אוניברסיטת תל אביב

Tel Aviv University

Conformational changes and ligand recognition in proteins: an evolutionary angle

Thesis submitted for the degree “Doctor of Philosophy” by

Aya Narunsky

Submitted to the senate of Tel Aviv University

July 2019

This work was carried out under the supervision of

Prof. Nir Ben-Tal

Acknowledgments

I would like to express my gratitude to all who helped on the research for my thesis.

My deepest gratitude goes to my advisor, Prof. Nir Ben-Tal, who helped shape my scientific thinking. Thank you for the empathy, trust and patience, and for the exact balance between enthusiasm and cautious.

Special thanks for Dr. Rachel Kolodny, for her pleasant guidance, the intriguing discussions and the ever-positive way of thinking.

Thanks to Dr. Sharon Ruthstein, for her kindness and support.

Thanks to all the current and the former members of the Ben-Tal lab, for the great time we had and for the coffee.

Thanks to the members of my doctoral accompanying committee: Dr. Chen Kesar, Prof. Gali Prag and Prof. Haim Wolfson. Your insightful remarks were well appreciated.

Last but not least I would like to thank my family, especially my parents and my husband Boris. I wouldn't have made it without your support.

Table of Contents

Abstract

A protein's ability to identify and bind its ligand(s) is essential to most biological functions, and often binding involves conformational changes. In the first part of my PhD I developed ConTemplate (http://bental.tau.ac.il/contemplate), a method and web-server that suggests known and alternative conformations for a query protein with at least one known structure, based on structurally similar proteins and their additional conformations. I show that most PDB proteins have multiple structures, and proteins sharing one similar conformation often undergo similar conformational changes. ConTemplate exploits these observations, and uses simplified representations for each of the PDB proteins to enable an efficient search for structurally similar proteins. In addition, it provides a network visualization of the models, to enable the user to identify pathways between conformations. A paper describing this work was published in 2015. In my second PhD project, in collaboration with Sharon Ruthstein (Bar-Ilan University), I used ConTemplate to study conformational changes in CueR (Cu export Regulator), a member of the MeR metalloregulator family. Upon Cu(I) binding, the protein initiates the transcription process of two other metalloregulators, which in turn have role in the export of copper out of the cell. The full mechanism of activation and inactivation of CueR was unknown, but previous studies suggested that it may involve ligand-induced conformational changes. Sharon carried out EPR measurements, which could not be fully explained by the known structures of CueR. I used ConTemplate to model alternative conformations of CueR, and indeed some of these correlated with Sharon's measurements. We used this set of alternative conformations to suggest a mechanism for the activation and function of CueR. A paper describing this work was published in 2017.

The first two parts of my thesis focused on conformational changes in proteins, which are often the response to ligand binding. In the third part, I studied how ligand binding and recognition has emerged, i.e., how proteins evolve to recognize their ligands? For convenience, I focused on adenine binding patterns, mostly within the context of larger molecules (such as ATP). I developed a computational pipeline which identifies the PDB ligands containing adenine, extracts complexes with these ligands, structurally superimposes them according to the alignments of their adenine fragments,

and detects the hydrogen bonds that mediate the interaction. I composed large datasets of protein-ligand complexes, and compared their adenine binding sites. I found that binding is often mediated by specific amino acid segments, referred to as 'themes', which occur repeatedly. Certain themes and combinations thereof are reused when binding the same adenine-containing cofactor. I identify proteins containing these themes as potential adenine binders. My analysis suggests that adenine binding has emerged multiple times in evolution. A paper describing this work was recently submitted for publication.

List of publications related to this thesis

1. Narunsky, A., Kessel, A., Solan, R., Alva, V., Kolodny, R., Ben-Tal, N. On the evolution of protein-adenine binding. *Submitted*.
2. Narunsky, A., Ben-Tal, N., & Kolodny, R. Navigating Among Known Structures in Protein Space. In Computational Methods in Protein Evolution (pp. 233-249). Humana Press, New York, NY. 2019.
3. Sameach, H.[#], Narunsky, A.[#], Azoulay-Ginsburg, S., Gevorkyan-Aiapetov, L., Zehavi, L., Moskovitz, Y., Juven-Gershon, T., Ben-Tal, N., Ruthstein, S. Structural and Dynamics Characterization of the MerR Family Metalloregulator CueR in its Repression and Activation States. Structure 25(7), 988-996, 2017.
4. Narunsky A, Nepomnyachiy, S., Ashkenazi H, Kolodny R, Ben-Tal N. ConTemplate suggests possible alternative conformations for a query protein of known structure. Structure 23(11), 2162–2170, 2015.

[#] - equal contributors

1 Introduction

Proteins are macromolecules involved in various functions in living organisms. These complex molecules are the result of 3.7-billions of years of evolution, constrained by their own physicochemical nature, as well as by the physicochemical characteristics of their environment. Thus, sequence and structure similarities between proteins is a testimony of this combination, and generally speaking, structural similarities are considered to be the result of the physicochemical pressure, while local sequence similarities indicate a common ancestor [1, 2].

It is useful to study the protein structure space to address questions regarding proteins' evolution and the forces which underlie it [3-6]. There are different ways and approaches to conduct such searches. Most generally, we can look for local vs. global similarities: when we examine local similarities, we enable moving in "small steps", which can indicate for example a more distant common ancestor. When we examine global similarities, we can focus more easily on the small differences between proteins that are otherwise similar, and study how they affect their function. In other words, different questions require different approaches and resolutions, and this must be taken into account when conducting the research. My research combines these approaches to tackle different questions regarding the function and evolution of proteins, focusing on protein-ligand interactions.

The Protein Databank (PDB) [7] consists of more than 100,000 entries, containing structural representations for about 20,000 different UniProt [8] proteins. This means it is highly redundant, and different entries represent different parts of proteins, different experimental conditions, etc., or, of course, different proteins altogether. In addition, it is not uniformly distributed, in that some protein families are represented more than others. This could be due to the experimental procedure, which enables researchers to derive the 3D-structures of some proteins more easily than others; or due to the interest in some families, which targets more effort in them comparing to other families. Either way, this redundancy must be taken into account and treated when looking for insights on the protein universe based on this dataset. Usually the redundancy can be limited based on the amount of

sequence identity between each two proteins in the dataset. The research methodology determines the entity and measure by which the redundancy removal process will be carried out. Different algorithms enable different entities, for example PISCES [9, 10] remove redundancy from a list of PDB entries and takes into account the quality of the structures, while CD-HIT [11] receives a list of sequences as its input, and selects representatives based on their length.

Domain classifications of the PDB aim to cluster together proteins based on similarity in their sequence or structure. These classifications include CATH [12, 13], SCOP [14] and the more recent and up-to date ECOD [15], as well as PFAM [16, 17] classification into sequence domains. The domains are sequences of ~100 amino acids, which current view holds as principle evolutionary units, whose combinations form the large diversity of proteins. However, a domain doesn't necessarily hold a specific function, and sometimes the function of the protein requires the combination of several domains acting together. This is, for example, the case of substrate binding proteins (SBP), where the binding occurs in a cleft between two domains connected by a hinge [18]. In this case, domain classifications will overlook this important functional aspect of the protein.

Protein function is the driving force of protein evolution, and of all protein functions, ligand binding, often involving conformational changes [19], is probably the most pervasive [1]. These conformational changes can vary from small changes, as, for example the well-characterized oxy and deoxy conformations of hemoglobin, which superimpose with a root-mean-square deviation (RMSD) of less than 1 Å [20], to large-scale changes or even domain-domain motions. Structural models of the different conformations of proteins can therefore help in deciphering their mechanism. For this reason, various methods have been developed to study the conformational space of proteins, and generally they can be divided into two approaches. The first approach aims to predict the missing conformations based on our understanding of the physicochemical forces which determine proteins' dynamics. Normal mode analysis and the force fields used in molecular dynamics are among the methods which apply this approach [21-27]. As it is difficult to accurately simulate the full dynamic range of proteins, these approaches may result in conformations which are biologically irrelevant. The

second, complementary, approach suggests using the wealth of data in the PDB to identify the missing conformations. This can be done by compiling and studying databases of conformational changes [19, 28-31] or using homology modelling tools with various templates [32, 33]. Domain classifications can also be used to this end, yet they may overlook domain-domain motions, which can be important in binding and catalysis, as in the example of SBPs mentioned above. To avoid the difficulties described above, I developed ConTemplate [34]: a method implemented as a web-server that suggests putative conformations to query proteins with a known structure, based on their similarity to other proteins, and alternative conformation of the latter. I showed that most of the PDB proteins are found in more than one PDB entry, and in addition, that two proteins sharing one conformation are likely to share others as well. ConTemplate suggests alternative conformations to the query protein based on the alternative structures of proteins that are structurally-similar to the query.

I used ConTemplate in two different cases where ligand binding involves large-scale conformational changes. The first case was basically a "proof of concept" for the approach, and I used it to suggest a pathway between the two well-established conformational states of the ribose-binding protein. In the second example, I used ConTemplate to suggest a structural interpretation to experimental results. This was carried out in collaboration with Sharon Ruthstein (Bar Illan University). Sharon studies the CueR (Cu export regulator) protein, a member of the MeR metalloregulator proteins. Metals are essential for many cell functions, yet at high concentration they may become toxic. Therefore, a tight regulation of their concentration is crucial for cell survival [35-38]. Metalloregulators are a family of proteins in bacteria which upon binding of their metal, initiate the transcription process of proteins that export the metal out of the cell [39, 40]. The CueR member of the family regulates Cu(I) concentration in the cell. It 'senses' it's ligand ions at an extremely high affinity of 10^{-21}M. Briefly, the protein is a homodimer, and in the absence of copper it binds to the DNA and bends it, so that RNA polymerase cannot interact with its sequence and the transcription process is repressed [35, 41]. When DNA-bound CueR also binds Cu(I), it induces the transcription of CopA [42], which moves the copper to the periplasm, and CueO [43], which oxidizes it to the less toxic

Cu(II). The mechanism of activation is well studies, yet the full dynamics of the proteins is required in order to understand its mechanism of inactivation [35, 44-48]. We aimed to get a better understanding of CueR's mechanism for transcription activation and repression, by gaining knowledge of its dynamic range in the presence and absence of Cu(I) and DNA. Sharon carried out Double electron-electron resonance (DEER) experiments, a pulsed electron paramagnetic resonance (EPR) technique, to study the dynamics of CueR [49-55]. This technique is used to measure the dipolar interactions between two (or more) electron spins, which can be interpreted to nanometer interspin distances in the range of 1.5-8.0 nm, the range relevant to CueR's conformational changes we were targeting. Sharon inspected the DEER spectroscopy of CueR in different states, and her results indicated that the available structures of CueR did not cover its full dynamic range. I used ConTemplate to model the CueR in the missing conformations. Next, Sharon added elastic-network simulations to simulate DEER measurements of these models, and compared these to the experimental results. This enabled us to suggest a mechanism for CueR's ability to regulate Cu(I)'s cellular concentration [56].

The structure and dynamics of current-day proteins are the final step in their evolutionary process. After characterizing the influence of ligand binding on these characteristics of proteins, I focused on tracing the evolutionary process itself, and on how ligand recognition and binding emerged. Nucleotide ligands are highly common in extant organisms, especially as enzyme cofactors. As such, they participate in many key metabolic reactions, including those that utilize energy and synthesize complex organic molecules. The commonness of nucleotides as enzyme cofactors may be explained by the putative dominance of catalytic RNA molecules on primordial Earth. This in turn suggests that protein-nucleotide interactions are among the oldest and most conserved interactions between proteins and small molecular ligands. Most nucleotide cofactors contain an adenine fragment, connected to other organic group(s). Although the adenine fragment is not the active component in these cofactors, it serves as a 'molecular handle' that increases the binding affinity and specificity between the nucleotide and the protein. For this reason, I decided to explore these interactions in depth.

To study the evolution of protein-adenine binding, one can relate two different types of information: (1) physicochemical binding patterns of proteins and adenine, and (2) evidence of common evolutionary origin. Once I characterize adenine-binding patterns, I can use these patterns to compare and reason about instances of adenine-binding proteins. Thus, I extend previous studies of motifs binding adenine (or adenine-cofactors) [57-59] to provide a more comprehensive catalogue of adenine-binding patterns. To deduce common evolutionary origin of protein parts, scholars study their sequence similarity [2, 60]. Here, I relate 'themes', i.e., recurring protein segments at the sub-domain level to protein-adenine binding patterns.

Previous work studied the chemical nature of protein-adenine binding. Adenine is a planar, triangle-like, molecule that features three edges; each can participate in hydrogen bonds with its environment: the Watson-Crick edge, the Hoogsteen edge and the sugar edge (Figure 1A). Studies in the 1980s and 1990s were first to suggest an adenine-binding motif [61, 62]; the motif is a carbonyl and amide groups within a protein loop that hydrogen bond the N6 and N1 nitrogen atoms of the Watson-Crick edge of adenine. In a series of papers from 2001 to 2003, Denessiouk and co-workers extended the search of this motif to other ATP-binding proteins, as well as to proteins that bind nicotinamide adenine dinucleotide (NAD), flavin adenine dinucleotide (FAD), S-adenosyl methionine (SAM), and CoA [57-59]. They demonstrated that, despite the different folds of the proteins, many of them bound adenine via the Watson-Crick edge using the aforementioned motif. The authors suggested a general scheme for the binding motif of adenine, which involved three amino acid positions, termed sites I, II, and III, and which are located on a loop. The sites hydrogen-bonded with adenine's N1 and N6 groups (Figure 1, B-D). The authors discovered three variations of the motif that differed in the sequence of the interaction sites along the loop. In the first variation, named the 'direct motif' (Figure 1B), sites I-III were ordered from the N' of the loop to the C'. In this motif, N6 hydrogen-bonded with the carbonyl group of site I, and N1 hydrogen-bonded with the backbone amide group of site III. In the second variation of the motif, named the 'reverse motif', the loop had a reverse orientation of sites I-III, and the interactions with adenine are a little different; N1 and N6 both

hydrogen-bonded with the backbone amide and carbonyl groups of site III, respectively (Figure 1C). The authors also described a third variation of the motif, which they found mostly in the adenine-binding sites of NAD-binding proteins, and which they named the 'Asp motif' (Figure 1D). In this variation, N1 hydrogen-bonded with position III, and N6 hydrogen-bonded with the side chain of a negatively charged amino acid (mostly aspartate) in position II. Subsequent studies validated these findings, emphasizing that adenine-binding involves only adenine's Watson-Crick edge [63-65].

To trace the evolutionary process by which protein families, superfamilies, and folds emerged and continue to evolve, scholars search for links between their sequences, structures, and functions [66-75]. The current view holds that domains are the principle evolutionary units, and that the large diversity of proteins emerged from their combinatorial shuffling [12, 14, 15, 17, 74]. In many instances (e.g., the SCOP family and superfamily level) proteins with overall similar sequences have similar function, possibly explaining their conservation in the evolutionary process. And yet, this does not explain how the domains themselves emerged. By one possible evolutionary mechanism, domains emerged from combinations of short peptides, originating in the RNA world [2, 68, 76-79]. Such a mechanism is supported by studies demonstrating various molecular functionalities of isolated peptides [80-82]. Specifically, these studies show that different peptides with up to and around 55 amino acids can bind small ligands, and even catalyze chemical reactions, by using recurring sequence- and structure-similar motifs (for example, the Walker motifs in P-loops [83]). These aforementioned sequence- and structure-motifs were identified by searching for common features of proteins sharing a similar function [79, 84-90]. In a recent study, Nepomnyachiy et al. introduced themes, sequences of 35-200 amino acids that are heavily reused by proteins, and suggested that they were used as evolutionary building blocks of protein domains [60]. While the authors were able to identify such themes, they did not link them to the function of the proteins, which underlies the evolutionary process. I related themes and protein-ligand interactions to learn how proteins evolved to recognize and bind their ligands.

I studied adenine binding from the two aforementioned perspectives: the physicochemical nature of the interactions, and their evolutionary origins. I first analyzed protein-adenine interactions in a representative set of 985 protein-adenine Protein Data Bank (PDB) complexes. I studied the binding patterns both from the perspective of adenine and of adenine-binding proteins. For the former, I superimposed the adenine fragments, taking advantage of their rigid and planar shapes and analyzed the specific patterns of the hydrogen bonding interactions between adenine and the interaction site. In conflict with previous studies, I found that all, rather than only some, hydrogen bond donors and acceptors in adenine can participate in the binding. I also found that water molecules often mediate protein-adenine binding. These findings are consistent in all the cofactors-interaction sites in the dataset. On the proteins side, the proteins have evolved in intricate ways to take advantage of these binding opportunities: For example, in ATP-binding proteins, which constitute the largest group of adenine-binding proteins, I found that different sequence families manifest different adenine-binding patterns. At the same time, there are different protein families that share extremely similar adenine-binding patterns. The complexity of adenine-binding evolution led me to further examine evolutionary patterns of adenine binding. I explored the possibility that protein-adenine interactions are mediated by certain themes, and attempted to delineate the structural and evolutionary relationship between these themes in different adenine-binding sites. I found that some themes are used for the binding of certain cofactors, whereas others are shared by different cofactors. Moreover, I found that proteins sharing the same themes in their adenine-interaction site tend to bind adenine via the same interaction patterns. This provides a direct link between themes and biological function at the atomic level, which further supports the putative role of themes as evolutionary building blocks of modern proteins.

2 Research aims

My research goal is to suggest insights into proteins' sequence-structure-function relationships, and how they affected the emergence of the protein universe. To do so I focused on ligand binding, which is the most fundamental and prevalent of all protein functions. I used structural comparisons to study the dynamic properties of proteins, and applied this to give structural interpretation to the experimental results of the copper regulator (CueR) in the presence and absence of its ligand. To focus on the evolution of proteins I studies how proteins evolved to recognize and bind adenine, a rigid molecule which is a fragment of many ligands that were present on Earth since the beginning of life. I made structural comparisons of adenine binding sites based on the adenine fragment, and in addition, I searched for themes in adenine binding sites – short sequences that are heavily reused by proteins. I find that the binding sites are enriched with specific themes.

The specific goals of my research were as follows:

1. To model how proteins alter their conformations in response to ligand binding.
2. To suggest insights to how the function of ligand binding emerged in the protein universe.

3 Methods and materials

3.1 Proteins' dynamics and conformational changes upon ligand binding

3.1.1 Studying the abundance of Conformations in the PDB

I counted the number of PDB chains containing the sequence of each chain in the PDB. To do this, for each PDB chain I ran a standard BLAST [91] search against the PDB, and collected all the hits with 100%, 99%, 95%, and 90% sequence identity, and full coverage.

To obtain a dataset of PDB structures of good quality, I used ASTRAL, and selected proteins with SPACI score higher than 0.4 [92]. The dataset included 77,663 PDB chains. I searched for the alternative conformations of the proteins in this dataset: I ran BLAST for each chain in the dataset, and selected hits where the product of the sequence identity and the mutual coverage was higher than 0.9. 56,255 chains in the dataset had other conformations with these criteria; I structurally aligned all the different conformations of each of the chains using the Kabsch algorithm [93, 94] on the Cα atoms of each conformation.

I searched for the largest conformational change each of the proteins in my dataset undergoes. I used PISCES [9, 10] to remove redundancy within this dataset. PISCES compares the sequences of the chains given as input, and when the sequence identity of two proteins is higher than a certain user-defined threshold, it removes from the database the one with the lower structural quality. I used the logs to change this, so that whenever two proteins shared a sequence identity above the threshold, the one undergoing the smaller conformational change was removed. This way, my dataset remains an up to 80% non-redundant representation of the largest conformational changes in the PDB.

I compared the available structures of all the proteins in my dataset that had two structures that the RMSD between them is above a certain threshold (2, 3, 4, 5 and 6Å). For each protein with conformations that met this criterion I composed an all-against-all structure alignment matrix including all the different structures for each of the proteins, and clustered it to distinguish between

the different conformations. That means that, for example, when I considered 4Å as the threshold for conformational change between two structures of the same protein, I was left with 246 proteins and their additional conformations, resulting in overall 516 structures. I structurally aligned all-vs.-all of these structures with the GESAMT structural-alignment tool [95]. When I found two structures that aligned with coverage of more than 70%, RMSD under 2Å and Q-score above 0.4, I considered them similar.

3.1.2 The ConTemplate Methodology

Structural neighbors in the PDB

I used FragBag [96], a fast method for comparing protein structures, to build a profile for each PDB chain with more than 40 residues. The library I used for the profiles contains 400 fragments of 11 amino acids (as recommended in [96]), so that each profile is a vector of length 400. FragBag aligns the protein backbone to each of the fragments in the library, and finds the number of times each fragment appears in the protein – this is the profile vector. For NMR structures I only considered the first model. Structurally similar proteins will have similar profiles. ConTemplate compares the profile of the query to the profiles of all the PDB proteins to find the nearest structural neighbors.

Through a trial-and-error process, I found that two PDB chains can be considered "close" structural neighbors if the cosine-distances between their FragBag profiles are 0.25 or less. Thus, when more than 5000 structural-neighbors meet this criterion, ConTemplate enables the user to repeat the run using all the close neighbors. This may result, however, in a longer run-time.

Step 1: Collecting the proteins that are structurally equivalent to the query

ConTemplate collects all the PDB structural neighbors of the query, as defined by their FragBag profiles. It then performs structure-alignment between the query and each of the nearest 5,000 neighbors using the GESAMT structure-alignment tool. The website suggests default parameters to consider another protein as a structural-equivalent of the query: RMSD threshold of 2Å, Q-score above

0.4, and alignment coverage of 80% of the amino-acids in the query. I found these thresholds through a trial and error process, and the user can modify them.

Step 2: Identifying and clustering additional conformations

ConTemplate uses BLAST [91] to detect the additional conformations of all of each structurally-equivalent protein detected in step 1. The similarity threshold used to define an additional conformation is set at 95% sequence identity, although the user may alter this threshold. The sequence of the structure equivalent and the sequence of its additional conformations (to which I will refer to as "Templates") may slightly differ (depending on the similarity threshold). ConTemplate uses MUSCLE [97] to align the two sequences. ConTemplate also uses the same procedure (and the same sequence identity threshold) to search for known conformations of the query in the PDB.

Clustering the suggested templates

ConTemplate uses the k-means clustering algorithm [98, 99] to cluster the templates found in the first step. To do so, it first creates Local Features Frequency profiles (LFF) [100] for each of the templates. Like FragBag, LFF also enables fast comparisons between protein structures, however it is sensitive to local changes, and so can be used to distinguish between structures representing different conformations. In LFF the internal distance matrix of the protein is divided into overlapping sub-matrices of fixed size. Each submatrix is compared to each of the submatrices in the features library (similar to comparing each fragment of the protein to all the fragments in FragBag's library), and a profile is derived. For NMR structures only the first model is considered for profiling. The features library used by ConTemplate (and recommended by [100]) includes 100 matrices of size 10×10. Once the profiles are built, ConTemplate clusters them into the user-defined k clusters. The representative template of each of the k clusters will be used in the next step as structural templates for model-building.

Step 3: Model building

ConTemplate uses Modeller [101] for comparative modelling the query sequence based on the template-structures found in the second step. In other words, ConTemplate models the query sequence based on the structures of each of the representative templates found in the second step.

Detecting a pathway between conformations

The pathway between the query and the target structures in the usage example was identified using the Cytoscape tool for network analysis [102]. Two nodes are connected by an edge if the RMSD between the respective models is under a certain cutoff (in the example, 2.5Å). The length of the edge is proportional to the RMSD, meaning shorter edge connects more-similar models. This cutoff depends on the RMSD between the models of interest: low cutoffs may break the network into separated components, while high cutoffs will form overly-crowded connected component and eventually disable the opportunity to find the pathway. ConTemplate provides the user with this network-view, and the Cytoscape session file can also be downloaded.

3.1.3 CueR dynamics upon ligand binding

Experimental procedure on CueR (performed by the lab of Dr. Sharon Ruthstein, Bar Ilan University)

The CueR protein was expressed and purified from BL21 competent *E. coli* cells. To produce wild-type strain, cells were transformed with pET-28a(+) expression vector forming CueR gene. This background was used for all mutant strains construction. The strains were constructed using kanamycin resistance – which enables to select cells with specific properties – again using pET-28a(+) expression vector. The cell growth conditions were standard 37°C in LB liquid media. Isopropyl-β-D-thiogalactopyranoside (IPTG) reagent was used to induce CueR expression.

DNA Preparation for EPR Measurements (performed by the lab of Dr. Sharon Ruthstein, Bar Ilan University)

Spin-labeling and centrifugation were carried out as described in the methods section of [56], for the removal of free spin-labels from the solution. At the end of this process, 0.1 mM KCN was added to the protein solution and CW-EPR spectra was generated, to verify no Cu(I) ions were found in the protein solution. The spectra with and without KCN were similar, and no activity was measured for the apo-protein solution.

To maintain anerobic conditions, the Cu(I) (Tetrakis (acetonitrile) copper(I) hexafluorophosphate) was added to the solution under nitrogen gas. Its concentration was up to 4-fold more than that of the CueR monomer, to verify binding and enable the measurement of the major conformational changes in the presence of DNA. Cu(II) EPR signal was never observed.

DNA fragment isolated from copA promoter, containing 237 base pairs, was used for the EPR measurements. This DNA includes a specific region known to bind CueR: -35/TTG ACCTTCCCCTTGCTGGAAGGTTTA/-10. The PCR was done on *E. coli* genomic DNA using specific primers: primer(+) (5'-CACCCGCAACTTAACTACAG-3') and primer(-) (3'-TTTAACGCAGTGACCGCA GG-5').

3.1.3.1 EPR measurements and simulations

The complete description of the EPR measurement is given in [56]. In each of the experiments, DEER measurements were acquired for 48-72 hours. The data collected were analyzed using the DeerAnalysis 2013 program [103], with the Tikhonov regularization and L-curve critetion [104, 105]. The fit of the time domain data was used to optimize the regularization parameter of the L-curve. While the data presented in [56] are after removal of the 3D homogeneous background, this background was negligible in most cases due to the low protein concentration. Thus, in most cases the DEER data presented are similar to the raw data.

For the elastic-network model simulations the 2013 MMM program was used [106]. This is an open-source Matlab (MathWorks^{TM}) package for Multiscale Modeling of Macromolecules, with a graphical user interface, used for the modeling of structural changes of protein based on experimental restraints. The user inputs the program with a PDB template, and can select sites for rotamer analysis

based on specific type of spin label. The computed rotamer distributions are then transformed into dipolar evolution time traces of DEER experiments. These simulated DEER data can be compared to the experimental DEER data, and so the data closest to the experimental one may be the best representation to the structure of the protein under the experimental conditions. The MMM program requires a complete structure. For the 1q05, 12 residues are missing: at the end of the C-terminus, and on the loop connecting helices 4 and 5 in one monomer, and 5 and 6 in the second. To predict the missing residues, and acquire full structure for the MMM simulations, I used MODELLER loop prediction tool, accessible through UCSF chimera [107].

3.1.3.2 Models of CueR in different conformational states

I used ConTemplate to search for template of alternative conformations of CueR, as described above. ConTemplate also uses the structural alignments between the query and structurally similar proteins to produce sequence alignments. When this alignment was sufficient (and did not have, for example, large number of gaps), I used it. When it wasn't, I used MUSCLE sequence aligner to align the query and the template. I used these alignments and the structures of the templates as input to MODELLER homology modelling tool, in order to produce model of the query in various conformations. Since I aimed at modelling the dimers in a certain configuration, I modeled each chain separately, then structurally aligned it to the relevant chain in the template.

In several cases, residues from the N-terminal binding domain were missing from template structures, which made it impossible to use them as inputs for the MMM program as this region was used for spin labelling. I used PDB structure 1q06 – a CueR structure – to model the missing residues. For the rest of the model I used the selected template.

3.2 On the emergence of adenine binding

3.2.1 The ComBind methodology

2D representations of the ligands. 3D coordinates of the ligands are downloaded from the PDB, and their 2D representations are acquired using the OpenBabel [108] chemistry toolbox.

Identify the adenine fragment in a ligand. To identify the adenine fragment of a ligand, ComBind calculates the distances (in Angstroms) between all the nitrogen atoms in a 2D representation of the ligand, and compares them to the distances between adenine nitrogen atom pairs. When the algorithm identifies a pair of identical distances, it shifts the ligand to the adenine to match the two nitrogen atoms. Next, it uses the Hungarian algorithm [109] to find the ligand atoms that have the closest proximity to the adenine atoms, and calculates the RMSD between these atoms and adenine. The ligand will be considered as containing adenine if this RMSD is small enough (less than 0.1Å). ComBind then uses the Kabsch algorithm [93, 94] to calculate a rotation and translation that optimally superimposes the adenine on one another, and uses it to transform the bound proteins. ComBind can be used with any rigid fragment of ligands.

Identify hydrogen bonds between the adenine and its environment. The polar interactions between the adenine and its surrounding (e.g., water molecules and amino acids) are identified using Arpeggio [110]. The ligand and the atoms that hydrogen bond with the adenine are extracted to a Pymol [111] session, and these atoms are referred to as the interaction site of adenine.

3.2.2 Composing the datasets

I collected all the proteins in the PDB that bind adenosine triphosphate (ATP) and its analogs (PDB ligands: ADP, AMP, ANP, ACP, DTP, AGS, DAT, APC, A12, AN2, ADX, M33), nicotinamide adenine dinucleotide (NAD) and its analogs (PDB ligands: 8NA, A3D, CNA, DND, NXX, NAP, NA0, NJP), flavin adenine dinucleotide (FAD) and its analogs (PDB ligands: 6FA, FAS, 5X8), S-adenosyl methionine (SAM) and its analogs (PDB ligands: SAH, SMM), and coenzyme A (CoA) and its analogs (PDB ligands: CAO,

COS, COZ, 1VU, ACO, BCO, IVC, ACO). I selected analogs that did not change the functional part of the ligands, and where the adenine moiety fragment remains unchanged. To remove redundancy from the dataset, I selected only the proteins found in a dataset containing all the PDB proteins with up to 30% sequence identity. Clustering was performed using the sensitive cluster mode of MMseq2 [112] at a length coverage of 70%. From the resulting clusters, one representative per cluster was chosen based on resolution, R-free factor, and completeness; when possible, crystal structures were preferred over NMR structures. The dataset includes 985 entries (see Table 3), 751 of them (76%) are structures of a very good quality (resolution under 2.5Å, free-R value under 0.25), in 113 (11%) entries the resolution is between 2.5-6.93Å and in the rest (121, 12%) of the structures the resolution is good (2.5Å or better) but the free-R value is between 0.25 and 0.3.

3.2.3 Network of binding patterns

To compare the adenine interaction sites in two different complexes, the closest binding site atoms are detected using the Hungarian algorithm. After the correlating atoms are detected, the RMSD between them is calculated using the Kabsch algorithm [93, 94]. I consider two interaction sites as 'similar' if this RMSD is under 0.3Å, and the corresponding atoms include at least 60% of the atoms in each of the interaction sites. I performed this calculation for all vs. all adenine interaction sites in the ATP datasets, and used Cytoscape [102] to visualize the network (Figure 18A). Each node in the network represents an ATP (or analog)-interaction site, where the adenine fragment has at least 3 interactions with its environment. Two nodes are connected by an edge if the respective interaction sites have a similar geometry. The length of the edge corresponds to the overall similarity between the two interaction sites, the shorter the edge is the more similar the two binding sites. Nodes not found in the main connected component, forming small clusters, were removed for a clearer view of the network.

Composing the themes dataset for adenine binding proteins (performed by Dr. Rachel Kolodny, Haifa University)

The procedure we used to generate the themes dataset for adenine binding proteins follows several steps:

Alignments: we started with HMM alignments for all the chains in our adenine-binding dataset. We filtered the alignments, keeping only those with E-value under 10^{-2}. For each chain we collected all the alignments to other proteins in the set.

Generate candidate segments for the themes: for each chain, we calculated variations of different minimal lengths: 30, 40, 50, 60, 70, 80 amino acids, and used a unified naming scheme.

Identify the connected components in the chain network: from the alignments generated in the first step we composed a network where each node is a protein chain, and two nodes are connected by an edge if the respective protein chains are aligned to each other. Next, we separate the chains into connected components. We perform the next step on each connected component separately.

Search and join the variations: for each chain in the adenine-binding dataset, and for each set of variations in the chain, we search for the connected component of that chain. Each chain will be a node in a graph. For each node, we listed all the variations found in it. Starting from a specific node C with a specific variation M_n1, residues number (s, e), we consider only edges that connect C via alignments that match the residues between (s, e). We enforce the edge to connect two nodes with alignment of approximately the same length as (s, e), and that 80% of the residues of the variation M_n1 were matched to some residues in the alignment.

Theme generation: Once we have the list of pairs of variations that are similar we group them into themes. The themes are the connected components in another graph, where the nodes are the variations we described earlier, and the edges are the similarity relationships between them. We assign each theme (or a connected component) a number. The theme is a set of protein fragments, and evidence of similarity between them.

Detecting themes in adenine interaction sites (performed Ron Solan, from the lab of Prof. Nir Ben-Tal, Tel-Aviv University)

To detect all the themes in our dataset we expanded the dataset of themes. First, we search for the themes in the UniProt database [8]: we used each theme as HMM and used HMMER [113], with a threshold of E-value smaller than 10^{-5}. In order not to lose the proteins that were initially included in the HMM of the theme, we made sure to add them to the resulting HMM. We applied this expansion process twice for each theme. Next, we used HMMER again on our adenine-binding dataset with each theme, and searched for the proteins in this database containing the theme.

3.2.4 Themes network

I listed all the themes found in the interaction site of each of the proteins in my dataset, according to the unified naming scheme described above. I compared all-vs.-all of the proteins in the dataset to search for all the themes shared by pairs of proteins: I created a network where each node represents a protein interaction site, where two nodes are connected by an edge if the corresponding binding sites have a shared theme, hence: amino acids which hydrogen-bonds to adenine in both proteins are part of the same theme (Figure 19A, only clusters with 10 or more nodes are shown). I used Cytoscape [45] to view the network.

3.2.5 Discover adenine-binding proteins in protein datasets

I used HMMER [113] to search for each of the themes shared by proteins in Figure 19 in the PDB/UniProt. When a theme was found in a dataset entry with E-value smaller than 10-5, I listed this entry as “suspected adenine binding protein”. When searching against the PDB, I used ComBind to check for adenine-containing ligands in this entry and created two lists: one of proteins which have a theme related to adenine binding, and also have adenine as part of their structure (possibly in the context of a larger ligand); the second list was formed of proteins which contain a theme related to the binding, but with no adenine in their structure. I used BLAST to search for all the proteins in the

second list against proteins from the first list, hence: to check whether PDB entries with no bound adenine may share similarity to the sequence of proteins which bind adenine. A protein will be considered probable to bind adenine if it shares at least 80% sequence identity, with 80% coverage, to a protein which is known to bind adenine.

4 Results

4.1 ConTemplate

4.1.1 PDB statistics: assessing the redundancy and enumerating the scale of conformational changes

I counted the number of occurrences for each chain in the PDB using BLAST [91] search with several sequence-identity thresholds (100%, 99%, 95% and 90%). This analysis showed that most PDB chains are found more than once, and often in multiple entries (Figure 2).

Next, I evaluated the scale of these conformational changes. When looking at two different structures representing the same protein, it is not trivial to determine whether the differences between them are the result of meaningful conformational changes, or reflect the differences in the experimental process by which each structure was obtained. Small changes could still be important; for example, the oxy and deoxy conformations of hemoglobin superimpose with a root-mean-square deviation (RMSD) of less than 1Å [20]. Thus, I started by evaluating the overall differences found between structures of the same proteins in the PDB. I collected 77,663 PDB chains with SPACI score [92] higher than 0.4 (indicating structures of high quality) with up to 80% sequence identity. I counted the number of appearances in the PDB for each chain, and evaluated the largest difference between two of its structures (Figure 3). Most of these differences were very small, 69% under 1Å. However, I also observed large-scale changes, as with the c-Src, where the difference between the active and inactive conformations is over 20Å [114, 115].

I composed datasets of proteins with conformations that differ from one another with RMSD of 2, 3, 4, 5 and 6Å (Table 1). The result for each of the datasets showed that in general, two proteins sharing one conformation are likely to have more conformations in common. I focused on the 4Å dataset, as it is a large enough difference to assume that two structures indeed represent conformational changes, and the dataset includes enough proteins. This dataset included 246 proteins in 526 different conformations. It is quite uncommon for protein pairs in this dataset to be structurally

similar (less than 1%), but when two proteins share one conformation with each other, they have additional conformation in common in 57% of cases (Figure 4).

4.1.2 The ConTemplate method

ConTemplate suggests putative conformations for a query protein structure using a three-step process (Figure 5). First step: ConTemplate searches for structural equivalents of the query, using the GESAMT structure-aligner [95]. To reduce the number of structural comparisons, ConTemplate only searches amongst the structural neighbors pf the query, as defined by their FragBag profiles [96]. At the end of the first step, ConTemplate gathers all the known conformations of the query, together with structure-based sequence alignments of the query and its structural equivalents. Second step: for each structural equivalent, ConTemplate runs a BLAST search [91] against the PDB to identify all the known conformations of this equivalent. All the collected conformations are then clustered, so that each cluster represents a different conformation. Third step: ConTemplate models the query according to the center of each of the clusters identified in the second step. The center of the cluster is the structure closest to the center, and it is used as template for the modelling. The sequence alignment is the structure-based sequence alignment obtained in the first step.

4.1.3 The ConTemplate web-server

Input

ConTemplate is a fully automated web-server. The user needs to upload the structure of the query (either as PDB ID or as a coordinates file). The thresholds for structural similarity between the query and the structural equivalents, and the sequence identity thresholds, can also be set by the user, but there are also default parameters for less-experienced users. Several detailed examples are available, to demonstrate the usage and the parameters selection of different use-cases.

Output

The web-server's output includes a list of all the known conformations of the query, and models of suggested conformations based on structural equivalents. The models are generated automatically based on the method describe above. The number of models is determined by the number of clusters. When the number of structures collected in the second step is smaller than the number of clusters set by the user, all the identified structures will be used as templates for modelling the query in the third step. When the number of structures is higher, only the cluster centers are used as templates. For each model, ConTemplate indicates the size of the cluster, the structural equivalent that was the origin of the template and the RMSD between this structural equivalent and the query. The size of the cluster may indicate whether the conformation it represents is biologically relevant, as the more relevant conformations are more likely to be shared by several proteins. The user can view the suggested conformations, aligned to the query, using JSmol. The models can also be downloaded.

Another important visualization the server offers is the similarity network view of the query and its models, using Cytoscape [102] and CyToStruct [116]. The conformations are represented as nodes connected by edges. The lengths of the edges indicate the RMSD between the models (meaning, short edges connect nodes representing similar models).

4.1.3.1 Case study

To demonstrate the abilities and limitation of ConTemplate, I use it to study conformational changes in the D-ribose-binding protein, a member of the periplasmic binding protein superfamily [117, 118]. Like other members of this family, this protein participates in the membrane transport protein, and It is located in the periplasmic space of the bacteria cell. While the sequences of members of this family may be very different, their structures are rather similar, composed of two domains connected by a hinge. The ligand binds to a cleft between the two domains, and induces a rotation around the hinge bringing the two domains closer together to what is called "the closed conformation". The RMSD between the open [119] and closed [120] conformations is 4.1Å. In the closed conformation, the protein interacts with the transport complex in the inner cell membrane. The high diversity in the

sequences of the periplasmic binding protein superfamily makes them a good choice to study ConTemplate's methodology. It enabled me to apply an artificial cutoff for a maximal amount of sequence identity between the query and templates collected in the second ConTemplate step, and by that to avoid the trivial result: selecting the known conformations of the query protein itself. In addition, the difference in the popularity of the two conformations amongst PDB proteins makes it an interesting example.

I started by querying ConTemplate with the open, ligand-free conformation (Figure 6, A-C). I set the number of clusters to two, and used the web-server's default parameters. That was enough to reproduce the closed, ligand-bound conformation with RMSD of 1.7Å from the actual known closed conformation. The model is based on the structural similarity between the query and the xylose-binding protein [121], another member of the superfamily, sharing 27% sequence identity with the query. The template is the closed conformation of the same protein. When I tried to set the number of clusters to higher numbers, and by that create larger number of models, all the suggested models represented either the open or closed conformation, but never an intermediate between them.

The reverse direction is, in fact, the more challenging one, as the open conformation is by far the less abundant one amongst PDB proteins, making it harder to detect (Figure 6, D-F and Figure 7). For this reason, I decide to increase the number of clusters in order to capture the open conformation. When the number of clusters was under nine, this conformation was not represented by any of the models. Overall, as the number of clusters is increased, ConTemplate produces models that describe the open conformation better. With nine clusters, one model represents an open conformation, but the RMSD between this model and the actual known open conformation is 4.2Å. With 20 models, the server already suggests a model that the RMSD between it and the actual open conformation is 2.2Å. This model is based on the open conformation of the D-allose binding protein [122], which shares 34% of its sequence with the query. As I increased the number of clusters, I found models which represent intermediates between the open and closed conformations (Figure 7). I used that to describe the transition between them, based on a pathway I identified using the Cytoscape network analysis tool

[102]. This pathway also matches the dominant mode of motion in an anisotropic network model analysis. I compared ConTemplate to other automatic homology modelling tools, such as Swiss-Model [32] and ModBase [33]. These tools only suggested models representing the abundant closed conformation, while the open, ligand-free conformation was never suggested.

4.2 CueR dynamics upon ligand binding

Once I validated the ConTemplate approach, meaning I showed it can suggest relevant conformations for query proteins, I could use it to expand the available conformations of other proteins of interest. For The copper export protein (CueR), only some of the conformations where available. Sharon Ruthstein's lab performed a variety of wet-lab experiments to measure the dynamic range of the protein, and I used ConTemplate to model different conformations of the dimer and check whether they could describe the experimental results.

4.2.1 Generation of mutants for spin labeling (performed by the lab of Dr. Sharon Ruthstein, Bar Ilan University)

The spin labels should be attached to cysteine residues, and for that reason we needed to generate two types of mutations: one, where the cysteine residues that aren't supposed to be labeled are mutated to alanine, and the second one, where residues in the locations we wish to use for labelling are mutated to cysteine. CueR is a dimer, where each monomer contains six helices (Figure 8). The N-terminal domain (which is also the DNA-binding domain) of each monomer is composed of helices α1 to α4. The C-terminal domain of each monomer includes the dimerization helix and the metal-binding site, and is composed of helices α5 and α6. Each monomer has four cysteine residues: C112, C120, C129 and C130. C112 and C120 are in the metal-binding site, and their mutations will result in a non-responsive protein. C129 and C130 are non-essential for the transcription process or the metal binding, and in addition, are accessible to spin labels. For this reason, both were mutated to alanine

(and experiments show this does not influence the transcription process). G11C, G57C and the combination: G11C+G35C mutations, in the DNA-binding domain, were used to measure the response of the protein to DNA sequence. In addition, M101C on α5 is used to measure the changes in the distances between the dimerization arms. These mutations do not alter the protein's secondary structure, as shown by Circular Dichroism spectra analysis (Figure 9). In addition, these mutants bind the promotor in a similar way to a wild-type protein.

4.2.2 Comparisons between Apo-CueR and CueR-Cu(I)-DNA, CueR-Cu(I), CueR-DNA (performed by the lab of Dr. Sharon Ruthstein, Bar Ilan University*)*

Sharon performed several DEER experiments on spin-labeled CueR (as described above). The presence of DNA and Cu(I) (which enables the protein to assume the "activation" state) had large effect on the DEER signals, and on the corresponding distance distribution (Figure 10). The analysis of the data suggests that the large conformational changes are in the region near the spin-labeled G11C sites upon DNA and Cu(I) binding (Figure 10A), while M101C undergoes only minor changes (Figure 10C), and G57C doesn't undergo any change upon the binding (Figure 10B).

The results of adding DNA alone (and thus obtaining the "repression" state of CueR) showed distance distribution that corresponds to larger distances than the Apo-CueR, but shorter than that of CueR-Cu(I)-DNA. This suggests that CueR can assume a different conformation for each of the states, and that there is an equilibrium between the different states, which is influenced by the concentration of Cu(I). This is in line with the population shift model. Crystal structures of CueR show variations to the inter-monomer distances of G11, which also supports Sharon's experimental results (Figure 11).

4.2.3 Exploring the intra-monomer conformational changes in the DNA N-terminal domain (performed by the lab of Dr. Sharon Ruthstein, Bar Ilan University)

Spin labels G11C and G35C were used to measure intra-monomer conformational changes. This means the G11C-G11C inter-monomer will also be measured, and indeed, the results of this give a bimodal

distribution. The peak of G11C-G11C was similar to the one detected in a previous measurement, and so the second peak was referred to the G11C-G35C intra-monomer distances. In the presence of DNA and Cu(I), the DEER measurements indicated distances which corresponded to the intra-monomer activation state. Small population of the protein's molecules also showed distance distribution which correlates to the apo-CueR, again supporting the population shift model suggested above. The DEER measurements show that the intra-monomer distance between G11C and G35C decreases significantly as a result of DNA and Cu(I) binding.

4.2.4 Comparisons of CueR crystal structures with computational models

The DEER data suggests that the known conformations of CueR do not cover the protein's full conformational space. Thus, we used computational approaches in order to suggest structural models of the missing conformations. First, we used MMM2015 [106] software with the distance constraints derived from the experiments to create structural models of the apo-CueR and CueR-Cu(I)-DNA complex. As this is a fairly small protein, five constraints were sufficient for the software to build a structural model. We compared the resulting structures with known crystal structures of the protein (Figure 12). For the CueR-Cu(I)-DNA, there are some differences between the crystal structures and the DEER measurements. However, the MMM2015 models may result in structures that are not biologically relevant, and so we decided to use the wealth of data in the PDB to build structural models of CueR. I used ConTemplate, and queried it with the CueR known crystal structure (pdb 1q05). ConTemplate suggested several templates, based on various members of the MerR metalloregulator family. We than used the MMM software to simulate the MTSSL attachment to specific locations in each model, and compute the corresponding distance distribution. These distributions were compared to the experimental results; this way I could suggest a structural interpretation to Sharon's experimental results (Figure 13).

The simulated distance distribution of several models was similar to that of the distribution calculated based on the DEER measurements. ConTemplate's models suggest that the regions that

undergo the largest conformational changes are those corresponding to the DNA-binding N-terminal domain. In addition, the DEER simulations of the models suggest that these regions have dynamic ranges of about 3.5±0.5nm, the largest distance distribution when compared to spin labels located on other regions of the molecule. This is also consistent with the experimental DEER data.

4.3 On the emergence of adenine binding in evolution

4.3.1 Analyzing protein-ligand interactions using ComBind

I developed ComBind (Common Binding) to detect polar (hydrogen bonding) interactions of a given rigid ligand or a fragment of a ligand, both of which will be referred to as 'fragment'. This allowed me to identify interaction patterns that are associated with the binding of specific fragments, and are shared or differ between protein families or folds that bind that same fragment. The ComBind pipeline is implemented in Matlab and includes a few steps. Briefly, it searches the PDB for all instances of ligands that contain the query fragment using a planar representation of the ligands. Then, ComBind downloads all the PDB entries containing these ligands, and superimposes their three-dimensional structures based on the query fragment (Figure 15). Next, ComBind uses Arpeggio [110] to identify polar interactions between the fragment and its interaction site in each of the PDB entries. Finally, ComBind creates a PyMOL [111] session containing the fragment and the atoms of the interaction sites (Figure 16, Figure 18).

4.3.2 The binding of the adenine fragment of different cofactors

I examined the binding of proteins to the rigid adenine fragment within larger ligands. First, I used ComBind on the original dataset of Denessiouk et al. [58], and verified that my results are consistent with theirs (Figure 16A). Their dataset included mainly proteins that bind adenosine triphosphate (ATP), with about 200 redundant proteins (from ~500 complexes, as some of the proteins have more than one structure in the PDB). Next, I compiled a dataset of 985 PDB protein-cofactor complexes, sharing at most 30% sequence identity with each other [112], and used ComBind to analyze the

complexes. The cofactors in my dataset included ATP (41% of the dataset), flavin adenine dinucleotide (FAD, 20%), S-adenosyl methionine (SAM, 17%), nicotinamide adenine dinucleotide (NAD, 13%), coenzyme A (CoA, 8%), and their analogs (Figure 17, the analogs are listed in the Methods). I found that, in contrast to previous studies, there are instances of proteins hydrogen bonding on any of the three edges of adenine, not only the previously documented Watson-Crick edge (Figure 18A). More specifically, the analysis identifies two interaction clusters of the binding protein and adenine's N6 group, one forming the known interaction with the Watson-Crick edge and the other forming another interaction with the Hoogsteen edge.

The 'direct', 'reverse', and 'Asp' motifs defined by Denessiouk and coworkers [57-59] are detected by my analysis in about half of the complexes in the dataset (Table 2). The rest of the complexes present various interaction patterns that do not correspond to a specific predefined motif. The 'reverse' and the 'Asp' motifs are the two most common motifs. In FAD-binding complexes, the 'reverse' motif is the most prevalent (over 60%), while the 'direct' motif is completely absent. In SAM-binding complexes, the 'reverse' motif is found in 17% of the complexes, and the 'direct' motif is almost completely absent. In NAD- and CoA-binding complexes, both the 'direct' and the 'reverse' motifs are relatively rare. The 'Asp' motif is more common than originally reported; it is found in 36% of the NAD complexes, and in 60% of the SAM complexes. In addition, I found this motif to be more variable than described; position II can be populated also by Glutamate, Serine and Cysteine residues. These variations are most common in binding adenine within the context of SAM complexes. With this extension, the motif is found in complexes binding all of the adenine-containing cofactors in my dataset, except CoA. The 'direct' motif is the least common among the three, and it is found mostly in ATP-binding complexes. In these complexes, the 'direct' and 'reverse' motifs are found in about 37% of the complexes, in almost equal numbers.

4.3.3 Adenine binding within the context of ATP

ATP-binding proteins are the most prevalent in my dataset (406 complexes, see Table 3). These proteins are diverse, spanning about a dozen different ECOD [15] folds. Because this suggests that they interact with the adenine fragment of ATP differently, I used ComBind to characterize the geometry and interaction patterns of their adenine interaction sites. Similar to the full dataset of adenine-binding proteins, here too I observed a tendency of the proteins to exploit the full hydrogen-bonding potential of adenine for binding (Figure 18B).

I compared the binding sites of the different ATP-protein complexes by identifying atoms in the proteins that participate in equivalent hydrogen bonds with the adenine fragment. Then, I extracted these atoms from all the ATP-binding complexes in the dataset and compared their spatial arrangement; each combination of atoms from a complex is defined as an interaction site. For these interaction sites, I created a network. The nodes represent sites that have at least three hydrogen bonds with adenine; only 238 (59%) of the interaction sites satisfy this condition. Edges connect similar interaction sites. I considered two sites similar if after I superimpose the adenine fragments, the interacting atoms on the adenine site are generally the same ones (at least 60% are in both, i.e., at least 2 out of 3, 3 out of 4, etc.), and that the RMSD of their hydrogen-bonding partners is lower than 0.3Å. Figure 19A shows that the resulting network has a large connected component of proteins that bind adenine via similar hydrogen bonds. The color of the nodes encodes the ECOD F-group classification of the binding protein, which is equivalent to PFAM's protein family assignment. Since ECOD's families are code-names and linked to PFAM's families, I used PFAM's assignments. Figure 19A shows that the adenine-binding patterns vary and are only sometimes shared by different proteins.

Proteins of the most similar sequences (i.e., from the same family) tend to cluster together. This is expected, as the sequence determines structure, including that of the binding site, and the latter determines the interaction site. I also saw that families of similar function cluster together (the same ECOD H-group – equivalent to SCOP's superfamily) – meaning that they share similar adenine

binding modes. A good example is given by protein kinases, where the adenine-interactions clusters of several kinase groups are all connected (Figure 19A, rectangle 1).

Interestingly, I also found that proteins with different sequences, folds, and even binding site topology (i.e., arrangement of secondary structures) sometimes still share the same interactions with adenine, and even use the same binding motif for that. For example, the two PFAM families, the "universal stress protein" family and the "IMPDH/GMPR" family, share only about 15% sequence identity with each other, adopt different folds with different secondary structure topology, and different overall structures. Nevertheless, both use the reverse motif to bind adenine, and both have water molecules forming the same hydrogen bonds with adenine (Figure 19A, rectangle 2, and Figure 19B), suggesting that they may have independently converged to adenine binding. The seemingly opposite trends demonstrated above indicate complex relationships between sequence, structure, and function within adenine-binding sites. Thus, to further explore these relationships, I studied the evolutionary building blocks that construct the binding sites.

4.3.4 Adenine binding and themes

To study the evolution of adenine-binding I looked for themes that participate in the binding and construction of the binding site and region; a theme participates in the binding if at least one of its residues hydrogen-bonds to the adenine fragment of the ligand. Figure 20 shows the similarities among the identified themes as a network. In this network, each node represents a protein-adenine complex, and edges connect nodes representing complexes whose protein part share a theme (Figure 20A). Specifically, I added an edge when the same theme appeared in both complexes, and in each theme, at least one amino acid hydrogen-bonds the adenine. The nodes are colored by the bound ligand. The Figure shows that proteins that share the same theme tend to bind the same cofactor. Moreover, when I examined all the large network clusters (of 10 nodes or more), I saw that each corresponds to a distinctive binding pattern of adenine (see Figure 20B). The participation of themes

in adenine's binding mode is briefly summarized below and demonstrated in Figure 21 and in Figure 22.

4.3.4.1 Clusters in the network of adenine-binding regions

There are 37 clusters in the adenine binding network, containing overall 528 nodes. I focused on clusters with 10 or more nodes, resulting in 10 clusters with 388 nodes. In four of these clusters water molecules are used for coordination with the adenine fragment.

4.3.4.2 Themes found in adenine-binding regions in proteins

There are 756 themes found in adenine-binding regions. Of these, 537 are found in more than one binding region, and so connect two protein-adenine complexes in the network. Figure 20 only shows the clusters with 10 or more adenine-protein complexes. 151 themes are shared by the complexes forming the network in the figure.

4.3.4.3 Clusters of ATP-binding proteins

There are five large clusters of ATP-binding complexes (Figure 20, clusters 1-5). The clusters differ in the themes (i.e., the reused protein sequence segments) they use for adenine binding. This translates to binding patterns that are cluster-specific, as shown in Figure 20B. For example: The binding pattern of cluster 1 is the direct motif (see Figure 21A). The binding pattern in all the complexes of cluster 4 is the reverse motif, with an additional interaction between the protein and adenine's N6 in the Hoogsteen edge (see Figure 21B). I found another cluster, formed entirely by proteins from PFAM's IMPDH/GMPR family, which shows an interesting variation to previously known motifs. All the proteins in the cluster bind adenine in the Watson-Crick edge using the reverse motif, but they also have an additional interaction between a backbone carboxyl group at 'position XV/XVI' (22/23 residues downstream to the residues forming the reverse motif) and adenine's N6 in the Hoogsteen edge (Figure 22A).

4.3.4.4 Clusters of other nucleotide-binding proteins

Clusters 6-10 of protein complexes show additional interesting patterns. First, I saw how a combination of themes can be used to form a complete interaction site. The binding sites in the FAD-binding sub-cluster use a combination of two themes to bind adenine (Figure 21C); both themes are always used together. In the NAD-binding sub-cluster, I saw proteins which use a combination of three themes to bind adenine (Figure 21D), but not all three themes are required to form the binding (meaning, proteins "mix and match" these three themes). Some of the clusters exhibit new adenine binding motifs. For example, the proteins in a cluster formed completely by SAM-binding proteins in PFAM's SET domain family bind adenine in a pattern similar to the 'reverse' motif, except that the interactions are between the backbone amide and carboxyl group of the amino acid in "position III" and adenine's N6 in the Hoogsteen edge and N7. There are additional hydrogen bonds between the protein and other adenine atoms, but this new "motif" is the only one shared by all the proteins in the cluster (Figure 20, A and B, cluster 9, and Figure 22B).

4.3.4.5 Themes used in adenine binding

In the following section I list all the themes shared by the clusters in the dataset. The clusters numbering follows the scheme in Figure 20.

<u>Cluster #1:</u>

The cluster is composed of 20 binding sites, the vast majority of them (18) bind adenine by the direct form, as described by Denessiouk and co-workers. The themes shared by proteins in the cluster are numbered 867, 873, 896, 1354, 1355, 1356, 1357, 1648, 1924, 2421, 2422, 2423, 2527.

<u>Cluster #2:</u>

The cluster is composed of 20 binding sites, almost all of them (18) bind adenine by the direct form, and in addition, most of the binding sites (16) have an additional interaction between adenine and

the amino group of a lysine residue with adenine's N7. The themes shared by proteins in this cluster can be divided into three groups: themes that cover the entire binding pocket (includes themes 226, 227, 229, 230, 393, 395, 858, 859, 861, 126, 1327, 1328, 1384, 1389, 2451, 2457), themes that represent the direct motif (includes themes 232, 1325, 1384, 1385, 1388, 1392, 2332, 2416, 2417, 2419, 2452, 2453), and a single theme shared by almost all of the binding sites that include the additional N7/lysine interaction (228).

Cluster #3:

The cluster is composed of 18 binding sites, in all of them adenine's N6 in the Watson-Crick edge binds to the carboxylate group of an aspartate residue, and the rest of the interactions of adenine and its environment are with water molecules. The themes shared by proteins in this cluster donate this aspartate (themes 21, 23, 35, 289, 290, 292, 659, 1007).

Cluster #4:

The cluster is composed of 11 binding sites, binding either ATP (or its analogs) or SAM using the reverse motif. In addition, all of them have an additional interaction between adenine's N6 in the Hoogsteen edge and a backbone carboxyl group at 'position XV/XVI' (22/23 residues downstream to the residues forming the reverse motif, Figure 22A). The themes shared by proteins in this cluster usually form the full binding region (themes 1403, 1405, 1406, 1455, 1458, 1460, 1753, 1754, 1755, 2061, 2095), or only position XV/XVI (themes 1404, 1456, 1457, 1459, 2060).

Cluster #5:

The cluster is composed of 28 binding site and can be roughly divided into three connected-components connected in a linear order. Generally speaking, the proteins in the cluster bind adenine mostly by the reverse form, with an additional interaction of adenine's N6 in the Hoogsteen edge. This binding mode is especially common amongst the proteins in the center of the cluster.

Cluster #6

This interesting cluster contains 241 binding sites and can be divided into several sub-clusters each containing a different ligand's binding sites: SAM, NAD and FAD. For the SAM and NAD, the

dominate binding mode is usually the 'Asp' form of the adenine-binding motif, where adenine's N6 in the Watson-Crick edge hydrogen bonds to the carboxylate group of an aspartate residue, with an additional interaction between the protein and adenine's N3 (for SAM) or N6 in the Hoogsteen edge (for NAD). For FAD, the most common binding mode involves the reverse form of the adenine-binding motif, with an additional interaction between the protein and adenine's N3.

Cluster #6A

The cluster is composed of 81 nodes, the vast majority of them (66) bind adenine in the 'Asp' form, where N6 in the Watson-Crick edge hydrogen bonds to either to the carboxylate group of aspartate, the amide group of asparagine or the hydroxyl group of serine. In addition, most of the binding sites in this cluster (59) have an additional interaction, where adenine's N3 hydrogen bonds to the amide group of another residue, that is found 25-30 amino acids upstream to the amino acid binding N1. The themes shared by proteins in this cluster can be divided into three groups: themes that create the 'scaffold' of the binding (themes 186, 793, 1756), a theme that donate the interactions between the protein and the Watson-Crick edge of adenine (318), and themes that donate the interaction with adenine's N3 (31, 317, 631).

"bridging nodes" between cluster #6A and cluster #6B

There are three nodes connecting the cluster of SAM binding proteins with NAD binding proteins. Two of the nodes represent NAD binding sites, the third node represents ATP binding site. For the two NAD binding proteins, the theme connecting the proteins to the SAM binding proteins creates a large scaffold for the entire binding of adenine (theme 944). It has some overlap with the themes connecting them to the NAD binding proteins cluster (theme 949). For the ATP binding protein, the same theme (theme 944) has a smaller coverage, hence it is shorter. However, it includes an interaction between a backbone amide group and adenine's N3, and the variation of the direct form of the adenine-binding motif as described in cluster #6A.

Cluster #6B

This cluster is mainly characterized by the variation of the direct form of the adenine-binding motif, as described in cluster #5A, which is found in 15 of the 17 binding sites composing the cluster. Themes shared by proteins in the cluster indeed represent this binding mode (themes 100, 281, 977, 1363).

Cluster #6C

The cluster is, again, characterized by the 'Asp' form (found in nine of the ten proteins in the cluster), but with an additional interaction of adenine's N3 with the protein (or with a water molecule, in three proteins), and in some cases, interactions of adenine's N6 in the Hoogsteen edge either with the residues of amino acids or with water molecules. The themes in the cluster can be divided into three groups: a theme involved in the 'Asp' form (theme 623), a theme involved in binding of adenine's N3 (theme 279), and themes involved in the binding of adenine's N6 in the Hoogsteen edge (themes 101, 498, 648, 831, 832, 962).

Cluster #6D

A large cluster, composed of 113 nodes, most of them represent FAD binding proteins. The vast majority of the proteins in the cluster (104) bind adenine using the reverse form of the adenine-binding motif. In addition, I noticed an additional interaction with adenine's N3, in 103 of the proteins in the cluster, and interactions with the rest of adenine's binding atoms with water molecules in its environment. The themes shared by proteins in this cluster are divided into two groups, according to the binding mode they represent: the first group contains themes representing the reverse motif (themes 39, 40, 46, 320, 332, 333, 1055, 1200, 1202, 1219, 1220, 1221, 1454, 1663, 1683, 1734, 1921, 1968, 2086, 2329), while the second group donates backbone interaction with adenine's N3 (themes 331, 335, 365, 61, 981, 1919). In addition, I noticed less common themes, representing interaction with adenine's N7 (themes 461 and 561).

Cluster #7:

This cluster, with 11 binding sites in total, is composed mainly of FAD binding sites, but also includes two ATP binding sites and two BAD binding sites. Most of the binding sites (8) bind adenine via the

'Asp' motif, except for two binding sites using the reverse motif, and another binding sites where the interaction with adenine's N6 in the Watson-Crick edge is mediated by a water molecule. Seven of the binding sites have an additional interaction between adenine's N7 and a backbone amide group, and six of the binding sites have an additional interaction between adenine's N6 in the Hoogsteen edge and a water molecule. The themes shared by the binding sites in this cluster donate the 'Asp' motif toether with the interaction with adenine's N3 (themes 89, 92, 485, 488, 491, 492, 1020, 1023, 1025).

Cluster #8:

The cluster is composed of ten SAM binding-sites, nine of them bind adenine via the reverse motif, and in seven of them there is an additional interaction between adenine's N6 in the Hoogsteen edge and a backbone carbonyl group (in another binding site this interaction is mediated by a water molecule). Themes shared by the binding sites in the cluster donate both interactions, in a large binding-site scaffold (themes 516), or only form the additional interaction with adenine's N6 (themes 2300, 2301, 2344).

Cluster #9:

This cluster is composed of ten SAM binding proteins, all of which belong to PFAM's "SET" family. Their adenine binding pattern is unique: it is very similar to the reverse motif, except here the interactions are between the backbone amide and carboxyl group of the amino acid in "position III" and adenine's N6 in the Hoogsteen edge and N7. The themes shared by proteins in this cluster form this motif (themes 1125, 1127, 1128, 1843, 1845, 2314, 2325). Other interactions that are found in this cluster are not represented by themes share by proteins in the cluster.

Cluster #10:

The cluster is composed of 21 FAD-binding proteins, all of which bind adenine by the reverse form of the adenine-binding motif. In all of them there is an additional interaction between adenine's N6 in the Hoogsteen edge, mostly with water molecule. In addition, in 17 of the binding sites there is an additional interaction between adenine's N7 and mostly a water molecule. The themes shared by

proteins in the cluster form the reverse motif (themes 254, 255, 256, 257, 776, 777, 778, 889, 895, 987, 1173 2166, 2169, 2170, 2179, 2397, 2398). It is noteworthy that many of these themes are quite long, some cover more than hundred amino acids.

4.3.4.6 Adenine-binding themes in the PDB

My analysis detected themes that tend to be a part of adenine-binding sites. However, this does not necessarily mean that the detected themes have been selected during evolution for this task. To estimate the degree of 'dedication' of the theme in Figure 20 to adenine binding, I searched for them in the entire PDB. 47% of the PDB entries that contained these themes also included adenine-containing ligands. Since these ligands appear in only 9% of the entire PDB, the results indicate that adenine-binding proteins are enriched with the themes in Figure 20. Furthermore, I found that about half of the PDB entries, which contained an adenine-binding theme but no ligand, shared at least 80% of their sequence with those that did contain adenine in their structure. This could indicate that although the structures of these proteins were determined without a bound ligand, their biological function requires interaction with an adenine-containing ligand. This further supports the evolutionary link between the themes and adenine binding. Encouraged by the results, I composed a list of 1,144,830 UniProt [8] entries in which I detected adenine-binding themes. The list includes almost 4% of the UniProt entries which I predict are likely to bind adenine (or an adenine-containing ligand).

5 Discussion

It is not practical to believe that proteins emerged by linking amino acids one at a time, as the number of possibilities is way too high (around 20^{100} possibilities to construct an average-sized domain), and most of which won't fold. In addition, the need to conserve function drives proteins' evolution. As protein's function is tightly connected with its dynamic flexibility, it can be used to study the physicochemical forces that underlie this range. Complementary, focusing on sequence similarities between proteins sharing the same molecular function can enable scientists to gain a better understanding of the evolution of molecular function. My PhD research navigated between these two approaches.

5.1 ConTemplate

Global structural similarities between proteins may indicate similar dynamic properties. ConTemplate is the first automatic implementation of this commonly used concept, and it exploits it together with the abundance of structures in the PDB to enable scholars to study functional aspects of query proteins. The templates it uses are limited to the structures found in the PDB. Since these templates are assumed by proteins, it increases the chances that they may be biologically relevant to the protein of interest as well, even though it limits the suggested conformations.

ConTemplate computes pairwise sequence alignment between the query and each of the templates based on the structural alignment between the proteins in the shared conformation. As structure is overall more conserved than sequence, that should produce better alignments that those based on sequence similarity alone [123]. This is the case especially when the template shares low sequence identity with the query - as, for example, with the ribose-binding protein demonstration. I used this example to test the usability of ConTemplate's approach, and also to analyze its limitations. To avoid the trivial outcome in which ConTemplate only finds the known structures of the ribose-binding protein, I reduced the sequence-identity threshold between the query and suggested

templates to be no more than 50%. This allowed me to test ConTemplate's suggested conformations in terms of their biological relevance for the query. With this I could also investigate the limitations of the web-server which are the result of the non-uniformed distribution of structures in the PDB, as expressed by the commonness of some of the conformations compared with others.

ConTemplate treats the differences in the popularity of some of the conformations in respect to others in two ways. First, when it gathers all the possible templates it clusters them. The clusters are formed based on the inner-distance matrix of the templates, so that eventually the clusters differ in the conformations they represent. Second, ConTemplate visualizes the distances between the suggested models [116], by creating a network where each node represents a model, and two nodes are connected if the RMSD between the two models is under 2.5Å, with an edge whose length corresponds to that distance. In other words, two nodes connected by a relatively short edge will represent two rather similar models. This visualization enables the user to see how many different conformations are suggested, and also to find pathways between different conformation-clusters. The user can set most of the similarity thresholds, as well as the number of clusters. Thus, increasing this number will make it easier to detect less common conformations (which could be less common as their structures are harder to obtain), and suggest pathways between conformations. Limiting the number of clusters will enable to find the "consensus conformations", hence conformations that are more easily obtained and shared by many PDB proteins.

When two or more conformations are known, it is interesting to study the pathway between them, as a mean to test our understanding of proteins' dynamics. Several works studied these transitions, using different approaches [124-129]. When ConTemplate suggests a pathway between two conformations, the conformations along this pathway can be milestones in the transition between the two conformations as detect by other tools. I am now investigating the feasibility of this approach in collaboration with Turkan Haliloglu (Bogazici University, Istanbul, Turkey).

5.2 Using ConTemplate to study the dynamics of the CueR copper-regulator

Once I validated ConTemplate's approach and the quality of its results, I could use it to search for unknown conformations of query structures. I this example, I used it to give structural interpretation to experimental data. This is another example of the way ligand binding has an influence of the conformations changes of proteins. In this work we employed DEER results with ConTemplate's templates and elastic-network analysis to better our understanding of the flexibility of CueR upon Cu(I) and DNA binding.

The Ruthstein lab spin-labeled the CueR protein in different locations to measure distances regarding conformational changes in different parts of the protein: in the DNA binding domain, the dimerization helices, and the N-terminal domain. Next, they sampled the protein in different states: bound/unbound to DNA and with or without Cu(I). The largest change corresponded to changes in the DNA-binding domain, where the distance between the DNA-binding arms of the two monomers increased upon DNA-binding in the presence of Cu(I), while the distance between the spin labels of the intra-monomer in the DNA-binding arms decreased. In other words, the two monomers spread further apart, and the two regions of the DNA-binding arms in each monomer got closer together. To get a better understanding of the described changes, they repeated the experiments, this time binding either DNA or Cu(I). The results from these experiments suggested a population shift model [1], where the protein can reside in three different conformations: apo-CueR state, repression state in which CueR binds to DNA but with no Cu(I), and activation state where CueR binds both DNA and Cu(I). CueR in cells is always bound to DNA, so the population shifts between CueR-DNA states with or without Cu(I). In other words, copper binding shifts the population from repression state into activation state, and so initiates the transcription. In addition, a comparison between ConTemplate's models also showed that the DNA binding domain is the most varied domain of the protein, which makes sense as the different templates used for modeling represent different members of the MerR metalloregulator family, each binding its DNA sequence at a very high affinity.

The combination of experimental and computational results enabled us to suggest a mechanism for the repression/activation of the transcription process regulated by the CueR. DNA binding to the apo-CueR in the repression state causes changes in the DNA-binding arms of both monomers, causing the α1 helices to grow further apart, and as a results the intra-monomer distances between the α1- α2 and α2- α3 regions get closer. Upon Cu(I) binding, the protein enters its activation state. The DNA binding domains of both monomers get closer, the two α1 helices get even more distant, but the α2- α3 domains get closer to the α1 helix. This can be thought of as a squeezing motion of the DNA-binding arms, which either bend the DNA (in the repression state) or enable it to spread (in the activation state). In the activation state it seems likely that the α1, α2 and α3 coordinate with the DNA and so can promote transcription. The protein constantly changes between the different conformational states, suggesting a population shift model where high copper concentration shifts the equilibrium towards the activation state.

This study enabled us to provide insights into the activation and repression mechanism of the CueR upon ligand binding. It shows the usability of ConTemplate in suggesting a structural interpretation to experimental data. In addition, it is an example of the way ligand binding has a major contribution to the dynamics of proteins, and that even in the same protein, different ligands, in different binding sites, can be of different sizes and have different chemical characteristics.

5.3 On the evolution of adenine binding

The first parts of my thesis dealt with conformational changes in protein, especially upon ligand binding. The rest of my thesis examines the evolution of protein-ligand interactions, taking the binding of the ancient ligand adenine as a convenient example. Two properties of protein-adenine interactions suggest that studying them is useful not only for understanding nucleotide binding, but also to decipher general evolutionary principles: (1) ligand binding is the hallmark of protein function and (2) these interactions are highly conserved. Adenine is part of numerous nucleotide cofactors which are involved in the most central pathways of metabolism in all three kingdoms of life (e.g., the Krebs cycle

[130]). I studied protein-adenine interactions to investigate a link between themes and biological function. I explored the patterns proteins use to bind the adenine fragment of different nucleotide cofactors.

To do this I constructed a large dataset of protein-nucleotide complexes available in the PDB. The quality of the structures in the dataset is quite diverse, with resolutions ranging between 0.74 and 6.93Å (only 31 structures have resolutions worse than 3Å) and R-scores ranging between 0.117 and 0.341. However, the vast majority of structures (above 75%) have resolutions of 2.5Å and R-scores of 0.25 or better. Furthermore, I observed the same main results in preliminary analyses of other datasets. In general, I am not particularly concerned with the interactions that I report, since I observe them in multiple structures, which, in essence, eliminates the possibility that they are erroneous. The only concern might be to miss interactions due to insufficiently high resolution.

In analyzing the noncovalent interactions between proteins and adenine I focused on hydrogen bonds for several reasons. First, hydrogen bonds are common in proteins, and specifically in protein-ligand interactions [1]. Second, hydrogen bonds are highly specific, due to the dependence of their free energy on their geometry (i.e., on bond length and angles). This, in turn, also makes hydrogen bonds easier to identify than other protein-adenine interactions, such as those involving π electrons (π-π and cation-π interactions). The latter interactions are weaker than canonical hydrogen bonds, less common, and their energy dependence on chemical and geometric characteristics is much more complicated. Obviously, ligand binding also involves van der Waals and non-polar interactions. Generally, these contribute to the affinity, while hydrogen bonding adds to specificity [1].

The results depend strongly on the geometric definitions of hydrogen bonds I use, which is a known problem [131-134]. I chose the definitions used by Arpeggio [110], which are in agreement with other commonly used tools. In addition, the geometric comparison of binding sites in my analysis is sensitive to the thresholds I use for similarity. Specifically, the thresholds that determine how close two hydrogen bond donors or acceptors from two binding sites must be to each other, to be

considered as equivalent. When I use more lax thresholds, the network grows much larger and it is harder to trace the clusters (Figure 23A). However, the binding sites are also considerably less similar in this case. I used relatively strict thresholds (low RMSD) to ensure that equivalent hydrogen bonds are detected, yet allowed some variability. Using more strict thresholds would limit the variability to the point where it breaks the network to numerous connected components (Figure 23B).

Previous studies focused on the interactions of proteins with the Watson-Creek edge of the adenine moiety, overlooking the interactions with the other edges of the molecule. My results, based on a 5-times larger and redundancy-reduced dataset, confirm the involvement of the interactions motifs proposed by previous studies in the binding of adenine, but show that proteins tend to interact with all the edges of the adenine moiety, and with various functional groups included in it. These tendencies, which I found in complexes of all the adenine-containing cofactors, are consistent with the evolutionary drive to optimize favorable interactions within binding sites, and by this, increase binding strength and specificity. On the protein side of the interactions, I identified the interaction motif suggested by Denessiouk and co-workers, which the authors subdivided to 'direct', 'reverse' and 'Asp' variations, based on the relative positions of the interacting amino acids. I identified this motif in the binding sites of all adenine-containing cofactors that were considered by Denessiouk and co-workers. However, in contrast to these studies, I found that the most common variation of the motif is the 'Asp', which is found in the binding sites of almost all of the cofactors in the dataset. I also found that water molecules often mediate binding interactions between groups that are too far from one another to interact directly, and, in some cases, replace amino acids as hydrogen bond donors and acceptors. Since I observed protein groups or water molecules, at equivalent positions that interact with the ligand (e.g., Figure 18), it could be that these water molecules are conserved in certain types of the complexes. Indeed, I found conserved water molecules in different adenine-binding families (e.g., see Figure 19C). The conservation of water molecules within protein binding and catalytic sites has been documented in many cases and has even been linked to the protein's function [135, 136].

The analysis of ATP-binding proteins showed that whereas the same binding geometry is often shared by proteins with similar sequences (i.e., protein families), it may also appear in proteins with very low sequence and structure similarity. Thus, ligand-binding patterns in proteins, as reflected by the interaction site and binding mode, seem to be more conserved than the overall sequence and structure. Therefore, such patterns may possibly be used to assign newly discovered proteins to their corresponding families. Clearly, a more extensive study of various protein families and their cognate ligands is required to determine the feasibility of this approach.

Interestingly, when I analyzed ATP binding to proteins that transport it across membranes, I found that the binding does not involve the adenine moiety, but rather the ribose and phosphate moieties of the nucleotide. The exclusion of adenine from the interactions should decrease the binding affinity of the transport proteins to ATP, which is in line with their biological function; while ATPases must bind their substrate strongly to catalyze its hydrolysis, proteins that transport ATP must bind it more weakly, to facilitate the transport process. If true, this explanation supports the mechanistic role of adenine as an 'interaction handle', which ensures binding of nucleotide cofactors to enzymes with sufficiently high affinity.

Ligand-binding patterns in proteins are determined by the structures and amino acid compositions of their binding and interaction sites. Therefore, to better characterize the determinants of the adenine-binding interactions I focused on the binding sites' building blocks, which determine the patterns of these interactions, as a mean to study protein evolution. The concept of themes, as proposed by Nepominyachy et al., challenges the widespread notion that only domain-level units constitute the evolutionary and functional building blocks of proteins; it suggests that evolutionary units may be much smaller than domains. My results support to this suggestion by demonstrating that certain themes, as defined by Nepominyachy et al., tend to appear in nucleotide-binding sites (and interaction sites), where they participate directly in the binding. Specifically, I found the following. First, certain themes are associated with the binding of specific nucleotide ligands (Figure 20), and their presence in a given protein predicts with high certainty that this protein binds adenine. Second,

most of the themes interact with nucleotide ligands in a specific way, and many of the themes constitute parts of known nucleotide-binding motifs (Figure 21, A and B), or even extend them (e.g., clusters 5 and 9 in Figure 20, and Figure 22, A and B). Third, certain theme combinations tend to appear together in interaction sites (Figure 21, C and D). Again, this further supports the notion of themes acting as evolutionary building blocks of proteins. Moreover, my work is the first to connect the themes, detected based on the frequency of certain sequences in proteins, to a molecular function. The results show that proteins containing certain themes are about 5 times more likely to bind adenine than a random protein in the PDB, and that about 4% of UniProt's proteins could bind adenine. This demonstrates how themes can be used in function prediction. It is also noteworthy that combinations of themes, which often recur in the binding sites of different proteins, form the same 3D-geometry and the physicochemical environment required for the binding.

5.4 Conclusions

My PhD research focused on the complex relationships between protein's sequence, structure and function, mostly from the perspective of ligand binding, being the most fundamental of all protein functions. In the first part of my PhD I developed ConTemplate, a method and web-server that suggests known and alternative conformations for a query protein with at least one known structure, based on structurally similar proteins and their additional conformations. I demonstrated ConTemplate using the ribose binding protein, which is a great example of how ligand binding induces large-scale conformational changes in proteins. In the second part of my PhD, I used this notion in a collaboration with Sharon Ruthstein (Bar Ilan University) to study conformational changes in CueR in response to its ligand binding. This study also demonstrated how ligands can have very different characteristics. In my third project, I delved into the binding itself, and aimed to study how proteins evolved to bind and recognize their ligands. As I intended to study proteins' evolution, focusing on nucleotide cofactor – which are ancient and important for key cellular functions – seems natural. In this study, I focused on two key aspects of protein-adenine interaction. The first is the physicochemical

nature of the binding, as reflected by the hydrogen bonding interactions between the protein and ligand, and by the structural motifs that support this binding. The second aspect is the evolutionary basis of the binding, as reflected by the role of highly reused sequence 'themes' in the binding.

Adenine offers many hydrogen donors and acceptors on the Watson-Crick edge, the Hoogsteen edge, and the sugar edge, and as I show, the protein may take advantage of various combinations of these, demonstrating the opportunistic nature of evolution. In addition, I correlate this knowledge with evolutionary trends in adenine-binding proteins, as reflected by their theme composition. That different geometries and binding patterns were found in the various proteins suggest that adenine binding emerged more than once in evolution. Future efforts can be directed in creating an evolutionary path that underlies the gradual construction of larger themes, and of related binding sites (and hence, interaction sites) from an initial finite set of short themes.

On a broader level, my results support the notion that themes are not merely recurrent protein sequences, but they could in fact be conserved functional units, or at least some of them are. It is tempting to suggest that ligand binding has emerged from combinations of such functional building blocks: segments of varied sizes, which formed interactions with the ligand, and so were adopted by evolution and coupled together to form larger interaction sites. The proteins in my dataset share sequence identity of 30% or lower only, and adopt different folds, yet many of them use the same themes to bind the shared adenine moiety, which could also suggest a convergent evolution process.

The procedure that I introduced here can be used to study the binding of virtually any ligand containing a rigid fragment. In addition, it can be used to explore themes that are involved in other molecular functions, e.g. catalysis, and by this to construct a functional "theme vocabulary". I demonstrated potential usage of such a vocabulary with the examples of adenine binding themes. Such a vocabulary can help the assignment of a function to newly discovered proteins, as shown in my UniProt predictions; or identify binding sites, as I found potential adenine-binding sites in the PDB. In addition, it can improve methods of molecular docking, and add functionalities to designed proteins.

This is the first work that links themes to a well-defined biological function, and by that sheds light on the relationship between the sequence, structure and function of the evolutionary building blocks forming proteins. It paves the way to future works in this field, which will ultimately help scientists reveal how the complex proteins universe has emerged.

6 Tables and figures

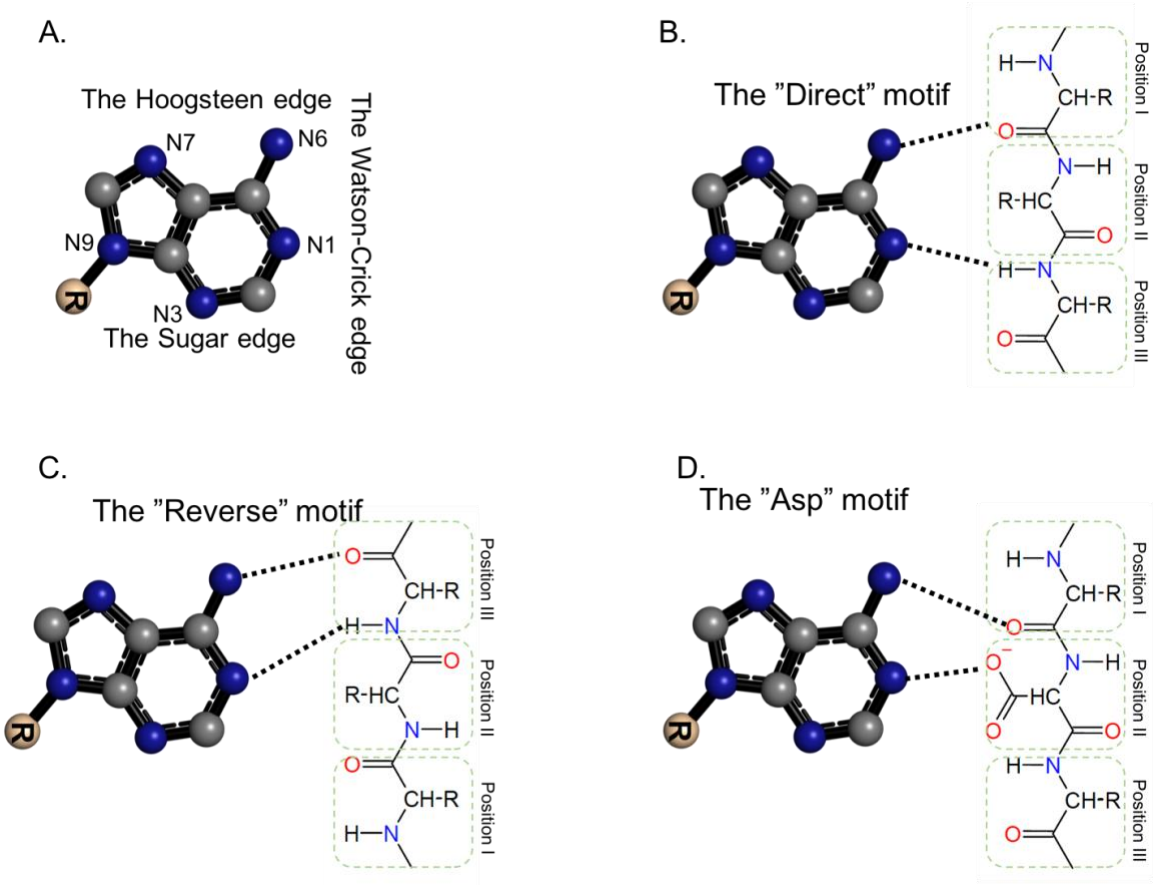


Figure 1. Previously documented adenine binding motifs (adapted from [58]). Only the three most common ones are presented: A: Adenine (within ADP) with the conventional atom numbering and binding edges. Carbon atoms are marked as grey spheres, nitrogen atoms in blue, oxygen atoms in red, and phosphate atoms in orange. B, C, and D: The 'direct', 'reverse', and 'Asp' motifs of adenine binding along the Watson-Crick edge. Hydrogen bonds are marked with dashed lines.

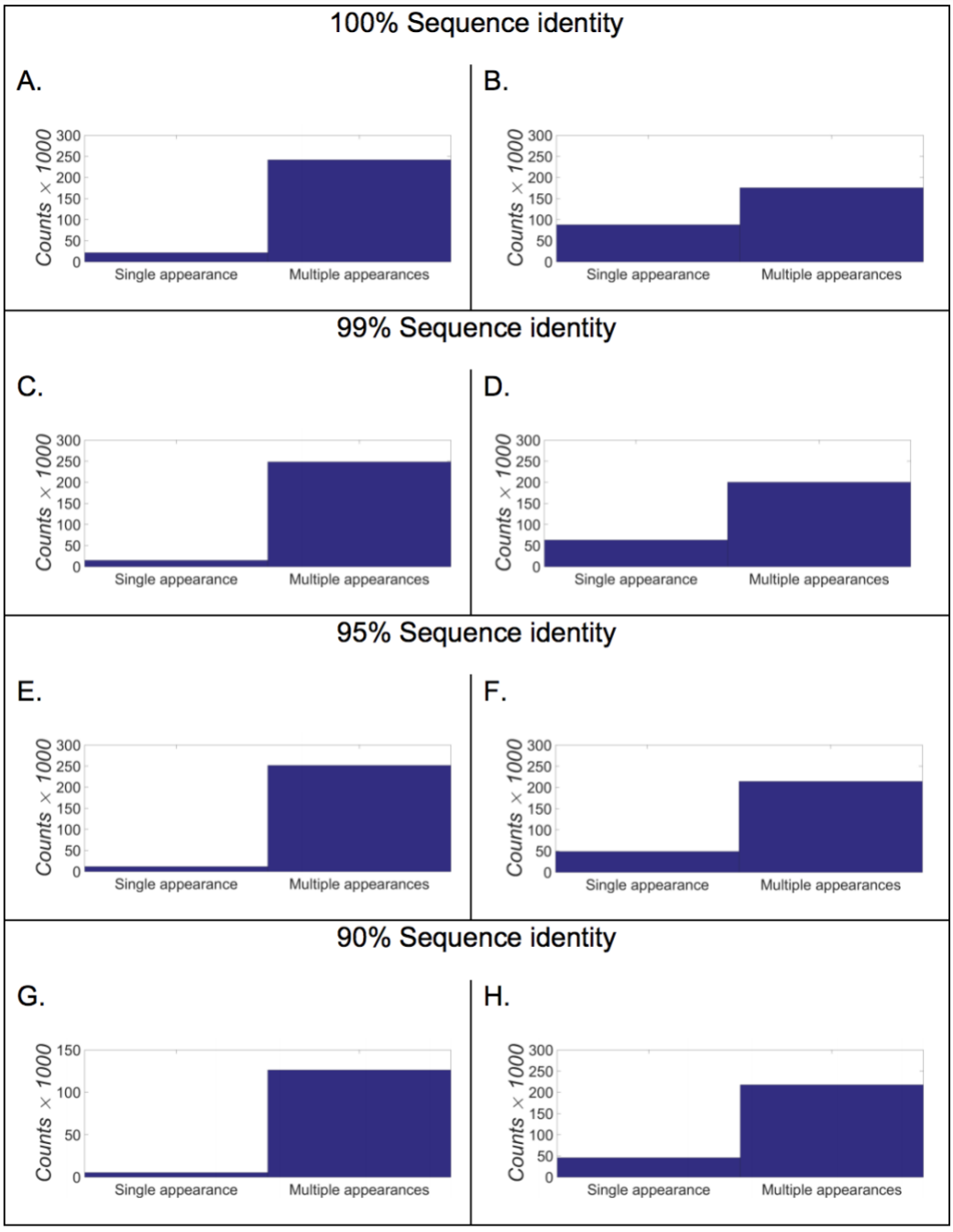


Figure 2. Most of the PDB chains appear in multiple entries (taken from [34]). A. The number of chains that are found only one time, in a single PDB entry, vs. these appearing two times or more in the PDB (100% sequence identity and full coverage). B. The number of chains that are found only in one PDB entry, vs. these appearing in two or more PDB entries. C & D. Same as "A" and “B” for 99% sequence identity. E & F. Same as "A" and “B”, for 95% sequence identity. G & H. Same as "A" and “B”,for 90% sequence identity.

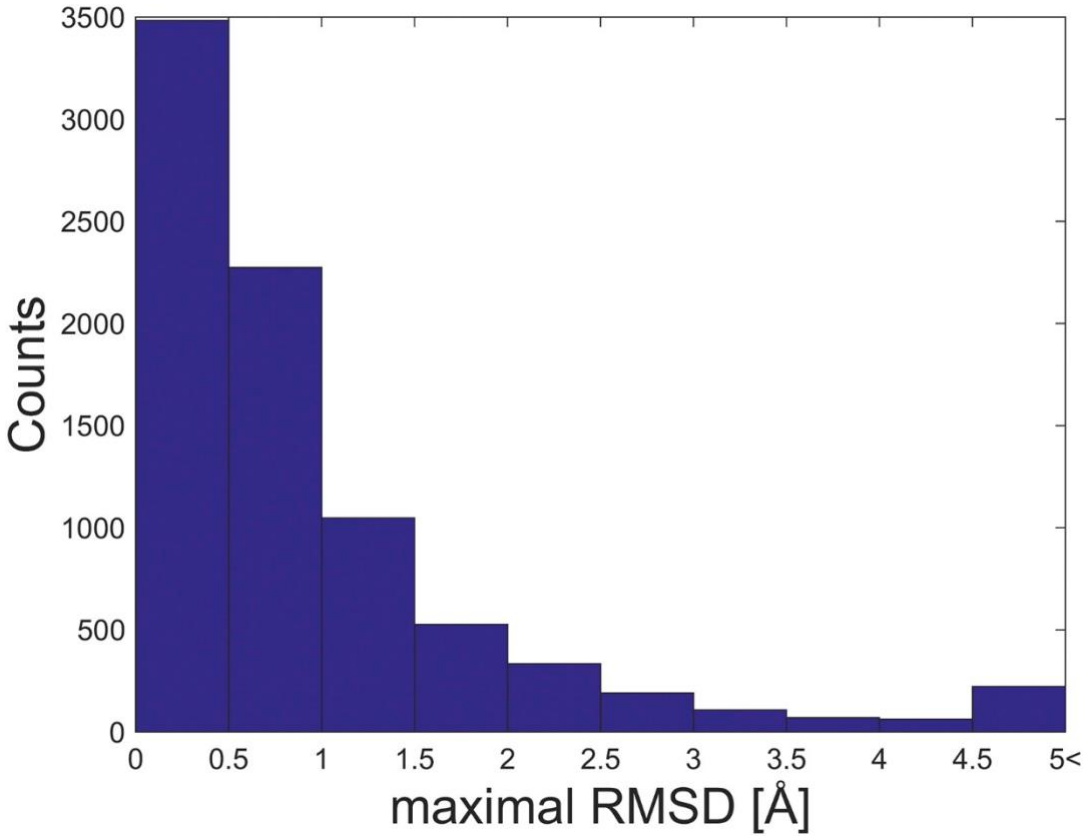


Figure 3. The maximal RMSD between two structures of the same protein (taken from [34]). The size of the maximal RMSD between two PDB chains representing the structure of the same protein, in a dataset containing 8,322 protein chains. For the vast majority of proteins, this RMSD is less than 1Å.

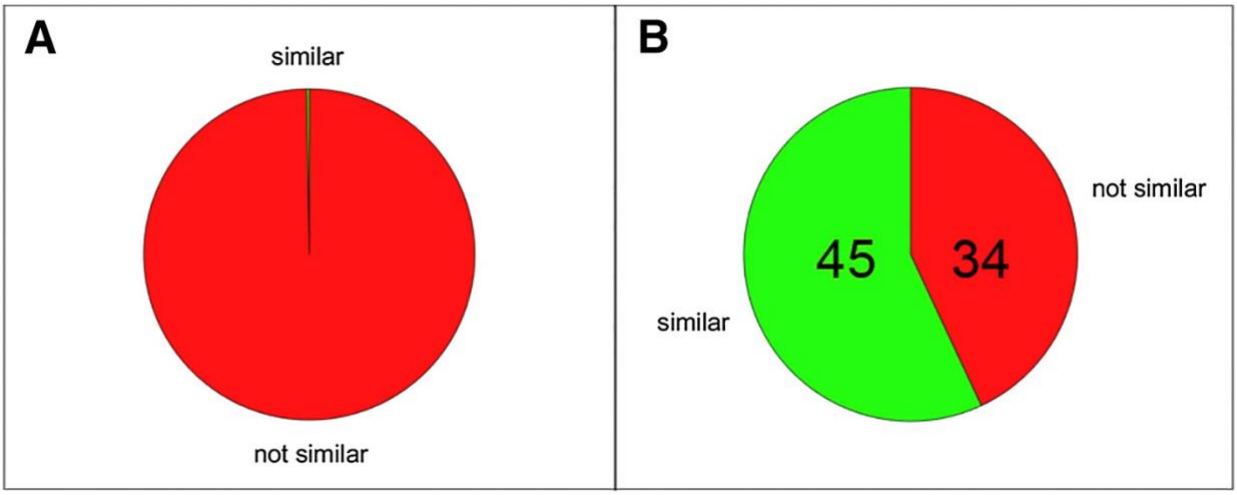


Figure 4. Proteins sharing one conformation often share others as well (taken from [34]). A. Protein pairs with at least one similar conformation in a collection of 246 proteins. 79 of the 30,135 pairs (0.3%) have a similar conformation. B. Of the 79 protein pairs in A that share one conformation, 45 have additional conformation(s) in common.

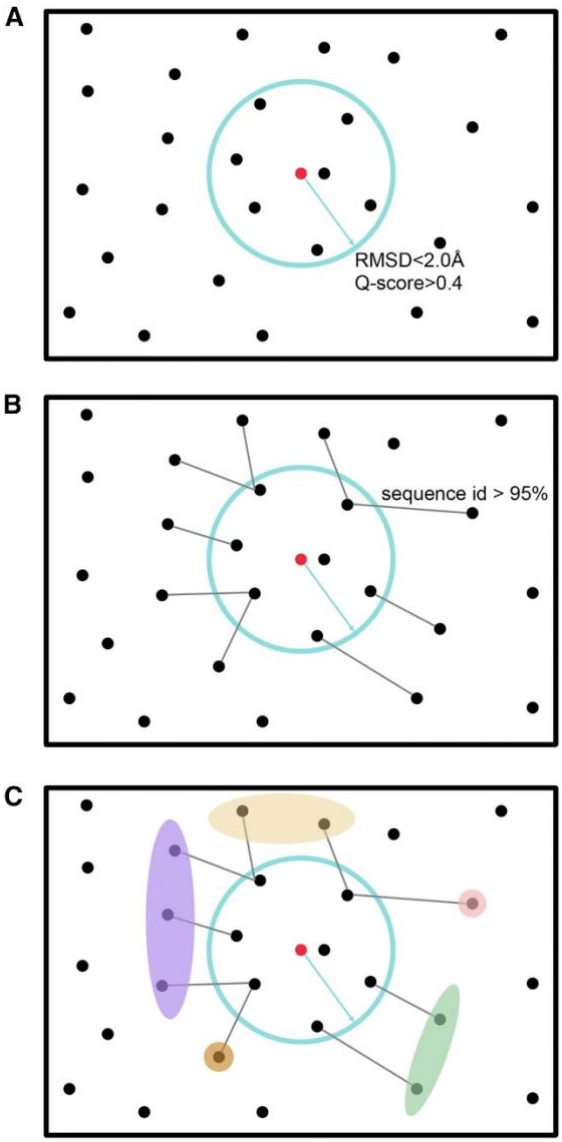


Figure 5. ConTemplate methodology (taken from [34]). The points represent all proteins' PDB structure, the red point represent the query structure. A. Searching for structurally similar proteins. ConTemplate uses the GESAMT structural-aligner to identify all PDB proteins that are structurally similar to the query. The blue circle represents the region in the protein-structure space where proteins are structurally-similar to the query. The radius of the circle corresponds to the user-defined similarity thresholds: the RMSD and the Q-score (default values are listed in the figure). B. finding all the additional conformations for each of the structurally-similar proteins. Using BLAST, ConTemplate identifies alternative conformations for each structurally-similar protein found in A, including the query. The grey edges connect points representing different structures of the same proteins. The

sequence-identity threshold for the search is user defined, the default threshold is listed in the figure. C. clustering the proteins found in B. five clusters are shown. Their centers (defined as the point with the shorter distance to all other points) are used to model the query in the conformation represented by the protein-structures in the cluster. The user can configure the number of clusters.

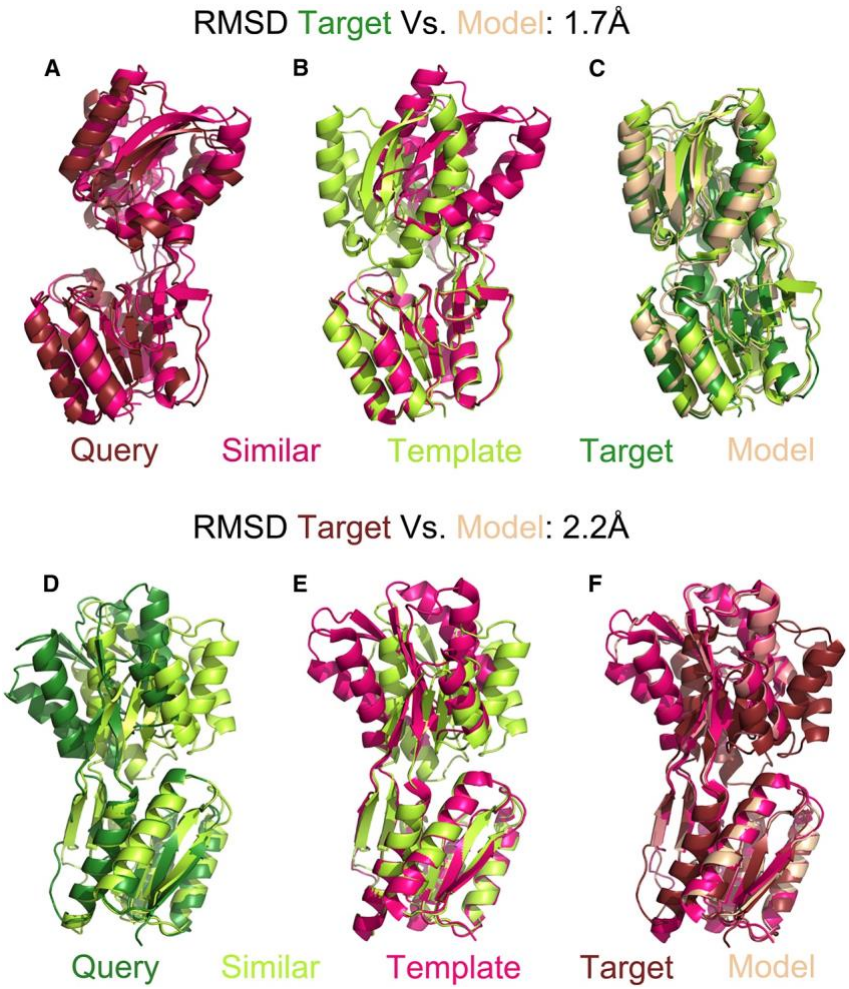


Figure 6. Using ConTemplate to model conformational changes of the ribose-binding protein (taken from [34]). Upper: Using the open (ligand-free) conformation as a query (Query, PDB: 1URP) and reproducing the closed, ligand-bound conformation (Target, PDB: 2DRI). Lower: Using the closed conformation as a query and reproducing the open conformation. The RMSD between the two conformations is 4.1Å. ConTemplates thresholds used for the demonstration: the maximal RMSD between the query and structurally-similar proteins is set to 2.0 Å, and the minimal Q score is set to 0.4. In the upper panel the number of clusters is set to 2. In the lower panel it is set to 20. A. and D. Searching for proteins that are structurally-similar to the query; only one is shown (Similar, PDB: 3M9X, 1RPJ). B. and E. Identifying the alternative conformations of the proteins detected in the previous step; only one alternative conformation is shown (Template, PDB: 3MA0, 1GUB). C. and F. Modelling putative conformations of the query based on the structures found in the previous step; only one model is shown (Model).

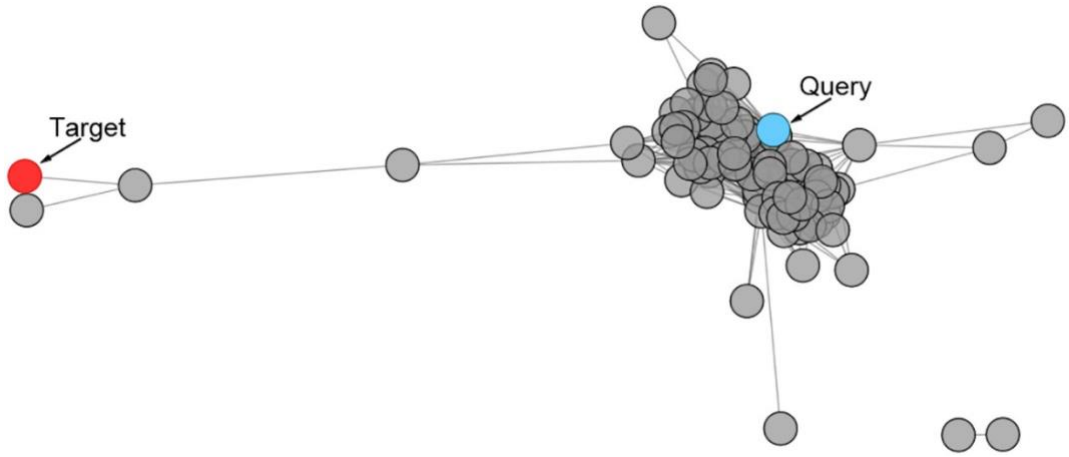


Figure 7. A network representation of ConTemplte's results can suggest pathways between conformations. The network is presented using Cytoscape (taken from [34]). The query conformation is in blue (the ribose-bound conformation of the ribose-binding protein). The target conformation is in red (the ligand-free conformation of the same protein). The gray nodes are ConTemplate's suggested models. Edges connect nodes where the corresponding structures have RMSD under 2.5Å. The length of the edge connecting two models is proportional to the RMSD between them. Outlier nodes (bottom-right) correspond to models obtained using templates that are biologically irrelevant to the query, and so could be considered irrelevant conformations to the query.

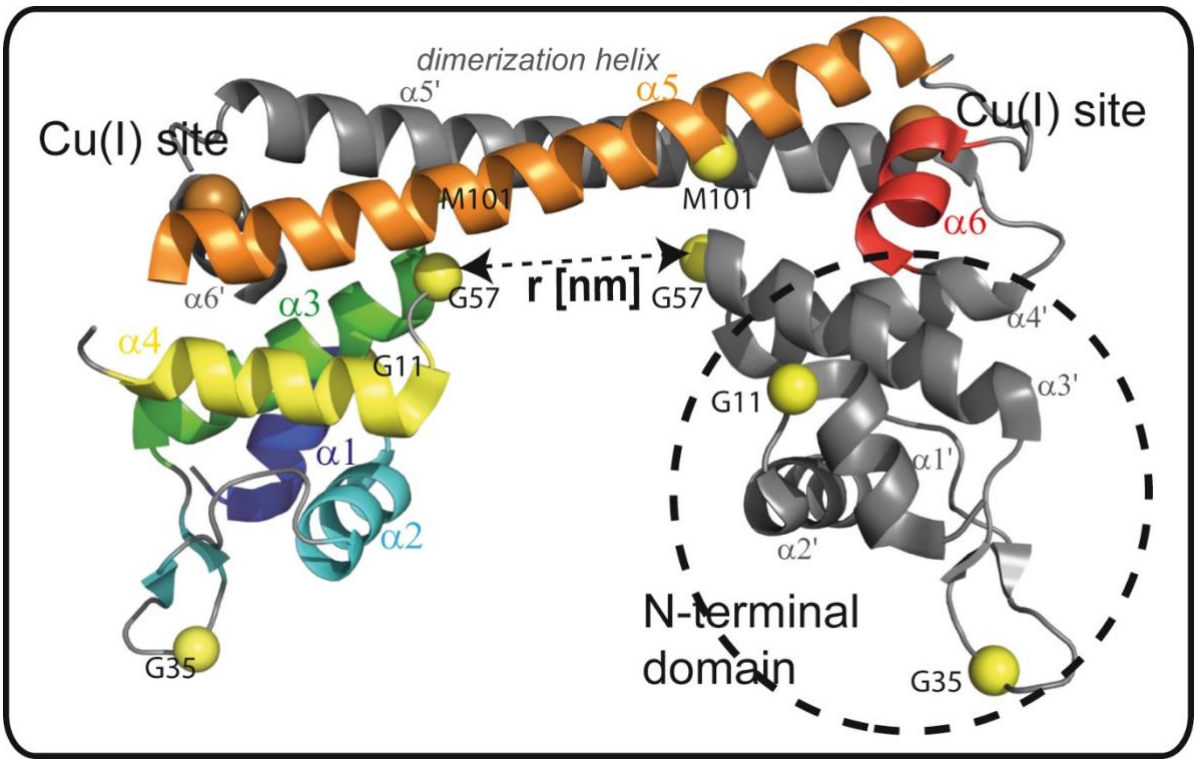


Figure 8. The CueR dimer (pdb 1q05) (taken from [56]). The yellow spheres mark the spin-labeled sites, orange spheres are the Cu(I) ions. The N-terminal domain of one monomer is encircled with dashed lines.

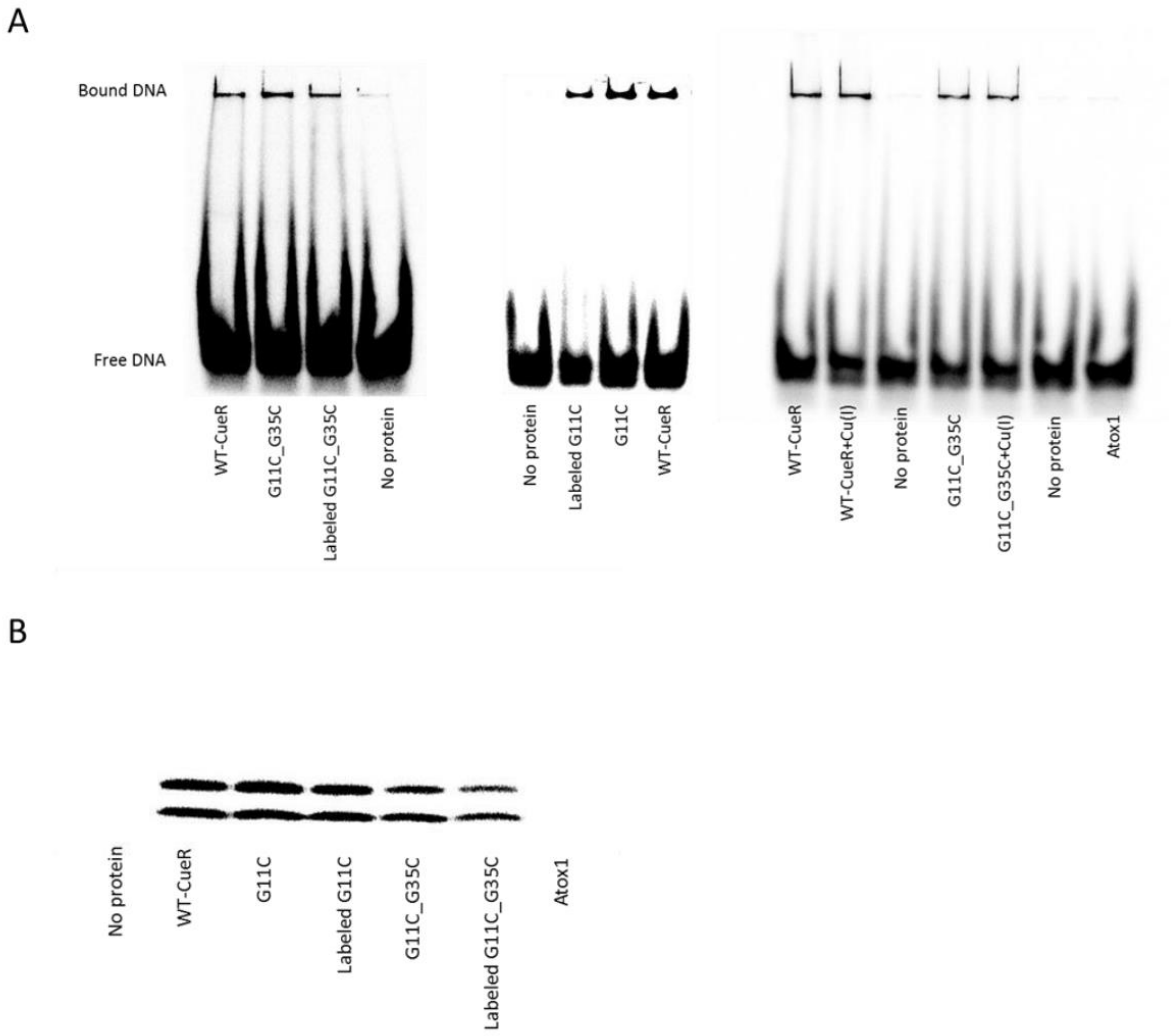


Figure 9. Mutations used for spin-labeling does not interfere CueR with binding to the promotor (taken from [56]). A. Electrophoresis mobility shift assays by fluorescence for CueR mutants. The control is Atox1, a copper chaperon which does not bind to DNA. B. Pull down experiments for CueR mutants.

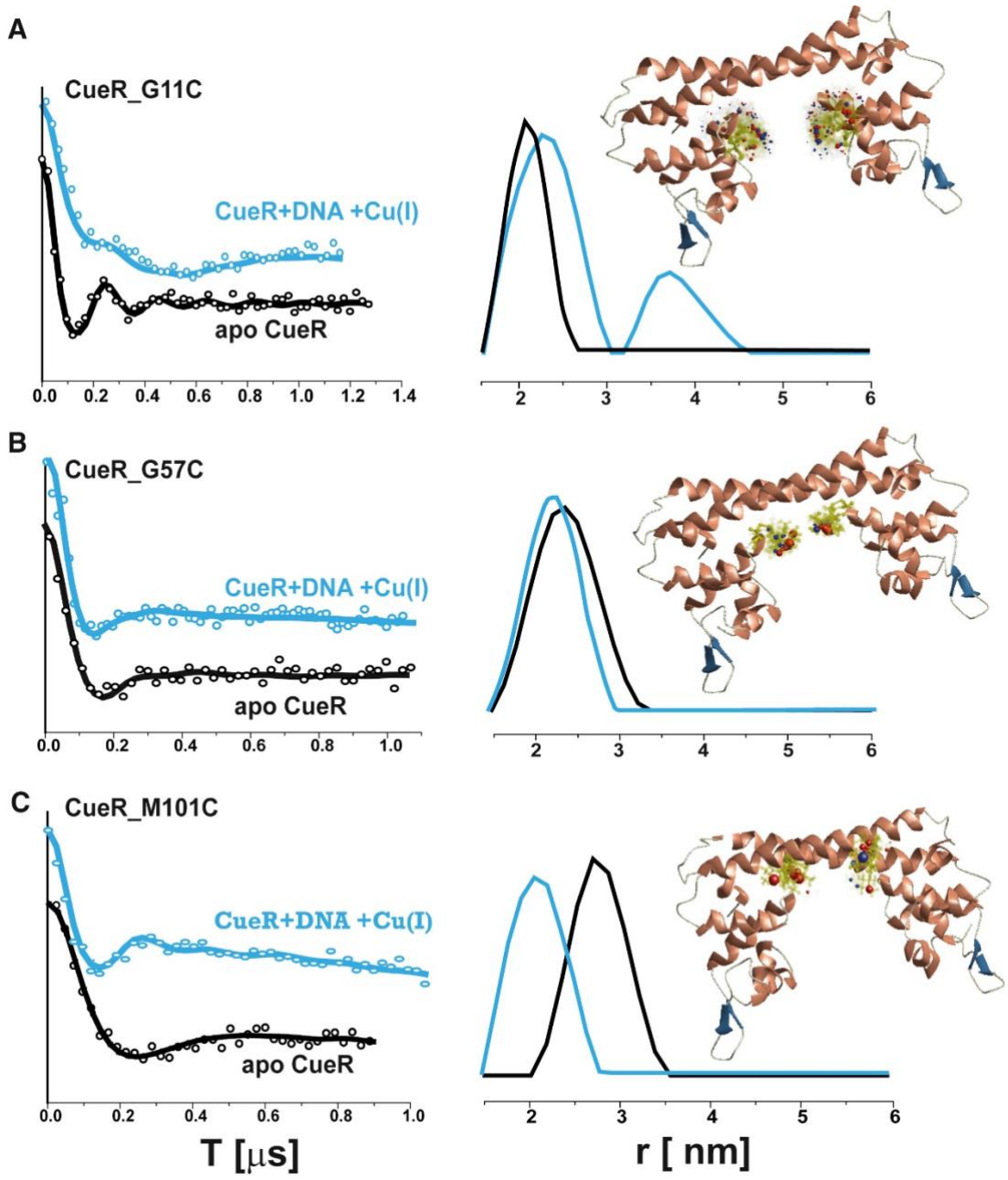


Figure 10. DEER measurements for the conformational changes of the CueR upon Cu(I) binding (taken from [56]). The DEER signals (left) and the corresponding distance distribution functions (right) measured for the apo-CueR (black curve) and for the complex CueR+DNA+Cu(I) (blue curve). To the right of the panels, there is the PDB structure of CueR dimer (pdb id 1q05), with the distribution of the spin-label conformations as calculated using the MMM 2015 software. The spin labels are attached to A. CueR_G11C, B. CueR_G57C, and C. CueR_M101C.

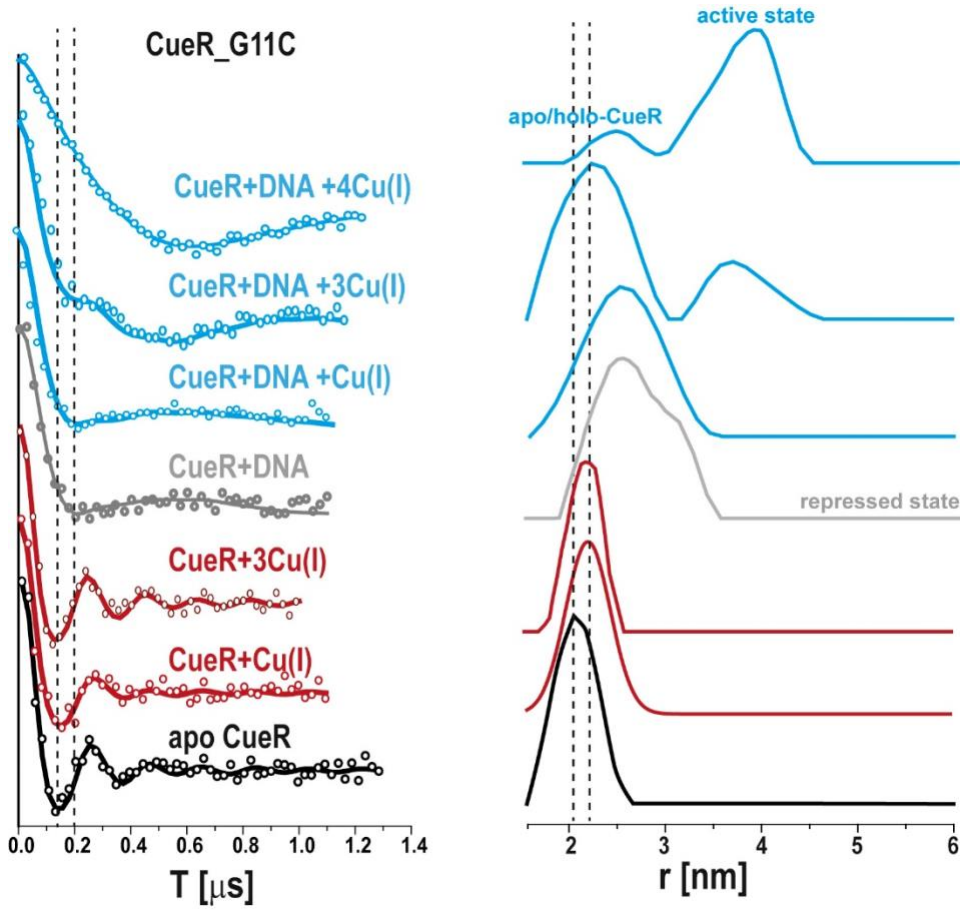


Figure 11. DEER measurements with varied DNA and Cu(I) concentration (taken from [56]). The DEER signals (left) and the corresponding distance distribution functions (right) measured for the CueR_G11C mutant in the presence and absence of DNA, with varied Cu(I) concentrations.

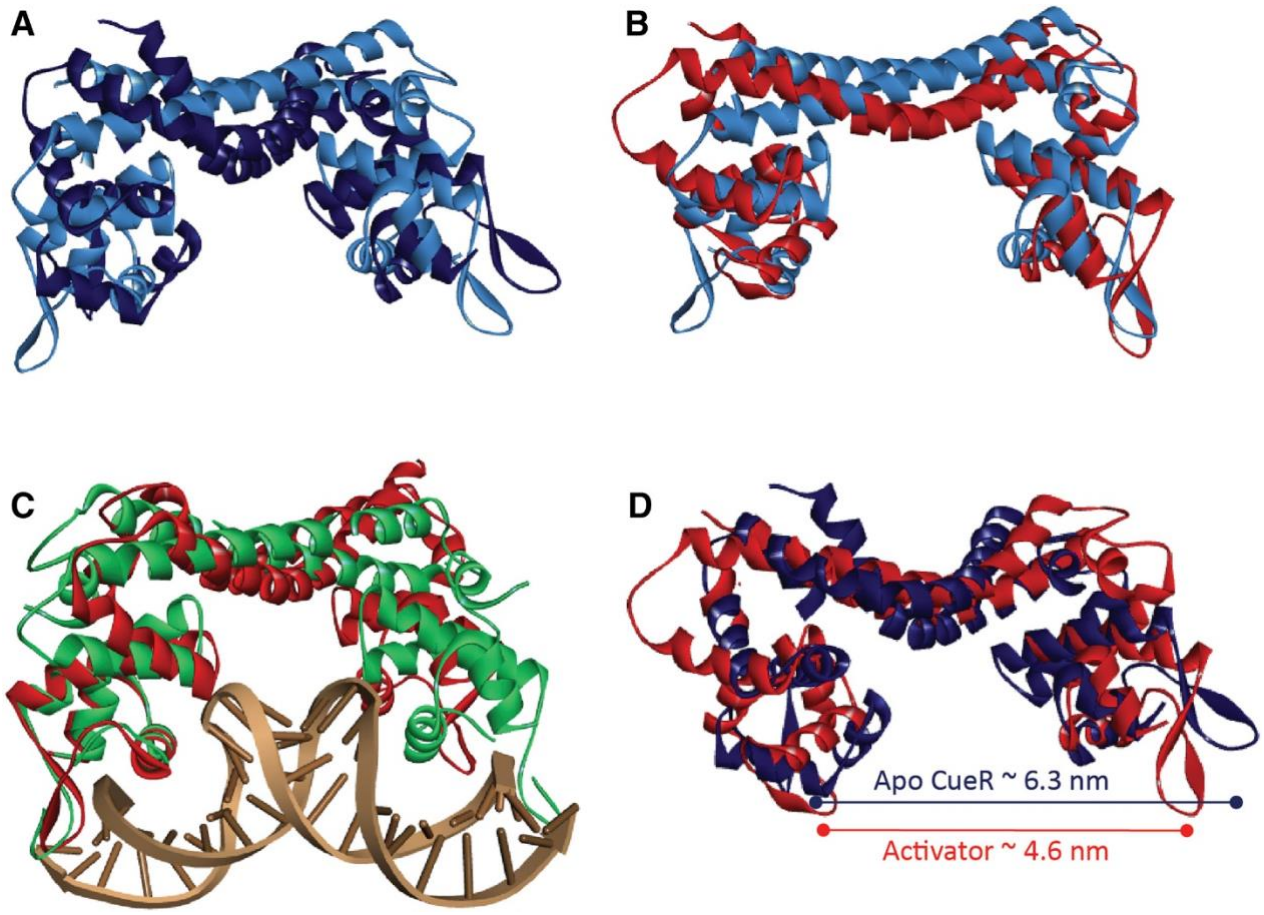


Figure 12. Elastic network models for the apo-CueR and CueR-DNA-Cu(I) structures, based on their DEER constraints (taken from [56]). A. The apo-CueR model-structure (dark blue) overlaid on the crystal structure of Cu(I)-CueR, pdb id 1q05 (light blue). B. The CueR-DNA-Cu(I) model-structure (red) overlaid on the crystal structure of Cu(I)-CueR, pdb id 1q05 (light blue). C. The CueR-DNA-Cu(I) model-structure (red) overlaid on the crystal structure of CueR-Ag(I)-DNA, pdb id 4wlw (green). D. The apo-CueR model-structure (dark blue) overlaid on the CueR-DNA-Cu(I) model-structure (red).

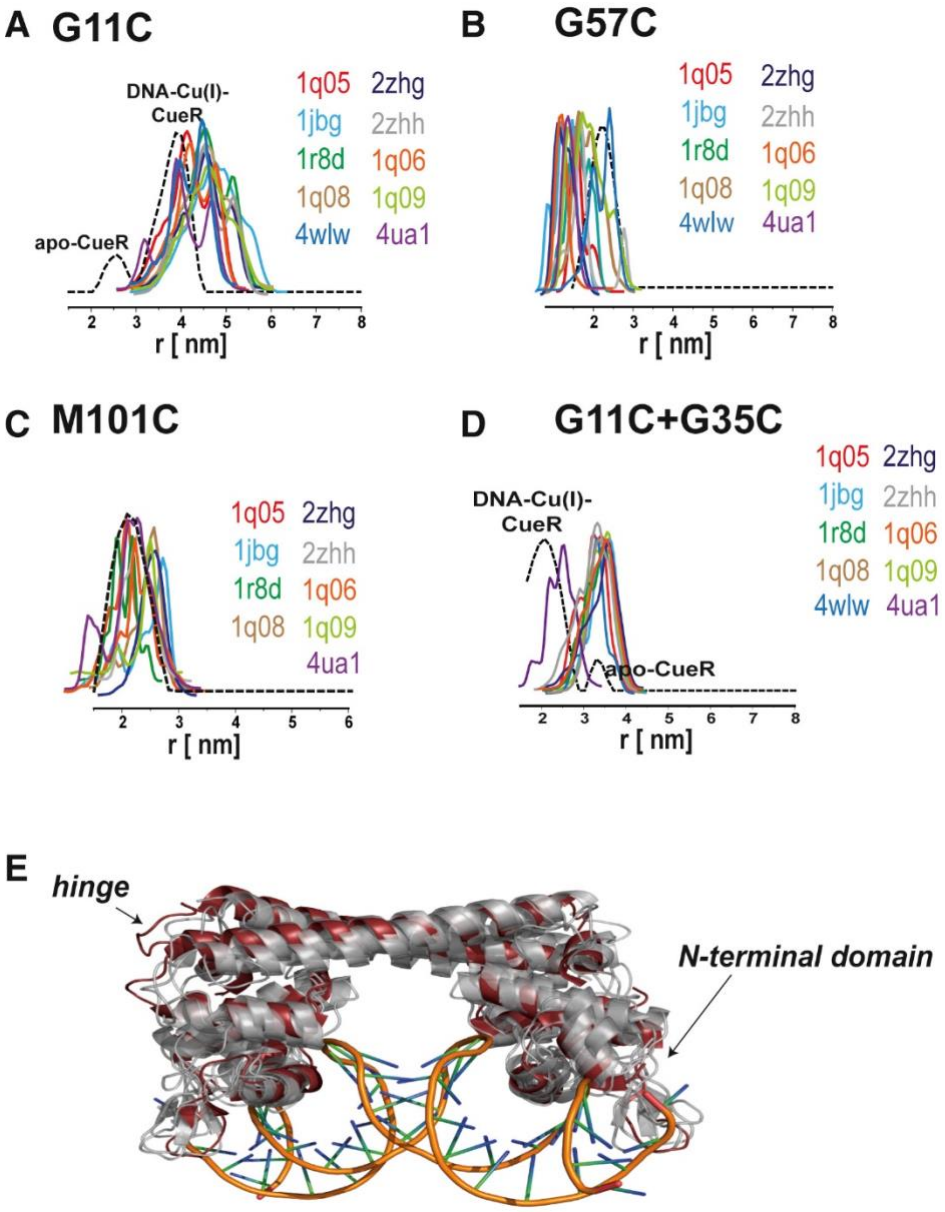


Figure 13. Using ConTemplate to predict the conformational changes and dynamics of the CueR (taken from [56]). The colored lines show simulated distance distribution functions calculated using the MMM software of models of CueR obtained using ConTemplate's templates. The black dashed line represents the DEER experimental data for CueR-DNA-Cu(I) mutants, with the spin labels: A. CueR_G11C, B. CueR_G57C, C. CueR_M101C, and D. CueR_G11C_G35C. E. PDB structure 1q05 overlaid on models based on 1r8d, 1q09, 2zhg and 4ua1 (gray). The black arrows mark the most dynamic regions, with the largest conformational changes.

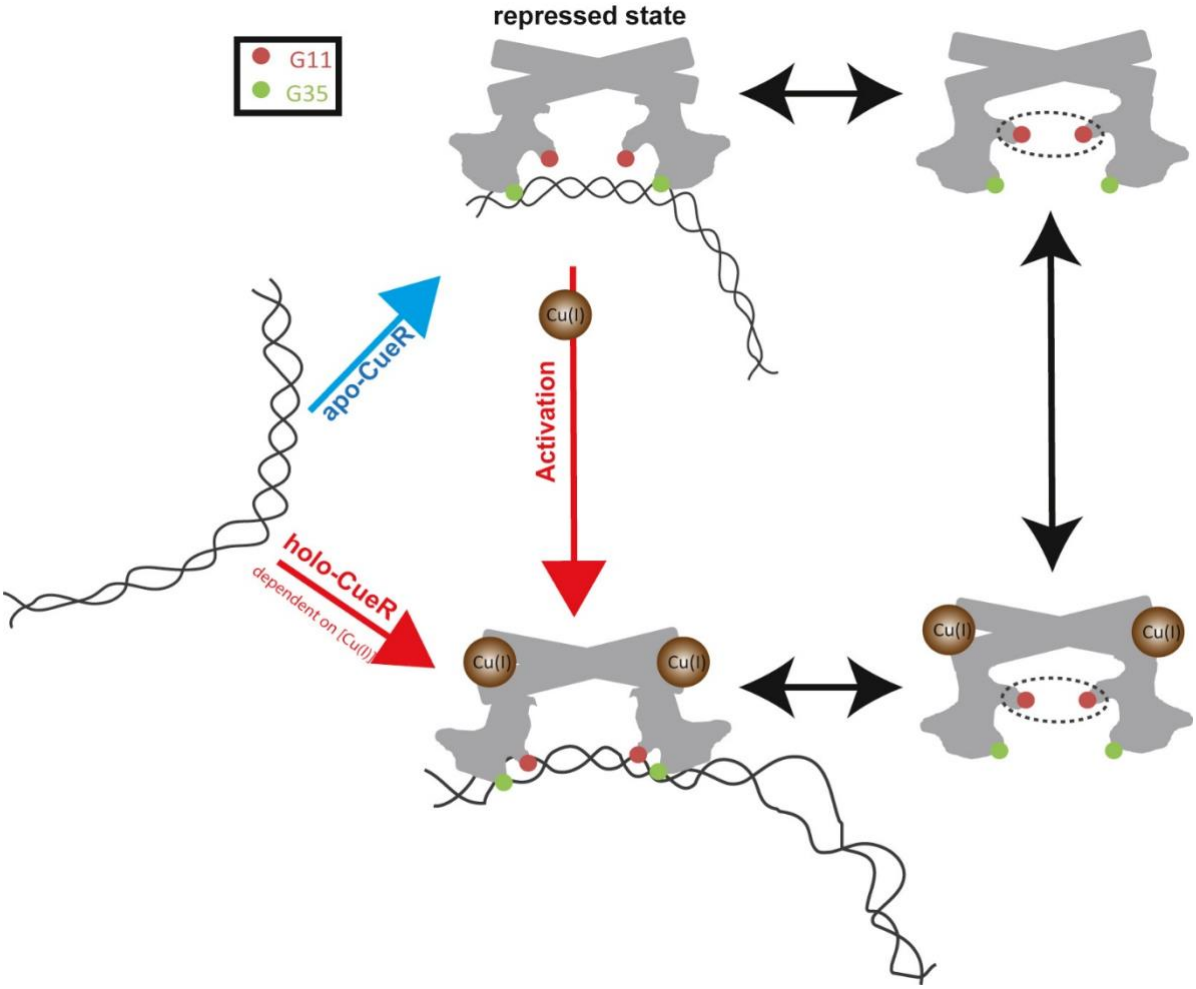


Figure 14. A suggested mechanism for CueR activation (taken from [56]). The red dots represent the locations of G11, the green dots represent G35. Upon activation, G11-G35 intra-monomer get closer, while the inter-monomer distance between the two G11 residues grows larger.

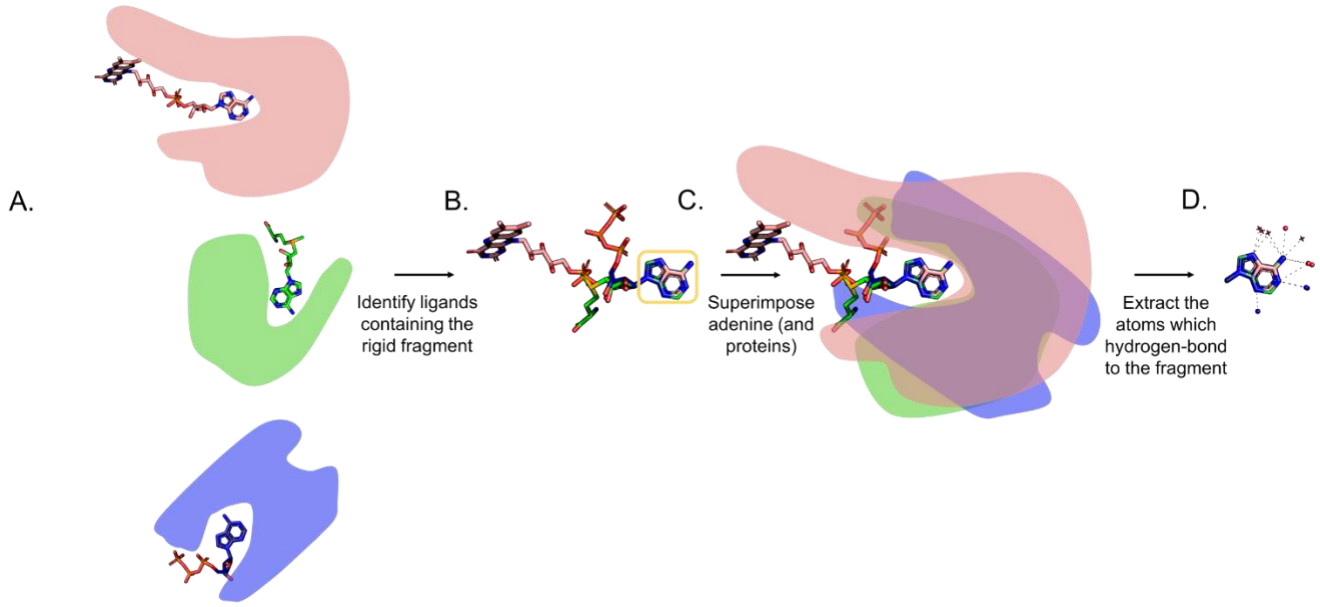


Figure 15. The ComBind methodology for detecting protein-fragment interactions among multiple and highly dissimilar proteins. The cofactors are represented as sticks, with the carbons of each cofactor colored differently than those of the others cofactors, where nitrogen, oxygen and phosphorus atoms are colored by type. The rigid fragment (in this example, adenine) is enclosed by the yellow rectangle. The proteins are superimposed based on the alignments between the rigid fragments. Protein nitrogen and oxygen atoms are shown as blue and red spheres, respectively, and water oxygens as red 3D '+' signs.

A.

B.

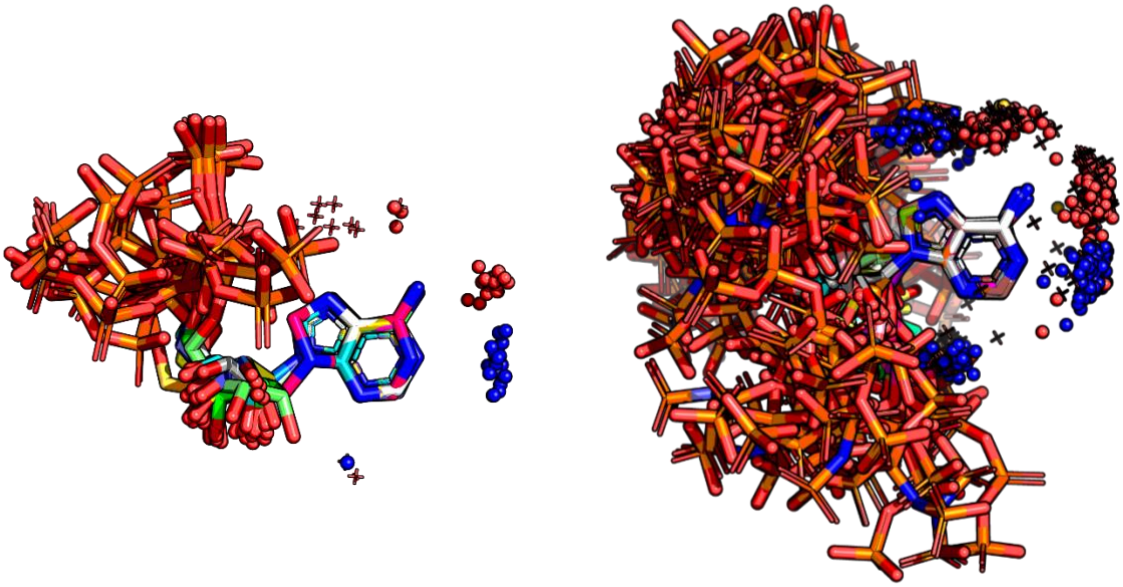


Figure 16. ComBind's results are consistent with the results found by Dennessiouk and Co-workers when using the same dataset. A. ComBind's results for the original dataset of ATP-binding sites recapitulate Dennessiouk and Co-workers observation that binding is mediated only via the Watson-Crick edge. B. However, ComBind's results for large ATP-binding sites dataset shows that the original dataset was not diverse enough to detect interactions with all adenine's edges.

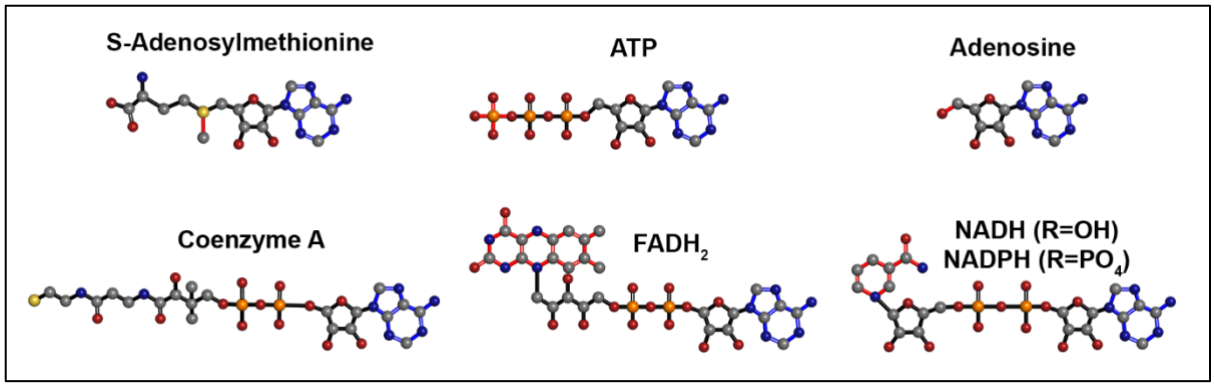


Figure 17. Adenine-containing nucleotide cofactors included in this work. The cofactors are represented as balls and sticks, with carbon atoms marked as grey spheres, nitrogen in blue, oxygen in red, phosphorus in orange and sulfur in yellow. The adenine fragment is marked using blue bonds to separate it from the functional groups, marked using red bonds.

A. B.

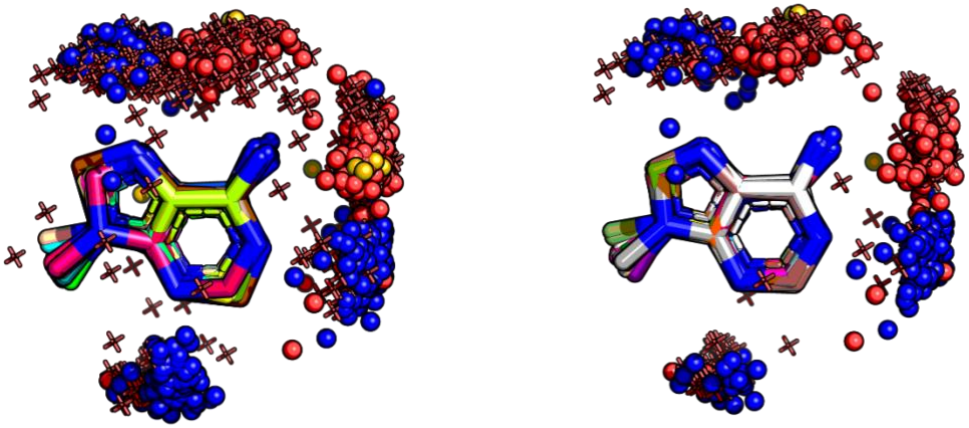


Figure 18. All the hydrogen donors and acceptors of adenine are used for protein binding. **A**. The hydrogen bonding interactions with adenine for a dataset of 985 ATP, NAD, FAD, SAM and CoA-binding proteins. Adenine carbon atoms are in various colors, as each of them originates from a different crystal structure. Protein oxygen atoms are shown as red spheres, nitrogen atoms are shown as blue spheres, sulfur atoms are shown as yellow spheres, and water oxygen are shown as red 3D '+' signs. **B**. The same as **A** focusing on the 406 ATP-binding proteins.

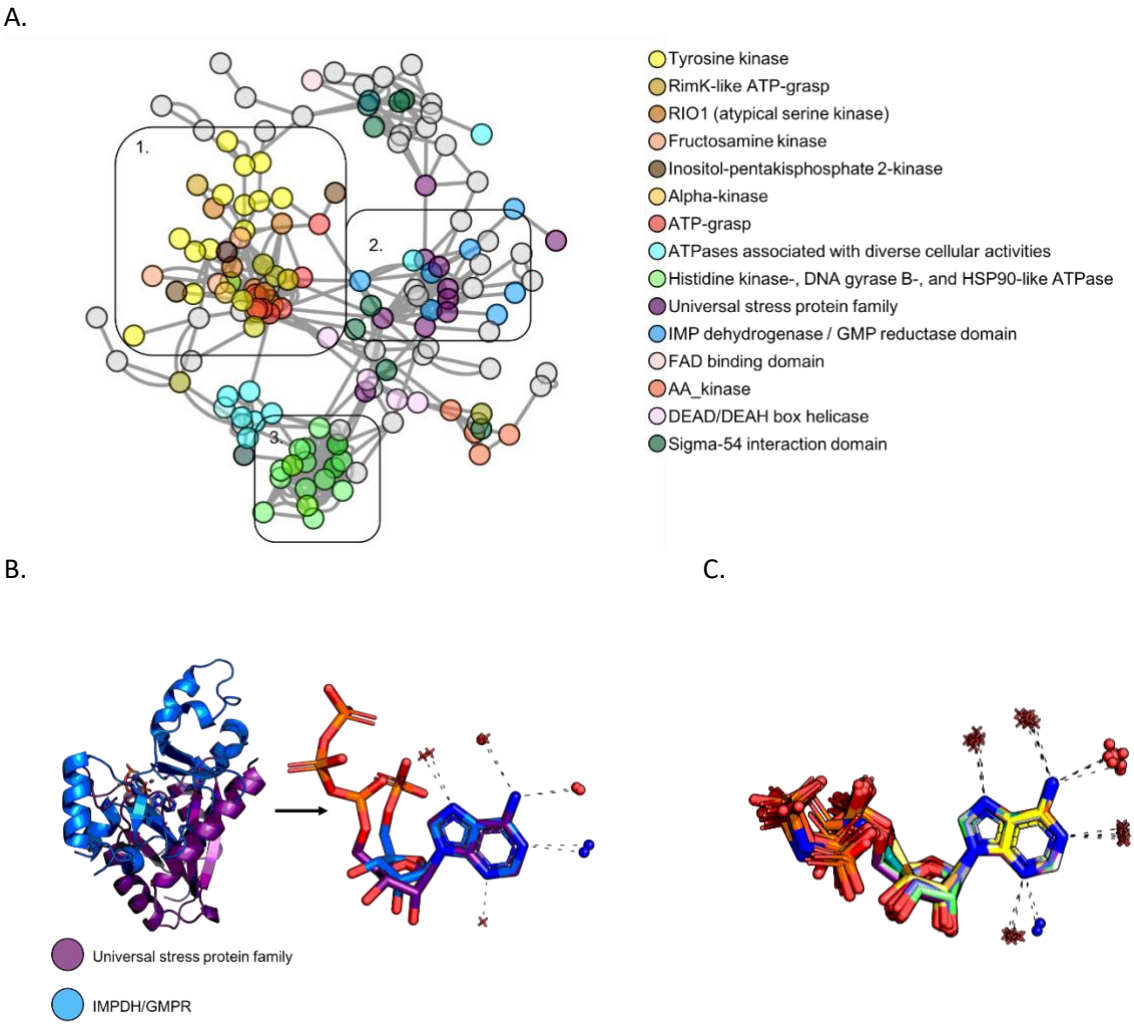


Figure 19. Adenine binding patterns show complex relationships between protein sequence, structure and function. **A**. Proteins of similar families and function tend to form similar adenine binding patterns. A network representation of protein-ATP complexes (colored circles) connected based on the geometry of their interaction regions. The nodes are colored according to the PFAM family assignment of the binding protein, only families represented by more than 3 nodes are colored, the rest are in gray. Similar colors indicate related families (i.e., various kinases, as shown in rectangle 1). Proteins with very different sequence and structure can nevertheless have similar adenine binding pattern, as shown in the intersecting families in rectangle 2, showing the cluster formed by the two PFAM families “Universal stress protein family” (in purple) and “IMPDH/GMPR” (in light blue). **B**. The left of the panel shows two proteins, one from each of these families (PDB 3fdx in purple and PDB 3lfr in blue), with different global structures, aligned according to the adenine fragment. The right of the

panel shows adenine and the atoms that hydrogen-bond to it. Protein nitrogen and oxygen atoms are shown as blue and red spheres, respectively, and water oxygen as red '+' s. The dashed lines represent the hydrogen bonds. **C**. Water molecules can be conserved in adenine binding, as shown from rectangle 3 in A, showing a cluster formed by PFAM family "Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase". The panel shows the 16 adenine interaction sites included in the cluster.

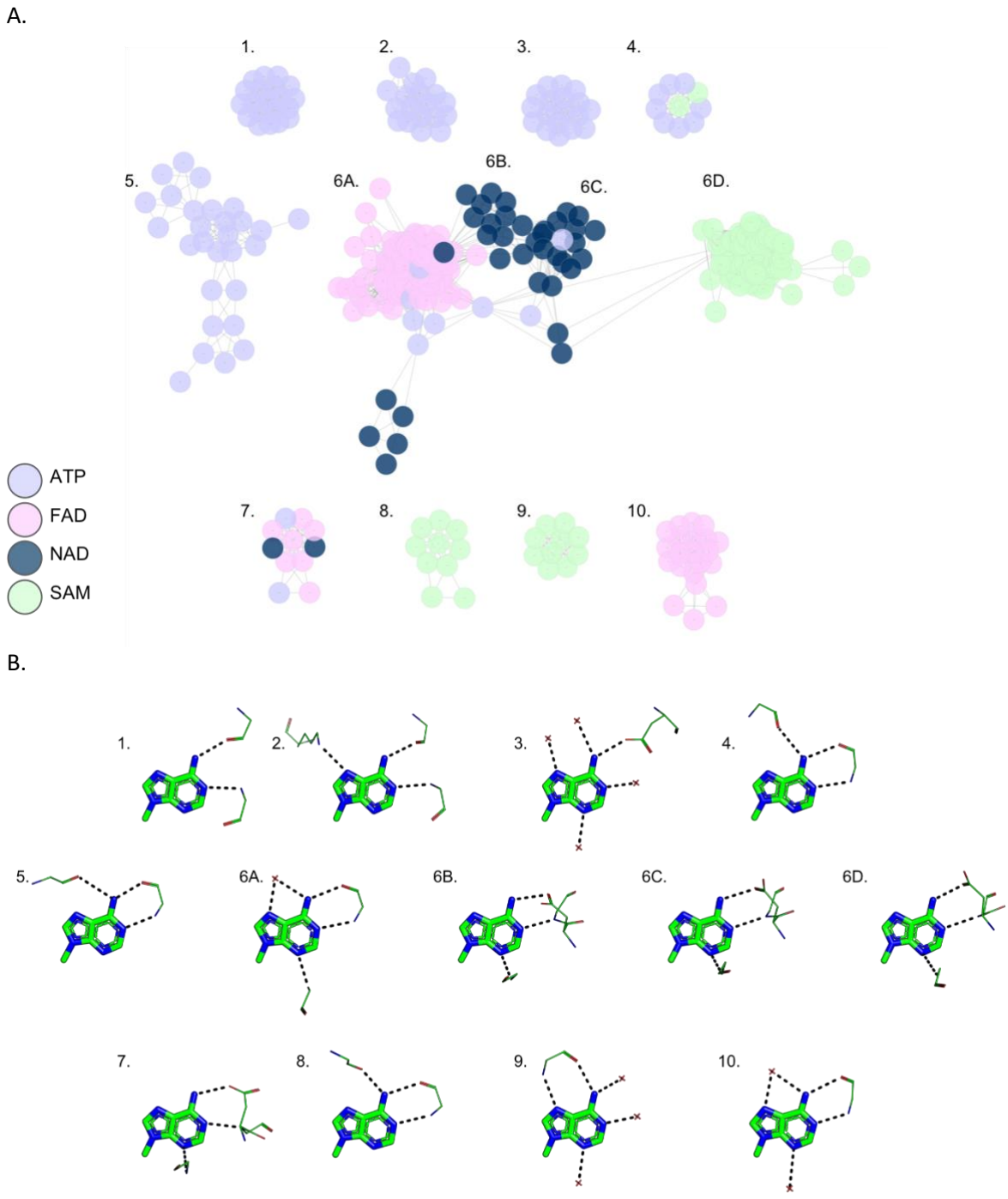


Figure 20. A network representation of themes in nucleotide cofactors-binding proteins shows that themes represent distinctive binding modes. **A**. Proteins use different themes to bind different ligands. Each node in the network represents a protein-ligand binding site. The nodes are colored according to the bound ligand, following the color-legend at the bottom-left of the panel. Connected nodes have at least one theme in common in the binding site. **B**. A representation of the binding

modes related to the clusters. The numbering of the binding modes corresponds to the cluster number in **A**. Only the adenine fragment is shown, protein atoms are shown in lines representation, side chains are shown only when mediating binding. The hydrogen bonds with adenine are shown as yellow dashed lines. The interacting atoms around adenine show the consensus binding mode of the cluster.

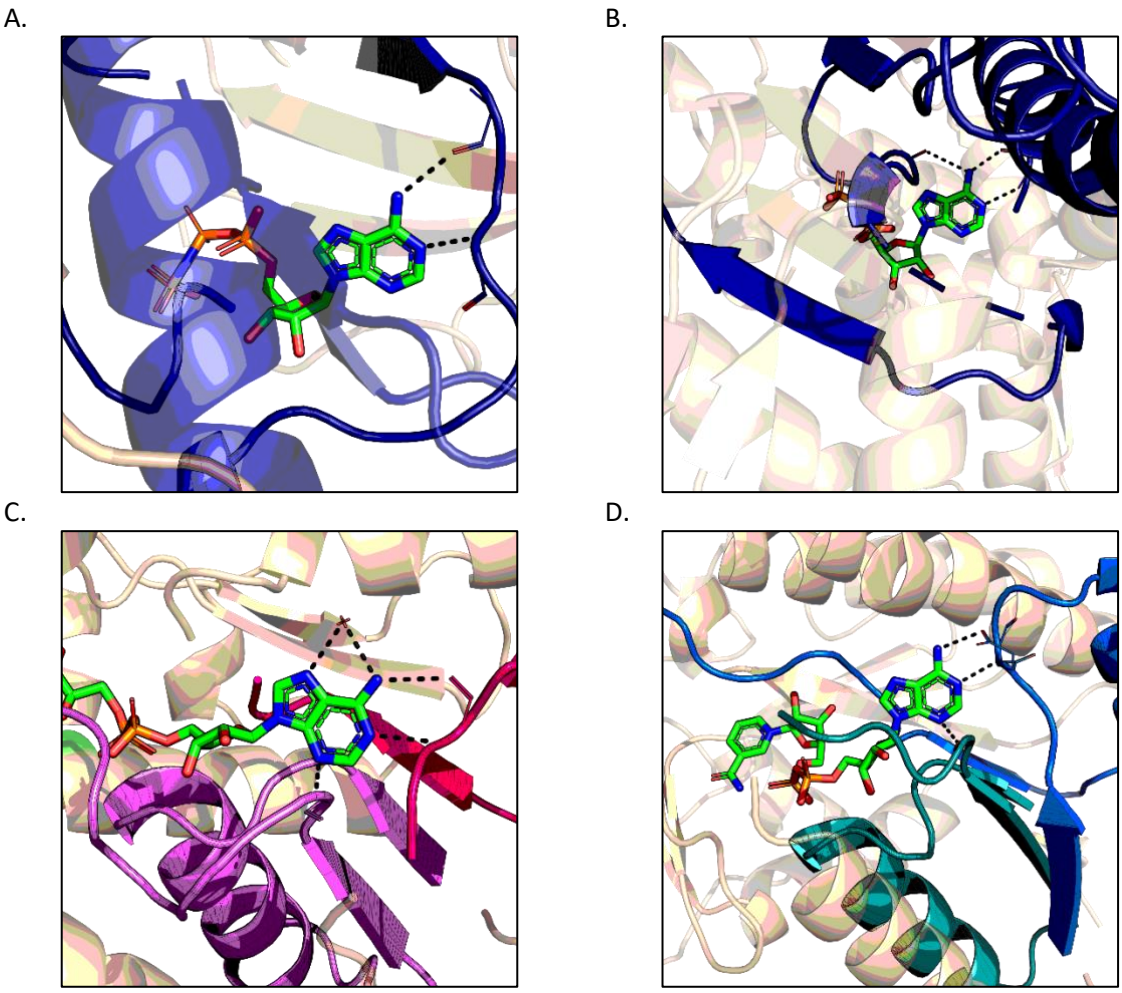
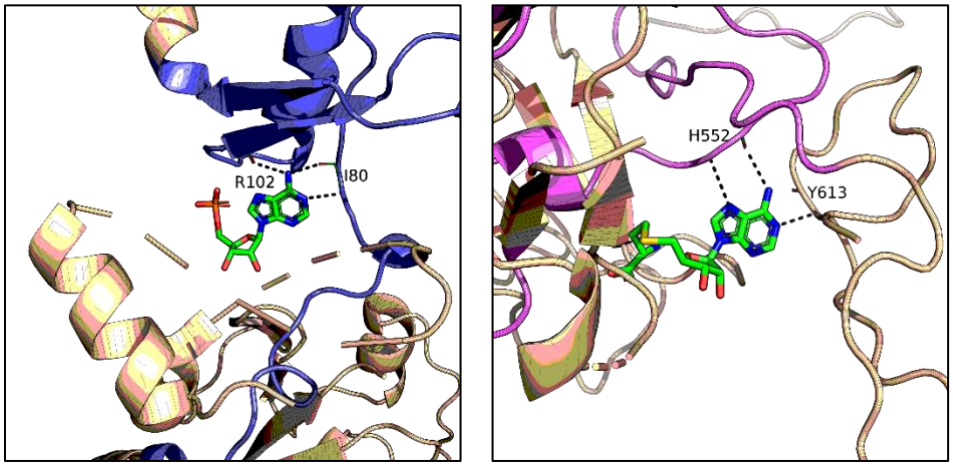


Figure 21. Themes can form the scaffolds for adenine-binding in proteins. The protein is shown in wheat with the themes highlighted in colors. The adenine-containing ligand is shown using bond-stick model, and the hydrogen bonds to specific amino acids of the themes and water molecules in black dashed lines. **A**. A theme representing the direct motif in ATP binding, as found in Figure 20A, cluster 1 (demonstrated using PDB 5ckw). **B**. A theme representing the reverse motif in ATP binding, with an additional interaction between the protein and adenine's N6 in the Hoogsteen edge, as found in Figure 20A, cluster 4 (demonstrated using PDB 1g41). **C**. A combination of two themes (purple and pink) composing adenine's binding site in FAD, as found in Figure 20A, cluster 6A (demonstrated using PDB 2gag). **D**. A combination of two themes (blue and cyan) composing adenine's binding site in NAD, as found in Figure 20A, cluster 6C (demonstrated using PDB 5u4q).

A. B.



C.

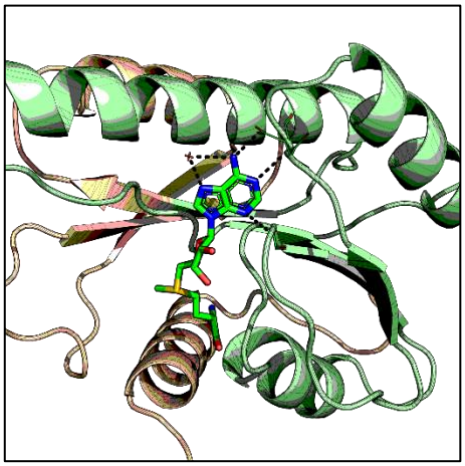


Figure 22. Themes can form adenine-binding patterns in proteins. The protein is shown in wheat with the themes highlighted in colors. The adenine-containing ligand is shown using bond-stick model, and the hydrogen bonds to specific amino acids of the themes and water molecules in black dashed lines. **A.** A theme representing the reverse motif with an additional interaction between adenine's N6 in the Hoogsteen edge and a carboxyl group at 'position XV/XVI' (here R102). Demonstrated using PDB 4hg0 **B.** A theme representing a variation of the reverse motif in the Hoogsteen edge, as found in PDB 4qeo. **C.** A relatively long theme creating a scaffold for adenine binding, as found in PDB 1ej0. The 'Asp' motif is used here, with an additional interaction between adenine's N3. A water molecule forming hydrogen bonds with adenine's Hoogsteen edge is also conserved.

A. B.

Figure 23. Different geometrical-similarity thresholds may lead to different network representations of protein-ATP complexes. A network representation of protein-ATP complexes (colored circles) connected based on the geometry of their interaction regions. The nodes are colored according to the PFAM family assignment of the binding protein, only families represented by more than 3 nodes are colored, the rest are in gray. The color scheme is the same as used in Figure 5A. **A.** Using lax thresholds leads to a larger network, with less noticeable clusters. Two nodes are connected by an edge if the RMSD between the binding sites is under 0.4Å and at least 60% of the interacting atoms are located in close proximity. **B.** Using strict similarity thresholds breaks the network to numerous connected components. Two nodes are connected by an edge if the RMSD between the binding sites is under 0.2Å and at least 70% of the interacting atoms are located in close proximity.

RMSD Threshold between Conformations (Å)	Maximal RMSD (Å)	MINIMAL Q-Score	Minimal Coverage (%)	No. of Proteins	No. of Protein-Pairs Sharing One Conformation	No. of Protein-Pairs Sharing Additional Conformations
2	1.5	0.5	80	1025	364	183
3	1.75	0.45	75	425	138	67
4	2.0	0.4	70	246	79	45
5	2.25	0.35	65	146	20	15
6	2.5	0.3	60	102	14	10

Table 1. Proteins sharing one conformation will often have more conformations in common (taken from [34]). The sequence identity between each proteins-pair is no more than 80%.

Cofactor	The 'direct' motif occurrence	The 'reverse' motif occurrence	The 'Asp' motif occurrence	Other	Total
ATP	77 (19%)	74 (18%)	10 (2.5%)	245 (60%)	406
FAD	0	128 (63%)	6 (3%)	69 (34%)	203
SAM	1 (0.5%)	29 (17%)	107 (61.5%)	37 (21%)	174
NAD	0	3 (2.5%)	42 (34%)	78 (63.5%)	123
CoA	2 (2.5%)	0	0	77 (97.5%)	79
	80	234	165	506	985

Table 2. Counts of the 'direct', 'reverse', and 'Asp' variations of the adenine-binding motif, found in the dataset. The total number (and percentage) of complexes of each type with the motifs are listed.

PDB ID	Chain ID	Ligand	R-free	Resolution	N1 H-bonds	N3 H-bonds	N6_WatsonCrick H-bonds	N6_Hoogsteen H-bonds	N7 H-bonds	ECOD X-group	ECOD F-group
12AS	B	AMP	0.287	2.2	S111/N		S111/O	E103/OE1		NO_X_NAME	AsnA
1A0I	A	ATP	0.341	2.6	K222/NZ	Y35/N	E32/OE1	I33/O	HOH/O	NO_X_NAME	DNA_ligase_A_M
1A9X	G	ADP		1.8	L6210/N		E6208/O	HOH/O		NO_X_NAME	ATP-grasp
1B37	B	FAD		1.9	V237/N	A36/N	V237/O			Rossmann-like	Amino_oxidase_1st
1B63	A	ANP	0.261	1.9	HOH/O	HOH/O	D58/OD2	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c_3
1BOO	A	SAH	0.283	2.8	S35/N /S35/OG	M295/N	D34/OD1			Rossmann-like	N6_N4_Mtase
1BXD	A	ANP								none	none
1C0P	A	FAD	0.15	1.2	V1162/N	R1035/N	V1162/O	HOH/O	HOH/O	Rossmann-like	DAO_1st
1CJA	B	AMP	0.272	2.9	V165/N		E163/O			NO_X_NAME	Act-Frag_cataly
1CQX	A	FAD	0.215	1.75		Q231/NE2		HOH/O	HOH/O	cradle loop barrel	FAD_binding_6
1CT9	B	AMP	0.297	2	V272/N		V272/O			HUP domain-like	Asn_synthase
1D2N	A	ANP	0.223	1.75	HOH/O			I516/O	I516/N	P-loop domains-like	TIP49_1st
1D4O	A	NAP	0.223	1.21	A167/N		D166/OD1		HOH/O	Rossmann-like	PNTB_C
1DDG	B	FAD	0.289	2.01	HOH/O	HOH/O			HOH/O	none	none
1E6U	A	NAP	0.17	1.45	L41/N	HOH/O		HOH/O	HOH/O	Rossmann-like	GDP_Man_Dehyd_1
1EJ0	A	SAM	0.232	1.5	F100/N	L84/N	D99/OD1	HOH/O	HOH/O	Rossmann-like	FtsJ
1EJ2	A	NAD	0.241	1.9	F125/N		F125/O	Y130/O	HOH/O	HUP domain-like	CTP_transf_like
1EP3	B	FAD	0.237	2.1	HOH/O	HOH/O	HOH/O			none	none
1EQ2	G	NAP	0.262	2				N92/O	N92/ND2	Rossmann-like	GDP_Man_Dehyd_1
1F0X	A	FAD	0.248	1.9	V262/N		V262/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_4
1F2U	C	ATP	0.282	1.6	V64/N	HOH/O	T62/O		HOH/O	P-loop domains-like	SMC_N_1
1F7L	A	COA	0.201	1.5		HOH/O		N84/O	HOH/O	Bacillus chorismate mutase-like	ACPS_1
1F9V	A	ADP	0.242	1.3		HOH/O	T694/O	N386/OD1	HOH/O	P-loop domains-like	Kinesin
1FNN	B	ADP	0.257	2		R204/NH2	P19/O	Y192/OH		Histone-like	TIP49_3rd_4
1FP2	A	SAH	0.233	1.4	M240/N		D239/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_1
1G41	A	ADP	0.276	2.3	I18/N		I18/O	V61/O		P-loop domains-like	Sigma54_activat
1G55	A	SAH	0.25	1.8	I57/N	V35/N	E58/OE1	HOH/O		Rossmann-like	DNA_methylase
1G60	A	SAM	0.221	1.74	C12/N	M242/N		HOH/O	HOH/O	Rossmann-like	N6_N4_Mtase
1GHE	B	ACO	0.23	1.55						none	none
1GM5	A	ADP	0.328	3.24						none	none
1GSA	A	ADP		2	L201/N		N199/O		K160/NZ	NO_X_NAME	RimK
1GTE	A	FAD	0.197	1.65	L261/N	K219/N	L261/O	HOH/O	HOH/O	Rossmann-like	Pyr_redox_2
1H5Q	D	NAP	0.209	1.5	V70/N		D69/OD1		HOH/O	Rossmann-like	adh_short_C2
1H72	C	ANP	0.207	1.8	K87/NZ		S101/OG	N62/OD1	V63/N	NO_X_NAME	GHMP_kinases_N
1HDO	A	NAP	0.158	1.15	V55/N		D54/OD1	HOH/O	HOH/O	Rossmann-like	NAD_binding_10
1HM9	B	ACO	0.219	1.75	HOH/O				HOH/O	none	none
1HP1	A	ATP	0.199	1.7	N431/ND2				HOH/O	NO_X_NAME	5_nucleotid_C
1HSK	A	FAD	0.223	2.3	V199/N	HOH/O	V199/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_4

1HYH	C	NAD	0.25	2.2	HOH/O	A54/N		HOH/O	HOH/O	Rossmann-like	Ldh_1_N
1I12	A	ACO	0.221	1.3	HOH/O			HOH/O	HOH/O	none	none
1I24	A	NAD	0.198	1.2	I76/N	N33/N	D75/OD2	N119/OD1	HOH/O/N119/ND2	Rossmann-like	GDP_Man_Dehyd
1I36	A	NAP	0.23	2	HOH/O					none	none
1I58	B	ADP	0.262	1.6	HOH/O		D449/OD2	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c
1IA9	A	ANP	0.294	2	M1721/N	HOH/O	E1719/O	E1718/OE2	K1646/NZ	NO_X_NAME	Alpha_kinase
1IN4	A	ADP	0.253	1.6	HOH/O	HOH/O	I28/O	Y180/OH		Histone-like	RuvB_N
1IQP	A	ADP	0.277	2.8						none	none
1J5P	A	NAD	0.26	1.9			E63/OE2			Rossmann-like	DapB_N
1JAY	A	NAP	0.206	1.65	HOH/O	HOH/O	HOH/O	HOH/O		none	none
1JG1	A	SAH	0.2	1.2	G149/N	R122/N	D148/OD1	I222/O	I222/N	Rossmann-like	Methyltransf_11_2
1JNR	A	FAD	0.202	1.6	I176/N	K57/N	I176/O	S242/OG	HOH/O	Rossmann-like	FAD_binding_3_1st
1JP4	A	AMP	0.18	1.69				HOH/O		none	none
1JR8	A	FAD	0.233	1.5		Y107/OH	C83/O	N87/OD1	N90/ND2	Four-helical up-and-down bundle	Evr1_Alr
1JU2	A	FAD	0.186	1.47	V217/N	R56/N	V217/O	HOH/O	HOH/O	Rossmann-like	GMC_oxred_N
1K0I	A	FAD	0.24	1.8	HOH/O	R33/N	HOH/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
1KAE	B	NAD	0.241	1.7		HOH/O				none	none
1KHT	C	AMP	0.253	2.5			G104/O	T91/OG1		P-loop domains-like	F_UNCLASSIFIED
1KJQ	B	ADP	0.214	1.05	V198/N	HOH/O	G196/O	E195/OE1	K155/NZ	NO_X_NAME	ATP-grasp
1KOL	B	NAD	0.206	1.65	HOH/O		H269/NE2	R267/O	R267/N	Rossmann-like	ADH_zinc_N
1KRH	A	FAD	0.249	1.5	HOH/O		F335/O		HOH/O	Other Rossmann-like structures with the crossover	NAD_binding_1
1L8Q	A	ADP	0.255	2.7	I89/N		I89/O			P-loop domains-like	Bac_DnaA_N
1LJ8	A	NAD	0.197	1.7	HOH/O	HOH/O	HOH/O	HOH/O	HOH/O	none	none
1LQT	B	FAD	0.153	1.05	V84/N	M41/N	V84/O	HOH/O/HOH/O	HOH/O/HOH/O	Rossmann-like	NAD_binding_8
1LTQ	A	ADP	0.261	2.33						none	none
1LW7	A	NAD	0.298	2.9			F199/O	V201/O		HUP domain-like	CTP_transf_like
1M15	A	ADP	0.14	1.2			S122/O	S282/OG		Glutamine synthetase-like	ATP-gua_Ptrans
1M4I	A	COA	0.205	1.5	HOH/O	HOH/O	HOH/O			none	none
1MIW	A	ATP	0.263	3	R157/NH2		D154/OD2		D112/N	PDEase-like	PolyA_pol_RNAbd
1MJH	A	ATP	0.254	1.7	V41/N		V41/O	HOH/O		HUP domain-like	Usp
1MO9	A	FAD	0.213	1.65	A158/N	R74/N	HOH/O	HOH/O	HOH/O	Rossmann-like	Pyr_redox_2
1N08	A	ADP	0.252	1.6	L99/N		L99/O	HOH/O	HOH/O	cradle loop barrel	Flavokinase
1N2X	B	SAM	0.226	1.9	Y82/N	V56/N	S81/OG	HOH/O		Rossmann-like	Methyltransf_5_1st
1N62	C	FAD	0.172	1.09	L167/N	HOH/O	L167/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_5
1N71	A	COA	0.245	1.8		HOH/O	E141/O		HOH/O	Nat/Ivy	Acetyltransf_10
1NN5	A	ANP	0.215	1.5	HOH/O	HOH/O	K182/O	HOH/O	HOH/O	P-loop domains-like	Thymidylate_kin
1NP7	B	FAD	0.233	1.9	N395/ND2			HOH/O	HOH/O	NO_X_NAME	FAD_binding_7
1NV8	A	SAM	0.253	2.2	F180/N			HOH/O		Rossmann-like	Methyltransf_11_2
1NYR	A	ATP	0.313	2.8	V378/N		V378/O	E367/OE2		NO_X_NAME	tRNA-synt_2b

1O94	A	ADP	0.212	2	M470/N		M470/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
1O97	C	AMP	0.229	1.6	V64/N		V64/O	HOH/O	HOH/O	HUP domain-like	ETF
1OAA	D	FAD	0.222	1.25	I307/N	T289/N/T289/OG1	D306/OD2	HOH/O		Rossmann-like	ETF_alpha
1OAA	A	NAP	0.222	1.25	L71/N	HOH/O	D70/OD1	HOH/O	HOH/O	Rossmann-like	adh_short_C2_1
1ORR	A	NAD	0.229	1.5	I59/N	N33/N	D58/OD1	HOH/O	HOH/O	Rossmann-like	GDP_Man_Dehyd
1P0H	A	ACO	0.237	1.6	HOH/O		HOH/O	HOH/O	HOH/O	none	none
1P3D	B	ANP	0.194	1.7	HOH/O		HOH/O	N295/OD1	N295/ND2	P-loop domains-like	Mur_ligase_M
1P5Z	B	ADP	0.197	1.6	F242/N	HOH/O	E240/O	HOH/O	HOH/O	P-loop domains-like	dNK
1P91	A	SAM	0.296	2.8	E96/N		C94/O	HOH/O		Rossmann-like	Methyltransf_11_1
1PJ5	A	FAD	0.198	1.61	V174/N	Q36/N	V174/O	HOH/O	HOH/O	Rossmann-like	DAO_1st
1PL8	D	NAD	0.222	1.9				HOH/O	HOH/O	none	none
1PN0	B	FAD	0.18	1.7		K43/N	HOH/O		HOH/O	Rossmann-like	FAD_binding_3_3rd
1PS9	A	FAD	0.243	2.2	V448/N	A404/N	V448/O			Rossmann-like	Pyr_redox_2
1PS9	A	NAP	0.243	2.2	HOH/O	HOH/O			HOH/O	none	none
1PVG	A	ANP	0.241	1.8	HOH/O		N99/OD1	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c_1
1QF9	A	ADP	0.244	1.7	HOH/O	HOH/O	R176/O	HOH/O	HOH/O	P-loop domains-like	AAA_33
1QHH	A	ATP	0.285	2.5		Q16/NE2				P-loop domains-like	UvrD-helicase
1QHX	A	ATP	0.238	2.5			K163/O			P-loop domains-like	CPT
1QR0	A	COA	0.278	1.9	HOH/O	N87/ND2	HOH/O	Y73/O		Bacillus chorismate mutase-like	KOG0945
1QSM	A	ACO	0.244	2.4						none	none
1R2J	A	FAD	0.263	2.1						none	none
1R30	A	SAM	0.3	3.4	V225/N		V225/O			TIM beta/alpha-barrel	Radical_SAM
1R6B	X	ADP	0.278	2.25	I189/N		I189/O	HOH/O	HOH/O	P-loop domains-like	Sigma54_activat
1R6D	A	NAD	0.216	1.35	I64/N	S38/N	D63/OD1		HOH/O	Rossmann-like	GDP_Man_Dehyd
1RJD	A	SAM	0.215	1.8	L176/N		D175/OD1	HOH/O	HOH/O	Rossmann-like	LCM
1RKX	D	NAD	0.266	1.8	I66/N		D65/OD1			Rossmann-like	GDP_Man_Dehyd
1RLZ	A	NAD	0.245	2.15	A343/N	T308/N	D342/OD1			Rossmann-like	DS
1RM6	E	FAD	0.173	1.6	L207/N	HOH/O	L207/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_5
1RSG	B	FAD	0.245	1.9	V223/N	A40/N	V223/O	G270/O		Rossmann-like	Amino_oxidase_1st
1RYI	A	FAD	0.214	1.8	V174/N	S35/N/S35/OG	V174/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
1S68	A	AMP	0.227	1.9	HOH/O			E34/O	I36/N	NO_X_NAME	RNA_ligase
1SAZ	A	ACP	0.267	2.5						none	none
1SBY	B	NAD	0.172	1.1	V64/N	HOH/O				Rossmann-like	adh_short_C2_1
1SEZ	B	FAD	0.293	2.9	V264/N	A44/N	V264/O			Rossmann-like	Amino_oxidase_1st
1SG6	A	NAD	0.25	1.7		HOH/O	F179/O	T139/O		Flavodoxin-like	DHQ_synthase_N
1SNY	A	NAP	0.284	1.75	L62/N	HOH/O	D61/OD2	HOH/O	HOH/O	Rossmann-like	adh_short_C2_1
1SVM	B	ATP	0.241	1.94	K550/N		R548/O	S430/O		P-loop domains-like	Polyoma_lg_T_C
1SXJ	E	ADP	0.306	2.85	S17/OG					P-loop domains-like	Sigma54_activat
1TF7	C	ATP	0.28	2.8						none	none

1TH8	A	ADP	0.253	2.4	HOH/O	HOH/O	D81/OD2	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c_2
1TIQ	B	COA	0.241	1.9			N137/OD1			Nat/Ivy	Acetyltransf_10_6
1TLL	A	FAD	0.272	2.3						none	none
1TV8	B	SAM	0.241	2.2	M197/N		M197/O	HOH/O	HOH/O	TIM beta/alpha-barrel	Radical_SAM
1U0J	A	ADP	0.237	2.1		HOH/O		HOH/O	HOH/O	none	none
1U2Z	C	SAH	0.246	2.2	F460/N		S459/OG	HOH/O	HOH/O	Rossmann-like	DOT1
1U8V	A	FAD	0.212	1.6			HOH/O			none	none
1U8X	X	NAD	0.221	2.05	HOH/O		HOH/O			none	none
1UA4	A	AMP	0.205	1.9	V429/N	HOH/O	V429/O	E195/OE2		Rossmann-like	ADP_PFK_GK
1UJ2	A	ADP	0.217	1.8			D213/OD2	HOH/O	HOH/O	P-loop domains-like	PRK
1UWK	B	NAD	0.179	1.19	A244/N/S198/OG	S198/N/S198/OG	N243/OD1	L275/O	L275/N	Rossmann-like	Urocanase
1UXY	A	FAD	0.246	1.8	I173/N	HOH/O	I173/O	HOH/O	S116/OG	FAD-binding domain-like	FAD_binding_4
1V25	A	ANP	0.24	2.3	HOH/O			G323/O		Other Rossmann-like structures with the crossover	AMP-binding_2nd
1VE3	B	SAM	0.27	2.1	A94/N	I68/N	D93/OD1			Rossmann-like	Methyltransf_11
1VPT	A	SAM	0.25	1.8	V116/N		V116/O			Rossmann-like	PARP_regulatory
1W07	B	FAD	0.249	2	HOH/O					none	none
1W0H	A	AMP	0.202	1.59	HOH/O	HOH/O	HOH/O	HOH/O	HOH/O/HOH/O	none	none
1W1O	A	FAD	0.215	1.7	I235/N		I235/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_4
1W5S	A	ADP	0.286	2.4	HOH/O	R227/NH2		Y215/OH		Histone-like	KOG2227
1WWZ	B	ACO	0.236	1.75				HOH/O		none	none
1WZN	A	SAH	0.233	1.9	V97/N	L71/N	D96/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_5
1X7D	A	NAD	0.19	1.6	HOH/O		HOH/O	HOH/O	HOH/O	none	none
1XDI	B	FAD	0.276	2.81	G117/N	C36/N	G117/O			Rossmann-like	Pyr_redox_2
1XDN	A	ATP	0.148	1.2	HOH/O	V88/N		E86/O		NO_X_NAME	RNA_ligase
1XDP	B	ATP	0.274	2.5	HOH/O		HOH/O			none	none
1XG5	B	NAP	0.206	1.53	L71/N		D70/OD1	HOH/O		Rossmann-like	adh_short_C2
1XQS	C	AMP	0.296	2.9						none	none
1Y0P	A	FAD	0.184	1.5	G278/N	K157/N	G278/O			Rossmann-like	FAD_binding_3_1st
1Y56	A	ATP	0.226	2.86	A181/N	HOH/O	A181/O		E136/N	Rossmann-like	FAD_binding_3_1st
1Y56	B	FAD	0.226	2.86	V172/N	K36/N	V172/O			Rossmann-like	DAO_1st
1Y8Q	B	ATP	0.248	2.25	I96/N	N118/ND2	S95/OG	HOH/O		Rossmann-like	ThiF_1
1YF3	A	SAH	0.27	2.29	F157/N	I51/N	H156/NE2	HOH/O		Rossmann-like	MethyltransfD12
1YOA	A	FAD	0.228	1.9			HOH/O			none	none
1YQT	A	ADP	0.225	1.9	HOH/O	HOH/O	HOH/O		HOH/O	none	none
1Z0S	B	ATP	0.231	1.7						none	none
1Z2I	A	NAD	0.265	2.2		HOH/O				none	none
1Z2N	X	ADP	0.205	1.2	I171/N		H169/O		K136/NZ	NO_X_NAME	Ins134_P3_kin_C
1Z45	A	NAD	0.245	1.85	L70/N	N43/N	D69/OD1	N110/OD1	N110/ND2/HOH/O	Rossmann-like	GDP_Man_Dehyd
1Z6T	B	ADP	0.244	2.21	V127/N		V127/O		HOH/O	P-loop domains-like	NB-ARC_1st

1ZAR	A	ADP	0.218	1.75	I182/N		E180/O	M179/SD	HOH/O	NO_X_NAME	RIO1
1ZH8	B	NAP	0.237	2.5		R41/N				Rossmann-like	GFO_IDH_MocA
1ZK4	A	NAP	0.176	1	S63/N		D62/OD2/D62/OD1	HOH/O	HOH/O	Rossmann-like	adh_short_C2
1ZR6	A	FAD	0.2	1.55	V195/N	HOH/O	V195/O	G106/O		FAD-binding domain-like	FAD_binding_4
1ZTH	C	ADP	0.262	1.89	I150/N		E148/O		HOH/O	NO_X_NAME	RIO1
2A14	A	SAH	0.197	1.7	V143/N		D142/OD1	HOH/O	HOH/O	Rossmann-like	NNMT_PNMT_TEMT
2A2C	A	ADP	0.201	1.65	N83/ND2		HOH/O	HOH/O	HOH/O	NO_X_NAME	GHMP_kinases_N
2A5Y	B	ATP	0.277	2.6			Y131/O	HOH/O		P-loop domains-like	NB-ARC_1st
2AOT	B	SAH	0.256	1.9	S120/N		HOH/O		HOH/O	Rossmann-like	Methyltransf_11_3
2AZN	A	NAP	0.253	2.7				V134/O	V134/N	Other Rossmann-like structures with the crossover	RibD_C
2B3T	A	SAH	0.303	3.1	W168/N	R141/N	D167/OD2			Rossmann-like	Methyltransf_11_2
2B69	A	NAD	0.162	1.21	V61/N	N36/N	D60/OD1	T94/OG1	HOH/O	Rossmann-like	GDP_Man_Dehyd_1
2B9E	A	SAM	0.222	1.65	F286/N		D285/OD1	HOH/O	HOH/O	Rossmann-like	Methyltr_RsmB-F_1
2B9W	A	FAD	0.19	1.95	HOH/O	R38/N	HOH/O		HOH/O	none	none
2BB3	A	SAH	0.274	2.27	L165/N		L165/O		HOH/O	Tetrapyrrole methylase C-terminal domain-like	F_UNCLASSIFIED
2BD0	C	NAP	0.213	1.7	I67/N		D66/OD1	T116/OG1		Rossmann-like	adh_short_C2_1
2BRY	B	FAD	0.222	1.45	F181/N	K115/N	F181/O	HOH/O	HOH/O	none	none
2BVF	B	FAD	0.208	1.92	V195/N		V195/O	G107/O		FAD-binding domain-like	FAD_binding_4
2C0C	A	NAP	0.216	1.45	HOH/O	HOH/O	HOH/O	HOH/O		none	none
2C2A	A	ADP	0.275	1.9	HOH/O	HOH/O	D411/OD1	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c
2C31	B	ADP	0.175	1.73	I326/N	I307/N	HOH/O			Rossmann-like	TPP_enzyme_M
2C43	A	COA	0.219	1.93		N118/ND2	G198/O	K99/O	HOH/O	Bacillus chorismate mutase-like	KOG0945
2C5A	A	NAD	0.14	1.4	L79/N	W59/N	D78/OD1	HOH/O	HOH/O	Rossmann-like	GDP_Man_Dehyd
2C5S	A	AMP	0.278	2.5	F209/N		F209/O			HUP domain-like	tRNA_Me_trans_1st
2C9O	B	ADP	0.257	2.2	V40/N		V40/O	Y366/OH		P-loop domains-like	TIP49_1st
2CDC	C	NAP	0.239	1.5	HOH/O					none	none
2CDU	A	ADP	0.223	1.8	V214/N		V214/O	HOH/O	HOH/O	Rossmann-like	Pyr_redox_2_1
2CDU	A	FAD	0.223	1.8	V81/N	M33/N	V81/O	HOH/O		Rossmann-like	Pyr_redox_2
2CE7	A	ADP	0.269	2.44	G164/N		G164/O			P-loop domains-like	Sigma54_activat
2CFM	A	AMP	0.231	1.8	K403/NZ		E247/OE1	I248/O	Y250/N	NO_X_NAME	DNA_ligase_A_M
2CHG	C	ANP	0.256	2.1	V20/N	HOH/O	V20/O	T49/O		P-loop domains-like	Sigma54_activat
2CJA	A	ATP	0.233	2.2	V348/N		V348/O	E338/OE2	HOH/O	NO_X_NAME	tRNA-synt_2b
2CUL	A	FAD	0.199	1.65	A91/N	Q34/N	A91/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
2CY2	A	ACO	0.227	2	HOH/O			K135/O/HOH/O		Nat/Ivy	Acetyltransf_10_5
2CYE	C	COA	0.242	1.9	HOH/O	HOH/O				none	none
2D0O	A	ADP	0.232	2	HOH/O	HOH/O	HOH/O		HOH/O	none	none
2D1C	B	NAP	0.238	1.8	N293/N	HOH/O	N293/O	H280/NE2		Other Rossmann-like structures with the crossover	Iso_dh
2DCL	A	AMP	0.29	2.28	V77/N		V77/O			Alpha-beta plaits	DUF190
2DPM	A	SAM	0.284	1.8	F178/N		D177/OD1			Rossmann-like	MethyltransfD12

2DVM	D	NAD	0.207	1.6					HOH/O	none	none
2E0N	B	SAH	0.266	2	V205/N		V205/O	P14/O		Tetrapyrrole methylase C-terminal domain-like	F_UNCLASSIFIED
2E1M	A	FAD	0.308	2.8	M354/N	A89/N	M354/O			Rossmann-like	Amino_oxidase_1st
2E5Y	A	ATP	0.25	1.92	D89/N	HOH/O	D89/O	HOH/O	HOH/O	Long alpha-hairpin	ATP-synt_DE
2EGV	A	SAM	0.245	1.45	L208/N		L208/O	Y211/O	L213/N	NO_X_NAME	Methyltrans_RNA_C
2EIS	B	COA	0.239	2.1				HOH/O		none	none
2F1K	C	NAP	0.222	1.55						none	none
2F69	A	SAH	0.174	1.3	HOH/O	HOH/O	E356/OE2	H297/O	H297/N	beta-clip	SET
2F6R	A	ACO	0.211	1.7			HOH/O			none	none
2F8L	A	SAM	0.211	2.2	G181/N	V155/N	D180/OD1	HOH/O		Rossmann-like	N6_Mtase
2FG9	A	FAD	0.245	2.2			S94/OG			cradle loop barrel	Pyridox_ox_2
2FIW	A	ACO	0.263	2.35			HOH/O	M2/SD	HOH/O	Nat/Ivy	Acetyltransf_10_1
2FK8	A	SAM	0.239	2	W132/N		E133/OE2	HOH/O	HOH/O	Rossmann-like	CMAS
2FNA	B	ADP	0.226	2	F15/N	K8/NZ	F15/O	HOH/O	HOH/O	P-loop domains-like	ATPase_2
2FT0	A	ACO	0.223	1.66	HOH/O		HOH/O	HOH/O		none	none
2G1U	A	AMP	0.175	1.5	A59/N	K38/N	D58/OD1	HOH/O		Rossmann-like	TrkA_N
2G8Y	B	NAD	0.186	2.15				HOH/O	HOH/O	none	none
2GAG	A	NAD	0.23	1.85	V205/N	E159/N	V205/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
2GAG	B	FAD	0.23	1.85	V197/N	K54/N	V197/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
2GB4	B	SAH	0.193	1.25	I130/N	I86/N	S129/OG		HOH/O	Rossmann-like	TPMT
2GF3	A	FAD	0.195	1.3	V173/N	A34/N	V173/O		HOH/O	Rossmann-like	DAO_1st
2GJ3	B	FAD	0.186	1.04	HOH/O	HOH/O	HOH/O		HOH/O	none	none
2GKS	A	ADP	0.24	2.31			T522/O/L525/O			P-loop domains-like	APS_kinase_1
2GM3	F	AMP	0.261	2.46	V53/N		V53/O			HUP domain-like	Usp
2GMH	B	FAD	0.254	2.5	A167/N	K72/N	A167/O			Rossmann-like	FAD_binding_3_1st
2GQT	A	FAD	0.209	1.3	V161/N	HOH/O	V161/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_4
2GRU	B	NAD	0.223	2.15			HOH/O	T128/O		Flavodoxin-like	DHQ_synthase_N
2GV8	B	FAD	0.259	2.1	V138/N	R39/N	V138/O		HOH/O	Rossmann-like	FMO-like_1st
2GXQ	A	AMP	0.168	1.2	HOH/O		T23/O		Q28/NE2	P-loop domains-like	DEAD
2H00	A	SAH	0.274	2	HOH/O		T127/O			Rossmann-like	Methyltransf_10
2H23	B	SAH	0.288	2.45	HOH/O			H243/O	H243/N	beta-clip	SET_5
2H88	A	FAD	0.206	1.74	A178/N	K49/N	A178/O	D232/OD2	HOH/O	Rossmann-like	FAD_binding_3_1st
2HA8	B	SAH	0.209	1.6	I149/N		I149/O		L158/N	NO_X_NAME	SpoU_methylase
2HMA	A	SAM	0.217	2.41	M38/N		M38/O			HUP domain-like	NAD_synthase
2HQ9	A	FAD	0.243	1.95						none	none
2HTI	A	FAD	0.262	2.5						none	none
2I0Z	A	FAD	0.248	1.84	V133/N	K33/N	V133/O	D176/OD2	HOH/O	Rossmann-like	FAD_binding_3_1st
2I79	B	ACO	0.251	2.1						none	none
2IGT	B	SAM	0.192	1.89	A191/N	A162/N	D190/OD1			Rossmann-like	Methyltransf_11_1
2IID	B	FAD	0.21	1.8	V261/N	A64/N	V261/O	HOH/O	HOH/O	Rossmann-like	Amino_oxidase_1st
2IO8	B	ADP	0.236	2.1	W571/N	HOH/O	Q569/O		K533/NZ	NO_X_NAME	GSP_synth_1st

2IPI	A	FAD	0.242	1.65	V203/N		V203/O	S67/OG	S67/OG	FAD-binding domain-like	FAD_binding_4
2IVD	B	FAD	0.283	2.3	V251/N	S40/N	V251/O			Rossmann-like	Amino_oxidase_1st
2IVN	A	ANP	0.224	1.65	N257/ND2		E176/OE1	HOH/O	HOH/O	Ribonuclease H-like	Peptidase_M22_C
2IW3	A	ADP	0.242	2.4	S82/OG		HOH/O	I41/O	N702/ND2/H43/N	Repetitive alpha hairpins	KOG0213
2IXA	A	NAD	0.22	2.3	HOH/O		N80/O			Rossmann-like	GFO_IDH_MocA
2IYV	A	ADP	0.178	1.35		HOH/O	R153/O	HOH/O	HOH/O	P-loop domains-like	SKI
2J4J	E	ACP	0.242	2.1	Y145/N		Y145/O			Other Rossmann-like structures with the crossover	AA_kinase
2J91	B	AMP	0.199	1.8				HOH/O		none	none
2J9L	E	ATP	0.295	2.3	T596/N		T596/O	G619/O		NO_X_NAME	IMPDH_2nd_1
2JAE	B	FAD	0.182	1.25	V261/N	A43/N	V261/O	HOH/O	HOH/O	Rossmann-like	Amino_oxidase_1st
2JCB	A	ADP	0.238	1.6		HOH/O	HOH/O	HOH/O		none	none
2JDI	B	ANP	0.22	1.9			Q430/O	HOH/O		NO_X_NAME	ATP-synt_ab_C
2JDI	D	ANP	0.22	1.9	HOH/O	HOH/O	HOH/O	HOH/O	HOH/O	none	none
2JHF	A	NAD	0.152	1			HOH/O/HOH/O	HOH/O/HOH/O	HOH/O	none	none
2JHP	A	SAH	0.289	2.5	G397/N		N396/OD1	N417/OD1	N417/ND2	Rossmann-like	Orbi_VP4_3rd
2JJQ	A	SAH	0.228	1.8	D326/N	S300/N	D326/OD1	HOH/O		Rossmann-like	Methyltransf_11_2
2KMX	A	ATP				H1086/ND1	E1081/OE2			NO_X_NAME	P-type_ATPase_Cu-like
2NV4	B	SAM	0.25	2.2	K59/N		D58/OD1	S116/OG	L113/N	cradle loop barrel	UPF0066
2O0J	A	ADP	0.244	1.8		HOH/O	Q138/O		Q143/NE2	P-loop domains-like	Terminase_6
2O23	B	NAD	0.154	1.2	V65/N		D64/OD1		HOH/O	Rossmann-like	adh_short_C2
2O7S	A	NAP	0.233	1.78						none	none
2O8B	B	ADP	0.281	2.75	I1109/N		I1109/O			P-loop domains-like	MutS_V
2OB0	B	ACO	0.204	1.8	HOH/O			HOH/O		none	none
2OJW	B	ADP	0.212	2.05	S257/OG		S257/OG	HOH/O	HOH/O	Glutamine synthetase-like	Gln-synt_C
2OLN	A	FAD	0.149	1.15	V176/N	R35/N	V176/O	HOH/O	R35/NH1	Rossmann-like	DAO_1st
2OLR	A	ATP	0.198	1.6		HOH/O	T455/OG1	I450/O		P-loop domains-like	PEPCK_ATP_C
2OZG	A	COA	0.215	2						none	none
2P09	A	ATP	0.195	1.65	M45/N	HOH/O	M45/O	HOH/O	G63/N	none	none
2P0A	A	ANP	0.218	1.9	I287/N	HOH/O	A285/O	E284/OE2	K248/NZ	NO_X_NAME	RimK
2P0W	A	ACO	0.228	1.9						none	none
2P35	A	SAH	0.214	1.95	L84/N	S62/N	D83/OD1	HOH/O		Rossmann-like	Methyltransf_11_3
2PAN	B	FAD	0.253	2.7	A322/N	I303/N	D321/OD2			Rossmann-like	TPP_enzyme_M
2PEZ	B	DAT	0.22	1.4				C207/O		P-loop domains-like	APS_kinase_1
2PGN	B	FAD	0.176	1.2	A322/N	T303/N	D321/OD1	HOH/O		Rossmann-like	TPP_enzyme_M
2PML	X	ANP	0.327	2.6	M120/N		E118/O			NO_X_NAME	Pkinase_Tyr
2PQ8	A	COA	0.231	1.45		HOH/O		HOH/O	HOH/O	none	none
2PR1	B	COA	0.255	3.2						none	none
2PV7	B	NAD	0.194	2	W135/NE1		HOH/O		HOH/O	Rossmann-like	F420_oxidored
2PXX	A	SAH	0.226	1.3	V114/N	Y89/N	D113/OD1		HOH/O	none	none
2PYW	B	ADP	0.197	1.9	L119/N	HOH/O	R117/O		HOH/O	NO_X_NAME	Fructosamin_kin

2PZM	B	NAD	0.221	2	V54/N		S53/OG	T91/OG1	HOH/O	Rossmann-like	GDP_Man_Dehyd
2Q14	G	ADP	0.252	2.2	HOH/O					none	none
2Q46	B	NAP	0.267	1.8	I57/N	HOH/O	D56/OD2		HOH/O	Rossmann-like	NAD_binding_10
2Q7D	B	ANP	0.236	1.6	I191/N		N189/O		K157/NZ	NO_X_NAME	Ins134_P3_kin_C
2QBU	B	SAH	0.27	2.1						none	none
2QCU	A	FAD	0.24	1.75	A172/N	A34/N		HOH/O		Rossmann-like	FAD_binding_3_1st
2QQ0	B	ADP	0.217	1.5	V139/N		V139/O	HOH/O	HOH/O	Glucocorticoid receptor-like	TK_C
2QZS	A	ADP	0.228	2.2	H356/N	HOH/O	G354/O	HOH/O	HOH/O	Other Rossmann-like structures with the crossover	Glycos_transf_1
2R0C	A	FAD	0.241	1.8	L138/N	Q37/N	L138/O		HOH/O	Rossmann-like	FAD_binding_3_3rd
2R85	B	AMP	0.186	1.7	V176/N	HOH/O	E174/O		K132/NZ	NO_X_NAME	DUF1297
2RGH	A	FAD	0.253	2.3	A197/N	M49/N	A197/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
2RIO	A	ADP	0.266	2.4	C91/N		E89/O			NO_X_NAME	Pkinase_Tyr
2UV4	A	AMP	0.207	1.33	A205/N	HOH/O	A205/O	A227/O		NO_X_NAME	IMPDH_2nd_1
2UYT	A	ADP	0.209	1.55	HOH/O			HOH/O	HOH/O	none	none
2V1U	A	ADP	0.297	3.1			P34/O	Y213/OH		P-loop domains-like	AAA
2V3A	A	FAD	0.202	2.4	V83/N	A37/N	V83/O	HOH/O	HOH/O	Rossmann-like	Pyr_redox_2
2V5Z	A	FAD	0.227	1.6	V235/N	A35/N	V235/O	HOH/O	HOH/O	Rossmann-like	Amino_oxidase_1st
2V8Q	E	AMP	0.237	2.1	L276/N		L276/O	R298/O	R298/NH1	NO_X_NAME	IMPDH_2nd_1
2VDW	E	SAH	0.265	2.7	N550/ND2		L549/O		R655/NH2	Rossmann-like	Methyltransf_11_3
2VFR	A	FAD	0.165	1.1	V169/N		V169/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_4
2VHJ	B	ADP	0.237	1.8					S292/OG	P-loop domains-like	NTPase_P4
2VOS	A	ADP	0.192	2	HOH/O		HOH/O	N303/OD1	N303/ND2	P-loop domains-like	Mur_ligase_M
2VOU	B	FAD	0.222	2.6	L120/N	R36/N	L120/O			Rossmann-like	FAD_binding_3_3rd_1
2VVM	A	FAD	0.205	1.85	V279/N	A70/N	V279/O	HOH/O		Rossmann-like	Amino_oxidase_1st
2WDS	A	COA	0.205	1.35		HOH/O		G81/O	HOH/O	Bacillus chorismate mutase-like	ACPS_1
2WK1	A	SAH	0.163	1.4	F178/N	S123/N	HOH/O	S202/OG		Rossmann-like	TylF
2WM3	A	NAP	0.197	1.85	Q59/N	HOH/O	D58/OD1	HOH/O/HOH/O	HOH/O	Rossmann-like	NmrA
2WPX	B	ACO	0.253	2.31			HOH/O			none	none
2WZB	A	ADP	0.186	1.47						none	none
2X0Q	A	ATP	0.221	1.96	HOH/O	HOH/O	HOH/O	HOH/O		none	none
2X3J	A	ATP	0.231	2	N509/ND2	HOH/O	HOH/O	HOH/O		NO_X_NAME	lucA_lucC
2X3N	A	FAD	0.22	1.75	HOH/O	Q36/N		HOH/O		Rossmann-like	FAD_binding_3_1st
2XDO	C	FAD	0.23	2.09	L139/N	R47/N	L139/O		HOH/O	Rossmann-like	FAD_binding_3_3rd_1
2XVM	B	SAH	0.23	1.48	L87/N	K60/N	D86/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_3
2XYQ	A	SAH	0.227	2	C115/N	L100/N		HOH/O		Rossmann-like	NSP13
2XZO	A	ADP	0.233	2.39			D470/O		Q475/NE2	P-loop domains-like	AAA_11
2Y27	B	ATP	0.194	1.6	HOH/O	A214/N	D235/OD1	I236/O	HOH/O	Other Rossmann-like structures with the crossover	AMP-binding_2nd
2Y5D	B	NAP	0.193	1.4	HOH/O					none	none
2YB1	A	AMP	0.207	1.9						none	none
2YQZ	A	SAM	0.235	1.8	A95/N	A69/N	D94/OD1			Rossmann-like	Methyltransf_11

2YUT	A	NAP	0.224	2.2	L51/N	HOH/O	D50/OD1	HOH/O	HOH/O	Rossmann-like	adh_short_C2_1
2YWL	B	FAD	0.214	1.6	V78/N	G32/N	V78/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
2YYY	B	NAP	0.198	1.85	N97/ND2/HOH/O		HOH/O			Rossmann-like	Gp_dh_N_2
2YZQ	A	SAM	0.243	1.63	I232/N	HOH/O	I232/O/HOH/O	Q254/O	Q254/NE2	NO_X_NAME	IMPDH_2nd_1
2Z08	A	ATP	0.245	1.55	A38/N/H37/ND1	HOH/O	A38/O	HOH/O		HUP domain-like	Usp
2Z6R	A	SAH	0.206	1.5	A209/N		A209/O	L10/O	HOH/O	Tetrapyrrole methylase C-terminal domain-like	PTZ00175
2ZB4	A	NAP	0.244	1.63	HOH/O	HOH/O	HOH/O	HOH/O		none	none
2ZE6	A	AMP	0.264	2.1		D173/N	S45/O	D33/OD1/D33/OD2	HOH/O	P-loop domains-like	IPT
2ZFN	A	ACO	0.229	1.9						none	none
2ZFU	A	SAH	0.231	2	M347/N		D346/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_8
2ZPA	B	ACO	0.273	2.35						none	none
2ZUE	A	ANP	0.253	2	V418/N	Y415/OH	HOH/O/V418/O	HOH/O		HUP domain-like	tRNA-synt_1d
2ZVB	A	SAH	0.235	2	A215/N		A215/O	P11/O	HOH/O/HOH/O	Tetrapyrrole methylase C-terminal domain-like	TP_methylase_C
2ZW5	A	COA	0.287	2.4						none	none
2ZWA	A	SAH	0.211	1.7	L197/N		D196/OD1	HOH/O	HOH/O	Rossmann-like	LCM
2ZXI	C	FAD	0.248	2.3	V126/N		V126/O			Rossmann-like	FAD_binding_3_1st
3A27	A	SAM	0.255	2.01	N155/N	K128/N	D154/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_2
3A4M	A	ADP	0.252	1.79	HOH/O			HOH/O	HOH/O	none	none
3A8T	A	ATP	0.235	2.37			T74/O	D62/OD2	I222/N	P-loop domains-like	IPT
3A99	A	ANP	0.206	1.6		HOH/O	E121/O		HOH/O	NO_X_NAME	Pkinase_Tyr
3AB8	A	ATP	0.218	1.7	V36/N	HOH/O/HOH/O	V36/O	HOH/O		HUP domain-like	Usp
3ABI	A	NAD	0.235	2.44	A49/N	V30/N	D48/OD2			Rossmann-like	Sacchrp_dh_NADP
3AFN	A	NAP	0.209	1.63	L66/N		D65/OD1	HOH/O	HOH/O	Rossmann-like	adh_short_C2
3AJE	A	ANP	0.238	1.8	HOH/O		T118/OG1	HOH/O	HOH/O	NO_X_NAME	Sua5_yciO_yrdC
3ALJ	A	FAD	0.207	1.48	A130/N	K42/N	A130/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
3AR4	A	ATP	0.281	2.15		HOH/O	E442/OE2			NO_X_NAME	Cation_ATPase
3AUY	B	ADP	0.296	2.7	K64/N		I62/O			P-loop domains-like	SMC_N_1
3AXB	A	FAD	0.205	1.92	V184/N	A35/N	V184/O	HOH/O	HOH/O	Rossmann-like	DAO_1st
3AYJ	B	FAD	0.131	1.1	V374/N	A95/N	V374/O	HOH/O		Rossmann-like	Amino_oxidase_1st
3C1M	A	ANP	0.246	2.3	Y236/N		Y236/O	HOH/O		Other Rossmann-like structures with the crossover	AA_kinase
3C4A	A	FAD	0.262	2.3	L121/N		L121/O	N142/OD1	N142/OD1	Rossmann-like	FAD_binding_3_1st
3C4N	B	ADP	0.245	2.4	A195/N	E69/N	A195/O	HOH/O		none	none
3C7A	A	NAD	0.279	2.1		HOH/O				none	none
3C96	A	FAD	0.197	1.9	V132/N	S36/N/S36/OG	V132/O	HOH/O		Rossmann-like	FAD_binding_3_1st
3C9U	B	ADP	0.219	1.48			Y101/OH	N119/O		Bacillus chorismate mutase-like	AIRS
3CIN	A	NAD	0.194	1.7			V87/O	S91/OG		Rossmann-like	NAD_binding_5
3CIS	D	ATP	0.268	2.9	A43/N		A43/O			HUP domain-like	Usp
3CR3	B	ADP	0.263	2.1	HOH/O		G115/O		HOH/O	NO_X_NAME	Dak2

3CW9	A	AMP	0.209	2	HOH/O		N302/OD1	I303/O	G281/N	Other Rossmann-like structures with the crossover	AMP-binding_2nd
3CWQ	A	ADP	0.282	2.47	Y160/N		K158/O			P-loop domains-like	AAA_31
3D1C	A	FAD	0.225	2.4	V110/N	K35/N	V110/O	HOH/O		Rossmann-like	FAD_binding_3_1st
3D2F	A	ATP	0.244	2.3		R346/NH1			HOH/O	Ribonuclease H-like	FtsA
3D2M	A	COA	0.28	2.21	HOH/O					none	none
3D36	B	ADP	0.215	2.03	HOH/O	HOH/O	D352/OD2	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c
3D8B	A	ADP	0.237	2	A404/N	HOH/O	A404/O	T445/O		P-loop domains-like	Sigma54_activat
3DCM	X	SAM	0.235	2	I161/N		I161/O	N168/O		NO_X_NAME	Methyltrn_RNA_4
3DDD	A	COA	0.221	2.25		HOH/O		HOH/O	HOH/O	none	none
3DDJ	A	AMP	0.206	1.8	Q148/N	HOH/O	Q148/O	R170/O	HOH/O	NO_X_NAME	IMPDH_2nd_1
3DH0	A	SAM	0.296	2.72			E97/O	E97/OE2	K6/NZ	Rossmann-like	Methyltransf_11_1
3DJE	A	FAD	0.197	1.6	V187/N		V187/O	HOH/O	HOH/O	Rossmann-like	DAO_1st
3DJL	A	FAD	0.179	1.7	HOH/O	N429/ND2	HOH/O			Bromodomain-like	Acyl-CoA_dh_1
3DK9	A	FAD	0.152	0.95	A130/N	S51/N	A130/O	HOH/O	HOH/O	Rossmann-like	Pyr_redox_2
3DLC	A	SAM	0.138	1.15	V101/N	F73/N	D100/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11
3DME	A	FAD	0.217	1.7	L173/N	A35/N	L173/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
3DMG	A	SAH	0.22	1.55	V289/N	D263/N	D288/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_2
3DOU	A	SAM	0.195	1.45	I84/N	L68/N	D83/OD1	HOH/O	HOH/O	Rossmann-like	FtsJ
3DTT	A	NAP	0.199	1.7	HOH/O				HOH/O	none	none
3DUW	B	SAH	0.234	1.2	A119/N	A91/N	HOH/O	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_4
3DXY	A	SAM	0.185	1.5	A122/N			HOH/O	HOH/O	Rossmann-like	Methyltransf_4
3DZV	A	ADP	0.204	2.57						none	none
3E18	B	NAD	0.237	1.95	HOH/O					none	none
3E23	A	SAM	0.183	1.6	F94/N	G73/N	HOH/O	HOH/O		Rossmann-like	Methyltransf_11_3
3E2Q	A	FAD	0.228	1.75			T457/O	HOH/O	HOH/O	TIM beta/alpha-barrel	Pro_dh
3E8S	A	SAH	0.176	2.1	Y102/N	G81/N	HOH/O	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_3
3EA0	A	ATP	0.252	2.2	D205/N	N177/ND2	P203/O	N177/OD1		P-loop domains-like	AAA_31
3EC6	A	FAD	0.227	1.6		HOH/O				none	none
3ECC	A	ADP	0.294	2.7			H54/O			P-loop domains-like	Bac_DnaA_N
3EHG	A	ATP	0.215	1.74	HOH/O	HOH/O	D320/OD2	HOH/O	K325/N	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c
3EHH	B	ADP	0.232	2.1	HOH/O	K325/N	D320/OD2	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c
3EPS	B	AMP	0.268	2.8			N377/O	E376/O	HOH/O	NO_X_NAME	AceK_C
3EPS	B	ATP	0.283	2.34	M419/N	K336/NZ	R417/O	E416/OE1		NO_X_NAME	AceK_C
3EWK	A	FAD	0.232	2.05						none	none
3EZ2	A	ADP	0.232	2.05	L346/N	HOH/O	P344/O	HOH/O	S121/OG	P-loop domains-like	AAA_31
3F5O	G	COA	0.237	1.7						none	none
3F8D	A	FAD	0.226	1.4	V92/N	E46/N	V92/O			Rossmann-like	Pyr_redox_2
3F8K	A	COA	0.221	1.84	HOH/O	K123/NZ			HOH/O	Nat/Ivy	Acetyltransf_10_6

3F9X	B	SAH	0.207	1.25	HOH/O	HOH/O		H299/O	H299/N	beta-clip	SET
3FBU	A	COA	0.224	1.8	HOH/O					none	none
3FCE	A	ATP	0.252	1.9	HOH/O	G270/N		T293/O	HOH/O	Other Rossmann-like structures with the crossover	AMP-binding_2nd
3FDX	A	ATP	0.209	1.58	V37/N	HOH/O	V37/O	HOH/O		HUP domain-like	Usp
3FKQ	A	ATP	0.218	2.1	Y331/N	N307/ND2	P329/O	N307/OD1	HOH/O	P-loop domains-like	AAA_31_1
3FRH	A	SAH	0.159	1.2	V158/N		D157/OD1	HOH/O	HOH/O	Rossmann-like	FmrO_C
3FS8	A	ACO	0.241	1.7			A199/O		HOH/O	Single-stranded left-handed beta-helix	Hexapep_2_1
3FWY	B	ADP	0.199	1.63	L237/N		P235/O	N211/OD1	N211/ND2	P-loop domains-like	AAA_31
3FWZ	B	AMP	0.202	1.79	A468/N	T448/N	N467/OD1	E494/OE2	HOH/O	Rossmann-like	TrkA_N
3FYH	A	ADP	0.227	1.9		HOH/O	HOH/O			none	none
3FZG	A	SAM	0.206	2	HOH/O	I138/N	HOH/O		Q188/NE2	Rossmann-like	FmrO_C
3G5S	A	FAD	0.178	1.05	V121/N	M32/N	V121/O	S138/OG	S138/N	Rossmann-like	FAD_binding_3_1st
3G5T	A	SAH	0.137	1.12	S98/N		HOH/O	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_2
3G89	A	AMP	0.21	1.5		HOH/O			HOH/O/HOH/O	none	none
3G89	A	SAM	0.213	2	A139/N	A112/N	E140/OE1	HOH/O	HOH/O	Rossmann-like	GidB
3GDH	A	SAH	0.175	2.11	F748/N	I720/N	D747/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_15
3GGD	A	SAH	0.175	2.11	G109/N	V85/N	D108/OD1			Rossmann-like	Methyltransf_11_5
3GNI	B	ATP	0.254	2.35	M150/N		S148/O			NO_X_NAME	Pkinase_Tyr
3GO6	B	ADP	0.203	1.98	HOH/O		HOH/O		HOH/O	none	none
3GON	A	ANP	0.207	1.9			HOH/O	HOH/O		none	none
3GQV	A	NAP	0.22	1.74	HOH/O	HOH/O	HOH/O	HOH/O		none	none
3GU3	B	SAH	0.212	2.3	A98/N	S72/N	D97/OD2			Rossmann-like	Methyltransf_11_1
3GWC	H	FAD	0.229	1.9		HOH/O				none	none
3H1Q	A	ATP	0.278	2.8			K191/O			Ribonuclease H-like	FtsA
3H2B	A	SAH	0.226	2	I150/N		HOH/O		HOH/O	Rossmann-like	Methyltransf_11_3
3H5N	B	ATP	0.256	1.9	I194/N	H215/N	I194/O	H215/O	HOH/O	Rossmann-like	ThiF_1
3HFW	A	ADP	0.185	1.92	HOH/O			S124/O		NO_X_NAME	ADP_ribosyl_GH
3HGM	A	ATP	0.251	1.9	V38/N	HOH/O	V38/O			HUP domain-like	Usp
3HRD	G	FAD	0.251	2.2	M169/N		M169/O	HOH/O		FAD-binding domain-like	FAD_binding_5
3HTX	A	SAH	0.288	3.1	I779/N		S778/OG			Rossmann-like	Methyltransf_11_6
3HZ6	A	ADP	0.193	1.65	HOH/O					none	none
3I33	A	ADP	0.199	1.3			HOH/O	HOH/O	R345/NH1	Ribonuclease H-like	MreB_Mbl_C
3I3L	A	FAD	0.261	2.2	V131/N	R34/N	V131/O	HOH/O		Rossmann-like	FAD_binding_3_1st
3IC9	C	FAD	0.233	2.15	A116/N	G36/N	A116/O	HOH/O	HOH/O	Rossmann-like	Pyr_redox_2
3IE7	A	ATP	0.239	1.6	HOH/O	HOH/O	HOH/O	E243/OE1	HOH/O	Rossmann-like	PfkB
3IEI	B	SAH	0.232	1.9	L172/N		D171/OD1	HOH/O	HOH/O	Rossmann-like	LCM
3IHG	C	FAD	0.255	2.49	L143/N	R36/N				Rossmann-like	FAD_binding_3_1st
3IHL	A	ADP	0.232	2.8	V247/N		H245/O			P-loop domains-like	CTP_synth_N
3IKH	D	ATP	0.24	1.88	HOH/O		HOH/O		HOH/O	none	none
3IO3	A	ADP	0.247	1.8	L312/N		P310/O	N268/OD1	N268/ND2	P-loop domains-like	AAA_31
3IQ0	A	ATP	0.255	1.79	HOH/O	HOH/O	HOH/O		HOH/O	none	none

3IV6	B	SAM	0.238	2.7	I96/N		D95/OD1		HOH/O	Rossmann-like	Methyltransf_11_5
3IWC	A	SMM	0.238	1.9				HOH/O		none	none
3J94	E	ATP	0.302	4.2	G548/O	I508/N				P-loop domains-like	TIP49_1st
3JQQ	F	FAD	0.308	2.2				HOH/O		none	none
3JS8	A	FAD	0.196	1.54	I207/N	HOH/O	I207/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_4
3JU8	B	NAD	0.196	1.82						none	none
3JYO	A	NAD	0.194	1	HOH/O	L159/N	HOH/O	HOH/O	HOH/O	Rossmann-like	Shikimate_DH
3K1J	B	ADP	0.235	2	I44/N		I44/O	T71/O/HOH/O		P-loop domains-like	Sigma54_activat
3KA7	A	FAD	0.223	1.8	V219/N	R31/N	V219/O	HOH/O		Rossmann-like	Amino_oxidase_1st
3KD6	B	AMP	0.237	1.88	W152/NE1			HOH/O	HOH/O	Rossmann-like	PfkB
3KEO	B	NAD	0.234	1.5			HOH/O		HOH/O	none	none
3KH5	A	ADP	0.294	2.1	V15/N	Q10/NE2	V15/O	R37/O		NO_X_NAME	IMPDH_2nd_1
3KH5	A	AMP	0.186	1.68	I234/N		I234/O	A256/O	HOH/O	NO_X_NAME	IMPDH_2nd_1
3KKZ	A	SAM	0.254	2.1	M105/N	F77/N	S104/OG	HOH/O		Rossmann-like	Methyltransf_11_1
3KLJ	A	FAD	0.254	2.1	A79/N		A79/O	HOH/O		Rossmann-like	Pyr_redox_2
3KQN	A	ADP	0.223	2.05						none	none
3L77	A	NJP	0.18	1.55		V60/N			HOH/O	Rossmann-like	adh_short_C2
3L8K	B	ADP	0.243	2.5	V109/N		V109/O			Rossmann-like	Pyr_redox_2
3L9W	A	AMP	0.234	1.75	A450/N	H430/N	D449/OD1			Rossmann-like	TrkA_N
3LBE	D	COA	0.188	1.7				T97/OG1		Thioesterase/thiol ester dehydrase-isomerase-like	4HBT
3LCC	A	SAH	0.177	1.8	V124/N	I96/N	D123/OD1		HOH/O	Rossmann-like	TPMT
3LCV	B	SAM	0.276	2	L183/N		D182/OD1			Rossmann-like	FmrO_C
3LFR	A	AMP	0.212	1.53	I74/N	HOH/O	I74/O	R96/O	HOH/O	NO_X_NAME	IMPDH_2nd_1
3LKM	A	AMP	0.224	1.6	L716/N	HOH/O	P714/O	E713/OE1	K645/NZ	NO_X_NAME	Alpha_kinase
3LL3	B	ADP	0.224	2	HOH/O		HOH/O	HOH/O		none	none
3LL5	C	ATP	0.238	1.99	Y169/N	G201/N	Y169/O			Other Rossmann-like structures with the crossover	AA_kinase
3LLK	C	FAD	0.271	2			W478/O		N485/ND2	Four-helical up-and-down bundle	KOG1731_2nd
3LLM	B	ADP	0.238	2.8						none	none
3LNB	A	COA	0.244	2.01				HOH/O/HOH/O	HOH/O	none	none
3LO8	A	FAD	0.155	1.05	HOH/O	HOH/O				none	none
3LOQ	B	AMP	0.225	2.32	V34/N		V34/O			HUP domain-like	Usp
3LOV	A	FAD	0.241	2.06	L256/N	A36/N	L256/O		Q288/NE2	Rossmann-like	Amino_oxidase_1st
3LRT	B	ADP	0.22	1.53						none	none
3LSJ	A	COA	0.276	2.3						none	none
3LZW	A	FAD	0.196	1.8	V90/N	S38/N	V90/O			Rossmann-like	Pyr_redox_2
3LZW	A	NAP	0.23	2.3	HOH/O	HOH/O				none	none
3M2T	A	NAD	0.238	1.85	HOH/O	S37/N	HOH/O			Rossmann-like	GFO_IDH_MocA
3M31	A	FAD	0.238	1.85			S228/O		HOH/O	NO_X_NAME	ERO1
3M6A	F	ADP	0.313	3.4	H323/N			Y492/OH		Histone-like	Lon

3M6W	A	SAM	0.191	1.3		HOH/O		HOH/O/HOH/O	HOH/O/HOH/O	none	none
3M84	A	AMP	0.183	1.7			D107/OD2	E141/O		Bacillus chorismate mutase-like	AIRS
3MB5	A	SAM	0.199	1.6	I154/N		D153/OD2		HOH/O	Rossmann-like	Methyltransf_11_4
3MFI	A	DTP	0.187	1.76		HOH/O		HOH/O	HOH/O	none	none
3MGD	B	ACO	0.228	1.9						none	none
3MKH	A	FAD	0.209	2			HOH/O			none	none
3MQG	D	ACO	0.217	1.43					HOH/O	none	none
3N3Y	B	FAD	0.223	2.31		HOH/O				none	none
3NBX	X	ADP	0.269	2.91	Y23/N		Y23/O			P-loop domains-like	Sigma54_activat
3ND1	A	SAH	0.22	1.5	V209/N		V209/O	T11/O	HOH/O/HOH/O	Other Rossmann-like structures with the crossover	TP_methylase_C_1,TP_methylase_N
3NDC	B	SAH	0.222	2	A193/N		A193/O	HOH/O/P10/O	HOH/O/HOH/O	Tetrapyrrole methylase C-terminal domain-like	TP_methylase_C
3NG7	X	FAD	0.204	1.95	V226/N	G32/N	V226/O	HOH/O	HOH/O	Rossmann-like	Amino_oxidase_1st
3NIX	C	FAD	0.264	2.6	V129/N	K36/N	V129/O			Rossmann-like	FAD_binding_3_1st
3NKS	A	FAD	0.232	1.9	V257/N	S35/N	V257/O		HOH/O	Rossmann-like	Amino_oxidase_1st
3NL6	C	ACP	0.271	2.61	I455/N					Rossmann-like	PfkB
3NLC	A	FAD	0.271	2.15	V233/N	R128/N	V233/O	T270/OG1	R128/NH1	Rossmann-like	FAD_binding_3_1st
3NUA	B	ADP	0.161	1.4	V85/N	HOH/O	K83/O	H70/ND1	D192/N	NO_X_NAME	SAICAR_synt
3NVZ	K	FAD	0.246	1.6	L404/N	HOH/O	L404/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_5
3NYC	A	FAD	0.16	1.06	A1171/N	R1033/N	A1171/O	HOH/O	HOH/O	Rossmann-like	DAO_1st
3NYQ	A	AMP	0.216	1.43	HOH/O	HOH/O		R283/O	S261/N	Other Rossmann-like structures with the crossover	AMP-binding_2nd
3NZT	A	AMP	0.196	2	S141/OG		HOH/O	HOH/O	HOH/O	Glutamine synthetase-like	Glu_cys_ligase
3O0F	A	AMP	0.192	1.94		HOH/O	S162/OG	G265/O	R161/NH1	TIM beta/alpha-barrel	PHP
3O6X	C	ADP	0.269	3.5						none	none
3O9Z	A	NAD	0.226	1.45	Q86/NE2					Rossmann-like	GFO_IDH_MocA
3OC4	A	FAD	0.251	2.6	V79/N	K33/N	V79/O	HOH/O	T109/OG1/HOH/O	Rossmann-like	Pyr_redox_2
3OFK	C	SAH	0.238	1.85	I99/N	V74/N	D98/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_3
3OHR	A	ADP	0.186	1.66			HOH/O			none	none
3OKX	B	SAM	0.182	1.8	R80/N	HOH/O	H79/ND1	T137/OG1	L134/N	cradle loop barrel	UPF0066
3OND	B	NAD	0.159	1.17	HOH/O		HOH/O		N327/ND2	Rossmann-like	AdoHcyase_NAD
3ORH	A	SAH	0.236	1.86	W117/N	C91/N	E118/OE1	HOH/O		Rossmann-like	Methyltransf_11_2
3OU2	A	SAH	0.221	1.5	L99/N	G76/N	D98/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_3
3OYZ	A	ACO	0.239	1.95			T16/O	L259/O	S17/OG	TIM beta/alpha-barrel	HpcH_HpaI
3OZ2	A	FAD	0.171	1.6	A124/N	K34/N	A124/O	E162/OE1	HOH/O/HOH/O	Rossmann-like	FAD_binding_3_1st
3P0K	A	FAD	0.234	1.47		HOH/O	M222/O	N226/OD1	HOH/O	Four-helical up-and-down bundle	Baculo_p33_C
3P2E	B	SAH	0.232	1.68	A87/N		E88/OE1	T109/OG1	HOH/O/L110/N	Rossmann-like	Methyltransf_11_2
3PEY	A	ADP	0.179	1.4	HOH/O	HOH/O	K114/O		Q119/NE2	P-loop domains-like	DEAD
3PFG	A	SAM	0.209	1.35	M102/N		D101/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_5
3PFQ	A	ANP	0.244	4	V423/N		E421/O			NO_X_NAME	Pkinase_Tyr

3PL8	A	FAD	0.19	1.35	C283/N	I77/N	C283/O	HOH/O		Rossmann-like	GMC_oxred_N
3PM9	D	FAD	0.227	2.57	I216/N		I216/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_4
3PP9	C	ACO	0.222	1.6	HOH/O					none	none
3PVC	A	FAD	0.231	2.31	L435/N	A295/N	L435/O			Rossmann-like	DAO_1st
3PVZ	D	NAD	0.233	2.1	I95/N	I65/N	D94/OD1	HOH/O		Rossmann-like	Polysacc_synt_2_1
3PYF	A	ANP	0.249	1.7	L71/N	HOH/O		D109/OD1		NO_X_NAME	GHMP_kinases_N
3Q87	B	SAM	0.256	2	L70/N	L52/N	D69/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_2
3QB8	B	COA	0.168	1.5					HOH/O	none	none
3QCP	A	FAD	0.245	2.3		Y428/OH	W373/O	N377/OD1	N380/ND2	Four-helical up-and-down bundle	Evr1_Alr
3QF7	B	ANP	0.191	1.9	R63/N	HOH/O	V61/O	D59/O	Y54/OH	P-loop domains-like	SMC_N_1
3QFT	A	FAD	0.227	1.4						none	none
3QJ4	A	FAD	0.26	2.5	V132/N	K35/N	V132/O	HOH/O		Rossmann-like	Amino_oxidase_1st
3QKT	D	ANP	0.244	1.9	V64/N	HOH/O	T62/O	E60/O	HOH/O	P-loop domains-like	SMC_N_1
3QVP	A	FAD	0.188	1.2	V250/N	S51/N	V250/O	HOH/O	HOH/O	Rossmann-like	GMC_oxred_N
3QVS	A	NAD	0.243	1.7	Y185/OH		T99/O	HOH/O	HOH/O	Rossmann-like	NAD_binding_5
3QXC	A	ATP	0.173	1.34	HOH/O/HOH/O	N175/ND2	HOH/O	N175/OD1		P-loop domains-like	AAA_26
3R1K	A	COA	0.224	1.95	HOH/O		E122/OE2			Nat/Ivy	Acetyltransf_1
3R9F	B	COA	0.199	1.2	HOH/O		HOH/O	HOH/O		none	none
3RC1	A	NAP	0.234	1.71			HOH/O	HOH/O	HOH/O	none	none
3RC3	A	ANP	0.203	2.08	HOH/O	HOH/O				none	none
3RFA	A	SAM	0.242	2.05	N312/N		N312/O			TIM beta/alpha-barrel	Radical_SAM
3RPE	B	FAD	0.117	1.1				T23/OG1		Flavodoxin-like	FMN_red
3RPZ	A	AMP	0.148	1.51	HOH/O	HOH/O				none	none
3RQ4	A	SAM	0.221	1.8	E230/N	HOH/O	C229/SG	H183/O	H183/N	beta-clip	SET
3RTA	A	ACO	0.192	1.95	R122/N		E124/O	R122/O	HOH/O	Rossmann-like	YjeF_N
3RUV	C	ANP	0.22	2.24						none	none
3S1S	A	SAH	0.268	2.35	V389/N			HOH/O/HOH/O	HOH/O	Rossmann-like	N6_Mtase
3S3T	C	ATP	0.221	1.9	V41/N	HOH/O	V41/O	HOH/O		HUP domain-like	Usp
3S5W	B	FAD	0.217	1.9	V130/N	K46/N	V130/O			Rossmann-like	K_oxygenase
3SFZ	A	ADP	0.298	3	V127/N		V127/O			P-loop domains-like	NB-ARC_1st
3SL2	A	ATP	0.234	1.61	HOH/O	HOH/O	D533/OD2	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c
3SMT	A	SAM	0.259	2.04	HOH/O	HOH/O	HOH/O	H278/O	H278/N	beta-clip	SET_5
3SSO	D	SAH	0.176	1.89	Q253/N	I235/N	D252/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_24
3SX2	B	NAD	0.159	1.5	V81/N	L43/N	D80/OD1			Rossmann-like	adh_short_C2
3SX6	A	FAD	0.22	1.8	A78/N	A35/N	A78/O	HOH/O	HOH/O	Rossmann-like	Amino_oxidase_1st,Pyr_redox_2_2
3T58	B	FAD	0.259	2.4			W481/O	N485/OD1	N488/ND2	Four-helical up-and-down bundle	Evr1_Alr
3T7A	A	ADP	0.206	1.7	M240/N	HOH/O	E238/O	E237/OE1	K187/NZ	NO_X_NAME	RimK
3TD7	A	FAD	0.253	2.21		HOH/O	Y114/O	N118/OD1	N121/ND2	Four-helical up-and-down bundle	Evr1_Alr
3THX	A	ADP	0.273	2.7	I651/N		I651/O			P-loop domains-like	MutS_V
3TJ7	D	AMP	0.245	2.1			HOH/O			none	none

3TM4	B	SAM	0.196	1.95	A277/N		D276/OD1		HOH/O	Rossmann-like	UPF0020
3TNJ	A	AMP	0.241	2	V40/N		V40/O	D42/OD2		HUP domain-like	Usp
3TOS	G	SAH	0.173	1.55	V167/N	T108/N	D166/OD1	HOH/O		Rossmann-like	TylF
3TTC	A	ADP	0.202	1.86			E296/O	HOH/O	HOH/O	NO_X_NAME	Sua5_yciO_yrdC
3TUI	C	ADP	0.303	2.9						none	none
3TUT	A	ATP	0.252	1.58	HOH/O		HOH/O	HOH/O	Q288/NE2	IF3-like	RTC
3UX8	A	ADP	0.245	2.1						none	none
3V97	A	SAH	0.234	2.2	V291/N	S263/N	D290/OD1	HOH/O	HOH/O	Rossmann-like	UPF0020
3VC1	A	SAH	0.194	1.82	M156/N		N155/OD1	HOH/O	HOH/O	Rossmann-like	CMAS
3VOT	A	ADP	0.241	1.8	I201/N	R275/NH2	Q199/O	E198/OE2	K151/NZ	NO_X_NAME	ATP-grasp
3VPB	A	ADP	0.235	1.8	I170/N	HOH/O	E168/O		K127/NZ	NO_X_NAME	RimK
3VRD	B	FAD	0.195	1.5	A77/N		A77/O	HOH/O	HOH/O	Rossmann-like	Pyr_redox_2
3VYW	B	SAM	0.255	2.49	A175/N	K134/N	D174/OD1	E202/OE2		Rossmann-like	Methyltransf_30
3W0O	A	ADP	0.212	1.5	A95/N	HOH/O	R93/O	HOH/O	HOH/O	NO_X_NAME	Fructosamin_kin
3WE0	A	FAD	0.254	1.9	V301/N	A76/N	V301/O			Rossmann-like	Amino_oxidase_1st
3WGT	A	FAD	0.23	1.88	V164/N	D37/N	V164/O	HOH/O	HOH/O	Rossmann-like	DAO_1st
3WJP	A	ANP	0.184	1.53						none	none
3WNZ	A	ADP	0.243	1.9	L229/N	HOH/O	E227/O	E226/OE1	K178/NZ	NO_X_NAME	ATP-grasp
3WS7	A	NAP	0.186	1.18	HOH/O	HOH/O		D69/OD1	HOH/O	Rossmann-like	NAD_binding_2,RNAse_Pc_like
3WT0	D	ATP	0.236	2		HOH/O				none	none
3WXY	B	COA	0.199	1.71	HOH/O		I267/O	HOH/O	HOH/O	Other Rossmann-like structures with the crossover	ACP_syn_III_C
3X01	B	AMP	0.279	2.15	V204/N		R202/O			NO_X_NAME	PIP5K
3X0V	B	FAD	0.188	1.9	V252/N	S41/N/S41/OG	V252/O	HOH/O	HOH/O	Rossmann-like	Amino_oxidase_1st
3ZJ0	A	ACO	0.234	1.8				R177/O		Nat/Ivy	Acetyltransf_10_5
3ZL8	A	ADP	0.22	1.65	HOH/O	HOH/O	HOH/O	N264/OD1	N264/ND2	P-loop domains-like	Mur_ligase_M
3ZWC	A	NAD	0.28	2.3						none	none
4A2A	A	ATP	0.243	1.8						none	none
4A2N	B	SAH	0.275	3.4	V116/N					NO_X_NAME	ICMT
4A4Z	A	ANP	0.275	2.4			E329/O			P-loop domains-like	DEAD_1
4A6D	A	SAM	0.204	2.4	F237/N		D236/OD1			Rossmann-like	Methyltransf_11_1
4AFF	A	ATP	0.14	1.05		K90/NZ		HOH/O		Alpha-beta plaits	P-II
4AT0	A	FAD	0.175	1.6	V205/N	R52/N	V205/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
4AVA	A	ACO	0.23	1.7	HOH/O					none	none
4AXD	A	ANP	0.246	2.05	H149/N	HOH/O	N147/O		HOH/O	NO_X_NAME	Ins_P5_2-kin
4AYT	A	ACP	0.248	2.85		HOH/O	D264/OD2			NO_X_NAME	ABC_membrane
4AZS	A	AMP	0.222	2.15	L311/N	HOH/O	E309/O			NO_X_NAME	Pkinase_Tyr
4AZS	A	SAM	0.222	2.15	I110/N	F83/N	E111/OE1			Rossmann-like	Methyltransf_11_1
4B1Y	B	ATP	0.174	1.29	HOH/O		HOH/O	HOH/O	HOH/O	none	none
4B4D	A	FAD	0.23	1.5	HOH/O		T118/OG1			Other Rossmann-like structures with the crossover	NAD_binding_1
4B5O	A	ACO	0.157	1.05	HOH/O			HOH/O	HOH/O	none	none

4BBY	D	FAD	0.24	1.9	I374/N	HOH/O	I374/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_4
4BJZ	A	FAD	0.201	1.51	V134/N	R37/N	V134/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
4BLO	I	ADP	0.244	2.8						none	none
4BMV	I	NAP	0.251	2.5	L63/N		D62/OD1			Rossmann-like	adh_short_C2
4BT1	A	ADP		16						none	none
4BT1	B	ADP		16			V140/O			P-loop domains-like	Sigma54_activat
4C12	A	ADP	0.197	1.8	HOH/O		HOH/O	N304/OD1	N304/ND2	P-loop domains-like	Mur_ligase_M
4C3S	A	NAD	0.174	1.64					HOH/O	none	none
4C3X	A	FAD	0.208	2	L195/N	K38/N	L195/O	D261/OD2	HOH/O	Rossmann-like	FAD_binding_3_1st
4C4A	A	SAH	0.19	1.7	S123/N	V95/N	S123/OG	S158/OG	HOH/O	Rossmann-like	PrmA_1
4C5C	B	ATP	0.203	1.4	L183/N	K144/NZ	K181/O	E180/OE1	HOH/O	NO_X_NAME	ATP-grasp
4C5K	B	ADP	0.165	1.4	F204/N		F204/O	N209/OD1		Rossmann-like	PfkB
4C69	X	ATP	0.261	2.28						none	none
4CKB	A	SAH	0.255	2.8	N550/ND2		L549/O		R655/NH2	Rossmann-like	Methyltransf_11_3
4CNG	B	SAH	0.185	1.1	I131/N		I131/O	P138/O	L140/N	NO_X_NAME	SpoU_methylase
4CPD	C	NAD	0.248	2.74						none	none
4CS4	A	ANP	0.176	1.35	L339/N	HOH/O	L339/O	E332/OE2	HOH/O	NO_X_NAME	tRNA-synt_2d
4CS9	B	AMP	0.208	2.01	K8/N		P6/O	HOH/O	HOH/O	NO_X_NAME	Pneumovirus_M2_N
4CYI	B	ATP	0.234	2.42	H340/N		D338/O	HOH/O	HOH/O	NO_X_NAME	Pkinase_Tyr
4D25	A	ANP	0.208	1.9	HOH/O		K202/O		Q207/NE2	P-loop domains-like	DEAD
4D2I	B	ANP	0.254	2.84			I460/O			P-loop domains-like	DUF853
4D79	B	ATP	0.183	1.77	V109/N		V109/O			Rossmann-like	ThiF_2
4D7E	A	FAD	0.281	2.4	V137/N		V137/O		HOH/O	Rossmann-like	K_oxygenase
4D86	A	ADP	0.238	2	HOH/O					none	none
4DCM	A	SAM	0.252	2.3	A290/N	E260/N	N289/OD1			Rossmann-like	Methyltransf_11_2
4DG8	A	AMP	0.234	2.15			N305/OD1	G306/O	G283/N	Other Rossmann-like structures with the crossover	AMP-binding_2nd
4DJA	A	FAD	0.18	1.45	N406/ND2	HOH/O	N406/OD1	HOH/O	HOH/O	NO_X_NAME	FAD_binding_7
4DKJ	A	SAH	0.216	2.15	I114/N		D113/OD1		R154/NH2	Rossmann-like	DNA_methylase
4DMG	A	SAM	0.208	1.7	A270/N	K244/N	HOH/O		HOH/O	Rossmann-like	Methyltransf_11_4
4DPL	B	NAP	0.194	1.9	HOH/O	HOH/O				none	none
4DTH	A	ATP	0.192	1.78	S12/OG	HOH/O	HOH/O	HOH/O		Glutamine synthetase-like	ACD
4DZZ	B	ADP	0.244	1.8	R169/N		T167/O	HOH/O	HOH/O	P-loop domains-like	AAA_31
4E0I	C	FAD	0.304	3		W183/NE1	C159/O	N163/OD1	N166/ND2	Four-helical up-and-down bundle	Evr1_Alr
4E2X	A	SAH	0.236	1.4	F155/N		F155/O	Y184/OH	HOH/O	Rossmann-like	Methyltransf_11_3
4EH1	A	FAD	0.219	2.2						none	none
4EHU	B	ANP	0.197	1.6	HOH/O	HOH/O	HOH/O		HOH/O	none	none
4EQS	B	FAD	0.201	1.5	V81/N	K34/N	V81/O	HOH/O	HOH/O	Rossmann-like	Pyr_redox_2
4EYS	A	AMP	0.198	1.58			E23/OE2		HOH/O	Flavodoxin-like	Peptidase_S66_N
4F8Y	A	FAD	0.183	1.8		R174/NH2	HOH/O			Flavodoxin-like	FMN_red
4FBC	D	AMP	0.214	1.7	L88/N	R178/NH1/R178/NH2	L88/O		HOH/O	RIP/Polo-box domain	RIP

4FDA	A	NAP	0.229	2.1	F67/N		D66/OD1	HOH/O		Rossmann-like	adh_short_C2_1
4FFL	A	ADP	0.203	1.5	V163/N	HOH/O	E161/O	E160/OE1	K131/NZ	NO_X_NAME	ATP-grasp
4FFL	A	ATP	0.174	1.2	V50/N	N73/ND2	D49/OD1		K32/N	Rossmann-like	ATPgrasp_Ter
4FGL	D	FAD	0.209	2.6	K201/NZ		HOH/O	S20/OG		Flavodoxin-like	FMN_red
4FO0	A	ATP	0.248	2.7				HOH/O	HOH/O	none	none
4FT4	A	SAH	0.268	2.89	A398/N					SH3	DNA_methylase,Chromo_1
4FWI	B	ATP	0.268	2.89			N61/OD1			P-loop domains-like	ABC_tran
4FZV	A	SAM	0.19	2	G238/N	L205/N	D237/OD1	HOH/O	HOH/O	Rossmann-like	Methyltr_RsmB-F_1
4GA6	B	AMP	0.226	2.21				D256/OD1		Other Rossmann-like structures with the crossover	Glycos_transf_3
4GNI	A	ATP	0.244	1.8			HOH/O		HOH/O	none	none
4GT8	A	ADP	0.199	1.51	HOH/O	HOH/O	D290/OD2	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c
4GT9	A	FAD	0.164	1.39		HOH/O	HOH/O	HOH/O	HOH/O	none	none
4GVX	D	NAP	0.179	1.5	L64/N	R39/N	E63/OE1	HOH/O	HOH/O	Rossmann-like	adh_short_C2
4GYI	A	ADP	0.226	2.2	V192/N	HOH/O	S190/O	M189/SD	HOH/O	NO_X_NAME	RIO1
4H4R	A	FAD	0.216	1.4	A82/N	D40/N	A82/O	HOH/O	HOH/O	Rossmann-like	Pyr_redox_2
4HA6	A	FAD	0.233	2.1	V234/N	A49/N	V234/O	HOH/O		Rossmann-like	GMC_oxred_N
4HB9	A	FAD	0.218	1.93	F131/N	R31/N	F131/O	HOH/O	HOH/O/HOH/O	Rossmann-like	FAD_binding_3_3rd_1
4HG0	A	AMP	0.28	3.1	I80/N		I80/O	R102/O		NO_X_NAME	IMPDH_2nd_1
4HR3	A	FAD	0.194	1.8	HOH/O			HOH/O		none	none
4HSE	A	ADP	0.266	2.2	I173/N		I173/O			P-loop domains-like	Sigma54_activat
4HSU	A	FAD	0.232	1.99	V598/N	A413/N	V598/O	HOH/O	HOH/O	none	none
4HTF	B	SAM	0.201	1.6	A102/N	L74/N			HOH/O	Rossmann-like	Methyltransf_11_1
4I4T	F	ACP	0.204	1.8	L186/N	K198/NZ	K184/O		K150/NZ	NO_X_NAME	TTL
4ID9	A	NAD	0.213	1.6	L45/N	L31/N	S44/OG	HOH/O		Rossmann-like	GDP_Man_Dehyd_1,RNase_Pc_like
4IHQ	B	ADP	0.214	2	HOH/O	HOH/O	T239/OG1	HOH/O	HOH/O	P-loop domains-like	T2SSE
4IJN	B	AMP	0.201	1.7						none	none
4INE	B	SAH	0.194	1.45	A282/N	L255/N	D281/OD1	HOH/O		Rossmann-like	Methyltransf_11
4IV9	B	FAD	0.245	1.95	V300/N	S70/N	V300/O	HOH/O	HOH/O	Rossmann-like	Amino_oxidase_1st
4IVG	A	ANP	0.203	1.75	HOH/O	HOH/O	D173/OD2	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c
4IWX	A	ADP	0.257	2.85			Y179/O	E178/OE1	K141/NZ	NO_X_NAME	RimK
4J8F	A	ADP	0.261	2.7		HOH/O				none	none
4J91	D	ADP	0.278	2.93	A57/N		N56/OD1			Rossmann-like	TrkA_N
4JRN	A	ANP	0.289	2.71	A359/N		M357/O	M356/SD		NO_X_NAME	Pkinase_Tyr
4JWH	A	SAH	0.257	2.04	L229/N		HOH/O	K241/O	L243/N	NO_X_NAME	tRNA_m1G_MT
4JWJ	A	SAH	0.251	1.76	L232/N		HOH/O	R244/O	L246/N	NO_X_NAME	F_UNCLASSIFIED
4JXR	B	ACO	0.137	1.15						none	none
4K36	B	SAM	0.195	1.62	L195/N	HOH/O	L195/O	Y21/O		TIM beta/alpha-barrel	Radical_SAM_1
4KDR	A	SAH	0.233	2	HOH/O	M86/N	N11/O	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_6
4KGD	B	FAD	0.15	1.06	A326/N	I307/N	D325/OD1	HOH/O	HOH/O	Rossmann-like	TPP_enzyme_M

4KGM	D	ATP	0.232	2.36	D46/N			HOH/O		Alpha-beta plaits	Thg1
4KRG	A	SAH	0.14	1.68	A106/N	F81/N	D105/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_3
4KXF	B	ADP	0.266	3.2	T135/OG1/T135/N		T135/O			P-loop domains-like	NACHT
4L2I	A	FAD	0.208	1.45	A323/N	K305/N	D322/OD1	HOH/O		Rossmann-like	ETF_alpha
4L2I	B	FAD	0.156	1.2	M61/N		M61/O			HUP domain-like	ETF
4L8A	A	COA	0.244	1.7	HOH/O/HOH/O	HOH/O	HOH/O/HOH/O	HOH/O/HOH/O	HOH/O/HOH/O/HOH/O/HOH/O	none	none
4LJ9	A	ACP	0.188	1.62	V561/N	HOH/O	V561/O	V599/O		P-loop domains-like	Sigma54_activat
4LRJ	A	ANP	0.17	1.5	I204/N		D202/O		HOH/O/HOH/O	NO_X_NAME	Pkinase
4LRT	D	COA	0.17	1.5	HOH/O		HOH/O	HOH/O	HOH/O	none	none
4LV5	A	ADP	0.228	1.7	A340/N	HOH/O	P338/O	M337/SD	HOH/O	NO_X_NAME	Pkinase_Tyr
4LYA	A	ATP	0.263	2.45	HOH/O	HOH/O	I1197/O	HOH/O	HOH/O	P-loop domains-like	FtsK_SpoIIIE
4M37	A	SAH	0.183	1.7	S154/N	G126/N	S154/OG		HOH/O	Rossmann-like	PrmA
4M6T	A	SAM	0.268	2.5	R169/NH1		A26/O	K28/O	I30/N	none	none
4M73	B	SAH	0.211	2	A218/N	I190/N	D217/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_2
4M7T	A	SAM	0.183	1.56	S150/N	HOH/O	S150/O	Y22/O	HOH/O	TIM beta/alpha-barrel	Radical_SAM_1
4M8O	A	DTP	0.237	2.2					HOH/O	none	none
4MIY	A	NAD	0.207	1.42	HOH/O	I36/N		HOH/O		Rossmann-like	GFO_IDH_MocA
4MOB	A	ADP	0.211	2.4	HOH/O					none	none
4MOB	A	COA	0.25	1.9			HOH/O		K234/NZ	Thioesterase/thiol ester dehydrase-isomerase-like	4HBT
4MPO	F	AMP	0.184	1.61		HOH/O	E131/OE1	HOH/O	HOH/O	beta-Grasp	NUDIX_1
4MV4	A	ACP	0.226	1.8	L204/N	HOH/O	K202/O	E201/OE2	K159/NZ	NO_X_NAME	ATP-grasp
4MZ7	B	DTP	0.27	2.2	HOH/O			HOH/O		none	none
4MZU	J	COA	0.27	2.2						none	none
4N1A	A	ATP	0.266	3.24						none	none
4N49	A	SAM	0.203	1.9	I336/N	L302/N	D335/OD1	HOH/O	HOH/O	Rossmann-like	FtsJ
4N67	A	ADP	0.194	1.55	Q190/NE2			HOH/O	HOH/O	NO_X_NAME	Fic
4NDH	A	AMP	0.196	1.85		K194/N	HOH/O	HOH/O		HIT-like	DcpS_C
4NDO	A	ATP	0.177	1.35	Y196/N		Y196/O	P225/O		Other Rossmann-like structures with the crossover	AA_kinase
4NE2	B	ADP	0.214	1.9	HOH/O			HOH/O	K303/NZ	P-loop domains-like	PRK
4NEC	C	SAH	0.19	1.5	A95/N	L69/N	D94/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_1
4NH0	A	ATP	0.246	2.9			T1310/O			P-loop domains-like	FtsK_SpoIIIE
4NHE	B	NAP	0.207	1.95		HOH/O		HOH/O	N164/ND2	FwdE/GAPDH domain-like	GFO_IDH_MocA_C
4NJH	A	SAM	0.174	1.9	D176/N	Q173/NE2	D176/O	F48/O	D50/N	TIM beta/alpha-barrel	Fer4_14
4NL4	H	ADP	0.254	2.65						none	none
4NM9	B	FAD	0.192	1.9		S333/OG	T328/O			TIM beta/alpha-barrel	Pro_dh
4NTC	A	FAD	0.223	1.9	I91/N	S44/N	I91/O			Rossmann-like	Pyr_redox_2
4NTD	A	FAD	0.159	1.6	V84/N	S39/N	V84/O	HOH/O	HOH/O	Rossmann-like	F_UNCLASSIFIED
4O4F	A	ATP	0.222	1.7	L88/N	HOH/O	E86/O			NO_X_NAME	IPK

4O59	O	NAD	0.209	1.52	HOH/O		R77/O		HOH/O	Rossmann-like	Gp_dh_N_2
4O5Q	A	FAD	0.188	2	A290/N	E243/N	A290/O	HOH/O		Rossmann-like	Pyr_redox_2
4O9U	B	NAP	0.284	6.93	A436/N		D435/OD1			Rossmann-like	PNTB_C
4OAV	D	ACP	0.229	2.1	C437/N		T435/O		HOH/O	NO_X_NAME	Pkinase_Tyr
4OCV	A	ANP	0.2	1.47	I105/N	HOH/O	K103/O			NO_X_NAME	Fructosamin_kin
4OHX	A	ADP	0.219	1.98			E16/OE2	HOH/O	HOH/O	jelly-roll	CLP1_N
4OI4	A	ATP	0.221	2.4			D33/OD2			jelly-roll	CLP1_N_1
4OKE	A	AMP	0.232	1.7						none	none
4OL9	A	NAP	0.176	1.85	W84/NE1				HOH/O	Rossmann-like	ApbA
4OM8	B	NAD	0.188	1.55		V33/N			HOH/O	Rossmann-like	3HCDH_N
4OMF	B	FAD	0.181	1.71	HOH/O		T77/O	K75/O	K75/N	NO_X_NAME	FrhB_FdhB_C
4OTP	A	ADP	0.243	2.7	I280/N		S278/O			NO_X_NAME	RIO1
4P6V	F	FAD	0.276	3.5						none	none
4P8N	A	FAD	0.23	1.79	I184/N		I184/O	HOH/O	HOH/O	FAD-binding domain-like	FAD_binding_4
4PAB	A	FAD	0.231	1.85	V212/N	K74/N	V212/O			Rossmann-like	DAO_1st
4PIO	B	SAH	0.18	1.51	F142/N	V114/N	D141/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_33
4PL0	A	ANP	0.266	2.7						none	none
4PL9	A	ADP	0.182	1.9		HOH/O	D513/OD2/C573/SG	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c
4PNE	B	SAH	0.22	1.5	A131/N		D130/OD1	HOH/O		Rossmann-like	Methyltransf_11
4PSW	A	COA	0.212	2.1						none	none
4PU5	A	ANP	0.175	1.83	F222/N		E220/O		HOH/O	NO_X_NAME	HipA_C
4PVK	A	FAD	0.16	1.3	V215/N	E210/N	V215/O	S83/OG	S83/OG	FAD-binding domain-like	FAD_binding_4
4PWY	A	SAH	0.191	1.9	W212/N	G179/N	D213/OD1			Rossmann-like	Methyltransf_11_2
4Q86	D	AMP	0.25	2.25	HOH/O		HOH/O			none	none
4QDI	A	ATP	0.253	1.8		N296/ND2	HOH/O	N296/OD1	HOH/O	P-loop domains-like	Mur_ligase_M
4QDJ	A	SAM	0.18	1.6	L121/N	I92/N	D120/OD1	Y140/OH		Rossmann-like	Methyltransf_11_1
4QEO	A	SAH	0.226	2	Y613/N			H552/O	H552/N	beta-clip	SET
4QI5	A	FAD	0.244	2.4	V440/N	K260/N	V440/O	HOH/O/HOH/O	HOH/O	Rossmann-like	GMC_oxred_N
4QOS	A	ADP	0.19	1.42	L8/N/L9/N	HOH/O	L9/O	HOH/O	HOH/O	P-loop domains-like	Sigma54_activat
4QPM	A	ADP	0.269	2.2	Y869/N		E867/O		HOH/O	NO_X_NAME	Pkinase_Tyr
4QPN	A	SAH	0.167	1.25	W133/N		HOH/O		HOH/O	Rossmann-like	Methyltransf_11_2
4QRE	A	ATP	0.24	1.7	I302/N	HOH/O	I302/O	M309/O		HUP domain-like	tRNA-synt_1g
4QVH	B	COA	0.191	1.75		HOH/O		K449/O	HOH/O	Bacillus chorismate mutase-like	EntD
4R29	A	SAM	0.234	2.31	HOH/O		HOH/O	E208/O		R3H domain-like	EUF08394
4R33	B	SAH	0.181	1.78	R286/N		R286/O	M101/O		TIM beta/alpha-barrel	BATS
4R39	C	ANP	0.221	2.6	HOH/O		D282/OD1	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c_2
4R3A	B	ANP	0.33	2.92			D282/OD2			ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c_2
4R3L	A	COA	0.245	1.84					HOH/O	none	none

4R3N	A	NAP	0.161	1.35	HOH/O					none	none
4R78	A	AMP	0.212	1.45	I91/N	HOH/O	E89/O		HOH/O	NO_X_NAME	Fructosamin_kin
4R7Y	B	ADP	0.295	2.7			Y1291/O			P-loop domains-like	MCM_AAA
4R7Z	D	ADP	0.314	3.8	Y291/N		Y291/O			P-loop domains-like	MCM_AAA
4REK	A	FAD	0.123	0.74	V250/N	M41/N	V250/O	HOH/O	HOH/O	Rossmann-like	GMC_oxred_N
4REP	A	FAD	0.223	1.97	V247/N	K32/N	V247/O			Rossmann-like	Amino_oxidase_1st
4RFQ	A	SAM	0.238	2.4	W270/N	Q216/NE2	E269/OE1			Rossmann-like	Methyltransf_11_2
4RG1	A	SAH	0.208	1.86	T342/N/T342/OG1		N341/OD1		HOH/O	NO_X_NAME	Methyltrn_RNA_3_1st
4RV7	B	ATP	0.256	2.8	L88/N		L88/O			NO_X_NAME	DisA_N
4S1H	A	ADP	0.219	1.6	I209/N		I209/O		HOH/O	Rossmann-like	PfkB
4TL6	A	ANP	0.208	1.76			D241/OD1	S89/OG		P-loop domains-like	RecA
4TND	A	ANP	0.205	1.8	M266/N	HOH/O	T264/O	HOH/O	HOH/O	NO_X_NAME	Pkinase_Tyr
4U1Q	B	SAH	0.213	2.09	I229/N		D228/OD2	HOH/O		Rossmann-like	Methyltransf_11_4
4U63	A	FAD	0.18	1.67	N383/ND2	HOH/O	N383/OD1	HOH/O	HOH/O	NO_X_NAME	FAD_binding_7
4U7O	B	AN2	0.239	2.39	HOH/O		D542/OD2	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c
4U7T	C	SAH	0.261	2.9	V687/N		D686/OD1			Rossmann-like	DNA_methylase_1
4U89	A	COA	0.166	1.4		HOH/O		K78/O	HOH/O	Bacillus chorismate mutase-like	EntD
4U9U	A	FAD	0.196	1.55						none	none
4U9V	B	ACO	0.206	1.78	HOH/O					none	none
4UCI	A	SAM	0.201	2.21	A1756/N	L1720/N	D1755/OD1			Rossmann-like	Methyltrans_Mon_2nd_1
4UDQ	B	FAD	0.216	1.6	V233/N	A37/N	V233/O	HOH/O	HOH/O	Rossmann-like	GMC_oxred_N
4USR	A	FAD	0.204	1.83	V101/N	E35/N	V101/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
4UUU	A	SAM	0.199	1.71	L423/N	HOH/O	L423/O	Q445/O	Q445/NE2	NO_X_NAME	IMPDH_2nd_1
4UUW	A	AMP	0.224	1.98	T21/N/N22/N		D20/OD1	HOH/O		Other Rossmann-like structures with the crossover	MoCF_biosynth
4V03	A	ADP	0.279	1.9	E207/N		P205/O	N174/OD1	N174/ND2	P-loop domains-like	AAA_31
4V1T	A	ADP	0.244	2.14	N536/ND2				Q415/NE2	dsRBD-like	YcaO_2nd
4WBD	A	ADP	0.198	1.77	HOH/O	HOH/O	HOH/O/HOH/O	HOH/O	HOH/O	none	none
4WCX	C	SAM	0.213	1.59	R273/N		R273/O	Y96/O		TIM beta/alpha-barrel	BATS
4WER	A	AMP	0.231	2.05				T40/O		Flavodoxin-like	DAGK_cat
4WJI	A	NAP	0.159	1.4	HOH/O	HOH/O	HOH/O	HOH/O	HOH/O	none	none
4WJM	A	ANP	0.195	1.7	HOH/O	HOH/O		HOH/O	HOH/O	none	none
4WQM	A	FAD	0.189	1.62	HOH/O	HOH/O				none	none
4WUB	A	ANP	0.204	1.75	HOH/O		D73/OD2	HOH/O	HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c_1
4WUV	A	NAD	0.172	1.55	V73/N	L47/N	N72/OD1	HOH/O		Rossmann-like	adh_short_C2
4WW7	A	AMP	0.213	1.67	L109/N		E107/O		HOH/O	NO_X_NAME	RIO1
4WXX	A	SAH	0.245	2.62	C1191/N	M1169/N	D1190/OD1		HOH/O	Rossmann-like	DNA_methylase
4X9M	A	FAD	0.204	2.4	V177/N	K34/N	V177/O			Rossmann-like	FAD_binding_3_1st
4XC8	A	BCO	0.222	3.25			F587/O			TIM beta/alpha-barrel	MM_CoA_mutase

4XDU	A	ADP	0.188	1.35	N98/N	HOH/O	A96/O	I258/O	I258/N	T-fold	ApbE_C
4XFJ	B	ANP	0.182	1.55	I35/N		I35/O		HOH/O	HUP domain-like	Arginosuc_synth_N_1
4XGU	E	ADP	0.264	2.3	I140/N		I140/O			P-loop domains-like	Sigma54_activat
4XHP	A	ADP	0.287	3.2						none	none
4XJX	A	ATP	0.243	2.4	HOH/O	Q276/NE2	V271/O		K220/NZ	P-loop domains-like	DEAD_1
4XLO	A	FAD	0.23	1.67	V201/N		V201/O	G113/O		FAD-binding domain-like	FAD_binding_4
4XNH	C	ACO	0.228	2.1						none	none
4XQC	A	NAD	0.182	1.27	V66/N		V66/O	HOH/O	HOH/O	Rossmann-like	Sacchrp_dh_C,Sacchrp_dh_NADP
4XYM	C	COA	0.237	1.9	HOH/O			HOH/O		none	none
4XYM	D	A12	0.209	1.74	A114/N	K60/NZ	E112/O			NO_X_NAME	ATP-grasp_2
4Y0X	A	ADP	0.196	1.5	V235/N	HOH/O	E233/O	M232/SD	HOH/O	NO_X_NAME	Pkinase_Tyr
4Y1B	A	NAP	0.186	1.8	HOH/O	HOH/O	D305/OD1		HOH/O	Rossmann-like	ADH_zinc_N
4Y9J	A	FAD	0.223	1.9	HOH/O	HOH/O/HOH/O				none	none
4YBN	B	FAD	0.223	1.9		S163/OG				cradle loop barrel	Pyridox_ox_2
4YBR	A	NAP	0.216	1.65				G10/O		HUP domain-like	CTP_transf_like
4YDS	A	ATP	0.248	2.3		HOH/O		HOH/O		none	none
4YHB	B	FAD	0.219	1.89	HOH/O	N110/ND2/HOH/O		HOH/O		cradle loop barrel	FAD_binding_9
4YJ1	A	ADP	0.203	2.05	V278/N		V278/O	HOH/O	HOH/O	P-loop domains-like	ABC_ATPase_C
4YLR	A	ADP	0.269	2.55	I191/N	K201/NZ	E189/O			NO_X_NAME	TTL
4YS0	A	ADP	0.226	1.9	HOH/O	HOH/O	R75/O		Q80/NE2	P-loop domains-like	SecA_DEAD
4Z24	A	FAD	0.204	2	V263/N	A38/N	V263/O	HOH/O	HOH/O	Rossmann-like	GMC_oxred_N
4Z9D	D	NAD	0.227	1.8			HOH/O			none	none
4ZBG	A	ACO	0.163	1.25						none	none
4ZCD	A	FAD	0.176	1.66	I128/N	K33/N	I128/O	HOH/O	HOH/O	Rossmann-like	Amino_oxidase_1st
4ZCD	A	NAD	0.262	2.6		HOH/O	S96/O/S96/OG	D98/OD2	HOH/O	Rossmann-like	Amino_oxidase_1st
4ZCF	C	AMP	0.174	1.5						none	none
4ZFV	B	ADP	0.142	1.2	D237/N	HOH/O	D237/OD1	HOH/O	HOH/O	Ribonuclease H-like	AnmK
4ZJU	A	NAD	0.254	2.8	V67/N	HOH/O	D66/OD1	HOH/O/HOH/O	HOH/O	Rossmann-like	adh_short_C2_1
4ZM6	B	ACO	0.254	2.8						none	none
4ZQX	A	ATP	0.153	1.46	S234/OG		S234/O	HOH/O	HOH/O	jelly-roll	CPV_Polyhedrin
4ZV3	C	COA	0.296	3.1			N283/OD1			Thioesterase/thiol ester dehydrase-isomerase-like	4HBT
5AGA	A	ANP	0.264	2.9						none	none
5AHK	B	FAD	0.174	1.55	L328/N	L309/N	E327/OE2	HOH/O	HOH/O	Rossmann-like	TPP_enzyme_M
5AUN	B	ADP	0.213	1.63	Y219/N	HOH/O	P217/O	N187/OD1	N187/ND2	P-loop domains-like	AAA_31
5AYV	A	NAP	0.189	1.65	HOH/O		HOH/O	HOH/O		none	none
5B7I	A	ADP	0.249	2.6					L428/N	P-loop domains-like	DEAD_1
5B7N	A	SAH	0.198	1.4			HOH/O	N244/OD1	N244/ND2	Phosphorylase/hydrolase-like	PNP_UDP_1
5BN3	A	ADP	0.21	2	A493/N		HOH/O/Q491/O	HOH/O	HOH/O	NO_X_NAME	ATP-synt_ab_C
5BP9	A	SAH	0.179	1.5	L131/N		N130/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_2

5BQ5	A	ADP	0.203	2.1	D74/N	HOH/O	D74/O			P-loop domains-like	Bac_DnaA_N
5BR4	A	NAD	0.147	0.91	HOH/O	HOH/O	HOH/O	T140/O		Flavodoxin-like	Fe-ADH_N
5BT9	A	NAP	0.188	1.5	L71/N		D70/OD1	E72/OE1	HOH/O	Rossmann-like	adh_short_C2
5BUK	B	FAD	0.212	1.95	V146/N	S35/N	V146/O			Rossmann-like	FAD_binding_3_1st
5BVA	A	FAD	0.224	1.87	V124/N	K37/N	V124/O	HOH/O		Rossmann-like	FAD_binding_3_1st
5C3C	B	ADP	0.283	2.8			R17/O			P-loop domains-like	Sigma54_activat
5C40	B	ACP	0.181	1.5	HOH/O		HOH/O	HOH/O	HOH/O	none	none
5C8T	B	SAM	0.265	3.2	Y368/N	A353/N	Y368/O			Rossmann-like	NSP11_2nd
5CCB	A	SAH	0.222	2	V164/N	F136/N	D163/OD1			Rossmann-like	Methyltransf_11_4
5CIY	A	SAH	0.188	1.59	I61/N	W41/N	D60/OD1	HOH/O	HOH/O	Rossmann-like	DNA_methylase
5CKW	B	ANP	0.258	2.49	I148/N		T146/O			NO_X_NAME	Pkinase_Tyr
5CUO	A	COA	0.221	1.54	N211/ND2		N211/OD1	HOH/O	HOH/O	cradle loop barrel	PTAC_1
5CVD	B	SAH	0.16	1.3	L119/N	I92/N		HOH/O	HOH/O	Rossmann-like	Methyltransf_11_3
5CZY	A	SAM	0.248	2.2	Y245/N	HOH/O	Y245/O	F195/O	F195/N	beta-clip	SET
5D0N	A	AMP	0.235	3.2						none	none
5D2E	A	NAP	0.216	1.72	I334/N		D333/OD1	HOH/O	HOH/O	Rossmann-like	adh_short_C2_1
5D4N	B	AMP	0.235	1.6				HOH/O	HOH/O/HOH/O	none	none
5D4V	C	SAH	0.208	1.6		I78/N		HOH/O	HOH/O	Rossmann-like	DUF1188
5DCU	A	NAP	0.179	1.4	V85/N		D84/OD1		HOH/O	Rossmann-like	GDP_Man_Dehyd_1
5DGK	B	ANP	0.261	2.89			T335/O	T335/OG1		P-loop domains-like	DUF927_C
5DJH	A	AMP	0.169	1.45	HOH/O		G206/O	HOH/O		Other Rossmann-like structures with the crossover	Inositol_P_C
5DLY	A	SAH	0.237	1.5	A114/N	S92/N	N113/OD1	HOH/O/HOH/O	HOH/O	Rossmann-like	Methyltransf_11_1
5DM3	D	ADP	0.274	2.6			S266/OG			Glutamine synthetase-like	Gln-synt_C
5DMH	A	ADP	0.185	1.8		HOH/O	HOH/O	A324/O		Other Rossmann-like structures with the crossover	DUF1357_C_1
5DNK	B	SAH	0.249	1.9	I130/N	L103/N	S129/OG	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_1
5DP2	A	NAP	0.131	0.96	HOH/O	HOH/O	HOH/O/HOH/O	HOH/O	HOH/O	none	none
5DQR	D	FAD	0.246	2.7			T185/O	K183/O	K183/N	NO_X_NAME	FrhB_FdhB_C
5DZC	A	ANP	0.255	2.3	V614/N		E612/O	T611/OG1		NO_X_NAME	Pkinase_Tyr
5DZT	A	AMP	0.275	2.2	I315/N		E313/O	E312/OE1	K274/NZ	NO_X_NAME	DUF4135
5E3I	A	ATP	0.205	2.2	Y124/N	HOH/O	Y124/O	E117/OE2	HOH/O	NO_X_NAME	tRNA-synt_His
5EAN	A	ADP	0.246	2.36			G626/O		Q631/NE2	P-loop domains-like	AAA_11_1
5EC0	A	ADP	0.228	2.2	HOH/O	HOH/O				none	none
5EOX	B	ADP	0.257	2.4						none	none
5EPE	A	SAH	0.167	1.9	A95/N	L67/N	D94/OD1	HOH/O	W114/NE1	Rossmann-like	Methyltransf_11_4
5EPV	D	ACP	0.242	2.51			E423/OE1			ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c_7
5ER9	B	FAD	0.22	1.69	W234/N	R48/N	D233/OD1		R258/NH1	Rossmann-like	Amino_oxidase_1st
5EZ7	A	FAD	0.228	2.4		S37/N		HOH/O		Rossmann-like	FAD_binding_3_1st
5F2K	A	SAH	0.218	1.6	F134/N		S133/OG	HOH/O	HOH/O	Rossmann-like	Methyltransf_7

5F5E	A	SAH	0.236	1.8	N3958/N		C3957/SG	H3907/O	H3907/N	beta-clip	SET
5F5N	A	NAD	0.166	1.3	L59/N	HOH/O	D58/OD1	HOH/O/HOH/O	HOH/O	Rossmann-like	NmrA
5FA8	A	SAM	0.212	1.3	F88/N	I60/N	N87/OD1			Rossmann-like	Methyltransf_11_4
5FBS	A	ADP	0.318	2.59	I186/N		Q184/O			NO_X_NAME	PPDK_N
5FLG	A	ANP	0.232	2.04	V146/N		V146/O	E31/OE2	HOH/O	NO_X_NAME	BioW
5FRD	B	COA	0.173	1.4	HOH/O	HOH/O	HOH/O	R117/O	HOH/O	Other Rossmann-like structures with the crossover	Hydrolase_4_1
5FS8	A	FAD	0.192	1.4	V233/N	E164/N	V233/O	HOH/O		Rossmann-like	Pyr_redox_2
5FTB	A	ANP	0.183	1.38	I5/N	HOH/O	I5/O	HOH/O	HOH/O	P-loop domains-like	AAA_19
5FTJ	C	ADP		2.3	G207/N		G207/O	HOH/O		P-loop domains-like	Sigma54_activat
5FVJ	B	ACO	0.225	1.7			E135/O			Nat/Ivy	Acetyltransf_10_8
5G3Y	A	ADP	0.15	1.18	HOH/O		Q199/O	HOH/O	HOH/O	P-loop domains-like	SKI
5GI7	A	COA	0.17	1.2			HOH/O	HOH/O		none	none
5GMD	A	AMP	0.192	1.5			HOH/O		HOH/O	none	none
5GQI	A	ATP	0.169	1.3						none	none
5GUT	A	SAH	0.23	2.1	C1194/N	M1172/N	D1193/OD1	HOH/O	HOH/O	Rossmann-like	DNA_methylase
5GV8	A	FAD	0.14	0.78	HOH/O	HOH/O	T1088/OG1			cradle loop barrel	FAD_binding_6
5GXU	B	FAD	0.248	2.3	HOH/O		S416/OG	L414/O	R519/NH2	NO_X_NAME	FAD_binding_1_2nd
5GY7	A	NAD	0.16	1.43	I59/N	N32/N	D58/OD1	N99/OD1		Rossmann-like	GDP_Man_Dehyd
5GZ3	B	NAP	0.229	1.59						none	none
5GZA	A	ADP	0.218	2	H150/N		E148/O			NO_X_NAME	Pkinase_Tyr
5H02	A	SAH	0.196	1.78	W115/N	G89/N		HOH/O		Rossmann-like	Methyltransf_11_4
5HE9	A	ADP	0.218	1.9			HOH/O	HOH/O	HOH/O	none	none
5HNV	A	ATP	0.21	1.41	Y121/N	HOH/O	Q119/O		HOH/O	NO_X_NAME	Pkinase_Tyr
5HR5	A	ADP	0.162	1.82	HOH/O	N168/ND2	HOH/O	HOH/O	HOH/O	P-loop domains-like	6PF2K
5HSA	H	FAS	0.205	2.35	V227/N		V227/O	HOH/O		Rossmann-like	GMC_oxred_N
5HTX	A	ADP	0.172	1.49	N395/ND2		HOH/O			Ribonuclease H-like	FGGY_C
5HW4	A	SAM	0.229	2.21	L198/N		L198/O	I21/O	HOH/O	Tetrapyrrole methylase C-terminal domain-like	TP_methylase_C
5I9E	C	ATP	0.284	2.8		Q240/NE2				Ribonuclease H-like	Actin_2nd
5IB0	G	ACO	0.207	1.65			HOH/O		HOH/O	none	none
5IDJ	A	ADP	0.278	3.01			D479/OD2			ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c
5IL2	A	SAH	0.199	1.61	I378/N	N549/N	D377/OD2			Rossmann-like	MT-A70
5IN4	D	NAP	0.176	1.6	L87/N	HOH/O	D86/OD1	Y123/OH	HOH/O	Rossmann-like	GDP_Man_Dehyd
5IQ4	A	FAD	0.206	1.5	V126/N	K39/N	V126/O	HOH/O	HOH/O/HOH/O	Rossmann-like	FMO-like_1st
5IQ4	A	NAP	0.206	1.5			HOH/O	HOH/O		none	none
5IRN	A	ADP	0.236	2.34	T233/N		T233/O	HOH/O		P-loop domains-like	NACHT
5IX1	A	ANP	0.23	2.6			D67/OD1		HOH/O	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase-like	HATPase_c_3
5IZ4	A	ADP	0.182	1.75	V67/N	HOH/O	D66/OD1	HOH/O	HOH/O	Rossmann-like	adh_short_C2
5J1J	B	ANP	0.189	1.55	D217/N	HOH/O	P215/O	N181/OD1	N181/ND2	P-loop domains-like	AAA_31

5J1S	A	ATP	0.188	1.4	F70/N		F70/O	T106/O	HOH/O	P-loop domains-like	Torsin
5JAJ	A	ADP	0.168	1.5	HOH/O		E2/O		Q7/NE2	P-loop domains-like	DEAD_3
5JBX	B	COA	0.145	1.1	L66/N		A64/O	A62/O		NO_X_NAME	ECH_1
5JCA	L	FAD	0.153	1.5	V227/N	A185/N			HOH/O	Rossmann-like	Pyr_redox_2
5JCA	S	FAD	0.18	1.7	M221/N	HOH/O	M221/O	HOH/O	HOH/O/HOH/O	NO_X_NAME	F_UNCLASSIFIED
5JCI	A	FAD	0.184	1.4	I96/N	K40/N	I96/O	HOH/O	HOH/O	Rossmann-like	Pyr_redox_2
5JDA	A	AMP	0.154	1.52			HOH/O	HOH/O	HOH/O	none	none
5JE2	B	SAH	0.166	1.33	V96/N	I70/N	D95/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11_1
5JGK	A	SAH	0.174	1.4	A110/N	L83/N	N109/OD1		HOH/O	Rossmann-like	Methyltransf_11_3
5JIC	A	ADP	0.174	1.4					HOH/O	none	none
5JJS	A	SAH	0.198	1.65	V132/N	K105/N	D131/OD1	HOH/O		Rossmann-like	FtsJ
5JLB	A	SAH	0.19	1.5	F1679/N	HOH/O	C1678/SG	H1629/O	H1629/N	beta-clip	SET
5JM8	D	ATP	0.249	2.2	N487/ND2		N487/OD1	HOH/O		NO_X_NAME	lucA_lucC
5JPH	C	COA	0.17	1.46	HOH/O			HOH/O	HOH/O	none	none
5JWC	H	FAD	0.221	2.05	C117/N		C117/O	HOH/O	HOH/O	Rossmann-like	Pyr_redox_2
5K04	B	COA	0.239	2.4				HOH/O	HOH/O	none	none
5K5Z	A	ANP	0.263	2.37	R257/N		S255/O	HOH/O	HOH/O	P-loop domains-like	AAA_31
5K8C	A	NAD	0.218	1.85	HOH/O		HOH/O	T140/O		Flavodoxin-like	Fe-ADH_N
5KF6	B	FAD	0.222	1.7			T393/O			TIM beta/alpha-barrel	Pro_dh
5KF6	B	NAD	0.206	1.49			HOH/O	HOH/O		none	none
5KF9	A	ACO	0.215	1.44	HOH/O					none	none
5KFZ	A	DTP	0.215	1.8						none	none
5KOX	A	FAD	0.224	1.6	V122/N	K32/N	V122/O	T156/OG1	HOH/O/K32/NZ	Rossmann-like	FAD_binding_3_3rd
5KPG	B	SAH	0.193	1.83	V186/N		D185/OD1	HOH/O	HOH/O	Rossmann-like	CMAS
5KPR	A	ANP	0.193	1.83	HOH/O/HOH/O				R325/NH1	Ribonuclease H-like	Fumble_C
5KSD	B	ACP	0.324	3.5				D372/OD2		NO_X_NAME	Cation_ATPase
5KTC	A	COA	0.247	1.8			Y174/OH	HOH/O	HOH/O	Nat/Ivy	Acetyltransf_3
5KWA	A	ADP	0.269	2.9	G255/N	Y452/OH	G255/O			P-loop domains-like	Sigma54_activat
5L22	A	ADP	0.276	3.15						none	none
5L2X	A	DTP	0.255	2.2		HOH/O				none	none
5L4L	A	NAP	0.149	1.2	L50/N	HOH/O	D49/OD1	HOH/O	HOH/O	Rossmann-like	NmrA
5LB3	E	ADP	0.234	1.8	HOH/O	HOH/O	S28/O		Q34/NE2	P-loop domains-like	DEAD_1
5LJW	B	ANP	0.206	1.8		HOH/O	HOH/O			none	none
5LKT	A	BCO	0.204	2.04		HOH/O	I1457/O/HOH/O			Nat/Ivy	HAT_KAT11
5LLT	B	DND	0.235	2.2		HOH/O			HOH/O	HUP domain-like	CTP_transf_like_1
5LRT	A	ADP	0.198	1.85				P230/O	R433/NH2	Glutamine synthetase-like	Pup_ligase_N
5LTJ	A	ADP	0.199	1.78					HOH/O	none	none
5LVO	A	ATP	0.157	1.09	A162/N	HOH/O	S160/O	HOH/O	HOH/O	NO_X_NAME	Pkinase_Tyr
5LY3	A	ADP	0.2	1.6	HOH/O	HOH/O	HOH/O/HOH/O		HOH/O	none	none

5M10	A	FAD	0.165	1.22	V112/N	K40/N	V112/O	HOH/O		Rossmann-like	FMO-like_1st
5M10	A	NAP	0.217	1.87	HOH/O	HOH/O	HOH/O	HOH/O	HOH/O	none	none
5M45	H	AMP	0.285	3.2	L406/N		N404/O	HOH/O		Ribonuclease H-like	Hydantoinase_A_C
5MB9	A	ATP	0.2	1.4						none	none
5MBX	A	FAD	0.202	1.9	V240/N	A38/N	V240/O			Rossmann-like	Amino_oxidase_1st
5MGZ	A	SAH	0.257	3.19	A97/N	L71/N	D96/OD1	E4/OE1		Rossmann-like	Methyltransf_11_1
5MIO	C	ANP	0.257	3.19						none	none
5MOG	D	FAD	0.224	2.77	I342/N	A135/N	I342/O			Rossmann-like	Amino_oxidase_1st
5MPT	A	SAH	0.217	1.65	I2045/N	L2020/N	N2044/OD1	HOH/O	HOH/O	Rossmann-like	Methyltransf_11
5MW8	A	ATP	0.271	2.4	L118/N		P116/O			NO_X_NAME	Ins_P5_2-kin
5NAK	A	FAD	0.21	1.5	L135/N	R38/N	L135/O	HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
5NC8	A	AMP	0.26	3.09	A457/N	H437/N	D456/OD1			none	none
5NCC	F	FAD	0.278	3.12	V298/N		V298/O			Rossmann-like	GMC_oxred_N
5O0X	A	FAD	0.209	2.2						none	none
5ODQ	G	FAD	0.219	2.15	HOH/O	K174/N		S500/OG		TIM beta/alpha-barrel	FAD_binding_3_1st
5SVK	B	ATP	0.228	2.77						none	none
5T39	A	SAH	0.154	1.1	V112/N	G85/N	D111/OD1	HOH/O		Rossmann-like	Methyltransf_11_1
5T95	B	NAP	0.184	1.69	HOH/O	HOH/O	E80/OE1		HOH/O	Rossmann-like	F420_oxidored
5TEY	A	SAH	0.224	1.8	I378/N	N549/N	D377/OD1	HOH/O	HOH/O	Rossmann-like	MT-A70
5THY	B	SAH	0.215	2.09	I256/N		D255/OD1	E257/OE2	HOH/O	Rossmann-like	Methyltransf_11_3
5TPR	A	NAD	0.183	1.7		HOH/O	L183/O	T143/O		Flavodoxin-like	DHQ_synthase_N
5TSH	F	ADP	0.218	2.3						none	none
5TT5	A	NAD	0.205	1.55	K290/NZ	HOH/O	E113/OE2	L114/O	L116/N	NO_X_NAME	DNA_ligase_aden
5TT6	A	ATP	0.259	2.19		E100/N		T98/O		NO_X_NAME	RNA_lig_T4_1
5TTJ	B	FAD	0.231	2.2	V279/N	A84/N	V279/O	HOH/O	HOH/O	Rossmann-like	Amino_oxidase_1st
5TUI	A	FAD	0.219	1.75	HOH/O	K35/N		HOH/O	HOH/O	Rossmann-like	FAD_binding_3_1st
5U4Q	B	NAD	0.211	1.5	L61/N	N32/N	D60/OD1			Rossmann-like	GDP_Man_Dehyd
5U5G	A	NAP	0.207	2.05	HOH/O	HOH/O	HOH/O	HOH/O	HOH/O	none	none
5U8U	D	FAD	0.156	1.35	G122/N		G122/O	HOH/O		Rossmann-like	Pyr_redox_2
5UI9	A	NAD	0.202	1.92	HOH/O	R38/N				Rossmann-like	GFO_IDH_MocA
5UJ7	A	ATP	0.281	3.39				Y681/OH		Histone-like	KOG1514_1st
5UJ7	D	ATP	0.217	2.55						none	none
5UJ7	F	ATP	0.211	2.3						none	none
5UNA	F	SAH	0.236	2.7	Y560/N	I475/N	N559/OD1			Rossmann-like	Methyltransf_11_2
5UV4	A	ANP	0.199	1.5		HOH/O	A839/O	HOH/O		NO_X_NAME	Pkinase_Tyr
5UX5	A	FAD	0.223	2.1						none	none
5UZX	A	NAP	0.223	2.1	I66/N		D65/OD1	HOH/O	HOH/O	Rossmann-like	adh_short_C2_1
5V1T	A	SAM	0.223	2.1	S279/N	HOH/O	S279/O	F123/O		TIM beta/alpha-barrel	Radical_SAM_1
5VAC	A	SAH	0.228	1.95	V374/N		V374/O	G316/O	G316/N	beta-clip	SET
5VSC	B	SAM	0.182	1.4	Q1169/N	HOH/O	C1168/SG	H1113/O	H1113/N	beta-clip	SET
5WGG	A	SAM	0.225	2.04	V284/N	R253/NH2	V284/O	Y110/O		TIM beta/alpha-barrel	Radical_SAM_1
5WGX	A	FAD	0.199	1.97	V168/N	K71/N	V168/O	HOH/O		Rossmann-like	FAD_binding_3_1st

5X40	A	ACP	0.177	1.45	HOH/O					none	none
5X62	B	SAH	0.243	2.2	F235/N		HOH/O		HOH/O	Rossmann-like	Methyltransf_11_3
5X7F	A	SAM	0.216	2	A120/N	I91/N			HOH/O	Rossmann-like	Methyltransf_11_4
5X8F	C	AMP	0.193	1.76	HOH/O		HOH/O	S285/O	HOH/O	Other Rossmann-like structures with the crossover	AMP-binding_2nd
5XUN	B	ACO	0.247	2						none	none
5Y4Z	A	ANP	0.216	1.3						none	none
6AO7	A	ACO	0.244	1.85					HOH/O	none	none

Table 3. The list of PDB chains used in the analysis, including their R-free and resolution values, the bound ligand, the atoms which participate in the hydrogen bonds with any of adenine's nitrogen atoms, and ECOD's X- and F-groups assignments of the binding atoms.

7 References

1. Kessel, A. and N. Ben-Tal, *Introduction to proteins: structure, function, and motions*. Second ed. 2018, Boca Raton, Fl.: Taylor & Francis LLC.
2. Lupas, A.N., C.P. Ponting, and R.B. Russell, *On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world?* J Struct Biol, 2001. **134**(2-3): p. 191-203.
3. Kolodny, R., et al., *On the universe of protein folds.* Annu Rev Biophys, 2013. **42**: p. 559-82.
4. Zeldovich, K.B. and E.I. Shakhnovich, *Understanding protein evolution: from protein physics to Darwinian selection.* Annu Rev Phys Chem, 2008. **59**: p. 105-27.
5. Trifonov, E.N. and I.N. Berezovsky, *Evolutionary aspects of protein structure and folding.* Curr Opin Struct Biol, 2003. **13**(1): p. 110-4.
6. Choi, I.G. and S.H. Kim, *Evolution of protein structural classes and protein sequence families.* Proc Natl Acad Sci U S A, 2006. **103**(38): p. 14056-61.
7. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.
8. UniProt Consortium, T., *UniProt: the universal protein knowledgebase.* Nucleic Acids Res, 2018. **46**(5): p. 2699.
9. Wang, G. and R.L. Dunbrack, Jr., *PISCES: recent improvements to a PDB sequence culling server.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W94-8.
10. Wang, G. and R.L. Dunbrack, Jr., *PISCES: a protein sequence culling server.* Bioinformatics, 2003. **19**(12): p. 1589-91.
11. Steinegger, M. and J. Soding, *Clustering huge protein sequence sets in linear time.* Nature Communications, 2018. **9**.
12. Dawson, N.L., et al., *CATH: an expanded resource to predict protein function through structure and sequence.* Nucleic Acids Res, 2017. **45**(D1): p. D289-D295.
13. Knudsen, M. and C. Wiuf, *The CATH database.* Hum Genomics, 2010. **4**(3): p. 207-12.

14. Murzin, A.G., et al., *SCOP: a structural classification of proteins database for the investigation of sequences and structures.* J Mol Biol, 1995. **247**(4): p. 536-40.

15. Cheng, H., et al., *ECOD: an evolutionary classification of protein domains.* PLoS Comput Biol, 2014. **10**(12): p. e1003926.

16. Finn, R.D., et al., *Pfam: the protein families database.* Nucleic Acids Research, 2014. **42**(D1): p. D222-D230.

17. Finn, R.D., et al., *The Pfam protein families database: towards a more sustainable future.* Nucleic Acids Res, 2016. **44**(D1): p. D279-85.

18. Berntsson, R.P., et al., *A structural classification of substrate-binding proteins.* FEBS Lett, 2010. **584**(12): p. 2606-17.

19. Kosloff, M. and R. Kolodny, *Sequence-similar, structure-dissimilar protein pairs in the PDB.* Proteins, 2008. **71**(2): p. 891-902.

20. Perutz, M.F., *Stereochemistry of cooperative effects in haemoglobin.* Nature, 1970. **228**(5273): p. 726-39.

21. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: Survey of methods for simulating the activity of proteins.* Chemical Reviews, 2006. **106**(5): p. 1589-1615.

22. Eyal, E., G. Lum, and I. Bahar, *The anisotropic network model web server at 2015 (ANM 2.0).* Bioinformatics, 2015. **31**(9): p. 1487-1489.

23. Flores, S.C. and M.B. Gerstein, *FlexOracle: predicting flexible hinges by identification of stable domains.* Bmc Bioinformatics, 2007. **8**.

24. Flores, S.C. and M.B. Gerstein, *Predicting protein ligand binding motions with the conformation explorer.* Bmc Bioinformatics, 2011. **12**.

25. Grant, B.J., A.A. Gorfe, and J.A. McCammon, *Large conformational changes in proteins: signaling and other functions.* Current Opinion in Structural Biology, 2010. **20**(2): p. 142-147.

26. Korkut, A. and W.A. Hendrickson, *A force field for virtual atom molecular mechanics of proteins.* Proc Natl Acad Sci U S A, 2009. **106**(37): p. 15667-72.

27. Laughton, C.A., M. Orozco, and W. Vranken, *COCO: a simple tool to enrich the representation of conformational variability in NMR structures.* Proteins, 2009. **75**(1): p. 206-16.

28. Gerstein, M. and W. Krebs, *A database of macromolecular motions.* Nucleic Acids Research, 1998. **26**(18): p. 4280-4290.

29. Juritz, E.I., S.F. Alberti, and G.D. Parisi, *PCDB: a database of protein conformational diversity.* Nucleic Acids Research, 2011. **39**: p. D475-D479.

30. Li, W., et al., *ChSeq: A database of chameleon sequences.* Protein Sci, 2015. **24**(7): p. 1075-86.

31. Monzon, A.M., et al., *CoDNaS: a database of conformational diversity in the native state of proteins.* Bioinformatics, 2013. **29**(19): p. 2512-4.

32. Biasini, M., et al., *SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information.* Nucleic Acids Research, 2014. **42**(W1): p. W252-W258.

33. Pieper, U., et al., *ModBase, a database of annotated comparative protein structure models and associated resources.* Nucleic Acids Res, 2014. **42**(Database issue): p. D336-46.

34. Narunsky, A., et al., *ConTemplate Suggests Possible Alternative Conformations for a Query Protein of Known Structure.* Structure, 2015. **23**(11): p. 2162-70.

35. Changela, A., et al., *Molecular basis of metal-ion selectivity and zeptomolar sensitivity by CueR.* Science, 2003. **301**(5638): p. 1383-7.

36. Finney, L.A. and T.V. O'Halloran, *Transition metal speciation in the cell: insights from the chemistry of metal ion receptors.* Science, 2003. **300**(5621): p. 931-6.

37. Robinson, N.J. and D.R. Winge, *Copper metallochaperones.* Annu Rev Biochem, 2010. **79**: p. 537-62.

38. Waldron, K.J., et al., *Metalloproteins and metal sensing.* Nature, 2009. **460**(7257): p. 823-30.

39. Brown, N.L., et al., *The MerR family of transcriptional regulators.* FEMS Microbiol Rev, 2003. **27**(2-3): p. 145-63.

40. Ma, Z., F.E. Jacobsen, and D.P. Giedroc, *Coordination chemistry of bacterial metal transport and sensing.* Chem Rev, 2009. **109**(10): p. 4644-81.

41. Outten, F.W., et al., *Transcriptional activation of an Escherichia coli copper efflux regulon by the chromosomal MerR homologue, cueR.* J Biol Chem, 2000. **275**(40): p. 31024-9.

42. Stoyanov, J.V., J.L. Hobman, and N.L. Brown, *CueR (YbbI) of Escherichia coli is a MerR family regulator controlling expression of the copper exporter CopA.* Mol Microbiol, 2001. **39**(2): p. 502-11.

43. Grass, G. and C. Rensing, *CueO is a multi-copper oxidase that confers copper tolerance in Escherichia coli.* Biochem Biophys Res Commun, 2001. **286**(5): p. 902-8.

44. Andoy, N.M., et al., *Single-molecule study of metalloregulator CueR-DNA interactions using engineered Holliday junctions.* Biophys J, 2009. **97**(3): p. 844-52.

45. Joshi, C.P., et al., *Direct substitution and assisted dissociation pathways for turning off transcription by a MerR-family metalloregulator.* Proc Natl Acad Sci U S A, 2012. **109**(38): p. 15121-6.

46. Philips, S.J., et al., *TRANSCRIPTION. Allosteric transcriptional regulation via changes in the overall topology of the core promoter.* Science, 2015. **349**(6250): p. 877-81.

47. Newberry, K.J. and R.G. Brennan, *The structural mechanism for transcription activation by MerR family member multidrug transporter activation, N terminus.* J Biol Chem, 2004. **279**(19): p. 20356-62.

48. Kumaraswami, M., K.J. Newberry, and R.G. Brennan, *Conformational plasticity of the coiled-coil domain of BmrR is required for bmr operator binding: the structure of unliganded BmrR.* J Mol Biol, 2010. **398**(2): p. 264-75.

49. Aitha, M., et al., *Conformational dynamics of metallo-beta-lactamase CcrA during catalysis investigated by using DEER spectroscopy.* J Biol Inorg Chem, 2015. **20**(3): p. 585-94.

50. Bhatnagar, J., et al., *Self-association of the histidine kinase CheA as studied by pulsed dipolar ESR spectroscopy.* Biophys J, 2012. **102**(9): p. 2192-201.

51. Freed, D.M., et al., *Monomeric TonB and the Ton box are required for the formation of a high-affinity transporter-TonB complex.* Biochemistry, 2013. **52**(15): p. 2638-48.

52. Jeschke, G., *DEER distance measurements on proteins.* Annu Rev Phys Chem, 2012. **63**: p. 419-46.

53. Joseph, B., et al., *Conformational cycle of the vitamin B12 ABC importer in liposomes detected by double electron-electron resonance (DEER).* J Biol Chem, 2014. **289**(6): p. 3176-85.

54. Klare, J.P., *Site-directed spin labeling EPR spectroscopy in protein research.* Biol Chem, 2013. **394**(10): p. 1281-300.

55. Sahu, I.D., et al., *DEER EPR measurements for membrane protein structures via bifunctional spin labels and lipodisq nanoparticles.* Biochemistry, 2013. **52**(38): p. 6627-32.

56. Sameach, H., et al., *Structural and Dynamics Characterization of the MerR Family Metalloregulator CueR in its Repression and Activation States.* Structure, 2017. **25**(7): p. 988-996 e3.

57. Denessiouk, K.A. and M.S. Johnson, *When fold is not important: a common structural framework for adenine and AMP binding in 12 unrelated protein families.* Proteins, 2000. **38**(3): p. 310-26.

58. Denessiouk, K.A. and M.S. Johnson, *"Acceptor-donor-acceptor" motifs recognize the Watson-Crick, Hoogsteen and Sugar "donor-acceptor-donor" edges of adenine and adenosine-containing ligands.* J Mol Biol, 2003. **333**(5): p. 1025-43.

59. Denessiouk, K.A., V.V. Rantanen, and M.S. Johnson, *Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins.* Proteins, 2001. **44**(3): p. 282-91.

60. Nepomnyachiy, S., N. Ben-Tal, and R. Kolodny, *Complex evolutionary footprints revealed in an analysis of reused protein segments of diverse lengths.* Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(44): p. 11703-11708.

61. Remington, S., G. Wiegand, and R. Huber, *Crystallographic Refinement and Atomic Models of 2 Different Forms of Citrate Synthase at 2.7-a and 1.7-a Resolution.* Journal of Molecular Biology, 1982. **158**(1): p. 111-152.

62. Kobayashi, N. and N. Go, *ATP binding proteins with different folds share a common ATP-binding structural motif.* Nature Structural Biology, 1997. **4**(1): p. 6-7.

63. Kuttner, Y.Y., et al., *A consensus-binding structure for adenine at the atomic level permits searching for the ligand site in a wide spectrum of adenine-containing complexes.* Proteins-Structure Function and Genetics, 2003. **52**(3): p. 400-411.

64. Shulman-Peleg, A., R. Nussinov, and H.J. Wolfson, *Recognition of functional sites in protein structures.* Journal of Molecular Biology, 2004. **339**(3): p. 607-633.

65. He, W., et al., *LibME-automatic extraction of 3D ligand-binding motifs for mechanistic analysis of protein-ligand recognition.* Febs Open Bio, 2016. **6**(12): p. 1331-1340.

66. Aziz, M.F., K. Caetano-Anolles, and G. Caetano-Anolles, *The early history and emergence of molecular functions and modular scale-free network behavior.* Scientific Reports, 2016. **6**.

67. Caetano-Anolles, G., K.M. Kim, and D. Caetano-Anolles, *The Phylogenomic Roots of Modern Biochemistry: Origins of Proteins, Cofactors and Protein Biosynthesis.* Journal of Molecular Evolution, 2012. **74**(1-2): p. 1-34.

68. Caetano-Anolles, G., et al., *The origin, evolution and structure of the protein world.* Biochem J, 2009. **417**(3): p. 621-37.

69. Koonin, E.V., R.L. Tatusov, and M.Y. Galperin, *Beyond complete genomes: from sequence to structure and function.* Current Opinion in Structural Biology, 1998. **8**(3): p. 355-363.

70. Koonin, E.V., Y.I. Wolf, and G.P. Karev, *The structure of the protein universe and genome evolution.* Nature, 2002. **420**(6912): p. 218-223.

71. Glasner, M.E., J.A. Gerlt, and P.C. Babbitt, *Evolution of enzyme superfamilies.* Current Opinion in Chemical Biology, 2006. **10**(5): p. 492-497.

72. Goncearenco, A. and I.N. Berezovsky, *Protein function from its emergence to diversity in contemporary proteins.* Physical Biology, 2015. **12**(4).

73. Romero Romero, M.L., A. Rabin, and D.S. Tawfik, *Functional Proteins from Short Peptides: Dayhoff's Hypothesis Turns 50.* Angew Chem Int Ed Engl, 2016. **55**(52): p. 15966-15971.

74. Toth-Petroczy, A. and D.S. Tawfik, *The robustness and innovability of protein folds.* Curr Opin Struct Biol, 2014. **26**: p. 131-8.

75. Tokuriki, N. and D.S. Tawfik, *Protein Dynamism and Evolvability.* Science, 2009. **324**(5924): p. 203-207.

76. Gilbert, W., *Origin of Life - the Rna World.* Nature, 1986. **319**(6055): p. 618-618.

77. Fetrow, J.S. and A. Godzik, *Function driven protein evolution. A possible proto-protein for the RNA-binding proteins.* Pac Symp Biocomput, 1998: p. 485-96.

78. Soding, J. and A.N. Lupas, *More than the sum of their parts: on the evolution of proteins from peptides.* Bioessays, 2003. **25**(9): p. 837-46.

79. Schiller, M.R., *The minimotif synthesis hypothesis for the origin of life.* J Transl Sci, 2016. **2**(5): p. 289-296.

80. Romero Romero, M.L., et al., *Simple yet functional phosphate-loop proteins.* Proc Natl Acad Sci U S A, 2018. **115**(51): p. E11943-E11950.

81. Akagawa, K. and K. Kudo, *Development of Selective Peptide Catalysts with Secondary Structural Frameworks.* Acc Chem Res, 2017. **50**(10): p. 2429-2439.

82. Carter, C.W., Jr., *Urzymology: experimental access to a key transition in the appearance of enzymes.* J Biol Chem, 2014. **289**(44): p. 30213-20.

83. Walker, J.E., et al., *Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold.* EMBO J, 1982. **1**(8): p. 945-51.

84. Alva, V., J. Soding, and A.N. Lupas, *A vocabulary of ancient peptides at the origin of folded proteins.* Elife, 2015. **4**: p. e09410.

85. Marcotte, E.M., et al., *A census of protein repeats.* J Mol Biol, 1999. **293**(1): p. 151-60.

86. Andrade, M.A., et al., *Homology-based method for identification of protein repeats using statistical significance estimates.* J Mol Biol, 2000. **298**(3): p. 521-37.

87. Aravind, L. and E.V. Koonin, *Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches.* J Mol Biol, 1999. **287**(5): p. 1023-40.

88. Milner-White, E.J. and M.J. Russell, *Functional capabilities of the earliest peptides and the emergence of life.* Genes (Basel), 2011. **2**(4): p. 671-88.

89. Lyon, K.F., et al., *Minimotif Miner 4: a million peptide minimotifs and counting.* Nucleic Acids Res, 2018. **46**(D1): p. D465-D470.

90. Chothia, C., et al., *Evolution of the protein repertoire.* Science, 2003. **300**(5626): p. 1701-1703.

91. Altschul, S.F., et al., *Basic Local Alignment Search Tool.* Journal of Molecular Biology, 1990. **215**(3): p. 403-410.

92. Brenner, S.E., P. Koehl, and R. Levitt, *The ASTRAL compendium for protein structure and sequence analysis.* Nucleic Acids Research, 2000. **28**(1): p. 254-256.

93. Kabsch, W., *Discussion of Solution for Best Rotation to Relate 2 Sets of Vectors.* Acta Crystallographica Section A, 1978. **34**(Sep): p. 827-828.

94. Kabsch, W., *A solution for the best rotation to relate two sets of vectors.* Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography, 1976. **32**(5): p. 922-923.

95. Krissinel, E., *Enhanced fold recognition using efficient short fragment clustering.* Journal of molecular biochemistry, 2012. **1**(2): p. 76.

96. Budowski-Tal, I., Y. Nov, and R. Kolodny, *FragBag, an accurate representation of protein structure, retrieves structural neighbors from the entire PDB quickly and accurately.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(8): p. 3481-3486.

97. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Research, 2004. **32**(5): p. 1792-1797.

98. Seber, G., *Dimension reduction and ordination.* Multivariate observations, 1984: p. 175-204.

99. Spath, H., *The cluster dissection and analysis theory fortran programs examples*. 1985: Prentice-Hall, Inc.

100. Choi, I.G., J. Kwon, and S.H. Kim, *Local feature frequency profile: A method to measure structural similarity in proteins.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(11): p. 3797-3802.

101. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints.* J Mol Biol, 1993. **234**(3): p. 779-815.

102. Saito, R., et al., *A travel guide to Cytoscape plugins.* Nat Methods, 2012. **9**(11): p. 1069-76.

103. Jeschke, G., *DeerAnalysis 2013 User Manual.* 2013.

104. Hansen, P., *The L-curve and its use in the numerical treatment of inverse problems. Computational Inverse Problems in Electrocardiology.* Advances in Computational Bioengineering, 2001. **5**: p. 119.

105. Tikhonov, A.N. *On the solution of ill-posed problems and the method of regularization*. in *Doklady Akademii Nauk*. 1963. Russian Academy of Sciences.

106. Polyhach, Y., E. Bordignon, and G. Jeschke, *Rotamer libraries of spin labelled cysteines for protein studies.* Phys Chem Chem Phys, 2011. **13**(6): p. 2356-66.

107. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

108. O'Boyle, N.M., et al., *Open Babel: An open chemical toolbox.* J Cheminform, 2011. **3**: p. 33.

109. Munkres, J., *Algorithms for the assignment and transportation problems.* Journal of the society for industrial and applied mathematics, 1957. **5**(1): p. 32-38.

110. Jubb, H.C., et al., *Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures.* J Mol Biol, 2017. **429**(3): p. 365-371.

111. *The PyMOL Molecular Graphics System, Version 2.0*. 2017.

112. Steinegger, M. and J. Soding, *MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets.* Nature Biotechnology, 2017. **35**(11): p. 1026-1028.

113. Eddy, S.R., *Accelerated Profile HMM Searches.* PLoS Comput Biol, 2011. **7**(10): p. e1002195.

114. Cowan-Jacob, S.W., et al., *The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation.* Structure, 2005. **13**(6): p. 861-871.

115. Xu, W., et al., *Crystal structures of c-Src reveal features of its autoinhibitory mechanism.* Mol Cell, 1999. **3**(5): p. 629-38.

116. Nepomnyachiy, S., N. Ben-Tal, and R. Kolodny, *CyToStruct: Augmenting the Network Visualization of Cytoscape with the Power of Molecular Viewers.* Structure, 2015. **23**(5): p. 941-948.

117. Shilton, B.H., et al., *Conformational changes of three periplasmic receptors for bacterial chemotaxis and transport: the maltose-, glucose/galactose- and ribose-binding proteins.* J Mol Biol, 1996. **264**(2): p. 350-63.

118. Quiocho, F.A., *Atomic structures and function of periplasmic receptors for active transport and chemotaxis.* Current Opinion in Structural Biology, 1991. **1**(6): p. 922-933.

119. Bjorkman, A.J. and S.L. Mowbray, *Multiple open forms of ribose-binding protein trace the path of its conformational change.* Journal of Molecular Biology, 1998. **279**(3): p. 651-664.

120. Bjorkman, A.J., et al., *Probing Protein-Protein Interactions - the Ribose-Binding Protein in Bacterial Transport and Chemotaxis.* Journal of Biological Chemistry, 1994. **269**(48): p. 30206-30211.

121. Sooriyaarachchi, S., et al., *Conformational changes and ligand recognition of Escherichia coli D-xylose binding protein revealed.* J Mol Biol, 2010. **402**(4): p. 657-68.

122. Chaudhuri, B.N., et al., *Structure of D-allose binding protein from Escherichia coli bound to D-allose at 1.8 angstrom resolution.* Journal of Molecular Biology, 1999. **286**(5): p. 1519-1531.

123. Yang, A.S. and B. Honig, *An integrated approach to the analysis and modeling of protein sequences and structures. III. A comparative study of sequence conservation in protein structural families using multiple structural alignments.* J Mol Biol, 2000. **301**(3): p. 691-711.

124. Das, A., et al., *Exploring the Conformational Transitions of Biomolecular Systems Using a Simple Two-State Anisotropic Network Model.* Plos Computational Biology, 2014. **10**(4).

125. Enosh, A., et al., *Generation, comparison, and merging of pathways between protein conformations: Gating in K-channels.* Biophysical Journal, 2008. **95**(8): p. 3850-3860.

126. Kim, M.K., R.L. Jernigan, and G.S. Chirikjian, *Efficient generation of feasible pathways for protein conformational transitions.* Biophys J, 2002. **83**(3): p. 1620-30.

127. Lei, M., et al., *Sampling protein conformations and pathways.* J Comput Chem, 2004. **25**(9): p. 1133-48.

128. Sfriso, P., et al., *Exploration of conformational transition pathways from coarse-grained simulations.* Bioinformatics, 2013. **29**(16): p. 1980-6.

129. Flores, S., et al., *The database of macromolecular motions: new features added at the decade mark.* Nucleic Acids Research, 2006. **34**: p. D296-D301.

130. Wagner, A., *Arrival of the fittest : solving evolution's greatest puzzle*. 2014, New York: Penguin Group. viii, 291 pages.

131. Bissantz, C., B. Kuhn, and M. Stahl, *A medicinal chemist's guide to molecular interactions.* J Med Chem, 2010. **53**(14): p. 5061-84.

132. McDonald, I.K. and J.M. Thornton, *Satisfying hydrogen bonding potential in proteins.* J Mol Biol, 1994. **238**(5): p. 777-93.

133. Berndt, K.D., P. Guntert, and K. Wuthrich, *Nuclear magnetic resonance solution structure of dendrotoxin K from the venom of Dendroaspis polylepis polylepis.* J Mol Biol, 1993. **234**(3): p. 735-50.

134. Baker, E.N. and R.E. Hubbard, *Hydrogen bonding in globular proteins.* Prog Biophys Mol Biol, 1984. **44**(2): p. 97-179.

135. Levy, Y. and J.N. Onuchic, *Water mediation in protein folding and molecular recognition.* Annu Rev Biophys Biomol Struct, 2006. **35**: p. 389-415.

136. Baron, R., P. Setny, and J.A. McCammon, *Water in cavity-ligand recognition.* J Am Chem Soc, 2010. **132**(34): p. 12091-7.

תקציר

היכולת של חלבון לזהות ולקשור את הליגנד (או הליגנדים) שלו הינה חיונית למרבית הפונקציות הביולוגיות, ולעיתים קרובות כרוכה בשינויי מבנה. בחלק הראשון של עבודת הדוקטורט שלי פיתחתי את ConTemplate: מתודולוגיה הממומשת ככלי אינטרנטי המציע אוסף מבנים ידועים ומודלים המייצגים חלבון-מטרה בעל מבנה ידוע, במגוון קונפורמציות, בהתבסס על חלבונים בעלי מבנה דומה, ועל שינויי המבנה שהם עוברים. הראיתי כי למרבית החלבונים בעלי המבנה הידוע (המאוגדים יחד בבסיס הנתונים PDB) יש יותר ממבנה ניסיוני אחד, וכי חלבונים הדומים בצורתם בקונפורמציה אחת בסבירות גבוהה יעברו שינויי מבנה דומים. ConTemplate משתמש בממצאים הללו, ובהצגות מפושטות של מבני חלבונים, בכדי לחפש באופן יעיל חלבונים בעלי מבנה דומה לחלבון המטרה. בנוסף, השרת מציע תצוגה נוחה של רשת המכילה את חלבון-המטרה, המודלים המוצעים והמרחקים ביניהם, באופן המאפשר למשתמש לזהות נתיבי-מעבר בין קונפורמציות שונות. מאמר המתאר את העבודה יצא לאור בשנת 2015. בחלקה השני של עבודת הדוקטורט שלי השתמשתי ב-ConTemplate בכדי ללמוד את מרחב הקונפורמציות של CueR, חלבון האחראי על ויסות כמות הנחושת בתא. עבודה זו נעשתה בשיתוף פעולה עם המעבדה של דר' שרון רוטשטיין (אוניברסיטת בר-אילן). CueR הינו חלבון ממשפחת-העל של חלבוני metalloregulator, האחראים לויסות הריכוז של מתכות שונות בתא. בעקבות קישור נחושת, החלבון מפעיל את תהליך השעתוק של שני metalloregulators אחרים, האחראים לפינוי נחושת אל מחוץ לתא. מנגנון ההפעלה והכיבוי של החלבון לא היה ידוע במלואו, אך מחקרים קודמים הציעו כי הוא כרוך בשינויי קונפורמציה של CueR. שרון ביצעה ניסויי EPR, בהם נמדדו המרחקים בין איזורים שונים בחלבון. תוצאות המדידה שלה הראו כי מגוון הקונפורמציות הידוע ל-CueR לא מתאר את טווח התנועה של החלבון במלואו. אני השתמשתי ב-ConTemplate בכדי למדל את החלבון בקונפורמציות נוספות, ובדיקות נוספות העלו כי חלקן מציעות מבנים התואמים את תוצאות הניסויים של שרון. השתמשנו בתוצאות הללו ובמודלים בכדי להציע מנגנון לאקטיבציה ולפעולת החלבון. המאמר המתאר את העבודה התפרסם בשנת 2017.

שני החלקים הראשונים של עבודת הדוקטורט שלי עסקו בשינויי מבנה בחלבון, שכאמור לעיתים קרובות הינם תגובה לקישור ליגנד. בחלקה השלישי של העבודה, התמקדתי בשאלה כיצד התפתחה היכולת של חלבונים לזהות ולקשור את הליגנד שלהם. לשם נוחות התמקדתי בתבניות הקישור של אדנין, בעיקר כחלק ממולקולות גדולות יותר (כדוגמת ATP). פיתחתי כלי חישובי (ComBind) המזהה מהם הליגנדים ב-PDB המכילים אדנין. ComBind שולף קומפלקסים המכילים את אותם הליגנדים, ומבצע סופרפוזיציה (עימוד מבני) שלהם על סמך

האדנין שבליגנד. לאחר מכן, ComBind משתמש בכלי חיצוני בכדי לזהות קשרי מימן בין האדנין לסביבתו. יצרתי אוסף גדול של קומפלקסים המכילים אדנין, והשוויתי את אתרי הקישור שלהם. גיליתי כי לעיתים קרובות הקישור מתווך על ידי רצפי חומצות אמינו חוזרניים, שנפוצים בחלבונים וקרויים themes. Themes מסוימים, וקומבינציות שלהם, נמצאים בשימוש תדיר באתרי קישור של אדנין. זיהיתי חלבונים נוספים ב-PDB המכילים את אותם ה-themes, ואני מציעה כי, בסבירות גבוהה, אותם החלבונים קושרים בעצמם ליגנדים המכילים אדנין. תוצאות המחקר מדגימות את האופי האופורטוניסטי של האבולוציה, ומציעות כי קישור אדנין הינה פונקציה שהתפתחה בחלבונים יותר מפעם אחת במהלך האבולוציה.

עבודה זו נעשתה בהדרכתו של

פרופ׳ ניר בן-טל

Tel Aviv University

אוניברסיטת תל אביב

שינויי קונפורמציה וזיהוי ליגנד בחלבונים: נקודת-מבט אבולוציונית

חיבור לשם קבלת התואר "דוקטור לפילוסופיה"

מאת איה נרונסקי

הוגש לסנאט אוניברסיטת תל אביב

יולי 2019