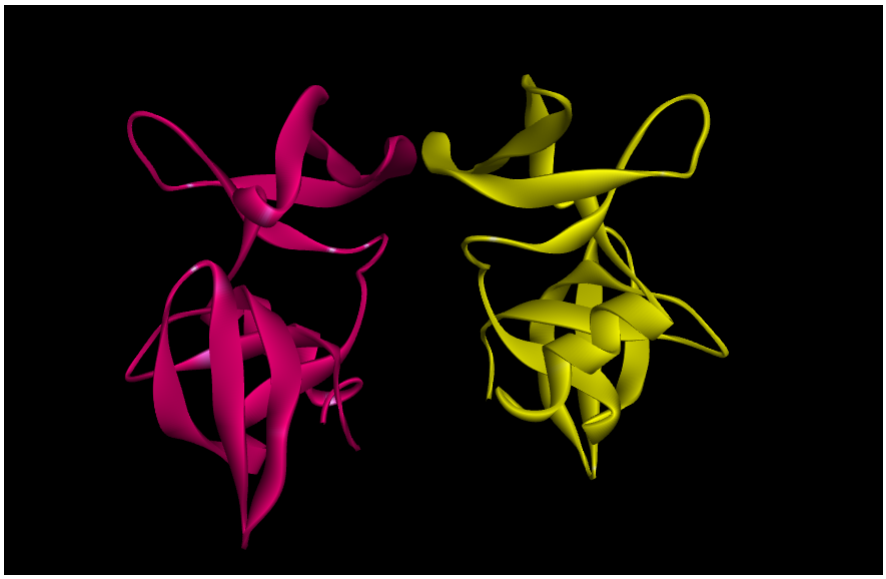

Figure 4.24: Tandem Tudor Domain of 53BP1 (PDB ID: 1XNI). According to prediction results 1XNI_A (yellow) and 1XNI_B (magenta) are both disordered.

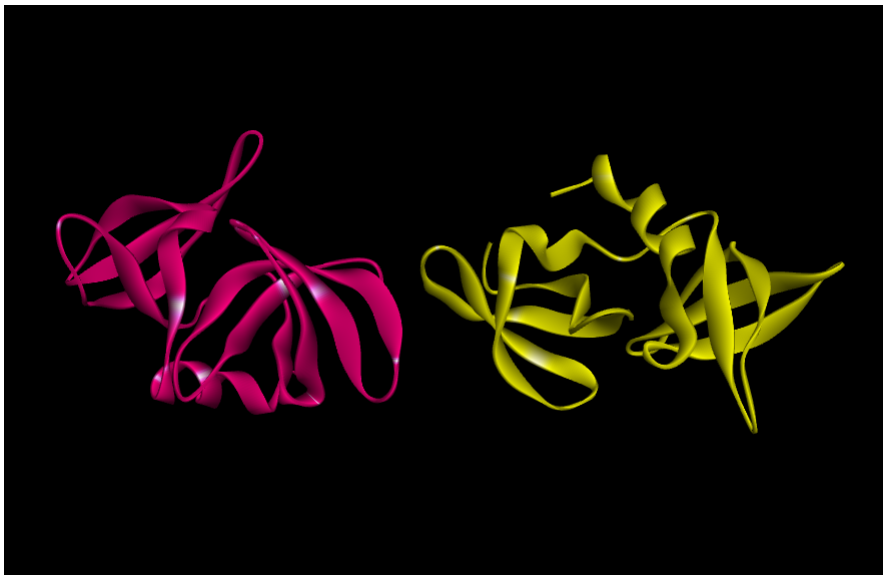

Figure 4.25: Tandem Tudor Domain of 53BP1 (PDB ID: 1XNI). According to prediction results 1XNI_A (yellow) and 1XNI_E (magenta) are both disordered.

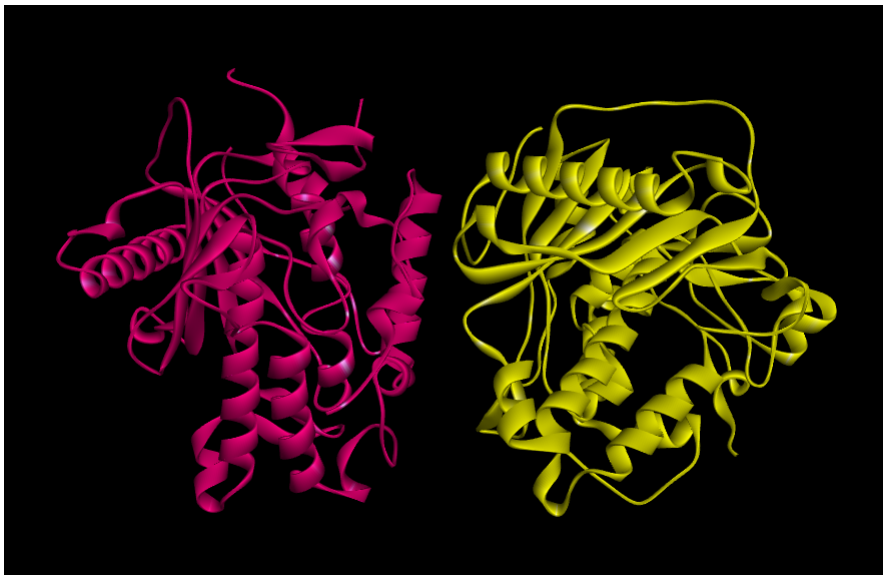

Figure 4.26: Crystal structure of the catalytic domain of an adenosine deaminase (PDB ID: 1ZY7). According to prediction results 1ZY7_A (yellow) is ordered and 1ZY7_B (magenta) is disordered.

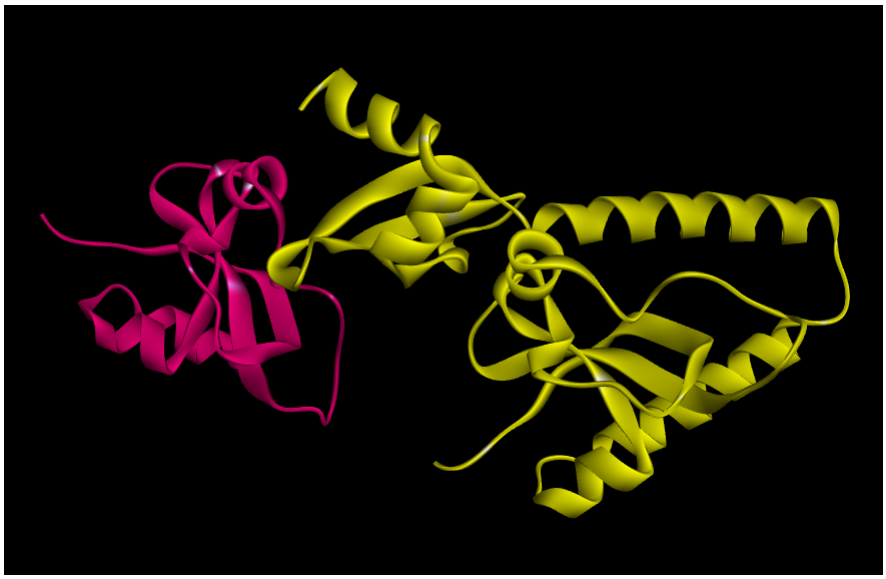

Figure 4.27: Tandem chromodomains of human CHD1 complexed with Histone H3 Tail containing trimethyllysine 4 (PDB ID: 2B2W). According to prediction results 2B2W_A (yellow) is disordered and 2B2W_C (magenta) is ordered.

Chapter 5

Conclusion

Understanding mechanism of unstructured proteins' interactions and differences between structured proteins' interactions is essential for finding how a disordered protein can specifically bind multiple partners. With identification of lot more disordered proteins it will be easier to see differences between them. In this work, we obtain that interfaces of IDPs and ordered proteins have different characteristics structurally and chemically.

In our dataset, number of IDPs that have more than 80% disordered residue fraction is 31. Most of the proteins in dataset have not got any disordered residue. Because of this situation it is hard to get good results. However, we chose randomly 150 of the ordered proteins that have 0% disordered residue for getting more reliable results. First of all, per residue areas of IDPs are generally larger than ordered proteins. If an ordered protein has same chain length with an IDP, surface and interface areas of IDP are presumably will be larger than ordered protein in a complex. Although IDPs undergo disorder to order conformational transition upon binding to target protein and become more stable, ordered proteins still more compacter from them.

In our results, we also claim that hydrophobic properties of IDPs are also differs significantly. It is known that ordered proteins have more hydrophobic residues than IDPs and they are using these hydrophobic residues to form hydrophobic core. However it is surprising that hydrophobicity of IDPs' interfaces more than ordered proteins. Instead of forming hydrophobic core and folding in 3D structures, IDPs are using their low number of hydrophobic residues to interact other proteins.

When we analyzed complexes of IDPs, surprisingly we observed that IDPs tend to bind ordered proteins. In our dataset partners of the proteins that have more than

50% disordered residues are generally have lower disordered residue fractions around 0-10%.

In conclusion, after analyzing IDPs and ordered proteins we claim that these two groups of complexes have characteristic differs. Not only their structural properties but also their chemical natures are specific to them.

As a future work, with larger dataset that have high number of unstructured proteins, these analyses can be replied. With larger number of disordered proteins that have more than 90-95% disordered residue fraction, it is possible to get better results. Also interaction energies and flexibility may be included in further works.

Bibliography

[1] P. J. Mitchell and R. Tjian. Transcriptional regulation in mammalian cells by sequence-specific dna binding proteins. *Science*, 245(4916):371–8, 1989.

[2] A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradovic. Intrinsic disorder and protein function. *Biochemistry*, 41(21):6573–82, 2002.

[3] A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic. Intrinsically disordered protein. *J Mol Graph Model*, 19(1):26–59, 2001.

[4] V. N. Uversky. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*, 11(4):739–56, 2002.

[5] O. Keskin, A. Gursoy, B. Ma, and R. Nussinov. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem Rev*, 108(4):1225–44, 2008.

[6] P. E. Wright and H. J. Dyson. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*, 293(2):321–31, 1999.

[7] H. J. Dyson and P. E. Wright. Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol*, 339:258–70, 2001.

[8] S. Vucetic, Z. Obradovic, V. Vacic, P. Radivojac, K. Peng, L. M. Iakoucheva, M. S. Cortese, J. D. Lawson, C. J. Brown, J. G. Sikes, C. D. Newton, and A. K. Dunker. Disprot: a database of protein disorder. *Bioinformatics*, 21(1):137–40, 2005.

[9] A. K. Dunker, Z. Obradovic, P. Romero, E. C. Garner, and C. J. Brown. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform*, 11:161–71, 2000.

[10] L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradovic, and A. K. Dunker. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol*, 323(3):573–84, 2002.

[11] P. Tompa. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays*, 25(9):847–55, 2003.

[12] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*, 337(3):635–45, 2004.

[13] A. P. Demchenko. Recognition between flexible protein molecules: induced and assisted folding. *J Mol Recognit*, 14(1):42–61, 2001.

[14] H. J. Dyson and P. E. Wright. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol*, 12(1):54–60, 2002.

[15] R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed, and P. E. Wright. Structural studies of p21waf1/cip1/sdi1 in the free and cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A*, 93(21):11504–9, 1996.

[16] H. J. Dyson and P. E. Wright. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*, 6(3):197–208, 2005.

[17] V. N. Uversky, J. R. Gillespie, and A. L. Fink. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, 41(3):415–27, 2000.

[18] J. F. Chang, K. Phillips, T. Lundback, M. Gstaiger, J. E. Ladbury, and B. Luisi. Oct-1 pou and octamer dna co-operate to recognise the bob-1 transcription co-activator via induced folding. *J Mol Biol*, 288(5):941–52, 1999.

[19] O. Wolstein, A. Silkov, M. Revach, and R. Dikstein. Specific interaction of tafii105 with oca-b is involved in activation of octamer-dependent transcription. *J Biol Chem*, 275(22):16459–65, 2000.

[20] A. A. Russo, P. D. Jeffrey, A. K. Patten, J. Massague, and N. P. Pavletich. Crystal structure of the p27kip1 cyclin-dependent-kinase inhibitor bound to the cyclin a-cdk2 complex. *Nature*, 382(6589):325–31, 1996.

[21] K. Gunasekaran, C. J. Tsai, S. Kumar, D. Zanuy, and R. Nussinov. Extended disordered proteins: targeting function with less scaffold. *Trends Biochem Sci*, 28(2):81–5, 2003.

[22] M. Gerstein. How representative are the known structures of the proteins in a complete genome? a comprehensive structural census. *Fold Des*, 3(6):497–512, 1998.

[23] L. W. Fisher, D. A. Torchia, B. Fohr, M. F. Young, and N. S. Fedarko. Flexible structures of sibling proteins, bone sialoprotein, and osteopontin. *Biochem Biophys Res Commun*, 280(2):460–5, 2001.

[24] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker. Sequence complexity of disordered protein. *Proteins*, 42(1):38–48, 2001.

[25] S. Vucetic, C. J. Brown, A. K. Dunker, and Z. Obradovic. Flavors of protein disorder. *Proteins*, 52(4):573–84, 2003.

[26] R. M. Williams, Z. Obradovi, V. Mathura, W. Braun, E. C. Garner, J. Young, S. Takayama, C. J. Brown, and A. K. Dunker. The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac Symp Biocomput*, pages 89–100, 2001.

[27] J. Liu and M. Lu. An alanine-zipper structure determined by long range intermolecular interactions. *J Biol Chem*, 277(50):48708–13, 2002.

[28] L. M. Iakoucheva, P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic, and A. K. Dunker. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res*, 32(3):1037–49, 2004.

[29] S. Longhi, V. Receveur-Brechot, D. Karlin, K. Johansson, H. Darbon, D. Bhella, R. Yeo, S. Finet, and B. Canard. The c-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the c-terminal moiety of the phosphoprotein. *J Biol Chem*, 278(20):18638–48, 2003.

[30] P. Tompa. Intrinsically unstructured proteins. *Trends Biochem Sci*, 27(10):527–33, 2002.

[31] R. S. Spolar and Jr. Record, M. T. Coupling of local folding to site-specific binding of proteins to dna. *Science*, 263(5148):777–84, 1994.

[32] K. I. Nakayama, S. Hatakeyama, and K. Nakayama. Regulation of the cell cycle at the g1-s transition by proteolysis of cyclin e and p27kip1. *Biochem Biophys Res Commun*, 282(4):853–60, 2001.

[33] P. Tompa and P. Csermely. The role of structural disorder in the function of rna and protein chaperones. *FASEB J*, 18(11):1169–75, 2004.

[34] J. J. Ward, L. J. McGuffin, K. Bryson, B. F. Buxton, and D. T. Jones. The disopred server for the prediction of protein disorder. *Bioinformatics*, 20(13):2138–9, 2004.

[35] C. Bracken, L. M. Iakoucheva, P. R. Romero, and A. K. Dunker. Combining prediction, computation and experiment for the characterization of protein disorder. *Curr Opin Struct Biol*, 14(5):570–6, 2004.

[36] J. N. Bright, T. B. Woolf, and J. H. Hoh. Predicting properties of intrinsically unstructured proteins. *Prog Biophys Mol Biol*, 76(3):131–73, 2001.

[37] L. M. Iakoucheva and A. K. Dunker. Order, disorder, and flexibility: prediction from protein sequence. *Structure*, 11(11):1316–7, 2003.

[38] D. T. Jones and J. J. Ward. Prediction of disordered regions in proteins from position specific score matrices. *Proteins*, 53 Suppl 6:573–8, 2003.

[39] R. Linding, L. J. Jensen, F. Diella, P. Bork, T. J. Gibson, and R. B. Russell. Protein disorder prediction: implications for structural proteomics. *Structure*, 11(11):1453–9, 2003.

[40] E. Melamud and J. Moult. Evaluation of disorder predictions in casp5. *Proteins*, 53 Suppl 6:561–5, 2003.

[41] P. Radivojac, Z. Obradovic, C. J. Brown, and A. K. Dunker. Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac Symp Biocomput*, pages 216–27, 2003.

[42] P. Radivojac, Z. Obradovic, D. K. Smith, G. Zhu, S. Vucetic, C. J. Brown, J. D. Lawson, and A. K. Dunker. Protein flexibility and intrinsic disorder. *Protein Sci*, 13(1):71–80, 2004.

[43] P. Romero, Z. Obradovic, and A. K. Dunker. Folding minimal sequences: the lower bound for sequence complexity of globular proteins. *FEBS Lett*, 462(3):363–7, 1999.

[44] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, and A. K. Dunker. Predicting intrinsic disorder from amino acid sequence. *Proteins*, 53 Suppl 6:566–72, 2003.

[45] J. Prilusky, C. E. Felder, T. Zeev-Ben-Mordehai, E. H. Rydberg, O. Man, J. S. Beckmann, I. Silman, and J. L. Sussman. Foldindex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, 21(16):3435–8, 2005.

[46] P. Romero, Z. Obradovic, C. Kissinger, J. E. Villafranca, and A. K. Dunker. Identifying disordered regions in proteins from amino acid sequences. pages 90–95, 1997.

[47] K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, A. K. Dunker, and Z. Obradovic. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J Bioinform Comput Biol*, 3(1):35–60, 2005.

[48] Z. R. Yang, R. Thomson, P. McNeil, and R. M. Esnouf. Ronn: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, 21(16):3369–76, 2005.

[49] J. Liu and B. Rost. Norsp: Predictions of long regions without regular secondary structure. *Nucleic Acids Res*, 31(13):3833–5, 2003.

[50] Z. Dosztanyi, V. Csizmok, P. Tompa, and I. Simon. Iupred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, 21(16):3433–4, 2005.

[51] C. J. Oldfield, Y. Cheng, M. S. Cortese, C. J. Brown, V. N. Uversky, and A. K. Dunker. Comparing and combining predictors of mostly disordered proteins. *Biochemistry*, 44(6):1989–2000, 2005.

[52] B. He, K. Wang, Y. Liu, B. Xue, V. N. Uversky, and A. K. Dunker. Predicting intrinsic disorder in proteins: an overview. *Cell Res*, 19(8):929–49, 2009.

[53] F. Ferron, S. Longhi, B. Canard, and D. Karlin. A practical overview of protein disorder prediction methods. *Proteins*, 65(1):1–14, 2006.

[54] R. Linding, R. B. Russell, V. Neduva, and T. J. Gibson. Globplot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res*, 31(13):3701–8, 2003.

[55] E. A. Weathers, M. E. Paulaitis, T. B. Woolf, and J. H. Hoh. Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein. *FEBS Lett*, 576(3):348–52, 2004.

[56] J. A. Wells. Systematic mutational analyses of protein-protein interfaces. *Methods Enzymol*, 202:390–411, 1991.

[57] T. Clackson and J. A. Wells. A hot spot of binding energy in a hormone-receptor interface. *Science*, 267(5196):383–6, 1995.

[58] A. A. Bogan and K. S. Thorn. Anatomy of hot spots in protein interfaces. *J Mol Biol*, 280(1):1–9, 1998.

[59] R. Beveridge, Q. Chappuis, C. Macphee, and P. Barran. Mass spectrometry methods for intrinsically disordered proteins. *Analyst*, 2012.

[60] N. Tuncbag, A. Gursoy, E. Guney, R. Nussinov, and O. Keskin. Architectures and functional coverage of protein-protein interfaces. *J Mol Biol*, 381(3):785–802, 2008.

[61] E. Cukuroglu, A. Gursoy, and O. Keskin. Hotregion: a database of predicted hot spot clusters. *Nucleic Acids Res*, 40(Database issue):D829–33, 2012.